Universitat Autònoma de Barcelona

Search for subsolar mass black holes in LIGO/Virgo using O3 data and the implementation of Machine Learning algorithms in the identification of Compact Binary Coalescence events

Alexis Menéndez Vázquez

Universidad Autónoma de Barcelona (UAB)

Department of Physics
Institut de Física d'Altes Energies (IFAE)

Search for subsolar mass black holes in LIGO/Virgo using O3 data and the implementation of Machine Learning algorithms in the identification of Compact Binary Coalescence events

PhD thesis by

Alexis Menéndez Vázquez

Director:
Prof. Dr., Mario Martínez-Pérez

Tutor:
Prof. Dr., Eduard Massó

Year 2022

Acknowledgements

It has been ten years since I started studying physics and mathematics and fifteen since I decided that physics was going to be my future career. It has been a long journey up to this point and, though sometimes my will to continue faltered, I am glad to have reached the end of the road. Nevertheless, this was not a lonely adventure. Along the way I was helped and guided by many people and I wish to express my gratitude to all of them. This wouldn't have been possible without your contribution.

First, I would like to thank my director, Dr. Mario Martínez-Pérez, for all his help throughout the PhD and the writing of this thesis. Thanks a lot for your patience and corrections that have greatly improved the quality of this work. I would also like to thank Dr. Lluisa M. Mir for always being there to give a helping hand in times of need. Finally, all the people at IFAE that were always willing to contribute and collaborate. Thanks a lot for the last few years.

Second, all the people from the LIGO-Virgo-Kagra collaboration, from which I have learnt so much about the interferometer, the analysis or interpretations. I wish to express my thanks to all the members of the PyCBC team. Special thanks to Dr. Bhooshan Gadre that allowed me to contribute to the subsolar mass search, and Dr. Gareth Cabourn Davies and Dr. Ian Harry that were always available to answer my questions. It was a pleasure to work with all of you and I wish you the best in your future endeavors.

Third, Dr. Alba Romero-Rodríguez and Dr. Christos Karathanasis, we started this together and I am glad to say that we saw the end together. One of the most fond memories I will always have is the close-knit family we created and I wish these bonds will remain forever. Alba, we met at the

beginning of the PhD and since then we have been inseparable. Even though sometimes we had our disagreements, we were always able to solve them and create an even stronger friendship. I will always admire your hard-working attitude and I am completely sure that you will go far in whatever you decide to do. Never doubt that. Christos, I will always remember fondly our long discussions about the best way to attack the O3 data, the Greek lessons and the uncertainty in our futures. I don't know what the future has in store for us, but I have no doubt that we will find a way to get the best outcome. An era has come to an end and the three of us have seen it end together. Thank you.

Me gustaría agradecer personalmente a Óscar y Rosabel por toda su ayuda durante el doctorado y estos últimos meses de escritura. Óscar, gracias por todas esas conversaciones que me permitían desconectar un poco en esos momentos de procrastinación. He aprendido mucho de tus historias y siempre recordaré esos días donde hablamos de todo y nada al mismo tiempo. Rosa, gracias por todos los consejos y ayuda en estos años. Has sido siempre la voz de la paciencia y empatía. Ha sido un placer haberte conocido.

A Ignacio Amores Sesar, desde que nos conocimos durante el bachillerato hemos mantenido una amistad que solo ha ido creciendo durante los años. Siempre recordaré con cariño todas nuestras aventuras durante los 5 años de carrera. Muchas gracias por todo tu apoyo durante estos años de doctorado y espero que, aunque estemos en distintos países, esta amistad se mantenga fuerte para el resto de mi vida.

A Sonia y Elena, aun recuerdo cuando nos conocimos, una simple casualidad de la cual ha surgido una gran amistad. Os considero dos de mis mejores amigas y vuestra ayuda y apoyo durante estos años ha sido inestimable. Muchísimas gracias por todo y nunca dudéis lo mucho que os aprecio. Siempre dicen que la verdadera amistad dura toda la vida, yo haré todo lo posible para que sea cierto.

A mis hermanos, *Mr* Ángel y Omar, habéis estado a mi lado desde el principio ¿Qué más se puede decir que no se haya dicho ya en estos años?. Vuestro apoyo ha sido fundamental y saber que siempre he podido contar con vosotros me ha ayudado en todos los momentos más complicados de estos años. Os quiero mucho y no os cambiaría por nada, muchas gracias por todo.

Por último, me gustaría agradecer a mi madre, Maria del Carmen Vázquez Antuña, sin la cual nada de esto habría sido remotamente posible. No hay suficientes palabras para expresar cuanto te agradezco todo lo que has hecho por mí. Espero que estas sean suficientes: Te quiero mucho y estoy orgulloso de que seas mi madre, muchísimas gracias por todo.

Finally, I would like to thank all the people that were not mentioned and helped me during the Ph.D. Thank you very much for everything.

Contents

Introduction

Gravitational waves have been theorized by General relativity since 1916 but it was not until September 14th 2015 that a confirmed detection was made by the LIGO interferometers. Since then, about 90 gravitational wave events have been detected by the LIGO and Virgo interferometers. This was made possible not only due to the increase in sensitivity of the interferometer network, but also due to our rising understanding of the population sources and noise distribution.

Gravitational waves provide a new window into the universe and our understanding of black holes. In particular, they allow us to probe different mass regions, as for example, the subsolar mass range. Even though no event has been found in this range, a possible detection would imply potential new physics and new formation channels.

The confirmed events have been produced by the coalescence of two compact objects (Black holes or neutron stars) and were detected by means of dedicated search pipelines. These pipelines use a technique, known as matched filter, that searches for the signal inside the data by correlating the data with a set of theoretical waveforms describing the gravitational waves. The efficiency of this method is limited by the computational resources available, as it requires a dense sampling of the searched parameter space, and the understanding of the noise in the detector. Nevertheless, the matched filter technique remains as a robust approach for the detection of gravitational waves.

The results of this thesis are divided into two parts: First, we present the results of the subsolar mass search carried out within the collaboration using the third observation run and matched filter. We will focus on two possible

formation channels of subsolar black holes, primordial black holes and dark matter black holes, and provide limits to their abundance in the universe. Second, we design a machine learning technique for the detection of gravitational waves from the coalescence of two compact objects. We cover a broad range of the parameter space and we search for gravitational waves in the second and third observation runs, where the results are also compared with the official catalog events.

The thesis is organized in the following way: In Chapter 1, the generation and emission of gravitational waves are explained. In Chapter 2, the instrument used for the detection of the aforementioned waves is described along with a comprehensive list of noise sources. In Chapter 3, the noise is characterized and all the elements needed to construct a search for CBC events are covered. In Chapter 4, the previous elements are used to search for gravitational waves produced by the collision of two compact objects with at least one subsolar component. In Chapter 5, a machine learning implementation for the prompt detection of gravitational waves is discussed and tested in data. Finally, the last chapter is dedicated to conclusions.

The work in this thesis led to the following publications in journals.

- Fiori I, Paoletti F, Tringali MC, Janssens K, Karathanasis C, Menéndez-Vázquez A, Romero-Rodríguez A, Sugimoto R, Washimi T, Boschi V, Chiummo A, Cieślar M, De Rosa R, De Rossi C, Di Renzo F, Nardecchia I, Pasqualetti A, Patricelli B, Ruggi P, Singh N. The Hunt for Environmental Noise in Virgo during the Third Observing Run. Galaxies. 2020; 8(4):82. `https://doi.org/10.3390/galaxies8040082`

- Ballester, O., et al (2022). Measurement of the stray light in the Advanced Virgo input mode cleaner cavity using an instrumented baffle. Class. Quant. Grav., 39(11), 115011. `https://doi.org/10.1088/1361-6382/ac6a9d`

- Andrés-Carcasona, M., et al (2022). An instrumented baffle for Advanced Virgo Input Mode Cleaner End Mirror. Submitted for publication to Phys. Rev. D. `https://arxiv.org/abs/2210.16313`

- Menéndez-Vázquez, A., Kolstein, M., Martínez, M. and Mir, Ll. M. (2021). Searches for compact binary coalescence events using neural

networks in the LIGO/Virgo second observation period. Physical Review D, 103(6). https://doi.org/10.1103/physrevd.103.062004

- Menéndez-Vázquez, A., Kolstein, M., Martínez, M. and Mir, Ll. M. Mir. Searches for Compact Binary Coalescence Events using Neural Networks in LIGO/Virgo Third Observation Period. In preparation.

- Abbott, R., et al (2022). Search for Subsolar-Mass Binaries in the First Half of Advanced LIGO's and Advanced Virgo's Third Observing Run. Physical Review Letters, 129(6). https://doi.org/10.1103/physrevlett.129.061104

- Abbott, R., et al. Search for subsolar-mass black hole binaries in the second part of Advanced LIGO's and Advanced Virgo's third observing run. In preparation.

- Cirone, A., Fiori, I., Paoletti, F., Perez, M. M., Rodríguez, A. R., Swinkels, B. L., Vazquez, A. M., Gemme, G., and Chincarini, A. (2019). Investigation of magnetic noise in advanced Virgo. Classical and Quantum Gravity, 36(22), 225004. https://doi.org/10.1088/1361-6382/ab4974

Chapter 1

Introduction to gravitational waves

In this chapter, we discuss the elements needed to understand the theoretical background describing gravitational waves (GWs) [1]. We first introduce some notions of general relativity [2] and the Einstein field equations. From these equations we derive the weak-field Einstein equations by assuming that we are interested in a region with a weak gravitational field. Finally, we deduce the existence of GWs by solving the equations outside a gravitational field source.

Next, we define a convenient gauge to describe these waves, the TT gauge, and how the GW can always be projected onto it. This gauge will be used to describe the emission of GWs from a generic source. Furthermore, we show that, at leading order, the emission is given by a quadrupole moment contribution. We will particularize this result in the case of a GW produced by the coalescence of two compact objects (a neutron star or black hole), and see how the dynamics of the system are affected by the emission of these waves.

1.1 Einstein's field equations

General relativity is a classical field theory of gravity [2] that states that the effect of gravity can be understood as the warping of spacetime. In other words, gravity is geometry and not a product of an attractive force as was

previously thought. It provides a unification between special relativity (SR) and Newton's law of gravitation.

The first building block of general relativity is known as "Einstein's Equivalence Principle", which states the following: *Any local physical experiment not involving gravity will have the same result if performed in a freely falling inertial frame as if it were performed in the flat spacetime of special relativity.*
In simpler terms, this means that from inside an isolated system it's not possible to distinguish between the physical effects produced by gravity and those produced by the acceleration of the system. Furthermore, this also implies that light has to be affected by gravity or, otherwise, the observer would be able to distinguish between the two situations by following the path of the light beam.

This notion motivated the idea of a geometrical interpretation of gravity, and was formalized by describing the spacetime as a four dimensional manifold [3, 4] with a metric. Gravity is produced by the curvature of said spacetime by a mass. Therefore, we need to start by describing what a metric tensor is [5].

Definition 1. *Let M be a smooth manifold, a metric at a point $p \in M$ is a $\binom{0}{2}$ tensor, g_p, from the space of all the tangent vectors at point p to $\mathbb{R}$ that varies smoothly with p and satisfies the following conditions.*

- *Bilineal: For $\alpha, \beta, \tau \in \mathbb{R}$ and A_p, B_p, C_p contained in the tangent space of the point p.*

$$g_p(\alpha A_p + \beta B_p, C_p) = \alpha g_p(A_p, C_p) + \beta g_p(B_p, C_p)$$

$$g_p(A_p, \beta B_p + \tau C_p) = \beta g_p(A_p, B_p) + \tau g_p(A_p, C_p)$$

- *Symmetric: $g_p(A_p, B_p) = g_p(B_p, A_p)$ for every element of the tangent space of p.*

- *Non-degenerate: For every $A_P \neq 0$ there is at least one B_p in the tangent space such that $g_p(A_p, B_p) \neq 0$.*

A metric tensor can be understood as the generalization of the notion of metric from metric spaces. The inner dot product from Euclidean spaces

is an example of a metric tensor. The metric is fundamental to understand general relativity as it provides a general framework for non-euclidean spaces, including curved spaces.

One of the most important metrics is the Minkowski metric, $\eta_{\mu\nu}$, defined in a four dimensional manifold that covers the whole $\mathbb{R}^4$. It corresponds to a flat spacetime with coordinates $(x^0, x^1, x^2, x^3) = (t, x, y, z)$ and it can be represented as a 4x4 matrix as in Eq. 1.1:

$$\eta = \begin{pmatrix} -c & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.1}$$

where c is the speed of light. If we take only the three spatial components (x, y, z), this metric acts as the inner dot product in an euclidean space.

The next concept we need to introduce to understand Einstein's field equations is the Stress-energy tensor, $T^{\alpha\beta}$, defined as follows [2]:

Definition 2. *The stress-energy tensor, $T^{\alpha\beta}$, is a tensor of order two defined as the flux of the α momentum component across a surface of constant x^β.*

In general relativity, the "α momentum component" is a component of the four momentum. The elements of this tensor can be interpreted as follows:

- T^{00}: Flux of zero momentum component, energy, across surfaces of constant time. This is the energy density.
- T^{0i}: Flux of energy across surfaces with x^i constant.
- T^{i0}: Flux of i momentum component across surfaces of constant t. This is the i momentum density.
- T^{ij}: Flux of i momentum component across surfaces with x^j constant. In Newtonian theory this is called the stress tensor.

It can be proved that the stress-energy tensor is symmetric and that conversation laws can be derived from it. In fact, the conservation of four momentum can be expressed as in Eq. 1.2.

In general relativity, this condition is generalized by using the covariant derivative, D_ν.

$$\begin{aligned} \partial_\nu T^{\mu\nu} &= 0 \\ D_\nu T^{\mu\nu} &= 0. \end{aligned} \tag{1.2}$$

With these two concepts we can create a covariant theory of gravity in which the source of the gravitational field, given by the stress-energy tensor, determines the metric. That is, we ask for solutions to Eq. 1.3:

$$O(g) = kT, \tag{1.3}$$

where $O(g)$ is a differential operator acting on the metric. This choice is inspired from the Newtonian case in which the gravitational field, ϕ, is given by the differential equation $\nabla^2\phi = 4\pi G\rho$. Since T is a $\binom{0}{2}$ tensor, the operator has to produce a tensor of the same rank. It can be shown that the general solution is given by Eq. 1.4:

$$O^{\alpha\beta} = R^{\alpha\beta} + \mu g^{\alpha\beta} R + \Lambda g^{\alpha\beta}, \tag{1.4}$$

where $R^{\alpha\beta}$ is known as the Ricci tensor, which represents how much the space differs from an euclidean space (and is defined as a contraction of the Riemann tensor on the first and third indices), $g^{\alpha\beta}$ are the contravariant components of the metric, μ and Λ are constants and R is the Ricci scalar. Using the equivalence principle, we can enforce local conservation of energy and momentum, which are conserved as stated by Eq. 1.2.

This implies that $D_\beta O^{\alpha\beta} = 0$ should hold for whatever metric tensor we are using. Hence, given that $D_\beta g^{\alpha\beta} = 0$ due to the definition of covariant derivative, we obtain Eq. 1.5:

$$D_\beta(R^{\alpha\beta} + \mu g^{\alpha\beta} R) = 0, \tag{1.5}$$

which is true for any metric as long as $\mu = -\frac{1}{2}$. If we substitute this value in Eq. 1.4, we get Eq. 1.6:

$$R^{\alpha\beta} - \frac{1}{2} g^{\alpha\beta} R + \Lambda g^{\alpha\beta} = kT^{\alpha\beta}, \tag{1.6}$$

the Einstein's field equations. These equations can also be written by introducing a symmetric tensor called the Einstein tensor, $G^{\alpha\beta}$, as in Eq. 1.7:

$$G^{\alpha\beta} + \Lambda g^{\alpha\beta} = kT^{\alpha\beta} \tag{1.7}$$
$$G^{\alpha\beta} = R^{\alpha\beta} - \frac{1}{2} g^{\alpha\beta} R = G^{\beta\alpha}.$$

Finally, by lowering and raising indexes, these equations can be expressed in covariant form, as in Eq. 1.8:

$$G_{\alpha\beta} + \Lambda g_{\alpha\beta} = kT_{\alpha\beta}. \tag{1.8}$$

Newton's law of gravitation can be recovered from this expression by assuming the following:

- Small gravitational field.
- The speed of the objects, v, is much smaller than the speed of light $|v| \ll c$.
- Static gravitational field.

This is known as the Newtonian limit and it allows us to estimate the value of 'k': $k = \frac{8\pi G}{c^4}$ where G is the gravitational constant and c is the speed of light. However, Λ will remain as a free parameter.

A simple solution to Einstein's equations is given by the curved spacetime defined by the following line element in spherical coordinates:

$$ds^2 = -\left(1 - \frac{2GM}{rc^2}\right) c^2 dt^2 + \left(1 - \frac{2GM}{rc^2}\right)^{-1} dr^2 + r^2(d\theta^2 + sin^2(\theta) d\phi^2), \tag{1.9}$$

which is known as the Schwarzschild metric [6] and describes the gravitational field of a star of mass M with zero electric charge and angular momentum. Furthermore, at $r \to \infty$ we recover the Minkowski metric. This metric provides the most general spherically symmetric solutions for Einstein's equations and, in particular, is used to describe a type of black hole, known as Schwarzschild black hole, that is defined only by the mass.

A black hole is a region of spacetime curved by a compact object with a very high mass that generates a gravitational field strong enough such that nothing can escape from it (including light). In a Schwarzschild black hole, the size of this region is given by the Schwarzschild radius:

$$R_s = \frac{2GM}{c^2}. \tag{1.10}$$

This definition is also valid for stars but, in that case, it occurs that $R_{Star} \gg R_s$ and light can escape its gravitational pull and be observed on Earth.

Another exact solution to Einstein's equations is given by Kerr metric [7] which describes a zero electric charge black hole rotating along the axis of angular momentum. This solution is axially symmetric along this axis and the black holes are described by two parameters, the mass, M, and the angular momentum, J. The momentum is generally given in terms of the dimensionless spin, a, defined as:

$$a = \frac{cJ}{GM^2}, \tag{1.11}$$

which is always lower than one for a black hole. This metric is given by the following line element:

$$\begin{aligned} ds^2 = &- \left(1 - \frac{R_s r}{\rho^2}\right) c^2 dt^2 - \frac{2GMR_s r \cdot a \cdot sin^2(\theta)}{c^2 \rho^2} c dt d\phi + \frac{\rho^2}{\Delta} dr^2 \\ &+ \left(r^2 + \left(\frac{GM}{c^2} a\right)^2 + \frac{G^2 M^2 R_s r a^2}{c^4 \rho^2} sin^2(\theta)\right) sin^2(\theta) d\phi^2 + \rho^2 d\theta^2, \end{aligned} \tag{1.12}$$

where ρ and Δ are defined as follows:

$$\begin{aligned} \rho^2 &= r^2 + \left(\frac{GM}{c^2} a\right)^2 cos^2(\theta) \\ \Delta &= r^2 - R_s r + \left(\frac{GM}{c^2} a\right)^2. \end{aligned} \tag{1.13}$$

In this case, ϕ corresponds to the angle around the axis of symmetry whereas r and θ are similar to the previous case. This metric reduces to a Schwarzschild

metric when $a = 0$. Furthermore, it is asymptotically flat, given that in the limit $r \to \infty$ we recover the Minkowski metric. A black hole described by this metric is called Kerr black hole.

A known formation channel for black holes is the gravitational collapse of stars [7] in the late stage of their evolution. These black holes are known as "Stellar" black holes, can have a mass up to hundreds solar masses and can further grow by absorbing matter from its surroundings.

The minimum mass of stellar black holes is roughly estimated to be around three solar masses. When the mass of the system is below three solar masses, neutron degeneracy (Pauli Exclusion Principle applied to neutrons) generates an effective pressure high enough to prevent gravitational collapse of the original star and, a neutron star (or white dwarf for low mass stars), is formed. Therefore, as will be explained in Chapter 4, detection of subsolar mass black holes would indicate new formation mechanisms or new physics.

1.2 Gravitational waves

In this section, we use the field equations shown in Eq. 1.7 to derive the expression for GWs. We will also discuss the approximations used and convenient gauge choices to describe the GWs.

1.2.1 Weak-field Einstein equations

The Einstein's field equations in Eq. 1.7 are not easily solved and, for this reason, approximations are made to work with them. One of the most natural approximation follows from answering the question: How would these equations look away from large masses? This introduces the notion of a "weak" gravitational field. We define 'weak' as a gravitational field such that the spacetime is almost flat. This means that the metric should take a form similar to the Minkowski metric (as in Eq. 1.1) with some small perturbations. Explicitly, this takes the shape seen in Eq. 1.14:

$$\begin{aligned} g_{\alpha\beta} &= \eta_{\alpha\beta} + h_{\alpha\beta} \\ |h_{\alpha\beta}| &\ll 1, \end{aligned} \tag{1.14}$$

which is the simplest way to express a near-flat spacetime metric. From special relativity, we know that the metric tensor transforms as in Eq. 1.15 under Lorentz transformations, $\Lambda^{\mu}{}_{\bar{\alpha}}$:

$$\begin{aligned} g_{\bar{\alpha}\bar{\beta}} &= \Lambda^{\mu}{}_{\bar{\alpha}}\Lambda^{\nu}{}_{\bar{\beta}}g_{\mu\nu} \\ \eta_{\bar{\alpha}\bar{\beta}} &= \Lambda^{\mu}{}_{\bar{\alpha}}\Lambda^{\nu}{}_{\bar{\beta}}\eta_{\mu\nu} \end{aligned} \tag{1.15}$$

Taking this into account in Eq. 1.14, we have that it transforms under Lorentz transformations as in Eq. 1.16:

$$\begin{aligned} g_{\bar{\alpha}\bar{\beta}} &= \eta_{\bar{\alpha}\bar{\beta}} + h_{\bar{\alpha}\bar{\beta}} \\ h_{\bar{\alpha}\bar{\beta}} &\coloneqq \Lambda^{\mu}{}_{\bar{\alpha}}\Lambda^{\nu}{}_{\bar{\beta}}h_{\mu\nu}, \end{aligned} \tag{1.16}$$

in which we define $h_{\bar{\alpha}\bar{\beta}}$ in such a way that $h_{\mu\nu}$ transforms in the same way as the metric tensors. However, it shall be noted that we have to be very careful with boosts as the first assumption we made was that we had small perturbations around the Minkowski metric, that is, the condition $|h_{\alpha\beta}| \ll 1$ should always hold. Furthermore, the transformation of coordinates given by Eq. 1.17 leaves Eq. 1.14 unchanged:

$$(x^{\alpha})' = x^{\alpha} + \xi^{\alpha}(x), \tag{1.17}$$

where ξ is a vector with components given by the position, x, such that $|\partial_{\beta}\xi^{\alpha}| \ll 1$. This is called a gauge transformation and corresponds to an infinitesimal local transformation. Under this transformation we have that $h_{\alpha\beta}$ changes as in Eq. 1.18:

$$h_{\alpha\beta} \to h_{\alpha\beta} - (\partial_{\beta}\xi_{\alpha} - \partial_{\alpha}\xi_{\beta}), \tag{1.18}$$

where we have defined $\xi_{\alpha} \coloneqq \eta_{\alpha\beta}\xi^{\beta}$. It should be noted that using only transformations such that $|\partial_{\beta}\xi^{\alpha}| \ll 1$ is due to the metric being the Minkowski metric. Under a different metric tensor, we would also need to ask for $|\xi^{\alpha}| \ll 1$.

To summarize, under the hypothesis of an almost flat space, we have that linearized theory is invariant under constant global translations, Lorentz transformation and infinitesimal local transformations.

Before introducing the linearized equations, it's convenient to define the trace reverse of $h_{\alpha\beta}$ as in Eq. 1.19:

$$\begin{aligned} \bar{h}^{\alpha\beta} &\coloneqq h^{\alpha\beta} - \frac{1}{2}\eta^{\alpha\beta}h \\ h &\coloneqq h^{\alpha}{}_{\alpha} = \eta^{\alpha\beta}h_{\beta\alpha}, \end{aligned} \tag{1.19}$$

where h is the trace. This new definition gets its name from the fact that the trace of $\bar{h}^{\alpha\beta}$ is $\bar{h} = \bar{h}^{\alpha}{}_{\alpha} = -h$. Furthermore, this can be used to define $h^{\alpha\beta}$ as the trace reverse of $\bar{h}^{\alpha\beta}$:

$$h^{\alpha\beta} = \bar{h}^{\alpha\beta} - \frac{1}{2}\eta^{\alpha\beta}\bar{h}. \tag{1.20}$$

This definition can be introduced into the Riemann tensor and Einstein tensor from Eq. 1.7 to get the linearized version of the Einstein tensor, as seen in Eq. 1.21:

$$G_{\alpha\beta} = -\frac{1}{2}[\partial^{\mu}\partial_{\mu}\bar{h}_{\alpha\beta} + \eta_{\alpha\beta}\partial^{\mu}\partial^{\nu}\bar{h}_{\mu\nu} - \partial^{\mu}\partial_{\beta}\bar{h}_{\alpha\mu} - \partial^{\mu}\partial_{\alpha}\bar{h}_{\beta\mu}], \tag{1.21}$$

where non linear terms of $h_{\alpha\beta}$ are ignored. This expression is very complex but it can be simplified by a clever choice of the gauge. From the transformations given by Eq. 1.18, we have gauge freedom that can be used for that purpose. Imposing what is known as the Lorenz gauge (sometimes called harmonic gauge, De Donder gauge or Hilbert gauge) seen in Eq. 1.22:

$$\partial_{\nu}\bar{h}^{\mu\nu} = 0, \tag{1.22}$$

we have that Eq. 1.21 reduces to Eq. 1.23.

$$\begin{aligned} G_{\alpha\beta} &= -\frac{1}{2}\partial^{\mu}\partial_{\mu}\bar{h}_{\alpha\beta} = -\frac{1}{2}\Box\bar{h}_{\alpha\beta} \\ \Box &= \left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2\right), \end{aligned} \tag{1.23}$$

where $\Box$ is known as the D'Alembertian or wave operator. This convenient choice of gauge exists thanks to the d'Alembertian operator being invertible. This property allows us to always be able to find a gauge transformation such that this condition holds [1, 2].

Finally, we can substitute Eq. 1.23 in Eq. 1.8 to get the weak field Einstein equations:

$$\Box\bar{h}_{\alpha\beta} = -\frac{16\pi G}{c^4}T_{\alpha\beta}, \tag{1.24}$$

where we have assumed that $\Lambda = 0$. In the Newtonian limit, the Newtonian gravitational field can be recovered from these equations.

The expression for GWs can be derived from the weak-field Einstein equations using that outside the source, in vacuum, the stress-energy tensor is zero: $T^{\alpha\beta} = 0$. This means that Eq. 1.24 takes the form seen in Eq. 1.25:

$$\Box\bar{h}_{\alpha\beta} = \left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2\right)\bar{h}_{\alpha\beta} = 0. \tag{1.25}$$

The Eq. 1.25 has the shape of a wave equation with waves propagating at the speed of light, c. This is known as the three dimensional wave equation and its solutions are the GWs. The magnitude of the GW signal is expected to be a very small number.

1.2.2 The transverse-traceless gauge

The solutions to the wave Eq. 1.25 are given by plane waves but they can be further simplified. In the first place, we should notice that coordinate transformations of the form $x^\mu \to x^\mu + \xi^\mu$ with $\Box\xi_\mu = 0$ are compatible with the Lorenz Gauge seen in Eq. 1.22. This is due to the trace reverse of $h_{\alpha\beta}$ transforming as in Eq. 1.26 under these transformations:

$$\begin{aligned} \bar{h}_{\mu\nu} \to & \bar{h}'_{\mu\nu} = \bar{h}_{\mu\nu} - (\partial_\mu\xi_\nu + \partial_n u\xi_\mu - \eta_{\mu\nu}\partial_\rho\xi^\rho) = \bar{h}_{\mu\nu} - \xi_{\mu\nu} \\ \partial^\mu\bar{h}_{\mu\nu} \to & (\partial^\mu\bar{h}_{\mu\nu})' = \partial^\nu\bar{h}_{\mu\nu} - \Box\xi_\mu. \end{aligned} \tag{1.26}$$

Therefore, given that $\xi_{\mu\nu}$ depends on four arbitrary functions, we can carefully choose them so that we add some constraints to $h_{\mu\nu}$. The gauge defined by these choices is known as the Transverse-Traceless (TT) gauge and is given by the conditions in Eq. 1.27:

$$\begin{aligned} h^{0\mu} &= 0 \\ h^i{}_i &= 0 \\ \partial^j h_{ij} &= 0. \end{aligned} \tag{1.27}$$

Components in this gauge will be indicated as h_{ij}^{TT}. By imposing these two gauges, we have effectively reduced the degrees of freedom of $h_{\alpha\beta}$ to two. However, it's important to remark that this derivation is based on the assumption that $T^{\alpha\beta} = 0$. Inside a source this is not true and hence the TT gauge cannot be used.

Under this gauge, the general solution to the wave Eq. 1.25 is a single plane wave propagating along the z-axis and is given by Eq. 1.28:

$$h_{ab}^{TT}(t,z) = \begin{pmatrix} h_+ & h_x \\ h_x & -h_+ \end{pmatrix} cos[w(t - \frac{z}{c})], \tag{1.28}$$

where w is the angular frequency of the wave and a, b are indices in (x, y) plane. The two elements h_+ and h_x are "plus" and "cross" polarization of the wave.

A GW outside the source can be projected into the TT gauge by using the Lambda tensor defined in Eq. 1.29:

$$\Lambda_{ij,kl}(\hat{\mathbf{n}}) = \delta_{ik}\delta_{jl} - \frac{1}{2}\delta_{ij}\delta_{kl} - n_j n_l \delta_{ik} - n_i n_k \delta_{jl} + \frac{1}{2}(n_k n_l \delta_{ij} + n_i n_j \delta_{kl} + n_i n_j n_k n_l), \tag{1.29}$$

where n_i are components of the unit vector, $\hat{\mathbf{n}}$, pointing towards the direction of propagation of the GW. This tensor allows us to define the transverse-traceless part of any symmetric tensor. In particular, the wave can be projected onto the TT gauge as in Eq. 1.30.

$$h_{ij}^{TT} = \Lambda_{ij,kl} h_{kl} \tag{1.30}$$

1.2.3 Reference frames for GW detection

The TT gauge explained in the previous section is a very convenient choice to describe a GW. However, it raises the following questions: How does the GW interact with a detector? Is this also the proper reference frame to use? In this section, we assume a detector is just defined by a set of test masses. A more in-depth explanation of how GWs are detected can be seen in Chapter 2.

The most obvious choice would be to work in the reference frame associated with the TT gauge, the TT frame. The physical meaning of this choice

can be seen by writing the geodesic equation assuming that the test mass is at rest at the proper time $\tau = 0$, as in Eq. 1.31:

$$\left[\frac{d^2x^i}{d\tau^2}\right]_{\tau=0} = -\left[\Gamma^i_{00}(\frac{dx^0}{d\tau})^2\right]_{\tau=0}, \tag{1.31}$$

where Γ^k_{ij} are the Christoffel symbols. Under the assumption of an almost flat space, the metric can be written as in Eq. 1.14 and, if we expand to first order in the perturbation, we have that the symbol can be written as in Eq. 1.32:

$$\Gamma^i_{00} = \frac{1}{2}(2\partial_0 h_{0i} - \partial_i h_{00}), \tag{1.32}$$

which is zero due to the TT gauge choices seen in Eq. 1.27. We conclude that, in the TT frame, if a particle was at rest before interacting with a GW it will remain at rest after it arrives. It is important to remark that this doesn't mean there is no effect, only that we cannot see a change in the coordinates. The reason is that the coordinates will stretch with the GW in such a way that the position of the test mass remains the same.

In this TT frame, the physical effects are seen when we work with proper distances or proper times. In general, if we have two events with a separation given by a vector $\mathbf{L}$, we have that the proper distance can be written as in Eq. 1.33:

$$\begin{aligned} s^2 &= L^2 + h_{ij}(t)L_iL_j \\ s &\approx L + h_{ij}(t)\frac{L_iL_j}{2L}, \end{aligned} \tag{1.33}$$

where we have approximated s to linear order in h_{ij}. We can get the geodesic equation in terms of the proper distance by differentiating twice over time as in Eq. 1.34:

$$\begin{aligned} \ddot{s} &\approx \frac{1}{2}\ddot{h}_{ij}L_j\frac{L_i}{L} = \frac{1}{2}\ddot{h}_{ij}L_jn_i = n_i\ddot{s}_i \\ \ddot{s}_i &\approx \frac{1}{2}\ddot{h}_{ij}L_j = \frac{1}{2}\ddot{h}_{ij}s_j, \end{aligned} \tag{1.34}$$

where we have used $\frac{L_i}{L} = n_i$, a unit vector, and that at first order in h_{ij} the approximation $L_i = s_i$ holds. These equations tell us that, in the TT frame, the proper distance between the two events oscillates due to the interaction

of the test masses with a GW. This is the basic idea behind the interferometers designed to detect GWs, as discussed in Chapter 2.

Unfortunately, though the TT frame provides a simple expression for the GW, it's not a good reference frame for experiments. The more convenient reference frame is what we know as the "detector frame". The main difference with the TT frame is that the position of the test masses changes with respect to our origin when a GW interacts with them. In this frame, the equation of the geodesic deviation, given by the difference between two close-by geodesics, ξ, can be reduced to Eq. 1.35:

$$\ddot{\xi}^i = \frac{1}{2}\ddot{h}^{TT}_{ij}\xi^j. \tag{1.35}$$

It should be noted that the term h^{TT}_{ij} corresponds to the GW in the TT gauge. This term appears due to an invariant tensor in linearized theory. Therefore, we write it in the TT frame in which we have a simpler expression. Furthermore, Eq. 1.35 also tells us that the interaction of a GW with a test mass in the detector frame is given in terms of a Newtonian force.

1.2.4 Gravitational waves emission

In this section, we discuss the mechanisms that generate GWs. We work under the hypothesis needed for the weak-field Einstein Eq. 1.24, that is, a weak gravitational field. Due to the linearity in $h_{\mu\nu}$ we can use Green's functions [8] to get a solution. Furthermore, if we impose an additional boundary condition, known as the Kirchoff-Sommerfeld boundary condition, we can reduce the expression to Eq. 1.36 by using a retarded Green's function.[1]. This condition is imposed so that the only physical solution remains. This solution assumes that the source of the perturbation originates from the past.

$$\bar{h}_{\mu\nu}(t,\mathbf{x}) = \frac{4G}{c^4}\int d^3x' \frac{1}{|\mathbf{x}-\mathbf{x}'|} T_{\mu\nu}(t - \frac{|\mathbf{x}-\mathbf{x}'|}{c}, \mathbf{x}'). \tag{1.36}$$

The energy-stress sensor depends on the retarded time $t_r = t - \frac{|\mathbf{x}-\mathbf{x}'|}{c}$. Outside the source, this solution can now be transformed into the TT gauge by using the Lambda tensor defined in Eq. 1.29:

[1] $G(x-x') = \frac{-1}{4\pi|\mathbf{x}-\mathbf{x}|}\delta(t_r - t')$ where t_r is the retarded time

$$h_{ij}^{TT}(t,\mathbf{x}) = \Lambda_{ij,kl}(\hat{\mathbf{x}})\frac{4G}{c^4}\int d^3x' \frac{1}{|\mathbf{x}-\mathbf{x}'|}T_{kl}(t-\frac{|\mathbf{x}-\mathbf{x}'|}{c},\mathbf{x}'), \tag{1.37}$$

where $\hat{\mathbf{x}}$ is a unitary vector along the direction of propagation. We are also limiting the integral to the spatial components of $T_{\mu\nu}$ due to the relations given by the conservation of 4 momentum in Eq. 1.2. However, solving this integral is not easy and more assumptions are needed. For this reason, we assume a source of radius 'd' such that $|\mathbf{x}| = r \gg d$. In practise, we are interested in the solution at our detector or, in other words, when $r \to \infty$ for a given time. Under these conditions, the term $|\mathbf{x}-\mathbf{x}'|$ can be written as:

$$|\mathbf{x}-\mathbf{x}'| = r - \mathbf{x}'\cdot\hat{\mathbf{x}} + O(\frac{d^2}{r})$$

Substituting this in Eq. 1.37 we get Eq. 1.38:

$$\begin{aligned} h_{ij}^{TT}(t,\mathbf{x}) &= \Lambda_{ij,kl}(\hat{\mathbf{x}})\frac{4G}{c^4}\int d^3x' \frac{1}{r-\mathbf{x}'\cdot\hat{\mathbf{x}}+O(\frac{d^2}{r})}T_{kl}(t-\frac{r-\mathbf{x}'\cdot\hat{\mathbf{x}}+O(\frac{d^2}{r})}{c},\mathbf{x}') \\ &\approx \frac{\Lambda_{ij,kl}(\hat{\mathbf{x}})}{r}\frac{4G}{c^4}\int d^3x' T_{kl}(t-\frac{r}{c}-\frac{\mathbf{x}'\cdot\hat{\mathbf{x}}}{c},\mathbf{x}'), \end{aligned} \tag{1.38}$$

where we ignore high order terms in r given the previous assumption about our detector being very far away from the source. If we write the stress-energy tensor using the Fourier transformation, we can transform the Eq. 1.38 to Eq. 1.39:

$$\begin{aligned} h_{ij}^{TT}(t,\mathbf{x}) &= \frac{\Lambda_{ij,kl}(\hat{\mathbf{x}})}{r}\frac{4G}{c^4}\int_{-\infty}^{\infty}\frac{dw}{2\pi}\tilde{T}_{kl}(w,\frac{w\hat{\mathbf{x}}}{c})e^{-iw(t-\frac{r}{c})} \\ T_{kl}(t-\frac{r}{c}-\frac{\mathbf{x}'\cdot\hat{\mathbf{x}}}{c},\mathbf{x}') &= \int\frac{dw}{2c\pi}\frac{d^3k}{(2\pi)^3}\tilde{T}_{kl}(w,\mathbf{k})e^{-iw(t-\frac{r}{c}+\frac{\mathbf{x}'\hat{\mathbf{x}}}{c})+i\mathbf{k}\cdot\mathbf{x}'} \end{aligned} \tag{1.39}$$

where w is the angular frequency and $\frac{w\hat{\mathbf{x}}}{c} = \mathbf{k}$ is the wave-vector. Let's assume now that the typical frequency inside the source is w_s, then the velocities inside should be of the order $v \sim w_s d$. This means that the frequency of the GW, w, will be of the same order and the reduced wavelenght[2], λ, should be close to $\lambda \sim \frac{c}{v}d$.

[2] $\lambda = \frac{c}{w}$

If we assume that the velocities of the source are much lower than the speed of light, or in other words, non relativistic, we have that $v \ll c$ and $\lambda \gg d$. In this situation, we can work on a macroscopic scale and ignore the finer details from the source. In particular, this means that the leading terms describing the GW radiation are given by the lowest multipole moments.

To derive this, we first need to realize that, under the low-velocity hypothesis, we have $\frac{w}{c}\mathbf{x}' \cdot \hat{\mathbf{x}} \leq \frac{w_s d}{c} \ll 1$. Where we have used that outside the source, $T_{\mu\nu} = 0$ and the integral is different from zero only when $|x'| \leq d$. Therefore, we can do a Taylor expansion of T_{kl} around $\frac{\mathbf{x}' \cdot \hat{\mathbf{x}}}{c}$ as in Eq. 1.40:

$$T_{kl}(t - \frac{r}{c} - \frac{\mathbf{x}' \cdot \hat{\mathbf{x}}}{c}, \mathbf{x}') \approx T_{kl}(t - \frac{r}{c}, \mathbf{x}') + \left[\frac{x'^i \hat{x}^i}{c} \partial_t T_{kl} + \frac{1}{2c} x'^i x'^j \hat{x}^i \hat{x}^j \partial_t^2 T_{kl} + \dots \right]_{t=t-\frac{r}{c}}. \tag{1.40}$$

Introducing this expansion in Eq. 1.38, we have Eq. 1.41, the fundamental equation for the multipole expansion:

$$h_{ij}^{TT}(t, \mathbf{x}) = \frac{\Lambda_{ij,kl}(\hat{\mathbf{x}})}{r} \frac{4G}{c^4} \left[S^{kl} + \frac{1}{c} \hat{x}_m \dot{S}^{kl,m} + \frac{1}{2c^2} \hat{x}_m \hat{x}_p \ddot{S}^{kl,mp} + \dots \right]_{t=t-\frac{r}{c}} \tag{1.41}$$

$$\begin{aligned} S^{kl} &= \int d^3x' T^{kl}(t, \mathbf{x}') \\ S^{kl,m} &= \int d^3x' T^{kl}(t, \mathbf{x}') x'^k \\ S^{kl,mp} &= \int d^3x' T^{kl}(t, \mathbf{x}') x'^k x'^p, \\ &\vdots \end{aligned}$$

where we have recursively defined new functions, the momenta of the stress tensor, and used that we can swap the integral in volume and time derivative. It should be noted that these functions are symmetric with respect to the indexes before or after the ',' but not with respect to swaps between the two types.

If we focus now only on the leading term from Eq. 1.41, we have:

$$h_{ij}^{TT}(t, \mathbf{x}) \approx \frac{\Lambda_{ij,kl}(\hat{\mathbf{x}})}{r} \frac{4G}{c^4} S^{kl}, \tag{1.42}$$

and using that S^{kl} can be written as in Eq. 1.43:

$$S^{ij}(t) = \frac{1}{2}\ddot{M}^{ij}(t) \tag{1.43}$$
$$M^{ij}(t) = \frac{1}{c^2}\int d^3x T^{00}(t,\mathbf{x})x^i x^j,$$

we can write Eq. 1.42 in terms of $M^{ij}(t)$. The correspondence given by Eq. 1.43 can be reached by using that the conservation of energy-momentum, given by Eq. 1.2, is valid in linearized theory. This new definition is a two index symmetric tensor and hence can be decomposed as in Eq. 1.44:

$$M^{ij} = (M^{ij} - \frac{1}{3}\delta^{ij}M_n n) + \frac{1}{3}\delta^{ij}M_{nn}, \tag{1.44}$$

as a traceless term plus the trace. This is a very useful decomposition due to one property of the Lambda tensor: $\Lambda_{ij,kl}\delta^{kl} = 0$. Hence, only the first term survives and will contribute. We can use this traceless term to define the quadrupole moment as in Eq. 1.45:

$$\begin{aligned} Q^{ij} &\coloneqq M^{ij} - \frac{1}{3}\delta^{ij}M_n n = \frac{1}{c^2}\int d^3x T^{00}(t,\mathbf{x})x^i x^j - \frac{1}{3c^2}\delta^{ij}\int d^3x T^{00}(t,\mathbf{x})r^2 \\ &= \int d^3x \frac{1}{c^2}T^{00}(x^i x^j - \frac{1}{3}r^2\delta^{ij}) = \int d^3x \rho(t,\mathbf{x})(x^i x^j - \frac{1}{3}r^2\delta^{ij}), \end{aligned} \tag{1.45}$$

where we have used $\rho(t,\mathbf{x}) = \frac{1}{c^2}T^{00}(t,\mathbf{x})$. Furthermore, in the non relativistic case, this term corresponds to the mass density of the source. Introducing Eq. 1.45 in Eq. 1.42 we have that the leading order term can be written as in Eq. 1.46:

$$h_{ij}^{TT}(t,\mathbf{x}) \approx \frac{\Lambda_{ij,kl}(\hat{\mathbf{x}})}{r}\frac{2G}{c^4}\ddot{Q}_{kl}(t - \frac{r}{c}) = \frac{1}{r}\frac{2G}{c^4}\ddot{Q}_{ij}^{TT}(t - \frac{r}{c}), \tag{1.46}$$

where we have used that the Lambda tensor defines the transverse-traceless part of any symmetric tensor. It should be noted from Eq. 1.46 that the leading term in the expansion is a quadrupole moment. There are no contributions from monopoles and dipoles in GWs emission. This expansion can be generalized using spherical harmonics, Y_{lm}, with $(l = 2, m = 2)$ corresponding to the quadrupole moment derived here. Further terms in the expansion are called "high order modes".

From Eq. 1.46, we can extract the expression for the two polarizations assuming that the GW is propagating along the z-axis. This is seen in Eq. 1.47:

$$h_+ = \frac{1}{r}\frac{G}{c^4}(\ddot{M}_{11} - \ddot{M}_{22}) \qquad (1.47)$$
$$h_- = \frac{1}{r}\frac{2G}{c^4}\ddot{M}_{12}.$$

These equations are valid only if the GW propagates along our z-axis. However, they can be generalized to an arbitrary direction by rotating the tensor M. In particular, assuming that in our reference frame the GW propagates in a direction $\hat{\mathbf{x}} = (sin(\theta)sin(\phi), sin(\theta)cos(\phi), cos(\theta))$ (where θ is the angle between our z-axis and this vector and ϕ is the angle between the projection in the (x, y) plane and the y-axis), we have that the most general formulation for the plus and cross polarization is given by Eq. 1.48:

$$\begin{aligned} h_+(t,\theta,\phi) = \frac{1}{r}\frac{G}{c^4}[&\ddot{M}_{11}(cos^2(\phi) - cos^2(\theta)sin^2(\phi)) \\ &+ \ddot{M}_{22}(sin^2(\phi) - cos^2(\theta)cos^2(\phi))) \\ &- \ddot{M}_{33}sin^2(\theta) - \ddot{M}_{12}sin(2\phi)(1 + cos^2(\theta)) \\ &+ \ddot{M}_{13}sin(\phi)sin(2\theta) + \ddot{M}_{23}cos(\phi)sin(2\theta)] \end{aligned} \qquad (1.48)$$

$$\begin{aligned} h_x(t,\theta,\phi) = \frac{1}{r}\frac{2G}{c^4}[&(\ddot{M}_{11} - \ddot{M}_{22})sin(2\phi)cos(\theta) + 2\ddot{M}_{12}cos(2\phi)cos(\theta) \\ &- 2\ddot{M}_{13}cos(\phi)sin(\theta) + 2\ddot{M}_{23}sin(\theta)sin(\phi)]. \end{aligned} \qquad (1.49)$$

Hence, once the tensor M is known and the assumptions made before are valid, we can estimate the two polarizations produced by a particular source at leading order in arbitrary directions. These equations will be used in Sec. 1.3 to understand the GW radiation produced by the coalescence of two compact objects.

The emission of GWs originates from different sources: Compact Binary Coalescence (CBC), stochastic background, burst, and continuous signals. Up to now, only the ones emitted by CBC events have been detected and the other potential sources are being actively searched for. In this thesis we focus on the search for signals produced by CBC events. An order of magnitude estimation of the magnitude of the GW signal can be obtained by assuming

a binary system of two $30M_\odot$ black holes at 10Mpc. Under these conditions, the emitted gravitational wave will reach a maximum value of $\approx 10^{-21}$. For microscopic objects, the GW turns out to be essentially undetectable.

1.3 Gravitational wave radiation emitted by CBC events

When two objects orbit around each other they emit radiation in the form of GWs. This produces a loss of energy, reducing the radius of the orbit until they collide and coalescence into a new object. If the binary system is composed of two compact objects (a black hole or a neutron star), and depending on the masses, the frequency of the GW emission is high enough that it can be detected using the techniques described in Chapter 2. As already pointed out, CBC events have produced the only confirmed GWs up to now, 90 events [9, 10, 11] since the first detection in 2015 [12]. In this section, we will go briefly over this family of GW events using the approximations and equations from Sec. 1.2.4.

Let's first assume that we have a binary system of masses m_1 and m_2 in a stable circular orbit. It can be shown that in the Center of Mass (CM) frame the tensor M can be written as in Eq. 1.50:

$$\begin{aligned} M_{11} &= \mu R^2 \frac{1 - cos(2w_s t)}{2} \\ M_{11} &= \mu R^2 \frac{1 + cos(2w_s t)}{2} \\ M_{12} &= \frac{\mu}{2} R^2 sin(2w_s t), \end{aligned} \tag{1.50}$$

where $\mu = \frac{m_1 m_2}{m_1 + m_2}$ is the reduced mass, w_s is the angular frequency of the source and R is the radius of the trajectory. Hence, given the entries of the tensor M we can estimate the two polarizations for this problem using Eq. 1.48. At the end, we get the following equation:

$$\begin{aligned} h_+(t, \theta, \phi) &= \frac{1}{r} \frac{G}{c^4} \mu w_s^2 R^2 \left(\frac{1 + cos^2(\theta)}{2} \right) cos(2w_s t - \frac{w_s r}{c} + 2\phi) \\ h_x(t, \theta, \phi) &= \frac{1}{r} \frac{2G}{c^4} \mu w_s^2 R^2 cos(\theta) sin(2w_s t - \frac{w_s r}{c} + 2\phi), \end{aligned} \tag{1.51}$$

from where we can deduce that, under the assumption of a stable circular orbit and non relativistic effects, the GW emitted is monochromatic with a frequency twice that of the source. Furthermore, due to both the angle ϕ and the distance to the detector, r, being constant, the term $2\phi - \frac{w_s r}{c}$ is just a constant phase inside the sine and cosine that can be removed by shifting our choice of start time. This means that Eq. 1.51 can be written as Eq. 1.52:

$$\begin{aligned} h_+(t,\iota) &= \frac{1}{r}\frac{G}{c^4}\mu w_s^2 R^2 \left(\frac{1+cos^2(\iota)}{2}\right) cos(2w_s t) \\ h_x(t,\iota) &= \frac{1}{r}\frac{2G}{c^4}\mu w_s^2 R^2 cos(\iota) sin(2w_s t), \end{aligned} \tag{1.52}$$

where we have also changed θ to 'ι' to better match the standard notation for the angle indicating the inclination of the source[3]. Assuming that the line of sight defines the z-axis of the source, a GW propagating towards us will have an angle between its direction of propagation and the line of sight given by θ.

The Eq. 1.52 for CBC events are usually written in terms of the chirp mass $M_c = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}$ and frequency of the GW $f_{gw} = \frac{2w_s}{2\pi}$. Furthermore, by Kepler's law, we can write R in terms of w_s as $R^3 = \frac{Gm}{w_s^2}$. To summarize, the GW emitted by two compact objects in a circular stable orbit is given by Eq. 1.53:

$$\begin{aligned} h_+(t,\iota) &= \frac{4}{r}\left(\frac{GM_c}{c^2}\right)^{5/3}\left(\frac{\pi f_{gw}}{c}\right)^{2/3}\left(\frac{1+cos^2(\iota)}{2}\right) cos(2\pi f_{gw} t) \\ h_x(t,\iota) &= \frac{4}{r}\left(\frac{GM_c}{c^2}\right)^{5/3}\left(\frac{\pi f_{gw}}{c}\right)^{2/3} cos(\iota) sin(2\pi f_{gw} t), \end{aligned} \tag{1.53}$$

which tells us that, at first order, the GW is defined only by the chirp mass, M_c, and not the individual values of the component masses. However, as was mentioned earlier, the system is losing energy by GW radiation and the radius of the orbit has to decrease to compensate. Hence, the orbit is not a stable circular orbit but a quasi-circular orbit and Eq. 1.53 does not completely describe the system.

[3]The inclination is defined as the angle between the line of sight and the direction of propagation

By including the loss of energy, it can be shown that the GW can be written as in Eq. 1.54:

$$h_+(t,\iota) = \frac{1}{r}\left(\frac{GM_c}{c^2}\right)^{5/4}\left(\frac{5}{c\tau}\right)^{1/4}\left(\frac{1+cos^2(\iota)}{2}\right)cos\left[-2\left(\frac{5GM_c}{c^3}\right)^{-5/8}\tau^{5/8}+\phi_0\right]$$
$$h_x(t,\iota) = \frac{1}{r}\left(\frac{GM_c}{c^2}\right)^{5/4}\left(\frac{5}{c\tau}\right)^{1/4}cos(\iota)sin\left[-2\left(\frac{5GM_c}{c^3}\right)^{-5/8}\tau^{5/8}+\phi_0\right], \tag{1.54}$$

where $\tau = t_{coal} - t$ is the time to coalescence, t_{coal} is the coalescence time and ϕ_0 is the coalescence phase. These equations are similar to the equations for a circular orbit but with one main difference: the frequency of the GW is not constant but changes with time due to the decrease in radius. In fact, the closer the GW is to the coalescence time, the larger the amplitude and frequency of the GW. This is known as chirp and an example can be seen in Fig. 1.1.

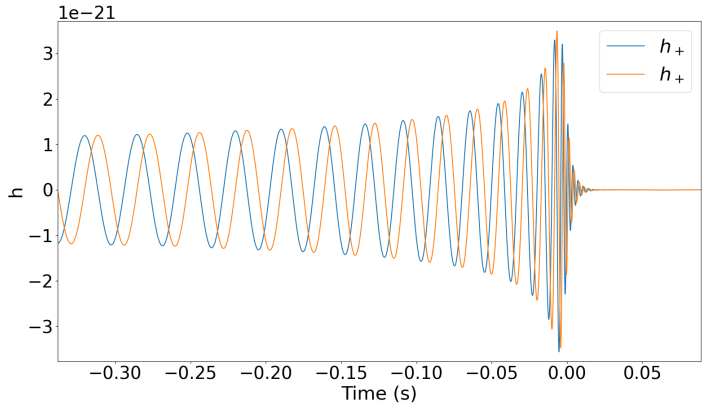

Figure 1.1: Plus and cross polarization for a GW produced by the coalescence of two black holes of 30 solar masses.

A GW produced by CBC events can be described by 15 parameters:

- Two component masses or chirp mass and reduced mass.
- 6 spin components in Cartesian or spherical coordinates. We have not taken spins into account here but the emission of GWs is affected by precession.
- Luminosity distance to the source.
- Sky position in terms of right ascension and declination. These two will be relevant for the detection and will be explored in Chapter 2.
- A reference time and phase, usually the merger time in the detector.
- Polarization and inclination angles. These angles correspond to the angles of the radiation frame and are used to define the propagation of the GW in a generic direction, see ϕ and θ in Eq. 1.48.

This derivation has been made by assuming a weak gravitational field at the source. This is approximately true when the radius of the orbit, r, is bigger than the ISCO (Innermost-stable circular orbit) radius, r_{ISCO} (minimum distance from which a circular stable orbit is allowed). The maximum frequency of this inspiral phase is given by the ISCO frequency, f_{ISCO}:

$$\begin{aligned} f_{ISCO} =& \frac{1}{6\sqrt{6}(2\pi)} \frac{c^3}{GM_T} \\ r_{ISCO} =& \frac{6GM_T}{c^2} = 3R_s, \end{aligned} \tag{1.55}$$

where M_T is the total mass of the system. For frequencies higher than this value, our hypothesis is no longer valid and we need to solve the full Einstein equations. These solutions are obtained by applying techniques from numerical relativity (NR).

Therefore, we can divide the evolution of a GW from a CBC event into three regions, as seen in Fig. 1.1:

1. Inspiral phase, when $f < f_{ISCO}$ the system has a quasi-stable circular orbit radiating energy as GWs. Therefore, the radius of the orbit is decreasing with time. This can be modelled using Newtonian approximations or including higher terms (known as post-Newtonian (PN) approximations [13]).

2. Merger, when $f > f_{ISCO}$ the system becomes relativistic and the full Einstein equations need to be solved. In this phase, the binary system coalesces at $\tau = 0$ and creates a new compact object.

3. Ring-down, after the coalescence, the new compact object emits GWs for a short time until it stabilizes and can be described again by perturbation theory.

A PN expansion is obtained by expanding the metric and energy-momentum tensor in terms of $\epsilon = \left(\frac{v}{c}\right)^2 \ll 1$ and allows us to solve the field equations under the assumption of slowly moving sources and a weak gravitational field. In particular, this approximation holds when $\left(\frac{v}{c}\right)^2 \sim \frac{R_s}{d} \ll 1$, where d is the orbital distance of the binary system. Each power of ϵ corresponds to one PN order in the expansion. For example, a 3.5PN approximation includes terms up to $\epsilon^{3.5} = \left(\frac{v}{c}\right)^7$. Schematically, the gravitational wave can be written as in Eq. 1.56:

$$h_{ij}(t, \mathbf{x}) = \frac{2G}{rc^4}(A_{ij} + \frac{1}{c}A_{ij}^{1PN} + \frac{1}{c^2}A_{ij}^{1PN} + \frac{1}{c^3}A_{ij}^{1.5PN} + \\ + \frac{1}{c^4}A_{ij}^{2PN} + \frac{1}{c^5}A_{ij}^{2.5PN} + \frac{1}{c^6}A_{ij}^{3PN} + \frac{1}{c^7}A_{ij}^{3.5PN} + \dots), \quad (1.56)$$

where each term is evaluated at the retarded time $t - \frac{r}{c}$. The quadrupole formula shown in Eq. 1.46 can be recovered from the leading term of the expansion. In the PN approximation, retardation effects[4] are considered small and is the correct tool in the near zone region of the non relativistic source. The condition describing this region is given by the wavelength of the emitted radiation: $r \ll \lambda$. PN expansion has been successfully used to solve several problems like the motion of N point-like objects (e.g. Solar System dynamics) or modelling the emission of GWs from CBC events. In the limit $\frac{v}{c} \to 0$ this approximation reduces to Newton's law.

In Fig. 1.2 a comparison between the predicted waveforms using NR calibrations and only PN terms can be seen. During the inspiral phase, the waveforms show similar amplitudes and frequencies (though with a small phase shift) but, when the frequency of the orbit is higher than the ISCO frequency and it gets closer to the merger, the PN approximation doesn't hold.

[4]Effects produced by the limited velocity of propagation of the radiation.

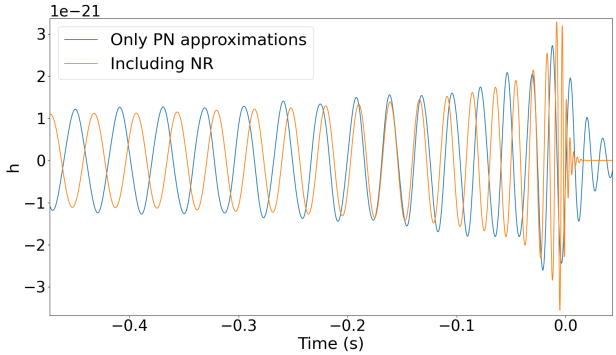

Figure 1.2: Comparison between the GW modelled by using only post-newtonian (PN) terms and including numerical relativity (NR) calibrations.

There is a huge effort to properly model the GWs from CBC events [14]. Several families of models have been developed to tackle this task, modelling the GW both in time and frequency domain. In Fig. 1.2 we show two of those models in the time domain. The waveform using only PN approximations [15, 16, 17, 18] is based on the model (template) "TaylorF2" in which the inspiral phase is modelled using a 3.5PN approximation. As previously stated, this approach holds reasonably well in regions of weak gravitational fields and no relativistic speeds. However, once the system becomes relativistic and the gravitational field intensifies, the approximation fails and different methods are needed. Nevertheless, for low mass mergers ($M_T < 30\mathrm{M}_\odot$) the PN approximations are valid due to the merger frequency being too high to be detected by the GW detectors.

More sophisticated approaches are done by using the results from NR simulations. These simulations are very expensive due to the computational cost increasing with the simulated number of cycles. An example can be seen in Fig. 1.2 where a waveform including NR calibrations, based on the frequency domain model 'IMRPhenomPv2', is used [19, 20, 21, 22, 23].

This combination is made in the region in which the PN approximation starts breaking down and NR simulation starts. Matching PN and NR simulations was suggested by studying the consistency between PN and NR simulations

for a variety of physical configurations and the PN amplitudes during mergers and ringdown. Depending on the procedure used to match the two results, different families of waveforms were developed.

In the case of waveforms from the phenomenological family, within the window in which both PN and NR are good approximations, the waveforms are matched (either in frequency or time domain) and used to interpolate a hybrid waveform. This hybrid waveform can be divided into three regions:

- Before the window where the waveform is purely PN approximation.
- Within the window where the waveform is interpolated from PN and NR simulations.
- After the window where the waveform is estimated purely from NR simulations.

Let's now assume that $\mathcal{D}$ is the space of intrinsic physical parameters and λ a point from that space. Then, for any point λ in $\mathcal{D}$, the physical waveform will be given by $h(t;\lambda)$. Assuming that we have N hybrid waveforms with known parameters, the objective is to create a phenomenological model, $h_{phenom}(t;\lambda)$, that interpolates between the hybrid waveforms with sufficient accuracy. However, a more convenient choice is to work with a set of phenomenological parameters, $\hat{\lambda}$, defined from a one-to-one mapping, $\mathcal{D} \to \hat{\mathcal{D}}$, to the space of phenomenological parameters $\hat{\mathcal{D}}$. Therefore, the phenomenological waveform can be written as $h_{phenom}(t;\hat{\lambda}(\lambda))$ and is usually expressed in terms of an amplitude and a phase: $\bar{h}_{phenom}(f) = A_{phenom}(f)e^{i\phi_{phenom}}(f)$. The phenomenological parameters that need to be estimated depend on the models used for the amplitude and phase.

Another important family of waveforms are the EOB (effective-one-body) waveforms [24, 25, 26, 27]. This waveform family is based on transforming the two body problem into a one body problem with an effective metric. This allows us to understand a non precessing binary system as a single particle moving in a deformed Schwarzschild spacetime. The deformation parameter is given by the symmetric mass ratio $\frac{m_1 m_2}{(m_1+m_2)^2}$. Furthermore, most of the parameters of this model can be estimated from PN approximations while the remaining are obtained from calibrating to NR simulations. An example of a waveform generated using this method is "SEOBNRv3", which also includes spin precession.

Chapter 2

Laser interferometry for gravitational waves detection

In Chapter 1, we presented the theory behind GWs and here we study how they are detected. The detection method is based on measuring the change in position of test masses by using laser interferometry [1]. We discuss the basic elements needed to understand a Michelson interferometer, how it couples to GWs and its limitations. We show that a Fabry-Perot interferometer is needed to detect GWs. Finally, we provide a comprehensive review of possible noise sources that limit the sensitivity of the interferometer.

2.1 Michelson interferometer

As we have seen in Sec. 1.2.3, a passing GW will change the distance between two objects. This motivated the search for GW by measuring the change in length of a bar due to the passing GW. However, this technique was unable to achieve the required level of precision and a different approach was needed: laser interferometers. The GW changes the arm length of the interferometer and consequently changes the interference pattern at the output. This produces a variation in the measured output power from which the GW is extracted.

As a first approximation, we consider a Michelson type interferometer (see Fig. 2.1). The interferometer consists of a monochromatic laser impinging on a beam splitter that will divide the beam in two. These two beams will

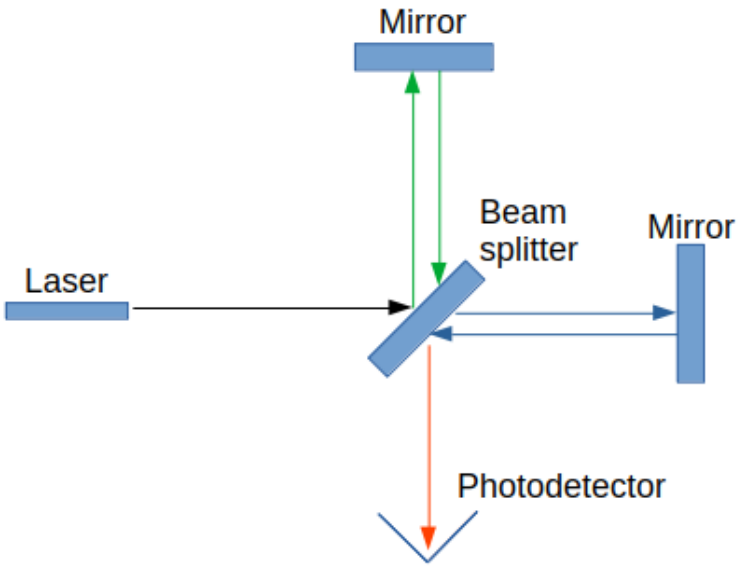

Figure 2.1: Layout of a Michelson interferometer

propagate along each arm until they reach a highly reflective mirror and bounce back towards the beam splitter. At this point, the beams will recombine and propagate towards the photodetector where the power will be measured.

Formally, given a beam described by the electromagnetic field:

$$E_{in}(t) = E_0 e^{-iw_L t + ik_L \cdot x},$$

with a laser frequency w_L and a wavenumber given by $k_L = \frac{w_L}{c}$, it will reach the beam splitter at a time t_0 and split into two beams. The beam going towards the upper mirror (propagating along the y-axis) will cover a distance L_y before bouncing back and reaching the beam splitter again at a time $t = t_0 + \frac{2L_y}{c}$. The second beam will reach the beam splitter at a time $t = t_0 + \frac{2L_x}{c}$. Therefore, the recombined beam at a time 't' will be generated by the superposition of the beam that reached the beam splitter at a time $t_0^{(x)} = t - \frac{2L_x}{c}$ and a second beam entering at a different time $t_0^{(y)} = t - \frac{2L_y}{c}$

Assuming that the beam splitter is at $(x, y) = (0, 0)$ we know that the initial phase of these two beams is $e^{-iw_L t_0^{(x)}} = e^{-iw_L t + 2ik_L L_x}$ and $e^{-iw_L t_0^{(y)}} = e^{-iw_L t + 2ik_L L_y}$. During free propagation, this phase will remain unchanged but there will be a change in the amplitude due to the effect of the mirrors/beam splitter. Taking into account that we are assuming perfect mirrors[1] and beam splitter[2] we have that the change in amplitude is of -1/2 for the first beam (x-axis) and 1/2 for the second (y-axis):

$$\begin{aligned} E_1^x =& \frac{-1}{2} E_0 e^{-iw_L t + 2ik_L L_x} \\ E_2^y =& \frac{1}{2} E_0 e^{-iw_L t + 2ik_L L_y}. \end{aligned} \tag{2.1}$$

The output beam is given by the superposition of these two electric fields, E_{out}:

$$\begin{aligned} E_{out} = E_1^x + E_2^y &= \frac{E_0}{2}(e^{-iw_L t + 2ik_L L_y} - e^{-iw_L t + 2ik_L L_x}) \\ &= \frac{E_0}{2} e^{-iw_L t + ik_L(L_x + L_y)}(e^{-ik_L(L_x - Ly)} - e^{ik_L(L_x - Ly)}) \\ &= -iE_0 e^{-iw_L t + ik_L(L_x + L_y)} sin[k_L(L_y - L_x)], \end{aligned} \tag{2.2}$$

where we have used $2L_x = (L_x + L_y) + (L_x - L_y)$ and $2L_y = (L_x + L_y) - (L_x - L_y)$. We can estimate the power detected at the photodiode by using that it is proportional to the absolute value squared of the impinging field:

$$P \propto |E_{out}|^2 = E_0^2 sin^2[k_L(L_y - L_x)], \tag{2.3}$$

which allows us to conclude that changes in the length of the arms can be measured by the photodiodes. Let's now assume that a GW propagating along the z-axis with only the plus polarization reaches the interferometer:

$$h_+(t) = h_0 cos(w_{gw} t),$$

then, in the TT gauge, the GW will affect the propagation of light between the free falling masses (mirrors). In this case, the line element is given by:

$$ds^2 = -c^2 dt^2 + [1 + h_+(t)]dx^2 + [1 - h_+(t)]dy^2 + dz^2, \tag{2.4}$$

[1]Perfect mirror: reflection coefficient of -1

[2]Perfect beam splitter: reflection coefficient of $1/\sqrt{2}$ for one side and $-1/\sqrt{2}$ for the other

and the photons will follow the path given by the null geodesic:

$$ds^2 = 0 = -c^2dt^2 + [1 + h_+(t)]dx^2 + [1 - h_+(t)]dy^2 + dz^2. \tag{2.5}$$

In particular, for the light propagating along the x-axis:

$$\begin{aligned} 0 &= -c^2dt^2 + [1 + h_+(t)]dx^2 \\ dx &= \pm cdt\frac{1}{\sqrt{1 + h_+(t)}} \approx \pm cdt[1 - \frac{1}{2}h_+(t)], \end{aligned} \tag{2.6}$$

where we have approximated to first order in h_0. In this case, the '+' sign corresponds to the propagation from the beam splitter to the mirror, while the '−' sign is from the mirror to the beam splitter. Assuming that the photon is at the beam splitter at a time t_0 and reaches the mirror at L_x at a time t_1, we have that:

$$\begin{aligned} L_x &= \int_0^{L_x} dx = \int_{t_0}^{t_1} cdt - \frac{c}{2}\int_{t_0}^{t_1} h_+(t)dt \\ &= c[t_1 - t_0] - \frac{c}{2}\int_{t_0}^{t_1} h_+(t)dt, \end{aligned} \tag{2.7}$$

where it will reflect and reach the beam splitter at a time t_2:

$$\begin{aligned} -L_x &= \int_{L_x}^{0} dx = -\int_{t_1}^{t_2} cdt + \frac{c}{2}\int_{t_1}^{t_2} h_+(t)dt \\ L_x &= c[t_2 - t_1] - \frac{c}{2}\int_{t_1}^{t_2} h_+(t)dt. \end{aligned} \tag{2.8}$$

These two equations can be added to obtain an expression of t_2 as a function of the starting time, t_0:

$$t_2 - t_0 = \frac{2L_x}{c} + \frac{1}{2}\int_{t_0}^{t_2} h_+(t)dt, \tag{2.9}$$

where we have recovered the term $\frac{2L_x}{c}$ plus a first order correction due to the GW. Now, using that at first order t_2 can be approximated as $t_2 \approx t_0 + \frac{2L_x}{c}$, we have that the integral is:

$$\int_{t_0}^{t_0+\frac{2L_x}{c}} h_+(t)dt = \int_{t_0}^{t_0+\frac{2L_x}{c}} h_0 cos(w_{gw}t)dt = \frac{h_0}{w_{gw}}[sin[w_{gw}(t_0 + \frac{2L_x}{c})] - sin(w_{gw}t_0)], \tag{2.10}$$

which can be simplified by using the identity $sin(\alpha+2\beta)-sin(\alpha) = 2sin(\beta)cos(\alpha+\beta)$:

$$\int_{t_0}^{t_0+\frac{2L_x}{c}} h_+(t)dt = 2\frac{h_0}{w_{gw}}sin(w_{gw}L_x/c)cos(w_{gw}(t_0+L_x/c)). \tag{2.11}$$

Substituting this derivation in the original equation:

$$\begin{aligned} t_2 - t_0 &= \frac{2L_x}{c} + \frac{1}{2}2\frac{h_0}{w_{gw}}sin(w_{gw}L_x/c)cos(w_{gw}(t_0+L_x/c)) \\ &= \frac{2L_x}{c} + \frac{L_x}{c}\frac{sin(w_{gw}L_x/c)}{w_{gw}L_x/c}h_0 cos(w_{gw}(t_0+L_x/c)) \\ &= \frac{2L_x}{c} + \frac{L_x}{c}\frac{sin(w_{gw}L_x/c)}{w_{gw}L_x/c}h_+(t_0+L_x/c). \end{aligned} \tag{2.12}$$

In the case of the y-axis, as we can see from the line element in Eq. 2.4, the derivation is the same but with the sign of $h_0(t)$ reversed. Therefore, we have that:

$$t_2 - t_0 = \frac{2L_y}{c} - \frac{L_y}{c}\frac{sin(w_{gw}L_y/c)}{w_{gw}L_y/c}h_+(t_0+L_y/c). \tag{2.13}$$

These equations can be used to estimate the phase of the recombined beam in the same way as before. The beams that will recombine at a time $t_2 = t$, will start the round trip at a time:

$$\begin{aligned} t_0^x &= t - \frac{2L_x}{c} - \frac{L_x}{c}h(t-L_x/c)\frac{sin(w_{gw}L_x/c)}{w_{gw}L_x/c} \\ t_0^y &= t - \frac{2L_y}{c} + \frac{L_y}{c}h(t-L_y/c)\frac{sin(w_{gw}L_y/c)}{w_{gw}L_y/c}, \end{aligned} \tag{2.14}$$

and the phase at the beam splitter $(x, y) = (0, 0)$ will be:

$$\begin{aligned} e^{-iw_L t_0^x} &= e^{-iw_L[t-\frac{2L_x}{c}-\frac{L_x}{c}h(t-L_x/c)\frac{sin(w_{gw}L_x/c)}{w_{gw}L_x/c}]} = e^{-iw_L(t-\frac{2L_x}{c})+i\Delta\phi_x(t)} \\ e^{-iw_L t_0^y} &= e^{-iw_L[t-\frac{2L_y}{c}+\frac{L_y}{c}h(t-L_y/c)\frac{sin(w_{gw}L_y/c)}{w_{gw}L_y/c}]} = e^{-iw_L(t-\frac{2L_y}{c})+i\Delta\phi_y(t)}, \end{aligned} \tag{2.15}$$

where $\Delta\phi_x$ and $\Delta\phi_y$ are the change in the phase produced by the passing GW:

$$\begin{aligned} \Delta\phi_x(t) &= h_0\frac{w_L L_x}{c}\frac{sin(w_{gw}L_x/c)}{w_{gw}L_x/c}cos(w_{gw}(t-L_x/c)) \\ \Delta\phi_y(t) &= -h_0\frac{w_L L_y}{c}\frac{sin(w_{gw}L_y/c)}{w_{gw}L_y/c}cos(w_{gw}(t-L_y/c)). \end{aligned} \tag{2.16}$$

During the free propagation, the phase will remain the same and only the amplitude of the electromagnetic field will change (see Eq. 2.1). Therefore, given that L_x and L_y should be similar (but not equal), we can write this field in terms of $L = \frac{L_x+L_y}{2}$ using that $2L_x = 2L + (L_x - L_y)$ and $2L_y = 2L - (L_x - L_y)$. Furthermore, inside $\Delta\phi_x$ and $\Delta\phi_y$ we approximate L by $L \approx L_x \approx L_y$:

$$\begin{aligned}
E_1^x(t) =& \frac{-1}{2} E_0 e^{-iw_L(t-\frac{2L}{c})+i\phi_0+i\Delta\phi_x(t)} \\
E_2^y(t) =& \frac{1}{2} E_0 e^{-iw_L(t-\frac{2L}{c})-i\phi_0+i\Delta\phi_y(t)} \\
\Delta\phi_x(t) =& h_0 k_L L \frac{sin(w_{gw}L/c)}{w_{gw}L/c} cos(w_{gw}(t-L/c)) \\
\Delta\phi_y(t) =& -h_0 k_L L \frac{sin(w_{gw}L/c)}{w_{gw}L/c} cos(w_{gw}(t-L/c)),
\end{aligned} \tag{2.17}$$

where we have defined $\phi_0 = k_L(L_x - L_y)$. Finally, the phase difference in a Michelson interferometer produced by a GW is:

$$\Delta\phi_{Mich} = \Delta\phi_x - \Delta\phi_y = 2\Delta\phi_x, \tag{2.18}$$

and the recombined electromagnetic field at the photodiode is given by:

$$E_{out}(t) = E_1^x(t) + E_2^y(t) = -iE_0 e^{iw_L(t-2L/c)} sin(\phi_0 + \Delta\phi_x(t)). \tag{2.19}$$

In the limit case $w_{gw}L/c \ll 1$ (large GW period compared to the length of the arms), we have that, at first order, the effect of the GW on the phase shift is given by the change of $L_x - L_y$:

$$h(t - L/c) = \frac{\Delta(L_x - L_y)}{L}, \tag{2.20}$$

where we have used that $\frac{sin(w_{gw}L/c)}{w_{gw}L/c} \to 1$ when $w_{gw}L/c \to 0$.

The effect in the power by the GW can be obtained from Eq. 2.19 in the same way as before:

$$\begin{aligned}
P = P_0 sin^2(\phi_0 + \Delta\phi_x(t)) \propto |E_{out}|^2 \\
P = \frac{P_0}{2}[1 - cos[2(\phi_0 + \Delta\phi_x(t))]] \\
P = \frac{P_0}{2}[1 - cos[2\phi_0 + \Delta\phi_{Mich}(t))]].
\end{aligned} \tag{2.21}$$

Hence, the detected power at the photodiode will be modulated by the GW. These oscillations will vary depending on $\Delta\phi_{Mich}(t)$ and, thus, we want this value to be large. This can be achieved by maximizing the non time dependent term:

$$\frac{w_L}{c}L\frac{sin(w_{gw}L/c)}{w_{gw}L/c} = \frac{w_L}{w_{gw}}sin(w_{gw}L/c)$$

That will be maximum when $w_{gw}L/c = \frac{\pi}{2}$. In other words, when the length of the arms is $L = \frac{\lambda_{gw}}{4}$ or equivalently:

$$L = \frac{c}{4f_{gw}} \approx 750km\left(\frac{100Hz}{f_{gw}}\right). \tag{2.22}$$

The Eq. 2.22 shows that a Michelson interferometer for GWs detection would need an arm length of hundreds of kilometers. This is not realistic and a different approach is needed; a Fabry-Perot interferometer.

2.2 Fabry-Perot resonator

As we have seen in Eq. 2.22, the detection of a GW using a Michelson interferometer is not possible due to the arm length needed being of the order of hundreds of kilometers. This problem can be solved by using a Fabry-Perot interferometer. A Fabry-Perot interferometer uses Fabry-Perot cavities to "fold" the path of the light, achieving a similar performance as a hundred arm length Michelson interferometer with a reduced size.

These cavities are constructed by two parallel mirrors facing each other, as seen in Fig. 2.2, in which we introduce an electric field, E_{in}. This electric field will transmit inside the cavity and propagate until the end mirror. Once it reaches this mirror, a fraction will reflect and another transmits outside the cavity. The fraction reflected will go back to the initial mirror and repeat. This creates an electric field inside the cavity, E_{cav}, that will be the superposition of all the beams bouncing inside.

Let's assume that our input beam is described by a plane wave, $E_{in} = E_0e^{-iw_Lt+ik_Lx}$, and that we have been sending a beam for some time. If it reaches the first mirror situated at $x = 0$ at a time $t = t_0$ we have that the incident electric field is $E_0e^{-iw_Lt_0}$. Using the Fresnel equations, the reflected

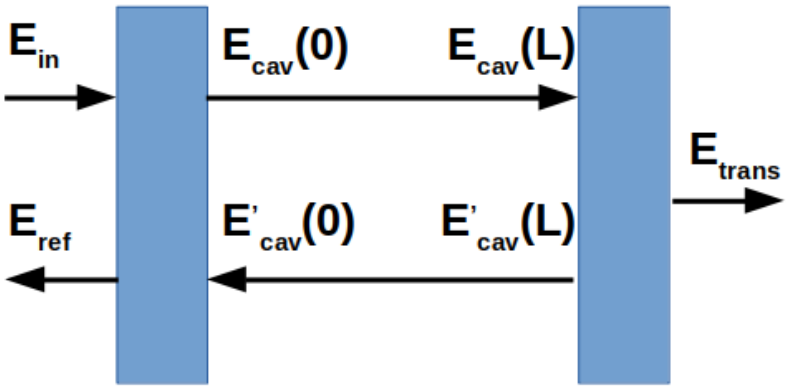

Figure 2.2: Schematic representation of a Fabry-Perot cavity with the electric field identified in each region. We assume that the first mirror is at $x = 0$ and the end mirror at $x = L$

beam is given by the reflection coefficient of the first mirror, r_1, times the electric field:

$$E_{ref}^{(0)} = r_1 E_0 e^{-iw_L t_0}. \tag{2.23}$$

However, given that the beam has been active for a while, as we have hypothesized, the reflected field will interfere with the one that reflected against the second mirror, came back and transmitted out of the cavity. This means that the beams had to reach the first mirror at a time $t_0 - \frac{2L}{c}$ so that it had time to travel through the cavity twice. Then, its electric field when it entered the cavity is given by $E_0 e^{-iw_L(t_0 - \frac{2L}{c})} = E_0 e^{-iw_L(t_0)} e^{2ik_L L}$, in which a new phase term appears due to the extra length travelled by the beam. If we take into account the reflection and transmission using the Fresnel coefficients, we have that this beam is given by Eq. 2.24, where t_1 is the transmission coefficient of the first mirror and r_2 is the reflection coefficient of the second:

$$E_{ref}^{(1)} = (-t_1^2 r_2 e^{2ik_L L}) E_0 e^{-iw_L t_0}. \tag{2.24}$$

In general, the n-th contribution is given by Eq. 2.25:

$$E_{ref}^{(n)} = (-t_1^2 r_2^n r_1^{n-1} e^{2ink_L L}) E_0 e^{-iw_L t_0}, \tag{2.25}$$

and at the end we have that the reflected beam is given by the sum of all the contributions, as in Eq. 2.26:

$$\begin{aligned} E_{ref} &= E_0 e^{-iw_L t_o}\left[r_1 - t_1^2 \sum_{n=1}^{\infty} r_1^{n-1} r_2^n e^{2ink_L L}\right] \\ &= E_0 e^{-iw_L t_o}\left[r_1 - t_1^2 r_2 e^{2ik_L L} \sum_{j=0}^{\infty} (r_1 r_2 e^{2ik_L L})^j\right] \\ &= E_0 e^{-iw_L t_o}\left[r_1 - t_1^2 r_2 e^{2ik_L L} \frac{1}{1 - r_1 r_2 e^{2ik_L L}}\right], \end{aligned} \tag{2.26}$$

where we have first changed the index of summation by $n = j + 1$ and then used the closed form for a geometric series given that $|r_1 r_2 e^{2ik_L L}| < 1$ and thus converges. The transmitted and cavity field can be obtained with a similar procedure. The end result can be seen in Eq. 2.27:

$$\begin{aligned} E_{ref} &= E_0 e^{-iw_L t_o}\left[\frac{r_1 - r_2(1 - l_1) e^{2ik_L L}}{1 - r_1 r_2 e^{2ik_L L}}\right] \\ E_{trans} &= E_0 e^{-iw_L t_o}\left[\frac{t_1 t_2 e^{2ik_L L}}{1 - r_1 r_2 e^{2ik_L L}}\right] \\ E_{cav}(0) &= E_0 e^{-iw_L t_o}\left[\frac{t_1}{1 - r_1 r_2 e^{2ik_L L}}\right] \\ E_{cav}(L) &= e^{ik_L L} E_{cav}(0) = E_0 e^{-iw_L t_o}\left[\frac{t_1 e^{ik_L L}}{1 - r_1 r_2 e^{2ik_L L}}\right], \end{aligned} \tag{2.27}$$

where we have used the relations between the Fresnel coefficients, $t_i^2 + r_i^2 = 1 - l_i$, assuming that the mirror 'i' has a loss given by l_i. It should be noted that all the end fields are proportional to the same factor, $\frac{1}{1 - r_1 r_2 e^{2ik_L L}}$. We can also observe that this factor becomes very large when the exponential is one and the reflection coefficients are very close to one. These points, given by the condition $2ik_L L = 2\pi n$ for integer n, are called resonances.

In a resonance, the beams that propagate through the cavity are interfering constructively and the resulting cavity and transmitted beam reach high values. The reflected beam is slightly more complicated due to the additional contribution of the exponential in the numerator. The power of the transmitted or cavity field as a function of $2k_L L$ reaches a maximum every

$2k_L L = 2\pi n$ for integer n (See Fig. 2.3). The distance between two consecutive maximums is given by Eq. 2.28, where we have used that $w_L = \frac{\pi c}{L} n$, obtained by using that $k_L = \frac{w_L}{c}$ and $2k_L L = 2\pi n$ for the maximum:

$$\Delta w_L = \frac{\pi c}{L}. \tag{2.28}$$

Δw_L is known as the free spectral range and along with the full width at half maximum of the peaks, δw_L, allows us to define the finesse, $\mathcal{F}$, of the cavity as in Eq. 2.29:

$$\begin{aligned} \delta w_L &= \frac{c}{L} \frac{1 - r_1 r_2}{\sqrt{r_1 r_2}} \\ \mathcal{F} &= \frac{\Delta w_L}{\delta w_L} = \frac{\pi \sqrt{r_1 r_2}}{1 - r_1 r_2}, \end{aligned} \tag{2.29}$$

which is used to characterize the performance of the cavity. In particular, if we write the reflected field in polar form, we have that changes in the phase around the resonance are given by Eq. 2.30:

$$\frac{\partial \phi}{\partial \epsilon} \approx \frac{2\mathcal{F}}{\pi} \tag{2.30}$$

with ϵ indicating small variations around the resonance, $2k_L L = 2\pi n + \epsilon$. It can be seen that the change in the phase around the resonance is given by the finesse of the cavity. Finally, a convenient parameter to define for further derivations is 'p', defined as in Eq. 2.31:

$$(1 - p_1) r_2^2 = (1 - p). \tag{2.31}$$

These elements describe an ideal Fabry-Perot cavity that can be used to construct a Fabry-Perot interferometer.

2.3 Gravitational wave coupling to a Fabry-Perot interferometer

In this section we are going to explore the coupling of GWs to a Fabry-Perot interferometer, as shown in Fig. 2.4, and how they can be detected. Let's assume that we have a beam divided in two perpendicular directions by a beam splitter. One of the beams will go up entering a Fabry-Perot cavity

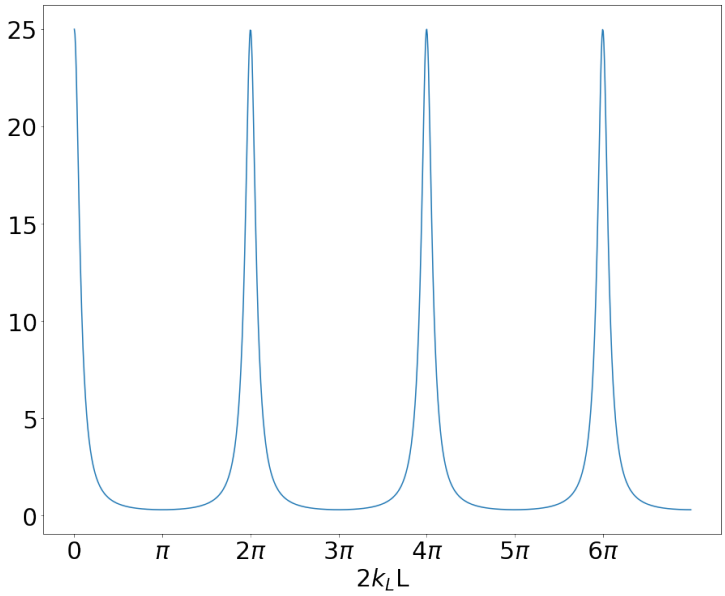

Figure 2.3: Transmitted field squared in units of $(E_0 t_1 t_2)^2$, as a function of $2k_L L$, for $r_1 r_2 = 0.8$. The maximums are reached when $2k_L L = 2\pi n$ for integer n.

aligned with the y-axis and the other will continue to the right, entering a different cavity aligned with the x-axis. Let's also assume that only the plus polarization of a GW enters the interferometer parallel to the z-axis.

Under these conditions, when a field enters the cavity and does a full trip, it will acquire an extra phase modulation due to the GW. The end result will be a field with two sidebands at $w_L \pm w_{gw}$, where w_{gw} is the angular frequency of the GW, and a central component at the carrier frequency w_L. These are the fields that will bounce back and forth through the cavity as seen before. In general, we have that the field in the left mirror is of the form:

$$A(t) = A_0 e^{-iw_L t} + \frac{1}{2} h_0 A_1 e^{-i(w_L - w_{gw})t} + \frac{1}{2} h_0 A_1 e^{-i(w_L + w_{gw})t}. \tag{2.32}$$

The presence of a GW also affects the time it would take to do a round trip. Let t be the time at which the field has finished one round trip, then this

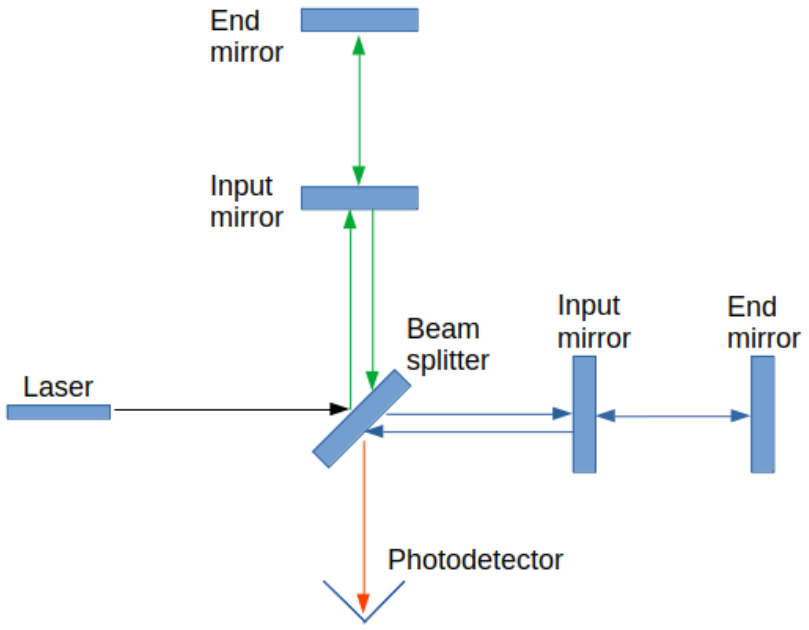

Figure 2.4: Layout of a Fabry-Perot interferometer.

field has entered the cavity at a time t_0 given by Eq. 2.33:

$$t_0 = t - \frac{2L}{c} - \frac{L}{c} h_0 cos(w_{gw}(t - \frac{L}{c}))sinc(w_{gw}\frac{L}{c}), \tag{2.33}$$

where we can observe the additional term given by the effect of the GW. If we perform similar derivations as before but including the GW we can generalize Eq. 2.27 to account for its effect. This can be used to estimate the phase shift in a Fabry-Perot cavity, as in Eq. 2.34:

$$\begin{aligned} |\Delta\phi_x| &= h_0 k_L L sinc(w_{gw}\frac{L}{c}) \frac{r_2(1-r_1^2-p)}{[r_2(1-p)-r_1]} \frac{1}{|e^{2iw_{gw}L/c} - r_1 r_2|} \\ &= h_0 k_L L sinc(w_{gw}\frac{L}{c}) \frac{r_2(1-r_1^2-p)}{[r_2(1-p)-r_1]} \frac{1}{[1+(r_1 r_2)^2 - 2r_1 r_2 cos(2w_{gw}\frac{L}{c})]^{1/2}}, \end{aligned} \tag{2.34}$$

which can be simplified further by assuming that the reflection coefficient of the second mirror, r_2, is one, $p \to 0$ and that we are interested in a

region where $\frac{w_{gw}L}{c} \ll 1$. The last assumption is obtained by using that our instrument is designed such that $\mathcal{F}\frac{w_{gw}L}{c} \to 1$, which is achieved by using large values for the finesse. Therefore, the second term should be smaller than one and the phase shift can be written as in Eq. 2.35:

$$|\Delta\phi_x| \approx h_0 2k_L L\frac{\mathcal{F}}{\pi}\frac{1}{[1+(4\pi f_{gw}\tau_s)^2]^{1/2}}$$
$$\tau_s = \frac{2L}{c}\frac{1}{1-r_1^2}, \tag{2.35}$$

where f_{gw} is the frequency of the GW in Hertz and τ_s is known as the storage time and describes the average time a photon spends inside the cavity. This approximation will hold as long as the condition from Eq. 2.36 is true:

$$f_{gw} \ll \frac{c}{2\pi L} \approx 12kHz\left(\frac{4km}{L}\right). \tag{2.36}$$

The total phase shift in the interferometer will be the difference between the contribution of each arm, that is, $\Delta\phi = \Delta\phi_x - \Delta\phi_y$. However, given that the space time interval in the TT frame is:

$$ds^2 = -c^2dt^2 + [1+h_+(t)]dx^2 + [1-h_+(t)]dy^2 + dz^2,$$

we have that the derivation will be the same but changing $h_0 \to -h_0$. This means that the phase shift in the cavity aligned with respect to the y-axis is $\Delta\phi_y = -\Delta\phi_x$ and the total phase shift is given by Eq. 2.37:

$$|\Delta\phi| = 2\Delta\phi_x \approx 4h_0k_LL\frac{\mathcal{F}}{\pi}\frac{1}{[1+(\frac{f_{gw}}{f_p})^2]^{1/2}} \tag{2.37}$$

$$f_p = \frac{1}{4\pi\tau_s}, \tag{2.38}$$

where f_p is known as the pole frequency. This phase shift will affect the interferometer pattern at the output of the interferometer and thus the GW will be detected. This formula describes the effect in the phase that a GW with only plus polarization coming along the z-axis will have. In a real situation, this is not always the case: the GW can reach the interferometer from an arbitrary direction and can include the other polarization.

The coupling between the interferometer and the incoming wave is given by the antenna pattern functions, $F_+(\theta, \phi)$ and $F_x(\theta, \phi)$. Under the approximation we used before, $\frac{w_{gw}L}{c} \ll 1$, it can be shown that these functions are described by the closed form given by Eq. 2.39:

$$\begin{aligned} F_+(\theta, \phi) &= \frac{1}{2}(1 + cos^2(\theta))cos(2\phi) \\ F_x(\theta, \phi) &= cos(\theta)sin(2\phi), \end{aligned} \tag{2.39}$$

where θ is the angle between the direction of propagation and the z-axis of the interferometer and ϕ is the angle between the x-axis and the projection onto the $x - y$ plane of this vector. These equations can be derived by rotating the frame in which the GW is propagating along the z'-axis into the detector frame using the angles defined previously, as we have seen in Eq. 1.48. Using these two functions, we can define the relative phase shift in the interferometer as in Eq. 2.40:

$$\frac{1}{2}(h_{xx} - h_{yy}) = F_+(\theta, \phi)h_+ + F_x(\theta, \phi)h_x. \tag{2.40}$$

These equations also show that for a given position relative to the incoming GW, the interferometer is blind in some regions of the sky. These regions are called blind spots and happen when the relative phase shift is zero. In other words, the relative displacement is the same in both directions and it gets compensated. For example, if we have a waveform with only cross polarization there will be a blind spot at $\theta = \frac{\pi}{2}$ due to $F_x(\frac{\pi}{2}, \phi) = 0 \ \ \forall \phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

The Eq. 2.37 allows us to get an estimate on the values of the phase shift we are expected to encounter in our interferometers. Let's assume that the interferometer interacts with a GW of amplitude $h_0 = 10^{-21}$, that the beam inside the cavity has a wave length of $\lambda_L = 1064$nm and that it has a frequency such that $f_{gw} \ll f_p$. Under these conditions and considering an interferometer with arm length $L = 4$km and a finesse of $\mathcal{F} = 200$, we have that the expected phase shift is:

$$|\Delta\phi| \approx 1.2 \cdot 10^{-8} \text{ rad},$$

which is a very small change in the phase of the beam.

2.4 Dark fringe

As discussed in the previous section, a passing GW will produce a small phase shift that will change the detected power at the output photodiode. Therefore, we need to define the working point of the interferometer so that we can extract the phase from the detected power. As we have seen in Eq. 2.21, in a Michelson interferometer, the power oscillates as a function of $\phi = \phi_0 + \Delta\phi_{gw}(t)$ where ϕ_0 is the working point of the interferometer and has to be chosen beforehand. If we want to maximize the change in power by a passing GW, this working point should be set to $\phi_0 = \frac{\pi}{4}$ as it would correspond to the point of maximum derivative (see Fig. 2.5 red point) and most sensitive to small displacements $\phi_0 \to \phi_0 + \Delta\phi_{gw}(t)$. This is, however, not the correct choice since this working point is not only sensitive to a passing GW but also to fluctuations in the power of the laser, which are indistinguishable from the phase changes by the wave. The signal is buried underneath the fluctuations of the DC contribution.

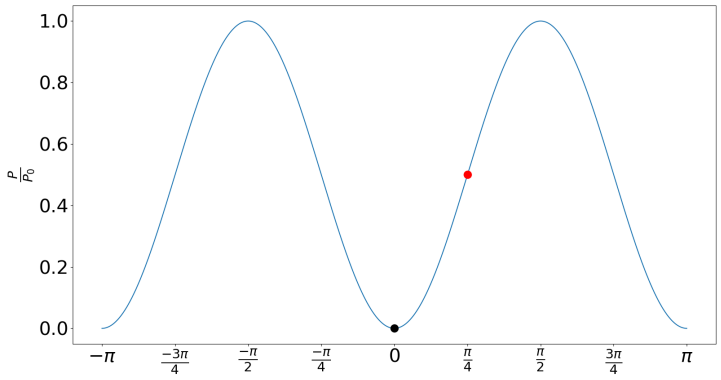

Figure 2.5: Power at the photodiode as a function of the phase of the beam in a Michelson interferometer. The black dot at $\phi = 0$ corresponds to the "Dark fringe" while the red dot at $\phi = \frac{\pi}{4}$ is the position with maximum derivative.

For this reason, a different working point is chosen: the dark fringe given by $\phi_0 = 0$ (see Fig. 2.5 black point) in which the detected output is zero (experiments with zero output when there is no signal are known as null instruments). At this working point, both P and the derivative are zero

and the output is insensitive to power fluctuations of the laser. However, in this working point, the response of the interferometer to a signal is no longer linear but quadratic: $\Delta P = O(h^2)$. Given the amplitudes we have seen in Sec. 1.2.4, with a maximum value of $\approx 10^{-21}$, it would be impossible to detect the GW. This is solved by introducing a phase modulation to the input laser. This modulation will add sidebands to the beam that will create a term linear in 'h' oscillating at the frequency of the modulator at the output photodiode. From this term we can extract the effect of the GW. Therefore, we can measure an output linear in 'h' that is insensitive to laser fluctuations. The modulated beam at the level of the beam splitter is described by the following electric field:

$$E_{In} = E_0 e^{-i(w_L t + \Gamma sin(\Omega_{mod} t))}, \tag{2.41}$$

where Γ is the modulation index and $\Omega_{mod} = 2\pi f_{mod}$ is the modulation frequency. Assuming that Γ is small, most of the sidebands will be suppressed. Therefore, we will focus on the first two sidebands around the carrier frequency, w_L. These sidebands have a frequency given by $w_\pm = w_L \pm \Omega_{mod}$.

The output beam with the carrier frequency, $(E_{out})_c$ will be similar to Eq. 2.2. The only difference is that the amplitude of this beam is now $E_0 J_0(\Gamma)$, where $J_n(\Gamma)$ are Bessel functions. Following the same method, we can derive the electric field of the sidebands at the output:

$$\begin{aligned}
(E_{out})_c &= -iE_0 J_0(\Gamma) e^{-iw_L t + ik_L(L_x+L_y)} sin[\frac{2\pi}{\lambda_L}(L_y - L_x)] \\
(E_{out})_\pm &= \mp iE_0 J_1(\Gamma) e^{-iw_\pm t + ik_\pm(L_x+L_y)} sin[2\pi \left(\frac{\Delta L}{\lambda_L} \pm \frac{\Delta L}{\lambda_{mod}}\right)].
\end{aligned} \tag{2.42}$$

From these equations we can deduce that in the case of symmetrical arms, $L_x = L_y$, both the carrier and sidebands are in the dark fringe: $(E_{out})_c = (E_{out})_\pm = 0$. This is solved by introducing an asymmetry, known as Schnupp asymmetry, in the length of the arms: we set $L_x - L_y$ to be an integer number of laser wavelengths. Under this hypothesis, only the carrier frequency will remain in the dark fringe and the electric field of the sidebands will be given by Eq. 2.43:

$$(E_{out})_\pm = -iE_0 J_1(\Gamma) e^{-iw_\pm t + ik_\pm(L_x+L_y)} sin[2\pi \frac{\Delta L}{\lambda_{mod}}]. \tag{2.43}$$

These equations can be used to derive the effect that a GW will have on the total output field. Assuming that GW with plus polarization and optimal direction reaches the interferometer, we have that the total output field will be given by Eq. 2.44:

$$\begin{aligned}(E_{out})_T = &- iE_0 e^{-iw_L t + 2ik_L L}[J_0(\Gamma)k_L L h(t) \\ &+ 2J_1(\Gamma) sin(2\pi \frac{\Delta L}{\lambda_{mod}}) cos(\Omega_{mod} t - \frac{4\pi L}{\lambda_{mod}})].\end{aligned} \tag{2.44}$$

Using that the power is proportional to $|(E_{out})_T|^2$ we find that there will be three terms contributing to the total measured power.

1. A term proportional to $h(t)^2$ that will be too small to be measured.
2. A term independent of $h(t)$ given by the modulation of the beam.
3. The mixed term that will be proportional to $h(t)$ oscillating at a frequency Ω_{mod}.

The linear term in h can be extracted from the output electric field by estimating the time-average of the product of the voltage on the output photodiode and a voltage $V_{osc} cos(\Omega_{mod} t - \alpha)$. The terms from the output field that are not oscillating as $cos(\Omega_{mod} t - \alpha)$ will be averaged to zero and only the linear term in h will remain.

2.5 Ground based interferometers

Following the principles explained in the previous sections, Fabry-Perots dedicated to the detection of GWs have been built in the past decades. The three main experiments are Advanced Virgo [28] (AdV - V1) located in Cascina, Italy; the two advanced LIGO [29] (aLigo) interferometers: the Hanford (H1) interferometer, located in Washington state, USA, and Livingstone (L1) interferometer, located in Louisiana, USA; and the Kagra [30] (K1) interferometer, located in Hida, Japan. The two aLigo interferometers have an arm length of 4km while AdV and Kagra have an arm length of 3km. A network of interferometers allows to detect GWs using coincidences between them to filter the noise. We will see more about this in Chapter 3.

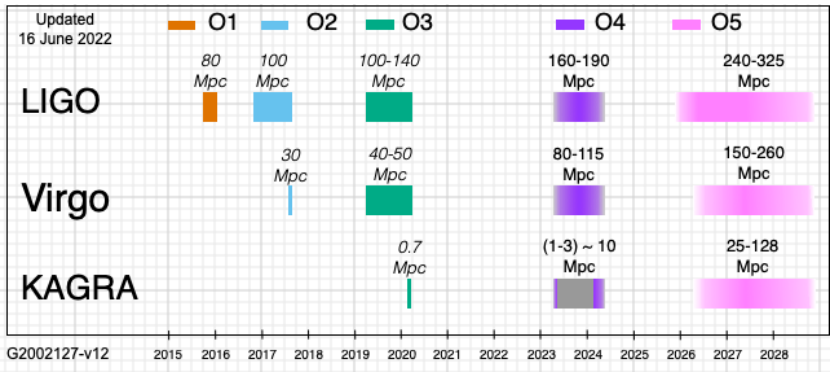

Figure 2.6: Observing plan with the expected sensitivity for the past and future observing runs. For O4, a one year observation period with one month commissioning break is planned for the aLigo and AdV experiments. Source: `https://observing.docs.ligo.org/plan/`

The anticipated observing schedule for the experiments is presented in Fig. 2.6. Apart from the four large inteferometers, there are also smaller instruments where many of the techniques and noise suppression methods have been developed and tested: GEO600 [31], a 600m arm-length interferometer located in Hannover, Germany, and TAMA300 [32], Tokyo, Japan, with an arm-length of 300m.

As we have seen in Sec. 2.4, the degree of precision required from these instruments is very high. This is achieved by installing photodiodes in critical positions to monitor the beam propagating within the cavity. The aforementioned photodiodes monitor the reflected and transmitted beams at different stages of the interferometer. An example of a real interferometer can be seen in Fig. 2.8, where the photodiodes, indicated by the letter 'B' followed by a number, are shown.

To improve the sensitivity of the detectors, the power circulating inside the interferometers is increased by a technique known as "Power recycling" [34]. This consists of creating a new cavity by adding a mirror, the Power Recycling mirror (PR), between the detector's input and the beam splitter.

Figure 2.7: Aerial photo of AdV (Left) and the Livingstone detector (Right) where the 3 and 4km long arms can be seen. Credit: aLigo and AdV.

The new mirror recycles the light that is reflected into the input laser by coherently reflecting it back to the interferometer, effectively increasing the power circulating inside the detector.

The sensitivity is further improved by introducing a mirror between the beam splitter and the output of the instrument, the Signal Recycling mirror (SR), to create a new cavity, the Signal Recycling Cavity (SRC) [35]. This mirror reflects the signal back into the interforemeter so that it adds coherently with more signal produced by the effect of the passing GW. The combination between PR and SR is known as dual recycling and was used by the aLigo interferometers during the third observation run (O3). Depending on the tuning of the SRC, dual recycling can be set to improve the sensitivity in different regions, broadband and narrowband. In broadband mode, the SRC is tuned to have maximum response at the carrier frequency, w_L. This will enhance the signal sidebands only if they are within the bandwidth of the cavity. In narrowband mode, the SRC is tuned to enhance one signal sideband, allowing it to improve the sensitivity to a narrower frequency region.

The sensitivity of these experiments is characterized by the spectral strain sensitivity, $\sqrt{S_n(f)}$, given by the ensemble average of the detector noise, n(t), as in:

$$\langle \tilde{n}^*(f)\tilde{n}(f')\rangle = \delta(f-f')\frac{1}{2}S_n(f) \tag{2.45}$$

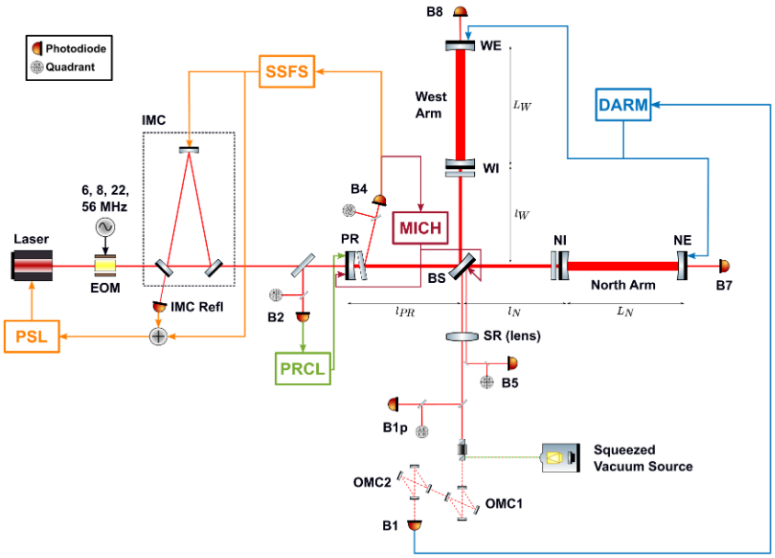

Figure 2.8: Schematics of AdV including photodiodes used to monitor the performance of the instrument. Taken from Ref. [33]

This sensitivity is constrained by different noise sources that affect the instrument at different frequencies. An example of the sensitivity of the AdV interferometer along with different noise sources is presented in Fig. 2.9. At low frequencies, the sensitivity of the interferometers is limited by seismic noise, while at high frequencies they are affected by quantum noise originating from the fact that a laser is composed by a discrete number of photons. This is discussed in detail in the following section.

2.6 Noise sources in ground based interferometers

In this section, we will systematically describe some of the noise sources that affect our instrument and limit its sensitivity. An example of a real sensitivity curve for AdV can be seen in Fig. 2.10.

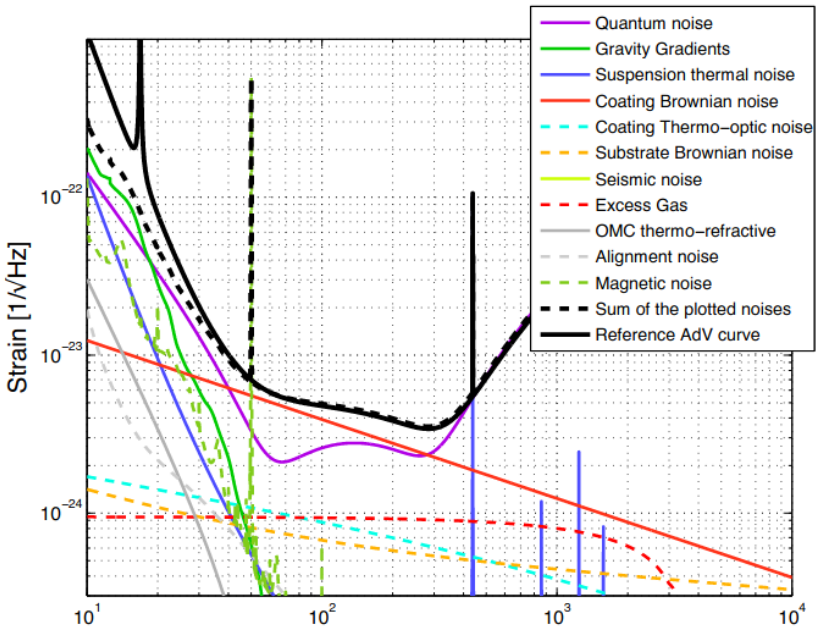

Figure 2.9: AdV reference sensitivity with some of the noise sources that contribute to it. Taken from Ref. [28]

2.6.1 Displacement noise

The displacement noise is produced by the movement of the components of the interferometer. At low frequency (<10Hz), this noise mostly has a seismic origin. However, it is not produced by a single source but several: from bad weather and human activity to micro-seismic background and earthquakes.

The displacement of the main components due to the micro-seismic background can be modelled as $x(f) \propto \frac{1}{f^2}$, which implies an effect on the sensitivity proportional to $1/f^2$. This is a severe limiting factor to the sensitivity of the interferometer and has to be attenuated. In the case of the AdV interferometer, the attenuation is achieved by hanging the key components from compound pendulums with multiple stages. Each pendulum, characterized by its resonance frequency f_0, attenuates the movement by a factor $\frac{f_0^2}{f^2}$ at frequencies $f \gg f_0$. Therefore, assuming a compound pendulum of N stages, an

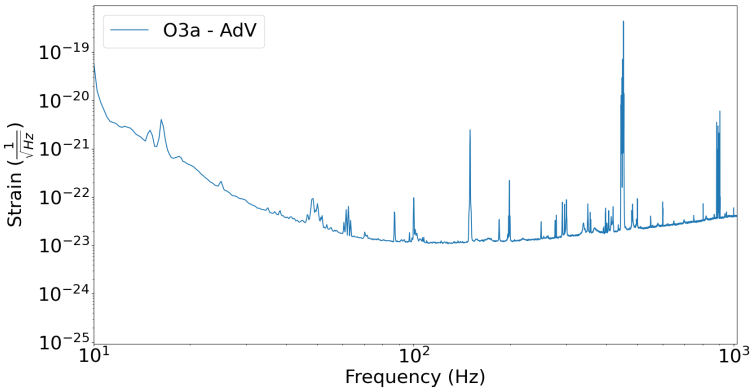

Figure 2.10: Real sensitivity of AdV for a small period of O3a.

attenuation of $\left(\frac{f_0^2}{f^2}\right)^N$ is achieved. However, the wires used to construct the pendulums have a resonance frequency that results in spectral lines, known as violin modes, at 300Hz (see Fig. 2.10).

Another source of noise at low frequencies is produced by the changes in the gravitational gradient around the test masses. These changes might appear, for example, from mass density fluctuations originated from a micro seismic, atmospheric turbulence or human activity. This is known as gravity gradient noise and limits the sensitivity of the interferometer at frequencies below $\approx 20Hz$. Due to the gravitational field being unknown, this noise can only be attenuated but not eliminated, providing an effective limit to the sensitivity of ground based detectors at low frequencies. Nevertheless, for current detectors, this is not the dominant contribution at those frequencies. Techniques designed to mitigate the contribution from this noise are implemented by monitoring the sources of gravitational gradients.

Thermal fluctuations within the key components are a source of noise, known as thermal noise, due to the vibrations they produce. These fluctuations originate from the high powered beam propagating within the cavities. In the case of the mirrors, the noise is produced by deformations or Brownian motions due to the temperature increase in the coatings. The suspension system is similarly affected by thermal fluctuations by inducing horizontal

and vertical motions. As a simple model, the displacement produced by these fluctuations can be modelled as:

$$x(w) = \frac{1}{w|Z(w)|}[4k_bTReZ(w)]^{1/2}, \tag{2.46}$$

where k_b is the Boltzmann's constant, T is the temperature and $Z(w)$ is known as the impedance.

2.6.2 Shot noise and radiation pressure

A fundamental contribution to the noise arises from the fact that our beam is composed by a discrete number of photons. This fundamental noise, also known as shot noise, will produce fluctuations in the observed power. A simple model of the effect of this noise on the sensitivity can be seen in Eq. 2.47:

$$S_n^{1/2}(f)|_{Shot} = \frac{1}{8\mathcal{F}L}\left(\frac{4\pi\hbar\lambda_L c}{\eta P_{bs}}\right)\sqrt{1+(f/f_p)^2}, \tag{2.47}$$

where P_{bs} is the recycled power at the level of the beam splitter and η is the efficiency of the photodetector used for detection. Due to the term f/f_p, the effect of shot noise at $f \ll f_p$ is almost constant. However, when the frequency becomes comparable to the pole frequency, it becomes linear in f. This noise is one of the limiting factors at frequencies higher than 1kHz.

Another uncorrelated source of noise originating from quantum light fluctuations is radiation pressure. When a beam hits an object, it exchanges momentum with the object. This exchange produces a pressure (radiation pressure) that affects the test masses and negatively impacts the sensitivity of the interferometer. In Eq. 2.48 a simple model of the strain due to radiation pressure can be seen:

$$S_n^{1/2}(f)_{Rad} = \frac{16\sqrt{2}\mathcal{F}}{M_{Mir}L(2\pi f^2)}\sqrt{\frac{\hbar P_{bs}}{2\pi\lambda_L c}}\frac{1}{\sqrt{1+(f/f_p)^2}}, \tag{2.48}$$

where M_{Mir} is the mass of the mirror. From this model we can observe that the radiation pressure will be more important at low frequencies and that it scales with the power inside the cavity. Comparing Eq. 2.48 and Eq. 2.47, it can be seen that these two mechanisms have an inverse dependence on the power.

Therefore, increasing the power will decrease the shot noise at higher frequencies while increasing the shot noise at lower frequencies. These two introduce a limitation to the sensitivity of the interferometer, known as optical readout noise, or quantum noise. If we add the two together, we can estimate the power of the laser needed to achieve the maximum sensitivity at a frequency of choice, f_0:

$$S_n^{1/2}(f)_{QN} = \frac{1}{L\pi f_0}\sqrt{\frac{\hbar}{M_{Mir}}}\left[(1+\frac{f^2}{f_p^2})+\frac{f_0^4}{f^4}\frac{1}{1+f^2/f_p^2}\right]^2 \tag{2.49}$$

$$f_0 = \frac{8\mathcal{F}}{8\pi}\sqrt{\frac{P_b s}{\pi\lambda_L c M}}.$$

It is not possible to optimize this noise for all frequencies, it can only be considered minimal at f_0 by imposing the condition given by Eq. 2.50:

$$\frac{f_0^2}{f^2} = 1+\frac{f^2}{f_p^2}, \tag{2.50}$$

where we will reach the optimal value of the quantum noise. This value is known as the standard quantum limit (SQL) and it can be shown that it is explained by the Heisenberg uncertainty principle due to the fact that there will be an uncertainty in the amplitude and the phase describing the light wave:

$$S_{SQL}^{1/2}(f) = \frac{1}{2\pi f L}\sqrt{\frac{8\hbar}{M_{Mir}}}. \tag{2.51}$$

This fundamental limitation can be, to some extent, overcome by a technique known as "Squeezing" [36, 37]. The basic notion is to prepare the light in such a way that the uncertainty is reduced in either the amplitude or the phase at the cost of increasing the other (squeezed states of light). The squeezed states of light are injected into the interferometer from the output side of the beam splitter (known as dark port) and are superimposed with the interferometer output containing the GW. Hence, the vacuum fluctuations are replaced with the squeezed vacuum states, allowing us to manipulate the measured uncertainty and reach below the SQL. An example can be seen in Fig. 2.11.

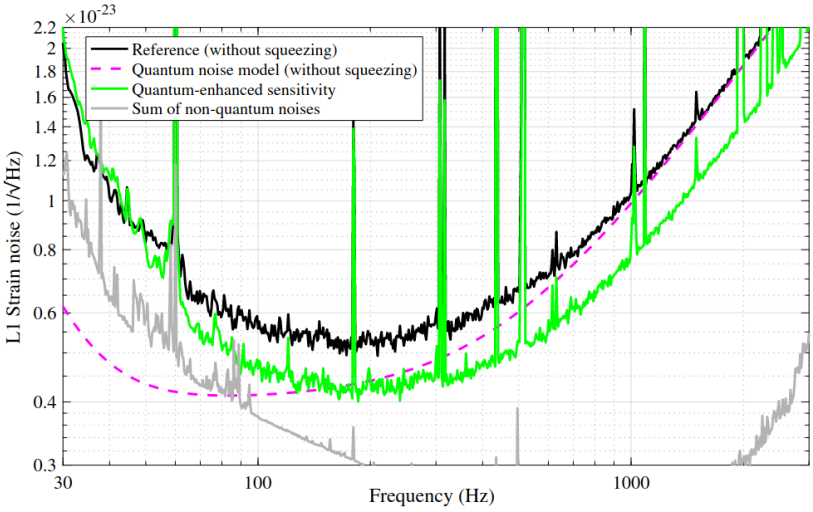

Figure 2.11: Squeezing in the L1 interferometer. The black line shows the sensitivity when no squeezing is applied while the green shows the amplitude during a period of O3. A model of the quantum noise is included in purple. A sum of all the remaining known (nonquantum) noises is included in grey. Taken from Ref. [38]

2.6.3 Environmental noise

Environmental noise [39] refers to the contributions to the noise from external environmental sources. It's composed of three different noises: scattered light, magnetic noise and acoustic noise.

Scattered light is produced by the light that scatters inside the interferometer and re-enters the main beam with an additional modulation given by the motion of the reflecting surface. The effect that it has on the sensitivity is modelled as in Eq. 2.52:

$$S_n^{1/2}(f) = C\left(K_\phi(f)\mathscr{F}\left[sin(\frac{4\pi}{\lambda}x(t))\right] + K_{\delta P/P}(f)\mathscr{F}\left[cos(\frac{4\pi}{\lambda}x(t))\right]\right), \tag{2.52}$$

where $\mathscr{F}$ is the Fourier transformation, $x(t)$ is the movement of the scatterer, C is the coupling constant that controls how much scattered light couples back into the interferometer and $K_\phi(f)$ and $K_{\delta P/P}(f)$ are transfer functions, from phase and amplitude respectively, to the noise of the interferometer.

These transfer functions are estimated by using optical models of the instrument, whereas C is estimated from the observed data. A characteristic pattern of scattered light are the arches that are produced in the time-frequency representation (spectrograms) of the sensitivity. An example of these arches can be seen in Fig. 2.12. This is a very difficult source to identify due to its non linear coupling to the sensitivity of the interferometer. In general, the identification is based on searching for arches as in Fig. 2.12 or excesses of power in photodiodes.

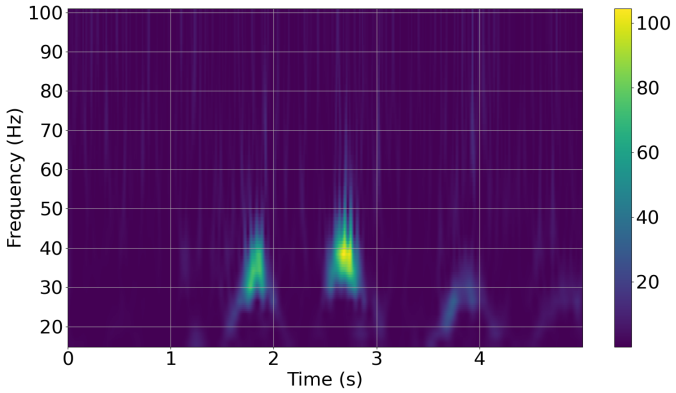

Figure 2.12: Scattered light seen in AdV during O3a. The characteristic arches from scattered light are clearly seen below 50Hz.

Magnetic noise is the noise produced by magnetic fields interacting with components of the instrument. In the interferometer there are several elements susceptible to magnetic fields, like magnets, coils or actuators that can be affected by an external magnetic field, producing movements in the test masses and worsening the overall sensitivity. The origin of these external magnetic fields can go from badly insulated cables to a lightning storm. Therefore, it is important to monitor the magnetic field in the interferometer by using magnetometers. This noise can also be a source of correlated noise among the interferometers, by means of Schumann resonances, which can create problems during the discrimination of signals from the background by enforcing coincidences between the interferometers (more in Chapter 3). This is particularly important for detecting GWs of stochastic origin due to the search being based in correlations between the interferometers.

Acoustic noise is generated by vibrations produced by acoustic waves, that is, vibrations produced by changes in density due to a sound wave propagating through the medium. This noise can originate from several sources, from humans to noisy machines. It can be decreased by improving the vacuum of the tubes and dampening possible sources using acoustic insulation materials.

During the commissioning break between the first segment (O3a) and the second segment (O3b) of the third observation run (O3), a campaign to mitigate sources of environmental noise was performed. This work can be seen in Appendix C and Ref. [39].

2.6.4 Other sources of noise

There are many more sources of noise inside the interferometers and all of them are modelled and understood as much as possible. Some of these sources are:

- Electronic noise due to conversion from digital to analog. The data acquisition systems are designed using digital-to-analogue converters that can introduce unwanted noise into the measured data due to the electronics. For that reason, these components are installed away from sensitive components of the interferometer.

- As seen in Fig. 2.8, in the detector there are several photodiodes monitoring the status of the instrument. Each of these is a possible source of noise due to dark currents.

- An interferometer has several mirrors that have to be aligned so that the beam can propagate inside the cavities. This introduces a source of noise produced by the uncertainty during the alignment of the system.

- The residual gas within the interferometer might produce fluctuations in the effective refractive index along the path. Due to these fluctuations, the optical path length will change (known as optical path noise), which might mask or imitate a signal.

2.7 Calibration and data quality

As we have seen in Sec. 2.5, the current interferometers are designed as Fabry-Perot interferometers with resonant cavities. This is achieved by using feedback loop [40, 41, 42, 43] (a loop in which part of the output is used as input for the next iterations). Therefore, the output strain has to be calibrated by taking into account all the readout electronics and actuation hardware affecting the mirrors through the suspension systems [44]. These effects are modeled as a function of the frequency for a total of three contributions: the actuation function $A(f)$, the sensing function $C(f)$ and the digital filters $D(f)$. A schematic representation can be seen in Fig. 2.13.

While the response of the digital filters is very well known, this is not the case for the sensing and actuation functions. The difference between the model and the measurement of these two functions is the origin of the calibration error and uncertainty. They have to be independently measured by using two beams produced by auxiliary lasers [45] (the photon calibrators), with known intensities modulated at known frequencies and amplitude, that will interact with the test masses producing radiation pressure and allowing to measure these functions. Once they are known, the strain is calibrated as in the following equation:

$$h(t) = \frac{1}{L}\left[C^{-1} * d_{err}(t) + A * d_{ctrl}(t)\right]. \tag{2.53}$$

These calibration measurements are regularly performed during an observation run along with additional monitoring of time dependent variables. Furthermore, calibration lines are continuously injected using the photon calibrators. These lines are removed from the strain [46]. Nevertheless, these lines wouldn't affect the search for CBC signals due to the data treatment techniques that will be explained in Chapter 3.

The calibrated data is processed to filter out periods of unreliable data, including periods with strong noise transients (glitches). These glitches manifest as excesses of power in the instruments and might be mistaken for GW signals. Therefore, possible paths for the noise to couple with the interferometer are monitored (e.g. environmental noise) and used to flag periods with an excess of noise. This is known as data quality investigations and the flagged segments are treated based on their severity, with the most polluted segments

being directly removed from the analysis. In the case of CBC, an example of a dangerous noise feature are lines of constant frequency, as they might be confused with the inspiral phase of a long signal [33, 47, 48, 49, 50, 51].

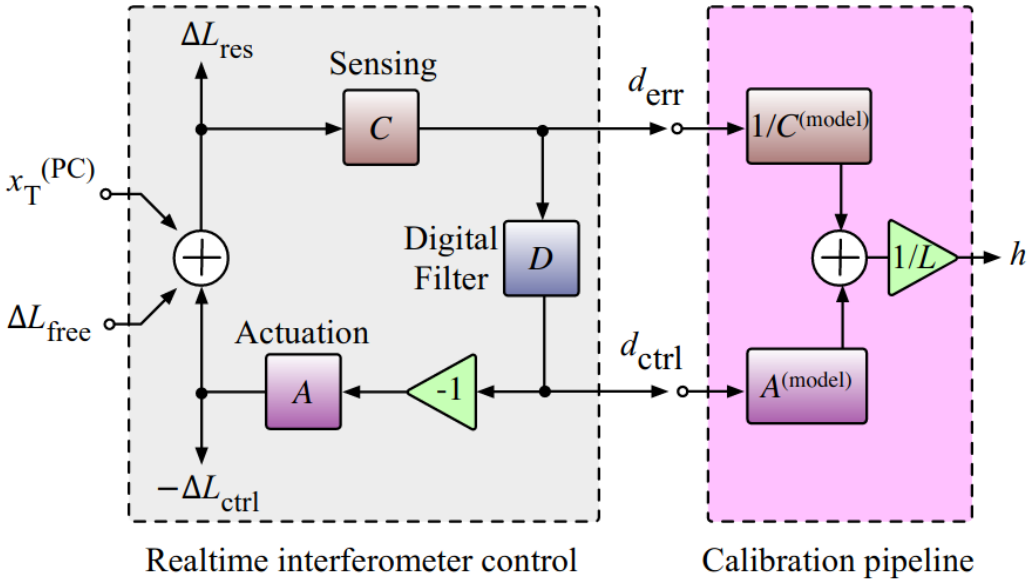

Figure 2.13: Schematic representation of the feedback loop used to control the differential arm length and the calibration pipeline. The unsuppressed change in differential arm length, ΔL_{free}, is the desired quantity and has to be estimated by using the models in A and C (purple box) and the data. This data is measured by the photodiodes and is given by ΔL_{res}. It corresponds to the differential arm length suppressed by the loops and can be used to obtain the error signal. The error signal, d_{err}, is the product of the sensing function, C, and ΔL_{res}. This error signal is passed through the digital filters and used as input to the actuators controlling the differential arm length. $x_T^{(PC)}$ denotes where the photon calibrators are used to test the mass mirrors. The output of the calibration pipeline is the desired strain signal $h(t) = \frac{\Delta L_{free}}{L}$. Taken from Ref. [52]

2.8 Instrumented baffle

As was discussed in Sec 2.6.3, scattered light is one of the noise sources affecting the current interferometers. Within the arms, scattered light is produced mainly by light scattered off the mirror surfaces. This effect was mitigated by installing low-reflective, low-scattering baffles, either suspended around the mirrors or connected to the ground in the vacuum tubes, that were able to absorb 99.5% of the scattered light. No active monitor of the scattered light inside the arms has been implemented.

In addition, Virgo lacks the information from photosensors, close to the test masses, that in the case of LIGO proved to be very useful in the pre-alignment of the arms. This motivated the design of instrumented baffles, to which photosensors are added, for O5. These baffles will be installed surrounding the test masses. A first instrumented baffle was installed around the end mirror of the Input Mode Cleaner (IMC) cavity at Virgo in April 2021. This baffle served as a demonstrator of the technology designed to instrument the baffles in front of the main mirrors. The information from the baffles will allow a better understanding of the scattered light distribution at low angles and monitor the contamination of the mirror surfaces. Furthermore, it will facilitate the pre-alignment and fine-tuning of the parameters of the interferometer. More information can be found in Appendix A.

Chapter 3

Data preprocessing, searches and parameter estimation

In this chapter, we discuss the methods used to analyze and process the data produced by the interferometers. In the first place, the basic notions needed to characterize the noise from the interferometers are explained. The technique used to extract the GW produced by CBC event is discussed and used to construct a search pipeline. Finally, the significance is assigned to each event detection.

3.1 Gaussian noise, stationarity and whitening

In Sec. 2.6 we have seen some of the noise sources that affect our interferometer and, in this section, we will formalize the treatment of the data from a statistical point of view [1, 52, 53]. The objective is to create a framework that can be used for the extraction of GWs contained in the data out of the noise.

The interferometer operates in continuous mode with a data taking sampling frequency of 16384Hz. The data is described using time series: arrays of data points with a time assigned. We can describe the interferometer output as a discrete time series, usually denoted as $\mathbf{n(t)}$, with samples $n_i = n(t_i)$, length N and that follows a joint probability distribution $p(\mathbf{n})$. This allows us to introduce statistical terms as the mean, $\hat{n}$, and covariance, C_{ij}.

There is an intrinsic limitation in the definition of the covariance. In our time series we have one and only one data point per time, which is not enough to estimate the covariance matrix, as in Eq. 3.1:

$$C_{ij} = \frac{1}{K-1}(n_i - \hat{n})(n_j - \hat{n}), \tag{3.1}$$

where i, j correspond to two time indexes and K is the total number of samples for that combination. It can be seen that when $K = 1$, that is, only one sample, the matrix is not defined. In general, the covariance matrix of a time series can only be estimated if for each point we have a vector and not a scalar, as is the case here. This can only be estimated if we assume a model for the noise distribution or if the features of the noise do not depend on time.

A common assumption is that the noise follows a multivariate Gaussian distribution, as in Eq. 3.2. A noise following this joint distribution is called Gaussian noise [54]:

$$p(\mathbf{n}) = \frac{1}{det(2\pi C)^{\frac{1}{2}}} exp\left(\frac{-1}{2}\sum_{ij}(n_i - \hat{n})(n_j - \hat{n})C_{ij}^{-1}\right). \tag{3.2}$$

The hypothesis of Gaussian noise will be used later on to extract GWs from CBC events by testing if the residual distribution, that is, data from which we have subtracted a CBC signal with parameters θ (modelled as in Sec. 1.3), follows this distribution.

Furthermore, during the analysis, the data is also assumed to be wide-sense stationary, or just stationary. A time series is called stationary if the mean and variance are finite and the covariance matrix only depends on the time lag between two points. In particular, this also means that the autocorrelation function[1], $R(t_1, t_2)$, depends only on the time lag, $\tau = t_1 - t_2$ (see Eq. 3.3). It shouldn't be confused with the concept of strictly sense stationary[2], which is a stronger assumption over the behaviour of the data:

$$R(t_1, t_2) = R(t_1 - t_2, 0) = R(\tau). \tag{3.3}$$

[1]Defined as the correlation of a signal with itself as a function of the delay.

[2]We say that a process is strictly sense stationary when the probability distribution does not change with the time lag τ.

Under the hypothesis of stationary noise, we can define the power spectral density (PSD), $S_n(f)$, as the Fourier transformation of the autocorrelation function, as seen in Eq. 3.4. While the definition given in Eq. 2.45 for the power spectral density is in general not well defined (it's not ensured that $n(t) \to 0$ when $t \to \pm\infty$), Eq. 3.4 doesn't have that problem due to the autocorrelation function going to zero as $\tau \to \pm\infty$:

$$S_n(f) = 2\int_{-\infty}^{\infty} R(\tau)e^{i2\pi f\tau}d\tau. \tag{3.4}$$

Under these conditions, the covariance matrix in the frequency space is $C_{ij} = \delta_{ij}S_n(f_i)$. This allows to introduce the term of white noise as $C_{ij} = \delta_{ij}\sigma^2$, that is, constant power at all frequencies.

Unfortunately, as we have seen in Sec. 2.6, the noise is not always stationary. At short time intervals ($\backsim$a few hundreds of seconds), the hypothesis holds and the noise can be considered stationary. In this case, the Fourier transformation becomes a powerful tool to characterize and handle our data. In particular through the usage of Fast Fourier Transformation (FFT) [55, 56, 57]. Furthermore, if the data is Gaussian, each frequency bin will be uncorrelated.

This introduces another complication given by the fact that a Fourier transformation requires the data to be periodic. To tackle this problem, window functions [58, 59] are used. A window function (see Fig. 3.1)is a function that is zero outside an interval and symmetric around its center. The importance of these functions is that they can be used to make the data periodic. Not applying a window function, which is equivalent to using the "Boxcar" window seen in Fig. 3.1 (left), before the FFT can produce a phenomenon known as spectral leakage [58] (see right Fig. 3.1), which will create fake correlations. These issues appear due to the Fourier transformation of a boxcar window, also known as step function, being the *sinc* function, $sinc(x) = \frac{sin(x)}{x}$, which decreases as $1/f^2$. The $1/f^2$ contribution dominates the whole spectrum and "leaks" at higher frequencies, producing incorrect results and fake correlations.

Using these window functions and the previous assumptions about stationarity of the data, we can estimate the power spectral density. Using a single FFT to estimate the PSD is sub-optimal due to the large variance of the resulting PSD. Therefore, more sophisticated methods are used for its

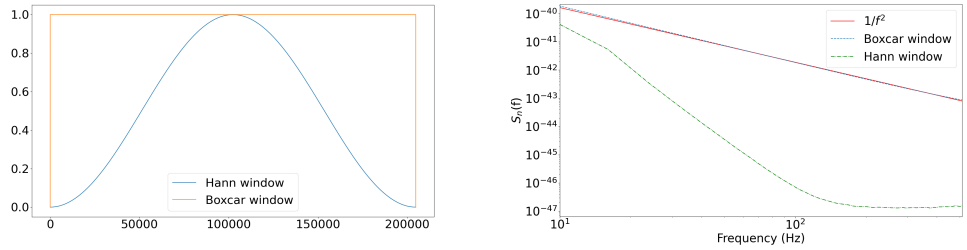

Figure 3.1: (Left) Examples of two window functions. (Right) Spectral leakage is produced by not using a window function to get periodic data.

estimation, with Welch's method being one of the most commonly used [60]. This method consists of splitting the initial segment into M segments (with D overlapping points) and estimating the spectral density of each segment by means of FFTs after being windowed. The resulting spectral densities are then averaged to reduce the variance of the individual estimations. This PSD can now be used to "whiten" [61, 62] our data, that is, transforming our data to equally powered frequency bins. This step is needed to suppress noise artifacts so that the GW can be extracted.

The idea is to transform our data to the frequency space by performing FFTs, divide by the square root of the PSD ($S_n^{\frac{1}{2}}(f)$, also knonw as the Amplitude Spectral Density or ASD), and go back to the time domain by using an inverse fast Fourier transformation. Schematically, all the steps are shown in Eq. 3.5:

$$h(t) \xrightarrow{\text{FFT}} h(\tilde{f}) \xrightarrow{\text{Whiten}} \tilde{h}_W(f) = \frac{h(\tilde{f})}{S_n^{\frac{1}{2}}(f)} \xrightarrow{\text{iFFT}} h_W(t). \tag{3.5}$$

Finally, using FFTs, we can produce spectrograms which allow us to get a visual representation of our data in time-frequency space. These are the basic elements to describe and treat the data from the interferometers and will be used in the following sections to extract the GW signal from the data.

3.2 Gating and noise subtraction

The data used not only have to be calibrated, as explained in Sec. 2.7, but also treated to deal with glitches and non Gaussian noise that pollute it [63, 64]. The most simple technique is known as "gating" (see Fig. 3.2) in which the glitch is smoothly set to zero using a window function. Gating is used in GWs searches to remove glitches with an excess of energy above a certain threshold before proceeding with the search. The threshold is carefully selected so that the likelihood of gating real GWs signal is negligible. Even though this method is useful in this context, it introduces problems in more delicate studies, like parameter estimation of the source properties. The reason is that these studies assume the Gaussianity of the data.

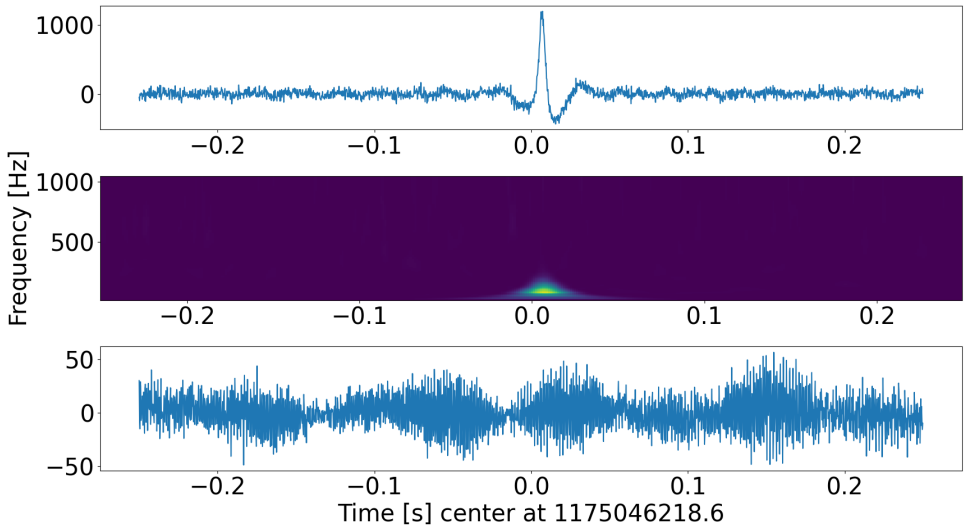

Figure 3.2: Glitch found in real O2 data gated so that it can be used for a search.

Based on this, new techniques for noise subtraction have been developed. The main idea behind these new methods is to model the offending glitch and subtract it from the data. This method was successfully used to remove the glitch in L1 for GW170817 [65] and is now extensively used in the recent catalogs [11, 66]. Furthermore, noise subtraction is also used to remove known noise sources, like calibration lines, electric power grid lines or noise coming from sources monitored by witness sensors.

3.3 Search of CBC gravitational waves

Using the techniques and elements defined in the previous sections, we extract the signal of CBC events from the data. These signals are typically buried underneath the noise, as seen in Fig. 3.3. In this section we will explain the technique and elements used in searches for GW signals from CBC events.

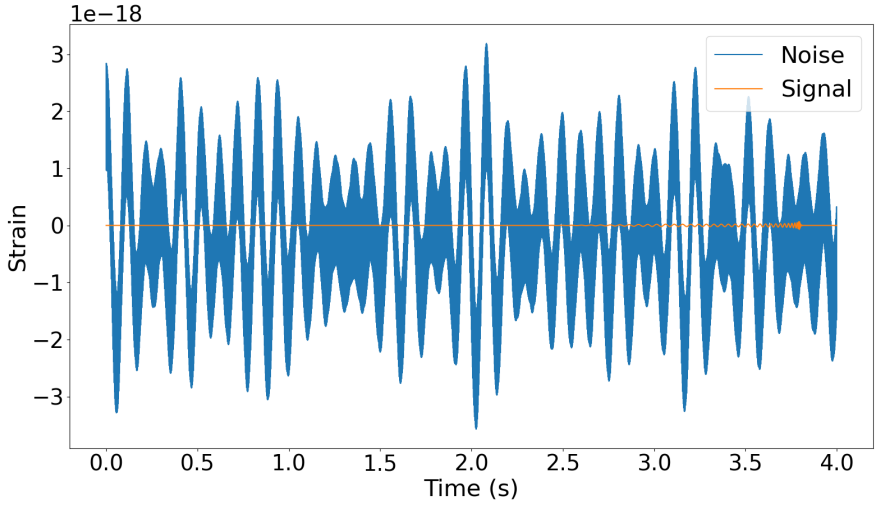

Figure 3.3: CBC signal with component masses (30,30)$M_{\odot}$ at 10Mpc under real noise from L1 interferometer.

3.3.1 Matched filter

Matched-filter [67] is a technique for searching for a known shape (template) from a segment of data with larger noise. This is a well known technique in several fields, like radio engineering. In this section we will prove this method is optimal for the search for a signal when the noise is Gaussian [1]. We start by correlating our data with a signal in it, $d(t)$, with a filter function, $G(t)$, as in Eq. 3.6:

$$\hat{d} = \int_{-\infty}^{\infty} dt d(t) G(t), \tag{3.6}$$

and we derive the filter function that maximizes the SNR for this data knowing the form of the GW, $h(t)$.

The contribution from the signal inside the noise is given by Eq. 3.7:

$$\begin{aligned} S &= \int_{-\infty}^{\infty} dt\langle d(t)\rangle G(t) = \int_{-\infty}^{\infty} dt\langle h(t)+n(t)\rangle G(t) = \\ &= \int_{-\infty}^{\infty} dt[\langle h(t)\rangle + \langle n(t)\rangle]G(t) = \int_{-\infty}^{\infty} dt\langle h(t)\rangle G(t) = \\ &= \int_{-\infty}^{\infty} df\langle \tilde{h}(f)\rangle \tilde{G}^*(f) = Re\int_{-\infty}^{\infty} df\langle \tilde{h}(f)\rangle \tilde{G}^*(f), \end{aligned} \tag{3.7}$$

where we use $\langle n(t)\rangle = 0$ (which can be assumed without loss of generality due to always being possible to subtract the average noise from our data) and Parseval formula[3] to do the transformation to frequency domain. The noise contribution is given by the rms value when there is no signal present and takes the form seen in Eq. 3.8:

$$\begin{aligned} N^2 &= [\langle \hat{d}^2(t)\rangle - \langle \hat{d}(t)\rangle^2]_{h=0} = [\langle \hat{d}^2(t)\rangle] = \\ &= \int_{-\infty}^{\infty} df \frac{1}{2} S_n(f)|\tilde{G}(f)|^2 = Re\int_{-\infty}^{\infty} df \frac{1}{2} S_n(f)|\tilde{G}(f)|^2, \end{aligned} \tag{3.8}$$

where we have used the relation between the auto correlation function and the power spectral density. Hence, the signal-to-noise ratio (SNR) is:

$$\frac{S}{N} = \frac{Re\int_{-\infty}^{\infty} df\langle \tilde{h}(f)\rangle \tilde{G}^*(f)}{\left[Re\int_{-\infty}^{\infty} df \frac{1}{2} S_n(f)|\tilde{G}(f)|^2\right]^{1/2}}. \tag{3.9}$$

The SNR can be written in a more compact form using the inner weighted product defined in Appendix E. This inner product is defined in such a way that regions in which the detector is more noisy will be suppressed by a larger weight. Using this product, we obtain Eq. 3.10:

$$\frac{S}{N} = \frac{(u|h)}{(u|u)^{1/2}}, \tag{3.10}$$

where $u(t)$ is a function such that its Fourier transformation is equal to:

$$\tilde{u}(f) = \frac{1}{2} S_n(f)\tilde{G}(f).$$

[3] $\int_{-\infty}^{\infty} |h(t)|^2 dt = \int_{-\infty}^{\infty} |h(f)|^2 df$

Using a geometrical interpretation in two dimensions, we know that the maximum is reached when the two vectors are parallel or, in other words, they differ by a constant. This translates into the solution seen in Eq. 3.11:

$$\tilde{u}(f) = \frac{1}{2} S_n(f) \tilde{G}(f) = \alpha \tilde{h}(f) \rightarrow \tilde{G}(f) = \kappa \frac{\tilde{h}(f)}{S_n(f)}, \tag{3.11}$$

which is the optimal filter for Gaussian noise [54] and, from a physical point of view, favors the frequency regions from the template which are less affected by the real noise, $S_n(f)$. Furthermore, in the case of white noise, we have that the filter reduces to simply the template, an expected outcome due to all the frequencies weighing the same. Introducing this filter in Eq. 3.10, we get the expression, as seen in Eq. 3.12, for the so-called optimal SNR:

$$\left(\frac{S}{N}\right)^2 = (h|h) = 4 \int_0^\infty \frac{|\tilde{h}(f)|^2}{S_n(f)} \tag{3.12}$$

3.3.2 Statistical approach to matched-filter

In the previous section, it was shown that the optimal filter to search for a signal is the signal itself. Therefore, in the search for CBC events [52], the filter is constructed using waveform models (See Sec. 1.3). The data is resampled to a lower frequency, e.g. 2048Hz, and split into segments of equal duration. The length of these segments (e.g: 512s) is selected by considering several factors, like the maximum duration of the searched signals. Each segment is correlated with the template and triggers are found based on the SNR distribution. A detailed description of this procedure is given in this section.

Let's assume that we have a period of data, $d(t)$, which might have a real signal in it. Given that GWs signals from CBCs are described by $h(t;\theta)$ for some parameters θ [68], the residual $r(t) = d(t) - h(t;\theta)$ should follow the model of the noise. Therefore, we can estimate how likely the data is to have a signal by estimating the probability of the residual following the noise distribution. In the case of Gaussian noise [54] we can write this likelihood as in Eq. 3.13:

$$L(d|h;\theta) = \frac{1}{det(2\pi C)^{1/2}} exp\left(-\frac{1}{2}\chi^2(d, h(t;\theta))\right), \tag{3.13}$$

in which C is the correlation matrix and $\chi^2(d,(t;\theta))$ is defined as in Eq. 3.14:

$$\chi^2(d,h(t;\theta)) = r\cdot C^{-1}\cdot r = \sum_{I,J,k,m}(d_{Ik}-h_{Ik})C^{-1}_{(Ik)(Jm)}(d_{Jm}-h_{Jm}), \quad (3.14)$$

where I and J iterate over the different detectors and k and m the data samples. In the case of uncorrelated noise between all the detectors in the network, we recover the previous definition for the covariance matrix, $C_{(Ik)(Jm)}(f) = \delta_{IJ}S^I_{km}(f)$. Furthermore, under the condition of stationary noise, we have that the correlation matrixes are diagonal and thus:

$$\chi^2(d,h(t;\theta)) = (d(t)-h(t;\theta)|d(t)-h(t;\theta)) = (r(t;\theta)|r(t;\theta)).$$

However, even if in most cases the noises are uncorrelated, there are noise sources that might be present in all the interferometers, like the ones produced by Schumann resonances or earthquakes (see environmental noise in Sec. 2.6.3).

This likelihood can be used to construct a hypothesis test.

H_0 : Only noise in the data.
H_1 : There is a signal with parameters θ.

And in terms of the likelihood we have that:

$$L(d|H_0;\theta) = L_0(d;\theta)$$

$$L(d|H_1;\theta) = L_1(d;\theta),$$

which can be used to construct a standard likelihood ratio statistic as in Eq. 3.15:

$$\Lambda(d;\theta) = \frac{L_1(d;\theta)}{L_0(d;\theta)}. \quad (3.15)$$

Assuming stationary Gaussian noise, the likelihood is given by Eq. 3.13 and the log ratio test takes the form of Eq. 3.16:

$$
\begin{aligned}
log\Lambda(d;\theta) &= log\left(\frac{exp\left(-\frac{1}{2}\chi^2(d,h(t;\theta))\right)}{exp\left(-\frac{1}{2}\chi^2(d,0)\right)}\right) = \\
&= -\frac{1}{2}(d(t)-h(t;\theta)|d(t)-h(t;\theta)+\frac{1}{2}(d(t)|d(t)) = \\
&= \frac{1}{2}((d(t)|d(t))-(d(t)|d(t))-(h(t;\theta)|h(t;\theta))+2(d(t)|h(t;\theta))) = \\
&= (d(t)|h(t;\theta))-\frac{1}{2}(h(t;\theta)|h(t;\theta)),
\end{aligned}
\tag{3.16}
$$

where the term $(d(t)|h(t;\theta))$ corresponds to the unnormalized matched-filter statistic seen in the last section. We are interested in the maximum of this log likelihood ratio and it's clear that it's reached when the residual is minimized. In the case of a signal, this minimization occurs when it's well described by the parameters θ. Using now that we can write the signal as in Eq. 3.17 [69]:

$$
h(t;\theta) = Ap(t;\theta)cos(\phi) + Aq(t;\theta)sin(\phi), \tag{3.17}
$$

where A is the amplitude, inversely proportional to the distance, ϕ is the phase of the signal in the detector and $p(t;\theta)$ and $q(t;\theta)$ are two orthogonal functions normalized with respect to the inner weighted product. These two functions are obtained by decomposing the GW signal in the in-phase (cosine) and quadrature (sine) components[4]. Plugging this into Eq. 3.16 we have that the likelihood ratio can be expressed as in Eq. 3.18:

$$
\begin{aligned}
log\Lambda(d;\theta) &= A\rho(t;\theta)cos(\phi-\psi)-\frac{1}{2}A^2 \\
\psi &\coloneqq arctan\frac{(d|q(t;\theta))}{(d|p(t|\theta))} \\
\rho(t;\theta) &\coloneqq \sqrt{(d|q(t;\theta))^2+(d|p(t;\theta))^2},
\end{aligned}
\tag{3.18}
$$

where $\rho(t;\theta)$ is the SNR time series for a template with parameters θ and peaks at times in which a signal like the template is likely to be.

[4]Any sinusoid signal with phase modulation can be decomposed in two components with an offset of $\pi/2$. This can be seen from the fact that $sin(A+B) = sin(A)cos(B) + sin(A+\pi/2)sin(B)$.

These peaks correspond to the mergers times. Finally, the log likelihood ratio is maximized when:

$$cos(\phi - \psi) = 1 \iff \phi = \psi$$
$$\rho(t;\theta) = A,$$

reaching the maximum value at $\frac{1}{2}\rho^2(t;\theta)$. In general, it is more convenient to express the time series as a complex time series, seen in Eq. 3.19:

$$z(t;\theta) = 4\int_0^\infty \frac{\tilde{d}(f)\tilde{p}^*(f;\theta)}{S_n(f)} e^{2\pi i f t} df, \tag{3.19}$$

where $\tilde{p}^*(f;\theta)$ is the Fourier transformation of the in-phase waveform. This complex time series, $z(t;\theta)$, is defined in such a way that $\rho = |z(t;\theta)|$ and $\phi = arg(z)$ [70].

There is a wide range to cover in the signal parameter space. Therefore, we need to create a template bank to search for signals from CBC events. The data segments mentioned previously are correlated with the elements from the template bank and the complex SNR time series per template and segment, given by Eq. 3.19, is estimated. The triggers are extracted from these distributions by setting a SNR threshold (e.g. 4). This step has to be performed for all the interferometers in the network, allowing us to define the network SNR as:

$$\rho^2_{Net} = \sum_{I \in Int} \rho^2_I.$$

where Int contains all the online detectors at the time of the trigger. An example of a SNR time series can be seen in Fig. 3.4, along with the effect of using a template different from the signal contained in the data.

3.3.3 Creating a template bank

Ideally, the whole signal parameter space would be covered with a dense grid so that the maximum SNR is reached. In practice, this is not a feasible option. A template bank so large would be impossible to manage and use with limited computed resources. Therefore, a more careful approach is needed.

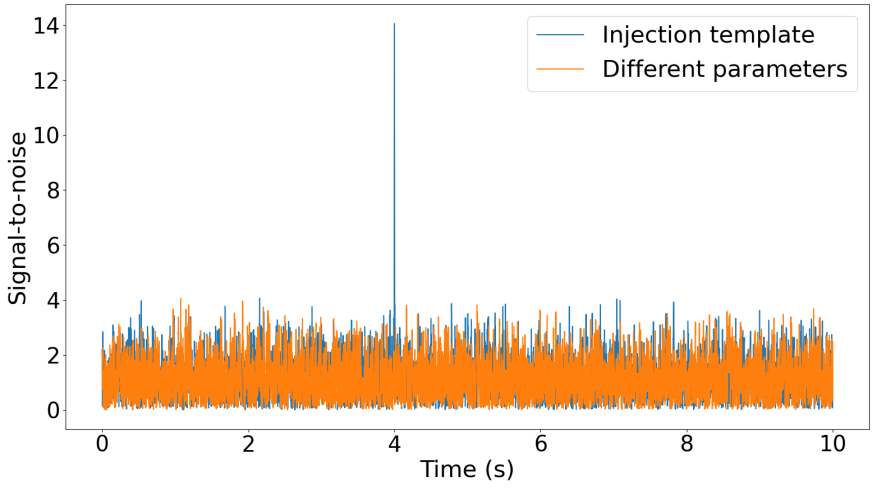

Figure 3.4: SNR time series using two different sets of parameters, the same as the injection and different ones.

The template bank is created by using the overlap and match between templates [71, 72, 73, 74, 75] to estimate how much SNR is lost by using two close-by (in parameter space) templates.

We define the overlap between two templates as seen in Eq. 3.20:

$$O(h_1, h_2) = (\hat{h}_1|\hat{h}_2) = \frac{(h_1|h_2)}{\sqrt{(h_1|h_1)(h_2|h_2)}}. \tag{3.20}$$

In general, it can be assumed that our templates are normalized with respect to the inner weighted product, that is, $(h_1|h_1) = 1$ and $(h_2|h_2) = 1$. The overlap quantifies how much of the first signal we would recover using the second, h_2, as template for the matched-filter method. It is important to note that all the waveforms related by time/phase offset are described by the same template. Therefore, we define the match as seen in Eq. 3.21:

$$M(h_1, h_2) = max_{\phi,t}(h_1|h_2(\phi, t)), \tag{3.21}$$

allowing us to define the mismatch as how much we are missing to reach the maximum possible overlap of one. That is, we define the mismatch between

the two templates as $1 - M$. This value quantifies the fraction of SNR we are losing by using h_2 to search for h_1.

Assuming that we have a grid of templates, we can estimate the match between all the templates and quantify how much SNR we are willing to lose due to the finite dimension of our grid. In particular, it's convenient to use a term called the "Fitting Factor" for a template h_s, defined in Eq. 3.22, that estimates the maximum agreement between that template and the rest of the bank:

$$FF(h_s) = max_{h \in \{h_n\}} M(h_s, h). \tag{3.22}$$

We can now impose limits on this fitting factor and construct a template bank in which there is always an overlap between two signals bigger than our threshold. These elements are used in dedicated algorithms to produce the template banks that will be used during the search. An example of a very simple template bank covering the mass range between $2M_\odot$ and $3M_\odot$ can be seen in Fig. 3.5. The distribution of templates within the bank is such that the minimum fitting factor is at least 0.97. The template bank shows two main features: the limits given by the condition $m_1 > m_2$ and a gap. The gap is an artifact produced by the algorithm used for the generation of the bank.

3.3.4 Signal consistency tests

The template bank defined in the previous section along with the matched-filter technique can be used for the search for GWs. However, as we have seen before, the noise is not always Gaussian and matched-filter won't be the optimal detection statistic. In fact, noise can produce transients that will trigger with some templates with high SNR. This can be alleviated by performing data quality activities, as seen in Sec. 2.7. There are more techniques that can be employed to further improve the significance of our search.

An important technique is the imposition of tests to characterize the deviation of the data from a model with signal in it, as explained in detail in Refs. [76, 77, 70, 69, 78]. These tests are called signal consistency tests and in CBC they allow us to discard many glitches from our triggers. The basic idea of these methods is to test the Gaussian behaviour of the residual after subtracting our template from the data.

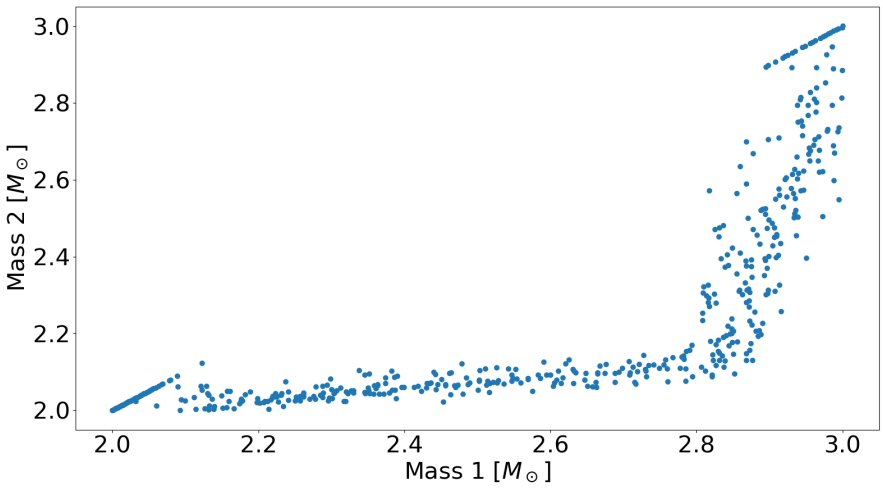

Figure 3.5: The template bank covers the mass range between $2M_\odot$and $3M_\odot$. The templates are distributed in such a way that, for a given template, there is always at least a second template with a match of 97%.

If triggers are still found, this indicates that the template is not properly describing the data and should be weighed down by a factor.

A test used by current search pipelines is the chi-square test [69]. This test uses the fact that the accumulation of SNR over time and frequency is different for glitches and real signals. The SNR of a signal will accumulate along the duration in a way consistent with the waveform. This is tested by dividing the template into 'p' orthogonal subtemplates at different frequency intervals, such that the SNR of each template over Gaussian noise is the same on average. Therefore, the signal to noise ratio, z, will be the sum of the signal to noise ratios from all the frequency bands, z_i:

$$z = \sum_{j=1}^{p} z_j. \tag{3.23}$$

From this definition we have that the quantity defined as follows:

$$\Delta z_j = z_j - \frac{z}{p}, \tag{3.24}$$

will have zero mean and sum zero by definition. We use these to define the χ^2 time frequency discriminator statistic as:

$$\chi^2 = \chi^2(z_1, ..., z_p) = p \sum_{j=1}^{p} (\Delta z_j)^2, \tag{3.25}$$

which will have an expected value of $\langle \chi^2 \rangle = p - 1$ given that the expected value of $(\Delta z_j)^2$ is:

$$\langle (\Delta z_j)^2 \rangle = \frac{1}{p}(1 - \frac{1}{p}). \tag{3.26}$$

In conclusion, if the signal matches the given template the reduced χ^2, defined as $\chi^2_r = \frac{\chi^2}{p-1}$, will be close to one. In any other situation the output value of this test can then be used to reweight our detection statistic so that cases in which the hypothesis is not true have a down-weighted SNR. This derivation was made in the simplest case of a signal defined by only one polarization and assuming that the noise is stationary. In general the signal is a linear combination of two polarizations and, in that case, the degrees of freedom are $2p - 2$. The reduced χ^2 is then given by:

$$\chi^2_r = \frac{\chi^2}{2p - 2}. \tag{3.27}$$

Finally, a good understanding of the noise of our interferometers [79] and the GWs models allows us to declare single detector detection. However, the most stringent signal consistency test is to require the GW to be detected by matched-filter in at least two independent interferometers [9], the so-called coincidence triggers. These triggers should be produced by templates with consistent parameters and the difference in the time of arrival should be lower than the maximum time of flight. This limits the detection of GWs to times in which there is at least two interferometers working.

3.3.5 False Alarm Rate (FAR)

If all the tests explained before are successfully passed by our candidate, the probability of it being produced by a noise transient is very small. However, it's necessary to estimate how significant this trigger really is. The False Alarm Rate (FAR) is introduced for this purpose. Formally the FAR for a candidate is defined as the rate of triggers due to noise with a detection

statistic bigger or equal than our candidate [80, 81]. In other terms, for a certain detection statistic X the FAR assigned to a candidate, with a value of the statistic x, is given by Eq. 3.28:

$$FAR(x) = \frac{N(X \geq x)}{T_{obs}}, \tag{3.28}$$

where $N(X \geq x)$ is the number of triggers produced by noise with detection statistic bigger than x and T_{obs} is the total observation time (usually expressed in years). Unfortunately, an analytical expression for $N(X \geq x)$ is in general not know, and we need to estimate it empirically. The simplest approach consists on time shifting the data of the different interferometers and re-doing the search again over the new time shifted data. If this shift is bigger than the possible time-on-flight for real signals (including timing errors, $\approx 150ms$ is the minimum shift allowed), then all the new triggers found by the search will be produced by noise and can be used to estimate the distribution. The time-shift is repeated for different shifts until we achieve the desired T_{obs}.

With this method we can obtain a map between the detection statistic we are using and the FAR. This map can be used to assign significance to all of our triggers. An example can be seen in Fig. 3.6, where we have estimated this distribution for a machine learning discriminant that we will use in Chapter 5. This plot shows two features originating from limits in the FAR: below discriminant values of $\backsim 0.1$ all the data used for the estimation is flagged as noise and the FAR remains constant. Instead, the constant line above $\backsim 0.95$, arises from the fact that the minimum FAR is bounded by the total observation time. From Eq. 3.28 it can be extracted that the minimum FAR is given by: $FAR_{Min} = \frac{1}{T_{obs}}$. Therefore, values below that point would require to extrapolate from the distribution.

A limitation of this approach is that it does not take into account the possible noise contributions that are common to all the interferometers, like the ones produced by Schumann resonances. It is thus fundamental to validate all the possible candidates by examining all the possible known noise sources using the witness channels.

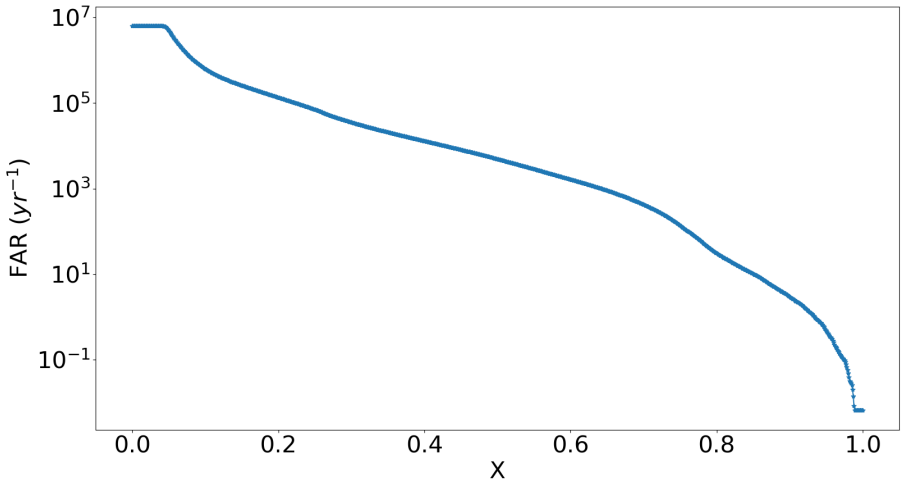

Figure 3.6: Map between the FAR and a detection statistic based in machine learning techniques.

A detection is claimed by selecting a threshold value for this FAR. For example, the O2 catalog paper used a FAR threshold of 0.01 yr^{-1}, or 1 per 100 years, for detection. A trigger with a FAR value below this threshold is considered a confident detection.

3.3.6 Volume-time and detector efficiency

Using the previous definition, we can estimate the efficiency of the search pipeline by injecting simulated signals following a given population model. As we have seen in Sec. 2.3, the efficiency of the detector is not uniform over the whole sky. Furthermore, the amplitude of the GW is also affected by the inclination of the source with respect to us. This translates into an effective sensitive volume modulated by the efficiency of the detector and the target population [82]. This sensitive volume is given by Eq. 3.29:

$$\langle VT \rangle = T \int dz d\theta \frac{dV_c}{dz} \frac{1}{1+z} \phi(\theta) \epsilon(z, \theta), \tag{3.29}$$

where T is the observation time, $\frac{dV_c}{dz}$ is the differential of co-moving volume with redshift, $\phi(\theta)$ is our population model and $\epsilon(z,\theta)$ is the efficiency of the search for parameters θ at redshift z. The efficiency is not generally known and has to be empirically estimated by using injections. The integral is solved by using Monte-Carlo integration.

Firstly, a population model, $\phi(\theta)$, from which we will extract the parameters of the injections that will be inserted into the data is chosen. For example, a population uniform in component masses is drawn from a probability density function given by:

$$\phi(m_1^d, m_2^d) = \frac{1}{(m_{max}^d - m_{min}^d)^2}, \tag{3.30}$$

where m_i^d are the masses in the detector frame[5]. These signals will be injected at a redshift z with a probability given by the distribution defined by Eq. 3.31:

$$p(z) = \frac{dV_c}{dz}\frac{1}{(1+z)V_0}, \tag{3.31}$$

where V_0 is defined as in Eq. 3.32:

$$V_0 = \int_0^{z_{max}} \frac{dV_c}{dz}\frac{1}{1+z}dz, \tag{3.32}$$

which corresponds to the total volume given the maximum redshift, z_{max}, of the used injections. The maximum redshift is selected using the expected sensitivity for the population model as reference. It should be noted that the redshift distribution doesn't depend on the parameters of the injection. This means that a signal with parameters θ might be injected at a redshift z from which cannot be recovered by the search, affecting the overall estimate of the sensitivity by reducing the statistics.

Once the injections are made, the search pipeline is passed through the data and the sensitive volume is estimated as in Eq. 3.33:

$$\langle V \rangle = V_0 \frac{N_{det}}{N_{Total}} = V_0 \epsilon, \tag{3.33}$$

[5]The relation between source frame masses and detector frame masses is given by: $m^d = (1+z)m$. Therefore, masses are bigger in the detector frame.

where N_{det} is the number of injections recovered by the search and N_{Total} is the total number of injections used. This sensitive volume depends on the population model used for the integration and will change if a different model is used. Therefore, a new population model requires a new injection campaign to estimate the efficiency and sensitivity.

An injection campaign is usually computationally expensive and this limits the available results. For that reason, a technique known as "reweighting" is used. This method consists of transforming the data from an initial population model, $\phi(\theta)$, to a different one, $\phi_f(\theta)$, by applying a set of weights to the sensitive volume. We assume that we have enough injections to cover the whole parameter space we are considering with our model and the regions in which the new model is. These weights are given by the ratio of the probabilities, as in Eq. 3.34:

$$w_i = \frac{\phi_f(\{m_1, m_2, \vec{s}_1, \vec{s}_2\}_i) p(z(d_i))}{\phi(\{m_1^d, m_2^d, \vec{s}_1, \vec{s}_2, d\}_i) J_i}, \tag{3.34}$$

where d is the luminosity distance of the injection, m_1 and m_2 are the component masses and $\vec{s}_1$ and $\vec{s}_2$ are the spins. By definition, the target population model, ϕ_f, is given in terms of the source frame component masses and the original model in terms of the detector frame masses: $m_1^d,\ m_2^d$. Therefore, we need to add a jacobian, J_i, to transform the initial population model from the detector frame to the source frame. This jacobian is given by Eq. 3.35:

$$J_i = (1 + z(d_i))^2 \left(\frac{\partial d}{\partial z}\right)_i. \tag{3.35}$$

The next step is to correctly estimate the maximum volume. In this case, the expression given by Eq. 3.32 is not valid due to the injections not following the population model. The correct way to estimate the volume is given by Eq. 3.36:

$$V_0^m = \int_{M_{c_{min}}}^{M_{c_{max}}} \int_{z_{min}(M_c)}^{z_{max}(M_c)} \phi_f(M_c) \frac{dV_c}{dz} \frac{1}{1+z} dz dM_c, \tag{3.36}$$

where we have used that the models can be written in terms of the chirp mass and that it's also the only parameter affected by the redshift. Therefore, the target population model is marginalized over the rest of the parameters and

defined in terms of the source frame chirp mass. The integral limits are given by the model (chirp mass) and the injections used (redshift).

This integral cannot be easily solved so Monte-Carlo integration is used. We start by drawing N_T samples of (M_c, z) from $\phi_f(M_c)$ and redshift distribution given by Eq. 3.31. These samples are then converted into detector frame chirp mass and luminosity distance: (M_c^d, d). Finally, we count how many of them are within the integral boundaries: N_{in}. By this method we have that the integral from Eq. 3.36 can be estimated as in Eq. 3.37:

$$V_0^m = V_0 \frac{N_{in}}{N_T}, \tag{3.37}$$

This volume can be used along with the weights given by Eq. 3.34 to estimate the sensitive volume assuming the population model, ϕ_f:

$$\langle V \rangle = V_0^m \frac{\sum_{i \in Detected} w_i}{\sum_{i \in Injected} w_i}, \tag{3.38}$$

where $i \in Detected$ iterates over the detected injections and $i \in Injected$ over the total injections.

3.3.7 Rate estimation

The sensitive volume described in Eq. 3.33 can be used to estimate the merger rate for the searched population. In first approximation, assuming that we have a sufficiently large number of confident detections, N_{Obvs}, we can approximate the rate density as in Eq. 3.39:

$$R = \frac{N_{Obvs}}{\langle VT \rangle}, \tag{3.39}$$

which can be derived by assuming that the observation of a GW is described by a Poisson process. This approximation, known as the foreground dominated statistics, is not valid when the number of detections is zero or not big enough. For this reason, in this section we describe a generic method to estimate the rate posterior [83, 84, 85]. We will follow in the next section with the particular case of no detection using the "loudest ranking statistic" [86].

Let's start by assuming that we have M events above a chosen threshold of our detection statistic, x, whose origin we don't know a priori (astrophysical or background). If we describe the process using an inhomogeneous Poisson process, we have that:

$$\begin{aligned}\frac{dN_s}{dx} &= s(x,\theta) = \hat{n}_s \hat{s}(x,\theta) \\ \frac{dN_b}{dx} &= b(x,\theta) = \hat{n}_b \hat{b}(x,\theta),\end{aligned} \tag{3.40}$$

where θ includes any parameter needed to define the distributions $s(x,\theta)$ and $b(x,\theta)$. These distributions can be used to define the cumulative number of triggers by integrating over the possible values of the detection statistic, x:

$$\begin{aligned}S(x,\theta) &= \int_{-\infty}^{x} dy s(y,\theta) \\ B(x,\theta) &= \int_{-\infty}^{x} dy b(y,\theta),\end{aligned} \tag{3.41}$$

and from that definition we extract the expected number of triggers detected from each type as $\hat{n}_s = S(\infty,\theta)$ and $\hat{n}_b = B(\infty,\theta)$. Furthermore, using an inhomogeneous Poisson process implies that the number of events from two non overlapping intervals are independent. The number of events for an interval, $x \in [x_1, x_2]$, is given by $S(x_2,\theta) - S(x_1,\theta)$ if it has astrophysical origin or $B(x_2,\theta) - B(x_1,\theta)$ if it is background.

As already mentioned, there is no way to know the origin of each trigger. This uncertainty is tackled by introducing a nuisance parameter, g_i, for each trigger as a flag: zero for background and one for signal. Therefore, our objective is to estimate the posterior given by Eq. 3.42 using Bayes' rule from Eq. B.2:

$$p(\{g_i\}, \hat{n}_s, \hat{n}_b, \theta | d, M) = \frac{L(d|\{g_i\}, M, \hat{n}_s, \hat{n}_b, \theta)\pi(\{g_i\}, M, \hat{n}_s, \hat{n}_b, \theta)}{p(d,N)}, \tag{3.42}$$

where 'd' corresponds to the time ordered data. Using now that our data is identically and independently distributed (i.i.d.) the likelihood can be written as the product of the density functions. In other words, we have that the likelihood is given by Eq. 3.43:

$$L(d|\{g_i\}, M, \hat{n}_s, \hat{n}_b, \theta) = \left[\prod_{\{i|g_i=1\}} \hat{s}(x_i,\theta)\right]\left[\prod_{\{i|g_i=0\}} \hat{b}(x_i,\theta)\right]. \tag{3.43}$$

Furthermore, from the law of total probability seen in Eq. B.1, we have that the prior can be written as:

$$\begin{aligned}\pi(\{g_i\}, M, \hat{n}_s, \hat{n}_b, \theta) &= p(\{g_i\}|M, \hat{n}_s, \hat{n}_b, \theta)p(M, \hat{n}_s, \hat{n}_b, \theta) \\ &= p(\{g_i\}|M, \hat{n}_s, \hat{n}_b, \theta)p(M|\hat{n}_s, \hat{n}_b, \theta)p(\hat{n}_s, \hat{n}_b, \theta) \\ &= p(\{g_i\}, M|\hat{n}_s, \hat{n}_b, \theta)p(\hat{n}_s, \hat{n}_b, \theta). \end{aligned} \tag{3.44}$$

Two of these priors can be chosen using the extra information we have from the problem. In the first place, for the term $p(\{g_i\}|M, \hat{n}_s, \hat{n}_b, \theta)$, we can use that the probability of the flag being one (zero) is $\frac{\hat{n}_s}{\hat{n}_s+\hat{n}_b}$ $\left(\frac{\hat{n}_b}{\hat{n}_s+\hat{n}_b}\right)$. Therefore, that distribution is given by the product of the probabilities as in Eq. 3.45:

$$\begin{aligned} p(\{g_i\}|M, \hat{n}_s, \hat{n}_b, \theta) &= \left[\prod_{i|g_i=1} \frac{\hat{n}_s}{\hat{n}_s + \hat{n}_b}\right] \left[\prod_{i|g_i=0} \frac{\hat{n}_b}{\hat{n}_s + \hat{n}_b}\right] \\ &= \left(\frac{\hat{n}_s}{\hat{n}_s + \hat{n}_b}\right)^{M_s} \left(\frac{\hat{n}_b}{\hat{n}_s + \hat{n}_b}\right)^{M_b}, \end{aligned} \tag{3.45}$$

where M_s is the number of elements with flag one and M_b with flag zero. Secondly, using the assumption that the number of events follows a Poisson process described by $\hat{n}_s + \hat{n}_b$, we have that $p(M|\hat{n}_s, \hat{n}_b, \theta)$ can be written as:

$$p(M|\hat{n}_s, \hat{n}_b, \theta) = \frac{(\hat{n}_s + \hat{n}_b)^M}{M!} exp[-(\hat{n}_s + \hat{n}_b)]. \tag{3.46}$$

Finally, for the term $p(\hat{n}_s, \hat{n}_b, \theta)$ we don't have extra information. Therefore, we decide to select a non informative prior and, in particular, we use Jeffrey's prior for the Poisson distribution:

$$p(\hat{n}_s, \hat{n}_b, \theta) = \frac{\alpha}{\sqrt{\hat{n}_s \hat{n}_b}} p(\theta), \tag{3.47}$$

where α is a normalizing constant. At the end we have that our full prior is given by Eq. 3.48:

$$\begin{aligned}\pi(\{g_i\}, M, \hat{n}_s, \hat{n}_b, \theta) &= \frac{\alpha}{\sqrt{\hat{n}_s \hat{n}_b}} p(\theta) \frac{(\hat{n}_s + \hat{n}_b)^M}{M!} exp[-(\hat{n}_s + \hat{n}_b)] \left(\frac{\hat{n}_s}{\hat{n}_s + \hat{n}_b}\right)^{M_s} \left(\frac{\hat{n}_b}{\hat{n}_s + \hat{n}_b}\right)^{M_b} \\ &= \frac{\alpha}{\sqrt{\hat{n}_s \hat{n}_b}} p(\theta) \frac{\hat{n}_s^{M_s} \hat{n}_b^{M_b}}{M!} exp[-(\hat{n}_s + \hat{n}_b)], \end{aligned} \tag{3.48}$$

where we have used that $M_b + M_s = M$. Now, by combining the prior from Eq. 3.48 and the likelihood in Eq. 3.43 we can construct the posterior distribution as in Eq. 3.49:

$$p(\{g_i\}, \hat{n}_s, \hat{n}_b, \theta | d, M) = \frac{\alpha}{p(d, M)M!} \left[\prod_{\{i|g_i=1\}} \hat{n}_s \hat{s}(x_i, \theta) \right] \left[\prod_{\{i|g_i=0\}} \hat{n}_b \hat{b}(x_i, \theta) \right] \times exp[-(\hat{n}_s + \hat{n}_b)] \frac{p(\theta)}{\sqrt{\hat{n}_s \hat{n}_b}}. \quad (3.49)$$

If we marginalize over the discrete nuisance parameter, $\{g_i\}$, we have that the marginal posterior distribution is given by Eq. 3.50:

$$p(\hat{n}_s, \hat{n}_b, \theta | d, M) \propto \prod_i \left[\hat{n}_s \hat{s}(x_i, \theta) + \hat{n}_b \hat{b}(x_i, \theta) \right] exp[-(\hat{n}_s + \hat{n}_b)] \frac{p(\theta)}{\sqrt{\hat{n}_s \hat{n}_b}}, \quad (3.50)$$

which can be used to estimate the distribution of the rates. For example, let's assume that we have chosen a threshold such that the number of triggers from an astrophysical origin is much bigger than from background. Under this assumption we have that $\hat{s}(x_i, \theta) \gg \hat{b}(x_i, \theta)$ for all elements. If the condition $\hat{n}_s \hat{s}(x_i, \theta) \gg \hat{n}_b \hat{b}(x_i, \theta)$ also holds, we have that the marginal posterior in Eq. 3.50 can be written as:

$$p(\hat{n}_s, \hat{n}_b, \theta | d, M) \propto \prod_i \left[\hat{s}(x_i, \theta) \right] \hat{n}_s^{M_s} exp[-(\hat{n}_s + \hat{n}_b)] \frac{p(\theta)}{\sqrt{\hat{n}_s \hat{n}_b}}. \quad (3.51)$$

If we marginalize over $\hat{n}_b$ we can write this as:

$$\begin{aligned} p(\hat{n}_s, \theta | d, M) &= \int p(\hat{n}_s, \hat{n}_b, \theta | d, M) d(\hat{n}_b) \\ &\propto \prod_i \left[\hat{s}(x_i, \theta) \right] \hat{n}_s^M exp[-(\hat{n}_s)] \frac{p(\theta)}{\sqrt{\hat{n}_s}} \int exp[-\hat{n}_b] \frac{1}{\sqrt{\hat{n}_b}} d(\hat{n}_b) \\ &\propto \prod_i \left[\hat{s}(x_i, \theta) \right] \hat{n}_s^M exp[-(\hat{n}_s)] \frac{p(\theta)}{\sqrt{\hat{n}_s}}, \end{aligned} \quad (3.52)$$

and if we assume there is no dependence in θ, we have that the maximum is reached when:

$$\begin{aligned}
&\frac{d}{d(\hat{n}_s)}\left[(M-\frac{1}{2})log(\hat{n}_s)-\hat{n}_s\right]=0\\
&(M-\frac{1}{2})\frac{1}{\hat{n}_s}-1=0\\
&\hat{n}_s=M-\frac{1}{2}
\end{aligned} \tag{3.53}$$

Using now that the number of events can be written in terms of rate density as $\hat{n}_s = R\langle VT\rangle$, we have that:

$$R=\frac{M-\frac{1}{2}}{\langle VT\rangle}\approx\frac{M}{\langle VT\rangle}, \tag{3.54}$$

where we have approximately recovered the rate for the foreground dominated statistics as in Eq. 3.39. In fact, the term $\frac{1}{2}$ appears due to the choice of prior in $p(\hat{n}_s, \hat{n}_b, \theta)$.

In the next section we will use this formalism to derive the estimation of an upper limit for the rate in the case of no detection.

3.3.8 Loudest event statistic

Sometimes the confident number of triggers from astrophysical origin is not large enough to provide a good estimation of the rate density and a different approach is needed to add constraints. This is done by using the "Loudest event statistic", which consists in using the loudest event, from a list ordered by a ranking statistics x, to estimate upper limits to the signal rate [87, 88, 86]. However, there is a drawback to this approach: we are discarding the information given by the other triggers. Nevertheless, this technique has been used to estimate rate upper limits in several searches [88, 89, 87, 90]. In this section we will explain the basic concepts of this method and derive the equation used to constrain the rate in the case of no signal.

Let's assume that from our list of events we are interested only in the loudest trigger, x_M. Then, under this hypothesis, the marginal posterior in Eq. 3.50 takes the form seen in Eq. 3.55:

$$p(\hat{n}_s, \hat{n}_b, \theta|d) \propto \left[\hat{n}_s \hat{s}(x_M, \theta) + \hat{n}_b \hat{b}(x_M, \theta)\right] \times exp[-\hat{n}_s(1 - \hat{S}(x_M, \theta)) - \hat{n}_b(1 - \hat{B}(x_M, \theta))] \frac{p(\theta)}{\sqrt{\hat{n}_s \hat{n}_b}}, \tag{3.55}$$

where the term inside the exponential, $\hat{n}_s(1-\hat{S}(x_M, \theta))$, indicates the number of signals/backgrounds expected above the ranking statistic of our loudest event. Let's now assume the following:

- The background distribution and $\hat{n}_b$ are known.
- We don't have extra information on $\hat{n}_s$ but instead of using Jeffrey's prior we use a uniform prior: $p(\hat{n}_s) \propto k$ with k constant.
- There is no dependence in θ.

Under these assumptions we have that Eq. 3.55 simplifies to Eq. 3.56:

$$p(\hat{n}_s|d) \propto \left[\hat{n}_s \hat{s}(x_M) + \hat{n}_b \hat{b}(x_M)\right] \times exp[-\hat{n}_s(1 - \hat{S}(x_M)) - \hat{n}_b(1 - \hat{B}(x_M))], \tag{3.56}$$

which can be integrated to estimate the normalization constant, κ:

$$\begin{aligned}
\kappa &= \int_0^\infty p(\hat{n}_s|d) d(\hat{n}_s) \\
&= e^{-\hat{n}_b(1-\hat{B}(x_M))} \left[\hat{s}(x_M) \int_0^\infty \hat{n}_s e^{-\hat{n}_s(1-\hat{S}(x_M))} d(\hat{n}_s) + \hat{n}_b \hat{b}(x_M) \int_0^\infty e^{-\hat{n}_s(1-\hat{S}(x_M))} d(\hat{n}_s)\right] \\
&= \frac{\hat{n}_b \hat{b}(x_M)}{1 - \hat{S}(x_M)} e^{[-\hat{n}_b(1-\hat{B}(x_M))]} \left[1 + \frac{\hat{s}(x_M)}{(1 - \hat{S}(x_M))\hat{n}_b \hat{b}(x_M)}\right].
\end{aligned} \tag{3.57}$$

Therefore, the posterior is:

$$\begin{aligned}
p(\hat{n}_s|d) &= \frac{1}{\kappa} \left[\hat{n}_s \hat{s}(x_M) + \hat{n}_b \hat{b}(x_M)\right] \times exp[-\hat{n}_s(1 - \hat{S}(x_M)) - \hat{n}_b(1 - \hat{B}(x_M))] \\
&= \frac{\hat{\epsilon}}{1 + \Lambda}(1 + \hat{n}_s \hat{\epsilon} \Lambda) e^{-\hat{n}_s \hat{\epsilon}},
\end{aligned} \tag{3.58}$$

where we have defined $\hat{\epsilon}$ and Λ as follows:

$$\hat{\epsilon} = 1 - \hat{S}(x_M) \tag{3.59}$$

$$\Lambda = \frac{\hat{s}(x_M)}{(1 - \hat{S}(x_M))\hat{n}_b \hat{B}(x_M)}. \tag{3.60}$$

$\hat{\epsilon}$ describes the detection efficiency or, in other words, the probability that a trigger louder than x_M has an astrophysical origin. Λ measures how likely it is for a trigger with ranking statistics x_M to have an astrophysical origin versus being background. In particular, $\Lambda \to 0$ indicates that the origin is almost sure background.

If we integrate this posterior to estimate a credible interval for $\hat{n}_s$ as in Eq. 3.61, we have that:

$$1 - \alpha = \int_0^n p(\hat{n}_s|d)d(\hat{n}_s) = 1 - \left[1 + \frac{n\hat{\epsilon}\Lambda}{1+\Lambda}\right] e^{-n\hat{\epsilon}}. \tag{3.61}$$

We are interested in the case where the loudest trigger is background. Therefore, by taking $\Lambda = 0$, we have that an upper limit for the number of events at confidence level $1 - \alpha$ is given by Eq. 3.62:

$$n = -\frac{log(\alpha)}{\hat{\epsilon}}, \tag{3.62}$$

which can be transformed into rate density by dividing over the observation time, T_0, and the maximum volume searched V_0:

$$R = \frac{n}{TV_0} = -\frac{log(\alpha)}{TV_0\hat{\epsilon}} = -\frac{log(\alpha)}{\langle VT \rangle}, \tag{3.63}$$

where we have transformed into sensitive volume using the efficiency as in Eq. 3.33. It should be noted that this efficiency is estimated by using the loudest event ranking statistics as threshold. Eq. 3.63 provides an estimation of the upper limit for the rate density at confidence limit of $1 - \alpha$ in the case of no detection. In particular, the 90% confidence interval ($1 - \alpha = 0.9 \to \alpha = 0.1$) is given by Eq. 3.64:

$$R_{90\%} = -\frac{log(0.1)}{\langle VT \rangle} = \frac{2.303}{\langle VT \rangle}. \tag{3.64}$$

This will be used in Chapter 4 to add constrains to the rate of CBC events with at least one subsolar mass component.

3.3.9 Creating a search pipeline

The elements introduced in the previous sections are used to construct a pipeline dedicated to the search for GW from CBC events. The purpose of these searches is to identify candidates from detector data and to provide an estimation of their statistical significance. In this section we will show how all the steps are merged together to construct these pipelines. We will use the PyCBC pipeline, explained in detail in Refs. [70, 69, 91, 92, 93, 94], as an example. A flowchart showing all the steps of a pipeline can be seen in Fig. 3.7.

The offline PyCBC pipeline [91, 92] is a matched-filter pipeline designed to search for coincident triggers in two or more detectors [94]. The pipeline's input is the calibrated strain data of the different detectors. The data is gated to remove non Gaussian components and noise transients. Furthermore, the worst periods of the data are filtered through data quality investigations. Depending on the objective of the search, a template bank that spans the astrophysical signal space is created. PyCBC is limited to searches of GW from binaries in quasi-circular orbits with aligned spins.

The data from each detector is matched filtered against the template bank independently. A list of triggers is created by searching for the maxima of the SNR time series that exceed a chosen threshold value. Given that many sample points in the SNR time series can exceed the threshold and a signal will have a single narrow peak, a time-clustering algorithm is applied. The SNR time series is divided into equal one second windows and the maximum SNR is identified. This list is subject to signal consistency tests, as described in Sec. 3.3.4, starting with the chi-squared test. The output of this test is used to down-weight triggers according to the following recipe:

$$\hat{\rho} = \begin{cases} \rho/[(1+(\chi_r^2)^3)/2]^{1/6}, & \text{if } \chi_r^2 > 1 \\ \rho & \text{if } \chi_r^2 \leq 1 \end{cases}, \tag{3.65}$$

where ρ is the matched filter SNR and $\chi_r^2 = \frac{\chi^2}{2p-2}$ is the reduced chi squared for each trigger. The re-weighted SNR is used to filter the triggers by setting a re-weighted SNR threshold. Now coincidence between the different interferometers taking data at the time of each trigger is enforced. The triggers should be observed at times compatible with the maximum travel time of the GW between interferometers, accounting for uncertainty in the measured time, e.g. 15ms for the two LIGO detectors.

Furthermore, only the triggers with the same template in all the detectors are allowed. Any trigger remaining is labeled as candidate event.

Each of these candidates event is assigned a detection statistic value, e.g. the network SNR, ranking the likelihood of the trigger being produced by a GW signal. The significance of each trigger is estimated by using the FAR as described in Sec. 3.3.5. The estimation of the significance requires a list of statistically independent candidates. To prevent noise transients and signals to generate correlated coincident candidates from all the triggers within a short time, a clustering step is done: If there are more than one coincident trigger within a time window, e.g. 10s, only the event with the highest detection statistics is taken into account. Finally, a list of significant candidates is produced.

The performance of the pipeline is monitored by regular injections into the data. These injections are also used to estimate the sensitivity of the search.

3.4 Parameter estimation

Assuming we have a detection, the next step would be to estimate the parameters of the source [95, 96, 46, 97, 98]. This is done by performing parameter estimation, using bayesian inference (See Appendix B), assuming a model and constructing a likelihood with our data. Parameter estimation can also be used to constrain parameters of models. In this section we will focus on parameter estimation for CBC sources but the technique for other searches follows the same principle.

The likelihood is constructed as in Eq. 3.13, where we have assumed that the residual should follow Gaussian noise if the model correctly describes the signal within the data. If we also assume that our data is stationary, the likelihood model is reduced to:

$$L(d|h;\theta) = \frac{1}{det(2\pi C)^{1/2}} exp\left(-\frac{1}{2}(d(t)-h(t;\theta)|d(t)-h(t;\theta)\right). \quad (3.66)$$

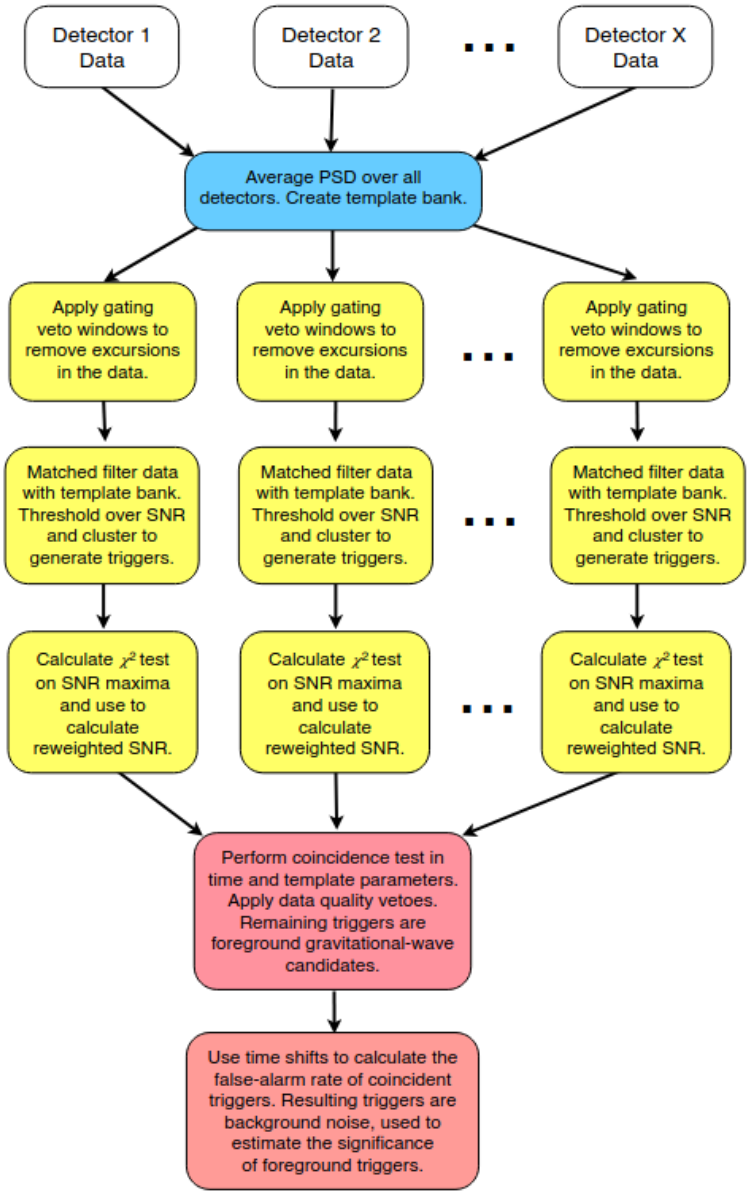

Figure 3.7: Flowchart showing the different steps of the PyCBC pipeline. Taken from Ref. [92]

And with this likelihood, we can construct the posterior probability for parameters θ using Bayes' rule, as in Eq. 3.67:

$$p(\theta|d) = \frac{L(d|h;\theta)p(\theta)}{Z}, \tag{3.67}$$

where $p(\theta)$ is the prior distribution of our parameters and Z is the evidence, used to normalize the posterior distribution. A key aspect to notice in this analysis is that it strongly depends on the choice of signal model for $h(t;\theta)$. This model dependence means that there might be systematics due to differences in the waveform used to model the signal.

The prior distribution is given by our knowledge of the parameters that describe the waveform from a CBC event, as we have seen in Chapter 1 and Chapter 2. For the most general search, we need to provide priors for:

- Two component masses or chirp mass and reduced mass. In detector frame.
- 6 spin components in Cartesian or spherical coordinates.
- Luminosity distance to the source.
- Sky position in terms of right ascension and declination.
- A reference time and phase, usually the merger time in the interferometer.
- Polarization and inclination angle.

The choice of the functional form of these priors is important and, usually, a non informative prior is used. For example, the masses are usually described by a prior uniform in component masses and the distances with a prior uniform in comoving volume (using that galaxies are distributed in the same way in the universe). The same considerations are given to the rest of the parameters.

The last step is to sample the posterior distribution to get an estimation of the source parameters. This is not an easy task as we need to estimate the distribution for 15 parameters, 11 if we assume aligned spins.

This translates into a staggering amount of 45 parameters, using 3 detectors and including parameters needed to describe the interferometers uncertainty, to estimate. This problem is solved by sampling as explained in Sec. B.5.

In particular, two methods are used, a mcmc approach using the Metropoli-Hasting algorithm (see Sec. B.5.1) and another method based on a different family known as nested sampling. An example of a posterior distribution for a five dimensional can be seen in Fig. 3.8. In the diagonal, the individual posterior distributions for five different parameters are shown. The dotted lines correspond to the 1σ intervals around the median value, indicated at the top of each distribution. The elements below the diagonal show the two-dimensional posteriors where the correlations between the parameters are seen. Contour lines for the 1σ (darkest), 2σ and 3σ (lightest) confidence regions are drawn.

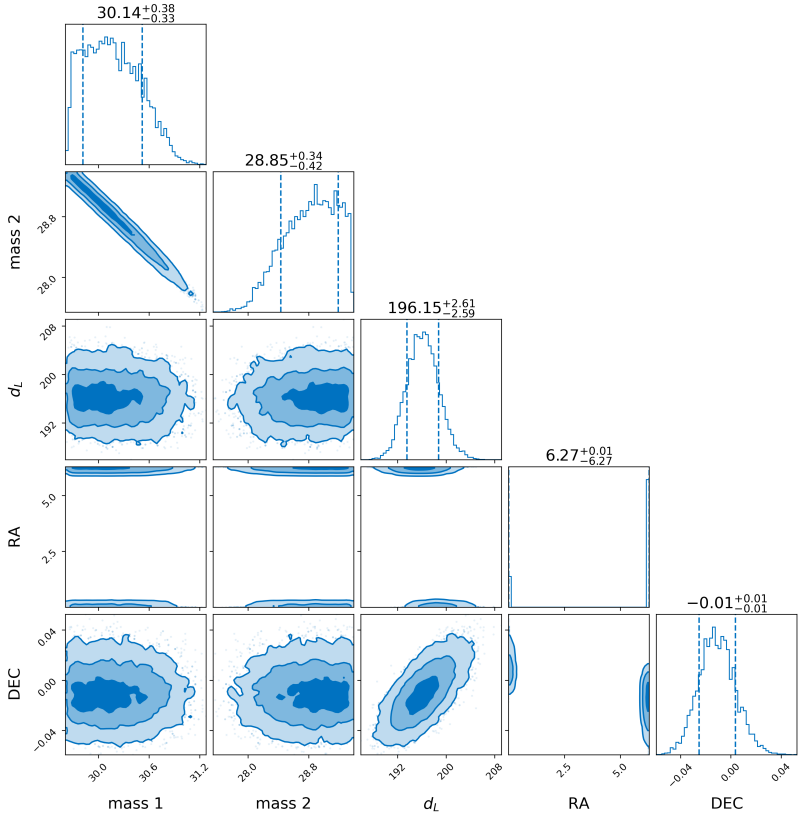

Figure 3.8: Parameter estimation for five parameters corresponding to the masses, distance and sky position (Right ascension+Declination). The darkest shaded region corresponds to the 1σ region while the lightest is the 3σ region.

Chapter 4

Sub solar mass searches with matched filtering

In this chapter, we will discuss the results of a search for GWs produced by the coalescence of two compact objects in which at least one of them has a subsolar mass. The search is motivated by the fact that there are no standard channels based on star evolution to produce subsolar mass compact objects. Therefore, such a detection would point to a new formation mechanism different from stellar evolution. We report the search performed by the Ligo-Virgo-Kagra (LVK) collaboration using the data from the third observation run (O3a+O3b) [89, 99]. The search includes an extended mass range with respect to the O1 and O2 subsolar mass searches [88, 87]. Outside the collaboration, the O3 data was used to independently search for subsolar mass objects, as detailed in Refs. [90, 100].

The results will be used to estimate the sensitive volume (see Sec. 3.3.6) and constrain the rate density of subsolar mass CBCs using the loudest event statistic (see Sec. 3.3.8). Two models will be constrained using these limits: fraction of primordial black holes (PBHs) in the universe and fraction of mass of atomic dark matter that collapses into a black hole [101, 102]. A more in-depth description of these models is presented in the following sections.

4.1 Primordial black holes

PBHs have been a source of interest since they were independently suggested by Zeldovich and Novikov in 1967 [103] and Hawking in 1971 [104] to account for at least a fraction of the universe's dark matter (DM). These objects would have been formed in the early universe with mass depending on the formation time [105, 106, 107].

A BH of mass M_{PBH} would have a Schwarzschild radius of $R_s = 2GM_{PBH}/c^2$ and a mean density inside that region of $\rho_s = \frac{M_{PBH}}{(4\pi R_s^3/3)}$. Assuming that the cosmological density at a time t after the Big Bang is given by $\rho_c \sim 1/(Gt^2)$, we have that in order for PBH to form we need densities of the same order inside the horizon, $\rho_c \sim \rho_s$. Therefore, the initial mass of a PBH would be of the order of the horizon mass, M_H, at the time, as in Eq. 4.1:

$$M_H \sim \frac{c^3 t}{G} \sim 10^{15} \left(\frac{t}{10^{-23}s}\right) g. \tag{4.1}$$

From this equation it can be deduced that the PBHs masses could span an enormous mass range. For example, at a time $t \sim 10^{-43}s$ they would have a mass of $\sim 10^{-5}g$ while at the time of the QCD transition, $t \sim 10^{-6}s$, they would have masses around $1\mathrm{M}_\odot$. Therefore, due to the lower mass limit of stellar mass black holes being $\approx 3\mathrm{M}_\odot$, a subsolar mass black hole would point to a possible primordial origin.

The small masses of PBHs motivated Hawking to study their quantum properties. He found that black holes radiate thermally at a temperature given by Eq. 4.2:

$$T = \frac{\hbar c^3}{8\pi G M k_b} \approx 10^{-7} \left(\frac{M}{\mathrm{M}_\odot}\right)^{-1} K, \tag{4.2}$$

losing mass and energy during the emission and ultimately evaporating on a timescale given by Eq. 4.3:

$$\tau(M) \approx \frac{\hbar c^4}{G^2 M^3} \approx 10^{64} \left(\frac{M}{\mathrm{M}_\odot}\right)^3 yr. \tag{4.3}$$

From this result it can be extracted that PBHs smaller than $10^{15}g$ (formed before $10^{-23}s$) would have evaporated by now and cannot contribute to dark

matter today. Since their conception, PBHs have been used to explain several cosmological features. For example, evaporating PBHs have been used to explain extragalactic and Galactic γ ray backgrounds or short period γ ray burst. Non evaporating PBHs have been used not only to explain DM but also lensing effects or GW detected events. Even though there are other explanations for these effects, they have been used to constrain the number of PBHs of a given mass.

In this work, we are going to focus on non evaporating PBHs in the subsolar mass range. These are expressed as constraints on the density fraction of PBHs with respect to the density of DM, that is, $f_{\rm PBH} = \Omega_{\rm PBH}/\Omega_{\rm DM}$. It should be noted that most of the constraints assume that the PBH mass spectrum is quasi-monochromatic (In other words, a distribution with width $\Delta M_{PBH} \leq M_{PBH}$). The mass distribution and abundance of PBHs depend on their formation mechanism and the fluctuations generated in the primordial power spectrum during inflation. In the next section some of the most relevant formation scenarios will be discussed.

4.1.1 Formation of PBHs

The most natural formation channel follows from the collapse of overdense regions in the early universe. Density fluctuations would enter the horizon stopping the expansion and collapsing. This would happen when the overdensity, δ, excess a critical value, δ_c, at the time of entering the horizon. In early calculation, this value was estimated to be 1/3 for the radiation era and spherically symmetric regions. Assuming that the fluctuations follow a Gaussian distribution with dispersion θ, then the fraction collapsing to PBH is estimated to be [108]:

$$\beta \approx Erfc\left[\frac{\delta_c}{\sqrt{2}\sigma}\right], \tag{4.4}$$

where 'Erfc' corresponds to the complementary error function[1] and indicates the excess above the mean value. Assuming a monochromatic spectrum, the end mass of the resulting PBH [109] on some mass scale K is given by the Eq. 4.5 when the collapse occurs near the critical point:

$$M = K(\delta - \delta_c)^{\eta}, \tag{4.5}$$

[1] $Erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt$

where K is of the order of the horizon mass and η depends on the equation of state. This allows the mass of the PBHs to extend down to small scales. The value of the critical density has been estimated several times, including more effects that might affect PBH formation. Recent numerical simulations suggest that it should be constrained between 0.4 and 2/3. However, given that the formation of PBHs originates from the tail of the density fluctuations, it is expected to be very sensitive to non Gaussianities in the density. Differing values for the equation of state of the universe have also shown to affect the value of the critical density.

The dependence of δ_c in the equation of state is particularly important during the QCD transition[2] that happened when the universe cooled down [110]. During this transition, the equation of state of the universe softened by around 30% boosting the PBH formation due to a decrease in the critical density of around 10%. As indicated by Eq. 4.1, the QCD transition would occur at a time with horizon masses close to $1M_\odot$, producing in turn a peak in the PBH mass distribution at that mass. In fact, this boost is not negligible between $0.1M_\odot$ and $100M_\odot$ which corresponds to the sensitive region of the current interferometers. Finally, inflation dynamics might also influence the formation of PBHs in the subsolar mass range, as explained in Refs. [111, 112, 113, 114, 115, 116, 117, 118, 119].

4.1.2 Binary formation and merger rate

The mechanism of PBH binary formation depends on the cosmological epoch at which they formed: before matter-radiation equality (point in which the energy density of matter and radiation are equal), binaries form when the PBHs decouple from the cosmic expansion and become gravitationally bound to each other (early binaries). After matter-radiation equality, they are dynamically formed inside PBH clusters (late binaries). The two families of PBH binaries contribute to the merger rate. However, the contribution from early binaries is expected to dominate the present rate.

[2]New results from lattice QCD indicate that this is not a phase transition but a crossover. There were no periods of phase coexistence and the thermodynamic quantities evolved smoothly.

An estimation of the merger rate of late binaries is given in Refs. [120, 121] and shown in Eq. 4.6.

$$R_{LB} = R_{clust} f_{\rm PBH} f(m_1) f(m_2) \frac{(m_1+m_2)^{10/7}}{(m_1 m_2)^{5/7}}, \tag{4.6}$$

where R_{clust} is an effective scaling factor incorporating the clustering properties of PBHs and $f(m_i)$ are the PBHs mass distribution.

In this thesis two early binary formation models will be considered. The models provide an analytical expression of the theoretical merger rate that will be used along the estimated merger rate limits to constraint $f_{\rm PBH}$.

3-body system + No suppression

The first model consists on a phenomenological model for early binary formation, following the derivation from Refs. [122, 123]. The model assumes a population of PBHs uniformly distributed in comoving volume and with equal mass. This model provides a simple approach that can be used to constrain generic PBH mass functions [124].

The model in Refs. [122, 125, 123, 126] assumes that the binaries are created when some PBHs in the early universe decouple from the cosmic expansion. Without external contributions, the nearest pair of PBHs will move closer until they collide. The collision is prevented by the gravitational field of the PBH closest to the pair. It will produce a torque that disrupts the collision and allows a binary system with a large eccentricity to form.

The new binary system will lose energy in the form of GW until it finally coalesces at some time t_c. Therefore, the probability of coalescence occurring in a time interval $(t, t+dt)$ is given by:

$$\mathrm{d}P = \begin{cases} \dfrac{3\, f_{\rm PBH}^{37/8}}{58} \left[f_{\rm PBH}^{-29/8} \left(\dfrac{t}{t_{\rm c}}\right)^{3/37} - \left(\dfrac{t}{t_{\rm c}}\right)^{3/8} \right] \dfrac{\mathrm{d}t}{t}, & t < t_{\rm c} \\ \dfrac{3\, f_{\rm PBH}^{37/8}}{58} \left[f_{\rm PBH}^{-29/8} \left(\dfrac{t}{t_{\rm c}}\right)^{-1/7} - \left(\dfrac{t}{t_{\rm c}}\right)^{3/8} \right] \dfrac{\mathrm{d}t}{t}, & t \geq t_{\rm c} \end{cases}, \tag{4.7}$$

and the coalescence time, t_c, is:

$$t_{\rm c} = \frac{3}{170}\frac{c^5}{(Gm_{\rm PBH})^{5/3}}\frac{f_{\rm PBH}^7}{(1+z_{\rm eq})^4}\left(\frac{8\pi}{3H_0^2\Omega_{\rm DM}}\right)^{4/3}, \tag{4.8}$$

where c is the speed of light, G the gravitational constant, $z_{\rm eq}$ is the redshift at matter-radiation equality ($z_{eq} \approx 3000$), $m_{\rm PBH}$ the mass of the black hole and H_0 the Hubble constant. The merger rate at present time is given by Eq. 4.9:

$$\mathcal{R}_{\rm PBH} = n_{\rm PBH}\left.\frac{{\rm d}P}{{\rm d}t}\right|_{t=t_0} = \frac{3H_0^2}{8\pi G}\frac{\Omega_{\rm DM}}{m_{\rm PBH}}\left.\frac{{\rm d}P}{{\rm d}t}\right|_{t=t_0}. \tag{4.9}$$

This model provides a simple estimation of the merger rate but doesn't take into account effects that might disrupt the binary at some point between the formation and the merger. A more refined model including effects that might suppress the merger rate is shown in the next section.

N-body system + Rate suppression

The previous model considered that the torque was fully provided by only one PBH but, recent studies, have found that the contribution from nearby PBHs is not negligible and a full N-body analysis is needed, as detailed in Refs. [127, 128, 129]. Furthermore, depending on the initial spatial distribution of PBHs, the binary might be disrupted shortly after formation due to the nearest PBHs. This results in a suppressed merger rate.

These effects are taken into account in Refs. [127, 130, 131] and estimated that the theoretical merger rate is given by Eq.

$$\frac{{\rm d}\mathcal{R}^{\rm PBH}}{{\rm d}\ln m_1 {\rm d}\ln m_2} = 1.6\times10^6\,{\rm Gpc}^{-3}\,{\rm yr}^{-1}\times f_{\rm sup} f_{\rm PBH}^{53/37} f(m_1)\times f(m_2)\left(\frac{m_1+m_2}{M_\odot}\right)^{-32/37}\left[\frac{m_1 m_2}{(m_1+m_2)^2}\right]^{-34/37}, \tag{4.10}$$

where f_{sup} is the suppression factor and $f(m)$ is the normalized PBH density distribution. The suppression factor also accounts for the effect of inhomogeneities in the surrounding matter and disruptions due to nearby PBH clusters absorbing the binary. In particular, it was found that the disruption of the binary due to a single PBH is dominant when $f_{\rm PBH} \ll 1$. The contribution from PBH clusters is expected to be more dominant when $f_{\rm PBH} \approx 1$.

The suppression factor has an important contribution in the case of a distribution of PBHs with the same mass or strongly peaked mass function. Under this hypothesis, the suppression factor is given by $f_{sup} \approx 2.4 \times 10^{-3} f_{\rm PBH}^{-0.65}$, highly suppressing the merger rate.

4.2 Dark matter black holes

The second source of subsolar objects that will be considered are dark matter black holes. These black holes are produced by the collapse of dark matter halos made of dark fermionic particles. Therefore, it is referred to as "Atomic dark matter".

Atomic dark matter models assume that dark matter might couple to new short or large range forces. In particular, we are going to focus on a long range force, similar to electromagnetism, mediated by a massless dark photon and with a strength given by a dark fine structure constant, α_D. In this model we assume that atomic dark matter is made of two dark fermionic particles of opposite charge: one with mass m_x and the other one with mass m_c such that $m_x > m_c$. Through the interaction with the new force, these two particles form bound states similar to atomic and molecular hydrogen [101, 102].

Once enough density of gas dark matter is accumulated in a region of space, it starts cooling down by several processes [132] (e.g. recombination or Bremmstrahlung), dissipating kinetic energy into dark radiation. This reduces the internal gas pressure of the system and, through the effect of gravity, the dark gas cloud collapses, forming a black hole. Due to the lack of pressure produced by nuclear fusion, dark matter clouds can only collapse into black holes. The minimum mass needed for this collapse to occur is given by the Jeans mass[3] for atomic dark matter as in Eq. 4.11:

$$M_{min} \approx 800 M_\odot \left(\frac{m_p}{m_x}\right)^{5/2} \left(\frac{m_c}{m_e}\right)^{1/2} \left(\frac{\alpha_D}{\alpha}\right)^{1/2}, \tag{4.11}$$

where m_p is the mass of the proton, m_e the electron mass, m_x and m_c the mass of the dark matter particles and α is the fine structure constant in the

[3]Critical mass at which a gas system starts collapsing due to gravity.

Standard Model. However, the minimum mass a dark matter black hole may have is given by the dark Chandrasekhar limit. This limit, shown in Eq. 4.12, corresponds to the Chandrasekhar limit for star formation including a scaling factor depending on the mass of the dark heavy particle, m_x:

$$M_{DC} = 1.457\mathrm{M}_{\odot}\left(\frac{\mathrm{m_p}}{\mathrm{m_x}}\right)^2. \tag{4.12}$$

In both situations, the mass can reach subsolar ranges depending on the values of m_x, m_c and α_D. Therefore, searches in the subsolar range are able to constrain these models and the fraction of dark matter that collapses into Dark matter Black Holes (DBHs), f_{DBH}.

Furthermore, given that the formation of DBHs is expected to be similar to the formation of Population-III stars[4], their parameters are used as reference.

4.2.1 Likelihood and posterior estimation

The analysis is done by means of Bayesian inference: estimating the posterior distribution given the sensitive volume of the search and modelled dark matter BH rates, $\mathcal{R}$. The posterior is given in Eq. 4.13:

$$P(f, \overline{\theta}|\mathcal{R}, VT) \propto P(f)P(\overline{\theta})\mathcal{L}(f, \overline{\theta}; \mathcal{R}VT), \tag{4.13}$$

where $\overline{\theta} = (M_{min}, b, r)$ includes two additional parameters, b and r, needed to fully model the distribution. These two parameters are used to model the initial mass function distribution of DBHs: 'b' corresponds to the slope and 'r' limits the maximum mass of the initial population. In particular, an exponential described by $d\mathcal{P}(m) \propto m^b$ with $m \in [M_{min}, M_{max}] = [M_{min}, rM_{min}]$ is assumed.

Uniformative priors large enough to include ranges from Population-III studies are used:

- b: uniform between [-1,2]. Compatible with values from Population-III studies for binary systems [133, 134, 135].

[4]Hypothetical population of high massive stars with low metalicity formed in the early universe. It is suggested that these stars might have started the production of elements heavier than hydrogen.

- r: uniform in log between [2,1000]. This range is large enough to include values from Population-III stars [135].
- $f_{\rm DBH}$: uniform between $[10^{-10}, 1]$.
- M_{min}: uniform between $[10^{-3}M_{\odot}, 3.1M_{\odot}]$.

The rate, $\mathcal{R}$, is modelled as in Eq. 4.14 and restricted to the search chirp mass range, $M_c \in [0.2M_{\odot}, 2.5M_{\odot}]$:

$$\mathcal{R}_i(M_c = m_i|f, \overline{\theta}) = P_i(m_i|t_{\rm m}, \overline{\theta}) \left(\frac{{\rm d}P(t_{\rm m} = 10 \text{ Gyr}|\overline{\theta})}{{\rm d}t} \right) \times \left(\frac{\rho_{\rm DM} \times f \times f_{\rm binary}}{\langle M \rangle} \right), \tag{4.14}$$

where $\rho_{\rm DM} = 3.3 \times 10^{19}\ M_{\odot}\ {\rm Gpc}^{-3}$ is the density of dark matter in the universe, $f_{\rm binary} = 0.26$ is the fraction of DBH binaries from the total amount of DBHs (a choice motivated by the results from Population-III stars [134]), $P_i(m_i|t_{\rm m}, \overline{\theta})$ is the chirp mass distribution given some parameters $\overline{\theta}$ and a merger time $t_{\rm m} = 10Gyr$, $P(t_{\rm m} = 10 \text{ Gyr}|\overline{\theta})$ is the probability of the merger time being $t_{\rm m}$ and $\langle M \rangle$ is the average mass of the DBHs given the initial mass function for some parameters $\overline{\theta}$.

Finally, we need to weight the sensitive volume estimated by the search over the allowed values of the mass ratio for a population $\overline{\theta}$:

$$VT_i(M_c = m_i|\overline{\theta}) = \int_1^{q_{\rm max}} \mathcal{P}(q|m_i, t_{\rm m}, \overline{\theta}) VT(m_i) {\rm d}q. \tag{4.15}$$

Assuming that the DBH event counts within a chirp mass bin follows a Poisson distribution, we have that the rate posterior in the limit of no detection is given by:

$$P(\mathcal{R}|VT) = VT(\overline{\theta}) \exp(-\mathcal{R}(f, \overline{\theta}) \times VT(\overline{\theta})), \tag{4.16}$$

which in turn can be used to estimate the likelihood, $\mathcal{L}(f, \overline{\theta}; \mathcal{R}VT)$, as in Eq. 4.17:

$$\mathcal{L}(f, \overline{\theta}; \mathcal{R}VT) = \prod_i \frac{\int_{\mathcal{R}_i}^{\infty} \mathcal{P}_i(\mathcal{R}|f, \overline{\theta}, VT_i) {\rm d}\mathcal{R}}{\int_0^{\infty} \mathcal{P}_i(\mathcal{R}|f, \overline{\theta}, VT_i) {\rm d}\mathcal{R}}, \tag{4.17}$$

where 'i' iterates over all the chirp mass bins used. The likelihood in Eq. 4.17 along with the prior choices explained before can be used to estimate the 4D posterior in Eq. 4.13. As a last step, the posterior is marginalized over 'b' and 'r' so that only the pair given by (M_{min}, f) remains.

The PBH models explained in Sec. 4.1.2 and the DBH model in Sec. 4.2.1 will be constrained by the results of the searches in the following sections. The search was divided into two parts: O3a and O3b. The O3b search includes modifications with respect to the O3a search.

4.3 Search of subsolar objects in the first half of the third observing run

In this section we will discuss the main results from the subsolar mass search performed by the LVK collaboration in the first half of the third observation run [89].

4.3.1 Data and search pipelines

The first half of the third observation run (O3a) covers the period from 1 April 2019 1500 UTC to 1 October 2019 1500 UTC. As explained in Sec. 2.7, the data is characterized and calibrated prior to the analysis. A detailed description of the calibration can be found in the catalog papers [43, 136, 10, 137]. The data used for this search includes an additional non-linear removal of spectral lines [138, 11] that uses the information from sensors monitoring the interferometer to subtract non-stationary noise. The search is made by using three different matched-filtering (see Chapter 3 and Sec. 3.3.9) pipelines: GstLAL [80, 139, 140], MBTA [141], and PyCBC [70, 69, 91, 92, 93, 94]

The offline PyCBC pipeline is explained in Sec. 3.3.9. The configuration used is the same as in the O3a GW catalog (GWTC2) [10]. For this search, a log-likelihood model as a ranking statistic was used. This model uses the SNRs, time delay, phase difference, estimated background and the network performance at the time of the coincidence to rank each coincidence trigger.

The Multi-Band Template Analysis (MBTA) pipeline is a matched-filter pipeline that splits the filtering into two frequency bands to reduce the computational cost [73, 142]. The data is re-sampled to 4096Hz and gated following a procedure similar to PyCBC [92] prior to the matched-filter step. The configuration used for this search is the same used during the third observation run [141] but with two differences: longer time periods for the FFTs (from seconds to hundreds of seconds) and longer update time for the PSD estimation.

The GstLAL [80, 139, 140] pipeline is a matched-filter pipeline that performs the filtering in the time-domain. Furthermore, it also applies time-domain signal consistency tests. Candidates are ranked according to a likelihood ratio statistic based on their SNR, time delay, phase difference, sensitivity, signal population model [143] and statistical data quality from the iDQ algorithm [51, 144]. However, in this case, population models of subsolar mass objects are not known and a uniform template density is used. No gating is applied and the ranking statistic and configuration are the same as in Ref. [10].

In the next sections we will show the results obtained by the three different pipelines and the estimated sensitive volume.

4.3.2 Template bank

As was shown in Sec. 3.3.3, a template bank is needed for a search using matched-filter. The same bank was used by the three search pipelines. The bank was constructed following the geometric placement algorithm, described in Ref. [71], with a minimum match of 0.97 over the frequency band 45Hz-1000Hz. This algorithm works by defining a metric on the waveform parameter space and describing the missmatch between two close points using the metric (see Sec. 3.3.3). This minimum match was selected so that the loss of signals due to a discrete template bank is no more than 10%. The signals were modeled using the `TaylorF2` waveform, as explained in Refs. [145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155]. This model includes phase terms up to 3.5PN order but doesn't include amplitude corrections (see Sec. 1.3).

The masses covered by the template bank can be seen in Fig. 4.1. These masses are in the detector frame. In other words, they are redshifted due to the propagation of the GW: $m^d = (1+z)m$. As it can be seen, the primary mass, m_1, ranges from $0.2M_\odot$ to $10M_\odot$ while the secondary mass, m_2, takes values between $0.2M_\odot$ to $1M_\odot$. The lowest mass limit was selected for consistency with previous searches using the first and second observation runs [87, 88]. There is an additional constraint in the minimum possible mass ratio, $q \equiv m_2/m_1$. This parameter is constrained to values in the interval $q \in [0.1,\ 1.0]$. Spins aligned with the orbital angular momentum are also taken into account for the template bank generation [156].

In order to reduce the amount of computational resources needed, the spin range covered is limited as follows:

- For a component mass, m_i, with $m_i \geq 0.5M_\odot$, the dimensionless component spin is allowed to go up to 0.9.
- For a component mass, m_i, with $m_i \leq 0.5M_\odot$, the maximum dimensionless component spin is limited to 0.1.

Furthermore, precession is not considered to limit the size of the template bank for the same reason. The resulting template bank is composed of more than 1.8 million templates, twice as large as the template bank used in previous searches.

The matched-filter (see Sec. 3.3.2) lower frequency is 45Hz instead of the 15Hz employed in other searches [10]. This value was selected so that the search is not computationally prohibitive due to the long duration of the templates. The higher value for the lowest frequency translates into a maximum loss of signal-to-noise ratio of 9% which, in turn, implies a reduction of sensitive volume of 24%. For future searches, the implementation of a variable lower frequency is being explored. This approach would limit the duration of the signals (e.g. to 512s) instead, allowing matched-filter to recover more SNR.

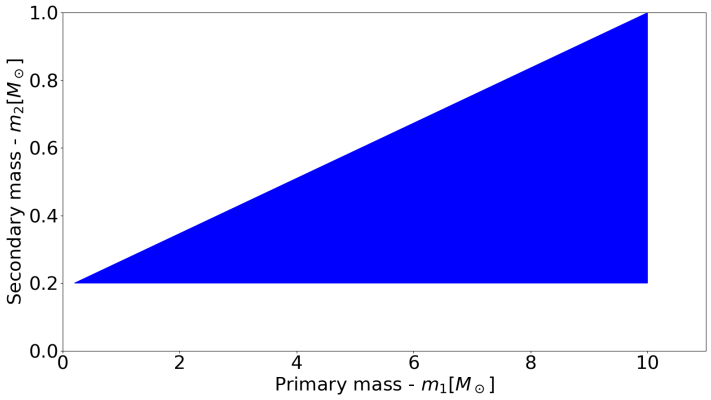

Figure 4.1: Template bank used in the O3a subsolar mass search by all the pipelines in the (m_1, m_2) plane.

4.3.3 Injections

An injection campaign per pipeline was made to estimate the sensitivity following the steps in Sec. 3.3.6. As seen in Eq. 3.29, a population model is needed to estimate the sensitive volume [95, 157, 158]. The choices used for this analysis are as follows:

- Masses: Source frame masses between $0.2M_\odot < m_1 < 10.0M_\odot$ and $0.2M_\odot < m_2 < 1.0M_\odot$. However, the masses in detector frame are larger than in the source frame ($m_1^d \in [0.22, 11.2]M_\odot$ and $m_2^d \in [0.22, 1.12]M_\odot$) and a detector frame mass cut was added so that the injections are covered by the template bank (given in terms of the detector frame masses).
- Aligned spins following the same distribution as the template bank.
- Sources uniform in comoving volume with random orientations and isotropically distributed.

Planck "TT,TE,EE+lowP+lensing+ext" cosmology [159] was used for all the calculations. These cosmological results are derived from the Planck observations of the temperature and polarization anisotropies of the cosmic microwave background. Additionally, reweighting, as explained in Eq. 3.38, was needed due to the original injection set not being distributed uniformly in comoving volume.

4.3.4 Results

Using the template bank defined previously, the O3a data was searched for GW from subsolar mass objects. No confident detection was found by the search pipelines. The most significant trigger had a FAR of 0.14 yr^{-1} which is not considered enough for claiming detection ($< 0.01\mathrm{yr}^{-1}$). Therefore, we will use the loudest ranking statistic, as in Sec. 3.3.8, to estimate an upper limit on the merger rate. We assume that this merger rate is constant as a function of the redshift, a reasonable assumption since the search is sensitive only to nearby sources ($z \lesssim 0.12$).

For this analysis, it is assumed that the sensitivity is, at first approximation, a function of the chirp mass [160], $\langle VT \rangle(M_c)$. Therefore, the sensitive volume is estimated in nine equally spaced chirp mass bins between $0.17M_\odot$ and $2.39M_\odot$. Each bin is denoted by the index 'i' and the associated sensitive-volume as $\langle VT \rangle_i$. The 90% confidence interval on the merger rate is estimated using the sensitive volume and Eq. 3.64.

The results of the search can be seen in Fig. 4.2 where the rate density upper limits as a function of chirp mass are shown for each pipeline. These upper limits are in the range $[220 - 24200]$ $\mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}$. Even though the pipelines generally agree, there are some discrepancies due to differences in the background estimation and ranking statistics used. These differences can lead to sensitive volume measurements that agree to within $\mathcal{O}(30\%)$. Furthermore, there is a loss of sensitivity in the upper chirp mass limit due to the injection set mass range being larger than the template bank.

Given these considerations, only the MBTA results, which shows a good compromise between the three pipelines, will be used to constrain the PBHs and DBHs models. Instrumental calibration errors [43] have been neglected for this work as in the GWTC 2.1 catalog and population papers detailed in Refs. [11, 161]. These errors are small in the sensitive region of interest with an expected maximum contribution of $\sim 10\%$ to the uncertainty of the sensitive volume.

In the following sections we will use these rate limits to add constraints to the two proposed models.

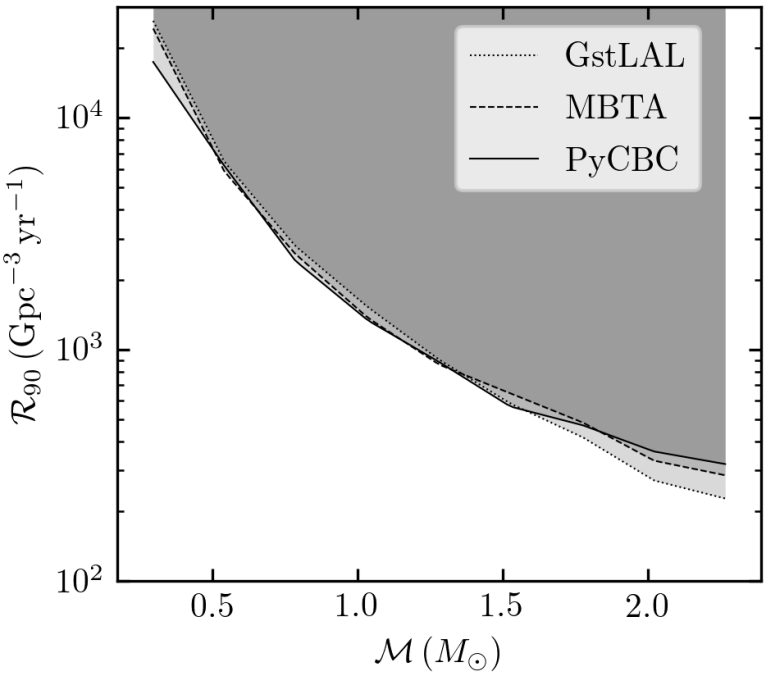

Figure 4.2: 90% confidence limit for the upper limit on the merger rate as a function of the chirp mass obtained by the different pipelines: GstLAL (dotted), MBTA (dashed) and PyCBC (solid). Taken from Ref. [89].

Limits in the fraction of primordial black holes

The estimated upper limits on the rate are used to constrain f_{PBH} assuming the PBH binary model given in Sec. 4.1.2 (3-body system + No suppression). The coalescence time given by Eq. 4.8 is evaluated using the Planck "TT,TE,EE+lowP+lensing+ext" cosmology [159], the same cosmology used to create the injection set.

We estimate f_{PBH} as a function of the component mass of the black holes in the binary, m_{PBH}. The model assumes a monochromatic distribution of masses ($q = 1$) and our injection set includes unequal mass binaries ($0.1 \leq q_{inj} \leq 1$), as indicated in Sec. 4.3.3. Therefore, given that the sensitivity of the detectors is mainly dependant on the chirp mass and not on the mass ratio, we assume that $R_{90}(\mathcal{M}, q = 1) \approx R_{90}(\mathcal{M})$ and limit the

estimation to masses $m_{1,2} < 1M_{\odot}$. This assumption allows us to use the upper limits seen in Fig. 4.2.

The results are shown in Fig. 4.3. Unexplored region of the parameter space is further constrained by our observations. It was found that, for a monochromatic model with masses between $0.2M_{\odot}$ and $1M_{\odot}$, an upper limit to the fraction of dark matter in PBHs is $f_{\mathrm{PBH}} \leq 6\%$. Fig. 4.3 also shows (model dependent) constraints from different methods and mass ranges. The high mass region ($m > 10M_{\odot}$) is constrained from observation of dwarf galaxies Eridanus II and Segue I. These constraints are obtained by observing the dynamics of dark matter dominated galaxies assuming that a fraction of the dark matter is made by massive black holes.

In the region below $10M_{\odot}$, limits were estimated by monitoring the Magellanic clouds for microlensing events caused by Massive Compact Halo Objects (MACHOS). A MACHO is a compact object with little (or no) radiation emission that might explain a fraction of dark matter. They are detected by the transient magnification of a star image produced when they pass in front of it (microlensing).

Constraining dark matter black holes

Constrains are estimated on the DBH model explained in Sec. 4.2. The analysis is performed using the likelihood and posterior model from Sec 4.2.1.

Using this method, the results presented in Fig. 4.4 are obtained. It was found that that the lowest upper limit is located at $M_{min} = 1M_{\odot}$ corresponding to $f_{\mathrm{DBH}} = 0.003\%$. Given the fact that the minimum mass is set by the Chandrasekhar limit, which depends on the mass of the heaviest fermionic particle [101], m_x, these limits can be used to constrain the range of values allowed for the mass of this particle. It was found that m_x is in the range $0.66GeV/c^2$ to $8.8GeV/c^2$. Below $M_{min} < 2 \times 10^{-2}\mathrm{M}_{\odot}$ no meaningful constraints can be set since, with the current sensitivity, none of the black holes in the population would be detected.

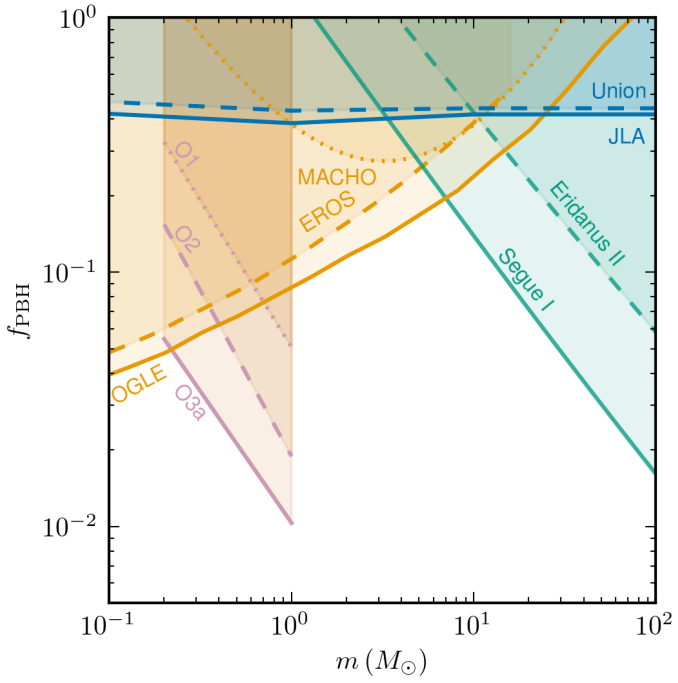

Figure 4.3: Upper limits on $f_{\rm PBH}$ as a function of the component mass of the black hole in the source frame for the model given in Sec. 4.1.2 (3-body system + No suppression). Previous results from the LVK collaboration are shown [87, 88] along with the new O3a results [89]. Constraints by non-GWs are shown as comparison: (Orange) Microlensing from MACHO [162],EROS [163], and OGLE [164]; (blue) supernova lensing (Union and JLA) [165]; (green) dynamical constraints from dwarf galaxies (Segue I and Eridanus II) observations [166, 167]. It should be noted that there is a model dependency on each constraint. Taken from Ref. [89].

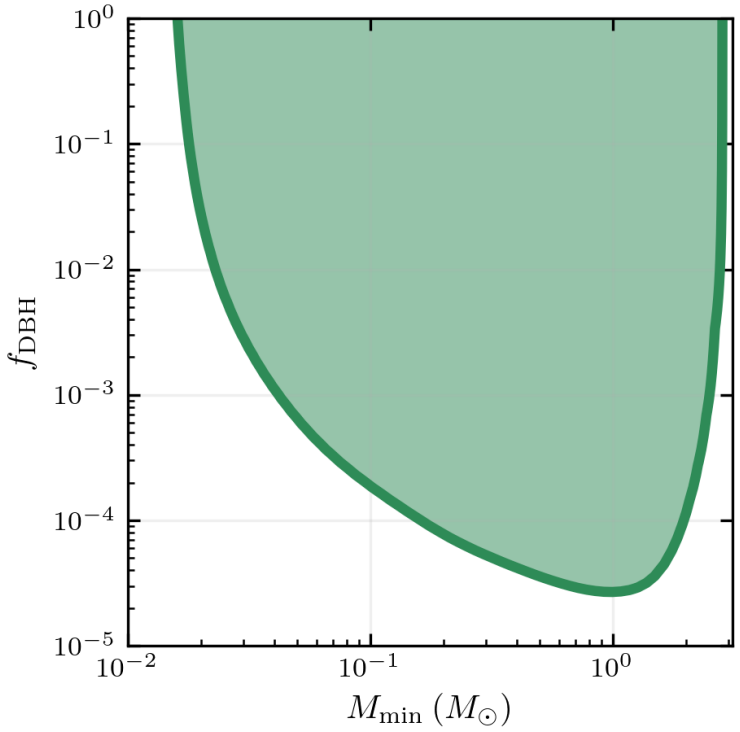

Figure 4.4: Limits on the fraction of dark matter that collapses into dark matter black holes, f_{DBH}, as a function of the minimum mass, M_{min}, in the source frame for the model given in Sec. 4.2.1. Taken from Ref. [89].

4.4 Search of subsolar objects in the second half of the third observing run

The subsolar mass search was extended to the second half of the third observing run, covering the period from 1 November 2019 1500 UTC to 27 March 2020 1700 UTC. The data was calibrated in the same way as in Sec. 4.3.1 and searched with the same pipelines. The template bank is the same as the one depicted in Sec. 4.3.2. Improvements with respect to the O3a search were made in the creation of the injection set and PBH model used. Furthermore, this search used the whole O3, increasing the observation time by a factor two with respect to the O3a search, to constrain the two models.

4.4.1 Injections

The injection set had slight modifications from what was used in O3a. For this search, a common injection set among all the pipelines was used. The set was designed as follows:

- Source component masses following a log-uniform distribution with $m_1 \in (0.19, 11.0)$ $M_\odot$ and $m_2 \in (0.19, 1.1)$ $M_\odot$.
- Component spins isostropically distributed with dimensionless spin up to 0.1.
- Distributed uniformly in comoving volume up to redshift $z = 0.2$. Therefore, no reweighting was needed for this analysis.

Furthermore, the sensitivity over the O3a period is re-estimated using this new injection set. This was done by covering the full O3 period with approximately 2 million signals.

4.4.2 Results

The O3b period was searched by cross correlating the template bank with the data. There is no statistically significant evidence of GW from a subsolar mass source in O3b. In Table 4.1 the three most significant candidates, down to a threshold of FAR $< 2\,\mathrm{yr}^{-1}$, are reported. No data quality issues were identified by a visual inspection around the time of the triggers. However, the number of triggers with their estimated FAR is consistent with what we would expect if no astrophysical signal was present in the data, given that three pipelines were used and an observation time of 0.34 yr.

FAR [yr^{-1}]	Pipeline	GPS time	m_1 [$M_\odot$]	m_2 [$M_\odot$]	χ_1	χ_2	Network SNR
0.20	`GstLAL`	1267725971.02	0.78	0.23	0.57	0.02	8.90
1.37	`MBTA`	1259157749.53	0.40	0.24	0.10	−0.05	10.25
1.56	`GstLAL`	1264750045.02	1.52	0.37	0.49	0.10	9.10

Table 4.1: The triggers with a FAR $< 2\mathrm{yr}^{-1}$ in at least one search pipeline. The parameters of the template associated with the trigger are shown for each candidate: m_1 and m_2, the redshifted component masses, and χ_1 and χ_2, the dimensionless component spin.

As in the case of the O3a search, the results from the search and the injection campaign were used to estimate the sensitive volume. Each pipeline used all injections with $q > 0.05$. A 90% confidence interval for the sensitive volume was estimated considering binomial errors in the efficiency, as in Eq. 4.18:

$$\delta(VT) = 1.645\sqrt{\frac{\epsilon(1-\epsilon)}{N_{inj}}}V_{inj}T_{obs}, \tag{4.18}$$

where ϵ is the efficiency, N_{inj} are the total injections, V_{inj} is the volume at the farthest injection and T_{obs} is the observation time. Though PyCBC and MBTA used the full injection set, GstLAL analyzed a subset, resulting in larger uncertainties in the sensitive volume. The population was split into nine chirp mass bins equally spaced within the range $0.16M_{\odot} \leq \mathcal{M} \leq 2.72M_{\odot}$.

The sensitive volume as a function of the chirp mass is presented in Fig. 4.5. The estimated sensitive volume is approximately twice as large as the one from O3a data, in agreement with the increased observation time. The drop in sensitivity in the highest chirp mass bin is due to the component masses contained within that bin being beyond the scope of the template bank. The re-estimated O3a sensitive volume, using the common injection set, was found to be compatible with the results shown in the previous analysis (see Sec. 4.3). Furthermore, between chirp masses of $1.3\mathrm{M}_{\odot}$ and $2.3\mathrm{M}_{\odot}$, the sensitive volume is compatible with the estimation presented in the GWTC-3 catalog [66].

An upper limit on the merger rate was obtained by using the loudest ranking statistic, as in Sec. 3.3.8. Each pipeline used the most significant candidate in the full O3 as the FAR threshold: $0.2yr^{-1}$, $1.4yr^{-1}$ and $0.14yr^{-1}$ for GstLAL, MBTA and PyCBC respectively. The results as a function of the chirp mass are shown in Fig. 4.6. Upper limits were also obtained as a function of the source component masses,(m_1, m_2), in Fig. 4.7. Similarly to the O3a search, these results will be used to constrain two models of subsolar mass objects.

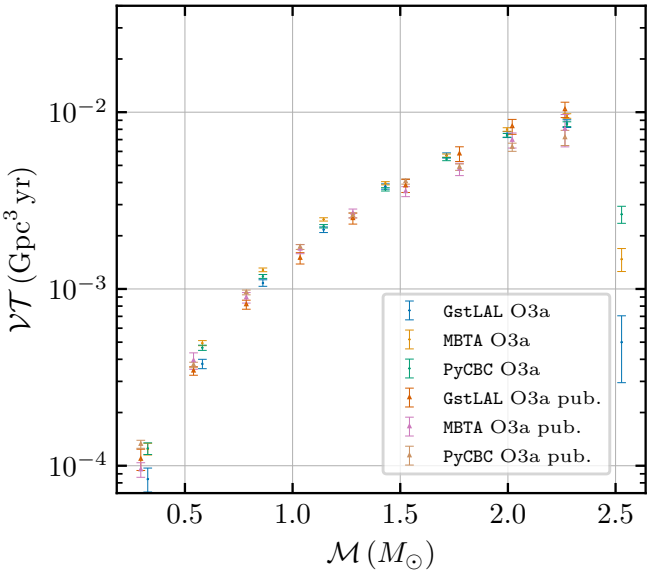

Figure 4.5: Sensitive volume as a function of the chirp mass for each pipeline. The results from the O3a analysis are shown as a comparison. Taken from Ref. [99].

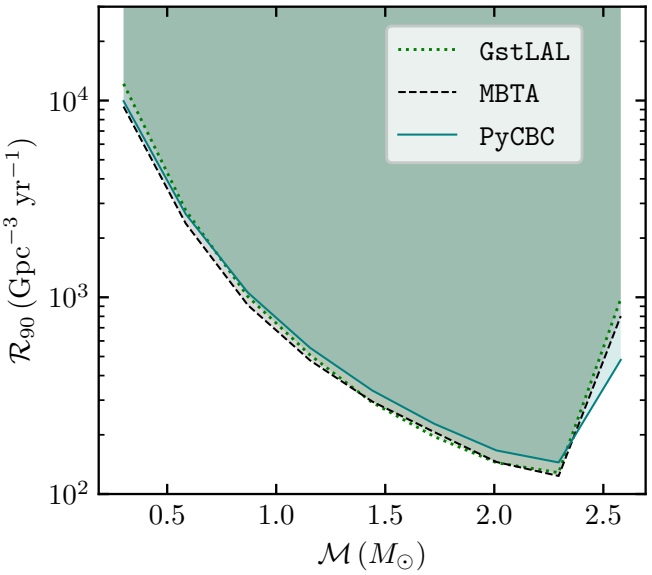

Figure 4.6: Upper limits for the merge rate as a function of the source frame chirp mass for the full O3. Taken from Ref. [99].

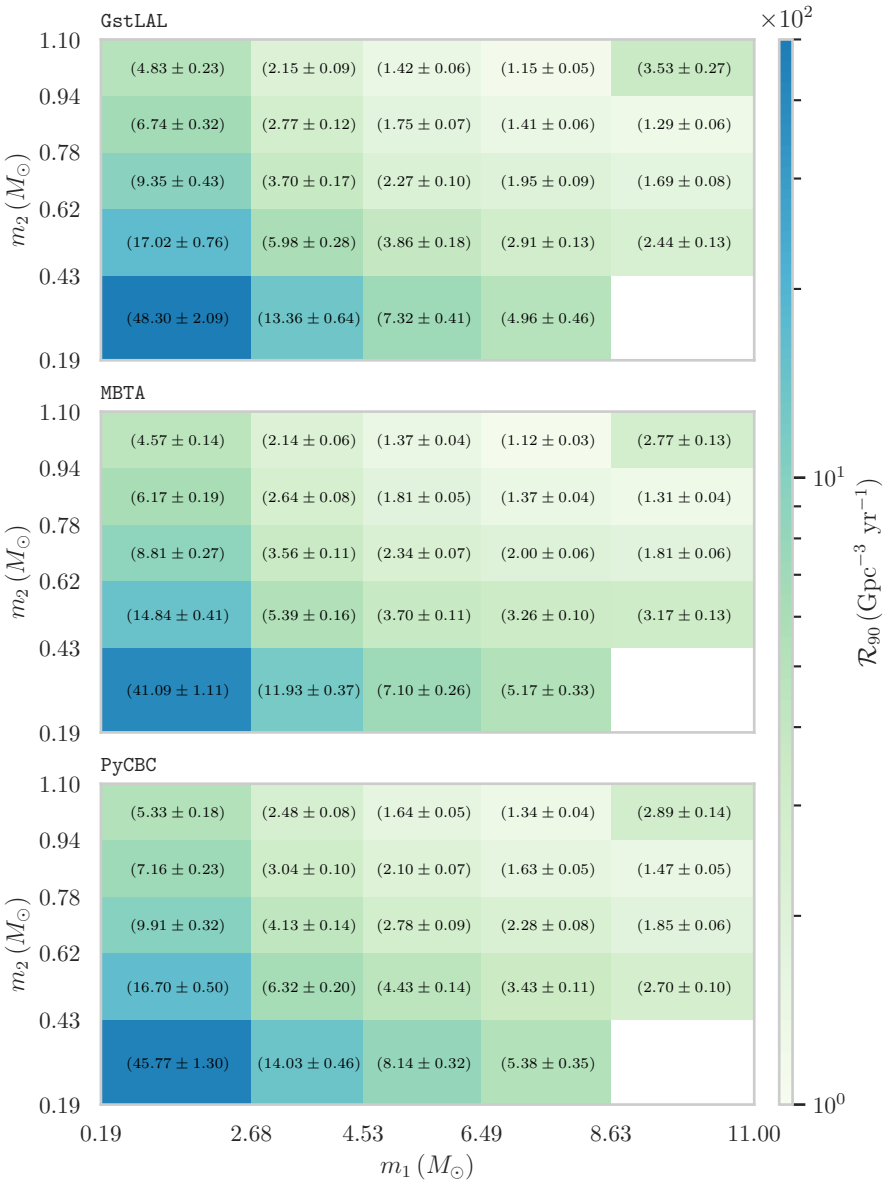

Figure 4.7: Upper limits for the merge rate as a function of the source frame component masses for the full O3. Taken from Ref. [99].

Limits in the fraction of primordial black holes

The estimated upper limits are used to constrain the PBH binary model given in Sec. 4.1.2 (N-body system + Rate suppression). The obtained results are less stringent with respect to the O2 and O3a limits due to the inclusion of suppression effects in the merger rate. Furthermore, due to the uncertainty and caveats in the derivation of the merger rate for early binaries, a merger rate for late binaries, given by Eq. 4.6, is also considered. These caveats include the initial spatial PBH distribution, assumed to be Poissonian, and PBH cluster dynamics. This allows us to show the variation in the PBH limits depending on the binary formation scenario.

The results can be found in Fig. 4.8, where $f_{\rm PBH}$ was not restricted to values lower than one as to show the results from late binaries. In the case of late binaries, no significant limits were found within the searched mass range. For early binaries, the scenario $f_{\rm PBH}$=1 becomes strongly disfavored up to $1M_\odot$. In particular, we have found that $f_{\rm PBH} < 0.5$ at $0.3M_\odot$ and is below 0.08 at $1M_\odot$. These results provide a looser constrain to $f_{\rm PBH}$ compared to the LVK O2 subsolar mass search [88] when no suppression factor was implemented. This is not the case when the suppression factor is included in the estimation.

The rates given in Fig. 4.7 can be used to constrain unequal mass merger rates. Unfortunately, these are more uncertain and model dependent. Unequal PBH mass distribution allows us to explain dark matter even if it has a small contribution in a particular mass range. However, it introduces the problem that the extended mass function should not violate the constraints at all mass scales. This is particularly important in the case of PBH formation extending over decades, as it would affect the mass of the collapsed PBHs (See Eq. 4.1).

Under these considerations, we can obtain a limit on an effective parameter:

$$F_{\rm PBH} \equiv \left(\frac{f_{\rm sup}}{2.4\times10^{-3}}\right) f(m_1) f(m_2) f_{\rm PBH}^{53/37} \ , \tag{4.19}$$

defined in such a way that it corresponds to the product of the mass distributions in the scenario of $f_{pbh} \approx 1$. Such definition allows us to define a model-independent limit since the uncertainties on the mass distribution and suppression factor are encompassed by $F_{\rm PBH}$.

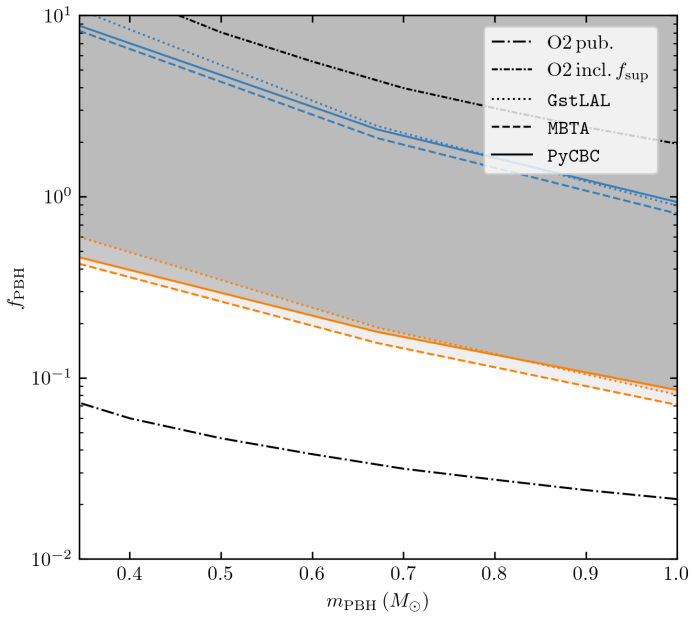

Figure 4.8: Constraints on $f_{\rm PBH}$ for a monochromatic mass function for early binaries (orange) and late PBH binaries (blue). In black, the results from O2 [87] are shown with and without the suppression factor. $f_{\rm PBH}$=1 is excluded for the first time in the whole subsolar mass range for early binaries. Taken from Ref. [99].

It was found that $F_{\rm PBH}$ is sensitive to the location on the $m_1 - m_2$ plane by using the limits shown in Fig. 4.7 and theoretical rates from Eq. 4.10.

In particular, this allows us to restrict possible PBH distributions in the SSM range for models with $f_{\rm PBH} = 1$ and a peak above $1M_\odot$. We find that some distributions with QCD enhanced features, as explained in Refs. [110, 116, 168, 169], are constrained in the range $f_{\rm PBH} \approx (0.1 - 1)$. This result implies that subsolar mass searches are complementary to searches in the solar mass range in order to distinguish viable PBH mass functions.

Constraining dark matter black holes

As in the case of the O3a search, we consider the model introduced in Sec. 4.2.1. The updated limits using the full O3 are shown in Fig. 4.9.

The strictest limit was derived at $1M_{\odot}$ for $f_{\mathrm{DBH}} < 0.0012 - 0.0014\%$, which corresponds to an improvement of a factor two from the previous O3a constrains. From these results we are also able to constraints the mass of the heavy dark matter particle, m_x, due to the relation with the Chandrasekhar mass for dark matter black holes given by Eq. 4.12. We have found that the mass is constrained in the following range: $1.1 GeV/c^2$ to $8.9 GeV/c^2$. No constraints are given below $0.01M_{\odot}$ due to the searches not being sensitive enough to support distributions within that region.

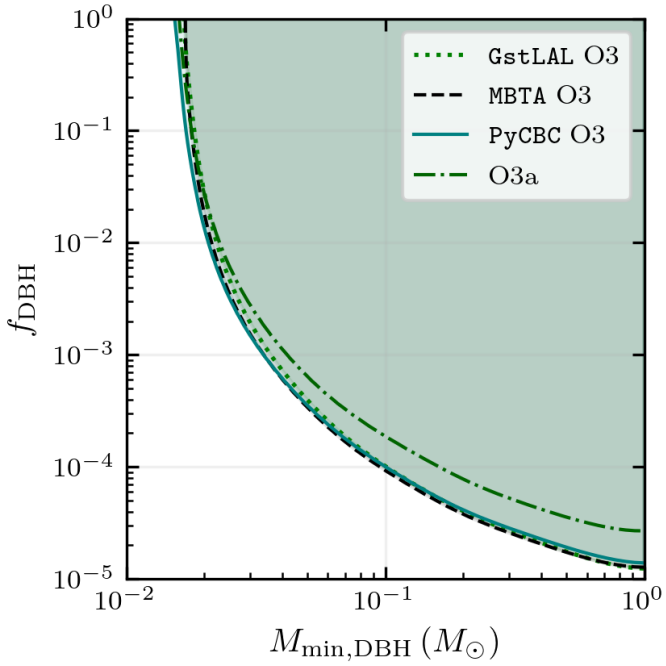

Figure 4.9: Constraints on f_{DBH} as a function of the minimum mass, M_{min} for the three pipelines using the full O3 for the model given in Sec. 4.2.1. The results from the previous O3a search are shown as a comparison. Taken from Ref. [99].

Chapter 5

Machine learning methods for early detection of gravitational waves

In this chapter, a machine learning implementation for the detection of GWs using O2 [170] and O3 data is explored. The implemented searches cover the mass range explored in the catalog papers [9, 10, 11, 66]. The major motivation for using these algorithms arises from the fact that a good performance can be achieved with less computational resources in a fast way, helping with low latency analysis and multi messenger astronomy by speeding up the detection of GWs. In Ref. [171] a review of machine learning method implemented for GW astronomy can be found.

The chapter is divided as follows: First, the basic elements to understand machine learning are discussed along with the architecture used for the search. Second, the training, testing and performance studies are explained. Finally, the whole O2 and O3 are scanned and the results are compared to the confirmed events from the released catalogs.

5.1 Introduction to machine learning

Machine learning algorithms [172] can be defined in broad terms as algorithms that, from a set of data, learn how to perform a task.

The basic concepts for every machine learning algorithm are:

- A task we want our algorithm to learn how to perform. For example, in a classification task the algorithm should learn how to classify an input in different categories. Another examples of tasks are denoising, regression or estimation of density functions.
- A function to quantitatively estimate the performance of the algorithm.
- A set of data from where it should learn how to perform the task.

During the learning process, the objective is to reach the highest performance possible. This is achieved by minimizing a function that adequately describes the desired task. This function, usually called loss or cost function [173], is a map between some values and a real number that describes the associated cost that, in an optimization problem, has to be minimized. This learning process is also referred as "Training". An example of a commonly used loss function for binary classification is the binary cross-entropy, seen in Eq. 5.1:

$$H_q(p) = -\frac{1}{N}\sum_{i}^{N} y_i log(p(y_i)) + (1 - y_1) log(1 - p(y_i)), \tag{5.1}$$

where y_i are the true labels of the input, $p(y_i)$ are the values found by the algorithm and N is the total number of inputs we are using. This cost function models how much information we are losing by describing the distribution 'q' using the estimated distribution 'p'.

Another example of a loss function is the Mean Squared Error (MSE), defined in Eq. 5.2:

$$MSE = \frac{1}{N}\sum_{i}^{N} (y_i - \hat{y}_i)^2, \tag{5.2}$$

generally used for regression problems where y_i is the real data point and $\hat{y}_i$ the value obtained by the algorithm. The minimum value of these loss functions is found by using optimization algorithms particularly suitable for complex problems, generally by using a numerically estimated gradient to move towards the minimum of the function.

Once the learning is finished, the performance can be studied over a new data set never used during the training, the testing set. This data set is used

to understand how it generalizes to new data and estimate the generalization error of the algorithm (error between estimation and real when the testing set is used). The ultimate objective of this procedure is to train an algorithm to perform a task over data never seen before with a good performance.

However, it might happen that even though the algorithm has a very good performance over the train data set, it is unable to provide good predictions when a new data set, different from the training set, is used. This is known as overfitting and occurs when the generalization error is too large in comparison to the training error. In general, we want our algorithm to learn in such a way that not only the training error is small, but that the gap with respect to the generalization error is also small. For this reason, the evolution of the training is monitored using another set, the "validation set", that is not used for the training.

5.1.1 Linear regression

In this section we will discuss a simple machine learning problem: linear regression. The objective is to find a set of parameters, $w \in \mathcal{R}^n$, that, for a given $x \in \mathcal{R}^n$ as input, predicts $\hat{y} \in \mathcal{R}$ as follows:

$$\hat{y} = w^T x + b, \tag{5.3}$$

where 'b' is the intercept, also called "bias", and the vector 'w' is given the name of weights. The next step is to define how to estimate the performance/loss function. In this case, we will use the MSE defined in Eq. 5.2, a loss function commonly used for regression problems.

Therefore, given a set of data (x^{train}, y^{train}), we want to find the set of weights such that the value in Eq. 5.4 is minimized:

$$MSE_{train}(w) = \frac{1}{m}\sum_{i}^{m}(y^{train} - \hat{y})^2 = \frac{1}{m}\sum_{i}^{m}(y^{train} - (w^T x + b))^2. \tag{5.4}$$

This problem can be solved both numerically and analytically by deducing the expression of the gradient as a function of the weights and searching for the value of w that makes it zero, that is, $\nabla_w MSE_{train} = 0$.

5.1.2 Layers and activation functions

Until now, we have been talking about machine learning algorithms but we have not yet defined how those are constructed. In general terms, a machine learning method is constructed by stacking blocks that transform an input into some output, usually called layers. Some of these layers (though not all of them) are defined by some weights that control this transformation. Inside a layer, these small blocks that perform transformations over the input are called "neurons". A neuron works as seen in Eq. 5.5:

$$y = w^T x + b. \tag{5.5}$$

However, a limitation of this approach is that the method is only able to learn linear relations. This is solved by composing the output 'y' with a different function, called the activation function, to introduce non linearities into the problem. In other words, the neuron takes some input 'x' and outputs 'y' which is then passed through a function that will decide if the neuron activates or not. The output of a neuron after passing through an activation function, σ, is given by Eq. 5.6:

$$y = \sigma(w^T x + b). \tag{5.6}$$

An example of an activation function is the threshold function, $\sigma(x)$, defined as $\sigma(x) = 1$ if $x > 0$ and $\sigma(x) = 0$ otherwise. This transforms the output into a Boolean variable. There are many activation functions and they have to be carefully chosen so that the algorithm is able to learn how to perform the task.

These layers, along with their activation functions, are used to define the machine learning algorithm, or architecture. Some of the most commonly used layers are:

1. Dense layers (also known as fully connected layers): This is the most common layer associated with machine learning. It refers to a layer in which all the neurons are connected to the neurons of the previous layer.

2. Convolutional layers: Applies a convolution to the input data and pass the result to the next layer. Widely used in image recognition to highlight important patterns before feeding them to the next layers.

3. Pooling layers [174]: Used to reduce the dimensions of the data by combining the output in some way, usually using the mean or maximum. This layer doesn't have weights.

4. Batch normalization layers [175]: Standardizes the input, x^{in}, as follows:

$$x^{out} = \frac{x^{in} - \mu(x^{in})}{\sqrt{Variance(x^{in})}}$$

   This layer doesn't have weights.

All of this makes neural networks very suitable to be represented in graph formats, as in Fig. 5.1, where all the connections can be shown along with the flow of the data.

5.1.3 Convolutional neural networks

A particular family of machine learning algorithms are the convolutional neural networks (CNNs) [172]. In broad strokes, these are algorithms in which at least one of the layers is a convolutional layer. They are specialized to be used in problems in which the data has a grid-like structure, like, for example, spectrograms and their time-frequency bins. A schematic representation of a CNN is shown in Fig. 5.2.

A convolutional layer applies a convolution to the input data, that is, it transforms the input as in Eq. 5.7:

$$O(t) = (f * w)(t) = \int f(\tau) w(t - \tau) d\tau, \tag{5.7}$$

where $f(\tau)$ is known as the input, 'w' as the kernel and $O(t)$ as the feature map. In the particular case when the variable 't' only takes integer values, we can define the discrete convolution as in Eq. 5.8:

$$O(t) = \sum_{n=\infty}^{\infty} f(n) w(t - n), \tag{5.8}$$

which is useful when the input is given in a grid. Notice that even though we have an infinite sum, in practise our data is limited and can be assumed to be zero outside the data points. The notion can be generalized to more

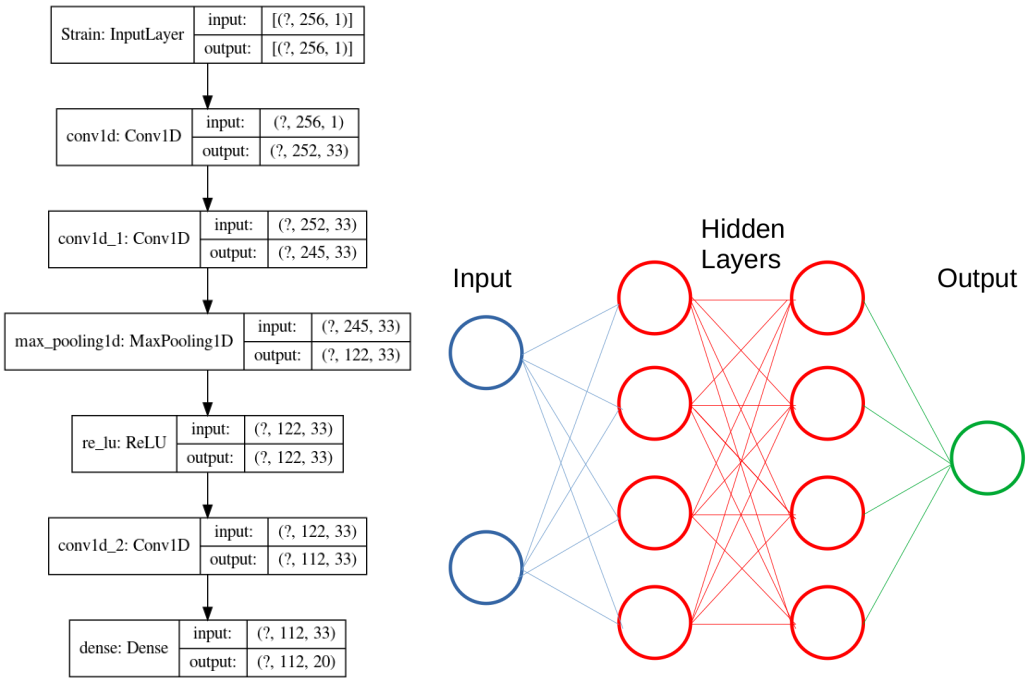

Figure 5.1: (left) Example of a graph from a convolutional neural network. The arrows point to the direction that the data takes and how its dimensions change with each one. The first dimension corresponds to the batch of the input, that is, how many examples we introduced. (right) Graphical representation of a neural network showing the input, output and the hidden layers in which the calculations are done.

dimensions by summing over the different axes. For example, for a two dimensional case, we have that the discrete convolution is given by:

$$O(i,j) = (f * K)(i,j) = \sum_m \sum_n f(m,n)K(i-m, j-n), \tag{5.9}$$

usually used in the case of two dimensional images. However, for most of the machine learning algorithms, what is really implemented is not the convolution but the cross-correlation, given by Eq. 5.10:

$$O(i,j) = (f * K)(i,j) = \sum_m \sum_n f(i+m, j+n)K(m,n). \tag{5.10}$$

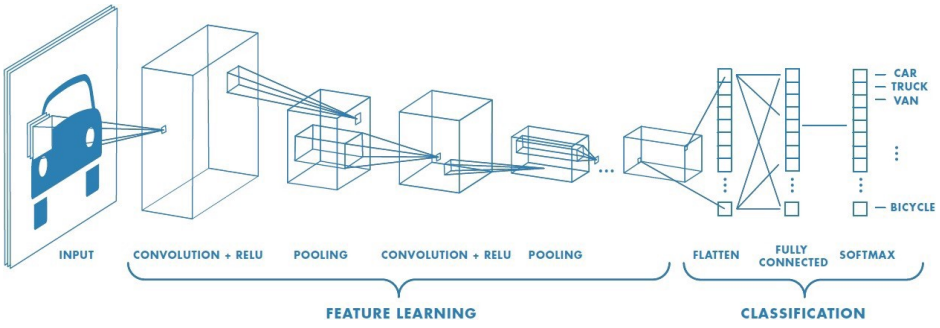

Figure 5.2: Representation of a convolutional neural network. Source: https://towardsdatascience.com/

This is not a problem in machine learning given that cross-correlation and convolution are related by how the elements in the kernel are distributed. During the learning phase, it will learn the optimal way to assign the kernel values and will produce the same output. For this reason, in machine learning, these two operations are called the same way. An example of a 2D convolution can be seen in Fig. 5.3.

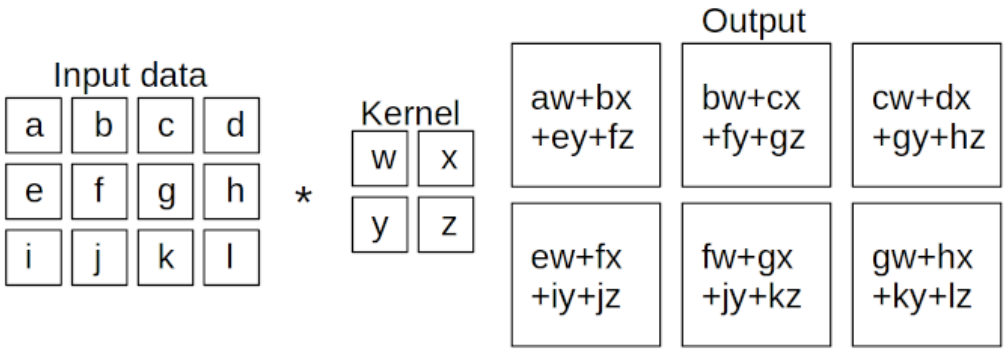

Figure 5.3: Discrete 2-D convolution with a 2x2 kernel that we slide through the elements of the input.

As we can see from Fig. 5.3, usually the kernel is smaller than the input data. The operations are done by moving the kernel through the input in independent steps along each axis.

This is usually known as the stride and, for example, in Fig. 5.3 we have a stride of $(1, 1)$. In other words, the kernel moves in steps of one element in each direction. There might be cases where the size of the input data is not big enough for the kernel size and stride used. In that situation, we add elements to the input data so that this operation is possible. This is known as "padding" and there are several ways to do it. For example, "zero padding" adds zeroes at the edges of the input.

The main advantage of convolutional layers is that the extraction of shapes and features from the input data can be achieved with small kernels, with respect to the input size, and reduces the amount of memory needed in comparison to traditional matrix multiplication. Furthermore, while in traditional dense layers each weight is used only once, in convolutional layers the elements of the kernel are used in all the positions. This parameter sharing means that it only needs to learn one set of parameters. This, again, reduces memory consumption in comparison to traditional problems.

Moreover, the input data can be introduced separated in different inputs, known as channels, each one with its own feature. The kernel can be applied to all the channels in the same way. For example, a 2D image can be decomposed into three colors: red, green and blue (RGB decomposition). Each color can be used as a different channel and passed through the convolutional layers which will apply the kernel to extract the feature map.

A common block of layers is done by stacking convolutional layers followed by a pooling layer. The reason is that pooling helps to achieve invariance under translations during the learning process [172, 174]. In other words, small shifts in the input will produce similar pooled outputs. This is useful for recognizing features from images, like faces or the characteristic chirp-like structure of GWs. Furthermore, pooling also reduces the dimensions of the data and the computational resources needed along with it.

5.2 The architecture: ResNet50

For this work we will use an architecture based on the ResNet (Residual Networks) architecture [176] to detect the GW produced by CBC events (See Fig. 5.4). This family of neural networks was designed to solve the vanishing gradient problem.

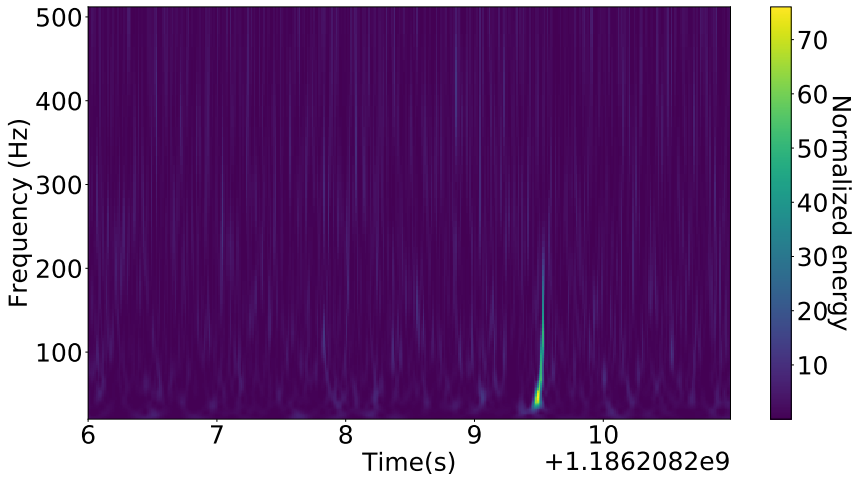

Figure 5.4: Injected BBH GW in the L1 interferometer as a two-dimensional image in time versus frequency with BH masses of 51 $M_\odot$ and 53 $M_\odot$ at a distance $D_L = 664$ Mpcs.

The vanishing gradient problem was a common issue that appeared in the training when too many layers were stacked on top of each other. During the training, the weights are updated using backgpropagation in such a way that the loss function is minimized. The gradient of each layer (gradient of the cost function with respect to the weights of the layer) is then used to estimate the direction and value to update the weights. However, when the network is too deep (many layers), this gradient becomes too small, reaching a point at which the update of the weights is insignificant, making it impossible to reach the optimal point.

These networks apply a method called skip connections, see Fig. 5.5, which allows the neural network to map the identity function (That is, a function

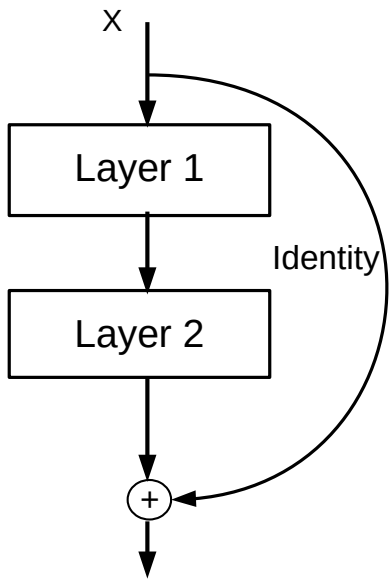

Figure 5.5: Graphical interpretation of skip connection

$I(x)$ such that $\forall x$ I(x) = x) and for the gradient to flow. Using this method, it has been shown that ResNet architectures have better training and test performance than smaller architectures. There are several versions of this architecture, but the most prominent are ResNet18, ResNet50 and ResNet101, with the number being the number of layers that particular architecture has.

For our work we will use ResNet50, shown in Table 5.1, as the default architecture with the binary cross-entropy as loss function and the optimization algorithm Adam [177]. The objective is for the neural network to classify the background input as zero and the signals as one.

5.3 Second observation run (O2)

In this section, the results of the search implemented using neural networks [170] in O2 open data [178] are discussed. The second observation run covers the period between November 30, 2016 up to August 25, 2017 with the AdV detector joining from August 1, 2017.

Layer name	Output size	Layer structure
conv1	112×112	7×7, 64, stride 2
conv2_x	56×56	3×3 max pool, stride 2 $\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times3$
conv3_x	28×28	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times4$
conv4_x	14×14	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times6$
conv5_x	7×7	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times3$
	1×1	Global average pool, 1-d fc, sigmoid
Hyper parameters		
Learning rate		0.01
Mini batch size		32
Maximum number of epochs		12
Optimizer		Adam
Loss function		Binary-cross entropy

Table 5.1: ResNet50 architecture and hyperparameters used during the training. Downsampling is performed by conv3_1, conv4_1, and conv5_1 with a stride of 2. Taken from Ref. [176].

5.3.1 Data preparation

Data from the three interferometers with a sampling rate of 4096Hz is used to construct the training and testing sets. The training set consists of 5s images with an even mix between background samples and simulated signal samples. An image duration of 5s was selected as it provides a good compromise between the size of the image and the duration of the GWs. Furthermore, special precaution was taken to avoid including any of the identified GWs events in the GWTC-1 catalog [9].

The parameters of the simulated signals are drawn following a uniform distribution between the limits given in Table 5.3.1. To constrain the size of the parameter space, we vary only seven parameters of the waveform: the masses, luminosity distance, sky position, polarization and inclination.

	Low mass	High mass
M_1, M_2 [$M_\odot$]	[0.2,2]	[25,100]
Dl [Mpc]	[1,50]	[100,1000]
RA [Rad]	[0, 2π]	[0, 2π]
Dec [Rad]	[0, π]	[0, π]
Pol [Rad]	[0, π]	[0, π]
Inc [Rad]	[0, $\frac{\pi}{2}$]	[0, $\frac{\pi}{2}$]

Table 5.2: Limits in O2 with masses in detector frame.

We use the IMRPhenomPv2 approximant as the waveform model. Moreover, the parameter space is divided into two mass regions: a low mass region and a high mass region. This choice was made to better train the neural network as the morphology of the signal changes greatly with the mass of the system. Low mass signals have longer tails and last longer in the sensitive band of the interferometer (see Fig. 5.6 left) while high mass signals are sharper in frequency (see Fig. 5.6 right).

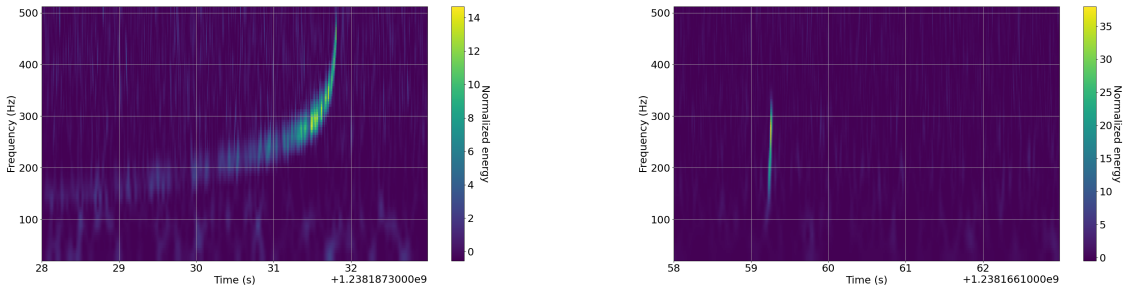

Figure 5.6: Example of a spectrogram as a two dimensional image in frequency versus time corresponding to two signals: low mass signal (left) and high mass signal (right).

In order to control the duration of the signal, a low frequency threshold of 80Hz (25Hz) is applied for the low mass (high mass) signals. The signal duration is limited to 5s counting backwards from the merger time to remove low frequency components that might confuse the neural network. Finally, the signals are randomly placed within the 5s window of each image.

In total, O(250000) signal and background samples are generated. These samples are whitened [61, 62] following the recipe from Eq. 3.5 so that the noise is reduced and the signals can be seen. From the whitened strain, spectrograms [179] with 400 bins in time and 100 bins in frequency are made. This choice of binning was taken as a good compromise between resolution and size of each image. The spectrograms are limited to the range [20,512]Hz corresponding to the most sensitive region of the interferometers. Due to the high variation between the values of the images (faint and loud signals are mixed in the data), the image is standardized as follows:

$$\hat{X} = \frac{X - \mu_X}{\sigma_X}, \tag{5.11}$$

where X is the matrix containing the pixels of the image and μ_X and σ_X are the mean and standard deviation of each image. The data sets are now ready to be split and used for training and testing.

5.3.2 Testing the performance

The training is made using the python package Keras [180] and Graphics processing units (GPUs) [181]. GPUs are used to speed up the training and testing of the neural network due to their capability to perform simultaneous computations. The images are randomly distributed in batches of 32 (batch size) evenly mixed between background and signal for a total of 122.000 images. From these, 10% are used as validation to check the performance during the training and select the best epoch for our analysis. The optimal epoch corresponds to the epoch with the best performance over the validation set.

For each mass region, the possibility of introducing data from more than one interferometer as input was studied. This should allow the neural network to recognize that a real signal should appear in all the input images and help to discriminate between them and glitches. Therefore, a neural network per combination of inputs is trained: H1-L1, H1-V1 and L1-V1. The possibility of using only one interferometer as an input was also explored but a performance worse than with a combination was achieved. For O2, the triple combination (H1-L1-V1) was not considered due to the small amount of data.

An example of the output of the neural network can be found in Fig. 5.7, where the discrimination between background (blue) and signal (orange) can be seen.

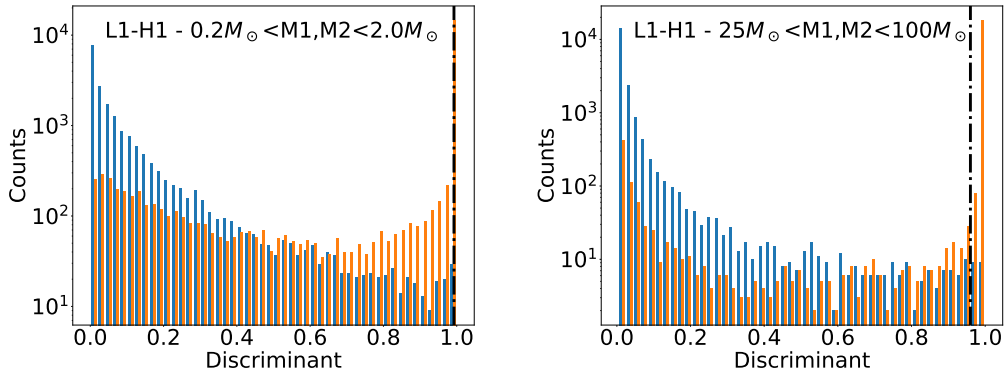

Figure 5.7: CNN output corresponding to H1-L1 for background (blue) and signal (orange). The distributions for the two implemented neural networks are shown: (left) low mass and (right) high mass. The dashed-dotted line indicates the threshold from Table 5.3. Taken from Ref. [170].

The performance of the NNs in terms of true positive rate and false positive rate can be seen in Fig. 5.8. The highest performance is achieved by using as input the data from H1-L1 which corresponds to the most sensitive combination.

From the curves in Fig. 5.8, a threshold for binary classification can be selected by limiting the number of false positives per day. However, the threshold was chosen by limiting the number of false positives estimated over a sample of the background to 25 events/day. The final values are in Table 5.3 and the thresholds in the first column will be used to claim detection in the following sections.

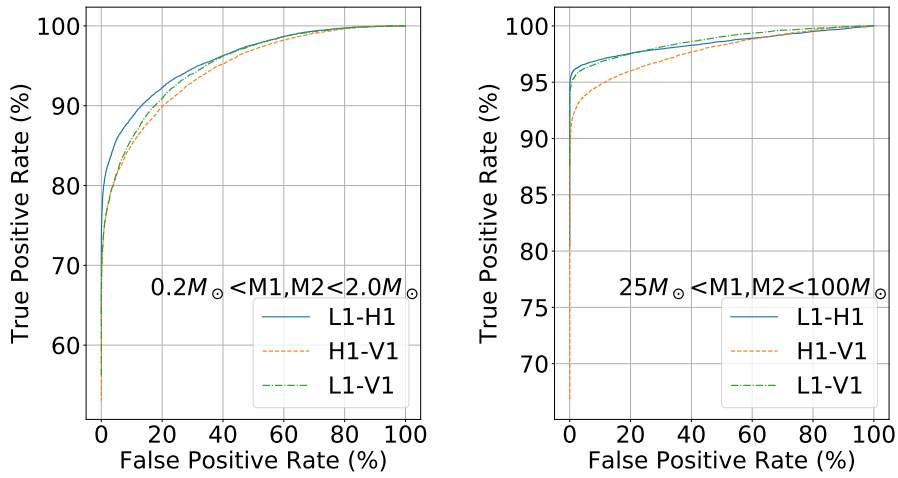

Figure 5.8: ROC curves for CNNs using pairs of interferometers inputs for (left) low mass and (right) high mass. Taken from Ref. [170]

	double interferometer channel		
	CNN discriminant (%) low/high mass	TP rate (%) low/high mass	FP rate (%) low/high mass
L1 – H1	99/96	74/95	0.06/0.09
L1 – V1	99/97	69/93	0.09/0.04
H1 – V1	99/98	66/88	0.03/0.05

Table 5.3: CNN discriminant and true positive (TP) and false positive (FP) rates from the ROC curves. Taken from Ref. [170]

5.3.3 Injection test

The performance of the neural networks is studied by preparing a data set consisting of only signals with a known optimal signal-to-noise ratio (SNR or ρ), defined in Eq. 3.12. The injected signals follow the same distribution as the training set but are re-scaled to a target SNR before being injected into real noise. This is achieved by multiplying the amplitude of the signal by the factor ρ^T/ρ where ρ is the original optimal SNR and ρ^T the target SNR. The signals were re-scaled to the same optimal SNR in all the interferometers as it allows for an easier implementation.

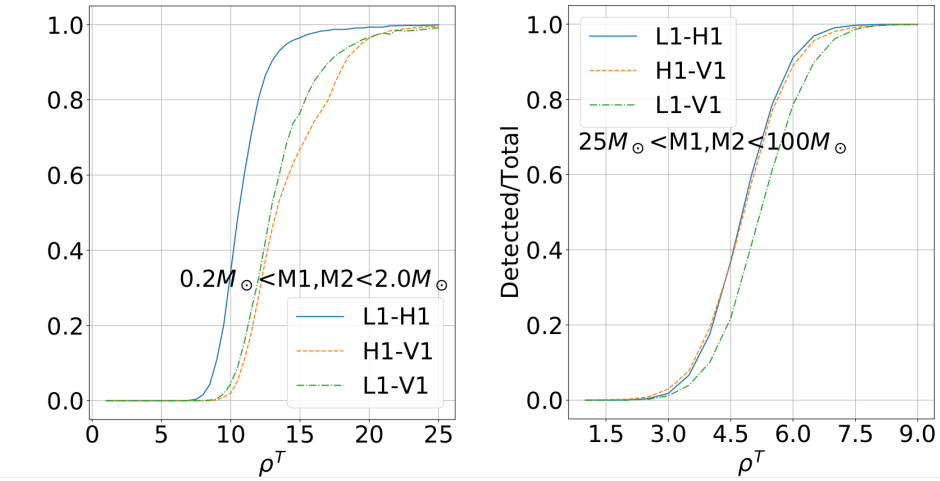

Figure 5.9: Injection test in O2 for both low mass (left) and high mass (right) for the three combinations. The x-axis is the SNR in one of the interferometers and not the network SNR, which will be a factor $\sqrt{2}$ bigger given the re-scaling. Taken from Ref. [170]

The results can be seen in Fig. 5.9 for both mass ranges. As expected, the efficiency increases rapidly with ρ^T reaching higher values at higher optimal SNR. We obtain that the highest efficiency is achieved using H1-L1, which is the most sensitive combination. The value of ρ^T for several efficiency values can be seen in Table 5.4.

double interferometer channel			
	ρ^T (50%) low/high mass	ρ^T (80%) low/high mass	ρ^T (99%) low/high mass
L1 – H1	11.0/5.0	12.0/6.0	19.0/7.0
L1 – V1	13.0/5.5	15.5/6.5	24.5/8.0
H1 – V1	13.5/5.0	17.5/6.0	23.0/7.5

Table 5.4: Values of ρ^T for different efficiencies. The high mass network becomes fully efficient at SNR higher than 8. In the case of the low mass, a SNR larger than 24 is needed. Taken from Ref. [170]

5.3.4 Search in O2 data

The neural networks described in the last sections were used to search for GW in O2 data. The O2 science period was sliced into images of 5s with an overlap of 2.5s between consecutive images. The overlap is motivated by the fact that the chirp of a GW could be cut between two consecutive images, making the detection more difficult. With this method, the chirp will always be fully contained in at least one of the images.

The results of the search can be found in Table 5.5. The L1-H1 CNNs are able to filter 99.9% of the background with a detection rate of 47 and 34 triggers per day for the low and high mass respectively. This rate is bigger than the desired by our choice of threshold but plausible due to glitches and periods of bad data in long stretches of time. In the case of H1-V1, only 0.07% were detected as triggers, resulting in a rate of approximately 25 events per day. L1-V1 high mass has shown a similar performance as H1-V1 whereas the low mass case performance has shown to be much worse, with a trigger rate of 180 per day.

CNNs Response to full O2 scan					
		low mass NN		high mass NN	
CNN	Images	Detected	Events/day	Detected	Events/day
L1-H1	4077233	5496	47	3973	34
H1-V1	584993	439	26	414	24
L1 -V1	601877	3078	178	445	26

Table 5.5: Results of the low and high mass searches over O2. Taken from Ref. [170].

The triggers found by the low mass and high mass CNNs were correlated with the results of the O2 single triggers from the pyCBC pipeline [92]. Almost all the images flagged by the neural networks were also detected by this pipeline with SNR between six and eight. This indicates a performance similar to the first step, in which single triggers are listed, of the dedicated pipelines. From this list, dedicated pipelines filter real triggers by imposing more conditions like coincidence between interferometers, detection by the same template in all the detectors and phase/time correlations.

The method has also proven capable of detecting the O2 (+O1) catalog events [9], as indicated in Table 5.6, for a total of 7 from 8 O2 events and 1 of 3 O1 events. In particular, this includes the NS-NS event, GW170817 [182], which was identified by the low mass neural network. Furthermore, all the non detected events have masses outside the training range.

CNNs response to O1+O2 catalog				
	low mass		high mass	
Event	CNN value	Detected (Y/N)	CNN value	Detected (Y/N)
GW170104	0.001	N	1.0	Y
GW170608	0.02	N	0.008	N
GW170729	0.1	N	1.0	Y
GW170809	0.15	N	1.0	Y
GW170814	0.01	N	1.0	Y
GW170817	1.0	Y	0.04	N
GW170818	0.003	N	1.0	Y
GW170823	0.05	N	1.0	Y
GW150914 (O1)	0.24	N	1.0	Y
GW151012 (O1)	0.06	N	0.95	N
GW151226 (O1)	0.29	N	0.08	N

Table 5.6: Summary of the CNNs response to the O1+O2 catalog events. A total of 9 from 11 O1+O2 events were detected. Taken from Ref. [170]

This method has demonstrated to be able to detect GWs from CBC events and filter almost all the noise from the data. However, to reduce the number of false positives, a more refined technique was needed to classify as detection. This improvement was done in O3 by using the FAR to further filter the images, as we will see in the following sections.

5.4 Third observation run

In this section we will discuss the results of the search implemented using neural networks in O3 data. As already discussed, the O3 analysis implements improvements with respect to the O2 search.

5.4.1 Data preparation

The data is processed almost in the same way as in Sec. 5.3.1 but with some differences. Firstly, the lower frequency for the low mass signals is changed to 45Hz to better match the values used in the subsolar mass searches [87, 88, 89]. Secondly, the upper mass limit in the low mass case is increased to $5M_\odot$. Finally, a new training data set with masses between $[5M_\odot,25M_\odot]$ is created so that the full mass range can be searched for GW candidates. The parameter space covered can be found in Table 5.7. Very asymmetric cases ($q < 0.05$) are excluded from the training and testing sets.

	Low mass	In-between	High mass
M_1, M_2 [$M_\odot$]	[0.2,5]	[5,25]	[25,100]
Dl [Mpc]	[1,100]	[1,1000]	[100,1400]
RA [Rad]	[0, 2π]	[0, 2π]	[0, 2π]
Dec [Rad]	[0, π]	[0, π]	[0, π]
Pol [Rad]	[0, π]	[0, π]	[0, π]
Inc [Rad]	[0, $\frac{\pi}{2}$]	[0, $\frac{\pi}{2}$]	[0, $\frac{\pi}{2}$]

Table 5.7: Limits of each parameter for the signal generation in O3. Masses are in the detector frame.

5.4.2 Testing the performance

The training is done in the same way as in previous sections. However, given the amount of data with three interferometers online and observing, in O3 we can explore the performance when using as input the three interferometers: H1-L1-V1. For the training, only O3a data was used but avoiding times with the GWTC-2 [10] catalog events. It was checked that the differences in sensitivity between O3a and O3b were negligible.

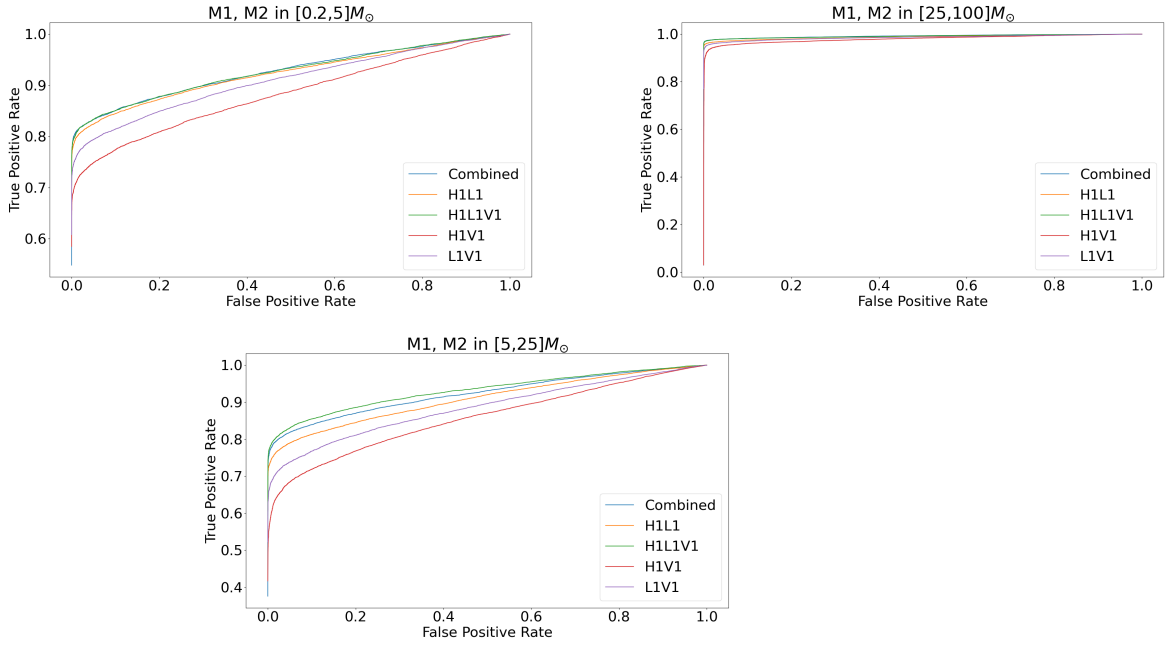

Figure 5.10: ROC curves for CNNs over O3 data using all the possible combinations between interferometers. (top left) shows the ROC curve for the low mass neural network, (top right) for the high mass network and (bottom) for the in-between neural network. At low false positive rate (<0.1) the highest true positive rate is achieved by the high mass neural network.

The results of the training and testing can be seen in Fig. 5.10. The best performance is achieved with the H1-L1-V1 combination. This was expected given the fact that more information was available to discriminate between background and signal. To further reduce the number of false positives detected, the outputs of the separate networks are combined into a single discriminant. A combination using the arithmetic mean, as in Eq. 5.12, was selected. The performance using different algorithms was studied but similar results were achieved. The combined discriminant is shown in the blue line "Combined" in Fig. 5.10.

$$X^{Comb} = \frac{1}{4}(X^{H1L1V1} + X^{H1L1} + X^{H1V1} + X^{L1V1}). \tag{5.12}$$

The combination suppresses the triggers produced by glitches affecting the interferometers. This effect can be seen in Fig. 5.11. The significance of

background triggers (left) at high discriminant is reduced due to the combined discriminant suppressing triggers not seen in all the networks. In the case of the signal distribution (right), there is a slight suppression of low SNR signals that are not fully detected by the network. This effect leads to a small decrease in the true positive rate, as seen in Fig. 5.10.

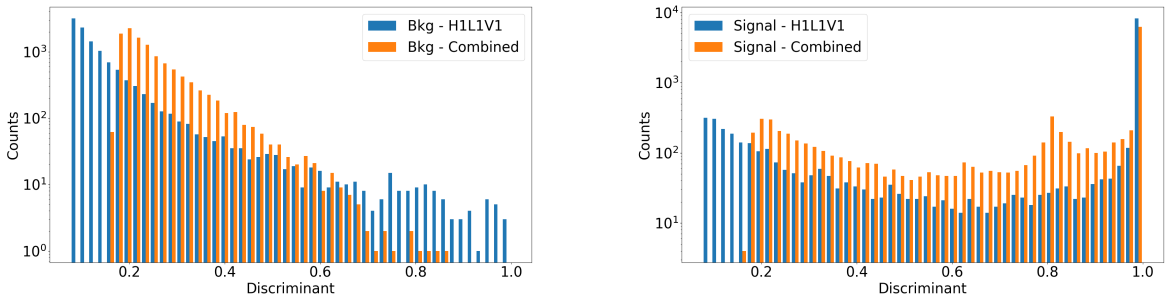

Figure 5.11: Comparison between the combined discriminant of all the neural networks and the one provided by the H1-L1-V1 neural network. (left) shows the background distribution and the suppression at high values of the discriminant. (right) the signal distribution is presented.

5.4.3 Estimation of the False Alarm Rate (FAR)

In this section, the estimation of the FAR used to assign a significance to the detected triggers will be discussed. This significance will be used to declare detection by selecting a FAR threshold. The relation between the FAR and our combined discriminant (Eq. 5.12) is empirically estimated using time-slides as explained in Sec. 3.3.5. The observation time reached by doing these slides is $\approx$ 150yr, which is enough to claim an early detection within the LVK collaboration. This analysis was carried out using resources provided by the Barcelona Supercomputing Center (BSC).

The FAR was estimated separately for O3a and O3b. A comparison between the two can be seen for the high mass case in Fig. 5.12. The distributions are similar up to higher values of the discriminant (lower values of FAR) where small differences are visible. The origin of this difference seems to be the increase in glitch rate between O3a and O3b, as indicated in Fig. 4 from the

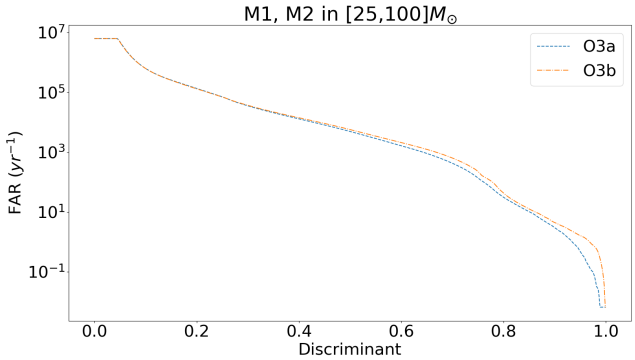

Figure 5.12: FAR distribution for the high mass case in both O3a and O3b.

GWTC-3 catalog in Ref. [66]. This interpretation is further motivated by the fact that the low mass neural network has better performance, as seen in Fig. 5.13 (left), for the O3b case. This neural network is able to better filter glitches due to their shape being very different from low mass signals.

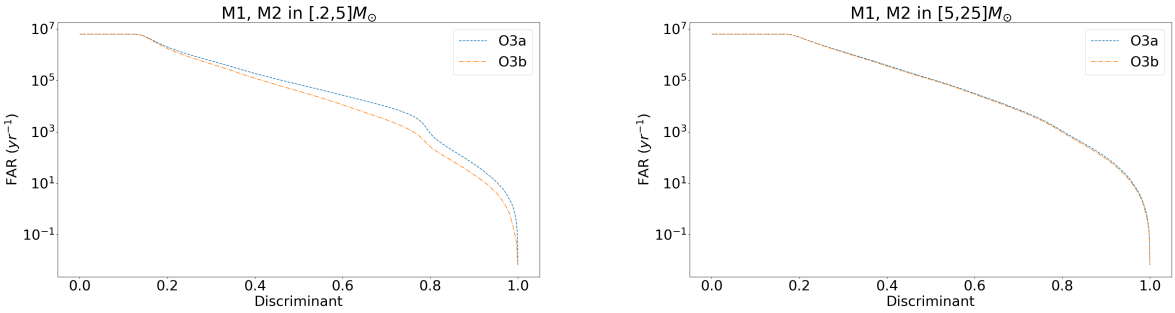

Figure 5.13: FAR distribution for the low mass (left) neural network and the in-between neural network (right) for O3a and O3b.

The effect of the noise suppression provided by the combined discriminant can be seen in Fig. 5.14. The combination allows us to reach lower values in FAR, leading to a higher significance of the signal events. Therefore, the combined output will be used to assign the significance of the analyzed images.

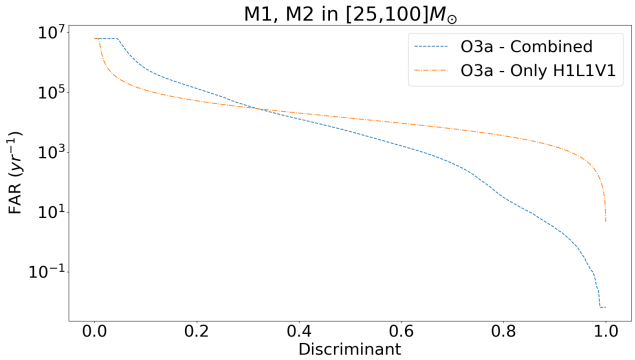

Figure 5.14: Effects on the FAR distribution using a combined discriminant or the value from H1L1V1 only. At 0.9 discriminant, the FAR obtained from the combination is $\approx 3yr^{-1}$ while the estimation using only H1L1V1 is $> 1500yr^{-1}$
.

5.4.4 Injection test

Following the method explained for O2, an injection set is generated to understand the performance of the neural networks. Some modifications were implemented with respect to the O2 analysis:

1. We use the same distribution for the parameters as the training set but instead of sampling uniform in distance (r) we sample uniform in comoving volume (r^3). This is in better agreement with the distribution of galaxies in the cosmos.

2. The simulated signals are not re-scaled to a target optimal SNR, we allow the values to vary freely so that the parameter space is fully sampled. This is different from the approach taken in O2 where the signals were scaled to a target SNR.

3. Injections are declared as detected based on a threshold in the assigned significance. The significance is estimated from the one-to-one map between the FAR and combined discriminant estimated in Sec. 5.4.3. Unless stated otherwise, 1 yr^{-1}, or 1 event per year, will be used as a threshold.

The efficiency is measured as a function of the source frame chirp mass and the network SNR of the injections. The masses are converted into source

frame by assuming Planck15 [159] cosmology from the astropy [183] python package. This is done by estimating the redshift from the luminosity distance and using it to transform the masses to source frame: $m_{source} = m_{detector}/(1+z)$. In total, 32000 injections per neural network were performed in O3a data.

The effect the FAR threshold has on the estimated efficiencies can be seen in Fig. 5.15. This figure shows the efficiency as a function of the chirp mass (left) and network SNR (right) for different thresholds: 0.01 yr^{-1}, 1 yr^{-1} and 10 yr^{-1}. Lower FAR values correspond to a more stringent condition in the significance of the image. This results in a reduced efficiency for a given chirp mass/network SNR bin. At $50M_{\odot}$, the efficiency decreases from 85% with a FAR threshold of 10 yr^{-1} to 79% when the threshold is lowered to 0.01 yr^{-1}. In the case of the network SNR, the efficiency at 12 network SNR is 71% with a FAR threshold of 10 yr^{-1} and decreases to 47% at a threshold of 0.01 yr^{-1}.

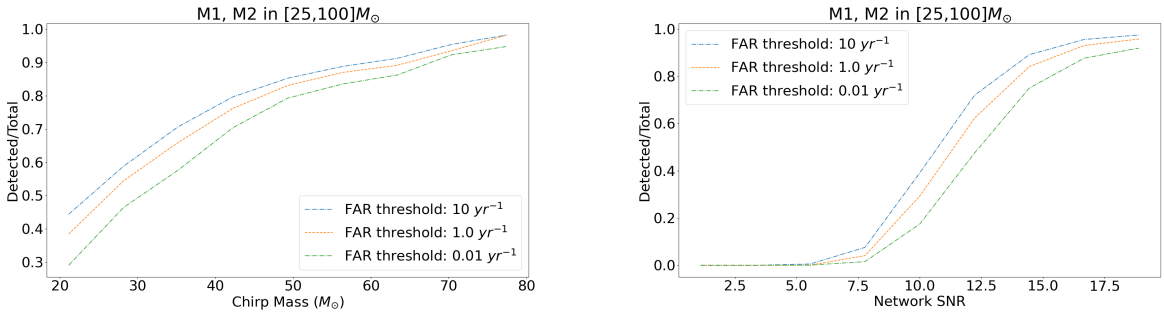

Figure 5.15: Efficiency as a function of the chirp mass (left) and network SNR (right) for high mass injections using three different FAR thresholds.

The efficiency of all our neural networks as a function of the network SNR can be seen in Fig. 5.16 for a FAR threshold of 1 yr^{-1}. The neural network dedicated to high masses peaks in efficiency much faster than the other two, reaching 90% at network SNR of ≈ 15. The low mass neural network reaches this point at a network SNR of ≈ 30.

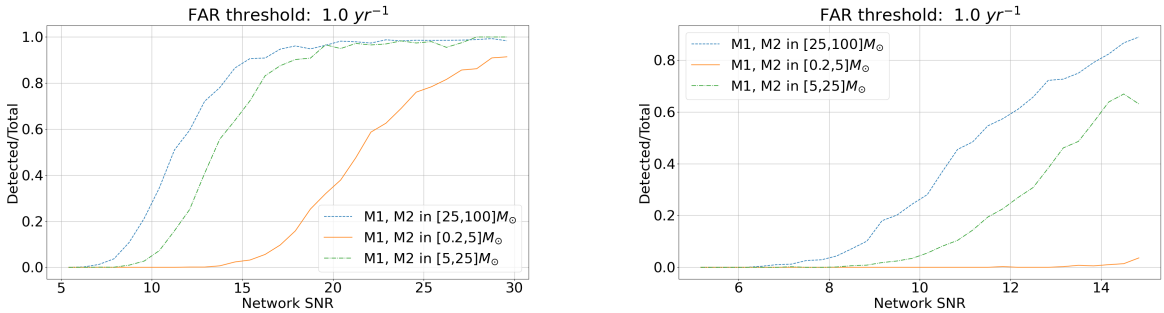

Figure 5.16: (left) Efficiency as a function of the network SNR for all the combinations for a FAR threshold of $1yr^{-1}$. (right) Zoom over a region of interest for the O3 catalog. The high mass neural network reaches an efficiency larger than 90% at network SNR of ≈ 15. The In-between neural network reaches this point at ≈ 20 while the low mass network is almost fully efficient at network SNR of ≈ 30.

The efficiency as a function of the chirp mass is shown in Fig. 5.17 for a FAR threshold of $1yr^{-1}$. The highest efficiency is achieved at the highest chirp mass values. In the case of the high mass neural network, we reach almost 90% efficiency for signals with chirp mass bigger than 60. The neural network working in the middle mass range is not very efficient in the lowest chirp mass range. This seems to indicate that this particular neural network has learnt to recognize the sharper frequency features of the higher mass signals instead of the longer tail of low mass GWs.

5.4.5 Search in O3 data

The neural networks are used to search the whole O3 (O3a+O3b) period. An overlap of 2.5s between consecutive images, as in Sec. 5.3.4 for O2, was used. The significance of each image is assigned using the FAR distribution estimated in Sec. 5.4.3. The estimated FAR is transformed into IFAR (Inverse False Alarm Rate) and compared with the number of events expected for a given IFAR assuming only background. The expected number of events is obtained from Poisson statistics. Finally, candidates are identified from excesses in the background distribution.

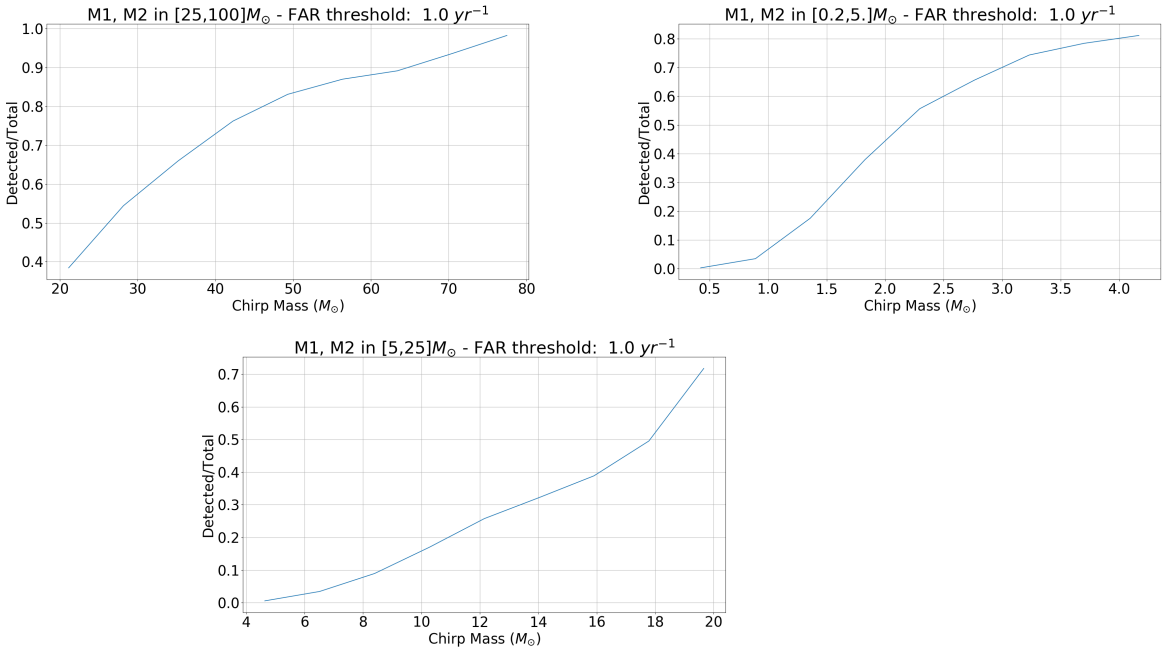

Figure 5.17: Efficiency as a function of the chirp mass for the three neural networks. The highest efficiency is reached at the upper bound of the training set for all the networks. (top left) shows the results for the high mass neural network, (top right) for the low mass neural network and (bottom) for the in-between neural network.

The results of the search can be found in Fig. 5.18. The distribution of the low mass and in-between masses neural network is compatible with the background distribution and, therefore, nothing has been detected in the scan. However, an excess can be seen in the distribution of the high mass neural network, both for O3a (top left) and O3b (top right), that indicates triggers detected with high significance.

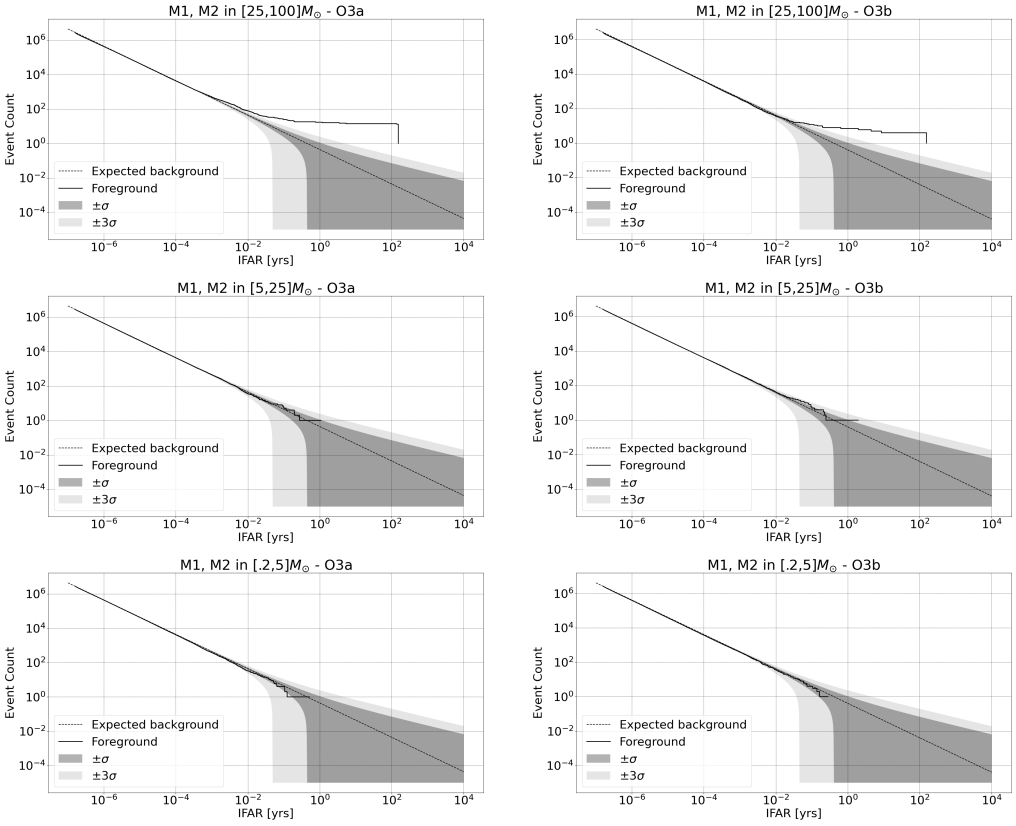

Figure 5.18: IFAR distribution compared to expected by pure background (dashed line). In the left column are the results for O3a and on the right for O3b. The excess seen at higher values of IFAR in high mass NN (top left and top right) indicate the presence of confident triggers. In the case of the low mass (mid left and mid right) and in-between neural networks (bottom left and bottom right), no confident trigger was found.

The main suspects for this excess are the events detected in the O3 catalog [10, 11, 66]. This interpretation is further motivated by the fact that once they are removed from the analysis the excess is no longer present, as seen in Fig. 5.19. Given that the analysis is limited to H1-L1-V1 times, 51 events are available from the full catalog: 30 from O3a and 21 from O3b. This result proves that the method is able to detect real events by searching the data.

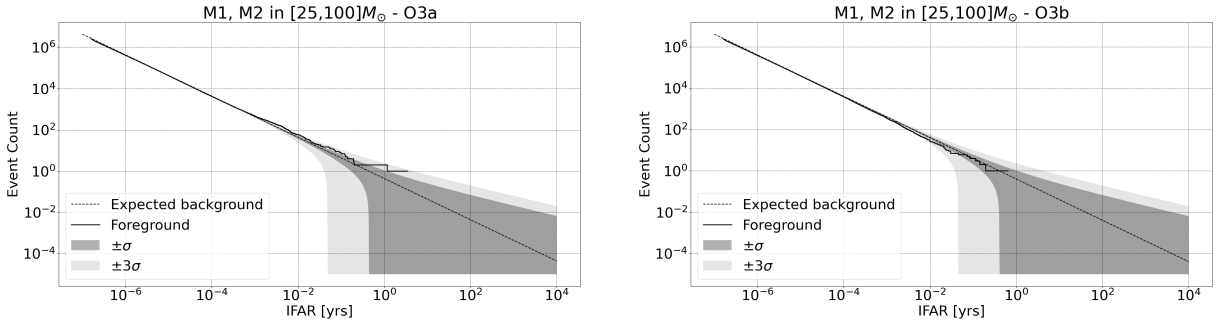

Figure 5.19: IFAR distribution compared to expected by pure background (dashed line) in which the catalog events have been removed. The O3a and O3b results are shown in the (left) and (right) figures respectively.

The results for all the catalog events can be seen in Appendix D, where we show the value of the discriminant and the associated FAR. A plot showing the FAR distribution vs the discriminant of the high mass neural network is shown in Fig. 5.20. This figure shows that all the detected events are concentrated at high discriminant (>0.95) which corresponds with the lower tail of the FAR distribution seen in Fig. 5.12.

The high mass neural network has detected 17 of the 51 events available, giving a detection efficiency of $\approx 33\%$ with a FAR threshold of 1 per year. As indicated in Fig. 5.15, the detection efficiency depends on the selected FAR threshold. Different threshold values have shown that the efficiency can be increased up to 50% with a reasonable admittance of false positives. In particular, a FAR threshold of 1 event per month provides an efficiency of $\approx 33\%$ (19 from 51 events) while a threshold of 1 per week has an efficiency of $\approx 50\%$ (26 from 51).

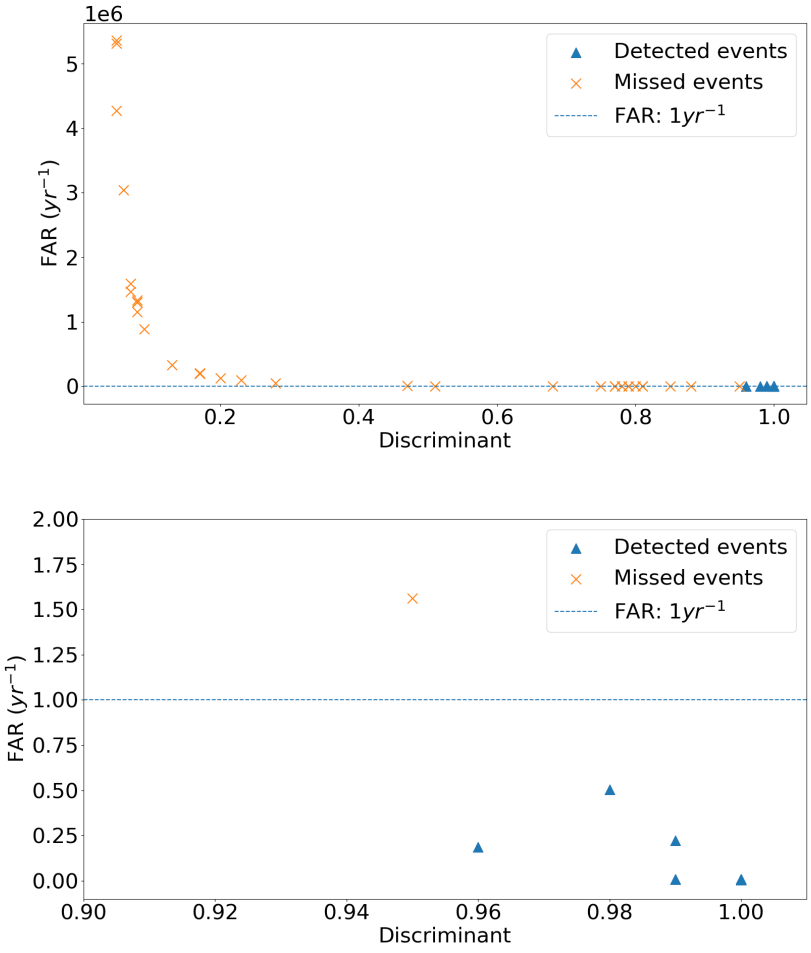

Figure 5.20: Catalog events distributed by their matched filter network SNR and chirp mass in the source frame. (top) shows all the catalog events. (bottom) zoom at high values of the discriminant (>0.9).

The previous estimation includes events outside the training range. Limiting the analysis to events within the training range, 16 of 34 events are detected, resulting in a detection efficiency of $\approx 47\%$. This value matches the expected outcome given the injection studies in Sec. 5.4.4 for signals with network signal-to-noise ratio between 10-15. As shown in Fig. 5.21 this is the region where the catalog events are mostly concentrated. Loosening the FAR threshold up to 1 per week, provides a detection efficiency of $\approx 69\%$ in the training range (24 of 35).

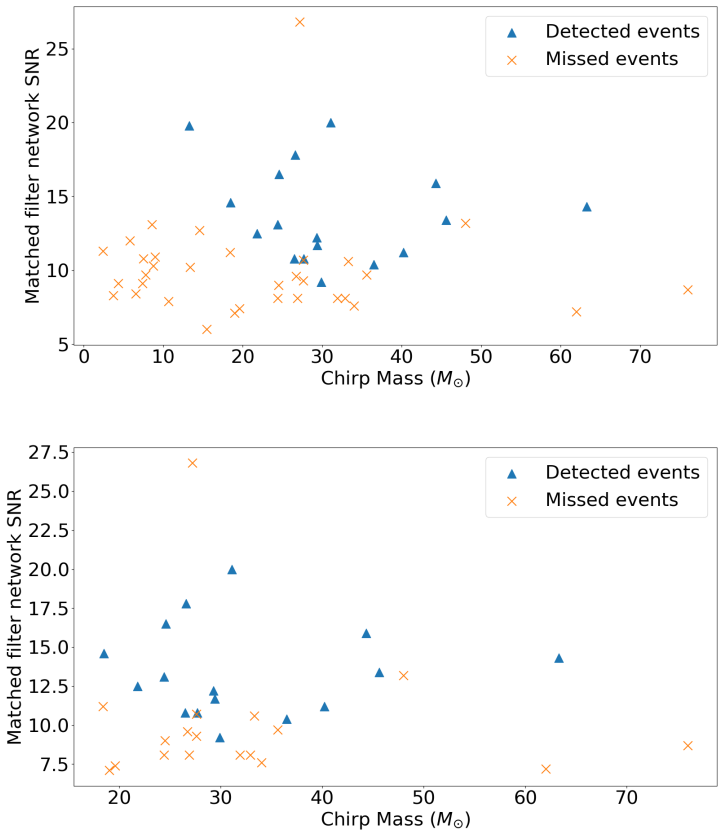

Figure 5.21: Catalog events distributed by their matched filter network SNR and chirp mass in the source frame. (top) shows all the catalog events. (bottom) only catalog events with source frame chirp mass bigger than $17M_\odot$.

The analysis indicates that the injection tests are accurate enough to understand the performance of the neural networks and predict their performance over catalog events. Furthermore, this new methodology allows us to better control the number of false positives by means of the FAR threshold.

5.5 Outlook

We have seen the application of a machine learning technique for the detection of CBC events in real data. This method has been tested in both O2 and O3 data and was able to detect real events and filter out most of the background.

In O3, the false positives were largely reduced by implementing the FAR, allowing a better control over the detection efficiency and false positive rate by means of a FAR threshold. Nevertheless, the implementation of an additional method dedicated to filter the background from glitches and non-Gaussian noises would further improve the performance of the network. One possibility would be a neural network dedicated to the search for glitches. More information can be found in Refs. [184, 185, 186]. Furthermore, the extra information would perhaps allow us to decouple the discriminant and run over double interferometer coincidence times, which has a larger duty cycle and reduces the number of missed real events.

Related to the training, from the injection tests seen in Fig. 5.17, we can also conclude that in some cases the training was too broad in scope, making the neural network unable to learn the important details. For future investigations, dividing into more regions, e.g. chirp mass of the source between [2.5, 4]$M_{\odot}$and [16,20]$M_{\odot}$, seems to be the course of action to maximize the efficiency of our neural networks.

Finally, it was found that the performance seemed to remain stable during the whole O3 (see Fig. 5.12). This coincides with our expectations as there was no significant change in sensitivity between O3a and O3b. Nevertheless, the possibility of retraining the neural network to account for important changes in sensitivity is possible and can be done on a short time scale.

Conclusions

In this thesis I show the results from a search of subsolar mass compact objects using a matched filter pipeline and a general search of CBC events using a machine learning implementation.

I present the design and installation of an instrumented baffle and the noise hunting activities that were carried out to improve the sensitivity of the instrument. The instrumented baffle installed in the IMC was able to detect scattered light within the cavity. These results will be used to design the instrumented baffles that will be installed surrounding the test masses for O5. The baffles will be able to monitor the scattered light within the main arms and facilitate the pre-alignment of the cavity.

The subsolar mass search was able to set upper limits on the rate density of subsolar mass events. The estimated range is between $[220-10000]\ \mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}$ depending on the chirp mass of the binary. The obtained values of the rate were used to constrain models of PBH binary formation and dark matter black holes. The dark matter black hole model was used to constrain the fraction of dark matter collapsing into black holes and the minimum mass of the aforementioned objects. The strictest limit was derived at $1\mathrm{M}_\odot$ for $f_{\mathrm{DBH}} < 0.0012 - 0.0014\%$. For PBHs, two different models for early binary formation were considered: a 3-body system that doesn't include suppression factors in O3a and a N-body system with suppression factors in O3b. Each model was used to constrain the fraction of dark matter collapsing into PBHs. The O3b results have found that, considering the N-body system model, $f_{\mathrm{PBH}} < 0.5$ at $0.3\mathrm{M}_\odot$ and is below 0.08 at $1\mathrm{M}_\odot$. These results strongly disfavor the f_{PBH}=1 scenario up to $1\mathrm{M}_\odot$.

The implemented machine learning methods were used to search for GW from CBC events. A wide mass range was covered but very asymmetric events were not considered. The search was first tested on O2 data and was able to detect catalog events, for a total of 7 from 8 O2 events and 1 of 3 O1 events. The NS-NS event GW170817 was detected by the low mass neural network. This technique was implemented with some improvements in O3. These improvements include the addition of FAR estimation to assign significance to each trigger. This allowed better control over the false positives detected by the neural network. Furthermore, an injection campaign was performed to understand the efficiency of the method. It was found that a detection efficiency of $\approx 69\%$ was obtained with a FAR threshold up to 1 per week.

Finally, the contribution to the subsolar mass searches was focused on the search using the PyCBC pipeline and the estimation of the sensitive volume. Future searches will be improved by taking into account the limitations found during this analysis. For example, a variable lower frequency for the matched filtering or correlation studies between the pipelines. Similarly, in the case of machine learning methods, an online implementation is being considered. Furthermore, the estimated efficiency can be used to constrain different GW emission models by performing a similar analysis as the subsolar mass search.

Appendix A

Instrumented baffle for scattered light control

As was discussed in Sec. 2.8, part of the planned upgrades of the AdV interferometer for O5 includes the installation of instrumented baffles around the main test masses. These new baffles will open up the possibility to monitor the scattered light within the cavity, the contamination of the mirror surfaces and facilitate the pre-alignment and fine-tuning of the parameters of the interferometer.

A demonstrator of the technology designed to instrument the baffles around the mirrors was installed in April 2021. The instrumented baffle was suspended surrounding the suspended end mirror of Virgo's IMC cavity [187]. In this Appendix, the first results of the detector are discussed and compared with the scattered light distribution expected from simulations.

A.1 Input Mode Cleaner

A beam propagating inside a cavity is described by a set of orthonormal solutions, known as the Hermite-Gauss modes TEM_{mn}, parameterized by two natural numbers (m, n). In particular, the $(0, 0)$ solution is given by a Gaussian beam (See Fig. A.1). Solutions with higher values of (m, n) are known as "Higher order modes". The interferometers are designed to work with Gaussian beams and, for that reason, they are filtered from higher order modes before being introduced into the cavity.

This filtering is done through a cavity tuned to resonate with the $(0, 0)$ mode, the Input Mode Cleaner (IMC).

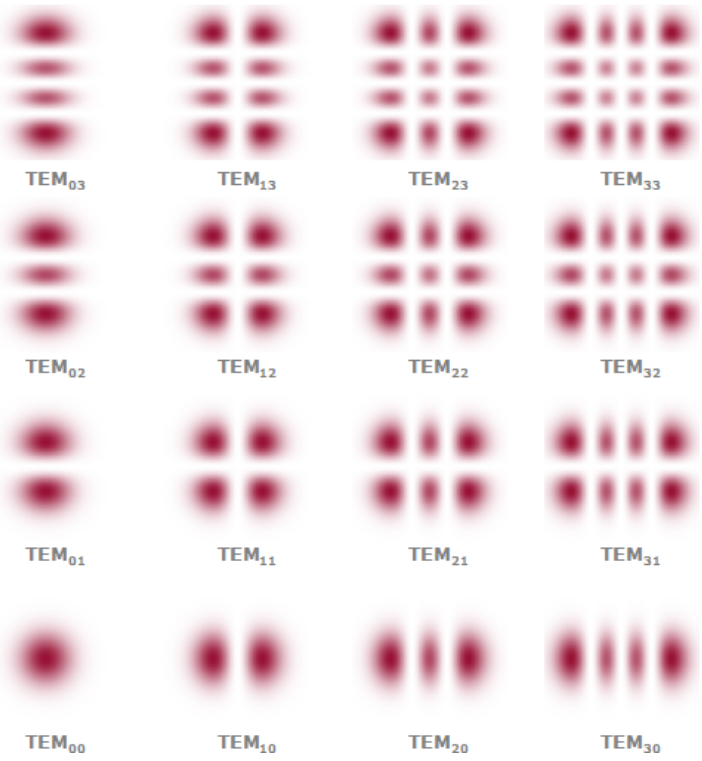

Figure A.1: TEM modes for different combinations of (m,n). Source: `https://www.rp-photonics.com/resonator_modes.html`.

The IMC cavity is a triangular cavity (see Fig. A.2) used to filter the beam before entering the main cavities of the interferometer [28]. This cavity consists of three suspended mirrors: MC1, MC2 and MC3. Both MC1 and MC3 are flat mirrors while MC2, the end mirror, is curved with a radius of curvature of 187m. This cavity has a length of $\sim$ 143 m and a finesse of 1200.

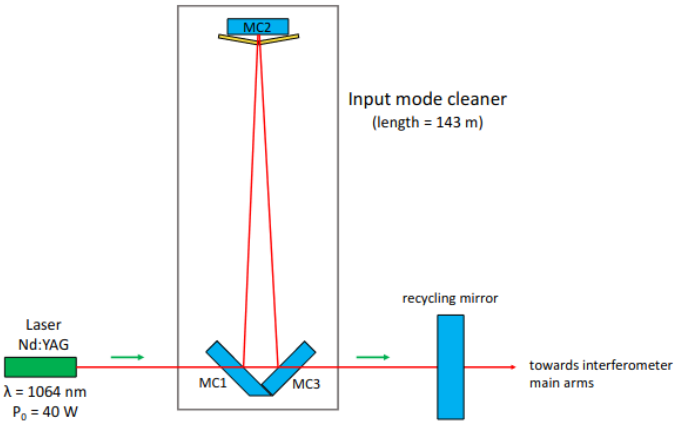

Figure A.2: AdV IMC cavity with half roundtrip length of $\sim 143m$. MC1 and MC3 are two flat mirrors while the end mirror (MC2) is curved. Taken from Ref. [188].

A.2 Instrumented baffle

A demonstrator of the technology designed to instrument the baffles around the mirrors was installed in April 2021 surrounding the end mirror (MC2 from Fig. A.2) of Virgo's IMC. This baffle has been taking data since and is able to detect variations of power inside the IMC cavity. It has an inner radius of 7cm and an outer radius of 17.5cm. The baffle is divided into two halves tilted 9 degrees with respect to the direction of the beam to avoid back reflections in the cavity. A total of 76 Si-based photosensors were installed in 4 mm diameter conical shaped holes centered around the sensors, such that the light does not resolve the sensor edge and the scattering of light with the hole geometry is minimized. Each sensor has a sensitive area of 6.97x6.97mm^2. Both the baffle and sensors surfaces include anti-reflective coating at 1064nm. All the components are certified for ultra high vacuum conditions.

A diagram showing the sensor holes and the baffle can be seen in Fig. A.3. Most of the sensors are located in two concentric rings at radii of 8.8 and 9.8 cm. The signals from the sensors are processed by 16 ADCs (8 per half baffle) averaging to a sampling frequency of 2 Hz with the potential to reach

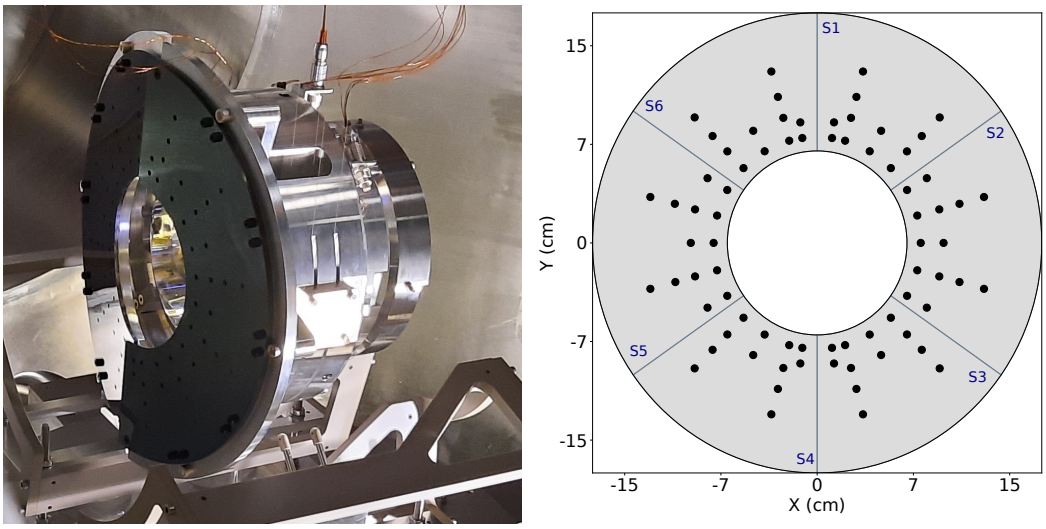

Figure A.3: (left) Instrumented baffle installed in the IMC tower. (Right) Representation of the baffle with the holes in which photosensors are added. The baffle is divided into six regions in such a way that the symmetry along the y-axis is preserved. Taken from Ref. [187].

up to 10 Hz in the future. Each ADC is instrumented with a temperature sensor and the system operates at a voltage of 3.3 V. The IMC cavity was checked after the installation and it was successfully restored.

A.2.1 Calibration

The sensors were calibrated using a dedicated laser setup shown in Fig. A.4. An infrared beam (1064nm) is used as the source of light. The beam was collimated so that its transverse size is limited to a diameter of 1.38mm at a focal distance of about 6mm. The setup was also used to study the performance of the sensors as a function of the temperature in the range between 23 ^{o}C and 40 ^{o}C. No significant dependence was observed. The laser is aligned to the center of each photodiode (PD) by using a red beam as reference. It was found that the response of the sensors does not depend on the exact location as long as the beam is fully within the sensitive region of the PD.

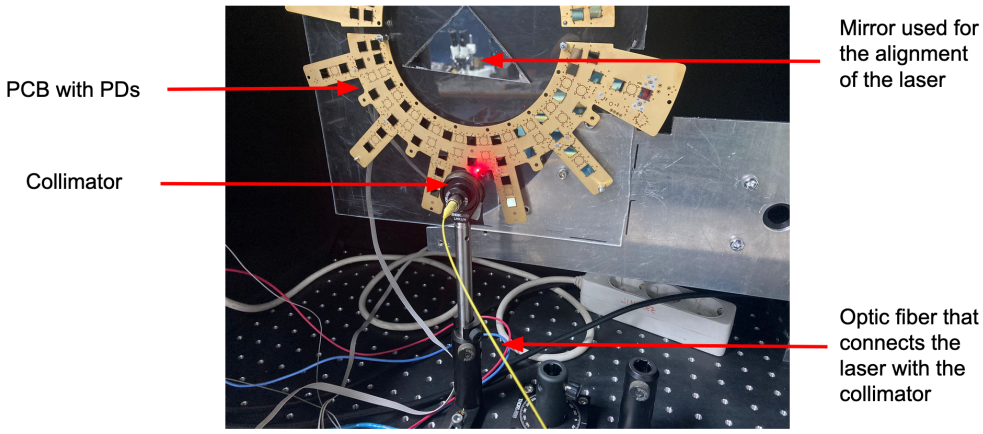

Figure A.4: Optical setup used to estimate the response of the photodiodes. The laser is collimated before being sent to the photodiodes. The beam is aligned to the center of each photodiode by using a red beam as reference.

The response of the photodiodes was studied by measuring the number of counts for different values of nominal laser power, as shown in Fig. A.5. The photodiodes were found to have a linear response at powers lower than 15mW, corresponding to the whole range of interest. These measurements were repeated for all the photodiodes and similar results were obtained. Interpolating the slope of the linear region, we obtain that the calibration factor is 6.2 ± 0.2 $\mu W/(ADC\ counts)$, as shown in Fig. A.6.

Due to the collimator, the power of the beam is attenuated and smaller than the initial value. Therefore, a calibrated photodiode was used to measure the real power illuminating the photodiodes. With this correction it was found that the absolute calibration factor is 4.6 $\mu W/(ADC\ counts)$, with an uncertainty of about 5%.

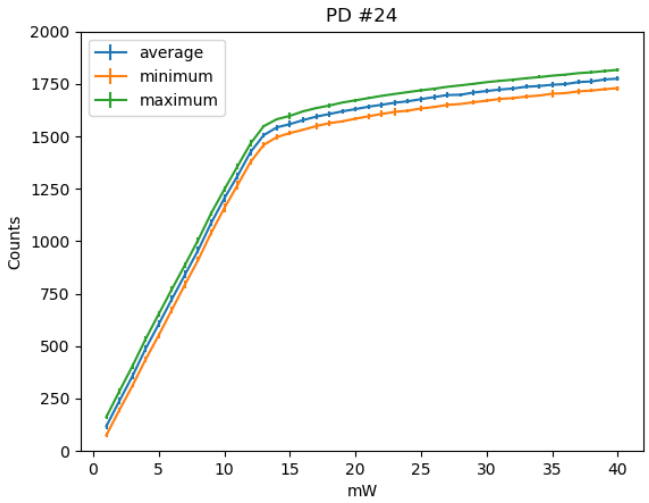

Figure A.5: Response of the photodiode as a function of the power of the laser. The linear region has a slope of $6.2 \mu W/(ADC\ counts)$.

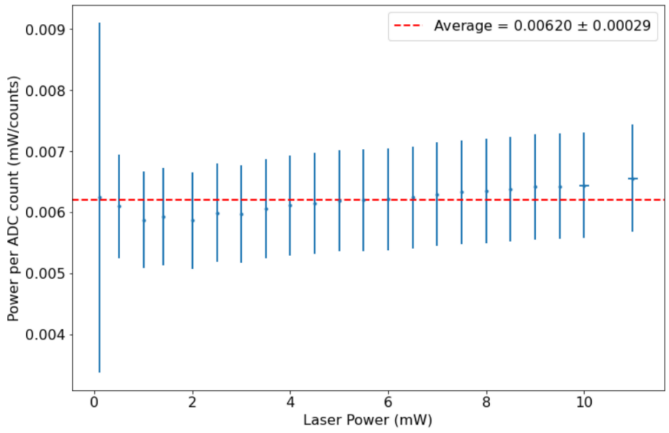

Figure A.6: Estimation of the calibration factor for different values of the (collimated) laser power. A value of

A.2.2 Temperature evolution

As was mentioned earlier, each ADC is instrumented with a temperature sensor to monitor the potential overheating of the front-end electronics. This was found to not be an issue thanks to the moderate operating voltage, efficient heat dissipation of the PCBs and the mechanical design.

The evolution of the temperature can be seen in Fig. A.7. After switching on the baffle, the system reaches thermal equilibrium 40 minutes later with temperatures in the range between 22 and 28^{o}C. The maximum temperature is reached by the ADC closest to the micro-controller electronics. This plateau is well in the safety zone of the photosensors.

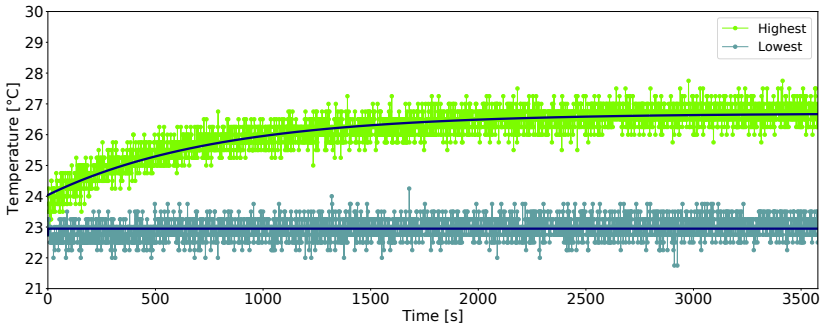

Figure A.7: Temperature evolution for a period of one hour. The highest temperature corresponds to the board from which the data is sent out of the baffle. A plateau is reached after around 40 minutes without leaving the safety threshold. Taken from Ref. [187].

A.2.3 Results

The baffle has been running and taking data periodically since the installation. In the absence of light inside the cavity, it has shown average noise levels limited up to seven counts, with an RMS of 0.01 to 0.16 counts depending on the sensor. In the presence of light, the SNR varies in the range between 1 and more than 10 times. The performance of the baffle was studied for long periods of time and was shown to remain stable.

The raw signals per photosensor averaged over one hour are shown in Fig. A.8 for two separate data sets.

The signals are concentrated at low radius in which sensors reach more than 100 counts. Furthermore, there is a left-right asymmetry with more power at negative x values. In addition, the power is concentrated in the plane tilted by about 15 degrees with respect to the nominal plane of the cavity. These effects remain in all the data. The performance of the baffle was tested by forcing small displacements of the IMC end mirror with respect to its nominal position. These changes produced a variation in the IMC output power and the appearance of higher order modes within the cavity. This translated into a variation of the total power measured by the baffle of about 5%, showing that the baffle is sensitive to the cavity. Furthermore, the baffle was able to detect fast unlocks inside the cavity, showing a sharp drop in counts for the duration of the unlock (at most a few seconds). Before the final calibration, the average noise in the absence of a beam is subtracted channel by channel.

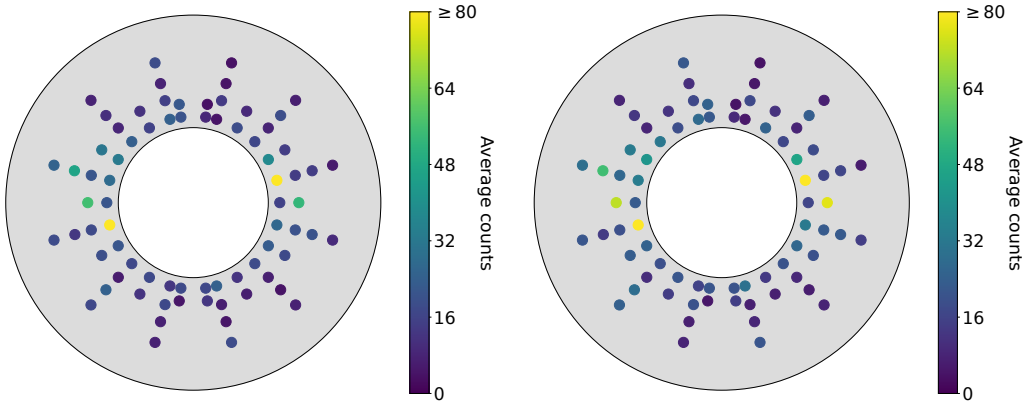

Figure A.8: Counts measured by the baffle when a beam is propagating inside the cavity. The data corresponds to two different periods of time. Taken from Ref. [187]

Comparison with simulations

The calibrated data are compared to simulations of the IMC cavity. These simulations were used to design the baffle and the position of the sensors [188, 189]. Following studies from Ref. [188], an input power of 28.5 ± 0.1W, corresponding to the average input power during the data taking period, was used and updated mirror maps of the IMC mirrors were considered. An additional uncertainty of 0.2% was included in the calibrated data due to oscillations of the input power. The simulation does not include thermal effects in the mirrors induced by the laser, that are expected to be small.

A similar representation to Fig. A.8 but using the simulated values can be seen in Fig. A.9, showing a general agreement with the observed data.

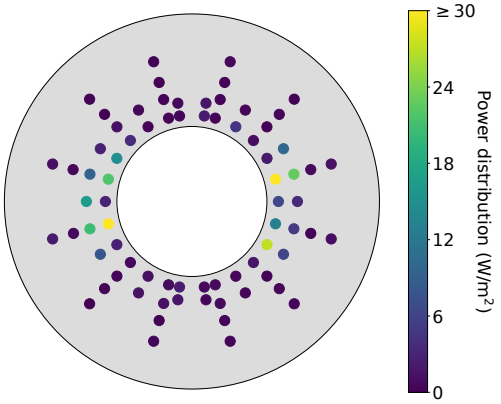

Figure A.9: Power per surface area (W/m^2) estimated by the simulation in the positions where the sensors are located. Taken from Ref. [187].

In Fig. A.10 a comparison between the measured distribution and simulations as a function of the radius, r, and the angle, ϕ, can be seen. In the simulations, the value is estimated at the center of each sensor and only the active area of the sensor is taken into account ($\phi \times 4mm^2$). The observed power varies between 1.1 W/m^2 and 52.3 W/m^2 at small r and 0.4 W/m^2 and 9.6 W/m^2 at very large r, showing a strong dependence on r.

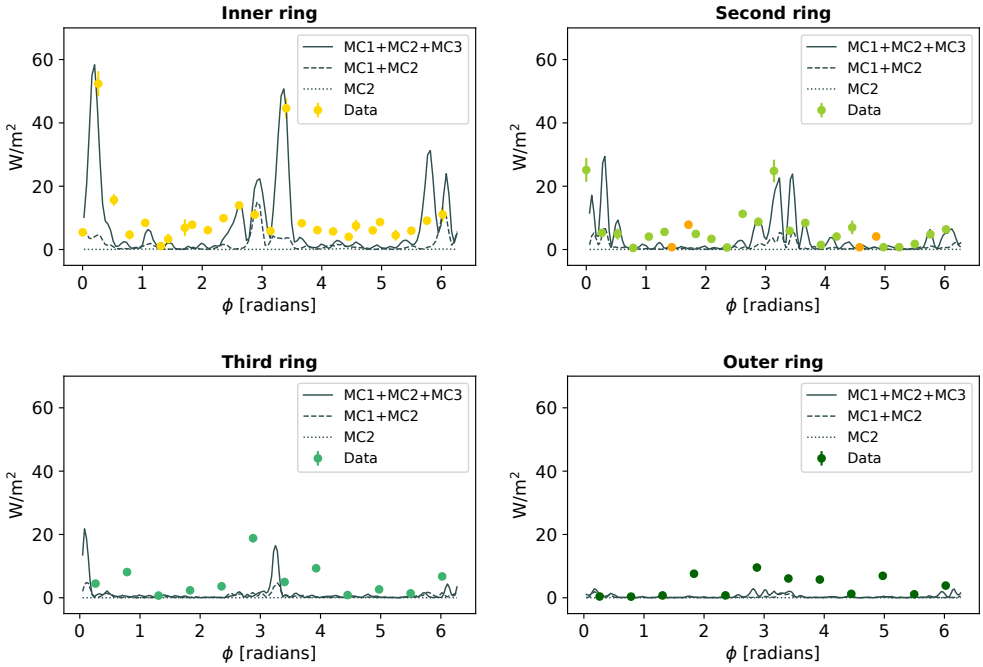

Figure A.10: Measured power (dots) as a function of the angle (ϕ) for the four rings vs the expected result from the simulations. It should be noted that there are four sensors with different radius. These are included in the second ring. The simulations were repeated for several cases: including the three mirrors (MC1+MC2+MC3) in the IMC cavity, removing the dihedron right mirror (MC1+MC2) or only including the end mirror (MC2) surface maps. Taken from Ref. [187].

A more in-depth study shows that systematically the power reaching the baffle is underestimated in simulations. This is shown in Table A.1, where the results are integrated over the sensors in each sextant. The power given by the simulations is normalized by the circulating power within the cavity (9695 W). Experimental data shows a total power in the sensors of about 6.6 mW while the simulation predicts around 4.4mW. Furthermore, the simulation does not completely support the left-right asymmetry observed.

Further studies have shown that the measured tilt in ϕ cannot be attributed to misalignment in the cavity. We conclude that the most probable origin of the disagreement is due to a description non accurate enough of the mirror surface maps. At a large radius, the predictions might also be affected by large angle scattering not included in the simulations. These topics will be studies using the data measured by the instrumented baffle.

Sextant	Data		Simulation
	Power (mW)	Ratio ($\times 10^{-5}$)	Ratio ($\times 10^{-5}$)
S1	0.47 ± 0.02	4.8 ± 0.2	1.0
S2	1.92 ± 0.10	19.8 ± 1.0	19.6
S3	0.55 ± 0.03	5.7 ± 0.3	1.6
S4	0.67 ± 0.03	6.9 ± 0.4	1.1
S5	2.25 ± 0.11	23.2 ± 1.2	20.6
S6	0.76 ± 0.04	7.8 ± 0.4	1.5

Table A.1: Measured power distribution in the baffle sextants (See right Fig. A.3) compared to simulations. The column "Ratio" shows the total power normalized by the circulating power. Taken from Ref. [187]

In summary, the baffle has proven itself able to directly detect scattered light inside the cavity. These results will serve to calibrate the simulations and demonstrate already the potential of instrumented baffles to detect defects in the mirrors and to improve the understanding of the scattered light inside ground-based gravitational wave experiments like Virgo.

Appendix B

Statistical model

In this appendix we explain the basic elements needed to understand the concepts of Bayesian inference used in this work [190, 191]. The main advantage of Bayesian inference over a frequentist approach is that it allows to update the model with new information. In broad strokes, a Bayesian inference process can be divided in three steps.

1. Define a model, consistent with the problem, that describes the full probability distribution of all the parameters involved.

2. Update the model using Bayes' rule and a given sample of data. This is known as the posterior distribution.

3. Evaluate the performance of the updated model and how the assumptions used fit the data.

The main focus of this appendix will be in the first two steps. In first place, we will go through an explanation to Bayes' rule and the terms involved. We will follow with an introduction to credible intervals and nuisance parameters and end with commonly used sampling methods.

B.1 Bayes' rule

The main objective of Bayesian analysis is to provide statements in terms of probability for the unknowns involved in a problem, be it a set of unknown parameters θ or unobserved data $\hat{d}$. These statements are made by conditioning our proposed model with real data, d.

Let's assume that we have a problem defined by some parameters θ and a joint probability distribution given by $p(\theta, d)$. Then, by using the law of total probability, we have that the joint distribution can be written as in Eq. B.1:

$$p(\theta, d) = p(d|\theta)p(\theta) = p(\theta|d)p(d), \tag{B.1}$$

from where we can extract the Bayes' rule by using the second equality, as in Eq. B.2.

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}. \tag{B.2}$$

Notice that the terms $p(\theta|d)$ and $p(\theta)$ are well defined only from a Bayesian perspective. This is due to θ not being understood only as a parameter of a distribution but also as a random variable.

The term $p(\theta|d)$ is known as the posterior probability and it should be understood as the probability of θ given the realization of the data, d. The terms on the right side of the Eq. are defined as follows:

- $p(\theta)$: This term encapsulates the information we have of our parameters before doing the experiment. For this reason it's usually referred as "prior distribution".
- $p(d|\theta)$: The probability of obtaining a particular realization of the data given a set of parameters θ. It should be noted that this is a function only of the parameters and not of the observed data.
- $p(d)$: A normalization constant for the posterior. It corresponds to the marginal distribution over θ of our joint probability model. There are several ways to refer to this term but the most common in this field is "evidence".

It should be noted that the observed data, d, affects the posterior only from the term $p(d|\theta)$ when seen as a function of θ. For this reason this term is usually called "Likelihood".

Given two values of the parameters θ_1 and θ_2, we can estimate which set of parameters is favored by the observed data by dividing the posterior distributions as in Eq. B.3:

$$\frac{p(\theta_1|d)}{p(\theta_2|d)} = \frac{p(\theta_1)}{p(\theta_2)}\frac{p(d|\theta_1)}{p(d|\theta_2)}. \tag{B.3}$$

This is known as the Bayes factor and is generally used to compare two models. In the case of using the same prior distribution the first term on the right side cancels out and this factor is given by the likelihood ratio between the two models.

B.2 Priors

We have seen that the estimation is heavily dependant on the prior choice. At first approximation we could try to use a prior that encapsulates all our knowledge about the parameters. This is however unfeasible when the problem is too complex or there are too many parameters, as the prior might become unwieldy.

A possible solution would be to use a prior that is as non informative as possible, for example a flat prior, $p(\theta) \propto constant$. This is one of the most used alternatives but it also has its own limitations:

- A flat prior is not always normalizable. For example a flat prior over the whole real number will integrate to infinity. However, this is not a problem if the posterior distribution can be normalized. These priors are known as "improper priors".
- They are not necessarily invariant under transformations of the variable, that is to say, a flat prior over a parameter doesn't mean that it will be flat over transformations of that parameter.

There has been several attempts to generalize the concept of noninformative priors and one of the most important is known as "Jeffreys prior" [192].

This prior is invariant under transformations of variable and is given as in Eq. B.4. In some cases it might be an improper prior:

$$p(\theta) \propto \sqrt{det(I(\theta))} \tag{B.4}$$
$$I(\theta)_{ij} = E\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} log(p(d|\theta))\Big|\theta\right],$$

where $I(\theta)$ is the Fisher information matrix. However, this choice of prior does not mean that the functional form is the same. In this sense being transformation invariant means that it assigns the same probability volume independently of the parametrization used.

Another convenient choice of priors are known as the conjugate priors for the likelihood $p(d|\theta)$. These are defined as the priors that produce a posterior in the same probability distribution family. For example, if we use a likelihood model defined by a Poisson distribution, a prior following a gamma distribution will result in a posterior also described by a gamma. These priors are convenient in the case that a closed form for the posterior is needed and simplify the process of normalization and sampling from it.

B.3 Nuisance parameters

We have seen that the first step for Bayesian inference is to define the model, $p(\theta, d)$, for a set of parameters θ that describe all the unknowns of the system (experiment). However, there might be situation in which some parameters have to be accounted to define the statistical model but are not of interest for the analysis. These are known as nuisance parameters. For example, assuming a Gaussian model with unknown variance, any model testing the mean of the distribution will treat the variance as a nuisance parameter.

In general, the parameter space can be divided in two types: $\theta = (\mu, \nu)$, where μ are the parameters of interest and ν are the nuisance parameters. The standard procedure to handle nuisance parameters is to marginalize them from the posterior. Therefore, we have that the marginalized posterior distribution will be:

$$p(\mu|d) = \int p(\mu, \nu|d) d\nu = \int p(\mu|\nu, d) p(\nu|d) d\nu, \tag{B.5}$$

where we can see that the marginalized posterior over the nuisance parameters depends on the conditional distributions, $p(\mu|\nu, d)$, times the posterior density $p(\nu|d)$. That is to say, a combination of the experimental data and the priors chosen to model the nuisance parameters.

B.4 Credible intervals

The posterior distribution can be used to define credible intervals. A credible interval is an interval in which the value of the parameter falls with a probability selected beforehand. A $(1-\alpha)$ credible interval is defined as in Eq. B.6:

$$\int_a^b p(\theta|d)d\theta = 1-\alpha. \tag{B.6}$$

This can be generalized to higher dimensions, known as credible regions, by integrating over a volume. There are several methods to select the bounds of these intervals. One of them is to define the interval such that we have the same probability above and below a selected point. These are called equal-tailed intervals and will contain the median of the parameter.

Credible intervals are the Bayesian analogue of frequentist confidence intervals. There are however some key differences in how to interpret them. From a frequentist point of view it doesn't make sense to consider the probability of a parameter being inside an interval, a parameter is a fixed value and will be inside or it will not. Therefore, a confidence interval with confidence level of $(1-\alpha)$, constructed from an estimator of the parameter, gives an interval in within the estimator will fall for subsequent repetitions with a probability of $(1-\alpha)$. Meanwhile, in Bayesian inference, the interpretation is as follows: given the observed realization of the data, the parameter will be inside the credible interval with a probability of $(1-\alpha)$.

B.5 Sampling

From Bayes' rule in Eq. B.2 we can obtain an analytical expression for the posterior distribution. However, in general, we are not interested in the posterior itself but in quantities and distributions obtained by manipulations over this posterior. For example, estimating mean values, marginalizing over

the nuisance parameters or estimating the posterior of only one parameter. All these require to integrate the posterior over the sample space which is doable with numerical integration only when the number of parameters that describe the model is small. This problem is tackled by using sampling techniques.

The main idea is to generate a sample big enough of the posterior such that all the quantities we want can be estimated using these samples. There are two possible ways to proceed: Independent and dependent sampling. However, independent sampling methods are not easy to generalize and dependent techniques are usually used. Many important algorithms are based in Markov chain simulations [193] (Also known as Markov chain Monte Carlo or mcmc).

A Markov chain is a sequence of random variables $x^1, x^2, x^3, ...$ such that for all i the conditional distribution of x^i with respect all the previous elements in the sequence only depends in x^{i-1}. An important characteristic of this chain is that when the number of samples is big enough it will converge to the sampled distribution. In mcmc methods the first point is arbitrarily selected while the following are obtained by drawing from a distribution chosen beforehand such that the Markov chain condition is fulfilled. The new points of the chain are corrected to better approximate the distribution we want to sample and, depending on how this step is performed, we have different algorithms. One of the most widely used is the Metropolis-Hasting algorithm.

B.5.1 Metropolis-Hasting Algorithm

The basic idea of this method is to perform random walks around the parameter space. In each iteration a new point is added to the chain if a condition is true and rejected otherwise. We first need to define a "transition function" or "proposal density function", $Q(\theta^i|\theta^{i-1})$, that will be used to perform the random walks by drawing the new candidates.

The algorithm to sample from a distribution $p(\theta|d)$ can be summarized as follows:

1. Draw or select a starting point θ^0.
2. for i = 1,2,3...:
   (a) Draw a proposal, θ^*, from $Q(\theta^*|\theta^{i-1})$.
   (b) Estimate the following ratio:

$$r = \frac{Q(\theta^{i-1}|\theta^*)p(\theta^*|d)}{Q(\theta^*|\theta^{i-1})p(\theta^{i-1}|d)}$$

   (c) Generate a random number, u, following a Uniform(0,1) distribution.
   (d) If $u < min(1, r)$ then the proposal is accepted and we set $\theta^i = \theta^*$. Otherwise we reject the proposal and set $\theta^i = \theta^{i-1}$.

In other words, we accept θ^* with a probability of min(1,r) and we reject otherwise. This should be repeated until we reach the desired number of samples. If the number of samples is large enough for the problem this Markov chain will converge to the posterior distribution. However, due to the first point being random, some elements at the beginning of the chain might not follow the posterior distribution and should be discarded. This is known as the burn-in period.

In the case of a symmetric transition function, that is, $Q(\theta^i|\theta^{i-1}) = Q(\theta^{i-1}|\theta^i)$, the ratio r is given by the ratio of the target distributions at each point. This version is also known as Metropolis Algorithm. A common choice for the transition function is a normal distribution centered around the previous element of the chain with standard deviation of σ: $\theta^*|\theta^{i-1} \sim N(\theta^{i-1}, \sigma)$ or more explicitly:

$$Q(\theta^*|\theta^{i-1}) = \frac{1}{\sigma\sqrt{2\pi}} exp(\frac{1}{2\sigma^2}(\theta^* - \theta^{i-1})^2),$$

where the spread, σ, is selected taking into account the complexity of the problem and the expected acceptance rate or efficiency of the sampling. In one hand, if the spread is much larger than the width of the target distribution, the acceptance rate will be very small and we will have many repeated

samples in the chain. On the other hand, if the spread is too small then the parameter space will be covered very slowly and it will take much longer to converge.

Appendix C

Mitigation of environmental noise in O3

In October 2019, between the first and second half of the third observation run, there was a commissioning period in which the sensitivity of the interferometers was slightly improved. One of the tasks during this period was the mitigation of noise from environmental sources (see Sec. 2.6.3) in Virgo [39].

In Virgo, all the possible causes of environmental noise in key areas of the interferometer are closely monitored. This includes temperature (thermometers), magnetic fields (magnetometers), acoustic noise (microphones) and displacement of critical components (accelerometers), among others. In this section we will discuss the steps that were taken to find and mitigate these contributions in preparation for the second half of the third observation run.

C.1 Methodology

Identifying the source of environmental noise, also known as "Environmental noise hunting", is not an easy task due to the high complexity of the interferometer. There are, however, some techniques that have proven useful.

In general, the search for environmental noise hunting is empirical. That is, the sensitivity of the interferometer is examined for features that might have been produced by noise. The most common features that are searched for are lines or bumps in the data that are new or not understood. These are

then compared with known interventions in the interferometer at the same time and, if no coincidence is found, they are further investigated. Dedicated tools are used to estimate the correlation of the noise with the witness sensors available to try to identify the source.

The possible sources are then explored by performing noise injections, switch off tests or measuring the contribution by using portable sensors. Noise injections are performed by simulating the causes of environmental noise. For example, we apply displacements with known frequencies to critical components to search for resonances or scattered light. However, due to the large number of candidates, switch off tests are the best solution to constrain the possible source of noise. During a switch off test, the suspected elements are disconnected one by one while the sensitivity of the interferometer is closely monitored. If the noise disappears, then the source has been found and we only need to understand how it couples to the interferometer.

C.2 Scattered light mitigation

Scattered light studies were performed during this period. Firstly, a switch off test of the heating, ventilation and air conditioning (HVAC) system was performed. This system controls the temperature and dust decontamination of key areas and produces both acoustic and seismic disturbances that might couple with the interferometer. During this test, it was found that noise features in the sensitivity were correlated with working HVAC units (see Fig. C.1), in particular inside the detection clean room and the laser lab room.

To constrain the location of the source inside the detection clean room, acoustic noise was injected by means of a loudspeaker. With this injection we found an increase of noise in the sensitivity between 46 and 50Hz and, with further testing, the noise was constrained to the bench. The source was finally identified and the noise dampened by optimizing the alignment of the bench.

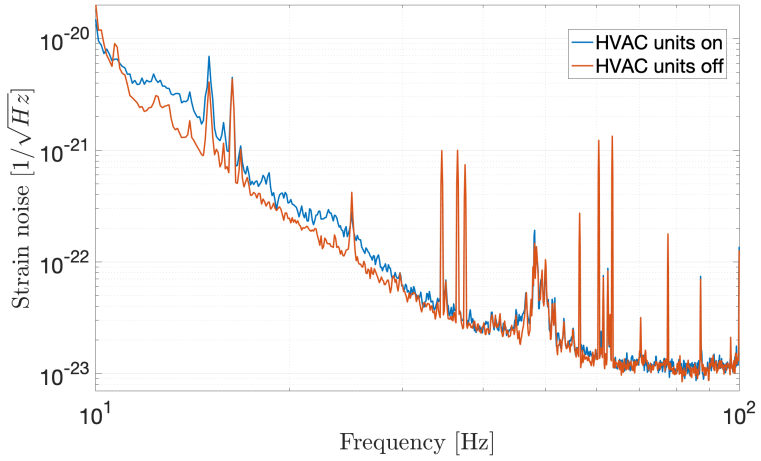

Figure C.1: Sensitivity before (blue) and after (orange) disconnecting all the HVAC units. The reduction of noise below 100Hz is clear. Taken from Ref. [39].

C.3 Electromagnetic noise studies

As it was mentioned in Sec. 2.6.3, understanding the effect of electromagnetic fields in our sensitivity is fundamental for some GWs searches. For that reason, magnetic fields [194] were injected on a weekly basis during the third observation run.

In this commissioning period, more injections were performed, both far field injections, to understand the coupling with ambient magnetic fields, and close field injections, to identify components susceptible to magnetic fields. The coupling between the magnetic field and the sensitivity was estimated as in Eq. C.1, where an upper limit was given when no visible effect was seen on the sensitivity during the injection. The coupling is considered measured when the effect on the nominal sensitivity is bigger by at least a factor two:

$$C(f)_{measured} = \sqrt{\frac{Y_{inj}^2 - Y_{bkg}^2(f)}{X_{inj}^2(f) - X_{bkg}^2(f)}} \tag{C.1}$$

$$C(f)_{upperlimit} = \frac{Y_{bkg}^2}{\sqrt{X_{inj}^2(f) - X_{bkg}^2(f)}}, \tag{C.2}$$

where Y is the sensitivity of the interferometer and X is the magnetic field as measured by the corresponding witness sensor, the magnetometers. The magnetometers are orientated in three perpendicular directions so that we can measure the three components of the magnetic field. We use as the total field the modulus of this vector. The coupling of the magnetic field for different periods can be seen in Fig. C.2. This coupling can be divided into two regions; a downward coupling below 100Hz and a rising slope above. The low frequency component is due to the coupling with the magnets attached to the mirrors. This coupling is expected to decay as f^3, which is the case for both the North End Building (NEB) and West End Building (WEB) but not the Central Building (CEB)[1]. Even though the reasons are unclear, it is reasonable to assume that this difference is due to coupling with more sources affected by the magnetic injection.

The rise at high frequency was investigated by performing close field injections around key components in CEB. Though we can't estimate the coupling with these studies, they are convenient to identify the elements more susceptible to magnetic fields. It was found that there is a significant coupling at 28Hz close to the input mirror of the north cavity, which is expected for the magnets controlling the mirror movements. There was also a significant coupling at 368Hz around the suspended detection bench. This indicates a possible coupling at higher frequencies: ambient magnetic fields affect the magnets acting on the bench and shaking it. In turn, this movement increases the amount of scattered light coupling back with the interferometer and limits the sensitivity between 300 and 600Hz.

[1]CEB: Building where the laser, beam splitter and input test masses are located; NEB: Building containing the end point of the north arm; WEB: Analogous to NEB but for the west arm.

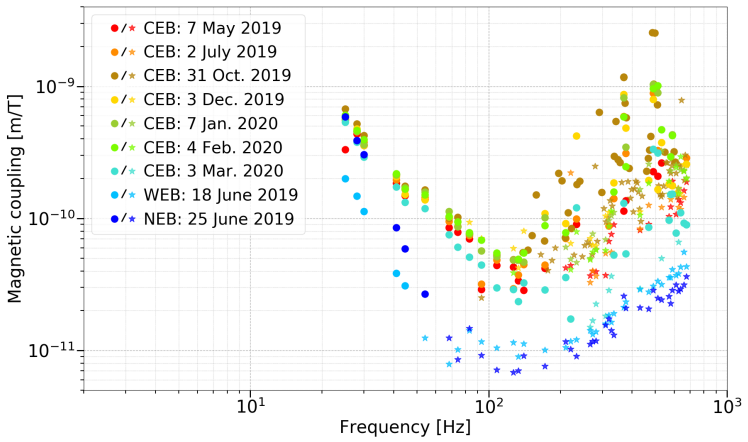

Figure C.2: Magnetic coupling for the Central Building (CEB), West End Building (WEB) and North End Building (NEB). A star corresponds to an upper limit and a star to measured value, as indicated by Eq. C.1. Taken from Ref. [39].

Using this coupling and the ambient magnetic field measurements, we can estimate the contribution to the noise from this source. This is shown in Fig. C.3. This projection has shown regions that will limit Virgo's sensitivity in the future. Many peaks are located at 50Hz (and subsequent harmonics), which correspond to the electrical frequency, but there are two, at 49.5Hz and 50.5Hz, that originated from a different source and limited the sensitivity during O3b. These were produced by currents flowing through a metal pipe into CEB and were successfully mitigated later.

These investigations are being used as reference for planning activities so that the noise budget from ambient magnetic fields is improved. For example, diaphragm baffles are going to be installed in the suspended detection bench to mitigate the scattered light coupling back into the interferometer and reduce the magnetic coupling above 100Hz.

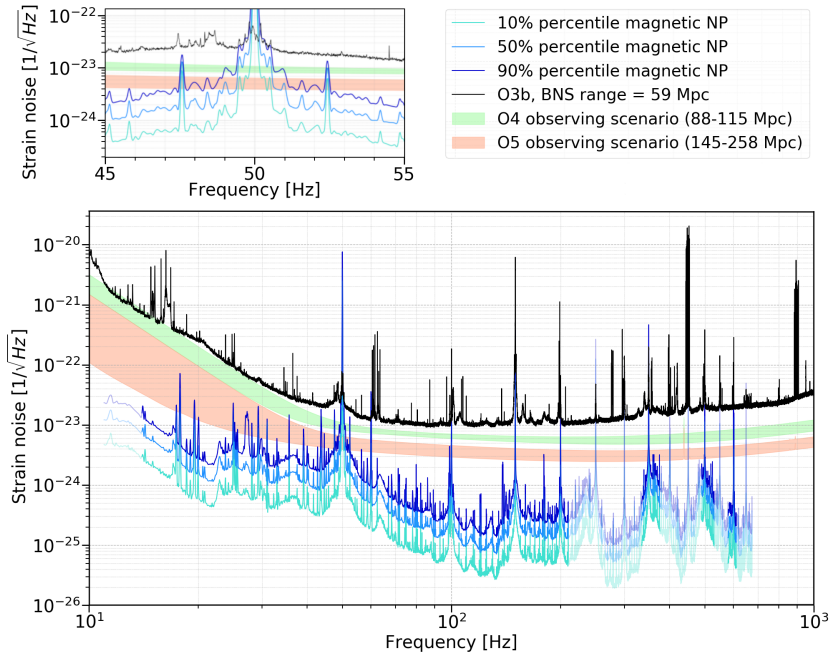

Figure C.3: Contribution to the noise from external magnetic field sources (NP). A continuous projection was obtained by using linear interpolation between the points used to estimate the coupling. Transparent regions correspond to upper limits. Taken from Ref. [39].

C.4 50Hz line and sidebands

As we have mentioned previously, the 50Hz line appears due to the electrical frequency in Italy. However, before O3, this line was unusually large and, though it was investigated, the reason was not found. At the end, it was decided to apply a correction using a feed forward control scheme to remove this noise from the sensitivity. The mitigation was done by applying a correction on the actuators that control the movement of the west end mirror. This correction is based on the voltage monitor of the uninterruptible power supply (UPS). This mitigation can be seen in Fig. C.4 (left).

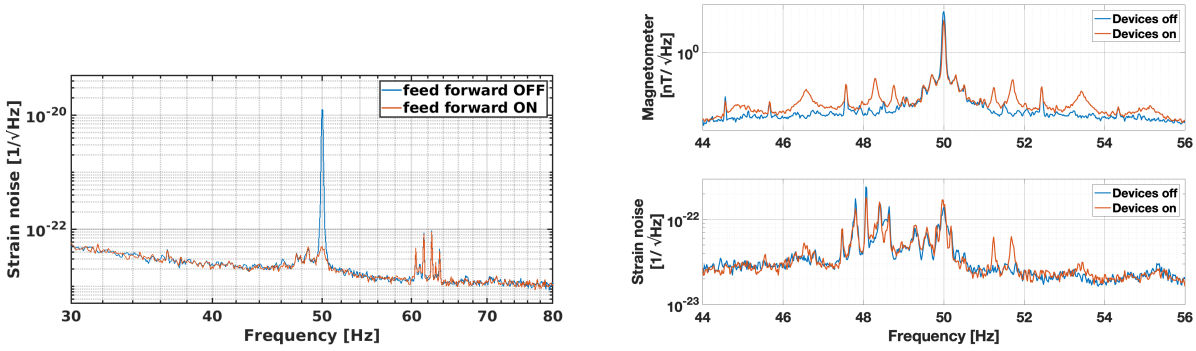

Figure C.4: (left) Feed foward control scheme applied to mitigate the 50Hz line. (right) The sidebands disappear from the magnetometer (top) and the strain (bottom) after switching off the infra-red laser beam profiles in the thermal compensation benches. Taken from Ref. [39].

Once the line was suppressed, sidebands around 50Hz were found in the residual noise, as seen in Fig. C.4 (right). In particular, two main families were found, one at $\pm 1.25 Hz$ and the other at $\pm 1.7 Hz$, both including harmonics. Using tools to correlate this noise with witness channels, it was found that there was a high correlation between these sidebands and the magnetic sensors and UPS monitor channels in the central building. This was followed by a search using a portable probe and a possible source was found in the thermal compensation benches, with the 1.2Hz coming from the west bench and the 1.7Hz from the north bench. After performing a switch off test, the origin of the noise was found. The culprits were devices used to monitor infra-red laser beam profiles that were installed to monitor the heating pattern profiles.

The devices were used only to monitor and the noise was successfully mitigated by switching them off during observation periods.

Appendix D

O3 scan results

In this appendix, the values of the discriminant and significance for all the O3 events are shown. Only the 51 events in H1-L1-V1 times have a value assigned. The discriminant assigned by the neural networks is given in the columns under the name "Discr" whereas the significance is given under the "FAR" columns. The acronyms correspond to the mass range covered by the neural network:

- HM: High mass - $[25M_{\odot}, 100M_{\odot}]$.
- MM: Masses in-between - $[5M_{\odot}, 25M_{\odot}]$.
- LM: Low mass - $[0.2M_{\odot}, 5M_{\odot}]$.

A detection is claimed when the FAR is lower than a preselected threshold.

Event	Discr HM	Discr LM	Discr MM	FAR HM	FAR LM,	FAR MM
GW190403_051519	0.79	0.18	0.5	37.344	>100	>100
GW190408_181802	1.0	0.2	0.88	0.007	>100	>100
GW190412_053044	0.99	0.24	0.9	0.007	>100	>100
GW190413_052954	0.77	0.15	0.73	68.624	>100	>100
GW190413_134308	0.88	0.29	0.41	5.01	>100	>100
GW190421_213856	-	-	-	-	-	-
GW190424_180648	-	-	-	-	-	-
GW190425_081805	-	-	-	-	-	-
GW190426_152155	0.06	0.55	0.22	>100	>100	>100
GW190426_190642	0.28	0.22	0.26	>100	>100	>100
GW190503_185404	0.99	0.3	0.76	0.007	>100	>100

GW190512_180714	0.79	0.18	0.8	39.744	>100	>100
GW190513_205428	1.0	0.14	0.76	0.007	>100	>100
GW190514_065416	-	-	-	-	-	-
GW190517_055101	0.96	0.16	0.43	0.185	>100	>100
GW190519_153544	1.0	0.18	0.4	0.007	>100	>100
GW190521_030229	1.0	0.15	0.49	0.007	>100	>100
GW190521_074359	-	-	-	-	-	-
GW190527_092055	-	-	-	-	-	-
GW190602_175927	0.78	0.17	0.67	53.144	>100	>100
GW190620_030421	-	-	-	-	-	-
GW190630_185205	-	-	-	-	-	-
GW190701_203306	1.0	0.15	0.49	0.007	>100	>100
GW190706_222641	1.0	0.22	0.47	0.007	>100	>100
GW190707_093326	-	-	-	-	-	-
GW190708_232457	-	-	-	-	-	-
GW190719_215514	-	-	-	-	-	-
GW190720_000836	0.07	0.14	0.58	>100	>100	>100
GW190725_174728	0.08	0.25	0.65	>100	>100	>100
GW190727_060333	1.0	0.16	0.76	0.007	>100	>100
GW190728_064510	0.17	0.54	0.97	>100	>100	5.035
GW190731_140936	-	-	-	-	-	-
GW190803_022701	0.85	0.16	0.35	10.01	>100	>100
GW190805_211137	0.81	0.2	0.46	26.524	>100	>100
GW190814_211039	-	-	-	-	-	-
GW190828_063405	1.0	0.24	0.64	0.007	>100	>100
GW190828_065509	0.17	0.2	0.46	>100	>100	>100
GW190909_114149	-	-	-	-	-	-
GW190910_112807	-	-	-	-	-	-
GW190915_235702	0.99	0.17	0.54	0.007	>100	>100
GW190916_200658	0.68	0.17	0.39	>100	>100	>100
GW190917_114630	0.05	0.21	0.2	>100	>100	>100
GW190924_021846	0.05	0.16	0.56	>100	>100	>100
GW190925_232845	-	-	-	-	-	-
GW190926_050336	0.75	0.2	0.2	>100	>100	>100
GW190929_012149	0.77	0.23	0.3	76.784	>100	>100
GW190930_133541	-	-	-	-	-	-
GW191103_012549	-	-	-	-	-	-
GW191105_143521	0.05	0.2	0.62	>100	>100	>100
GW191109_010717	-	-	-	-	-	-
GW191113_071753	0.07	0.19	0.33	>100	>100	>100

GW191126_115259	-	-	-	-	-	-
GW191127_050227	1.0	0.19	0.65	0.006	>100	>100
GW191129_134029	-	-	-	-	-	-
GW191204_110529	-	-	-	-	-	-
GW191204_171526	-	-	-	-	-	-
GW191215_223052	0.81	0.15	0.49	32.971	>100	>100
GW191216_213338	-	-	-	-	-	-
GW191219_163120	0.09	0.16	0.34	>100	>100	>100
GW191222_033537	-	-	-	-	-	-
GW191230_180458	0.98	0.17	0.63	0.505	>100	>100
GW200112_155838	-	-	-	-	-	-
GW200115_042309	0.08	0.17	0.23	>100	>100	>100
GW200128_022011	-	-	-	-	-	-
GW200129_065458	0.95	0.15	0.62	1.562	>100	>100
GW200202_154313	0.08	0.33	0.6	>100	>100	>100
GW200208_130117	0.99	0.27	0.38	0.223	>100	>100
GW200208_222617	0.51	0.22	0.44	>100	>100	>100
GW200209_085452	0.78	0.23	0.46	86.132	>100	>100
GW200210_092254	0.08	0.39	0.62	>100	>100	>100
GW200216_220804	0.47	0.17	0.47	>100	>100	>100
GW200219_094415	0.8	0.15	0.39	45.064	>100	>100
GW200220_061928	0.13	0.18	0.31	>100	>100	>100
GW200220_124850	-	-	-	-	-	-
GW200224_222234	1.0	0.25	0.95	0.006	>100	12.614
GW200225_060421	-	-	-	-	-	-
GW200302_015811	-	-	-	-	-	-
GW200306_093714	-	-	-	-	-	-
GW200308_173609	0.23	0.2	0.31	>100	>100	>100
GW200311_115853	1.0	0.18	0.94	0.006	>100	22.981
GW200316_215756	0.2	0.16	0.19	>100	>100	>100
GW200322_091133	0.17	0.14	0.48	>100	>100	>100

Appendix E

Inner weighted product

We define the inner weighted product for two real functions, $b(t)$ and $g(t)$, by a weight $S_n(f)$ as in Eq. E.1:

$$(b|g) \coloneqq \int_{-\infty}^{\infty} \frac{\tilde{b}^*(f)\tilde{g}(f)}{\frac{1}{2}S_n(f)}df = 2\int_0^{\infty} \frac{\tilde{b}(f)\tilde{g}^*(f) + \tilde{b}^*(f)\tilde{g}(f)}{S_n(f)}df. \tag{E.1}$$

It should be noted that $S_n(f) > 0$ and the numerator will be real due to being complex conjugate terms. For this reason another popular way to write this product is as seen in Eq. E.2:

$$(b|g) = 4Re\int_0^{\infty} \frac{\tilde{b}(f)\tilde{g}^*(f)}{S_n(f)}df. \tag{E.2}$$

Proposition 1. *The operation defined in Eq. E.1 is an inner product.*

Proof. Let $b(t)$ and $g(t)$ be two real functions and $S_n(f)$ a positive definite function.

1. Symmetry:

$$(b|g) = 2\int_0^{\infty} \frac{\tilde{b}(f)\tilde{g}^*(f) + \tilde{b}^*(f)\tilde{g}(f)}{S_n(f)}df = 2\int_0^{\infty} \frac{\tilde{b}^*(f)\tilde{g}(f) + \tilde{b}(f)\tilde{g}^*(f)}{S_n(f)}df =$$
$$= \int_0^{\infty} \frac{\tilde{g}(f)\tilde{b}^*(f) + \tilde{g}^*(f)\tilde{b}(f)}{S_n(f)}df = (g|b)$$

2. Linearity: Let a be any real number

$$(ab|g) = 2\int_0^{\infty} \frac{a\tilde{b}(f)\tilde{g}^*(f) + \overline{a}\tilde{b}^*(f)\tilde{g}(f)}{S_n(f)}df = 2a\int_0^{\infty} \frac{\tilde{b}(f)\tilde{g}^*(f) + \tilde{b}^*(f)\tilde{g}(f)}{S_n(f)}df = a(b|g)$$

3. Positive-definite:

$$(b|b) = 2\int_0^\infty \frac{\tilde{b}(f)\tilde{b}^*(f) + \tilde{b}^*(f)\tilde{b}(f)}{S_n(f)}df = 4\int_0^\infty \frac{|\tilde{b}(f)|^2}{S_n(f)}df \geq 0$$

And $(b|b) = 0 \iff b = 0$ due to the integrand being definite positive.

□

A physical interpretation of this inner product is that regions in which the detector is more noisy will have a bigger weight and contribute less to the total product.

Bibliography

[1] M. Maggiore. *Gravitational Waves: Volume 1: Theory and Experiments*. OUP Oxford, 2007.

[2] B. Schutz. *A First Course in General Relativity*. Cambridge University Press, 2009.

[3] M. Spivak. *A Comprehensive Introduction to Differential Geometry: Manifolds*. A Comprehensive Introduction to Differential Geometry. Publish or Perish, Incorporated, 1979.

[4] B.F. Schutz and D.B.F. Schutz. *Geometrical Methods of Mathematical Physics*. Geometrical Methods of Mathematical Physics. Cambridge University Press, 1980.

[5] J.A. Schouten. *Tensor Analysis for Physicists*. Dover Books on Physics Series. Dover Publications, 1989.

[6] V.P. Frolov and A. Zelnikov. *Introduction to Black Hole Physics*. OUP Oxford, 2011.

[7] M. Maggiore. *Gravitational Waves: Volume 2: Astrophysics and Cosmology*. OUP Oxford, 2018.

[8] S.S. Bayin. *Mathematical Methods in Science and Engineering*. Wiley, 2006.

[9] B.P. Abbott et al. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *Phys. Rev. X*, 9(3):031040, 2019.

[10] R. Abbott et al. Gwtc-2: Compact binary coalescences observed by ligo and virgo during the first half of the third observing run. *Physical Review X*, 11(2), jun 2021.

[11] The LIGO Scientific Collaboration and The Virgo Collaboration. GWTC-2.1: Extended catalogue of Compact Binary Coalescences Observed by LIGO 2 and Virgo During the First Half of the Third Observing Run. *In preparation*, 2021.

[12] B.P. Abbott et al. GW150914: The Advanced LIGO Detectors in the Era of First Discoveries. *Phys. Rev. Lett.*, 116(13):131103, 2016.

[13] Luc Blanchet. Gravitational radiation from post-newtonian sources and inspiralling compact binaries. *Living Reviews in Relativity*, 17(1), feb 2014.

[14] Mark Hannam. Status of black-hole-binary simulations for gravitational-wave detection. *Classical and Quantum Gravity*, 26(11):114001, may 2009.

[15] Alessandra Buonanno, Bala R. Iyer, Evan Ochsner, Yi Pan, and B. S. Sathyaprakash. Comparison of post-newtonian templates for compact binary inspiral signals in gravitational-wave detectors. *Physical Review D*, 80(8), oct 2009.

[16] Thibault Damour, Bala R. Iyer, and B. S. Sathyaprakash. Comparison of search templates for gravitational waves from binary inspiral. *Phys. Rev. D*, 63:044023, Jan 2001.

[17] Michael Boyle, Duncan A Brown, and Larne Pekowsky. Comparison of high-accuracy numerical simulations of black-hole binaries with stationary-phase post-newtonian template waveforms for initial and advanced LIGO. *Classical and Quantum Gravity*, 26(11):114006, may 2009.

[18] Manuel Tessmer and Achamveedu Gopakumar. On the ability of various circular inspiral templates to capture inspiral gravitational waves from compact binaries having tiny orbital eccentricities, 2008.

[19] L. Santamarí a, F. Ohme, P. Ajith, B. Brügmann, N. Dorband, M. Hannam, S. Husa, P. Mösta, D. Pollney, C. Reisswig, E. L. Robinson, J. Seiler, and B. Krishnan. Matching post-newtonian and numerical relativity waveforms: Systematic errors and a new phenomenological model for nonprecessing black hole binaries. *Physical Review D*, 82(6), sep 2010.

[20] Mark Hannam, Patricia Schmidt, Alejandro Bohé , Leïla Haegel, Sascha Husa, Frank Ohme, Geraint Pratten, and Michael Pürrer. Simple model of complete precessing black-hole-binary gravitational waveforms. *Physical Review Letters*, 113(15), oct 2014.

[21] Sebastian Khan, Katerina Chatziioannou, Mark Hannam, and Frank Ohme. Phenomenological model for the gravitational-wave signal from precessing binary black holes with two-spin effects. *Physical Review D*, 100(2), jul 2019.

[22] Sascha Husa, Sebastian Khan, Mark Hannam, Michael Pürrer, Frank Ohme, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. i. new numerical waveforms and anatomy of the signal. *Phys. Rev. D*, 93:044006, Feb 2016.

[23] Sebastian Khan, Sascha Husa, Mark Hannam, Frank Ohme, Michael Pürrer, Xisco Jimé nez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. a phenomenological model for the advanced detector era. *Physical Review D*, 93(4), feb 2016.

[24] A. Buonanno and T. Damour. Effective one-body approach to general relativistic two-body dynamics. *Physical Review D*, 59(8), mar 1999.

[25] Thibault Damour and Alessandro Nagar. Improved analytical description of inspiralling and coalescing black-hole binaries. *Physical Review D*, 79(8), apr 2009.

[26] Alessandra Buonanno, Yi Pan, Harald P. Pfeiffer, Mark A. Scheel, Luisa T. Buchman, and Lawrence E. Kidder. Effective-one-body waveforms calibrated to numerical relativity simulations: Coalescence of nonspinning, equal-mass black holes. *Physical Review D*, 79(12), jun 2009.

[27] Yi Pan, Alessandra Buonanno, Andrea Taracchini, Lawrence E. Kidder, Abdul H. Mroué , Harald P. Pfeiffer, Mark A. Scheel, and Béla Szilágyi. Inspiral-merger-ringdown waveforms of spinning, precessing black-hole binaries in the effective-one-body formalism. *Physical Review D*, 89(8), apr 2014.

[28] F Acernese, M Agathos, K Agatsuma, D Aisa, N Allemandou, A Allocca, and et al. Advanced virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, dec 2014.

[29] The LIGO Scientific Collaboration. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, mar 2015.

[30] The KAGRA Collaboration. Overview of kagra: Detector design and construction history, 2020.

[31] H Lück, C Affeldt, J Degallaix, A Freise, H Grote, M Hewitson, S Hild, J Leong, M Prijatelj, K A Strain, B Willke, H Wittel, and K Danzmann. The

upgrade of GEO 600. *Journal of Physics: Conference Series*, 228:012012, may 2010.

[32] The TAMA collaboration. Stable operation of a 300-m laser interferometer with sufficient sensitivity to detect gravitational-wave events within our galaxy. *Physical Review Letters*, 86(18):3950–3954, apr 2001.

[33] F. Acernese, M. Agathos, A. Ain, S. Albanesi, A. Allocca, and et al. Virgo detector characterization and data quality during the o3 run, 2022.

[34] E McClelland David. An overview of recycling in laser interferometric gravitational wave detectors. *Australian Journal of Physics*, 48:953–970, 1995.

[35] Brian J. Meers. Recycling in laser-interferometric gravitational-wave detectors. *Phys. Rev. D*, 38:2317–2326, Oct 1988.

[36] The LIGO Scientific collaboration. Enhanced sensitivity of the ligo gravitational wave detector by using squeezed states of light. *Nature Photonics*, 7(8):613–619, jul 2013.

[37] Moritz Mehmet and Henning Vahlbruch. The squeezed light source for the advanced virgo detector in the observation run o3. *Galaxies*, 8(4), 2020.

[38] M. Tse et al. Quantum-enhanced advanced ligo detectors in the era of gravitational-wave astronomy. *Phys. Rev. Lett.*, 123:231107, Dec 2019.

[39] Irene Fiori, Federico Paoletti, Maria Concetta Tringali, Kamiel Janssens, Christos Karathanasis, Alexis Menéndez-Vázquez, Alba Romero-Rodríguez, Ryosuke Sugimoto, Tatsuki Washimi, Valerio Boschi, Antonino Chiummo, Marek Cieślar, Rosario De Rosa, Camilla De Rossi, Francesco Di Renzo, Ilaria Nardecchia, Antonio Pasqualetti, Barbara Patricelli, Paolo Ruggi, and Neha Singh. The hunt for environmental noise in virgo during the third observing run. *Galaxies*, 8(4), 2020.

[40] Julia Casanueva Diaz. *Control of the gravitational wave interferometric detector Advanced Virgo*. Theses, Université Paris-Saclay, September 2017.

[41] B.P. Abbott et al. Calibration of the Advanced LIGO detectors for the discovery of the binary black-hole merger GW150914. *Phys. Rev. D*, 95(6):062003, 2017.

[42] J. Abadie et al. Calibration of the LIGO Gravitational Wave Detectors in the Fifth Science Run. *Nucl. Instrum. Meth. A*, 624:223–240, 2010.

[43] Ling Sun et al. Characterization of systematic error in Advanced LIGO calibration. *Class. Quant. Grav.*, 37(22):225008, 2020.

[44] N A Robertson, G Cagnoli, D R M Crooks, E Elliffe, J E Faller, P Fritschel, S Goszligler, A Grant, A Heptonstall, J Hough, H Luumlck, R Mittleman, M Perreur-Lloyd, M V Plissi, S Rowan, D H Shoemaker, P H Sneddon, K A Strain, C I Torrie, H Ward, and P Willems. Quadruple suspension design for advanced LIGO. *Classical and Quantum Gravity*, 19(15):4043–4058, jul 2002.

[45] S. Karki, D. Tuyenbayev, S. Kandhasamy, B. P. Abbott, T. D. Abbott, E. H. Anders, J. Berliner, J. Betzwieser, C. Cahillane, L. Canete, C. Conley, H. P. Daveloza, N. De Lillo, J. R. Gleason, E. Goetz, K. Izumi, J. S. Kissel, G. Mendell, V. Quetschke, M. Rodruck, S. Sachdev, T. Sadecki, P. B. Schwinberg, A. Sottile, M. Wade, A. J. Weinstein, M. West, and R. L. Savage. The advanced LIGO photon calibrators. *Review of Scientific Instruments*, 87(11):114503, nov 2016.

[46] The LIGO scientific collaboration. Improving astrophysical parameter estimation via offline noise subtraction for advanced LIGO. *Physical Review D*, 99(4), feb 2019.

[47] The LIGO Scientific Collaboration. LIGO detector characterization in the second and third observing runs. *Classical and Quantum Gravity*, 38(13):135014, jun 2021.

[48] L. K. Nuttall. Characterizing transient noise in the LIGO detectors. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2120):20170286, apr 2018.

[49] The LIGO Scientific Collaboration and The Virgo Collaboration. Characterization of transient noise in advanced LIGO relevant to gravitational wave signal GW150914. *Classical and Quantum Gravity*, 33(13):134001, jun 2016.

[50] Florent Robinet, Nicolas Arnaud, Nicolas Leroy, Andrew Lundgren, Duncan Macleod, and Jessica McIver. Omicron: A tool to characterize transient noise in gravitational-wave detectors. *SoftwareX*, 12:100620, jul 2020.

[51] Reed Essick, Patrick Godwin, Chad Hanna, Lindy Blackburn, and Erik Katsavounidis. iDQ: Statistical Inference of Non-Gaussian Noise with Auxiliary Degrees of Freedom in Gravitational-Wave Detectors. *Machine Learning: Science and Technology*, 2(1):015004, dec 2020.

[52] Benjamin P Abbott et al. A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals. *Class. Quant. Grav.*, 37(5):055002, 2020.

[53] Jolien Creighton and Warren Anderson. Gravitational-wave physics and astronomy: An introduction to theory, experiment and data analysis. *Gravitational-Wave Physics and Astronomy: An Introduction to Theory, Experiment and Data Analysis*, 09 2011.

[54] Piotr Jaranowski and Andrzej Krolak. Gravitational-Wave Data Analysis. Formalism and Sample Applications: The Gaussian Case. *Living Rev. Rel.*, 8:3, 2005.

[55] K.R. Rao, D.N. Kim, and J.J. Hwang. *Fast Fourier Transform - Algorithms and Applications*. Signals and Communication Technology. Springer Netherlands, 2011.

[56] E. O. Brigham and R. E. Morrow. The fast fourier transform. *IEEE Spectrum*, 4(12):63–70, 1967.

[57] James Cooley and John Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.

[58] F.J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

[59] Tukey J W. An introduction to the calculations of numerical spectrum analysis. *Analysis of time series*, pages 25–46, 1967.

[60] P. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.

[61] Elena Cuoco, Giovanni Calamai, Leonardo Fabbroni, Giovanni Losurdo, Massimo Mazzoni, Ruggero Stanga, and Flavio Vetrano. *Classical and Quantum Gravity*, 18(9):1727–1751, apr 2001.

[62] Elena Cuoco, Giovanni Losurdo, Giovanni Calamai, Leonardo Fabbroni, Massimo Mazzoni, Ruggero Stanga, Gianluca Guidi, and Flavio Vetrano. Noise parametric identification and whitening for LIGO 40-m interferometer data. *Physical Review D*, 64(12), nov 2001.

[63] B.P. Abbott et al. Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914. *Class. Quant. Grav.*, 33(13):134001, 2016.

[64] The LIGO scientific collaboration. Identification and mitigation of narrow spectral artifacts that degrade searches for persistent gravitational waves in the first two observing runs of advanced LIGO. *Physical Review D*, 97(8), apr 2018.

[65] B.P. Abbott et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16):161101, 2017.

[66] The LIGO Scientific Collaboration, The Virgo Collaboration, and The KAGRA Collaboration. Gwtc-3: Compact binary coalescences observed by ligo and virgo during the second part of the third observing run, 2021.

[67] G. Turin. An introduction to matched filters. *IRE Transactions on Information Theory*, 6(3):311–329, 1960.

[68] M. V. van der Sluys, C. Röver, A. Stroeer, V. Raymond, I. Mandel, N. Christensen, V. Kalogera, R. Meyer, and A. Vecchio. Gravitational-wave astronomy with inspiral signals of spinning compact-object binaries. *The Astrophysical Journal*, 688(2):L61–L64, oct 2008.

[69] Bruce Allen. χ^2 time-frequency discriminator for gravitational wave detection. *Phys. Rev. D*, 71:062001, 2005.

[70] Bruce Allen, Warren G. Anderson, Patrick R. Brady, Duncan A. Brown, and Jolien D. E. Creighton. FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries. *Phys. Rev. D*, 85:122006, 2012.

[71] Ian W. Harry, Alexander H. Nitz, Duncan A. Brown, Andrew P. Lundgren, Evan Ochsner, and Drew Keppel. Investigating the effect of precession on searches for neutron-star-black-hole binaries with Advanced LIGO. *Phys. Rev. D*, 89(2):024010, 2014.

[72] Drew Keppel. Balancing act of template bank construction: Inspiral waveform template banks for gravitational-wave detectors and optimizations at fixed computational cost. *Physical Review D*, 87(12), jun 2013.

[73] Stanislav Babak. Building a stochastic template bank for detecting massive black hole binaries. *Class. Quant. Grav.*, 25:195011, 2008.

[74] Tito Dal Canton and Ian W. Harry. Designing a template bank to observe compact binary coalescences in advanced ligo's second observing run, 2017.

[75] Javier Roulet, Liang Dai, Tejaswi Venumadhav, Barak Zackay, and Matias Zaldarriaga. Template bank for compact binary coalescence searches in gravitational wave data: A general geometric placement algorithm. *Physical Review D*, 99(12), jun 2019.

[76] Cody Messick, Kent Blackburn, Patrick Brady, Patrick Brockill, Kipp Cannon, Romain Cariou, Sarah Caudill, Sydney J. Chamberlin, Jolien D. E. Creighton, Ryan Everett, Chad Hanna, Drew Keppel, Ryan N. Lang, Tjonnie G. F. Li, Duncan Meacher, Alex Nielsen, Chris Pankow, Stephen Privitera, Hong Qi, Surabhi Sachdev, Laleh Sadeghian, Leo Singer, E. Gareth Thomas, Leslie Wade, Madeline Wade, Alan Weinstein, and Karsten Wiesner. Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data. *Physical Review D*, 95(4), feb 2017.

[77] Samantha A Usman, Alexander H Nitz, Ian W Harry, Christopher M Biwer, Duncan A Brown, Miriam Cabero, Collin D Capano, Tito Dal Canton, Thomas Dent, Stephen Fairhurst, Marcel S Kehl, Drew Keppel, Badri Krishnan, Amber Lenon, Andrew Lundgren, Alex B Nielsen, Larne P Pekowsky, Harald P Pfeiffer, Peter R Saulson, Matthew West, and Joshua L Willis. The pycbc search for gravitational waves from compact binary coalescence. *Classical and Quantum Gravity*, 33(21):215004, Oct 2016.

[78] Alexander H Nitz. Distinguishing short duration noise transients in LIGO data to improve the PyCBC search for gravitational waves from high mass binary black hole mergers. *Classical and Quantum Gravity*, 35(3):035016, jan 2018.

[79] T A Callister, J B Kanner, T J Massinger, S Dhurandhar, and A J Weinstein. Observing gravitational waves with a single detector. *Classical and Quantum Gravity*, 34(15):155007, jul 2017.

[80] Cody Messick et al. Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data. *Phys. Rev. D*, 95(4):042001, 2017.

[81] Kipp Cannon, Chad Hanna, and Jacob Peoples. Likelihood-ratio ranking statistic for compact binary coalescence candidates with rate estimation, 2015.

[82] Vaibhav Tiwari. Estimation of the Sensitive Volume for Gravitational-wave Source Populations Using Weighted Monte Carlo Integration. *Class. Quant. Grav.*, 35(14):145009, 2018.

[83] Will M. Farr, Jonathan R. Gair, Ilya Mandel, and Curt Cutler. Counting and confusion: Bayesian rate estimation with multiple populations. *Physical Review D*, 91(2), jan 2015.

[84] The LIGO Scientific Collaboration and The Virgo Collaboration. THE RATE OF BINARY BLACK HOLE MERGERS INFERRED FROM ADVANCED LIGO OBSERVATIONS SURROUNDING GW150914. *The Astrophysical Journal*, 833(1):L1, nov 2016.

[85] The LIGO Scientific Collaboration and The Virgo Collaboration. SUPPLEMENT: "THE RATE OF BINARY BLACK HOLE MERGERS INFERRED FROM ADVANCED LIGO OBSERVATIONS SURROUNDING GW150914" (2016, ApJL, 833, l1). *The Astrophysical Journal Supplement Series*, 227(2):14, nov 2016.

[86] Rahul Biswas, Patrick R. Brady, Jolien D. E. Creighton, and Stephen Fairhurst. The Loudest event statistic: General formulation, properties and applications. *Class. Quant. Grav.*, 26:175009, 2009. [Erratum: Class.Quant.Grav. 30, 079502 (2013)].

[87] B.P. Abbott et al. Search for Subsolar-Mass Ultracompact Binaries in Advanced LIGO's First Observing Run. *Phys. Rev. Lett.*, 121(23):231103, 2018.

[88] B.P. Abbott et al. Search for Subsolar Mass Ultracompact Binaries in Advanced LIGO's Second Observing Run. *Phys. Rev. Lett.*, 123(16):161102, 2019.

[89] R. Abbott et al. Search for subsolar-mass binaries in the first half of advanced ligo's and advanced virgo's third observing run. *Phys. Rev. Lett.*, 129:061104, Aug 2022.

[90] Alexander Harvey Nitz and Yi-Fan Wang. Search for Gravitational Waves from High-Mass-Ratio Compact-Binary Mergers of Stellar Mass and Subsolar Mass Black Holes. *Phys. Rev. Lett.*, 126(2):021103, 2021.

[91] Tito Dal Canton 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, 2014.

[92] Samantha A. Usman et al. The PyCBC search for gravitational waves from compact binary coalescence. *Class. Quant. Grav.*, 33(21):215004, 2016.

[93] Alexander H. Nitz, Thomas Dent, Tito Dal Canton, Stephen Fairhurst, and Duncan A. Brown. Detecting binary compact-object mergers with gravitational waves: Understanding and Improving the sensitivity of the PyCBC search. *Astrophys. J.*, 849(2):118, 2017.

[94] Gareth S. Davies, Thomas Dent, Márton Tápai, Ian Harry, Connor McIsaac, and Alexander H. Nitz. Extending the PyCBC search for gravitational waves from compact binary mergers to a global network. *Phys. Rev. D*, 102(2):022004, 2020.

[95] Eric Thrane and Colm Talbot. An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models. *Publ. Astron. Soc. Austral.*, 36:e010, 2019. [Erratum: Publ.Astron.Soc.Austral. 37, e036 (2020)].

[96] J. Aasi et al. Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network. *Phys. Rev. D*, 88:062001, 2013.

[97] Nelson Christensen and Renate Meyer. Using markov chain monte carlo methods for estimating parameters with gravitational radiation data. *Phys. Rev. D*, 64:022001, May 2001.

[98] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Vousden, and L. Wade. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Physical Review D*, 91(4), feb 2015.

[99] B.P. Abbott et al. Search for subsolar-mass black hole binaries in the second part of Advanced LIGO's and Advanced Virgo's third observing run. In preparation.

[100] Alexander H. Nitz and Yi-Fan Wang. Broad search for gravitational waves from subsolar-mass binaries through ligo and virgo's third observing run, 2022.

[101] Sarah Shandera, Donghui Jeong, and Henry S. Grasshorn Gebhardt. Gravitational Waves from Binary Mergers of Subsolar Mass Dark Black Holes. *Phys. Rev. Lett.*, 120(24):241102, 2018.

[102] Divya Singh, Michael Ryan, Ryan Magee, Towsifa Akhter, Sarah Shandera, Donghui Jeong, and Chad Hanna. Gravitational-wave limit on the Chandrasekhar mass of dark matter. *Phys. Rev. D*, 104(4):044015, 2021.

[103] Ya. B. Zel'dovich and I. D. Novikov. The Hypothesis of Cores Retarded during Expansion and the Hot Cosmological Model. *Sov. Astron.*, 10:602, February 1967.

[104] Stephen Hawking. Gravitationally collapsed objects of very low mass. *Mon. Not. Roy. Astron. Soc.*, 152:75, 1971.

[105] Bernard Carr, Kazunori Kohri, Yuuiti Sendouda, and Jun'ichi Yokoyama. Constraints on primordial black holes. *Reports on Progress in Physics*, 84(11):116902, nov 2021.

[106] Pablo Villanueva-Domingo, Olga Mena, and Sergio Palomares-Ruiz. A brief review on primordial black holes as dark matter. *Frontiers in Astronomy and Space Sciences*, 8, may 2021.

[107] Bernard Carr and Florian Kühnel. Primordial black holes as dark matter candidates. *SciPost Physics Lecture Notes*, may 2022.

[108] B. J. Carr. The primordial black hole mass spectrum. *Astrophysical Journal*, 201:1–19, October 1975.

[109] J. C. Niemeyer and K. Jedamzik. Near-critical gravitational collapse and the initial mass function of primordial black holes. *Physical Review Letters*, 80(25):5481–5484, jun 1998.

[110] Christian T. Byrnes, Mark Hindmarsh, Sam Young, and Michael R. S. Hawkins. Primordial black holes with an accurate QCD equation of state. *JCAP*, 08:041, 2018.

[111] Sébastien Clesse and Juan García-Bellido. Massive Primordial Black Holes from Hybrid Inflation as Dark Matter and the seeds of Galaxies. *Phys. Rev.*, D92(2):023524, 2015.

[112] Juan García-Bellido and Ester Ruiz Morales. Primordial black holes from single field models of inflation. *Phys. Dark Univ.*, 18:47–54, 2017.

[113] Chris Pattison, Vincent Vennin, Hooshyar Assadullahi, and David Wands. Quantum diffusion during inflation and primordial black holes. *JCAP*, 10:046, 2017.

[114] Jose Maria Ezquiaga, Juan García-Bellido, and Ester Ruiz Morales. Primordial Black Hole production in Critical Higgs Inflation. *Phys. Lett. B*, 776:345–349, 2018.

[115] Jose María Ezquiaga, Juan García-Bellido, and Vincent Vennin. The exponential tail of inflationary fluctuations: consequences for primordial black holes. *JCAP*, 03:029, 2020.

[116] V. De Luca, G. Franciolini, and A. Riotto. NANOGrav Data Hints at Primordial Black Holes as Dark Matter. *Phys. Rev. Lett.*, 126(4):041303, 2021.

[117] Matteo Braglia, Dhiraj Kumar Hazra, Fabio Finelli, George F. Smoot, L. Sriramkumar, and Alexei A. Starobinsky. Generating PBHs and small-scale GWs in two-field models of inflation. *JCAP*, 08:001, 2020.

[118] Zihan Zhou, Jie Jiang, Yi-Fu Cai, Misao Sasaki, and Shi Pi. Primordial black holes and gravitational waves from resonant amplification during inflation. *Phys. Rev. D*, 102(10):103527, 2020.

[119] Shi Pi and Misao Sasaki. Primordial Black Hole Formation in Non-Minimal Curvaton Scenario. *arXiv:2112.12680*, 12 2021.

[120] Sébastien Clesse and Juan García-Bellido. GW190425, GW190521 and GW190814: Three candidate mergers of primordial black holes from the QCD epoch. 2020.

[121] Khun Sang Phukon, Gregory Baltus, Sarah Caudill, Sebastien Clesse, Antoine Depasse, Maxime Fays, Heather Fong, Shasvath J. Kapadia, Ryan Magee, and Andres Jorge Tanasijczuk. The hunt for sub-solar primordial black holes in low mass ratio binaries is open. 5 2021.

[122] Kunihito Ioka, Takeshi Chiba, Takahiro Tanaka, and Takashi Nakamura. Black hole binary formation in the expanding universe: Three body problem approximation. *Phys. Rev. D*, 58:063003, 1998.

[123] Misao Sasaki, Teruaki Suyama, Takahiro Tanaka, and Shuichiro Yokoyama. Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914. *Phys. Rev. Lett.*, 117(6):061101, 2016. [Erratum: Phys.Rev.Lett. 121, 059901 (2018)].

[124] Nicola Bellomo, José Luis Bernal, Alvise Raccanelli, and Licia Verde. Primordial Black Holes as Dark Matter: Converting Constraints from Monochromatic to Extended Mass Distributions. *JCAP*, 01:004, 2018.

[125] Takashi Nakamura, Misao Sasaki, Takahiro Tanaka, and Kip S. Thorne. Gravitational waves from coalescing black hole MACHO binaries. *Astrophys. J. Lett.*, 487:L139–L142, 1997.

[126] Ryan Magee, Anne-Sylvie Deutsch, Phoebe McClincy, Chad Hanna, Christian Horst, Duncan Meacher, Cody Messick, Sarah Shandera, and Madeline Wade. Methods for the detection of gravitational waves from subsolar mass ultracompact binaries. *Phys. Rev. D*, 98(10):103024, 2018.

[127] Yacine Ali-Haïmoud, Ely D. Kovetz, and Marc Kamionkowski. Merger rate of primordial black-hole binaries. *Phys. Rev. D*, 96(12):123523, 2017.

[128] Bradley J. Kavanagh, Daniele Gaggero, and Gianfranco Bertone. Merger rate of a subdominant population of primordial black holes. *Phys. Rev. D*, 98(2):023536, 2018.

[129] Ville Vaskonen and Hardi Veermäe. Lower bound on the primordial black hole merger rate. *Phys. Rev. D*, 101(4):043015, 2020.

[130] Gert Hütsi, Martti Raidal, Ville Vaskonen, and Hardi Veermäe. Two populations of LIGO-Virgo black holes. *JCAP*, 03:068, 2021.

[131] Zu-Cheng Chen and Qing-Guo Huang. Merger Rate Distribution of Primordial-Black-Hole Binaries. *Astrophys. J.*, 864(1):61, 2018.

[132] Matthew R. Buckley and Anthony DiFranzo. Collapsed Dark Matter Structures. *Phys. Rev. Lett.*, 120(5):051102, 2018.

[133] Thomas Greif, Volker Springel, Simon White, Simon Glover, Paul Clark, Rowan Smith, Ralf Klessen, and Volker Bromm. Simulations on a Moving Mesh: The Clustered Formation of Population III Protostars. *Astrophys. J.*, 737:75, 2011.

[134] Athena Stacy and Volker Bromm. Constraining the Statistics of Population III Binaries. *Mon. Not. Roy. Astron. Soc.*, 433:1094, 2013.

[135] Tilman Hartwig, Marta Volonteri, Volker Bromm, Ralf S. Klessen, Enrico Barausse, Mattis Magg, and Athena Stacy. Gravitational Waves from the Remnants of the First Stars. *Mon. Not. Roy. Astron. Soc.*, 460(1):L74–L78, 2016.

[136] F. Acernese et al. Calibration of Advanced Virgo and Reconstruction of the Gravitational Wave Signal $h(t)$ during the Observing Run O2. *Class. Quant. Grav.*, 35(20):205004, 2018.

[137] D. Davis et al. Ligo detector characterization in the second and third observing runs. *Classical and Quantum Gravity*, 38(13):135014, jun 2021.

[138] Gabriele Vajente, Yiwen Huang, Maximiliano Isi, Jenne C. Driggers, Jeffrey S. Kissel, Marek J. Szczepanczyk, and Salvatore Vitale. Machine-learning nonstationary noise out of gravitational-wave detectors. *Phys. Rev. D*, 101(4):042003, 2020.

[139] Surabhi Sachdev et al. The GstLAL Search Analysis Methods for Compact Binary Mergers in Advanced LIGO's Second and Advanced Virgo's First Observing Runs, 2019.

[140] Chad Hanna et al. Fast evaluation of multidetector consistency for real-time gravitational wave searches. *Phys. Rev. D*, 101(2):022003, 2020.

[141] F Aubin, F Brighenti, R Chierici, D Estevez, G Greco, G M Guidi, V Juste, F Marion, B Mours, E Nitoglia, O Sauter, and V Sordini. The mbta pipeline for detecting compact binary coalescences in the third ligo-virgo observing run. *Classical and Quantum Gravity*, 38(9):095004, apr 2021.

[142] Stephen Privitera, Satyanarayan R. P. Mohapatra, Parameswaran Ajith, Kipp Cannon, Nickolas Fotopoulos, Melissa A. Frei, Chad Hanna, Alan J. Weinstein, and John T. Whelan. Improving the sensitivity of a search for coalescing binary black holes with nonprecessing spins in gravitational wave data. *Phys. Rev. D*, 89(2):024003, 2014.

[143] Heather Kin Yee Fong. *From simulations to signals: Analyzing gravitational waves from compact binary coalescences*. PhD thesis, Toronto U., 2018.

[144] Patrick Godwin et al. Incorporation of Statistical Data Quality Information into the GstLAL Search Analysis. 10 2020.

[145] 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, 1991.

[146] Luc Blanchet, Thibault Damour, Bala R. Iyer, Clifford M. Will, and Alan G. Wiseman. Gravitational radiation damping of compact binary systems to second postNewtonian order. *Phys. Rev. Lett.*, 74:3515–3518, 1995.

[147] Eric Poisson. Gravitational waves from inspiraling compact binaries: The Quadrupole moment term. *Phys. Rev. D*, 57:5287–5290, 1998.

[148] Thibault Damour, Piotr Jaranowski, and Gerhard Schaefer. Dimensional regularization of the gravitational interaction of point masses. *Phys. Lett. B*, 513:147–155, 2001.

[149] Balazs Mikóczi, Matyas Vasuth, and Laszlo A. Gergely. Self-interaction spin effects in inspiralling compact binaries. *Phys. Rev. D*, 71:124043, 2005.

[150] Luc Blanchet, Thibault Damour, Gilles Esposito-Farese, and Bala R. Iyer. Dimensional regularization of the third post-Newtonian gravitational wave generation from two point masses. *Phys. Rev. D*, 71:124004, 2005.

[151] K. G. Arun, Alessandra Buonanno, Guillaume Faye, and Evan Ochsner. 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, 2009. [Erratum: *Physical Review D* **84**, 049901 (2011)].

[152] Alessandra Buonanno, Bala R. Iyer, Evan Ochsner, Yi Pan, and B. S. Sathyaprakash. Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors. *Phys. Rev. D*, 80:084043, 2009.

[153] Alejandro Bohé, Sylvain Marsat, and Luc Blanchet. Next-to-next-to-leading order spin-orbit effects in the gravitational wave flux and orbital phasing of compact binaries. *Class. Quantum Grav.*, 30:135009, 2013.

[154] Alejandro Bohé, Guillaume Faye, Sylvain Marsat, and Edward K. Porter. Quadratic-in-spin effects in the orbital dynamics and gravitational-wave energy flux of compact binaries at the 3PN order. *Class. Quantum Grav.*, 32(19):195010, 2015.

[155] Chandra Kant Mishra, Aditya Kela, K. G. Arun, and Guillaume Faye. Ready-to-use post-Newtonian gravitational waveforms for binary black holes with nonprecessing spins: An update. *Physical Review D*, 93(8):084054, 2016.

[156] Duncan A. Brown, Ian Harry, Andrew Lundgren, and Alexander H. Nitz. Detecting binary neutron star systems with spin in advanced gravitational-wave detectors. *Phys. Rev. D*, 86:084017, 2012.

[157] Ilya Mandel, Will M. Farr, and Jonathan R. Gair. Extracting distribution parameters from multiple uncertain observations with selection biases. *Mon. Not. Roy. Astron. Soc.*, 486(1):1086–1093, 2019.

[158] Salvatore Vitale, Davide Gerosa, Will M. Farr, and Stephen R. Taylor. Inferring the properties of a population of compact binaries in presence of selection effects. 7 2020.

[159] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016.

[160] Curt Cutler and Eanna 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, 1994.

[161] R. Abbott et al. Population properties of compact objects from the second ligo-virgo gravitational-wave transient catalog. *The Astrophysical Journal Letters*, 913(1):L7, may 2021.

[162] R. A. Allsman et al. MACHO project limits on black hole dark matter in the 1-30 solar mass range. *Astrophys. J. Lett.*, 550:L169, 2001.

[163] P. Tisserand et al. Limits on the Macho Content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds. *Astron. Astrophys.*, 469:387–404, 2007.

[164] L. Wyrzykowski, J. Skowron, S. Kozłowski, A. Udalski, M. K. Szymański, M. Kubiak, G. Pietrzyński, I. Soszyński, O. Szewczyk, K. Ulaczyk, R. Poleski, and P. Tisserand. The OGLE view of microlensing towards the Magellanic Clouds - IV. OGLE-III SMC data and final conclusions on MACHOs. *Monthly Notices of the RAS*, 416(4):2949–2961, October 2011.

[165] Miguel Zumalacarregui and Uros Seljak. Limits on stellar-mass compact objects as dark matter from gravitational lensing of type Ia supernovae. *Phys. Rev. Lett.*, 121(14):141101, 2018.

[166] Savvas M. Koushiappas and Abraham Loeb. Dynamics of Dwarf Galaxies Disfavor Stellar-Mass Black Holes as Dark Matter. *Phys. Rev. Lett.*, 119(4):041102, 2017.

[167] Timothy D. Brandt. Constraints on MACHO Dark Matter from Compact Stellar Systems in Ultra-Faint Dwarf Galaxies. *Astrophys. J. Lett.*, 824(2):L31, 2016.

[168] Bernard Carr, Sébastien Clesse, Juan Garc'ia-Bellido, and Florian Kuhnel. Cosmic conundra explained by thermal history and primordial black holes. *Phys. Dark Univ.*, 31:100755, 2021.

[169] Karsten Jedamzik. Consistency of Primordial Black Hole Dark Matter with LIGO/Virgo Merger Rates. *Phys. Rev. Lett.*, 126(5):051302, 2021.

[170] A. Menéndez-Vázquez, M. Kolstein, M. Martínez, and Ll.M. Mir. Searches for compact binary coalescence events using neural networks in the ligo/virgo second observation period. *Physical Review D*, 103(6), Mar 2021.

[171] Elena Cuoco, Jade Powell, Marco Cavaglià , Kendall Ackley, Michał Bejger, Chayan Chatterjee, Michael Coughlin, Scott Coughlin, Paul Easter, Reed Essick, Hunter Gabbard, Timothy Gebhard, Shaon Ghosh, Leïla Haegel, Alberto Iess, David Keitel, Zsuzsa Márka, Szabolcs Márka, Filip Morawski, Tri Nguyen, Rich Ormiston, Michael Pürrer, Massimiliano Razzano, Kai Staats, Gabriele Vajente, and Daniel Williams. Enhancing gravitational-wave science with machine learning. *Machine Learning: Science and Technology*, 2(1):011002, dec 2020.

[172] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Adaptive Computation and Machine Learning series. MIT Press, 2016.

[173] S. Raschka and V. Mirjalili. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, 3rd Edition*. Packt Publishing, 2019.

[174] Y-Lan Boureau, J. Ponce, and Yann Lecun. A theoretical analysis of feature pooling in visual recognition. pages 111–118, 11 2010.

[175] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.

[176] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[177] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[178] The LIGO Scientific Collaboration and the Virgo Collaboration. Open data from the first and second observing runs of advanced LIGO and advanced virgo. *SoftwareX*, 13:100658, jan 2021.

[179] Judith C. Brown. Calculation of a constant q spectral transform. *The Journal of the Acoustical Society of America*, 89(1):425–434, 1991.

[180] Francois Chollet et al. Keras. 2015.

[181] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[182] B.P. Abbott et al. GW170817: Measurements of neutron star radii and equation of state. *Phys. Rev. Lett.*, 121(16):161101, 2018.

[183] Astropy Collaboration and Astropy Contributors. The astropy project: Building an open-science project and status of the v2.0 core package. *The Astronomical Journal*, 156(3):123, aug 2018.

[184] Massimiliano Razzano and Elena Cuoco. Image-based deep learning for classification of noise transients in gravitational wave detectors. *Classical and Quantum Gravity*, 35(9):095016, apr 2018.

[185] Rahul Biswas, Lindy Blackburn, Junwei Cao, Reed Essick, Kari Alison Hodge, Erotokritos Katsavounidis, Kyungmin Kim, Young-Min Kim, Eric-Olivier Le Bigot, Chang-Hwan Lee, John J. Oh, Sang Hoon Oh, Edwin J. Son, Ye Tao, Ruslan Vaulin, and Xiaoge Wang. Application of machine learning algorithms to the study of noise artifacts in gravitational-wave data. *Phys. Rev. D*, 88:062003, Sep 2013.

[186] Marco Cavaglia, Kai Staats, and Teerth Gill. Finding the origin of noise transients in ligo data with machine learning. *Communications in Computational Physics*, 25(4), 2019.

[187] O Ballester, O Blanch, L Cardiel, M Cavalli-Sforza, A Chiummo, C García, J M Illa, C Karathanasis, M Kolstein, M Martínez, A Menéndez-Vázquez, Ll M Mir, J Mundet, A Romero-Rodríguez, D Serrano, and H Yamamoto. Measurement of the stray light in the advanced virgo input mode cleaner cavity using an instrumented baffle. *Classical and Quantum Gravity*, 39(11):115011, may 2022.

[188] A. Romero-Rodríguez, A. Allocca, A. Chiummo, M. Martínez, Ll. M. Mir, and H. Yamamoto. Determination of the light exposure on the photodiodes

of a new instrumented baffle for the Virgo input mode cleaner end-mirror. *Class. Quant. Grav.*, 38(4):045002, 2021.

[189] H Yamamoto, A Romero, and T Edo. Sis20 - overview. *LIGO-T2000311-v2*, 2021.

[190] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman and Hall/CRC Texts in Statistical Science. Taylor and Francis, 2013.

[191] L. Wasserman and L.A. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer, 2004.

[192] Harold Jeffreys. An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society of London Series A*, 186(1007):453–461, September 1946.

[193] Matt Richey. The evolution of markov chain monte carlo methods. *American Mathematical Monthly*, 117:383–413, 05 2010.

[194] A Cirone, I Fiori, F Paoletti, M M Perez, A R Rodríguez, B L Swinkels, A M Vazquez, G Gemme, and A Chincarini. Investigation of magnetic noise in advanced virgo. *Classical and Quantum Gravity*, 36(22):225004, oct 2019.