A MACHINE LEARNING APPROACH TO THE SEARCH FOR GRAVITATIONAL WAVES EMITTED BY LIGHT SYSTEMS

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Science

Interactions Fondamentales en Physique et en Astrophysique
Département d'Astrophysique, Géophysique et Océanographie
Faculté des Sciences
Université de Liège
Belgique.

Grégory Baltus

2022

Artificial intelligence is moving in two direction: One is to teach us how to learn and the other is to make us learn by creating curiosity about the world [1].

- GPT-2 transformer-based language model

[1] An artificial intelligence, developed by OpenAI, generated this quote (`https://boredhumans.com/quotes.php`). And now, you understand why I don't use an AI to write this thesis.

Members of the thesis jury

Supervisor:
Jean-René Cudell

President:
Dominique Sluse

Secretary:
Maxime Fays

Members of the jury
Giacomo Bruno
Sarah Caudill

Abstract

With GW170817, gravitational waves have shown themselves to be very useful for multi-messenger astronomy. Combining the information from multiple channels such as gravitational waves, gamma-rays, neutrinos, etc. can lead to great physics. Contrarily to the electromagnetic telescopes, a gravitational wave interferometer surveys the entire sky. They do not have to focus on a small portion of the celestial sphere as do standard telescopes. It is also known that for binary neutron stars, the electromagnetic counterpart is produced during the last phase of the merger, whereas the gravitational wave signal can be detected several minutes before these last stages. If one is able to detect this signal before the merger and infer the sky location, gravitational wave astronomy can then send an alert and produce a sky map indicating where the astronomer can point their telescopes to see an electromagnetic counterpart.

The standard technique to detect these compact binary coalescences is matched filtering. The principle is to compute a template bank of pre-computed waveforms and match them with the data strain coming from the LIGO and Virgo interferometers. This thesis starts by illustrating a matched filter search with a project to detect long signals coming from sub-solar coalescence.

Recently, some matched filtering pipelines have started to adapt their method to search for gravitational waves with only the early stage of the signal. Other methods are beginning to be developed for this type of research. This thesis presents new methods, based on machine learning, to detect the early phase of a binary neutron star merger. We have developed multiple convolutional neural networks looking directly at the strain data of the detector to detect binary neutron stars before the merger.

The last step to produce an early warning for the astronomer is to create a sky map indicating the location of the event. We therefore shortly discuss how to accomplish this through a machine learning method for the whole signal, and also mention how it can be adapted to the early part of the signal.

Application des réseaux neuronaux à la recherche d'ondes gravitationnelles émises par des systèmes légers

Résumé

Avec GW170817, les ondes gravitationnelles ont montré qu'elles étaient d'une grande utilité pour l'astronomie multi-messagers. La combinaison des informations provenant de plusieurs canaux tels que les ondes gravitationnelles, les rayons gamma et les neutrinos peut mener à de grandes découvertes physiques. Contrairement aux télescopes calssiques, un interféromètre à ondes gravitationnelles scrute l'ensemble du ciel. Il n'a pas à se focaliser sur une petite portion de la sphère céleste comme le font les télescopes standards. On sait également que pour les binaires d'étoiles à neutrons, la contrepartie électromagnétique est produite lors de la dernière phase de la fusion des deux corps, alors que le signal des ondes gravitationnelles peut être détecté plusieurs minutes avant cette dernière étape. Si l'on est capable de détecter ce signal avant la fusion et d'en déduire la localisation dans le ciel, l'astronomie des ondes gravitationnelles peut alors envoyer une alerte et produire une carte du ciel indiquant où l'astronome peut pointer ses télescopes pour pouvoir observer une contrepartie électromagnétique.

La technique standard pour détecter ces coalescences binaires compactes est le filtrage adapté (*matched filtering*). Le principe est de construire une banque de modèles de formes d'ondes et de les faire correspondre avec les données provenant des interféromètres LIGO et Virgo. Cette thèse commence par illustrer une recherche utilisant le *matched filtering* avec un projet de détection de signaux longs provenant de coalescences d'objets ayant une masse inférieure à celle du soleil.

Récemment, certains pipelines au algorithmes de *matched filtering* ont commencé à adapter leurs méthodes de recherche d'ondes gravitationnelles en tenant compte uniquement des premiers stades du signal, appellé *l'inspiral*. D'autres techniques commencent également à être développées pour ce type de recherche. Cette thèse contribue à cette démarchant en présentant de nouvelles méthodes, basées sur l'apprentissage automatique (*machine learning*), pour détecter *l'inspiral* d'une fusion d'étoiles à neutrons. Nous avons développé de multiples réseaux neuronaux analysant les données des interféromètres pour détecter les binaires d'étoiles à neutrons avant leur fusion.

La dernière étape pour produire une alerte pour l'astronome est de créer une carte du ciel indiquant la localisation de l'événement. Cette thèse présente une méthode pour obtenir cette localisation à l'aide de *machine learning* et de réseaux neuronaux en utilisant l'entièreté du signal. Il est indiqué comment cette méthode peut être modifiée pour obtenir cette localisation avec seulement *l'inspiral*.

Acknowledgements

Les qualités d'un bon promoteur sont, selon moi : l'intérêt, le dévouement, la capacité à offrir des clés essentielles au développement de la recherche. Je peux m'estimer chanceux, car mon promoteur ne réunissait pas seulement ces caractéristiques, mais également toutes celles qui font de lui un bon leader et le meilleur des mentors. Grâce à toi Jean-René, j'ai réussi à réaliser une thèse dans un domaine dont je ne connaissais rien, j'ai acquis des compétences que je ne pensais pas pouvoir acquérir si vite, et je me sens grandi après ces 4 années. Je te remercie aussi de m'avoir soutenu et compris lors des moments difficiles, spécialement durant la première année. Peu ont la chance de pouvoir discuter avec leur promoteur comme un ami, c'était mon cas, je te remercie pour toutes nos discussions intéressantes et enrichissantes sur des sujets qui vont bien au-delà de la thèse et de la science.

I would like to thank my dear college and friend Atri. My ping pong, Linux, and physics master! I learn so much with you, thanks a lot! Sharing an office with you was just amazing. One of the thing I am grateful for, during this research, is that I met a lot of smart people, you are definitively one of them. The discussion about physics and the world that we have had together stimulated me a lot, and after this thesis, I am even more curious about all of that. Thanks once again! Come back to Belgium soon, so we can drink a beer together.

Il va aussi falloir remercier celui du bureau d'à côté : le bon vieux Vincent. Je ne pense pas qu'il puisse exister une autre personne aussi cool que toi pour être le deuxième PhD du groupe. Si mes années de thèse ont été aussi amusantes, c'est grâce à toi et à nos délires. Et tout ça, en étant stimulant et inspirant pour la recherche. Mais toi et moi ce n'est pas encore fini, il faut que l'on gagne encore quelques milliards ensemble !

On continue dans le couloir pour arriver à Max. Une semaine en Pologne ensemble, et après le premier jour, j'avais déjà compris que ça deviendrait vite plus qu'un collègue, mais un vrai pote. Je ne compte plus le nombre de soirées que j'ai passé chill chez toi après le boulot ! Merci pour tout ça, mais aussi pour tous les tuyaux, les trucs et astuces et les codes qui m'ont facilité la vie pendant cette thèse !

I also have to thank all the other colleagues: Prasanta, Rami, Maxim Latin, and Ica. You all contribute to the great ambiance that we have in our small group on the top of B5a! I'm really happy that my first job was with such a great team.

J'étais d'abord le PhD que l'étudiant va voir pour demander conseil, ensuite le doctorant un peu plus âgé à qui le petit nouveau parle avec respect, puis je suis devenu son collaborateur, et aujourd'hui, c'est lui le sage à qui je m'adresse quand je ne sais pas quelque chose. Justin bien que tu ne sois pas à Liège, tu étais un vrai collègue et ami pour moi. Merci pour tonaide sur cls projets. Si j'arrive au bout de cette thèse, c'est aussi énormément grâce à toi ! Travailler avec toi était un vrai plaisir !

I want to say a big thanks to Sarah Caudill, she helps us to start our little group in Liege. Thanks to you our group is deeply connected with Nikhef and Utrecht. All the work in this thesis is because of you: in a way or another, you put me in touch with great people with whom I could work and do good science.

I also want to thank Melissa. Just like Justin, it was a real pleasure to work with you ! It is amazing to have nice collaborators that become friends and work with them ! Also, a special thank for this very nice time in Valencia !

I also thank the other members of the jury: Giacomo and Dominique. It was a real pleasure to discuss with you in multiple occasion during these four years.

Un énorme merci à Greg et à Alexandra (et Cindy aussi) qui ont pris le temps de relire des parties de ma thèse et de corriger un grand nombre de fautes d'Anglais ou de formulation.

If this thesis, and the article in it, were so fun to work on, it is mainly because of the people I have worked with. So, I also would like to thank Alex Kolmus, Amit Reza, Arsh Narola, Sébastien Clesse, Antoine Depasse, Khun Sang Phukon, and all the others.

Si je suis arrivé où je suis maintenant, c'est surtout à cause de deux personnes. Deux personnes qui ont toujours été là pour moi, deux personnes qui m'ont donné le goût des études et ont forgé ma volonté. Deux personnes qui ont aussi fait tout ce qu'il était possible pour que je puisse faire mes études dans les meilleures conditions. Ils étaient là dans les moments durs comme dans les moments de joie. Je leur dois tout ce que je suis et tout ce que j'ai. Merci Bé et Gust, les meilleurs parents que l'on puisse rêver d'avoir.

Je dois évidemment dire merci à mon être humain préféré, la personne qui me connaît le mieux et à qui je fais le plus confiance. Tu ne t'en rendais probablement pas compte, mais quand j'étais mal, que ce soit pendant mes études ou ma thèse, personne n'était plus fort que toi, Loïc, pour me remonter le moral. Les soirées passées à deux à mon studio font partie de mes meilleurs souvenirs de ces quatre dernières années. Je te remercie du fond du cœur d'être ce que tu es frangin !

Je dois aussi remercier la bande des semis. Je les connais depuis que j'ai commencé mes études

de physique, et ils sont devenus mes amis les plus chers. Rien que pour eux, ça valait la peine d'avoir commencé mon parcours à l'université. Merci Arnaud, Jérémy, Sangio, Math et Lio d'avoir été là durant toutes ces années et de m'avoir autant stimulé avec toutes nos discussions qui auront aussi inspiré mon travail !

Je voudrais aussi remercier un autre groupe d'amis qui comme les autres m'a fait adorer ces années à l'université. Phil, Ben, Greg, Val, Art et Max, toutes ces années n'auraient clairement pas été aussi fun si je ne vous avais pas connu.

Merci aussi à mon frère d'armes Pierre Baibai. C'est aussi parce que tu m'as accompagné aux cours de QFT à l'époque que je suis où je suis aujourd'hui !

Avant de remercier la dernière personne, il faut que je remercie cette thèse elle-même. En effet, elle m'aura permis de l'utiliser comme phrase d'accroche pour rencontrer la personne qui a une place toute privilégiée dans mon cœur ! On dit souvent que l'écriture de la thèse est un des moments les plus stressants et pénibles du doctorat. Grâce à toi, Cindy, cette période aura été pour moi la plus belle de ces quatre années. L'écriture de cette thèse, je te la dédicace à toi tout particulièrement. Merci d'avoir été tant là pour moi, de m'avoir tant stimulé et aidé, et surtout de m'avoir fait sentir aussi bien durant ces 6 derniers mois. Cette thèse finit en véritable apothéose grâce à toi !

Table of Acronyms

- BH: Black Hole
- BBH: Binary Black Hole
- BNS: Binary Neutron Star
- CBC: Compact Binary Coalescence
- CE: Cosmic Explorer
- C.L.: Confidence Interval
- CNN: convolutional neural network
- DL: Deep Learning
- DM: Dark Matter
- EB: Early Binary
- EM: Electromagnetic
- ET: Einstein Telescope
- FAP: False Alarm Probability
- FAR: False Alarm Rate
- FN: False Negative
- FP: False Positive
- FP_t: False Positive over Time
- GCN: Gamma-ray Coordinates Network
- GR: General Relativity
- GRB: Gamma Ray Burst
- GW: Gravitational wave
- LB: Late Binary

- LIGO: Laser Interferometer Gravitational-wave Observatory
- MAAE: Mean Absolute Angular Error
- MCMC: Markov Chain Monte Carlo
- ML: Machine Learning
- MMA: Multi-Messenger Astronomy
- MSE: Mean Squared Error
- MVG: Multi Variate Gaussian
- NLL: negative log-likelihood
- NN: Neural Network
- NSBH: Neutron Star Black Hole
- OTW: Observational Time Window
- PBH: Primordial Black Holes
- PISNR: Partial Inspiral Signal to Noise Ratio
- PSD: Spectral Density
- SNR: Signal to Noise Ratio
- TAP: True Alarm Probability
- TBM: Time Before the Merger
- TN: True negative
- TP: True Positive
- VMF: Von Mises Fisher

Contents

Chapter 1

Introduction to gravitational waves

1.1 General relativity

In 2016 the LIGO-Virgo collaboration announced the first direct observation of a gravitational wave (GW) [1]. This earned Rainer Weiss, Kip Thorne, and Barry Barish a Nobel Prize as they played a key role in the development of the Laser Interferometer Gravitational-wave Observatory (LIGO). GWs are a prediction of the General Relativity (GR) of Einstein and can be seen as waves propagating via the vibration of space-time [2].

When sea waves propagate through water and a sound wave propagates through the air, a gravitational wave has no other substrate than space-time itself. Contrary to Newton's theory and even to special relativity, space-time is something that can be deformed in GR. In it, it is the energy of the matter which curves space-time, and this deformation is interpreted as gravity. For example, without any matter, space-time is flat, and the trajectory of the objects are straight lines, but if a massive object, like the earth, is present, space-time is deformed and the object follows this deformation, which makes it orbit the earth and fall on it. In a sense, these objects continue to follow straight lines, but they are not on a flat entity but on a curved one, in that case, they are called geodesics. Fig. 1.1 illustrates geodesics on different curved spaces. This idea is at the heart of GR, massive matter tells space-time how to curve, and space-time tells matter how to move. These two key principles are represented in GR by two sets of equations, respectively, Einstein's equations and the geodesic equations. To understand truly what a gravitational wave is, it is mandatory to have a basic understanding of GR and especially of these two sets of equations.

Before starting with the geodesic equations, which rule the trajectory of free objects within space-time, one must first introduce the means by which the space-time is modeled in GR and the different mathematical tools used in this theory. Space-time is seen as a manifold of 4 dimensions in GR. It is described by a second-order tensor, called the metric. This mathematical object is used to define the notion of distance between two points of a manifold. One can define an infinitesimal distance ds between two points, as follows:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \tag{1.1}$$

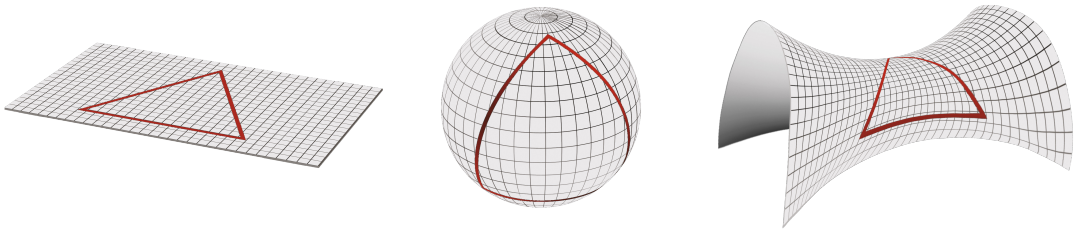

Figure 1.1: The first image represents three geodesics forming a triangle on a flat space. The second and third images represent respectively the three same geodesics on a sphere and on a hyperbolic geometry.

where x^μ represent the coordinates of space-time (t, x, y, z) and $g_{\mu\nu}$ is the metric tensor. Note that here Einstein's sum convention is used, meaning that if the same index appears once up and once down, there is an implicit summation on this index. The metric tensor is of capital importance in GR because it describes how space-time is deformed. The goal of Einstein's equations is to determine the values of this metric tensor.
The metric is something already used in special relativity. In that case, the value of the metric is the same at each point and is equal to:

$$\eta = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.2}$$

This special metric is called the Minkowski metric, it describes the flat space used in special relativity.

Going back to GR with a general metric $g_{\mu\nu}$, at each point of space-time, one can define a tangent space, where one can construct vectors and tensors. To compare vectors which are defined on different tangent spaces, one has to introduce the covariant derivative ∇_μ of a vector V:

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\rho} V^\rho \tag{1.3}$$

where $\Gamma^\nu_{\mu\rho}$ is a connection coefficient. This is an important object because the curvature of the space is described through this coefficient. Note that the connection coefficient is not a tensor. It is further to be noted that this connection is not unique. Nevertheless, only one of these coefficients is torsion-free (meaning $\Gamma^\nu_{\mu\rho} = \Gamma^\nu_{\rho\mu}$) and metric compatible (meaning $\nabla_\alpha g_{\mu\nu} = 0$). It is expressed as:

$$\Gamma^\alpha_{\mu\nu} = \frac{1}{2} g^{\alpha\beta} (\partial_\mu g_{\nu\beta} + \partial_\nu g_{\mu\beta} - \partial_\beta g_{\mu\nu}) \tag{1.4}$$

this special connection is the one usually used in GR and is called the Christoffel symbol. It is through this object that the link will be made between the curvature of space-time and the metric. From there, we have all the elements needed to introduce the geodesics and their equations. A

geodesic is a generalization, in curved space, of a straight line in flat space. Just like a straight line, a geodesic is the path with the minimal distance between two points. The mathematical equation of a geodesic is:

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda} = 0 \tag{1.5}$$

where λ is an affine parameter that parametrizes the path of the geodesic. Just as in Newton's theory free objects follow straight lines, in GR free objects follow geodesics. Consequently, the geodesic equations are the equations of motion. Within these equations appears $\Gamma^\alpha_{\mu\nu}$, and so through Equation (1.4) appears the metric. Before being able to solve the equations of motion, one needs to know the metric tensor. To do so, one needs to solve Einstein's equations.

Let us write these Einstein's equations and detail each term individually:

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

This set of equations are the Einstein equations, and, as mentioned earlier, the goal of these equations is to predict the metric $g_{\mu\nu}$. There are multiple terms within this equation and we shall discuss each.

The first term is known as the Ricci tensor, which is a contraction of the Riemann tensor, which is defined as follows:

$$R^\rho_{\theta\mu\nu} = \partial_\mu\Gamma^\rho_{\nu\theta} - \partial_\nu\Gamma^\rho_{\mu\theta} + \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\theta} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\mu\theta} \tag{1.7}$$

The Riemann tensor is an interpretation of the curvature of space-time. More precisely, it characterizes the curvature at each point of a manifold. To give a small intuition to the reader of why this tensor is linked to the curvature, it is opportune to mention the notion of parallel transport. In a flat space, if you define a vector at a point of space, you can "parallel transport" this vector to another point, and whichever the path that you decide to take, the final vector will be the same. This is illustrated on the first image of Fig. 1.2. This is not the case within a curved space, as shown in the second image of Fig. 1.2, where choosing two different paths to parallel transport a vector leads to a final vector that is not the same. This notion of parallel transport is useful in highlighting the fact that there is curvature. Without going into the details (for that see [3]), the Riemann tensor is constructed in such a way that it characterizes this variation of the parallel transported vector on an infinitesimal loop [3]. In a sufficiently small region, the spacetime can be considered flat, and an infinitesimal loop can be constructed with two vectors A^μ, B^ν. The purpose is to study the parallel transport of a vector V^ρ on this loop. First, let us parallel transport V^ρ in the direction of A^μ, then along B^ν, then backward, first following A^μ, then B^ν, as illustrated in Fig. 1.3. The change δV^ρ between the parallel transported vector and the initial one is express through the Riemann tensor:

$$\delta V^\rho = R^\rho_{\sigma\mu\nu}V^\sigma A^\mu B^\nu \tag{1.8}$$

From the Riemann tensor one can construct two other objects, the Ricci tensor $R_{\mu\nu}$ and the

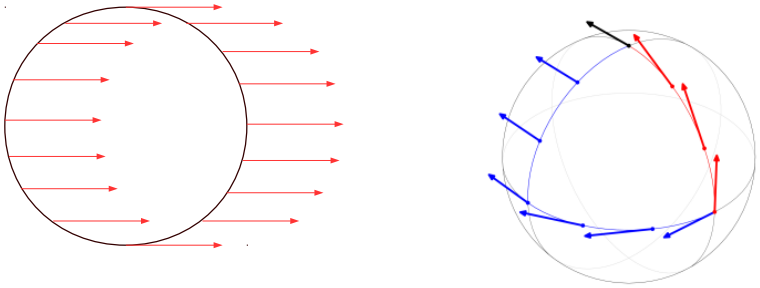

Figure 1.2: *Left*: Illustration of the parallel transport on a flat space. *Right*: Illustration of the parallel transport on a sphere. In that case, if one transports the black vector following the blue path, or following the red path, one ends up with a different vector.

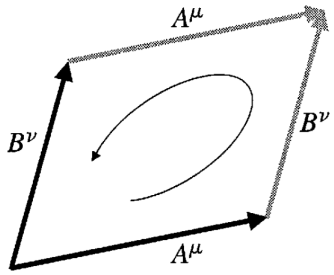

Figure 1.3: Illustration of an infinitesimal loop defined by two vectors A^μ and B^ν.

Ricci scalar R, which both appear in Einstein's equations (1.6) and are also linked to the curvature. The Ricci scalar is the contraction of the Riemann tensor:

$$R_{\mu\nu} = R^{\lambda}_{\mu\lambda\nu} \tag{1.9}$$

From this reduced tensor one can define the Ricci scalar:

$$R = R^{\mu}_{\ \mu} \tag{1.10}$$

Hence we already know the two first terms of Einstein's equations (1.6), these terms represent the curvature of the space-time. In the last term of the left-hand side of the equation, one can see the cosmological constant Λ and consider its effects on the expansion of the universe.

On the right-hand side of Einstein's equations, one can see the gravitational constant G, the speed of light c, and most notably the stress-energy tensor $T_{\mu\nu}$, which represents matter:

$$T_{\mu\nu} = \begin{pmatrix} u & \rho_x & \rho_y & \rho_z \\ \rho_x & P_{xx} & \sigma_{xy} & \sigma_{xz} \\ \rho_y & \sigma_{yx} & P_{yy} & \sigma_{yz} \\ \rho_z & \sigma_{zx} & \sigma_{zy} & P_{zz} \end{pmatrix} \tag{1.11}$$

where u is the energy density, ρ_i are the momentum densities, P_{ij} are the pressures and σ_{ij} are the shear stresses. Hence, the component $T_{\mu\nu}$ represents how much momentum in the μ direction flows into the ν direction through a given point in space-time. Einstein's equations tell us how the curvature (on the left-hand side of Equation (1.6)) reacts to the matter (on the right-hand side). As mentioned, these equations are solved in order to obtain the value of the metric $g_{\mu\nu}$.

Contrarily to special relativity or Newton's theory where one must choose an inertial reference system, in GR, one can choose any reference system. GR is a gauge theory, by choosing the coordinate system one fixes the gauge.

1.2 Gravitational waves

1.2.1 Theoretical definition

Now that the basic concepts of GR have been presented, we can discuss what a gravitational wave is. A GW is a local perturbation of the metric g which propagates through space. This perturbation obeys a wave equation. More formally, a gravitational wave is notably a particular solution of the linearized Einstein equation (the linearization is done on $h_{\mu\nu}$). Let us start by introducing this linearized form.

Einstein's equations are non-linear, and most of the time they do not have analytic solutions. But if one considers that we are in an almost Minkowski flat space, with just a weak perturbation of the metric, one can demonstrate that it is possible to linearize these equations and find some analytic solutions [1]. In fact, those solutions are the gravitational waves. So one considers that we are in a Minkowski space-time $\eta_{\mu\nu}$ with a small correction $h_{\mu\nu}$ due to a weak gravitational field:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \qquad |h_{\mu\nu}| << 1. \tag{1.12}$$

In order to arrive at the linearized Einstein equations, one must substitute all $g_{\mu\nu}$ (also those appearing in the Ricci tensor and in the Ricci scalar) in Equation (1.6) by this Equation (1.12) and keep only the first order in $h_{\mu\nu}$, that gives:

$$\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu} \tag{1.13}$$

where $\Box = \partial_\mu \partial^\mu$ is the d'Alembertian operator and $\bar{h}_{\mu\nu}$ is defined as:

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h \qquad \text{with } h = \eta^{\mu\nu} h_{\mu\nu} \tag{1.14}$$

To obtain the Equation (1.13), one must also work in the Lorentz gauge, which means that the following condition is satisfied:

$$\partial^\mu \bar{h}_{\mu\nu} = 0 \tag{1.15}$$

[1] It is possible to derive the GWs from a non-flat background metric, but for the simplicity of the demonstration, we consider here a Minkowski flat space.

These Equations (1.13) are the linearized Einstein equations. In the absence of any matter, they become:

$$\Box \bar{h}_{\mu\nu} = 0 \tag{1.16}$$

This is a wave equation for a wave traveling at the speed of light. A solution is for example a plane monochromatic wave propagating in the z direction:

$$\bar{h}_{\mu\nu} = A_{\mu\nu} \cos(\omega(t - z/c) + \varphi_0) \tag{1.17}$$

where, $\omega = 2\pi f$, with f the frequency of the wave, φ_0 a phase, and $A_{\mu\nu}$ a tensor defining the amplitude of the wave. In principle, $\bar{h}_{\mu\nu}$ is a symmetric 4×4 tensor, so it should have 10 independent components. Nevertheless, it is possible to reduce this number of freedom to two, by appropriately choosing the gauge. For further details we refer to [2, 4]. Note that with the right choice of gauge, one can have $\bar{h}_{\mu\nu} = h_{\mu\nu}$. Consequently, there exists a gauge, called the transverse-traceless gauge or TT gauge, that is such that Equation (1.17) can be rewritten as:

$$h^{TT}_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cos(\omega(t - z/c) + \varphi_0) \tag{1.18}$$

when we are in this TT gauge, we add a superscript TT on $h_{\mu\nu}$. With that, we have finally defined what a gravitational wave is. As one can see, a GW can be plus-polarized if $h_\times = 0$, cross-polarized if $h_+ = 0$, or elliptically polarized in the other cases.

1.2.2 Effect on matter

Now that gravitational waves are properly defined, let us explain how it is possible to detect them. First, one must understand their effect on matter. Let us consider that we are in a flat space-time with a gravitational wave, another way to say that is to write the metric:

$$g_{\mu\nu} = \eta_{\mu\nu} + h^{TT}_{\mu\nu} \tag{1.19}$$

$$= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 + h_+ \cos(\omega(t - z/c) + \varphi_0) & h_\times \cos(\omega(t - z/c) + \varphi_0) & 0 \\ 0 & h_\times \cos(\omega(t - z/c) + \varphi_0) & 1 - h_+ \cos(\omega(t - z/c) + \varphi_0) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.20}$$

To see the effect on the matter, one can write the interval:

$$\begin{aligned} ds^2 = &- c^2 dt^2 + \big(1 + h_+ \cos(\omega(t - z/c) + \varphi_0)\big) dx^2 \\ &+ \big(1 - h_+ \cos(\omega(t - z/c) + \varphi_0)\big) dy^2 \\ &+ \big(2h_\times \cos(\omega(t - z/c) + \varphi_0)\big) dx dy \\ &+ dz^2 \end{aligned} \tag{1.21}$$

More intuitively, consider first that $h_\times = 0$. In this case, the only differences with the flat Minkowski space-time are the terms in dx^2 and dy^2. One can see that the interval in the x direction of space will be stretched by a value of $h_+ \cos(\omega(t - z/c) + \varphi_0)\big)$ compared with the case without a GW, and the interval in the y direction will be contracted by a value $-h_+ \cos(\omega(t - z/c) + \varphi_0)\big)$. In other words, a plus-polarized GW will compress the interval in the direction perpendicular to its propagation, while it will stretch the other perpendicular direction by the same amount. Intuitively, it is thus possible to detect a GW if one can measure the difference of variations of distance in the two perpendicular directions. Now, let us see what happens if one considers a cross-polarized GW. In that case, one observes that only the term $dxdy$ varies from the standard Minkowski space. To understand what happened better, the easiest way is to change the coordinate system. Let us define the axes x' and y' as the rotation of 45° of x and y. With this coordinate system, ds is written:

$$ds^2 = -c^2dt^2 + \big(1 + h_\times \cos(\omega(t - z/c))\big)dx^2 + \big(1 - h_\times \cos(\omega(t - z/c))\big)dy''^2 + dz^2 \tag{1.22}$$

one can see that we obtain the same situation as in the case of a plus-polarization. Now, space is stretched in a direction at 45° of x and y, while it will be compressed in the perpendicular direction. Fig. 1.4 shows the effect of a plus and a cross-polarization on a disc of matter. One can also mention that usually in nature waves are a mixture of a cross and a plus-polarization, so most of the GWs are elliptically polarized.

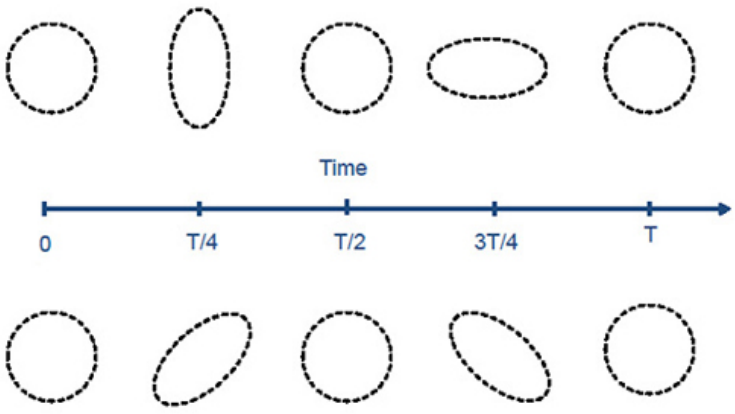

Figure 1.4: *Top*: Illustration of the effect of a plus-polarized GW on a circle of matter. *Bottom*: Illustration of the effect of a cross-polarized GW on a circle of matter.

1.2.3 Detector

Now that the effects of a GW on matter are known, we can think about how we can detect them. As we just discussed, they modify the distances in two typical directions, a way to make them stand out them is to measure the distance in two perpendicular directions. This is precisely the idea of the LIGO and Virgo interferometers. These detectors are based on the principle of a Michelson interferometer. This type of instrument is made of two perpendicular arms of the same length. At their intersection, a laser beam is split in two and each part goes through an arm. At the end of those arms, the beam is reflected on a mirror and goes back. The two beams are then sent to the same detectors, and the interference between them can be translated as the difference in the distances that each beam has covered. Fig. 1.5 illustrates how a GW detector works. If there is no GW, in principle, the interference should not change, because the distance in the arms remains constant. But if a GW passes through the detector, the distance in one direction will be periodically stretched while the other should be contracted, so the interference will change periodically, depending on the polarization of the GW. What is really measured is a phase shift of the laser beams, which represents this difference in distance between the two arms. Even if the principle is simple, in practice detecting GW is a challenge. This stems from the fact that GWs are usually very weak, and their effect on matter is extremely small. As we will see in the next section, the typical value of h_+ or $h_\times$ is around 10^{-22} meters. Consequently, the variation of distance that the GW detector must be able to detect is tiny, and compared to this faint signal, the amount of noise in the detector is huge. A lot of techniques must be deployed to reduce this noise and be able to see the gravitational wave. We will not enter into the details about the different types of noises, or the different hardware techniques to diminish them. Nevertheless, let us already introduce the Power Spectral Density (PSD), which is the curve that represents the sensitivity of the detector.

Let us say that $s(t)$ is the output time series of the detector. So the PSD $S_n(f)$ is defined such as:

$$\overline{\tilde{s}(f)\tilde{s}(f')} = \frac{1}{2}S_n(f)\delta(f - f') \tag{1.23}$$

where the bar denotes averaging over noise realizations, the δ represents the Dirac delta function, f represents the frequency and the tilde means that the Fourier transform is used, so for example $\tilde{s}(f)$ is defined as:

$$\tilde{s}(f) = \int_{-\infty}^{+\infty} s(t)e^{-2\pi itf}dt \tag{1.24}$$

The PSD shows the sensitivity of the detector as a function of the frequency. Fig. 1.6, represents the amplitude spectral density (ASD) of the LIGO and Virgo detectors during the third run of observation [5]. The ASD is defined as the square root of the PSD. Note that, the smaller the curve, PSD or ADS, the smaller the noise, and the higher the sensitivity of the detector. It is to be mentioned that a lot of different types of noise appear in that kind of detector, for example, the photon radiation pressure applied by the laser on the mirror, the Newtonian noise produced by terrestrial gravity fluctuations, the thermal noise coming from the Brownian motion of the atoms in the mirrors and their suspensions, etc. We refer the interested reader to [6] for more information

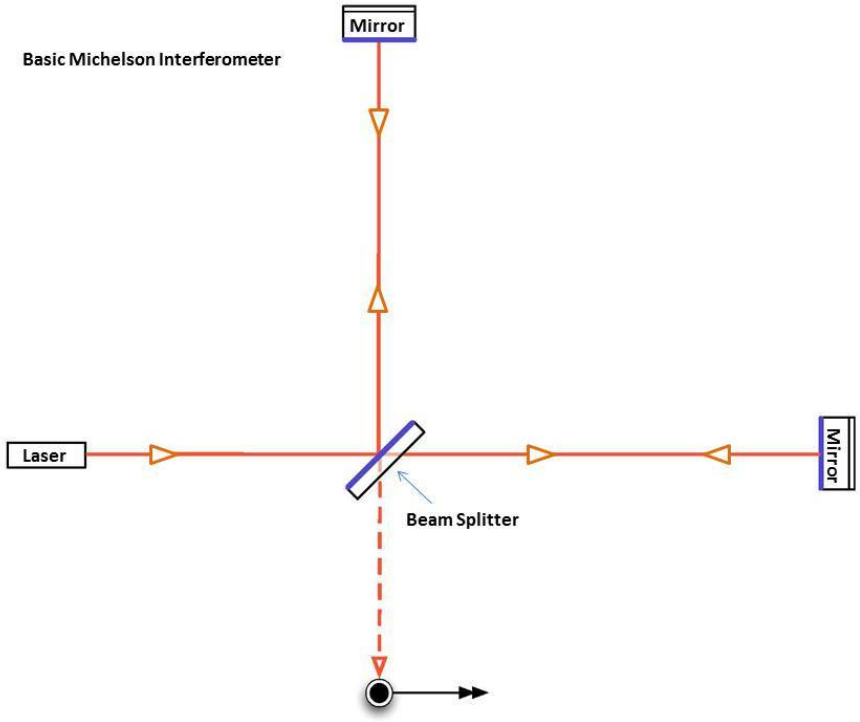

Figure 1.5: Illustration of the basic principle of a Michelson interferometer. The light emitted by the laser is split in the beam splitter, the light travels through the arms hits the mirror, and finally arrives at the receptor, where one can evaluate the interference and then the difference in distance between the two arms.

about the different types of noise and the detector hardware.

1.2.4 Compact binary coalescence

Now that it is know how to detect gravitational waves, it is time to discuss the different sources of GWs. Every system that has a variation of the quadrupole moment emits gravitational waves. To be more precise, the second and the third derivatives with respect to time of the quadrupole moment M^{ij} must be non-zero. It is possible to show that in the presence of matter, meaning that the wave equation for GWs is Equation (1.13), the solution depends on $\ddot{M}$, and so at a distance r from the source, h_+ and $h_\times$ become

$$h_+ = \frac{1}{r}\frac{G}{c^4}(\ddot{M}^{11} - \ddot{M}^{22}) \tag{1.25}$$

$$h_\times = \frac{2}{r}\frac{G}{c^4}\ddot{M}^{12} \tag{1.26}$$

for the details of this see [4]. Even if it is true that a lot of systems can generate GWs, in practice just a very few can produce GWs strong enough to be detected on earth. Indeed, a light object

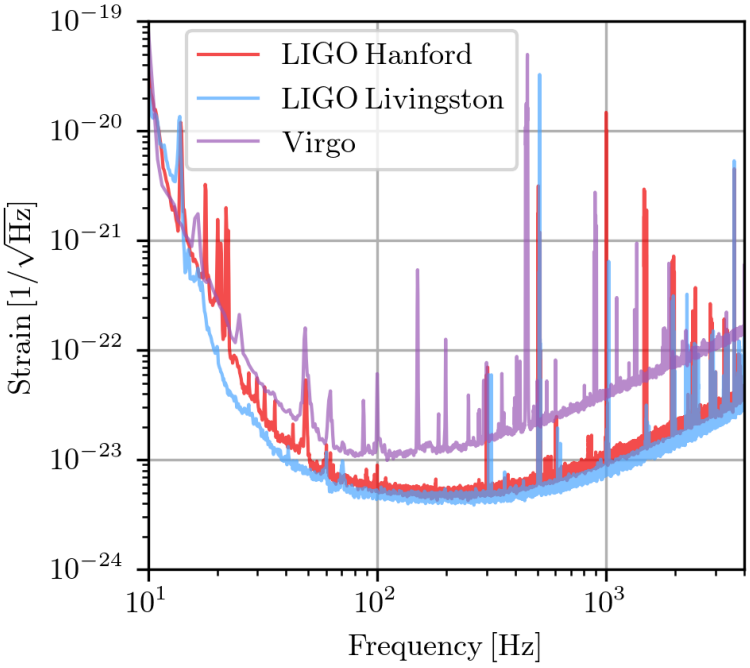

Figure 1.6: The amplitude spectral density during the third run of observation [5, 7, 8] for the two LIGO and the Virgo detectors.

hardly distorts space-time, and therefore only a very massive and dense object, like a binary black hole system or a binary neutron star system, can distort the fabric of space-time to generate detectable gravitational waves.
In gravitational wave searches, it is common to divide the sources of detectable GW in four different categories:

- compact binary coalescence
- continuous waves
- stochastic background
- burst

The Compact Binary Coalescence (CBC) is the source that will be of interest for this thesis. Those are two massive and compact objects rotating around each other before colliding and merging into a new more massive object. Only the densest astrophysical objects, such as black holes or neutron stars, can produce such signals [5, 7, 8]. This type of signal will be described in further detail, but first, let us explain the other types of sources. Continuous waves are all the astrophysical sources that emit GW on an ongoing basis, which means longer than weeks or months. Rotating neutron stars could for example emit such waves [9]. Another candidate for continuous waves is the creation of dark matter around a rotating black hole. One can show that if there exists dark matter consisting of light bosons, there is a possibility that these particles form a cloud around

a black hole by stealing its spin. When this cloud starts to collapse, it could emit gravitational waves [10]. There may exist a background noise in the detector that has a GW origin, it is called the stochastic gravitational-wave background. There are plenty of possible sources for this background, they range from a superposition of many unresolved sources coming from every direction in the sky to relic stochastic backgrounds from early Universe phase transitions, inflation or cosmic string [11]. The burst sources are all the other astrophysical sources that generate GWs. That could be supernovae [12], but also accretion-disk instabilities [13], non-axisymmetric deformations in magnetars [14] or gamma-ray bursts [15]. In general, the burst sources are not well modeled. To detect these kinds of signals, one generally looks at some excess power in the detector output.

All the events currently detected by the LIGO-Virgo-KAGRA collaboration, during the process of writing this thesis, come from the first source: the compact binary coalescence. This thesis focuses on this type of signal, so let us discuss them and study some of their proprieties. Fig. 1.7 represents the stellar graveyard. Even if most of the CBC events detected so far are Binary Black Holes (BBH), a few Binary Neutron Star (BNS) have been detected as well as two Neutron Star Black Hole (NSBH). A difference between BNS and BBH is the length of the signal. Indeed, BBHs are more massive than BNSs, and as we will see that the duration of the detectable signal depends on the masses of the system. Another difference between those two types of GWs is the tidal deformability. Under the effect of an external tidal field, the neutron star can be deformed, those deformations are called the tidal deformability, and a signature of this deformation can be found in the GW waveform [16].

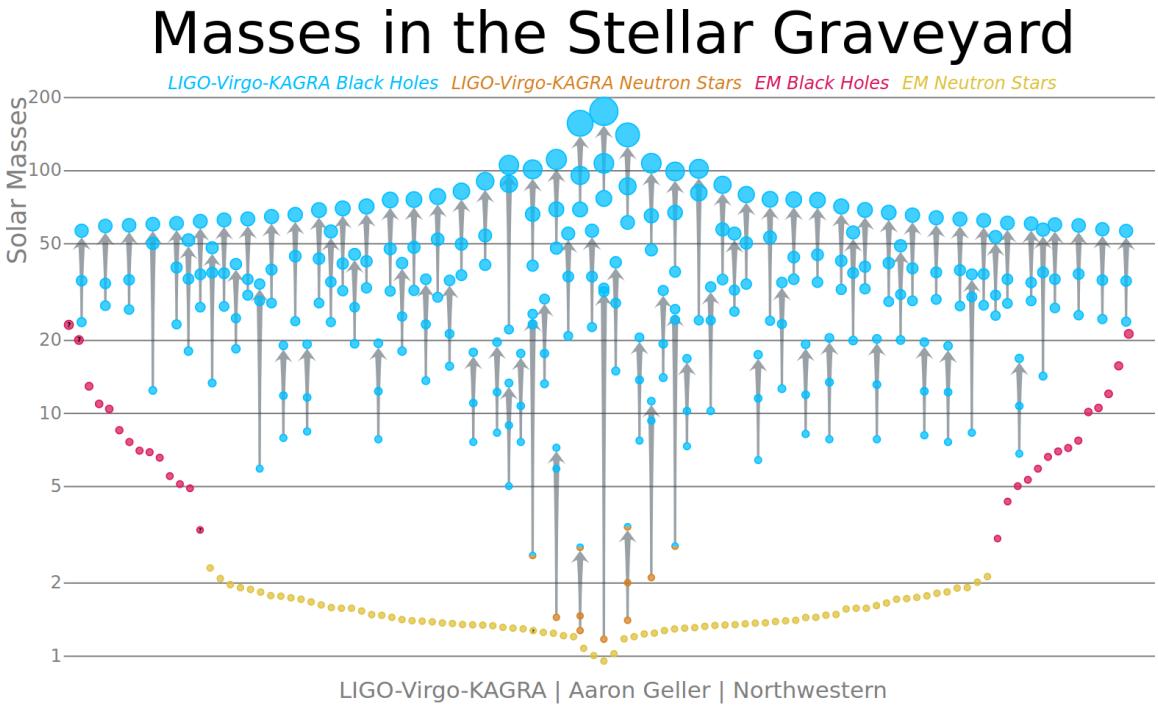

Figure 1.7: The stellar graveyard after the third run of observation [5, 7, 8].

A CBC signal can be decomposed into three different phases, the inspiral, the merger, and the ringdown, see Fig. 1.8. The inspiral is the phase where the two objects turn around each other. When they are too close, the system does not anymore have a stable orbit and the two objects plunge towards each other and merge into a new, more massive object, this stage is called the merger. The newly created object is initially in an excited stage, it will vibrate and emit GWs before reaching a stable mode, where no GWs are emitted. The merger and the ringdown are very short parts of the signal less than one second long. On the other hand, the detectable inspiral can be way longer, up to several minutes. As we will see, the duration of this phase depends on the mass of the system and thus is not the same for BBH as for BNS. In GR, the two-body problem does not have, in general, an analytic solution. Nevertheless, there exists a formalism called the post-Newtonian expansion, which approximates GR and can be used to describe the early inspiral phase with reasonably good accuracy. This formalism and the approximation that it supposed cannot anymore be used for the very late inspiral and the merger. One needs to resort to numerical simulation of GR to model this part of the signal. Finally, for the ringdown, it is possible to use a perturbative treatment in the BBH case.

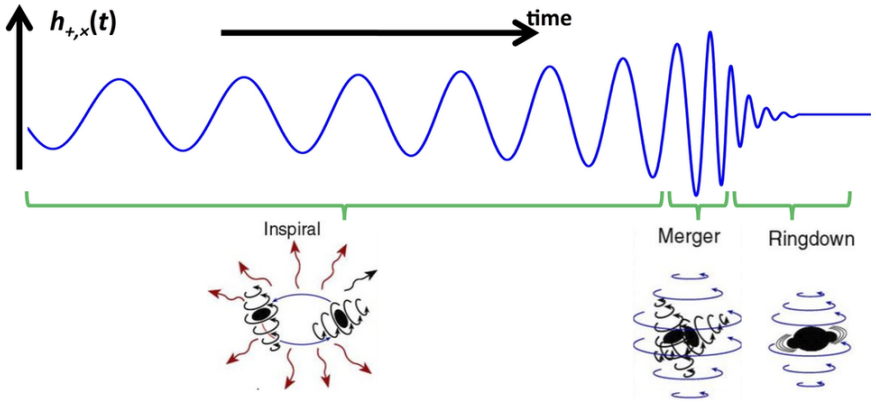

Figure 1.8: A Compact Binary coalescence signal. This figure comes from `https://www.soundsofspacetime.org/the-basics-of-binary-coalescence.html`

As this master thesis, will mostly focus on the early inspiral, let us show how it is possible to model it with the first order of the post-Newtonian formalism. To do so, first consider two massive particles of masses m_1 and m_2 rotating around each other in a circular orbit, at a distance R of each other. In that case, and if one chooses a coordinate system where the origin is the center of mass, it is possible to show that the quadrupole moment takes the form:

$$M^{ij}(t) = \mu R^2 \hat{e}^i \hat{e}^j \tag{1.27}$$

where $\mu = m_1 m_2/(m_1 + m_2)$ is the reduced mass and the vector $\hat{\mathbf{e}}$ is the unit vector pointing from the center of mass to the object of mass m_1. As previously, one can also choose the z axis to point

to the observer. In that case one can define the inclination angle ι, such as :

$$\hat{\mathbf{e}} = (\cos(\omega t), \cos(\iota)\sin(\omega t), \sin(\iota)\sin(\omega t)) \tag{1.28}$$

with ω the orbital angular frequency. Let us also mention here that ω, R and the masses of the system are not independent and are linked through Kepler's third law:

$$R^3 = \frac{GM}{\omega^2} \tag{1.29}$$

By injecting Equation (1.27) into Equations (1.25) and (1.26) and taking into account Equation (1.29), one can determine the value of h_+ and $h_\times$:

$$h_+ = -\frac{4}{r}\Big(\frac{G\mathcal{M}_c}{c^2}\Big)^{5/3}\Big(\frac{\omega}{c}\Big)^{2/3}\frac{1+\cos^2(\iota)}{2}\cos(2\omega t_{\text{ret}}) \tag{1.30}$$

$$h_+ = -\frac{4}{r}\Big(\frac{G\mathcal{M}_c}{c^2}\Big)^{5/3}\Big(\frac{\omega}{c}\Big)^{2/3}\cos(\iota)\cos(2\omega t_{\text{ret}}) \tag{1.31}$$

where, $t_{\text{ret}} = t - r/c$ is the retarded time and $\mathcal{M}_c$ is the chirp mass defined as:

$$\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}} \tag{1.32}$$

In Equations (1.30) and (1.31), one can observe that the orbital frequency arises in the cos and sin with a factor 2, meaning that the gravitational waves radiate with twice the orbital frequency of the system. In other words, one can obtain the gravitational wave frequency f_{gw} from the orbital frequency of the system $f_{orb} = \omega/(2\pi)$:

$$f_{gw} = 2f_{orb} \tag{1.33}$$

which also means that $\omega = \pi f_{gw}$.

Equations (1.30) and (1.31) are not completely true, for GWs carry energy away. One can show that the power carried by a GW is [4]:

$$\frac{dE_{GW}}{dt} = \frac{32}{5}\frac{c^5}{G}\Big(\frac{G\mathcal{M}_c\omega}{c^3}\Big)^{10/3} \tag{1.34}$$

Consequently, the two objects are not really in a circular orbit. The distance R will decrease with time, because of the loss of energy carried by the GW. This also means, as seen in Equation (1.29), that ω will increase. This ought to be taken into account in the expression of h_+ and $h_\times$. For that reason, let us first define the angle Φ which is the angle made by a fixed line and the line that joins the two particles, in other terms Φ is the angle used to define the orbital frequency ω:

$$\omega = \dot{\Phi}(t) \qquad \Phi(t) = \int^t \omega(t')dt'.$$

We also made an approximation that the change of ω is weak, this is valid when R is sufficiently large, typically when in the inspiral phase. This approximation lets us rewrite $\omega \to \omega(t_{\rm ret})$ and $\omega t_{\rm ret} \to \Phi(t_{\rm ret})$. The amount of power carried away by the gravitational wave, must be equal to the power of the orbital energy:

$$\frac{dE_{orb}}{dt} = -\frac{dE_{gw}}{dt} \tag{1.35}$$

where the orbital energy is:

$$E_{orb} = \frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2 - \frac{Gm_1 m_2}{R} \tag{1.36}$$

where v_1 and v_2 are respectively the speeds of particles 1 and 2. This equation can be rewritten as follows, using the Equation (1.29):

$$E_{orb} = -\frac{1}{2}(G^2 \mathcal{M}_c^5 \omega^2)^{1/3} \tag{1.37}$$

Thus one can rewrite Equation (1.35) with Equation (1.37) and (1.34):

$$\dot{f}_{gw}(t_{\rm ret}) = \frac{96}{5}\pi^{8/3}\Big(\frac{G\mathcal{M}_c}{c^3}\Big)^{5/3} f_{gw}^{11/3}(t_{\rm ret}) \tag{1.38}$$

which is a differential equation for the gravitational wave frequency, one can show that this equation is solved by:

$$f_{gw}(t) = \frac{1}{\pi}\Big(\frac{G\mathcal{M}_c}{c^3}\Big)^{-5/8}\Big(\frac{5}{256}\frac{1}{\tau(t)}\Big)^{3/8} \tag{1.39}$$

with $\tau(t) = t_c - t$, where t_c is the coalescence time, i.e. the time when the frequency formally diverges: $t \to t_c$ implies $f_{GW} \to \infty$, which also implies through Kepler's third law that $R \to 0$. In a real scenario, the objects are not point particles, but astrophysical objects, thus they have a size, and one never reaches this divergence. It is possible to see through the full general relativity that the innermost stable circular orbit (ISCO) is reached when the distance between the two objects is $R_{ISCO} \simeq 6GM/c^2$.

This Equation (1.39) is important in this work because it links the frequency of the wave and the time before the merger $\tau(t)$. As we will see in Chapter 3, one of our primary goals will be to detect the GW before the merger. To do that, we will have to look for a GW with a maximum frequency that can be interpreted as a typical time before the merger (assuming that the chirp mass is constant). Let us take a moment to study the significance of Equation (1.38) and its solution, Equation (1.39). From the differential equation for f, one can see that the variation of frequency as a function of the frequency, evolves as a power law with the frequency, meaning that the higher the frequency, the faster the frequency evolution. But the variation of frequency also depends on the chirp mass: the heavier the objects, the faster the evolution of frequency. The right-hand plot of Fig. 1.9 shows this double dependency. From Equation (1.39), one can see that running through a variation of frequency, takes a longer time if the chip mass is lower. This behavior is illustrated in the left-hand plot of Fig. 1.9. This means that the duration of the detectable inspiral is longer for lighter objects. In practice, most of the detections made so far are BBHs and have a detectable

inspiral which is very short, one second or less for most of them. But for BNSs, the detectable inspiral is way longer. If the minimal detectable frequency is about say 20 Hz, the inspiral of a BNS made of two objects of $1.4M_\odot$ lasts around three minutes. And if you can go down to 10 Hz, the inspiral could last for almost twenty minutes. This is one of the reasons why this thesis focuses more on BNSs than BBHs for the early detection. The other interest of BNSs, is obviously that they can emit electromagnetic counterparts, which will be discussed in Section 3.2.

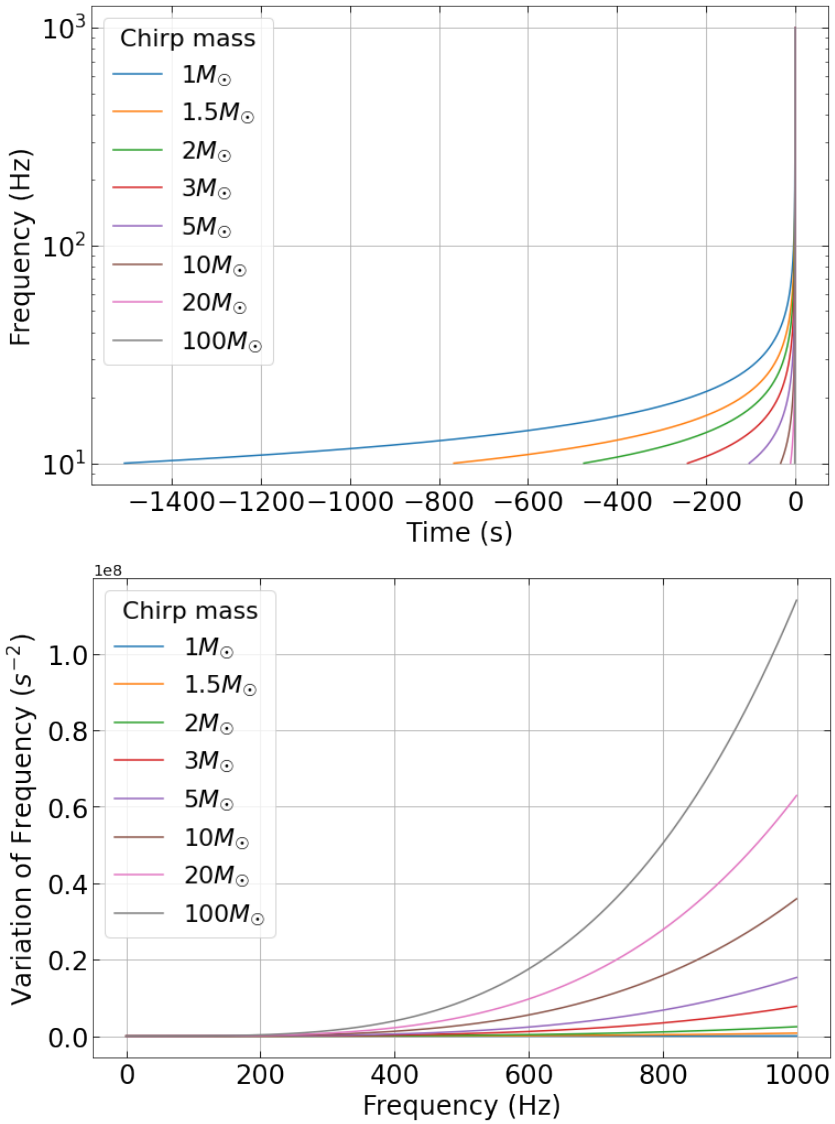

Figure 1.9: *Top*: The evolution of the frequency of the inspiral as a function of time for different chirp masses. *Bottom*: The variation of frequency as a function of the frequency of the inspiral for different chirp masses.

Let us go back to our problem to find the expression of h_+ and $h_\times$. With Equation (1.39) one can finally write the equation for Φ, where $\Phi_{gw} = 2\Phi$ because it has been seen that $f_{gw} = 2f_{orbi}$.

Using Equation (1.39) one obtains:

$$\Phi_{gw}(t) = -2\Big(\frac{5G\mathcal{M}_c}{c^3}\Big)^{-5/8}\tau^{5/8}(t) + \Phi_c \tag{1.40}$$

where Φ_c is the phase when $t = t_c$. One can finally rewrite h_+ and $h_\times$ for the inspiral:

$$h_+ = -\frac{4}{r}\Big(\frac{G\mathcal{M}_c}{c^2}\Big)^{5/3}\Big(\frac{\pi f_{gw}(t_{\text{ret}})}{c}\Big)^{2/3}\frac{1+\cos^2(\iota)}{2}\cos(\Phi_{gw}(t_{\text{ret}})) \tag{1.41}$$

$$h_+ = -\frac{4}{r}\Big(\frac{G\mathcal{M}_c}{c^2}\Big)^{5/3}\Big(\frac{\pi f_{gw}(t_{\text{ret}})}{c}\Big)^{2/3}\cos(\iota)\sin(\Phi_{gw}(t_{\text{ret}})) \tag{1.42}$$

With those expressions, one knows how much the space will be deformed by the passage of a GW coming from a CBC signal. Just to give us an intuition of the order of magnitude of this deformation, let us quickly do a calculation, if one considers the parameters of GW170817 [17], and that one wants to estimate h one second before the merger, we obtain $h_+ \sim \mathcal{O}(10^{-24})$. Thus, one can see that the effect of a GW on earth is tiny. It is why we need such huge interferometers. It also explains why so much effort is made to reduce the noise in the GW detectors.

Chapter 2

Standard searches for sub-solar objects

2.1 Matched filtering

In this chapter, matched filtering is introduced, which is the standard technique to detect GWs coming from CBCs. To illustrate this technique, a paper, in which we have participated is presented. It is a matched filtering search for the coalescence of sub-solar objects. Before presenting this paper, the key concepts of matched filtering and the sub-solar objects are introduced.

2.1.1 Antenna pattern

The output of GW detectors is a time-series, which characterizes the difference in length between the two arms of the interferometer. Let us denote the output of one detector as $s(t)$. In general, this time-series is just composed of noise $n(t)$, but if a GW goes through the detector, the output is the sum of the noise plus the gravitational wave $h(t)$:

$$s(t) = n(t) + h(t) \tag{2.1}$$

The $h(t)$ that is measured by the detector depends on the sky position of the source compared to the interferometer. As seen in the previous chapter, a gravitational wave has two polarizations, h_+ and $h_\times$. What one can measure with one detector is the quantity:

$$h = F_+(\theta, \phi, \psi)h_+ + F_\times(\theta, \phi, \psi)h_\times \tag{2.2}$$

where F_+ and $F_\times$ are the antenna patterns defined such as:

$$F_+(\theta, \phi, \psi) = \frac{1}{2}(1 + \cos^2(\theta))\cos(2\phi)\cos(2\psi) - \cos(\theta)\sin(2\phi)\sin(2\psi) \tag{2.3}$$

$$F_\times(\theta, \phi, \psi) = \frac{1}{2}(1 + \cos^2(\theta))\cos(2\phi)\sin(2\psi) + \cos(\theta)\sin(2\phi)\cos(2\psi) \tag{2.4}$$

where (θ, ϕ) are respectively two angles defined in Fig. 2.1, that can be linked to the declination and the right ascension angles. In other words, the position of the sources on the celestial sphere, ψ is the polarization angle, which is defined in Fig. 2.1.

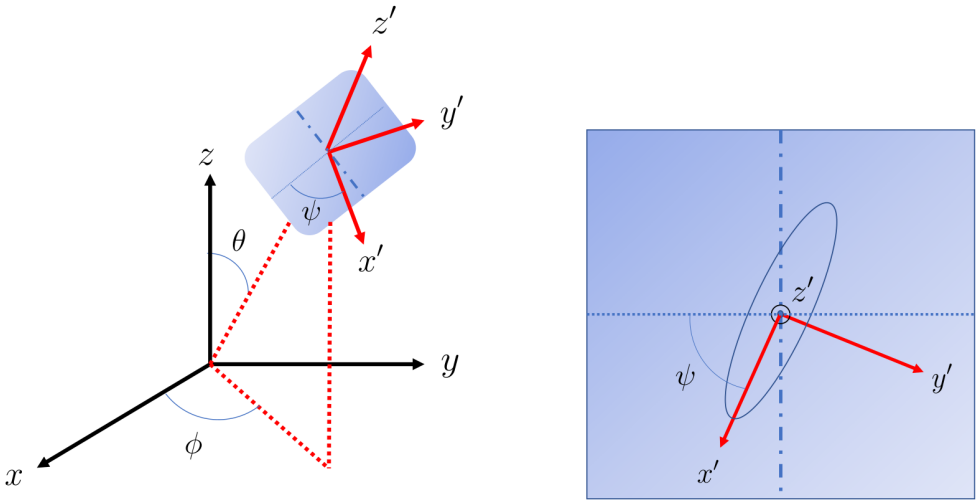

Figure 2.1: *Left*: The (x, y, z) coordinate basis of the detector, defined such that θ and ϕ represents the position of the sources in the celestial sphere. By a simple rotation, these two angles can be rewritten through the well-known declination and right ascension angles. The (x', y', z') is the coordinate system of the source: the z' axis is chosen to be on the line connecting the detector and the sources, while the x' one is along the major axis of the projection onto the sky of an orbit. *Right*: The plan tangent to the celestial sphere at the point of the center of an orbit. The ellipse represents the projection of an orbit onto this plan. The x' axis is along the major axis of the ellipse, and the z' one is perpendicular to the plane. The horizontal dotted line is parallel to the xy plane, and is used to define the polarization angle ψ.

2.1.2 The signal-to-noise ratio

This quantity h is several orders of magnitude smaller than the noise present in the LIGO and Virgo interferometers. Nevertheless, it is still possible to detect such a signal with some filtering methods. First of all, to give an intuition to the reader, let us say that one wants to filter $s(t)$ with a template h_{temp}. If one takes the integral from 0 to T of $s(t)h_{temp}(t)$ divided by T, one can make the GW signal stand out [4]:

$$\frac{1}{T}\int_0^T s(t)h_{temp}(t)dt = \frac{1}{T}\int_0^T (n(t)+h(t))h_{temp}(t)dt = \frac{1}{T}\int_0^T n(t)h_{temp}(t)dt + \frac{1}{T}\int_0^T h(t)h_{temp}(t)dt \tag{2.5}$$

If h_{temp} is the correct template, it means that $h(t) = h_{temp}(t)$, and the last term becomes $\frac{1}{T}\int_0^T h^2(t)dt$. $h(t)$ and $n(t)$ are oscillating functions. More precisely $n(t)$ can be seen as a random walk process. In this case, the first integral in the right-handed size of Equation (2.5) goes as $\sim (\frac{\tau_0}{T})^{1/2} n_0 h_0$, in which h_0 and n_0 are respectively the characteristic amplitude of h and n. τ_0

is a timescale associated with $h(t)$, which can be approximated as the period of the GW. On the other hand, the second integral of the right-handed side of Equation (2.5) depends only on the template, and this integral goes as $\sim h_0^2$. If $T \to \infty$ the first integral goes to zero, while the second stays constant at h_0^2. It is clear that only the first integral contributes if $s(t)$ contains only noise. Therefore, for $T \to \infty$ one has $\frac{1}{T}\int_0^T s(t)h(t)dt \to 0$. In the case in which a GW is present into the data, one obtains $\frac{1}{T}\int_0^T s(t)h(t)dt \to h_0^2$ for $T \to \infty$. With this simple Equation (2.5), one obtains, in theory, a way to distinguish between data containing only noise and data containing noise plus a GW. Nevertheless, it is important to be aware that, in reality, it is not possible to choose $T = \infty$: in reality, a GW signal has a finite length. However, this simple technique still works if one has:

$$h_0 \gg (\frac{\tau_0}{T})^{1/2} n_0 \tag{2.6}$$

In practice, for a BNS system, $T \sim 100$ s, the frequency f monotonically increases, so one can take τ_0 to be the period at 100 Hz, which means $\tau_0 \approx 0.01$ s. Thus, one has $(\frac{\tau_0}{T})^{1/2} \approx 10^{-2}$, which means that one can detect a signal h even if it is two order of magnitude smaller than the noise. Matched filtering techniques are based on a similar idea, nevertheless, in practice, matched filtering pipelines used more sophisticated methods than Equation (2.5). For example, these pipelines also take into account the sensitivity of the detector.

Let us first introduce the inner product, defined as:

$$\langle a|b\rangle = 4\int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}df \tag{2.7}$$

where $S_n(f)$ is the PSD defined in the previous chapter in Equation (1.23), the $*$, means that the complex conjugate is taken and the tilde over a and b signifies that those objects are in the Fourier domain, i.e.:

$$\tilde{a}(t) = \int_{-\infty}^{+\infty} a(t)e^{-2\pi i f t}dt \tag{2.8}$$

If the object a is real, it can be expressed by the Fourier components as:

$$a(t) = 2\int_0^{+\infty} \Re(\tilde{a}(t)e^{2\pi i f t})df. \tag{2.9}$$

From the inner product, Equation (2.7), one can define the signal-to-noise ratio as ρ [18, 19]:

$$\rho^2 = \frac{|\langle s|h_{temp}\rangle|^2}{\langle h_{temp}|h_{temp}\rangle} \tag{2.10}$$

the reason why this equation for the SNR is not simply $\langle s|h_{temp}\rangle$ is because this last object still depends on the phase at a reference time (i.e. the time when it enters the frequency band of the detector). One can show that it is possible to minimize the phase by taking the absolute value [20]. This means that with one template it is possible to search for every gravitational wave, whatever the phase at the time when it enters the frequency band of the interferometer. This is an advantage since you need to compute fewer templates to search for gravitational waves. Now that the reason

for the absolute value is understood, it is necessary to explain why the denominator appears in this definition of the SNR. The quantity $\langle s|h_{temp}\rangle$ also depends on the amplitude of the template h_{temp}, and this amplitude depends mainly on the chirp mass of the system but also on its sky position and the distance of the source. If you imagine that you have an event, and then the same one but at a further distance, the signal will be exactly the same except for the amplitude. The denominator $\langle h_{temp}|h_{temp}\rangle$ is there to make the SNR independent of the amplitude, meaning that with one template, it is possible to search for a typical signal whatever the distance of this signal. From expression (2.10), one can define the SNR $\rho(t)$ time series:

$$\rho^2(t) = \frac{4}{\langle h_{temp}|h_{temp}\rangle}\int_0^\infty \frac{\tilde{s}^*(f)\tilde{h}_{temp}(f)}{S_n(f)}e^{2\pi i f t}df \tag{2.11}$$

The basic idea of matched filtering techniques is then to compute the SNR over all the data and look for some peaks in this SNR time series.

As seen in the previous chapter, the shape of a signal is mainly driven by the masses of the system. Some other parameters as the spin and the tidal deformability also modify the behavior of the signal to a lesser extent. Computing the SNR of Equation (2.11) allows us to search with a single template for any corresponding signal, whatever the distance of the sources and the phase are. Yet, in order to search for any GW whatever the other parameters, masses, spins, etc., it is required to compute all the corresponding templates, and for each of them, compute Equation (2.11). This is what a matched filtering pipeline is. First of all, a template bank is constructed in such a way that the templates generated cover properly all the desired parameter space. Since the masses and the other parameters are continuous, it is not possible to produce a template for each set of parameters. In order to cover the parameter space, the templates are generated in such a way that the difference of SNR given by the match of a template by one of its adjacent ones and the SNR compared to the match of the template by itself is smaller than a typical threshold (typically 3%). [19]. In Fig. 2.2, one can see that the template bank is more densely populated at low masses. This is easily understandable, since a lighter object will have a longer signal, which leads to more oscillation and to a higher mismatch than for a heavier object, as shown in the frequency evolution Equation (1.39).

2.1.3 The ξ^2 statistics

In real noise, the data contain glitches and artifacts can mimic a gravitational wave and could potentially result in a peak of high value in the SNR time series, just as a GW would. To distinguish between real GW events and these glitches, standard searches, like gstlal [21], use the ξ^2 statistic. The idea is to compare the peak of SNR in the real data with the auto-correlation of the template $R(t)$. In the absence of noise, if $R(t)$ is rescaled to the peak SNR, those two objects must match exactly. The ξ^2 statistic is computed on a δt window around the peak (assumed to be at $t = 0$):

$$\xi^2 = \frac{\int_{-\delta t}^{+\delta t}|z(t) - z(t_0)R(t)|^2 dt}{\int_{-\delta t}^{+\delta t}(2 - 2|R(t)|^2)dt} \tag{2.12}$$

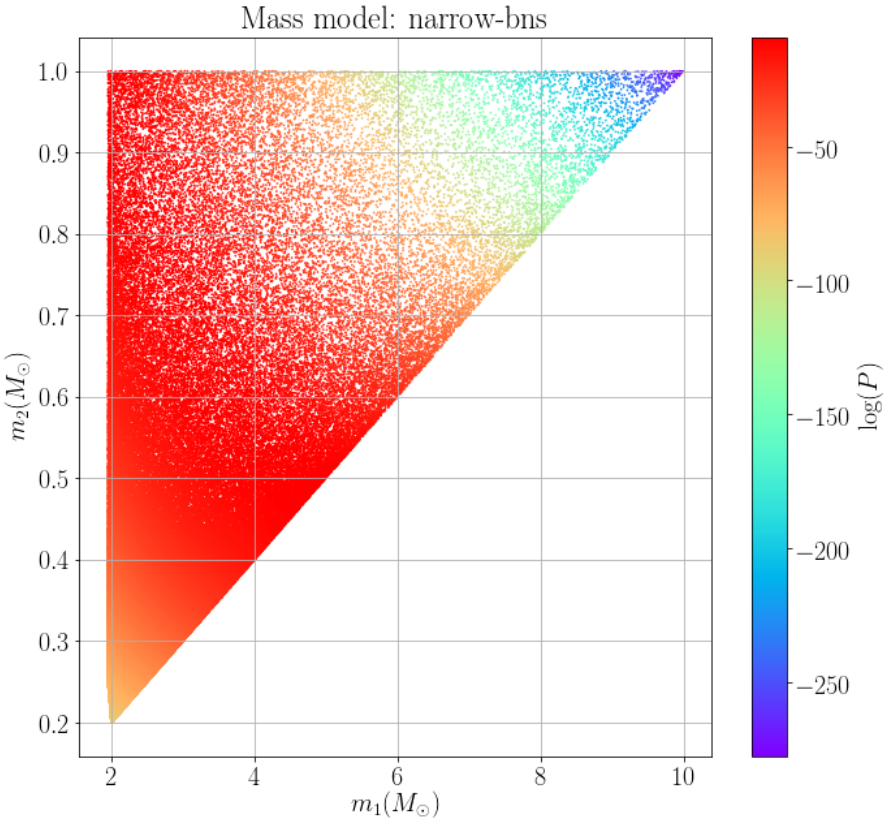

Figure 2.2: The template bank that was used in our paper to search for sub-solar objects. We focus on asymmetric binaries, where only one of the components has a sub-solar mass. The mass ratio ($q = m_2/m_1$) is fixed to be over 0.1, which explain why only the top left triangle is populated with templates. The minimal match is 0.97.

To constitute a trigger, the SNR must be higher than a certain threshold, and the ξ^2 must be smaller than another threshold.

2.1.4 The false alarm rate

Another way to discard glitches from proper GWs is to use the coincidence trigger [18, 21]. If a gravitational wave passes through the earth, it should go through all the detectors, unlike the glitches, which are linked to a detector. Therefore, all the triggers that appear only in one detector can already be discarded as potential GW candidates. Only the others, called *coincidence triggers*, appear at least in two detectors in a time frame smaller than the one that a GW will take to travel between those interferometers. To be considered as a potential candidate, each peak of SNR in the different detectors that constitute the coincident trigger must result from the match with the same (or a very similar) template. When this is the case, they are called coincident events. It is also useful to define a single SNR for all the detectors, called the network SNR:

$$\rho_{net}(t) = \sqrt{\sum_{i}^{N} \rho_i^2(t)} \tag{2.13}$$

while the index i represents the different detectors and N is the total number of detectors (three at the time when we are writing this thesis).

Even a coincident event is not enough to claim the detection of a GW. Indeed, it also happens that glitches occur at the same time in multiple detectors. To distinguish between the two types of coincident triggers, one can build a score of confidence, which is called the False Alarm Rate (FAR). This score represents the number of times that such a coincident event appears, because of noise fluctuation and glitches for a given period of time (usually expressed in years). To define properly the FAR, it is mandatory to introduce the probability distribution $p(\rho_{net})$ that such a trigger could appear in the data which contains only noise. Let us explain how one can estimate this probability $p(\rho_{net})$: we call *background triggers* all the coincident events that are due to noise fluctuation or glitches, and *foreground triggers* are those which are actually due to a gravitational wave. Here, the purpose is to distinguish between these two types of coincident events. Even though it is not possible to remove the background triggers, it is yet possible to characterize them, and especially their distribution $p(\rho_{net})$. Unlike the background triggers, it is possible to make the foreground triggers disappear. To do so, let us imagine that one has only two detectors and one can time-shift one with respect to the other. For example, let us say that the new output $s'(t)$ in one of the detectors is:

$$s'(t) = s(t + \Delta t_{shift}) \tag{2.14}$$

where $s(t)$ is the original output in this detector and where Δt_{shift} is a chosen time which is longer than the one needed by a GW to travel between the two interferometers. Thus, all the coincident events could not be due to a GW, so that they are all background triggers. It is also possible to follow the same principle with three detectors by choosing two different Δt_{shift}. This technique allows us to generate a lot of data only with background triggers. When one has generated X years of this type of data, called *background data* (unlike the data used for the search, which is called *foreground data*), one can estimate the probability distribution $p(\rho_{net})$ by looking at the number of coincident events which have a typical SNR value. Fig. 2.3 represents the estimation of $p(\rho_{net})$. If one obtains a trigger in the foreground data with an SNR value of Y, one can make the integration of the $p(\rho_{net})$ over the SNR value of Y estimated in the background data, see Fig. 2.3. This is called the FAR, and it can be interpreted as the number of triggers due to noise fluctuation with an SNR higher than Y for X years. It is usually this specific statistic which is used to distinguish between a trigger due to noise and a trigger due to a GW.

In reality, the matched filtering pipelines like gstLAL [21] used a likelihood-ratio ranking statistic to rank all the coincident triggers., and the FAR is computed with this likelihood-ratio instead of the SNR. A likelihood ratio can be expressed as [22]:

$$\mathcal{L} = \frac{P(\overrightarrow{D}_H, \overrightarrow{O}, \overrightarrow{\rho}, \overrightarrow{\xi^2}, \overrightarrow{\phi}, \overrightarrow{t}|s)}{P(\overrightarrow{D}_H, \overrightarrow{O}, \overrightarrow{\rho}, \overrightarrow{\xi^2}, \overrightarrow{\phi}, \overrightarrow{t}|n)} \tag{2.15}$$

the arrow over the different letters represents a vector in which each component belongs to one

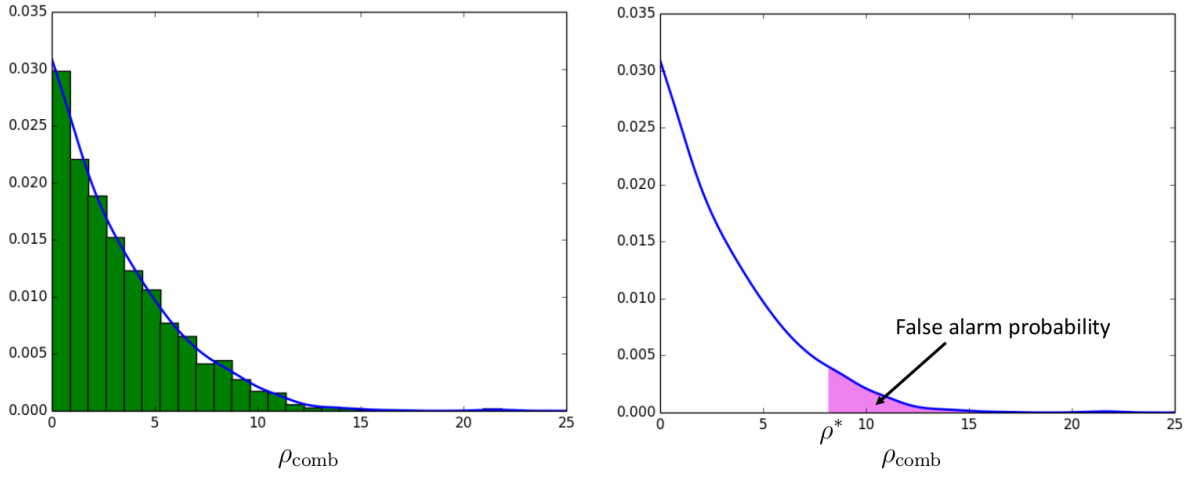

Figure 2.3: *Left:* The detection probability for the network SNR ρ_{net} estimate with the normalized histogram of the background noise. *Right:* The fit of the histogram and the definition of the FAR, as the integral of this curve, with the minimal bound being the observed SNR. This figure comes from [4].

detector. $\overrightarrow{D}_H$, represents the distance horizon, i.e. the sensitivity for each detector, $\overrightarrow{O}$ represents the set of detectors available at the time of the coincident trigger, $\overrightarrow{\rho}$ represents the SNR, $\overrightarrow{\xi^2}$ the ξ^2 veto, $\overrightarrow{\phi}$ is the estimated phase, $\overrightarrow{t}$ is the estimated time of the merger in each detector. The denominator in Equation (2.15) is the probability that all the characteristics mentioned just above occur if pure noise is assumed. The numerator is in the same probability, but if one assumes that an event is present in the data. In this thesis, we will not describe how this likelihood ratio is computed; we refer the reader to [22] for more information. What is important to know is that the FAR can be defined with the likelihood ratio instead of just being defined with the SNR. In fact, in most of the searches, the FAR is computed with this ranking statistic.

2.2 Sub-solar objects and primordial black holes

To illustrate the standard matched filtering techniques, we present a paper in which we used this method to search for CBC signals coming from sub-solar compact objects. Before introducing the paper, let us first discuss the motivation behind this type of search, and also explain what the sub-solar object that we are looking for are.

2.2.1 Primordial balck holes

As mentioned earlier, with the current detectors, it is expected to find GW coming from CBC systems in which the objects are dense enough, i.e. neutron stars and black holes. When a supergiant star with a mass between 10 and 25 $M_\odot$ comes to the end of its life, this one will collapse into a neutron star. This process is quite well understood and predicts that the mass of the neutron star

should be roughly between 1 and 2.3 $M_\odot$ [23, 24, 25, 26]. Except for black holes (or hypothetical exotic objects), the neutron stars are the densest objects that exist in our universe. On the other hand, it is known that black holes can also be produced in a similar process, with the exception that the supergiant star must be heavier. Once again, there exists a limit on the minimal mass that a black hole formed with such a scenario could take, which is around 4 $M_\odot$ [27, 28, 29]. A priori, black holes or neutron stars that are lighter than one solar mass should not exist. One can wonder: what is the interest in searching for CBC objects which are lighter than one solar mass, (called sub-solar objects). There are some theories where some special black holes can have sub-solar masses, for example, the PBH [30]. Even though we will focus on this specific one, let us mention that some other theories, such as the accretion of dark matter into a neutron star [31], allow the existence of sub-solar black holes. The major interest in searching for sub-solar objects is to confirm or put some constraints on these theories.

PBHs are black holes that could have been created by the collapse of large over-densities in the very early universe. At that stage, the universe was very dense, and some quantum fluctuation led to some variations of density. This phenomenon is the origin of the anisotropies in the Cosmic Microwave Background (CMB). In theory, if some fluctuations are high enough, it is possible that the over-density collapses into PBH. In the simplest model, the condition for an over-density $\delta\rho$ to become a PBH is to have the length of the instability which is larger than what the so-called Jeans length defines as:

$$L_j = \sqrt{\frac{\pi c_s^2}{G\rho}} \tag{2.16}$$

where c_s is the speed of sound of pressure disturbances, ρ is the density, and G is the gravitational constant. The Jeans length is the critical oscillation wavelength at which a small perturbation to a self-gravitating medium of uniform density becomes unstable to gravitational collapse [32, 33].

PBHs are supposed to be nearly collisionless, stable, in the universe since its earliest stage, which makes the PBH a good candidate for dark matter. Several studies such as microlensing have put some constraints on the fraction of dark matter that can be made of PBH [34], as we will make it explicit it in the paper, that we present here. However, those constrains are often model-dependent and do not take into account the newest mass distribution of the PBH. If one does so, most of the constraints can be avoided [35].

Unlike stellar black holes, the range of allowable masses for the PBHs is way wider. Even though the exact mass distribution depends on the theory used to model the early universe, most of these theories predict PBHs with a broad mass range. That makes the PBH natural candidates to be the seed for the intermediate-mass and supermassive black holes [36]. It has been shown that if one considers the equation of state for the QCD phase transition, it affects the mass distribution of the PBHs [37]. Fig. 2.4 shows that most of the PBHs must have a sub-solar mass. Yet, a second peak is also present around the value of 20 solar masses, which coincides in a way with the mass distribution of the black holes found by the LIGO and Virgo. One can hypothesize that some of these black holes have a primordial origin [30].

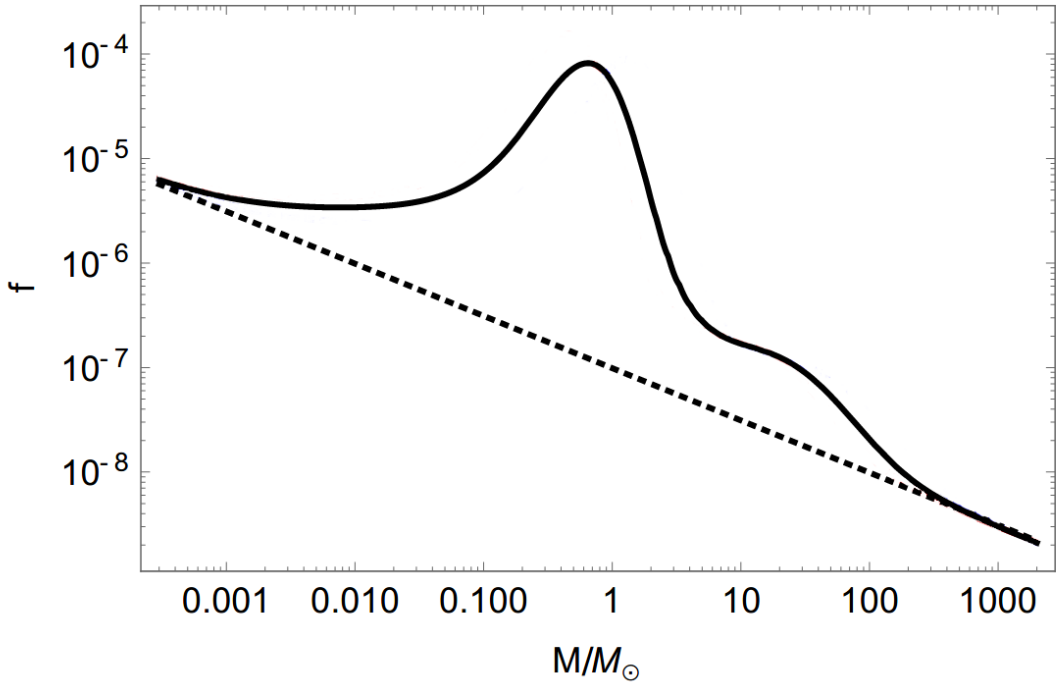

Figure 2.4: The mass distribution of the PBH formed during the phase transition is shown for a scale-invariant density power spectrum. The dashed line represents the mass distribution if there is no phase transition. This figure comes from [37], for more information about the different curves, we refer the reader to this article.

If the mass distribution of PBHs in Fig. 2.4 is correct, most of the PBHs should be sub-solar. There is a major interest to search for sub-solar objects. Let us first mention that the searches [7, 8, 38] do not take into account the sub-solar region because there is no way to form black holes and neutron stars of these masses by non-speculative astrophysical processes. As mentioned in section 2.1, the lighter the masses are, the longer the waveform is, which implies that more templates have to be used to cover correctly all the mass space. That also means that more computational power is needed to perform the matched filtering. It is also another reason why in [7, 8, 38] they do not include a search for sub-solar objects. Nevertheless, this type of search has been made for the first and second runs of observation [39, 40, 41].

2.2.2 Previous sub-solar searches

The paper that we are presenting is a follow-up to the previous searches for sub-solar objects [39, 40]. We will summarize quickly those two papers and the major results before presenting our work. The first sub-solar search [39] focuses on the first run of observation O1. They used a template bank of 500 332 template waveforms which was constructed to cover a mass range going from 0.19 to 2.0 $M_\odot$ with 97% fidelity. They also did not consider spin in the bank. Even though it is not totally realistic, this is assumed to reduce the computational cost of such a search. According

to Equation (1.39) the duration of a CBC signal of component masses of 0.19 $M_\odot$ with a minimum frequency of 20 Hz has a duration of almost 80 minutes. Creating a template bank with such a long signal and using it in the context of matched filtering searches is too computationally expensive. It is the reason why the author of [39] chose a maximum frequency of 45 Hz, which leads to a duration of around 500 s for a system with two objects of 0.19 $M_\odot$. Using this template bank, no viable gravitational wave candidate has been found in O1. Nevertheless, with these results, they were able to put a constraint on the binary merger rate. To do so, they used the loudest event statistics [19, 42]. Let us quickly describe how this method works.

If a search does not find any event, it is possible to put a limit on the rate at which these events occur. It is common to use the loudest event statistic to do so. Let us say that ρ^* is the SNR threshold at which a trigger is recorded, and $\mathcal{R}$ is the mean rate at which an event arrived at the set of detectors with a value of SNR $\rho > \rho^*$. One can model the arrival of a signal as a Poisson process. The probability of detecting an event with $\rho > \rho^*$ within time T can be written as:

$$P(\rho > \rho^* | \mathcal{R}) = 1 - e^{-\mathcal{R}T\varepsilon(\rho^*)} \tag{2.17}$$

where $\varepsilon(\rho^*)$ is the detection efficiency, meaning the ratio between detected signals and incident signals at the threshold ρ^*. This formula is not accurate, because it does not take into account the fact that there are background triggers. Let us denote by P_B the probability to observe no trigger with a higher SNRρ than ρ^*. In that case, one can rewrite Equation (2.17) such as:

$$P(\rho > \rho^* | \mathcal{R}, B) = 1 - P_B e^{-\mathcal{R}T\varepsilon(\rho^*)} \tag{2.18}$$

One can use this equation to set up an upper limit on the rate. For example, one can set up a rate $\mathcal{R}_{90\%}$. This corresponds to the fact that, if one assumes such a rate, there is a 90% chance that an event with a higher SNR value than ρ^* occurs during the time of the search. One can compute this rate using Equation (2.18). In that case, one obtains:

$$\mathcal{R}_{90\%} = \frac{2.303 + \ln(P_B)}{T\varepsilon(\rho_{max})} \tag{2.19}$$

Note that, in this equation, P_B is the probability that no event in the background has a SNR higher than ρ_{max}. To compute this quantity, one needs to know P_B and $\varepsilon(\rho_{max})$. It is possible to evaluate the second one using a Monte Carlo simulation, by injecting some BNS signals into the data and seeing how many are possible to recover. Estimating P_B is more tricky. One can argue that the background triggers recorded in the background data (the data created by time shift, as explained in section 2.1) can be used. Nevertheless, these triggers are known to have big variations depending on the details of the search. Therefore, it is very difficult to quantify them accurately. A standard method is then to set P_B to 1. In this case, $\mathcal{R}_{90\%}$ is maximized with respect to P_B and the estimation of the limit is then conservative. In addition to that, given the definition of P_B and the fact that ρ_{max} is the loudest event observed, it is statistically unlikely that the real $\mathcal{R}_{90\%}$ could be much under the one used with the approximation $P_B = 1$.

Let us go back to the description of [39]. With this loudest event statistic method, they set up a rate limit for the sub-solar event, which is illustrated in Fig. 2.5. Note that the rate is given per Gpc^3 per year. They finally put some constraints on the fraction of dark matter that sub-solar PBHs can provide, and this is illustrated in Fig. 2.6. Note that this constraint depends on the mass distribution of the PBHs. In their paper, they assumed a uniform mass distribution for the PBHs. In our paper, it is shown that using a different mass distribution leads to other constraints.

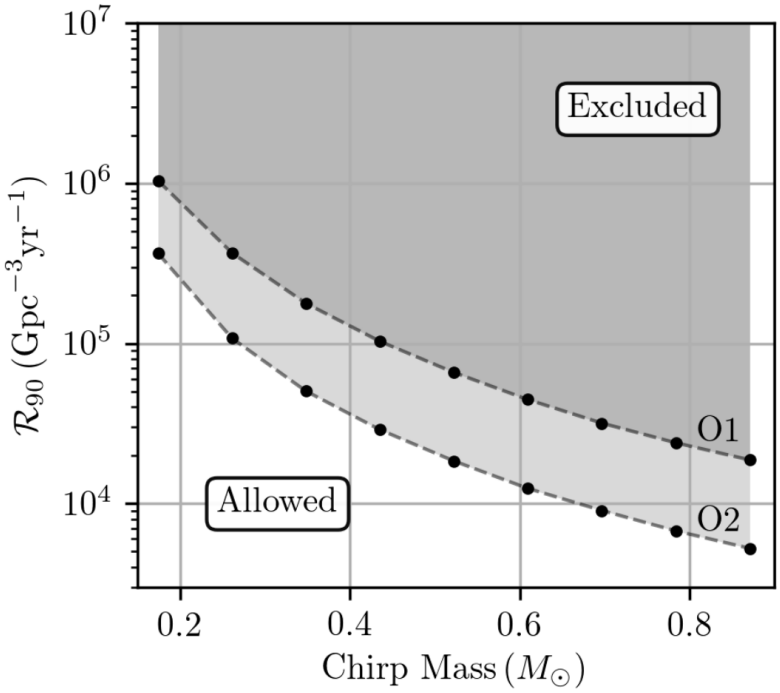

Figure 2.5: The constraints on the merger rate for sub-solar binaries as a function of the chirp mass. The curves are made for equal mass binaries. This graph comes from [40]

The second paper [40] is very similar to the previous one, but it has two major changes. First, they search in the data of the second run of observation O2. Then, they use another more complete template bank, in which they include the spin effect: the authors consider templates in which the objects have aligned or anti-aligned spins with a magnitude of 0.1 or less. This inclusion leads to a bigger template bank of 992 461 waveforms. Once again, no event has been found during this search. As in the first paper, constraints on the rate of the sub-solar merger, and the fraction of dark matter that they can contribute to were computed, see Fig. 2.5. There are now some papers that have searched for sub-solar in O3 [41], but as our search focused on O2, we will not discuss them in this thesis and leave them for the interested readers.

Now, we will present our search for sub-solar black holes. This search is an extension of [40], in which the data analyzed is also the one of O2, but the template bank is different. In this search, we consider asymmetric binaries in which one of the components is a sub-solar object and the other

one is a more massive object. Since no significant event is found, we also compute a rate limit for this type of CBC event, and for the fraction of dark matter that PBHs can contribute to. Unlike [40], we used a non-uniform PBH mass distribution coming from [37] and we also considered other methods for the formation of PBH binaries.

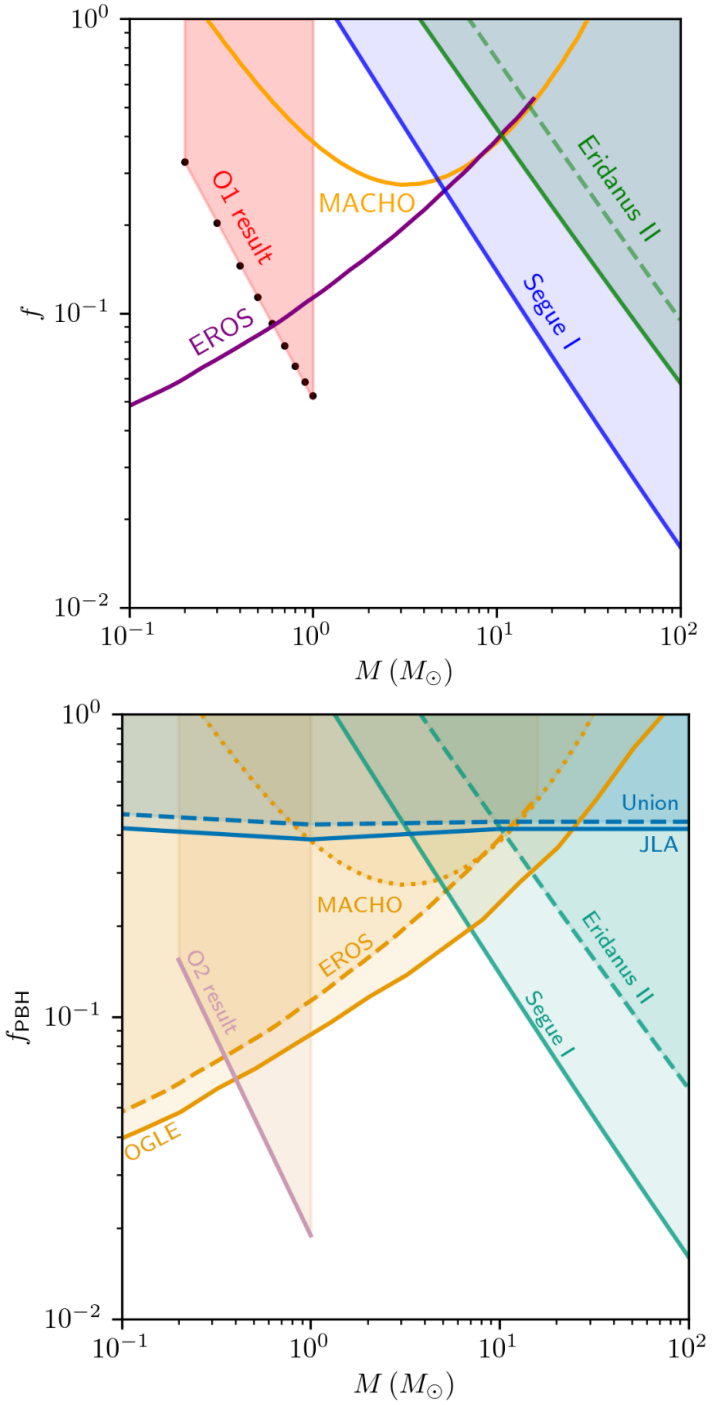

Figure 2.6: *Left*: Constraints of the first run of observation O1 on the dark matter that PBH can contribute as a function of the masses if a monochromatic distribution is assumed. This graph comes from [39] *Right*: Same as the left one, but for the second run of observation O2, this graph comes from [40].

2.3 Article 1

The hunt for sub-solar primordial black holes in low mass ratio binaries is open

Khun Sang Phukon, Gregory Baltus, Sarah Caudill, Sebastien Clesse, Antoine Depasse, Maxime Fays, Heather Fong, Shasvath J. Kapadia, Ryan Magee, Andres Jorge Tanasijczuk

Abstract

We perform a search for binary black hole mergers with one subsolar mass black hole and a primary component above $2M_\odot$ in the second observing run of LIGO/Virgo. Our analysis therefore extends previous searches into a mass region motivated by the presence of a peak in any broad mass distribution of primordial black holes (PBHs) around $[2-3]M_\odot$ coming from the equation of state reduction at the QCD transition. Four candidate events are found passing a SNR>8 and a false alarm rate (FAR) threshold of 2 per year, although none are statistically significant enough for being clear detections. We first derive model independent limits on the PBH merging rates assuming a null result of the search. Then we compare them to recent scenarios in which PBHs can constitute up to the totality of the Dark Matter, explain LIGO/Virgo mergers and the possible observation of a stochastic gravitational-wave background by NANOGrav. We find that these models still pass the rate limits and conclude that the analysis of the O3 and O4 observing runs will be decisive to test the hypothesis of a primordial origin of black hole mergers.

2.3.1 Introduction

The first detection of a gravitational wave (GW) event (GW150914) [1] has ushered astronomy into a GW era. Since then more than forty binary black hole (BBH) mergers [7, 43, 44, 45, 46, 47, 48, 49, 38] have been observed by Advanced LIGO [6] and Advanced Virgo [50], revealing some intriguing and mostly unexpected properties of BBHs: large masses, low spins, the existence of black holes (BHs) in the low mass gap [44] or in the pair-instability mass gap [43] and with low mass ratios [44]. Elucidating the origin of LIGO-Virgo BHs has emerged as an important research topic. One of the most attracting possibilities is that some are primordial black holes (PBHs) [30, 35, 51].

PBHs may have formed in the early Universe due to the gravitational collapse of large overdensities [52, 53, 54, 55], e.g. coming from inflation [56, 57, 58, 59], and could explain from a fraction to the totality of the dark matter (DM). There exist a series of astrophysical and cosmological limits on their abundance, covering almost all possible masses (see e.g.[60, 61] for reviews on PBHs and [34, 62, 63] for recent developments), as well as possible clues in observations [64, 65]. Those limits and observations are however model dependent, some are debated and today the status of PBHs to explain all the DM and GW observations is still controversial.

Contrary to stellar BHs, there is no physical process preventing the formation of PBHs lighter than the Chandrasekhar mass [66] or in the pair-instability mass gap [67, 68]. The subsolar and intermediate mass ranges are thus ideal targets to distinguish between stellar and primordial origins. If some of the observed BH mergers are primordial, one can expect a relatively broad mass distribution of PBHs spreading within these interesting regions. In such a case, the known thermal history of the Universe at the QCD epoch and the resulting transient reduction of the equation-of-state unavoidably leads to features in the PBH mass distribution [37, 69, 70], in particular a peak around the solar-mass and a bump around $30M_\odot$ [37]. They could explain the LIGO/Virgo merging rates [65, 71], a series of OGLE microlensing events due to BHs the low mass gap [72], boost the clustering of DM at high redshift and thereby explain unexpected correlations in the infrared and X-ray backgrounds [73] and evade the microlensing limits [65]. All this provides strong theoretical and observational motivations to extend previous subsolar searches in the first [39] and second [40] observing runs of LIGO/Virgo (referred as O1 and O2) and to search for binaries combining a subsolar black hole and a primary component above $2M_\odot$, eventually coming from the QCD peak.

In this *letter*, we perform an extended search on O2 data, assuming one subsolar component between $0.19\mathrm{M}_\odot$ and $1M_\odot$ and a primary component between 1.95 and $11\mathrm{M}_\odot$. Our analysis is therefore complementary to previous searches, either restricted to $2\mathrm{M}_\odot$ for the primary component [39, 40] or considering even lower mass ratios [74]. We derive new model-independent merger rate limits in the (m_1, m_2) plane of the two component masses[1]. Furthermore, whereas previous limits on $f_{\rm PBH}$ – defined as the PBH abundance with respect to DM – relied on the merging rates of early PBH binaries from [51], we include the rate suppression due to binary disruption by early forming clusters, matter inhomogeneities and nearby PBHs, put in evidence with N-body simulations [76]. We consider in addition the merging rate of late PBH binaries formed in clusters. For

[1] When our work was almost completed, an independent search has been released with a similar mass range [75]. Besides providing an independent confirmation of their results, our analysis also reveals a few candidate events and includes new limits for some motivated realisations of the PBH mass function.

Table 2.1: The candidates of the search with a SNR > 8 and a FAR $< 2\,\mathrm{yr}^{-1}$. Here we report the FAR, $\ln \mathcal{L}$, the UTC time of the event (date and hours), and recovered template parameters.

FAR [yr^{-1}]	$\ln \mathcal{L}$	UTC time	mass 1 [$M_\odot$]	mass 2 [$M_\odot$]	spin1z	spin2z	Network SNR	H1 SNR	L1 SNR
0.1674	8.457	2017-03-15 15:51:30	3.062	0.9281	0.08254	-0.09841	8.527	8.527	-
0.2193	8.2	2017-07-10 17:52:43	2.106	0.2759	0.08703	0.0753	8.157	-	8.157
0.4134	7.585	2017-04-01 01:43:34	4.897	0.7795	-0.05488	-0.04856	8.672	6.319	5.939
1.2148	6.589	2017-03-08 07:07:18	2.257	0.6997	-0.03655	-0.04473	8.535	6.321	5.736

the PBH mass distribution, we include thermal features and consider scenarios based on a (nearly) scale invariant primordial power spectrum. We argue that the derived limits are relatively conservative and should apply to most other broad mass models. We conclude that searches in the third and fourth observing runs of LIGO/Virgo will be decisive to test the hypothesis of a primordial origin of BBH mergers.

2.3.2 Search, candidates and rate limits

We analyze O2 public data from single and coincident observation time of the two LIGO detectors, comprising ∼193 days of analysis time. This search uses the GstLAL-based inspiral pipeline [21, 22, 77, 78, 79, 80] with configurations and procedures as outlined in [7]. Data from each detector are matched-filtered [20, 78, 81, 82, 83, 84] with an (anti-)aligned low-spin template bank of 195468 template waveforms modelled with the frequency-domain waveform-approximant, TaylorF2 [85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95]. Components have primary mass $m_1 \in [1.95, 11.0]\mathrm{M}_\odot$, subsolar mass $m_2 \in [0.19, 1.0]\mathrm{M}_\odot$ and spin magnitude $\chi_{i,z} \in [0, 0.1]$. The total masses ($\mathrm{M}_{\mathrm{tot}}$) and mass ratios ($q = m_2/m_1$) of the search are limited between $\mathrm{M}_{\mathrm{tot}} \in [2.2 - 11.0]\,\mathrm{M}_\odot$ and $q \geq 0.1$, respectively. The template bank construction uses the stochastic method of Refs. [96, 97, 98, 99] with a minimal match of 0.97. Following Ref. [100], we perform matched-filtering of the data in the frequency band 45-1024 Hz in order to reduce the computational burden of filtering data.

Candidates are ranked using a likelihood-ratio $\mathcal{L}$ ranking statistic [22, 79, 101]. Here we note, in contrast to Ref. [7], our pipeline does not include data-quality penalties to rank single-detector candidates [102, 103]. This ranking statistic also provides an algebraic procedure to estimate the significance of each event [79, 104]. The significance of an event is expressed in terms of a false alarm rate (FAR), e.g. defined in Section IV.C of Ref. [79].

Search results are presented in Fig. 2.7. Table 2.1 reports four candidates with a signal-to-noise ratio SNR$>$ 8 and a FAR$<$ $2\,\mathrm{yr}^{-1}$. Time-frequency representations of the whitened data around each candidate are provided in the Appendix for different resolutions and do not reveal any clear GW signature. The first two candidates are single detector triggers, implying their lower FAR values may have more uncertainty. The remaining candidates have high FARs considering the number of days analyzed and in comparison to GW events reported in [7, 38]. With these considerations, it is not clear if any of the four candidates are real GW events; future parameter estimation investigations will explore this further. For the rest of the paper, we assume a null result of the search.

Following methods in Refs. [105, 106], we place upper limits on the rate of mergers of binaries with a sub-solar mass component for the discrete mass pairs shown in Fig. 2.9. The merging rate limits are presented in Fig. 2.9 and Tab. 2.2.

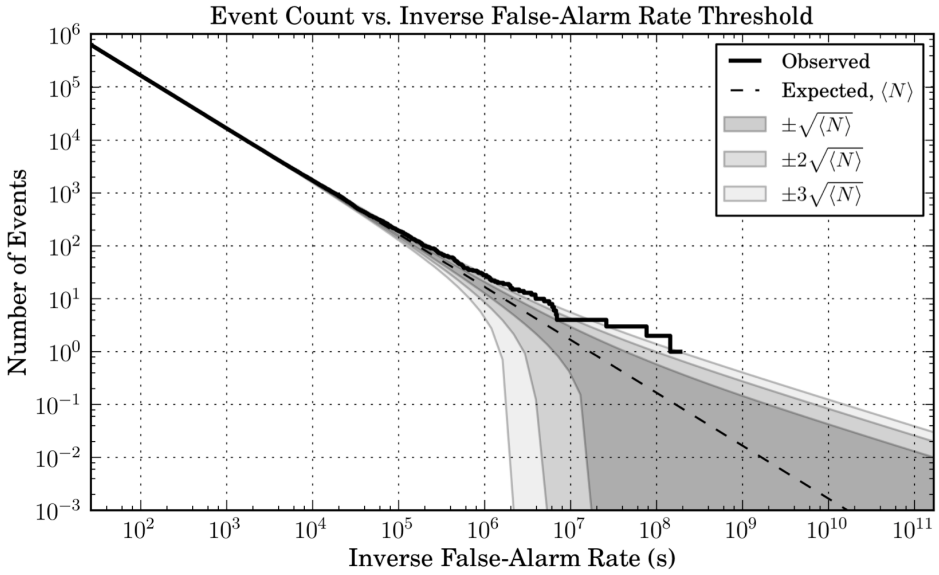

Figure 2.7: Results of the extended sub-solar PBH search in O2, in terms of number of events as a function of their inverse FAR. The dashed line is the expected distribution of background triggers, with the gray bands indicating uncertainties in multiples of the standard deviation for a Poisson distribution. The four candidates reported in Table 2.1 lie slightly above the 3σ limit.

2.3.3 PBH mass and rate distribution

The recent wide mass distribution proposed by Carr el al in [65] has been considered in our analysis (referred as mass model 1). It includes the expected features from the QCD transition and relies on an almost scale invariant primordial power spectrum of density fluctuations, enhanced compared to cosmological scales, with a spectral index $n_{\rm s} = 0.97$ (values of $n_{\rm s}$ between 0.96 and 0.98 are realistic). The PBH peak is located at the solar-mass scale and there is no overproduction of PBHs at smaller and larger masses. Even if $f_{\rm PBH} = 1$ the model can evade the microlensing limits because of PBHs clusters seeded by the Poisson fluctuations in the PBH distrubtion at formation [71]. Assuming an almost scale invariant primordial power spectrum allows us to set relatively conservative limits, because peaked spectra or with a negative running of the spectral index typically lead a lower number density of PBHs both at high and low mass. For comparison, we have also considered the model proposed by De Luca et al in [107] (referred as mass model 2), with $n_{\rm s} = 1$ and a cut-off mass around $10^{-14} M_\odot$. It produces a DM fraction made of stellar-mass PBHs of order 10^{-4}. But the merging rates are much lower than the limits obtained from this subsolar search and we do not set any significant limit on this model. The two density distributions $f(m_{\rm PBH})$ (normalized such that $\int f(m) {\rm d} \ln m = 1$) are represented on Fig. 2.8 for two plausible values of the ratio between the PBH and the Hubble horizon masses at formation, $\gamma = 0.8$ and $\gamma = 0.2$. The first value leads to a peak around $2.5 M_\odot$ motivated by GW190425 [108] and GW190814 [44]. The second one is obtained by considering the turnaround scale [37] in the PBH gravitational collapse. These models are further motivated because of the stochastic GW background generated at second order by the overdensities at the origin of PBHs [109, 110, 111, 112, 113], coinciding with the possible

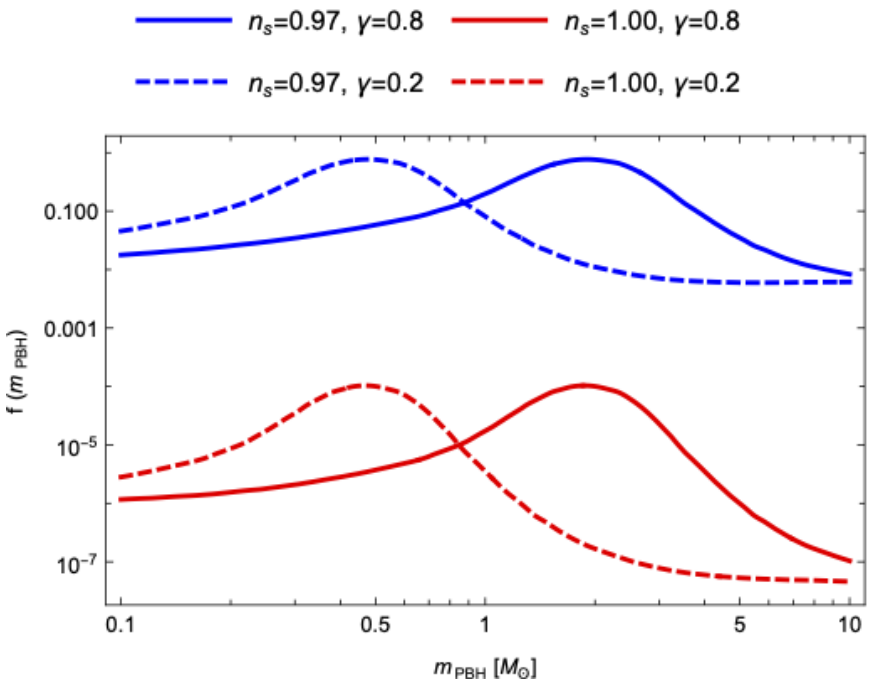

Figure 2.8: PBH mass functions $f(m_{\rm PBH})$ for the mass models 1 (blue lines) and 2 (red lines) with $\gamma = 0.8$ (solid lines) and 0.2 (dashed lines).

NANOGrav observation [114]. PBH binaries can form through two channels whose associated merging rates are briefly introduced below.

Merging rate of Early Binaries (EB). PBH binaries are formed in the early Universe, between the PBH formation time and the matter-radiation equality. A fraction of them will merge nowadays with a merging rate distribution (per unit logarithmic mass of the two black holes) given by [71, 76, 115, 116, 117, 118, 119]

$$R_{\rm EB} = \frac{1.6\times10^6}{\rm Gpc^3 yr} f_{\rm sup} f_{\rm PBH}^{34/37} \left(\frac{m_1+m_2}{M_\odot}\right)^{-32/37} \times \left[\frac{m_1 m_2}{(m_1+m_2)^2}\right]^{-34/37} f(m_1) f(m_2) \ ,$$

where m_1 and m_2 are the two binary component masses. N-body simulations have shown that the original merging rates from [51] are somehow suppressed [76], which motivates the introduction of a suppression factor $f_{\rm sup}$ plausibly ranging between 10^{-3} and 0.1. Our assumption for $f_{\rm sup}$ in relation with previous works of [76, 119] is described in Appendix. We did not consider the merging rate of disrupted binaries, estimated in [120] for a monochromatic mass function and which may surpass the EB rate if $f_{\rm PBH}$ is close to one, because it is unclear how to extend these estimations for wide mass distributions like the ones we consider here.

Merging rate of Late Binaries (LB). PBH binaries also form by tidal capture in PBH clusters, which gives the following rate distribution [71],

$$R_{\rm LB} \equiv R_{\rm clust.} f_{\rm PBH} f(m_1) f(m_2) \frac{(m_1+m_2)^{10/7}}{(m_1 m_2)^{5/7}}, \tag{2.20}$$

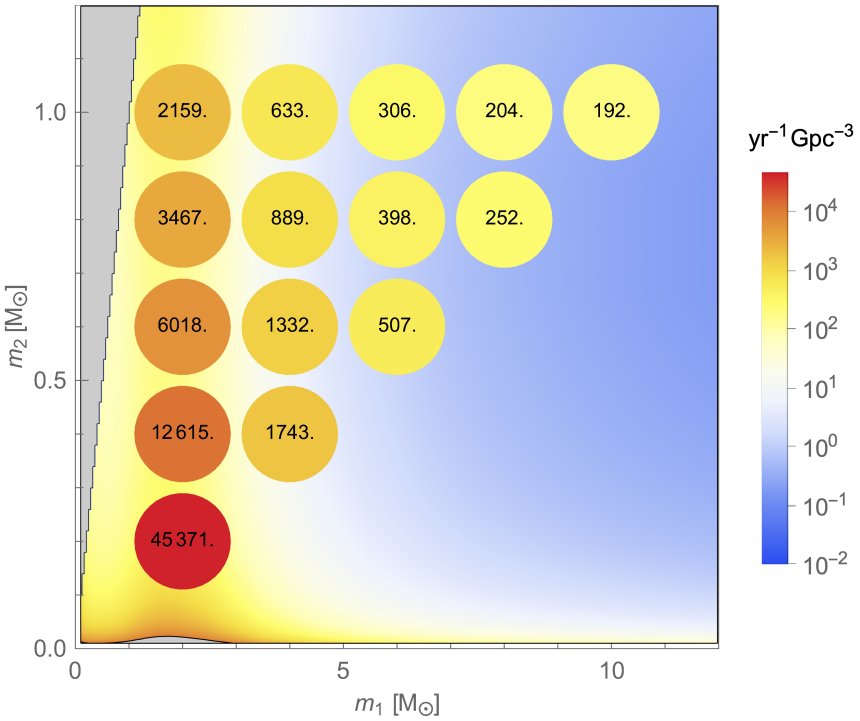

Figure 2.9: The disk color and numbers inside represent the merger rate limits at 90% C.L, for component masses between 2 and 10 $M_\odot$ for the primary component and between 0.2 and 1 $M_\odot$ for the secondary component, with mass ratios $q > 0.1$, in respective bin sizes of 2 and 0.2 $M_\odot$. The color scale behind represents the predictions of the EB merging rates for our mass model 1, assuming $n_s = 0.97$, $\gamma = 0.8$ and $f_{\rm PBH} = 1$.

where $R_{\rm clust.}$ is an effective scaling factor that incorporates the PBH clustering properties. We consider $R_{\rm clust} \approx 420\,{\rm yr}^{-1}{\rm Gpc}^{-3}$ as a benchmark, following [71], and comprised within $[80 - 1770]$[2] for our estimations of uncertainty bands. This value is larger than one may expect from standard halo mass functions [30] but is realistic if one takes into account the additional clustering due to Poisson fluctuations in the initial PBH separation and the relaxation time of small DM halos. This way, LBs compete with the rate of EBs and could also explain the rates of GW190425, GW190814 and GW190521 [71].

2.3.4 Constraints on PBH models

The EB merging rate predictions for mass bin widths $\Delta m_1 = 2M_\odot$ and $\Delta m_2 = 0.2M_\odot$, $f_{\rm PBH} = 1$ and $\gamma = 0.8$ have been represented in Fig. 2.9 with the corresponding rate limits. The other cases typically lead to lower rates in the mass region probed by the search. A similar figure for LBs is included in the Appendix, where we also compare the rates for $m_1 < 2M_\odot$ to the limits from [39, 40]. In all cases, we find that the subsolar limits do not yet exclude that $f_{\rm PBH} = 1$. Therefore more data will be needed to probe these broad-mass PBH models. For EBs the rate predictions are quite close to the current limits. If some candidates were to be real GW events, they could be explained by EBs given the uncertainties on the rate suppression parameter $f_{\rm sup}$. But it is less

[2]This range for $R_{\rm clust}$ is obtained when imposing the 90% C.L. rate limits from GW190425 at $m_1 = 2.6M_\odot$ and $m_2 = 2.0M_\odot$ with $\gamma = 0.8$ in the mass model 1.

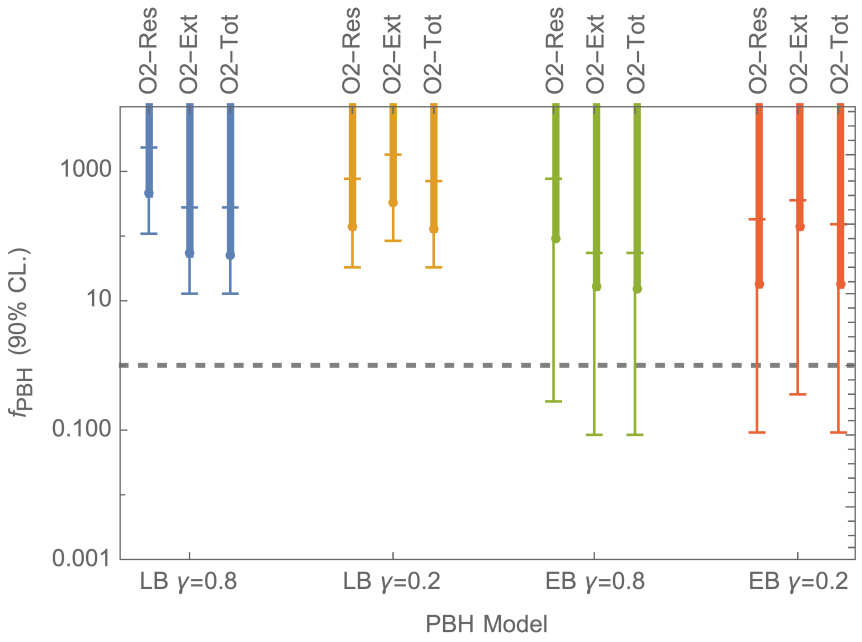

Figure 2.10: Limits (90% C.L.) on $f_{\rm PBH}$ for the mass model 1, for late (LB) and early (EB) binaries for $\gamma = 0.8$ and 0.2, from the search results in O2 data restricted to $m_1 < 2M_\odot$ [40] (O2-Res), this search (O2-Ext) and their combination (O2-Tot). The error bars estimate the uncertainties from the rate parameters $R_{\rm clust}$ and $f_{\rm sup}$ as described in Appendix. The dashed horizontal lines denotes $f_{\rm PBH} = 1$.

likely that they come from LBs. A firm detection would therefore help to distinguish the possible mass distributions and binary formation channels. For each case, we have computed an upper limit on $f_{\rm PBH}$, shown in Fig. 2.10 with error bars corresponding to the estimated rate uncertainties. For this purpose, we assume a Poissonian distribution of GW events and build a simple χ^2 function for $f_{\rm PBH}$,

$$\chi^2 = \sum_{\text{bins}\, i,j} \left[\frac{R_{\rm EB/LB}(f_{\rm PBH}, m_{1,i}, m_{2,j}) \Delta m_1 \Delta m_2}{m_{1,i} m_{2,i} \times \mathcal{R}_{1\sigma}(m_{1,i}, m_{2,j})} \right]^2 + \sum_{\text{bins k}} \left[\frac{R_{\rm EB/LB}(f_{\rm PBH}, m_{{\rm ch},k}) \Delta m_{\rm ch}}{m_{{\rm ch},k} \times \mathcal{R}_{1\sigma}(m_{{\rm ch},k})} \right]^2$$

that combines the (1σ) rate limits in all the mass bins of the present search and the limits in the chirp mass bins of the previous O2 search [40]. $R_{\rm EB/LB}(f_{\rm PBH}, m_{{\rm ch},k})$ denotes the integrated merger rate per unit logarithmic chirp mass. A more accurate analysis would be to calculate the rate limits from a large number of injections, distributed along m_1 and m_2 according to the expected rate distribution for each PBH mass model. However this would have been much more computationally expansive. Given the relatively large theoretical uncertainties on the rate limits, this was beyond the scope of the present paper and is thus left for future work. For $\gamma = 0.8$, we find that this search greatly improves (by about one order of magnitude) the limits on $f_{\rm PBH}$, compared to previous sub-solar searches. But as expected, the improvement is marginal if the peak of the mass function lies well below $2M_\odot$, as in the case $\gamma = 0.2$. For our benchmark modelling of the rates, the combined 90% C.L. limits on $f_{\rm PBH}$ lie between 10 and 100. As a consequence, probing $f_{\rm PBH} < 1$ will be possible with an improved detector sensitivity between two and four. Other

spectral shapes for the primordial fluctuations (lognormal, broken power-law...) are possible but typically lead to a suppression of low mass ratio binaries compared to a scale-invariant spectrum. Therefore the derived limits are relatively conservative.

Finally, we have reanalized the O2 limits on $f_{\rm PBH}$ for monochromatic mass models in [40] with our merging rate prescriptions. Due to the EB rate suppression, these limits are much less stringent and we find that O2 data do not yet exclude $f_{\rm PBH} = 1$, both for EBs and LBs, as shown in Fig. 2.13 in Appendix.

2.3.5 Conclusion

Broad-mass PBH models with a peak at the solar-mass scale from the QCD transition have been proposed to explain the DM and at least some GW events. Motivated by these models, an extended search of mergers of subsolar BHs between 0.1 and $1M_\odot$ in binaries with a primary component between 2 and $10M_\odot$ has been performed on the data from the second observing run (O2) of Advanced LIGO. The search has revealed four candidates (two being single detector triggers) with a total SNR > 8 and a FAR $< 2\,{\rm yr}^{-1}$, but whose statistical significance is not high enough to claim for a detection. Further investigation (parameter estimations, mass model dependence) is ongoing to clarify their status.

Assuming a null result of the search, new merging rate limits are set in this mass range, shown in Fig. 2.12, which complement previous limits restricted to $2M_\odot$ for the primary component. We emphasize that the search results and rate limits are model independent. But in order to translate them in limits on the PBH abundance, one has to assume a mass function and a rate model. Compared to previous studies, we tried to be more realistic while remaining relatively conservative, by choosing a wide mass distribution including inevitable thermal effects from the QCD transition, as well the latest prescriptions for the PBH merging rates.

When they are compared to the theoretical predictions for PBH merging rates, we find that our limits still allow $f_{\rm PBH} = 1$ for both early and late binaries (respectively PBH binaries formed before matter-radiation equality and in PBH clusters). This seems to contradict some conclusions of [40] where more stringent limits on $f_{\rm PBH}$ were obtained. The main difference comes from the inclusion of the rate suppression of early binaries seen in N-body simulations [76]. Nevertheless, we emphasize that rate predictions are within the range of the O3 or future O4 and O5 observing runs. Subsolar searches in these data sets should therefore ideally include binaries with low mass ratios.

Ultimately, detecting a subsolar black hole is the best way to distinguish primordial and stellar BHs, which would therefore have groundbreaking implications in cosmology and high-energy physics. The fact that we found a series of candidates might be a first sign of a subsolar BH population, opening their hunt with longer and more sensitive strain time series.

Acknowledgments

S.C. and K. S. P. are supported by the research program of the Netherlands Organisation for Scientific Research (NWO). G.B. is supported by a FRIA grant from the Belgian fund for research FNRS-F.R.S. S.C. aknowledges support from the FNRS-F.R.S and the Franqui foundation. This

research has made use of data, software and/or web tools obtained from the Gravitational Wave Open Science Center (https://www.gw-openscience.org/), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459.

2.3.6 Appendix

Appendix A: Injections and sensitive space-time volume

We estimate the sensitive space-time volume $\langle \mathrm{VT} \rangle$ of our search by using 15 populations of simulated, non-spinning binary sources, spanning the (m_1, m_2) plane of our search. Binaries in a given population have the same component masses and their distributions are isotropic in sky locations and source orientations, and uniform in co-moving distances. We simulate $\mathcal{O}(10^6)$ sources using the time-domain `TaylorT4` waveforms with phase corrections at 3.5 PN order [89] and inject them in data during the observation period $T_{\rm obs}$ of O2. We search for all these injected signals to measure $\langle \mathrm{VT} \rangle$ for each population. Subsequently, we use the estimated $\langle \mathrm{VT} \rangle$ to calculate upper limits on the merger rate. We carry out this analysis for a uniform mass model search and present the sensitive space-time volumes and the corresponding rate limits in Table 2.2, and on Fig. 2.9.

Appendix B: EB merger rate suppression factor $f_{\rm sup}$

For the suppression factor $f_{\rm sup}[f_{\rm PBH}, f(m_1), f(m_2)]$ in the EB merging rates of Eq. 2.20 we have used the latest analytical prescriptions from Ref. [119], adapted to our broad PBH mass distributions, with two contributions

$$f_{\rm sup} = S_1 \times S_2 \ . \tag{2.21}$$

The first factor is given by

$$S_1 = 1.42 \left[\frac{(\langle m^2_{\rm PBH} \rangle / \langle m_{\rm PBH} \rangle^2)}{\bar{N} + C} + \frac{\sigma^2_{\rm M}}{f^2_{\rm PBH}} \right]^{-21/74} {\rm e}^{-\bar{N}}. \tag{2.22}$$

It takes into account the binary disruption by both matter fluctuations with a (rescaled) variance $\sigma^2_{\rm M} \simeq 0.005$, and by the number of nearby PBHs $\bar{N}$ at a distance smaller than the maximal comoving distance for such nearby PBH to fall onto the binary before matter-radiation equality.

Table 2.2: Sensitive space-time volume $\langle \mathrm{VT} \rangle$ and rate upper limit $\mathcal{R}_{90\%}$.

$(m1, m2)\, M_\odot$	$\langle \mathrm{VT} \rangle$ (Gpc^{3}yr)	$\mathcal{R}_{90\%}$ (Gpc^{-3}yr^{-1})
2.0, 0.21	5.56×10^{-5}	44,368
2.0, 0.40	2.00×10^{-4}	12,333
2.0, 0.60	4.19×10^{-4}	5,882
2.0, 0.80	7.28×10^{-4}	3,389
2.0, 0.99	1.17×10^{-3}	2,111
4.0, 0.41	1.45×10^{-3}	1,704
4.0, 0.60	1.89×10^{-3}	1,302
4.0, 0.80	2.84×10^{-3}	869
4.0, 0.99	3.99×10^{-3}	618
6.0, 0.61	4.97×10^{-3}	496
6.0, 0.80	6.35×10^{-3}	389
6.0, 0.99	8.25×10^{-3}	299
8.0, 0.81	1.00×10^{-2}	246
8.0, 0.99	1.23×10^{-2}	200
9.8, 0.99	1.31×10^{-2}	188

This number, for any mass function, has been estimated as

$$\bar{N} = \frac{m_1 + m_2}{\langle m_{\rm PBH} \rangle} \frac{f_{\rm PBH}}{f_{\rm PBH} + \sigma_M} \ . \tag{2.23}$$

In Eqs. (2.22) and (2.23), $\langle m_{\rm PBH} \rangle$ is the mean PBH mass and $\langle m^2_{\rm PBH} \rangle$ is the corresponding variance. The function C in Eq. 2.22 encodes the transition between small and large $\bar{N}$ limits. A good approximation (with an estimated accuracy of 7% [119] compared to the N-body simulations of [76] for a monochromatic or a lognormal mass function) is given by

$$C \simeq \frac{f^2_{\rm PBH} \langle m^2_{\rm PBH} \rangle}{\sigma^2_{\rm M} \langle m_{\rm PBH} \rangle^2} \times \left\{ \left[\frac{\Gamma(29/37)}{\sqrt{\pi}} U\left(\frac{21}{74}, \frac{1}{2}, \frac{5 f^2_{\rm PBH}}{6 \sigma^2_{\rm M}} \right) \right]^{-74/21} - 1 \right\}^{-1} \tag{2.24}$$

where Γ is the Euler function and U is the confluent hypergeometric function. For an extended mass function spanning several decades of masses, this approach generically gives $\bar{N} \gg 1$, which would lead to a huge rate suppression for EBs and, as a consequence, LB mergers would be dominant. However it is unrealistic because it is unlikely that PBHs with a much smaller mass than the one of the binary components can ionize this binary. Instead, given that we have a sharp peak in the mass function, we use the approximation $\bar{N} \simeq 2$ and $\langle m^2_{\rm PBH} \rangle / \langle m_{\rm PBH} \rangle^2 \simeq 1$ that corresponds to the monochromatic limit. For $f_{\rm PBH} = 1$, one gets $S_1 \approx 0.2$. Given the uncertainties on the exact value of $\bar{N}$, we estimate the possible error on S_1 as the following. As a lower bound, we consider a value of S_1 ten times lower than for $\bar{N} = 2$. As an upper bound, we consider the limit $\bar{N} \to 0$, which gives

$$S_1^{\rm max} = \left(\frac{5 f^2_{\rm PBH}}{6 \sigma^2_{\rm M}} \right)^{\frac{21}{74}} U\left(\frac{21}{74}, \frac{1}{2}, \frac{5 f^2_{\rm PBH}}{6 \sigma^2_{\rm M}} \right) \ . \tag{2.25}$$

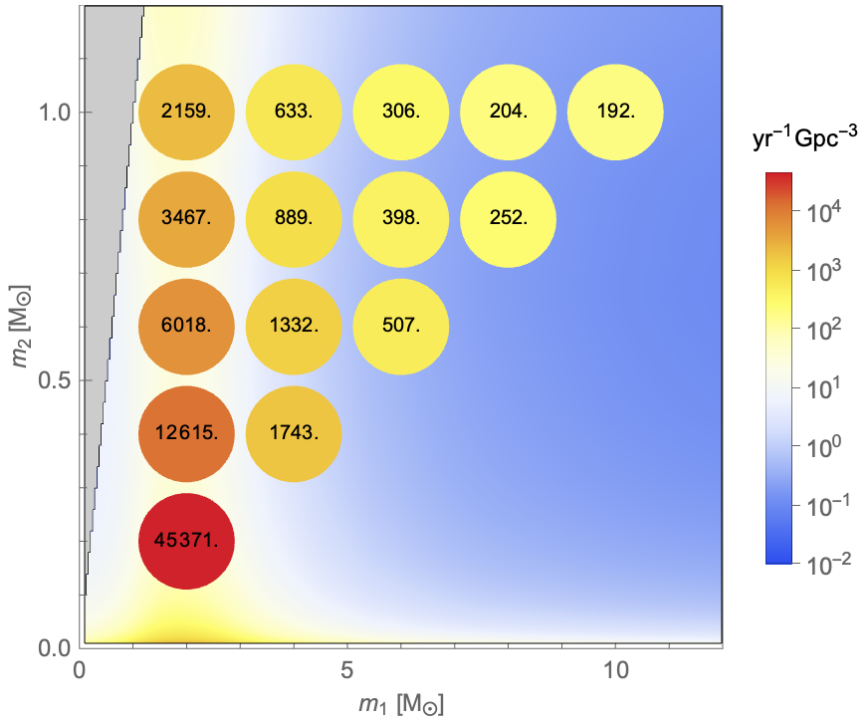

Figure 2.11: Merger rate predictions for LBs and the mass model 1 with $\gamma = 0.8$ and $f_{\rm PBH} = 1$ in mass bin sizes $\Delta m_1 = 2M_\odot$ and $\Delta m_2 = 0.2M_\odot$. The colored disks represent the 90% rate limits from this extended search in O2 data, as in Fig. 2.9.

For $f_{\rm PBH} > 0.1$, one gets $S_1^{\rm max} \simeq 1$ and there is no suppression in the EB merger rate. These lower and upper bounds on S_1 have been used to calculate the uncertainties in the $f_{\rm PBH}$ limits reported in Fig. 2.10 for our different mass models.

The second factor $S_2(f_{\rm PBH})$ comes from the binary disruption in early-forming clusters and can be approximated today by

$$S_2 \approx \min\left(1, 9.6\times10^{-3} f_{\rm PBH}^{-0.65} e^{0.03\ln^2 f_{\rm PBH}}\right). \tag{2.26}$$

Since we are interested by subsolar BHs at typical distances that are still relatively small compared to cosmological scales, one can safely neglect the redshift dependence in S_2 from [119]. For $f_{\rm PBH} = 1$, one gets $S_2 \simeq 0.01$, such that one gets $f_{\rm sup} \simeq 0.002$. Such a value is also motivated observationally, since it can reproduce the merging rates of GW190521, GW190814 and GW190425 [71].

Finally, we point out that if $f_{\rm PBH} \sim 1$, the merger rates from perturbed binaries could exceed the rate of non-perturbed binaries [120], however there is no simple prescription that could be used for extended mass functions and these rates also have large uncertainties. Therefore we did not consider perturbed binaries in our work. It can be also pointed out that subtle general relativistic effects may also highly suppress the rate of EBs [123] if the metric around each PBH is similar to the Thakurta metric. However this is not the case for a perturbed Mc Vittie metric, as shown in [124].

With these assumptions, the merging rates of EBs and LBs for $f_{\rm PBH} = 1$, $\gamma = 0.8$ or $\gamma = 0.2$ are shown in Figs. 2.9, 2.12 and 2.11 for the mass model 1. These rates are compared to the rate

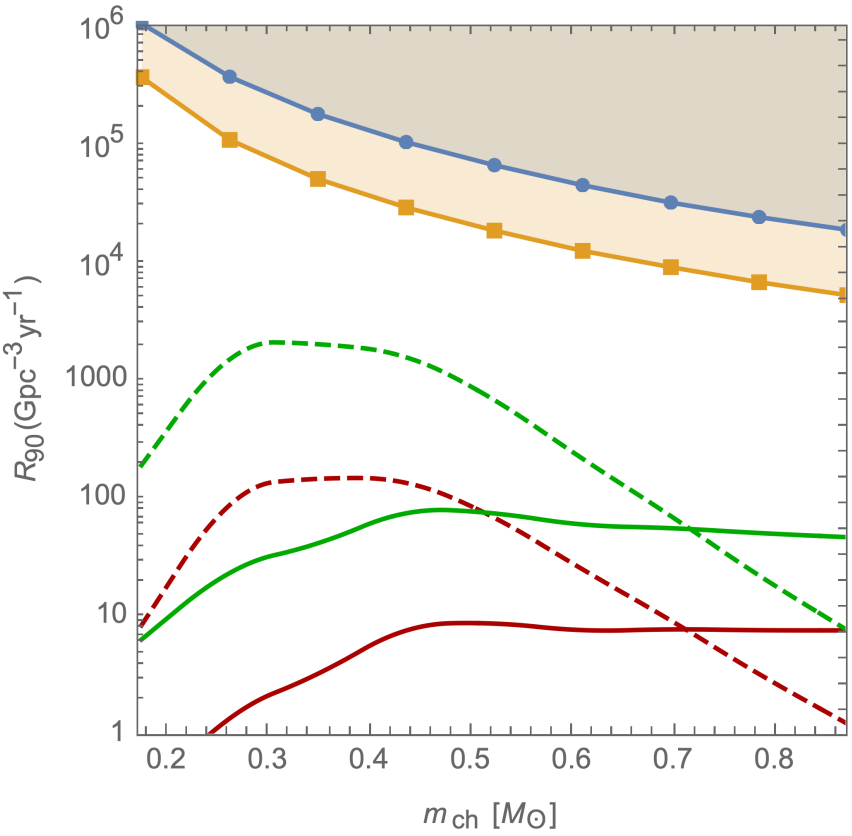

Figure 2.12: The previous O1 (blue line) and O2 (orange line) 90% C.L. limits on the merger rate as a function of the binary chirp mass [121]. Rate predictions for EBs (green) and LBs (red) for the mass model 1 and $f_{\mathrm{PBH}} = 1$, with $\gamma = 0.8$ (solid lines) and 0.2 (dashed lines).

limits from the extended search in O2 data.

Appendix C: Comparison with previous limits

The rate predictions have also been compared to previous limits restricted to $m_1 < 2M_\odot$ [39, 40], for both broad-mass and monochromatic models. On Fig. 2.12 we have represented the rates obtained for the mass model 1 for EBs and LBs, as a function of the binary chirm pass, assuming bins with $\Delta m_{\mathrm{ch}} = 0.15M_\odot$.

The previous O2 limits on f_{PBH} for monochromatic mass models can also be re-analysed with our assumptions for the merging rates of EBs and LBs, using the same method as [125]. These limits are shown in Fig. 2.13 and for EBs they are two orders of magnitude less stringent than previously found, an effect due to the rate suppression factor f_{sup}. We find a limit of $f_{\mathrm{PBH}} \lesssim 1$ only for EBs and masses between $0.8M_\odot$ and $1M_\odot$ but without firmly excluding $f_{\mathrm{PBH}} = 1$ given the rate uncertainties.

Appendix D: Time-frequency analysis

We perform time-frequency analysis of data around the trigger-times of the four candidates events reported in Table 2.1 for quantitative inspection about presence of GW signal. The scalograms or

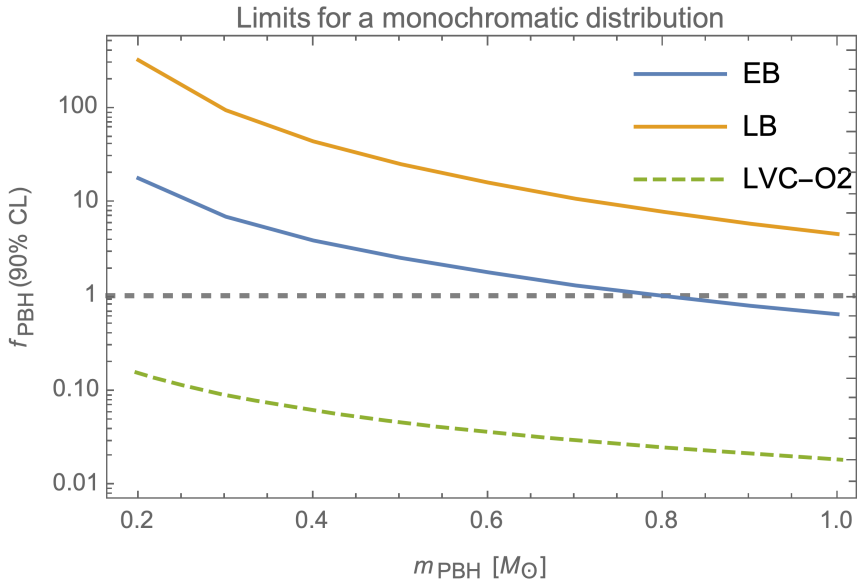

Figure 2.13: Subsolar limits on f_{PBH} for a monochromatic mass model, for late binaries (LB), early binaries (EB) in our benchmark rate models and as estimated for EBs in the previous search [40] without including EB rate suppression effects.

q-scans are obtained from the strain channel data stretches of the two LIGO detectors centered at trigger times of the four candidates. No q-scan provides clear visual indications of chirpy signatures or the presence of GW signal in the time-frequency representation of data. A similar conclusion is obtained for a search of long-duration bursts 80 seconds around the triggers using the strain channel `DCH-CLEAN_STRAIN_C02`, after spectral whitening each 4 seconds with 2 seconds of overlap.

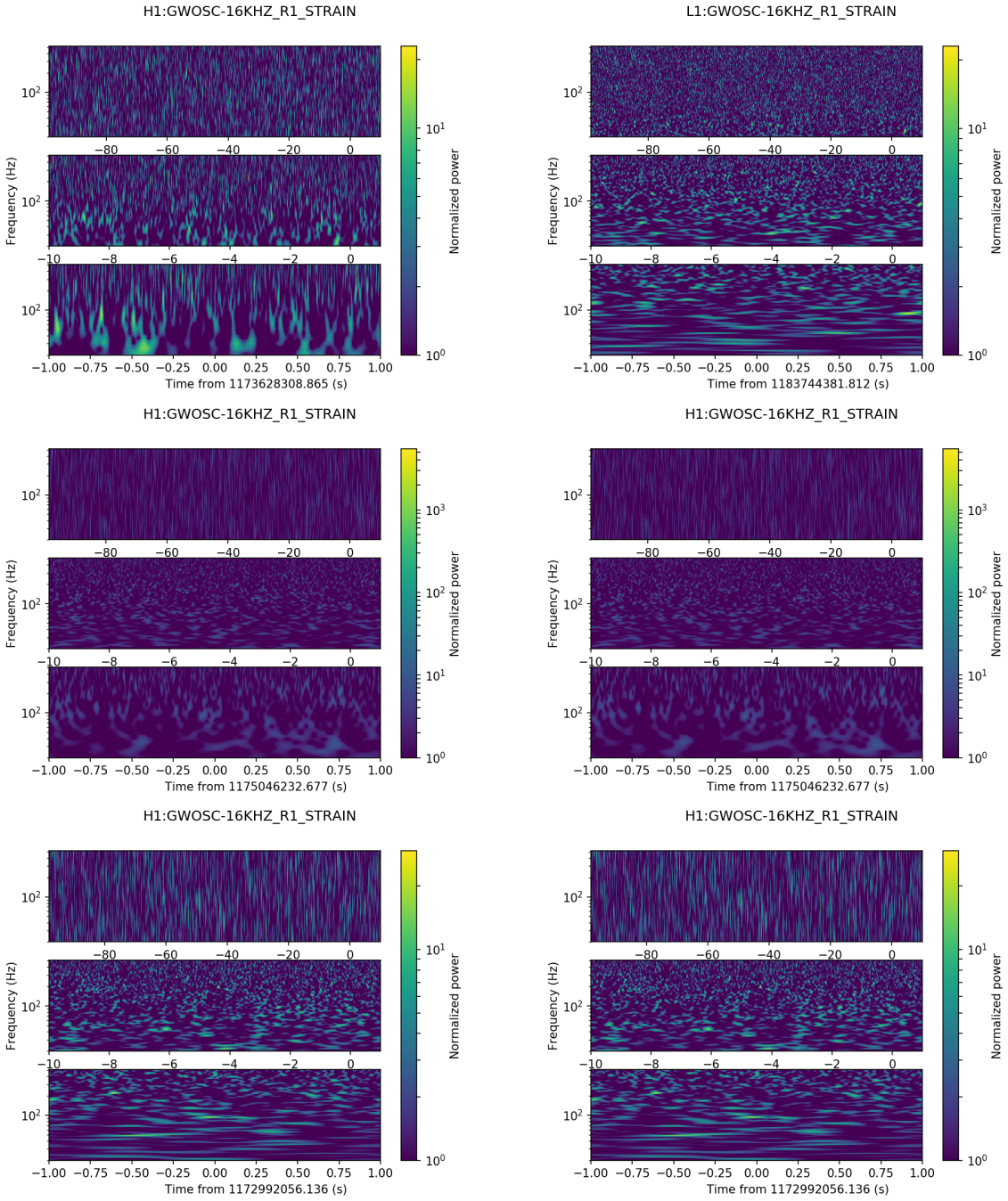

Figure 2.14: Q-scans of strain-channel data centered at trigger-times of the four candidates of Table 2.1: Candidate 1 in H1 (top left), Candidate 2 in L1 (top right), Candidate 3 (middle panels), Candidate 4 (bottom panels). For each trigger, three q-scans with different time intervals are shown. The colored scale represents the normalized power. The plots are generated using PyCBC [122] software. Here, Candidate 1 and Candidate 2 are single detector triggers and observed by LIGO Hanford (H1) and LIGO Livingston (L1), respectively. Data channel name is displayed on top of each plot.

Chapter 3

Detection searches with machine learning

3.1 Deep learning

As mentioned in Section 2.1, longer signals need a bigger template bank. In principle, it is possible to detect a long signal well before the merger has been recorded by the detector. If a GW detector can detect a signal before the merger, it could be very useful for Multi-Messenger Astronomy (MMA). Indeed, it could send an alert to the astronomers with the sky location of the event. Thus, they would have time to point their telescopes in the right direction and perhaps observe an electromagnetic counter-part or to store part of their data before a finer analysis is made. This is particularly useful for the BNSs, in which this kind of system has long signals and can emit light [126, 127]. This idea is one of the main motivations of this thesis and of the other articles that we will present here.

As mentioned, dealing with long signals can be a problem in matched filtering searches. Even though a lot of upgrades have been implemented to speed up these searches [18, 22, 21, 77, 128, 129], they are still computationally heavy and slower than other techniques such as Deep Learning (DL). The next two papers [130, 131] that we will present used DL techniques to detect a GW signal coming from the early inspiral of a BNS. This work was the main focus of the thesis and we have produced two papers on this subject. While the first is a proof of concept with multiple limitations, the second introduces the improvements and has been set up to run in a more realistic scenario.

3.1.1 Neurons and neural networks

Before presenting these papers, we will give a small introduction to DL. Machine learning is an ensemble of techniques that allow the software to learn by itself and increase its own performance at predicting outcomes without being explicitly programmed to do so. DL is a subbranch of machine learning, which uses neural networks. Deep learning can deal with a lot of different problems, such as classification, data generation, etc. We start by explaining it with a simple regression problem.

To understand how it works, let us make an analogy with linear regression. One has some input x and one has to predict the y value. One can represent this problem as a cloud of points in 2D space x, y as shown in Fig. 3.1. The easiest model that we can think about is a linear regression. One can construct such a model by finding the optimal parameters m and p, such that the equation $y = mx + p$ predicts y given x well. To do so, it is needed to introduce a loss function l, which, will be minimized with respect to m and p, to find the optimal linear regression. For example, one can use, as a loss function, the Mean Squared Error (MSE):

$$l = \frac{1}{N}\sum_{i=0}^{N} e_i^2 = \frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2 \tag{3.1}$$

where y_i and x_i are respectively the y and x components of the point i, and $\hat{y}_i$ is the prediction of the linear regression at a value of x_i. In this particular case of linear regression with such a loss, there exists an analytic solution to find the optimal m and p. When the best line is found, according to the cloud of points, one can use it to estimate y given some new x.

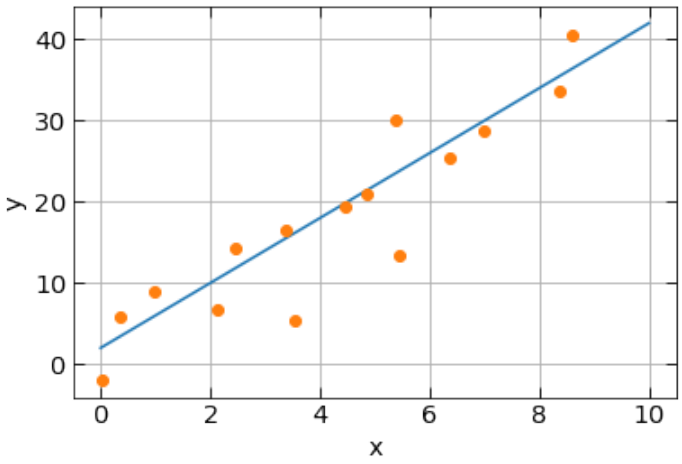

Figure 3.1: Illustration of a linear regression. The blue line is the linear regression with an equation of $y = mx + p$, in which m and p are parameters that need to be determined by minimizing the MSE.

The principle of deep learning is the same, except that the model is a neural network, which is a mathematical differentiable function [1] with way more parameters than a linear regression. The easiest and simplest neural network to describe is the one based on dense (or fully connected) layers, illustrated in Fig. 3.2. The basic block of a neural network is a neuron and its mathematical expression is:

$$h_0 = AF(\sum_i w_i x_i + b) \tag{3.2}$$

[1] The functions used in a neural network do not have to be differentiable everywhere. For example, the ReLU, see Fig. 3.3, is not differentiable in 0, but it is very regularly used in deep learning because in practice you are never precisely at 0.

where $\{x_i\}$ is the component i of the input vector, $\{w_i\}$ is a parameter called *weight*, b is also a parameter called the *bias* and AF represents an activation function. An artificial neuron is inspired by the one of the brain. In the brain, a neuron is connected to a lot of others and activates itself only if the signal received from the others is strong enough. This feature is represented by the activation function, which could be anything that mimics the true neuron activation. Fig. 3.3 illustrates different types of activation functions. In our work, we mainly used the Sigmoid and the ReLU, which have respectively the following equations:

$$\sigma(x) = \frac{1}{1 + exp^{-x}} \tag{3.3}$$

and the ReLU, which is defined as:

$$ReLU(x) = \begin{cases} x & \text{if } x \text{ bigger than } 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

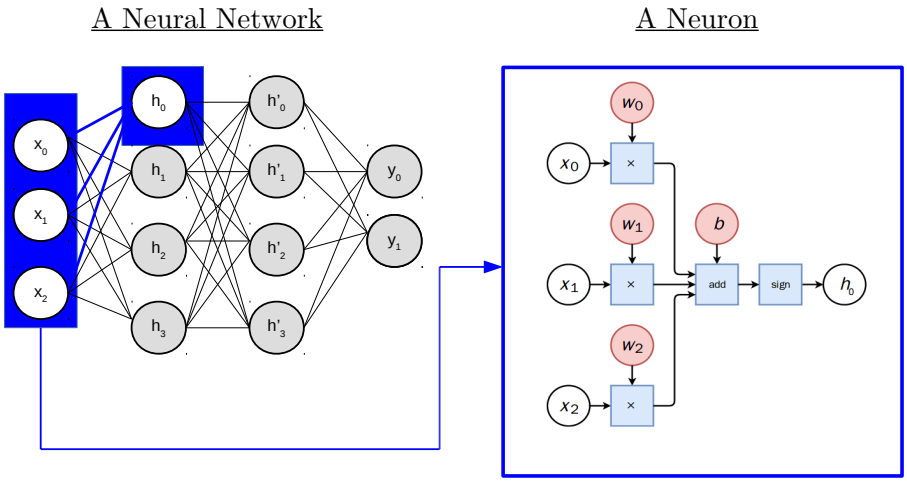

Figure 3.2: *Left*: Illustration of a neural network made of three linear layers, x represents the input, h and h' represent the value of the neurons, respectively of the first and second hidden layer, and y is the final output of the neural network. *Right*: Illustration of a neuron. Sign represents the activation function; the equation of a neuron is given in (3.2).

A neural network (made of fully connected layers) is just multiple neurons, organized in layers, in which each neuron takes as input the neurons of the previous layers, see Fig. 3.2. Thus, one can express a layer of neurons as a vector:

$$\overrightarrow{h^{\mu}} = AF(\mathbf{w}^T \overrightarrow{h^{\mu-1}} + b^{\mu}) \tag{3.5}$$

$\mathbf{w}^T$ represents the transposed matrix of the weight of the whole network, $\overrightarrow{h^{\mu}}$ is the value of the neurons of the layers μ, and b^{μ} is the bias at the layers μ. Let us remark here that each layer can be expressed as a function of the previous ones. Therefore, the output of the whole network is given

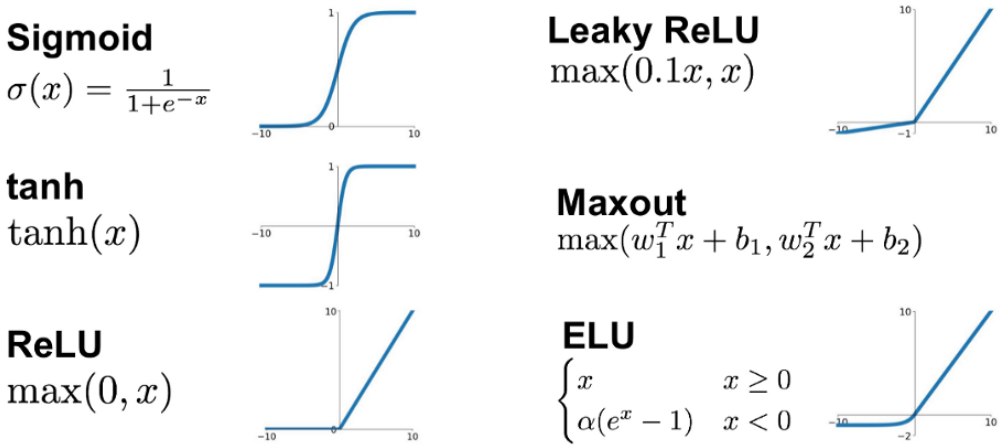

Figure 3.3: Representation of different activation functions.

by Equation (3.5), in which μ is the final layer. Just like our problem with the linear regression, all we have to do is to define a loss for our problem and minimize it with respect to $\mathbf{w}$ and $\overrightarrow{b}$. Because of the activation function, a neural network is highly non-linear, and thus, it does not have an analytical solution. Nevertheless, it is still a differentiable function. A well-known technique to be able to find at least a local minimum of the loss is the gradient descent.

3.1.2 Gradient descent

The gradient descent technique is an iterative process. It starts by evaluating the loss for a given set of values of the parameters $\mathbf{w}$ and $\overrightarrow{b}$. Then, it makes a local approximation of this loss, and minimize it. Finally, the real loss is re-evaluated with the new value of the parameters and it repeats the process. Fig. 3.4 represents this gradient descent cycle. $\mathcal{L}$ denotes the loss and the whole parameters ($\mathbf{w}$ and $\overrightarrow{b}$) is denoted θ. We start by evaluating $\mathcal{L}$ at a set value of the parameters θ_0. It is possible to approximate the loss around that point θ_0 with a first order approximation:

$$\hat{\mathcal{L}}(\theta_0 + \epsilon) = \mathcal{L}(\theta_0) + \epsilon \nabla_\theta \mathcal{L}(\theta_0) + \frac{1}{2\gamma} ||\epsilon||^2 \tag{3.6}$$

$\hat{\mathcal{L}}$ represents the approximate loss, ϵ is a small variation in the parameters space, and γ is a free parameter that one needs to choose, known as the *learning rate*. The last term is added so that the Equation (3.6) is a parabola tangent to the true loss at the point θ_0 as shown in Fig. 3.4. The learning rate γ controls the size of the parabola. The goal is to find the ϵ that minimizes the approximate loss and re-evaluate it at this new point $(\theta_0 + \epsilon)$. To find this minimum, one simply has to solve $\nabla_\epsilon \hat{\mathcal{L}}(\theta + \epsilon) = 0$. If one does so, one can obtain the value of ϵ:

$$\epsilon = -\gamma \nabla_\theta \mathcal{L}(\theta_0) \tag{3.7}$$

At each step of the gradient descent, the value of the parameters is changed to:

$$\theta_{n+1} = \theta_n - \gamma \nabla_\theta \mathcal{L}(\theta_n) \tag{3.8}$$

where n represents the number of steps already done. Training a neural network is nothing else than performing this gradient descent until one finds at least a local minimum in the parameters space that allows the network to perform correctly its task. Equation (3.7) depends on the parameter γ, which is an important parameter. If it is too big, one can miss a minimum by just jumping from one side to the other. If it is too small, the algorithm just converges too slowly to the minimum. This is one of the parameters that one can adjust during the training, or with some optimizer, but we will discuss this later.

$\mathcal{L}$ is the loss computed for all the training sets. For example, if one uses the mean squared error (MSE), one needs to compute the MSE for each piece of data, and just as in the case of the linear regression, the total loss $\mathcal{L}$ will be the sum (or the average) of these MSE losses. Computing the loss and the gradient for the whole dataset can take time and computational memory. In a lot of cases, the computer has not enough memory to load the whole dataset and perform one step of gradient descent over it. Because of this, it is very common not to compute the loss over the whole dataset before making a step of gradient descent, but only over a smaller sub-set called the *batch*[2]. Using a smaller batch size allows us to explore the parameters space in a more erratic way than using the full data set. On the one hand that helps to avoid local minima, but on the other hand, it could also take more time to converge. In practice, using a higher batch size is a good idea to increase the speed of the training, and, at least in our case, the batch size is usually limited by the amount of memory that we could to use.

Multiple tricks can be added to the principle of the gradient descent to converge faster or to avoid local minima, the algorithm that computes this gradient descent with all the tricks is called the *optimizer*. For example, the one we used in our search is Adam (for Adaptive Moment Estimation) and AdaMax [132]. We do not enter into the detail here, but the idea behind Adam is that at each step, the learning rate is adapted by a factor that depends on a running average of both the gradients and the second moment of the gradient.

3.1.3 Convolutional neural networks

Until now, we have taken the example of a neural network composed of dense layers. Yet, in deep learning, it is possible to invent a lot of different types of layers, and searching for new architectures is still an active field of research. In a single network, it is also common to combine multiple types of layers. In our papers, the network is composed of convolution with some dense layers only at the end. Therefore, let us describe how convolutional layers work.

The problem with dense layers is that the number of parameters can be huge for large and deep

[2]When using a batch of size 1, so with only one piece of data, is called a *stochastic gradient descent*.

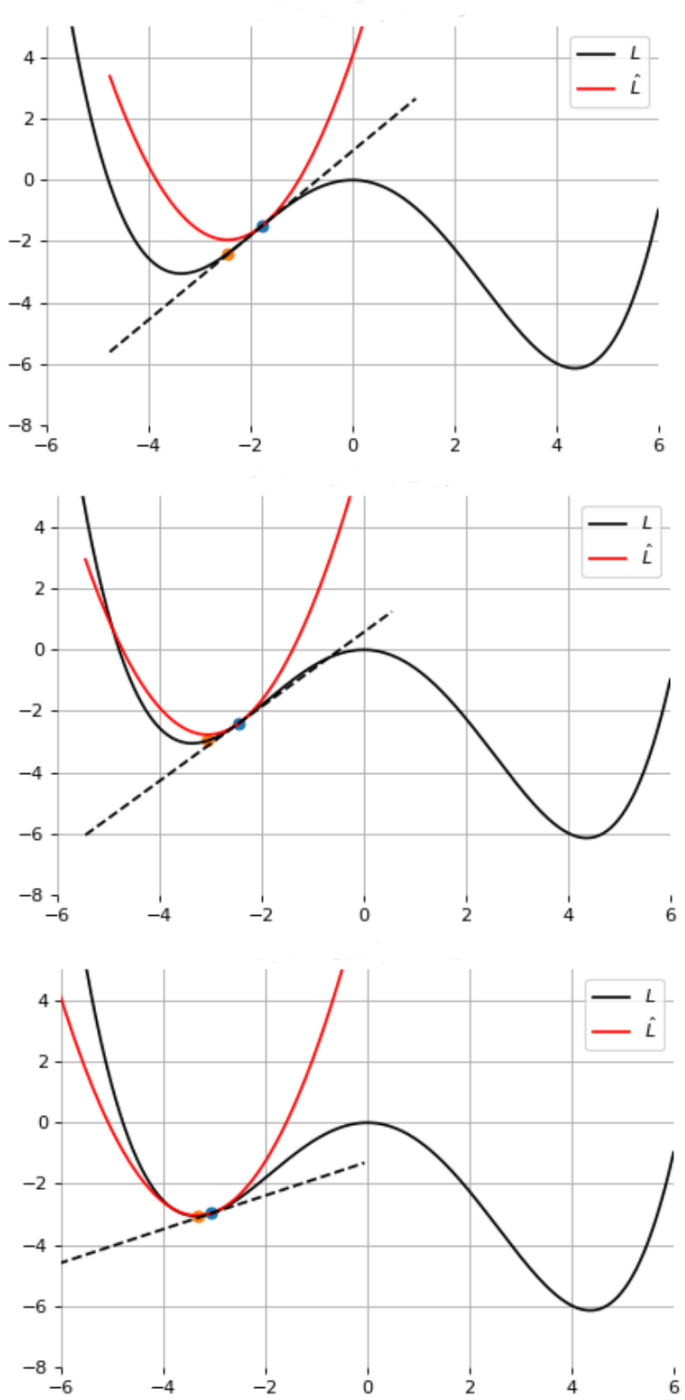

Figure 3.4: Representation of the gradient descent techniques for 3 different steps. The black curve is the real loss, the red one is the first-order approximation, and the dashed line is the tangent to the real loss at the point θ_n.

networks, and it could take a very long time to train it properly. In addition, the necessary memory depends on the number of parameters. As each neuron of one layer is connected to every neuron of the next one in a dense layer, the information of the spatial position of the input is destroyed in a dense layer. This can be a problem, for example, if one wants to do image recognition, the

information of where is which part of the image is important. To solve both of these problems, convolutional layers have been created. They are usually conceived in 2 dimensions because they are especially useful for image recognition. Nevertheless, we introduce them with the 1D version since we deal with 1D strain data in our work. The concept of convolutional layers can be easily adapted to higher dimensions.

In a convolution layer, one neuron of this layer is not connected to all those of the previous layer, but only to a few of them. As you can see in Fig. 3.5 which represents the principle of a convolutional layer, the first neuron of the second layer is only connected to the first three of the previous layer. The second neuron takes as input the second, the third, and the fourth, and so on. One can also see in Fig. 3.5, that every neuron has the same weight in a same layer. In this example, there are only three weights on this layer, which is less than in a dense layer of the same size. The set of weights in a layer is called the *kernel*, and so the number of weights is the *kernel size*.

It is also common to stack convolutional layers in parallel, just as shown in Fig. 3.6. In that case, each "parallel layer", which are called the *channels*, has the same kernel size, but the kernel (meaning the weights) could be different. In Fig. 3.6, the number of parameters depends on the kernel size and must be multiplied by the number of channels. Let us interpret this kind of layer in 2 dimensions, in which one can have a simpler intuition. Let us imagine that a convolutional neural network is trained to recognize cats in pictures. The kernel of a channel in the first layers will activate itself only if it sees straight lines. Another kernel from another channel in the same layer will be triggered only if it sees corners. In the first layer, each neuron of the different pattern is trained to trigger if they see very basic and small patterns. In the second layer, a neuron will trigger only if it sees a more complex pattern, such as a rectangle or another simple form. The deeper you go into the network, the more complex, abstract, and macroscopic the representation of the neuron is. In the very last layers of our supposed network, a neuron can represent eyes, a mouth, a tail, and in the final layer, a cat. In reality, it is very difficult to say exactly what a neuron represents because it could be very abstract.

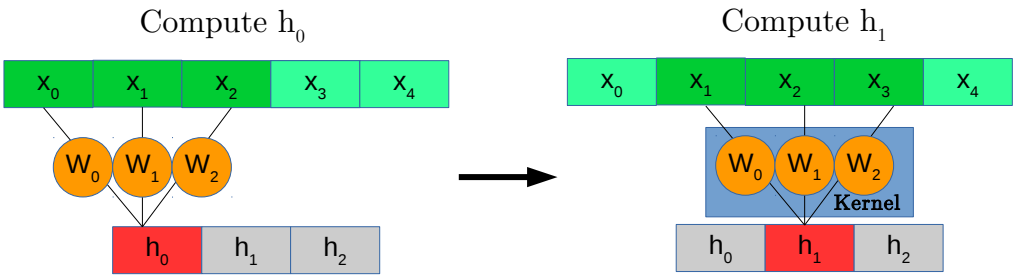

Figure 3.5: Representation of a convolutional layer. The first neuron h_0 is only connected to the first three of the previous layers. The second neuron h_1 is connected to x_1, x_2, x_3, but the kernel to compute h_0 or h_1 or any other neuron of this layer is the same. The kernel size is three in this example.

In a convolutional layer, there are multiple hyperparameters which one can play with. We have already seen the number of channels, the kernel size, which represents the number of connections that exist between one neuron and the previous layer. Yet, they are more hyperparameters. For example, there is what it is called the *stride*. The stride represents the step of the kernel between a neuron and the next one. For example, in Fig. 3.7, the first neuron is connected to the first, second, and third neurons of the previous layer, but the second one is connected to the third, fourth and fifth ones. This represents a stride of two because the step is two in that specific case, while in Fig. 3.5 the step was one. Unlike a dense layer, in which you can choose the output length, the output length in a convolution layer is a function of the kernel size, the stride, and the input length (the formula is given in Equation (3.9)). One can wish to keep the same size for input output. To do so, one introduces another hyperparameter: the padding. The idea of padding is to increase the size of the input artificially during the convolution by adding some zeros in the beginning and at the end of the input. For example, in Fig. 3.8, one sees that the first neuron will take 0 value, the first and the second neurons of the previous layer as input the due to padding. The number of padding is just the number of zeros added at the beginning and at the end of the input. A last feature of the convolution is the dilation, which has been introduced by [133]. The idea is to leave a space between the connection of a neuron and those of the previous layers. For example, in Fig. 3.9 the first neuron takes the first, the third, and the fifth neurons of the previous layer as input, which illustrates a dilation, while in Fig. 3.5, there is a dilation of zero, which means that there is no dilation. As mentioned, it is possible to know the length of the output L_{out} by taking into account all of these features. It can be expressed as follows:

$$L_{out} = \lfloor \frac{L_{in} + 2l_{pad} - l_{dil}(l_{ker} - 1) - 1}{l_{str}} + 1 \rfloor \tag{3.9}$$

where L_{in}, l_{pad} is the padding, l_{dil} is the dilation, l_{ker} is the kernel size and l_{str} is the stride. The symbol $\lfloor \rfloor$ represent the inferior entire part.

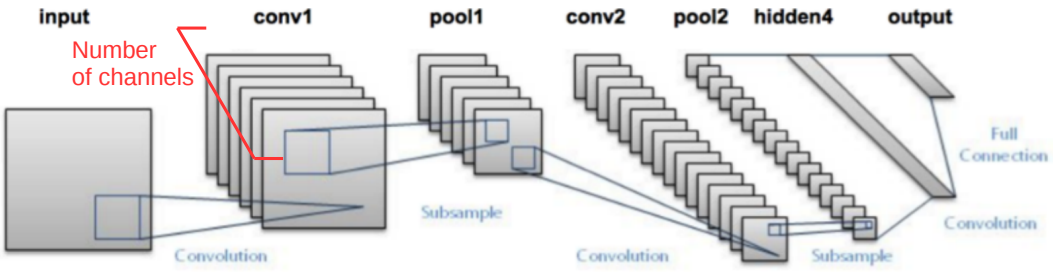

Figure 3.6: Representation of a 2D convolutional neural network followed by two dense layers. The neural network takes only one image as input, but the first layers have six channels.

In some situations, one can wish to reduce the output size of a convolutional layer: for example, in order to reduce the memory required to train the network. To do so, one can use pool layers. They are not trainable layers, which means that they have no parameter of their own which needs

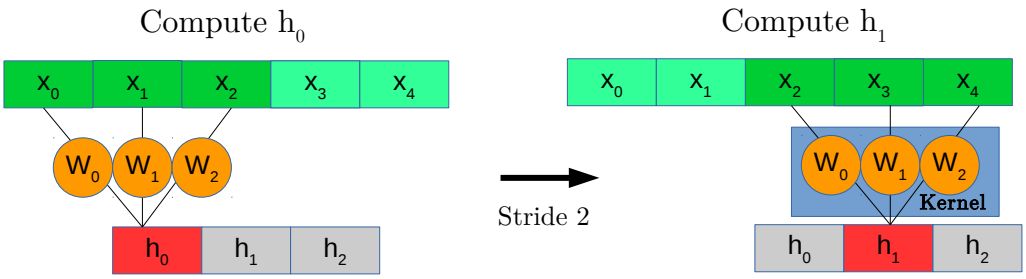

Figure 3.7: Representation of the principle of a convolutional layer, with a kernel size of 3 and a stride of 2.

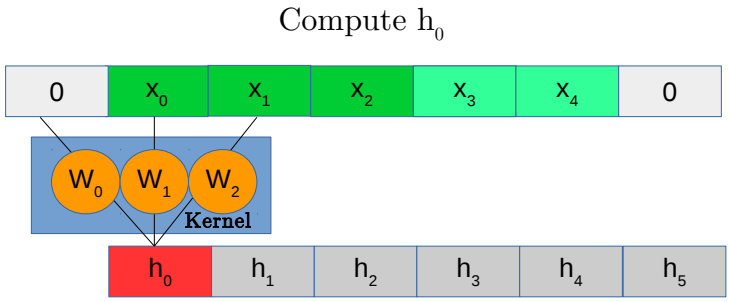

Figure 3.8: Representation of the principle of a convolutional layer, with a padding of one.

to be fine-tuned by the gradient descent technique. A pool layer works similarly to a convolutional layer in the sense that it also has a kernel, and that it is possible to use padding, stride, and dilation. Le us start by explaining the MaxPool layers, but there are others. In a MaxPool layer, the value of a neuron will be the maximum of those of the kernel, as illustrated in Fig. 3.10. There also exist average pool layers, in which one takes the average over the kernel instead of taking the maximum value of the kernel. As mentioned, the pooling layer is often used to reduce the size of the input after a convolution. To do so, the stride is often higher than the one in the convolution. To determine the length of the layer, one can use formula (3.9).

Training a neural network can be challenging. A way to facilitate it is to normalize the data. A well-known type of normalization in deep learning is batch normalization [134]: it consists of ensuring that the variance of the data in a batch is the same for each piece of data. During the training, it shifts and rescales according to the mean and the variance computed on the batch. When the network is trained, the batch normalization will shift and rescale according to the empirical moments estimated during the training. The mathematical expression of the batch normalization is:

$$y = \frac{x - E[x]}{\sqrt{Var[x] + \epsilon}}\gamma + \beta \tag{3.10}$$

where x is the input, $E[x]$ and $Var[x]$ are respectively the mean and the standard deviation computed on the whole batch, ϵ is an arbitrary small constant added to avoid division by 0 and γ and

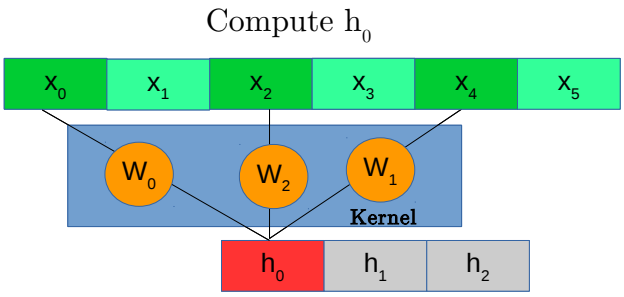

Figure 3.9: Representation of the principle of a convolutional layer, with a dilation of one.

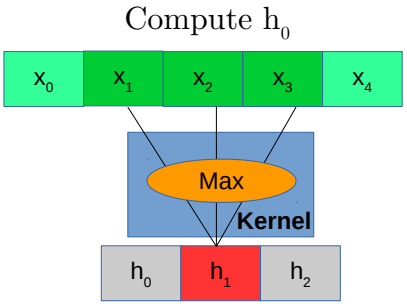

Figure 3.10: Representation of the principle of a MaxPool layer. Here h_1 will take the value of the maximum between x_1, x_2, x_3

β are learnable parameters.

A final tool that we need to introduce here is the Softmax layer. Since the two papers that we will present here deal with a classification problem, we want the output of the network to represents a vector in which each component is the probability that the input belongs to a particular class. The component of the output vector should be between zero and one, and the sum should be equal to one. A Softmax will transform every vector into a vector which has those two proprieties. The mathematical expression of a Softmax is:

$$SM(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{3.11}$$

where x_i is the i^{th} component of the input vector. The exponentials are there to ensure that all the value become positive.

In deep learning, it is common to see every operation made to the initial input as a layer, i.e. when one does a batch normalization to the data, it is common to say that one first applies a layer of batch normalization. The same stands for the activation function. For example, if one builds a dense layer or a convolutional layer, and one chooses to use a ReLU as activation function, it is common to say that one first has a dense (or convolutional) layer and then a layer of ReLU. This only means that one used the ReLU as activation function. One can define one's own network

depending on the problem by stacking different layers. For example, in the two papers that we will present, the network is composed of batch normalization layers followed by a few blocks of convolution, ReLU, pool layers, and finally attached to some dense layers at the end.

3.2 Early alert and multi-messenger astronomy

Before presenting our two papers on machine learning techniques for pre-merger alerts, let us introduce multi-messenger astronomy (MMA) and why early alerts of GWs are important in this field. MMA is astronomy based on the observation of a same event with different channels such as gravitational waves, electromagnetic waves, neutrinos, and cosmic rays. MMA is not new, in 1940, cosmic rays were detected coming from solar flares [135]. In 1987, a supernova detected with an electromagnetic telescope was also detected through neutrinos [136]. More interesting for us, in 2017, the Fermi Gamma-ray Burst Monitor saw a gamma-ray burst six minutes after the LIGO Virgo collaboration announced that a low latency pipeline had detected a BNS event known as GW170817 [127]. This was the first example of MMA with GWs. Using multiple channels to observe an event can lead to a better understanding of some very inserting physics such as the r- and s-processes which are at the origin of heavier elements in the Universe. It can also lead to a better measurement of the Hubble constant and to new tests of GR, such as a measurement of the speed of GWs [137, 138, 139, 140, 141, 142]. In principle, gravitational waves could play another important role in this MMA field. The electromagnetic signal emitted by a BNS occurs at the very end of the inspiral and at the merger [143]. On the contrary, the GW detectors can detect the early inspiral of BNSs. In addition, the LIGO and Virgo interferometers can survey the whole sky at the same time, which is not the case for the light telescopes, which can only focus on a small portion of the sky. Because of these two characteristics, GWs can serve as an early alert for the other messengers. If one detects the early inspiral of a BNS and if one can deduce its sky location, one can send an alert to the astronomer and suggest surveying the portion of the sky in which the BNS is, so that they can perhaps record an electromagnetic counterpart.

Such a scenario has not occurred yet. This is mainly because the pipelines were looking for the signal as a whole, not just for the early inspiral. This is not the case anymore since some pipelines have adapted their method to be able to produce pre-merger alerts. For example, the gstLAL team has now set up a method to search for the early inspiral of BNSs [144]. With this method, it performs multiple searches at the same time. Like all the matched filtering pipelines, it computes a template bank. Yet, instead of performing the matched filter with this template bank once, it does it six times. Each time, the filtering is done with a different maximum frequency, meaning that each search looks for a different time before the merger. Ref. [144] ha shown that they can recover some events into design sensitivity noise around one minute before the merger. PyCBClive [18, 145], another pipeline based on matched filtering is also able to produce pre-merger alerts with a similar strategy: where it uses truncated waveform to search for the early inspiral part.

However, these adapted matched filtering techniques have some disadvantages. As explained,

they sometimes need to run multiple analyses at the same time to check at different times before the merger. This results in a need for more computational resources. We also should remember here that these techniques already need parallelization over multiple CPUs to be able to perform the matched filtering with all the templates of the bank. This is a reason why some people started to think about other methods to detect CBC events and to produce pre-merger alerts. Some of those techniques include machine learning and convolutional neural network. The first paper to use this kind of algorithm was [146], in which CNN was trained as a classifier to distinguish between noise and noise plus CBC signal coming from massive BBHs. In the following years, multiple groups have tried to improve this method and adapt it to BNS, or have just tried to consider a more realistic scenario or increase the performance [147, 148, 149, 150, 151]. Later, some papers used these neural networks to produce early alerts [152, 153, 130, 131]. We present here one of the first papers on this subject [130], which is a proof of concept to search for the early inspiral with a CNN. Then, we present a second paper [131], which deals with the upgrades on this method, and which works on a more realistic scenario.

3.3 Article 2

Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal

Grégory Baltus, Justin Janquart, Melissa Lopez, Amit Reza, Sarah Caudill, Jean-René Cudell

Abstract

GW170817 has led to the first example of multi-messenger astronomy with observations from gravitational wave interferometers and electromagnetic telescopes combined to characterise the source. However, detections of the early inspiral phase by the gravitational wave detectors would allow the observation of the earlier stages of the merger in the electromagnetic band, improving multi-messenger astronomy and giving access to new information. In this paper, we introduce a new machine-learning-based approach to produce early-warning alerts for an inspiraling binary neutron star system, based only on the early inspiral part of the signal. We give a proof of concept to show the possibility to use a combination of small convolutional neural networks trained on the whitened detector strain in the time domain to detect and classify early inspirals. Each of those is targeting a specific range of chirp masses dividing the binary neutron star category into three sub-classes: light, intermediate and heavy. In this work, we focus on one LIGO detector at design sensitivity and generate noise from the design power spectral density. We show that within this setup it is possible to produce an early alert up to 100 seconds before the merger for the best-case scenario. We also present some future upgrades that will enhance the detection capabilities of our convolutional neural networks. Finally, we also show that the current number of detections for a realistic binary neutron star population is comparable to that of matched filtering and that there is a high probability to detect GW170817- and GW190425-like events at design sensitivity.

3.3.1 Introduction

On August 17, 2017 the first gravitational wave (GW) from a binary neutron star (BNS) system was observed by the Laser Interferometer Gravitational Wave Observatory (LIGO) [6] and by the Virgo detector [50].

The Fermi Gamma-Ray-Burst Monitor (Fermi-GBM) [154] and the INTEGRAL satellite [155] detected the associated γ-ray signal 1.7 s after the coalescence. This event, called GW170817, provided the first direct evidence of a link between these mergers and short γ-ray bursts. In addition, it gave an extra confirmation of the existence of GWs and initiated the era of multi-messenger astronomy (MMA) with GWs [17, 127, 156, 157]. The combined detection of multiple messengers allows us to improve our understanding of complex astrophysical phenomena, such as the r- and s-processes at the origin of heavier elements in the Universe. This improvement of MMA will also allow a better measurement of the Hubble constant and novel tests of General Relativity, such as a measurement of the speed of GWs [137, 138, 139, 140, 141, 142]. A key element in MMA is the time delay between the detection of a GW and the identification of the location of its source. In fact, to discover new physics it is necessary to detect the source in the electromagnetic band at times close to the merger. However, the time of response varies from one telescope to another, so that a detection in the GW channels should be made early to leave enough time for the electromagnetic observatories to focus on the source. For instance, the Swift Observatory [158] is able to focus on a sky position in only seconds (around 15 s for Swift's Burst Alert Telescope [158]). Thus, it is advantageous to detect gravitational waves from the inspiral, before the merger, to enable prompt detection of the merger event in the electromagnetic band.

The signal of a GW coming from compact binary coalescence (CBC) is composed of three parts: the inspiral (when the orbital motion of the two objects radiates away energy and the orbit shrinks), the merger (when they touch and join), then by a ringdown (when the newly formed body returns to its ground state). As the signal enters in the detector, the signal-to-noise-ratio (SNR) accumulates. Due to the low frequency and small amplitude of the signal during the early inspiral, the SNR accumulation is slow, which hinders the detection of the signal. It becomes observable when the frequency of the early inspiral enters into the sensitivity band of the detector. Upgrading the current detectors and building the next generation of interferometers, such as Cosmic Explorer (CE) and the Einstein Telescope (ET) [159, 160, 161], will increase significantly the sensitivity through, among other things, the reduction of the noise. As a consequence the frequency threshold will also decrease, leading to the detection of signals with longer inspirals and to a larger number of detections. Up to now, both the low-amplitude early inspiral and the high-amplitude late inspiral have been needed to detect BNSs. Some studies already investigate the question of early warning for the next generation of interferometers [162].

The standard methodology employed to search for GWs relies on *matched filtering* techniques. A large bank of template waveforms is built. The templates are then correlated with the input data of the detector over their sensitivity band, extracting the signals from the detector noise. The standard matched filtering can be computationally expensive [78, 79, 163, 164], but some pipelines such as GstLAL [21], PyCBCLive [18], MBTAOnline [165] and SPIIR [129] are adapted to run in low latency and obtain fast candidate detection, also referred to as trigger.

Matched filtering techniques usually consider the whole template (meaning the template over all the sensitive frequencies of the detector) to correlate it with the signal. However in the context of MMA it is necessary to employ only the pre-merger information of the template for prompt alerts. In this line of thought, recent advances to perform matched filtering with only a fraction of the inspiral have been made in Ref. [144], where the authors have implemented a GstLAL-based pipeline that produces pre-merger alerts.

It enabled them to compute the matched filter, the false-alarm rate, and the sky localisation using only the information in the low-frequency band of the template, corresponding to the early inspiral. This showed that this method could detect signals as early as one minute before the merger. However, the number of early alerts issued is lower than the total number of detections based on full waveforms.

Due to the computational complexity of matched filtering and the increasing amount of events related to the future upgrades of the detectors, alternative approaches to overcome the challenges of MMA are under development. In particular, the use of Machine Learning (ML) methods has sparked the interest of several authors, who have built Deep Learning (DL) algorithms. These algorithms are able to capture complex non-linear relationships in the data by composing hierarchical internal representations. The main advantage of these methodologies is that the prediction task is performed rapidly since most of the computations are made during the training stage [146]. Several studies have shown the power of these algorithms for the detection of GW in low latency, obtaining a sensitivity similar to that of matched filtering techniques [146, 147, 148, 150, 153, 166]. Other recent papers [167, 168, 169, 170, 171] focus on the parameter estimation for CBC events and other applications of ML for gravitational wave astronomy. Moreover, in [146] and [166] the authors have presented the generalisation ability of convolutional neural networks (CNNs) by training with a data set of non-spinning waveforms and obtaining a high performance when testing with precessing systems.

In [153], only part of the template is used in their DL approach. The authors employed a pre-trained Resnet-50 network [172] to classify time-frequency maps. The data were acquired from the detectors after an extra preprossessing step to build a spectrogram of the data. When computing such figures, depending on the desired resolution, we found that the time to produce a single map could vary from ~ 0.5 seconds [3].

In this paper, we propose to use the 1-D whitened strain as the input data of a CNN for pre-merger alert so that we bypass the computation of the spectrograms. The goal of the algorithm is to perform a binary classification task to differentiate inputs that contain a GW from inputs that do not. The classification is made independently for three different categories of objects: light, intermediate and heavy BNS. The templates with a GW contain only the early inspiral part of the waveform, which is embedded in colored Gaussian noise, made with the noise power spectral density (PSD), corresponding to the design sensitivity of LIGO and given in PyCBC [122]. Different categories have different observational time windows (OTW), i.e. the duration of inspiral seen by each CNN is different depending on the category. A short CNN, as in [146, 147, 166], is implemented for each category. We stress that this work is a proof of concept to show the promises

[3] All the tests on CPU were done on a Intel(R) Core(TM) i7-8650U CPU.

of this type of neural networks. The optimisation of their performance and the inclusion of multiple detectors will be considered in further studies.

This paper is organised as follows. In the first section, we discuss the data generation. The second section details the methodology used to design and train the networks. The third section is devoted to our different results and discusses them. Finally, we draw our conclusions and consider future enhancements to be brought to our detection system.

3.3.2 Data generation

SNR and Partial-Inspiral SNR

The final output of a matched filtering algorithm is the signal-to-noise ratio (SNR). It measures the match between the template and the data. Mathematically, the SNR (ρ) [20] is defined as

$$\rho = \left(4\,\mathfrak{Re} \left(\int_{f_{min}}^{f_{max}} \frac{\tilde{d}(f)\tilde{h}^*(f)}{\mathcal{P}(f)} df \right) \right)^{1/2} , \tag{3.12}$$

where $\tilde{h}^*(f)$ and $\tilde{d}(f)$ are respectively the complex conjugate of the Fourier transform of the template and the Fourier transform of the data. $\mathcal{P}(f)$ is the noise power spectral density. Here, f_{min} is the minimal frequency in the detector sensitivity band and f_{max} is the maximum frequency considered, typically the Nyquist frequency, i.e. half of the sampling frequency.

The SNR represents how well a typical template h matches the data d, which is the addition of a GW signal and noise. A matched-filtering-based search finds the template that maximises the SNR and is optimal for Gaussian stationary noise and an exactly known signal. For this type of noise, when the signal does not contain a GW, the SNR fluctuates around a mean value. Nonetheless, if a GW enters the detector, the SNR increases and when that exceeds a predefined SNR threshold, a candidate trigger event is recorded.

However, the noise from the detectors is neither Gaussian nor stationary, making the search more complex. For example, glitches (spurious noise variations in the detector band) can occur and lead to a peak in SNR which can mimic a GW trigger. To avoid the detection of noise, the matched filtering-based pipelines often require the detection to be in coincidence in different detectors. Additionally, more elaborate tests also exist, such as the χ^2 test [173], that can downrank the noise artefacts in the final candidate lists. The confidence one has about the detection is also often translated by a false-alarm-rate that gives the frequency at which noise fluctuations lead to the same ranking statistic value. This ranking is a multi-variate statistic that includes different statistics such as SNR and χ^2 [79].

The optimal SNR is obtained when the template is matched with itself [20]:

$$\rho_{opt} = \left(4\,\mathfrak{Re} \left(\int_{f_{min}}^{f_{max}} \frac{|\tilde{h}(f)|^2}{\mathcal{P}(f)} df \right) \right)^{1/2} .. \tag{3.13}$$

This value represents the loudness of the signal in the detector.

In the context of pre-merger analysis, only part of the inspiral is considered, and the loudness of the signal is not represented anymore by the optimal SNR. Instead, we define it by a partial

inspiral SNR (PI SNR), which has the same definition as the optimal SNR in Eq. (3.13), but where the template is now the partial template containing only the early partial inspiral (h_{PI}). In the frequency domain, it is equivalent, for a given waveform, to replace the f_{max} in Eq.(3.13) by the maximum frequency reached by the template in the part of the inspiral considered. Typically, this frequency will be below 50 Hz (instead of thousands usually), reducing the value of the integral.

The SNR increases more rapidly around the late inspiral and the merger than during the early inspiral. Fig. 3.11 shows this behaviour as it represents the value of the PI SNR as a function of the fraction of the signal that is taken into account.

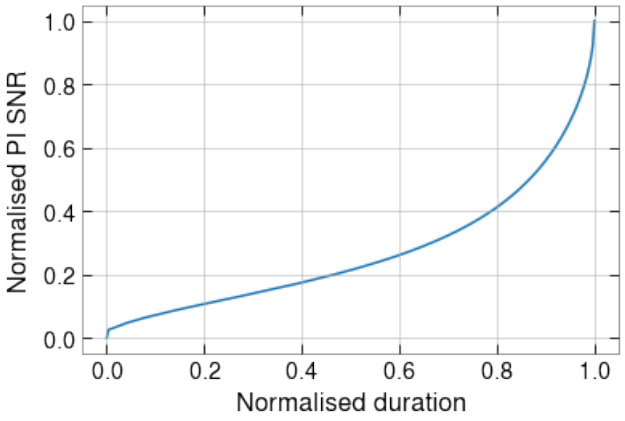

Figure 3.11: Evolution of the PI SNR as a function of the duration of the early inspiral for a BNS with component masses of 1 $M_\odot$. On the vertical axis, the PI SNR is normalised by the optimal SNR, and on the horizontal axes, duration of the early inspiral is normalised by the duration of the full template.

The behaviour of the PI SNR comes from the relation between frequency and time. At the lowest order in velocity, one finds:

$$f(t) = \frac{1}{\pi}\left(\frac{G\mathcal{M}_c}{c^3}\right)^{-5/8}\left(\frac{5}{251}\frac{1}{(t_m - t)}\right)^{3/8}, \tag{3.14}$$

where $f(t)$ is the frequency at time t, $\mathcal{M}_c$ is the chirp mass defined in terms of the component masses m_1 and m_2 of the system: $\mathcal{M}_c = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$ and t_m is the time of the merger. This behaviour is illustrated in Fig. 3.12, which shows the full and partial templates and their frequency evolution.

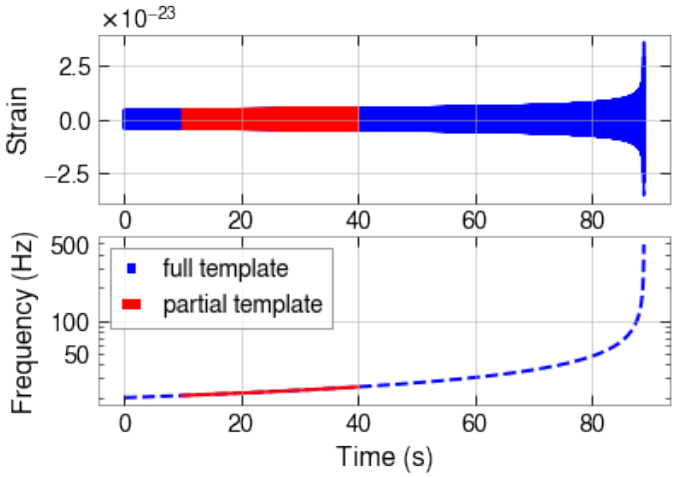

Figure 3.12: The top panel represents an intermediate BNS template where the component masses are $m_1 = m_2 = 2M_\odot$. The bottom panel shows the frequency evolution of the template with time. In both plots, the inspiral part considered for our ML-based approach is coloured in red.

BNS categories

The duration of the observable CBC signal depends mainly on the chirp mass $\mathcal{M}_c$. Indeed, at the lowest order in velocity[4], the duration of the signal is given by [174]

$$\tau(s) \simeq \frac{3}{(\mathcal{M}_c/M_\odot)^{\frac{5}{3}}}\left[\left(\frac{100\text{ Hz}}{f_{low}}\right)^{\frac{8}{3}} - \left(\frac{100\text{ Hz}}{f_{high}}\right)^{\frac{8}{3}}\right], \tag{3.15}$$

where f_{low} is the lowest frequency in the detector sensitivity band and f_{high} is the highest frequency reached by the binary (approximated by the frequency of the innermost stable orbit). From this expression it is clear that, for a fixed lowest frequency f_{low}, if the chirp mass $\mathcal{M}_c$ increases, the duration of the detectable signal shortens.

Furthermore, at the lowest order in velocity, the SNR also has a simple expression [175]:

$$\rho \simeq \frac{1}{2}\sqrt{\frac{5}{6}}\frac{1}{\pi^{\frac{2}{3}}}\frac{c}{D}\left(\frac{G\mathcal{M}_c}{c^3}\right)^{\frac{5}{6}}\sqrt{I}\, g(\theta,\phi,\psi,\iota). \tag{3.16}$$

In this expression, c is the speed of light, D is the luminosity distance, G is the gravitational constant, $\mathcal{M}_c$ is the chirp mass, I is the frequency integral

$$I = \int_{f_{min}}^{f_{max}} \frac{(f')^{-7/3}}{P(f')}\, df', \tag{3.17}$$

and $g(\theta,\phi,\psi,\iota)$ is a function that depends on the orientation of the orbital plane and on the sky position through the antenna pattern of the detectors [176]. From Eqs. (3.15) and (3.16) one can see that if the chirp mass decreases, the optimal SNR of the signal decreases while its duration

[4]In the early inspiral, the strong field effects and the velocities are rather small, which means that the expression derived for the lowest order in velocity approximates well the behaviour of the binary systems.

BNS	light	intermediate	heavy
$\mathcal{M}_c$ ($M_\odot$)	1.13 - 1.56	1.56 - 2.09	2.09 - 2.61
f_{low} (Hz)	20	20	20
Duration (s)	100 - 180	65 - 100	45 - 65
OTW (s)	80	50	30
Fraction of signal	0.44 - 0.8	0.5 - 0.77	0.46 - 0.66
Early alert before merger (s)	20 - 100	15 - 50	15 - 35

Table 3.1: Summary of the CBC merger types for the different CNNs. A different OTW is considered for each category because of the difference in duration of the signals depending on the component masses. The fraction of signal corresponds to the time passed by the signal in the frame w.r.t the total duration. The early alert time is the duration between the end of the OTW and the end of the signal. For each OTW, the minimum and maximum component masses are 1 and 3 $M_\odot$.

increases.

As we can observe in Fig. 3.11, the PI SNR depends on the fraction of the signal considered, as well as on the highest frequency reached within the observation time. Therefore, observing the signal for a longer time would lead to a higher PI SNR, making the signal easier to detect. However, we also want to detect the signal as early as possible in order to have an efficient pre-merger alert system. This leads to a trade-off in our method, as we want to have a high PI SNR, but also prompt detections.

Since we know that the time evolution of the amplitude of the signals will be different depending on the masses, we split the BNS set into three different categories: light, intermediate and heavy BNS. For each of these categories, we use a different observation time window (OTW), meaning that we train the networks on a different length of data. Hence, our algorithm consists of 3 CNNs, one for each category and input size. Note that the OTW is a hyperparameter that will be tuned in a later work. A discussion of the influence of this parameter will be discussed in section IV.

Table 3.1 summarises the characteristics of the different categories, which are classified according to their chirp mass $\mathcal{M}_c$. In order to give an intuition for the masses of the objects present in each category, we give the highest and lowest chirp masses of each category and the component masses for an equal-mass system[5]. For each of the categories, in addition to the constraint on the chirp mass, we also restricted the individual component masses to be between $1\,M_\odot$ and $3\,M_\odot$, which corresponds to a broad mass range for neutron stars. Note also that spin effects are absent at this order in $\frac{v}{c}$, so that we considered only non-spinning BNS.

Data set generation

The inputs of the neural networks are 1-dimensional whitened time series, made of Gaussian noise generated from the design sensitivity PSD of Advanced LIGO (aLIGO) with a GW added in some cases. Indeed, the network is trained as a classifier between an *event class* (noise + template) and a *noise class* (only noise). The GW data analysis and generation has been performed with the

[5]Non-equal mass systems were also considered during the training and testing of our networks.

PyCBC package [122].

We start by generating 120 seconds of coloured Gaussian noise. Then, a non-spinning BNS waveform is injected into it. The approximant used is *SpinTaylorT4* [177] and it is generated with a minimum frequency of 20 Hz.

By default, we employ the optimal sky localisation considering only the plus polarisation of the GW aligned with the arms of the interferometer to generate training, validation and testing sets. Note however that when we later will test the performance of the networks with realistic BNS populations, similarly to [174], the sky location will not be the optimal one anymore.

Since our objective is to train the networks on the early inspiral part of the waveforms, we select the desired OTW for the generated strain and its PI SNR is computed. In Fig. 3.13, we plot the waveform embedded in Gaussian noise. The vertical red lines represent the portion of the strain in the OTW. Note that the frame we use is always starts at the beginning of the 120 second injection it is taken from. Finally, we whiten the stretch of data under consideration and normalise its amplitude by dividing all the points by the maximum amplitude in absolute value. Therefore, all points are in $[-1, 1]$. For the frames containing a GW, the event characteristics, such as the distance, are chosen so that the PI SNR distribution covers a wide range.

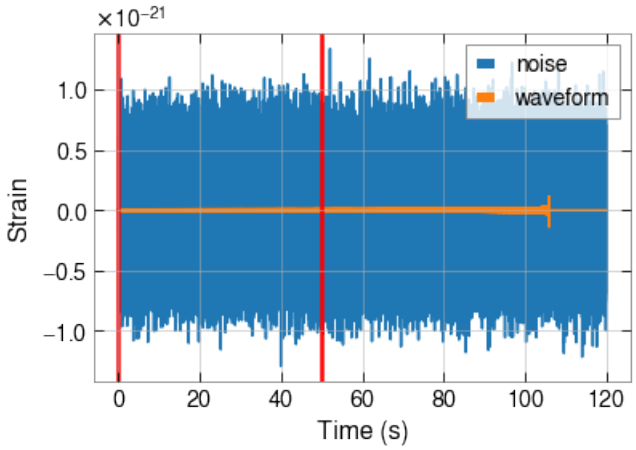

Figure 3.13: Representation of the noise and the injected waveform before the whitening. The CBC signal corresponds to a BNS where both component masses are $1.8\,M_{\odot}$ and the binary is placed at a luminosity distance of 100 Mpc. At the time of training and testing of the CNNs, we do not pass this full frame to the network, but only the first 50 s (denoted by the two red lines), which is the chosen OTW length for this BNS category.

3.3.3 Methodology

Architecture of the CNN

The goal of this search is to perform a binary classification task, to distinguish the OTWs with GW signals from those without, with a short CNN, similarly to [146, 147, 166]. The CNNs were implemented with the *PyTorch* package [178]. We use cross entropy as the loss function and ADAMAX as the optimizer, which is a variant of ADAM, based on the infinity norm [132]. Several

hyperparameters such as the learning rate, the batch size, the numbers of layers, the kernel size, were tested, but in this work we only report on the ones that provided the best performance.

After several trials, we found that the best performance with the minimum computational cost was acquired for 5 convolutional layers. It was found that a bottleneck structure, i.e. starting with a large kernel size, making it smaller in the middle and enlarging it again afterwards, yielded the best results. We represent the best-performing architecture in Fig. 3.14 and in Appendix A. The output of the network is a probability vector which contains the probabilities of the template belonging to the event class, where the event is present into the noise, or to the noise class otherwise. The classification task is performed according to a predefined threshold, which is associated with the False Alarm Probability (FAP).

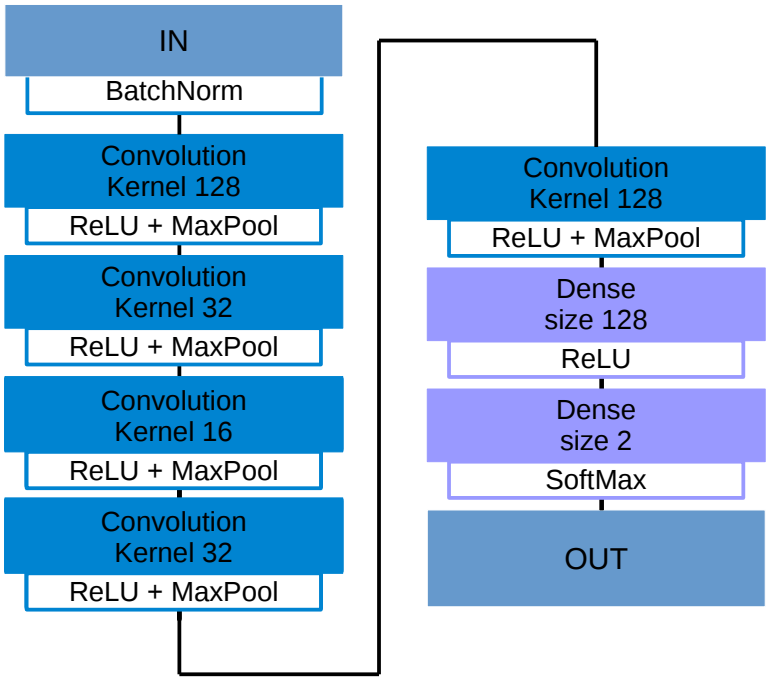

Figure 3.14: Architecture of the best performing CNN for all the categories. From one BNS type to the other, one needs to adapt the input size.

Training and testing of our neural networks

For each category we have a predefined OTW, given in Table 3.1. Due to the varying size of the inputs we perform a binary classification task with a tuned replication of the CNN for each BNS category. The data set is balanced, i.e. it contains 4000 frames of the noise + signal class, and 4000 frames of the noise class, where each of the frames corresponds to an OTW, built as described in the previous section. We employ 80% of the data set for training and 20% for validation. The performance of the network in the training and the validation sets is compared to avoid overfitting. Finally, we test the network with 2000 frames of the noise + signal class, where the events are

chosen to fall into the distance range considered for each BNS category, and 2000 frames of the noise class. More information about the distributions of the data set can be found in Appendix B.

To assess the performance of each neural network, we classify its output for a given data frame into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), according to the standard *confusion matrix* [179].

We also define the True Alarm Probability (TAP) and the False Alarm Probability (FAP) as follows:

$$TAP = \frac{TP}{TP + FN} \qquad FAP = \frac{FP}{TN + FP}. \tag{3.18}$$

The TAP corresponds to the number of noise + signal classified as such over all the number of frames that belong to the noise + signal class, whereas the FAP represents the number of noise frames which are misclassified over the number of frames that belong to the noise class. The performance of the networks will be evaluated based on the TAP for a fixed FAP, which is related to the threshold discussed in the previous subsection.

For this paper, we decided to present all the results for an FAP of 1%. This can be considered to be high if compared with the current GW searches, but we want to insist on the fact that this work is a proof of concept and that our pipeline uses only one detector. We expect that, by considering coincident triggers in N_d detectors, the FAP will roughly go as 0.01^{N_d}. This is an approximation where we assume that the three channels are independent and that, at each instant, each CNN has 1% chance to claim a false detection.

3.3.4 Results and discussion

In this section, we first discuss the performance of the three networks. Then, we report on the results of our method when applied to a realistic population of BNSs. Finally, we discuss a first attempt at curriculum learning, which is promising for the future.

Performance of the CNNs

In Fig. 3.15 we plot the TAP as a function of the distance in Mpc and the PI SNR for each category individually. We see that the network trained on heavier objects is able to reach higher distances. From Eq. (3.16) we can observe the same behaviour, as for smaller chirp masses we need to decrease the luminosity distance in order to keep the same SNR value. We obtain the best performance for the heavy BNS category. The intermediate and low categories have very similar performance, where we see that the 2σ interval of the two categories overlaps when considering the PI SNR. We also note that, since the architecture of the network has been optimised for the heavy category, it is expected that it performs best for this BNS category.

Note that the CNNs are sensitive to the accumulation of the signal. To confirm this, we trained and tested the networks on data with low frequency cut-offs set to higher values than the usual 20 Hz. This is a way to reduce the PI SNR of the injected signal while maintaining the same maximum amplitude. For the testing set, we obtained an 88% TAP for a cut-off at 20 Hz, and 71% for a low-pass frequency at 26 Hz, showing that the CNN is sensitive to the PI SNR for a fixed maximum amplitude. Similarly to matched filtering, a CNN is designed to recognise patterns and,

in this context, the larger PI SNR means that the signal is present for a longer time.

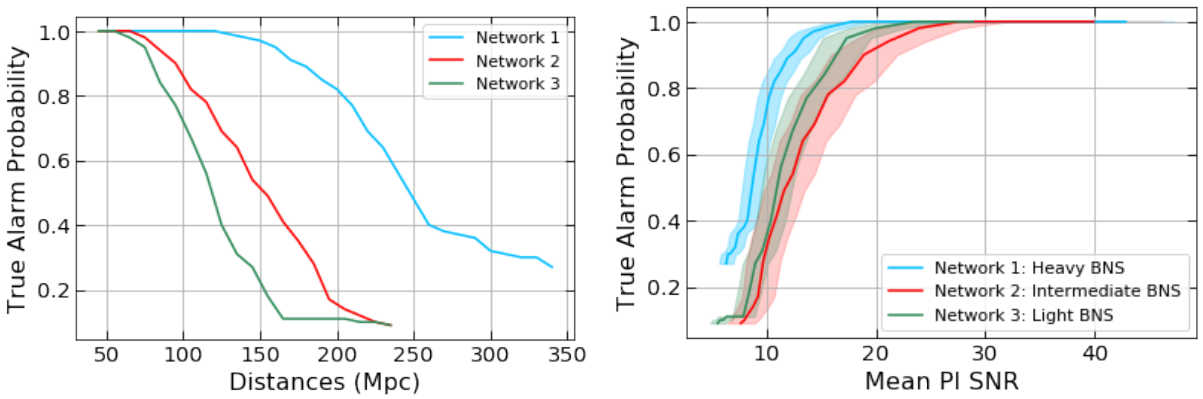

Figure 3.15: The left panel represents the results of the three networks, each trained on its category, as a function of the distance. In the second panel, we compute the mean μ PI SNR and its standard error ϵ, as $\mu(PI\ SNR) \pm \epsilon(PI\ SNR)$ for each distance and a confidence of 2σ, represented by the coloured band. For each graph the FAP is fixed at 0.01.

From Fig. 3.15, we see that Network 1 is able to reach distances larger than 60 Mpc before its TAP has a departure from 100%. As the first BNS detected GW170817 was located at a distance of the order of 40 Mpc, our method will have a high probability to detect similar signals when present in noise at design sensitivity. This means that our method is able to recover realistic signal from Gaussian noise when only the inspiral part is present. Network 2, which is trained on intermediate BNSs, is able to have a better performance at higher distances, which is expected based on the chirp mass - PI SNR relation. Finally, Network 3 has a TAP of 100% even for a distance of 125 Mpc, meaning that the efficiency of detection is still high for distances similar to that of GW190425, the second BNS discovered by the LIGO-Virgo collaboration [108].

We now perform a series of tests to evaluate the influence of the length of the OTW. Indeed, this is an important hyperparameter that represents the fraction of the signal seen for a given event. It needs to be optimised to have as many detections as possible while keeping a long enough delay between the trigger and the merger time. In Fig. 3.16, we show the TAP for Network 3 when using different OTW. As expected, a larger OTW increases the TAP, but is associated with a shorter time lapse before the merger.

We also test whether a network trained on a certain category is able to find signals that belong to a different category. We concentrate on Network 3, which is trained to detect heavy BNSs, and check whether it is capable of detecting intermediate BNSs. For this, we increase the OTW of intermediate BNSs to 30 s, to be able to feed the data set to Network 3. We find that the TAP decreases significantly. Network 2, which is trained to detect intermediate BNSs, yields a TAP of $\sim 68\%$, while Network 3 reaches only $\sim 16\%$. This is also understandable in terms of PI SNR, as the reduction of the OTW duration leads to a decrease in the PI SNR, and we already established that this is the key parameter for detection.

We now compare the time needed for our CNN to analyse one frame with the time needed for matched filtering. When applying matched filtering on a 50 s frame, similar to those passed to the

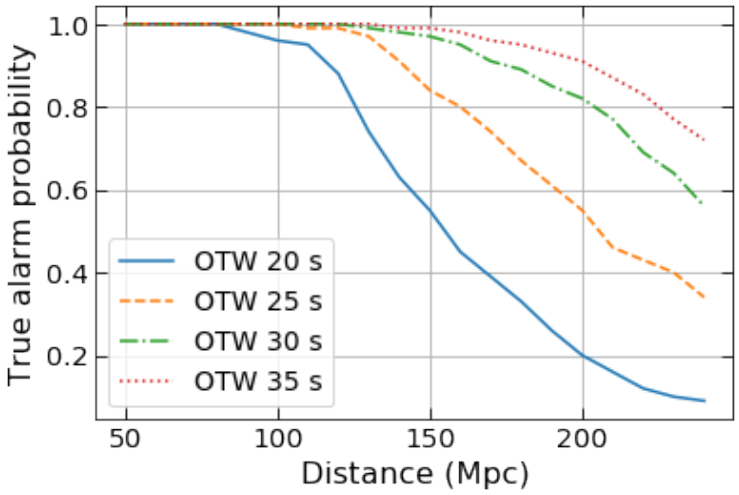

Figure 3.16: Representation of the performance of the CNN trained on the heavy BNS systems for different OTWs. One sees that a longer window gives a higher number of detections. However, it also means the detection happens closer to the merger time. The mean times before merger are 35, 30, 25, and 20 seconds for the 20, 25, 30, and 35 seconds OTW, respectively.

CNNs and with only the optimal template, the computation time is $\sim 0.05\ s$. This is just the bare minimum time needed to get the SNR in matched filtering. In this traditional method, several templates are tested and the trigger is not only assigned an SNR, but also other statistics, such as the FAR. As a consequence, the time to get the final information is longer [21]. Analysing the same frame using our CNN on a *Nividia GeForce RTX 2070 SUPER* GPU, we get the probability of an inspiral to be present in $\sim 0.005\,s$. Therefore, the time needed to analyse the frame and get a prediction probability is improved by a factor of 10.

Test on a realistic population of BNS

In order to have a better grasp on the performance of our networks with respect to matched filtering, we also test them on a simulated realistic population of BNS systems. Therefore, we compute both the optimal SNR and the PI SNR for each BNS with a high-frequency cut-off of 32 Hz, similarly to what was done in [144]. When performing the run with a high-frequency cut-off and the test with our CNNs, we only consider the events that have a full matched filtering SNR higher than 8. This basic computation is performed for the high rate presented in [180]

The cut-off frequency of 32 Hz has been chosen to give results comparable to those in [144], while having in-band times that correspond to the OTWs defined in Table 3.1.

The population synthesis is performed using the code of [174], with minor changes in order to suit our framework. For example, the PSD employed is the same as for the noise generation, the low frequency cut-off used is 20 Hz, and we generate the equivalent of 5 years of data.

One shortcoming of this procedure to generate a realistic population of BNSs is that, although it is fast, it is based on analytical approximations. As a consequence, we do not inject the signals in noise to compute the SNR, and are not able to compute the matched filtering false alarm rate

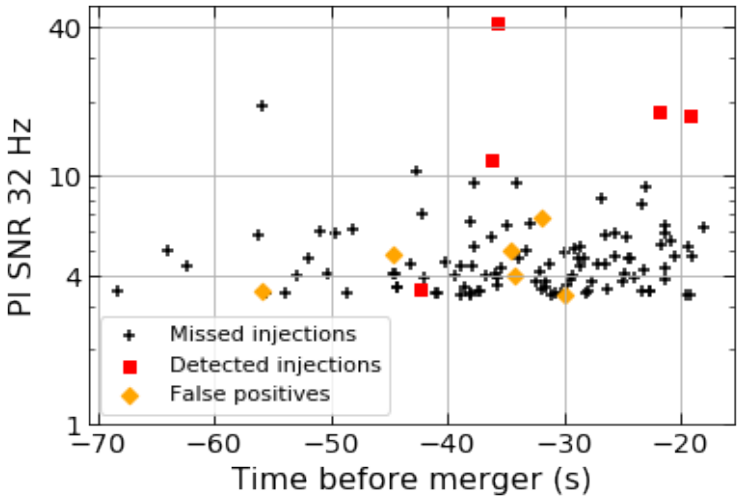

Figure 3.17: The PI SNR for a low-pass filter at 32 Hz for each BNS with a full SNR higher than 8. The black crosses represent the events missed by the CNNs, and the red squares are the events correctly found. The orange diamonds are triggers that correspond to noise fluctuations (false positives).

(FAR) for such frames. So, we cannot use the criterion of Ref. [144] (namely an SNR threshold followed by an FAR) and the direct comparison is non-trivial. Our procedure confirms the difficulty to detect those events with matched filtering methods.

Once we have selected the events based on the analytical approach, we inject them in design-sensitivity noise and pass the frames to the CNNs. Fig. 3.17 represents the events detected, those missed and those misidentified. We also generate the same noise for each event but without injecting the BNS in it. We test our networks on these pure noise frames to highlight the false positives. As shown in Fig. 3.15, the networks detect most of the BNSs which have a PI SNR sufficiently high. We also want to emphasise the fact that matched filtering applied for pre-merger alert also needs the PI SNR to be above a threshold to lead to a trigger. This threshold depends of the framework, and the number of detectors included. We can see that, if one chooses an SNR threshold of 8, our results are comparable to those of matched filtering. Nevertheless CNNs are much faster[6].

A key feature employed in [144] is the network of detectors. Requiring coincident detections in the different detectors helps to remove signals due to noise artefacts. Another advantage is that the signal can accumulate in several detectors simultaneously. Additionally, the sky localisation is found using the data in the three detectors [181]. For a neural network, the input will have a certain number of channels, one for each detector. Then, the input will be convolved through the network, finding relationships between the different channels. This should decrease the FAP of our detector network and enable us to find the sky localisation. This will be explored in a future work.

[6]Here, we neglected the latency needed for the data transfer. It would be, in the worst case, comparable to that of [144].

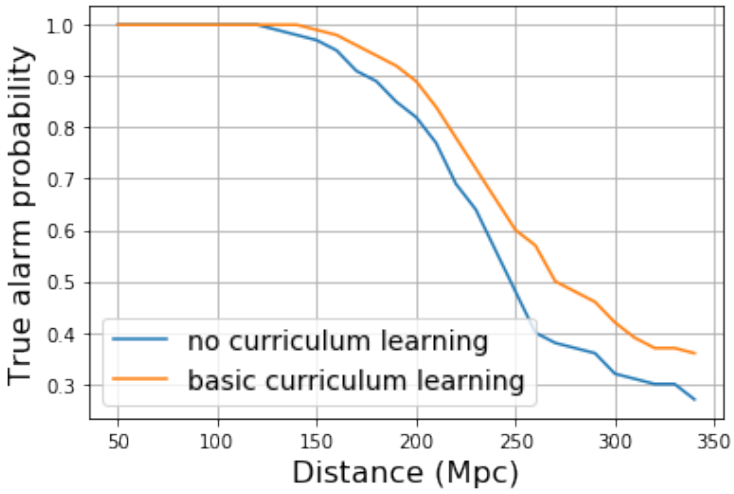

Figure 3.18: Comparison of the TAP as a function of the distance for the GW sources with and without basic curriculum learning for the heavy BNS class. One sees that even a very rudimentary curriculum learning setup helps improve the TAP at higher distances. Note that the blue curve is the same than in Fig. 3.15

Basic curriculum learning exploration

Aside from the architecture, another key factor in the development of DL algorithms is the training procedure. From the population analysis we conclude that the networks see the loudest events, i.e. those with the highest PI SNR in the OTW (or the highest SNR in the detector for the full template). The networks have been trained on a very wide distance range for the events (hence a wide PI SNR range), but it is hard for them to detect smaller PI SNR, as we can see in Fig. 3.15. A way to overcome this obstacle is by training the CNNs with *curriculum learning*. The main idea is to train the network on batches of PI SNR, first on the easy examples, namely the frames with highest PI SNR. Then the difficulty is increased iteratively by decreasing the PI SNR, until the hardest examples are reached, namely the frames with the lowest PI SNR (see [166] or [182] for an example).

With this idea in mind, we generate an extra batch of training data with higher distances and lower PI SNR. Thus, we train on the first data set, store the weights, then train on the newly generated set starting with the previously stored weights. The results of this test can be seen in Fig. 3.18. It can be observed that the TAP increases significantly even if we are using only one extra data batch. As a consequence, we expect the efficiency of our networks to increase substantially once they are trained through the curriculum learning methodology.

3.3.5 Conclusion

In this work, we have introduced a new approach based on short CNNs for pre-merger alert. We have shown that it is possible to detect BNS events when only part of the early inspiral is present in the data stretch under consideration. For this purpose, we have introduced three different neural networks, each trained on a particular range of chirp masses for the BNS systems. Such developments are important in the context of MMA, as the prediction stage is computationally

less expensive and usually faster than for traditional matched filtering. We have also shown that our method is able to recover signals coming from a realistic BNS population simulated at design sensitivity, and compared our detection statistics to those obtained with current matched filtering pipelines. In addition, we also suggested some improvements in the training method, as well as in the structure of our CNNs, to enhance their performance further, leading the way to a pre-merger alert system that would be competitive.

This paper was presented as a a proof of concept and we will continue to build upon this basis to upgrade our networks and get an even better performance. The next steps, which will probably require more complex networks, are the consideration of multiple interferometers and the implementation of sky localisation. Furthermore, curriculum training will be systematically deployed, as this will allow us to train on a bigger dataset with smaller PI SNR. Indeed, the training set currently has a minimum PI SNR around 8. With curriculum learning it will be possible to lower this value. A fourth CNN trained to retrieve the full BNS signal regardless of the category will be built. This will complete the pipeline as the events that are not detected based only on their inspiral would still be found in low-latency. Another recent approach [152] used CNNs to remove some non-linearly coupled noises and detect the early phase of a gravitational wave signal. This further reinforces the case for using CNNs in early alert systems.

It has already been shown in various works that ML-based algorithms can help GW astronomy. In this work, we have shown that it can also be used to solve one of the challenges that will arise in the future, namely the early detection of BNS mergers in the context of MMA. However, we still want to improve the performance and add some features, such as sky location. These are the next milestones which will probably require more complex networks and more advanced training methods.

Acknowledgments

The authors thank Michal Bejger, Chris Messenger and Andrew Miller for their useful comments, as well as Maxime Fays, Vincent Boudart and Chris Van Den Broeck for useful discussions. G.B. is supported by a FRIA grant from the Fonds de la Recherche Scientifique-FNRS, Belgium. J.R.C. acknowledges the support of the Fonds de la Recherche Scientifique-FNRS, Belgium, under grant No. 4.4501.19. M.L., S.C, A.R, and J.J are supported by the research program of the Netherlands Organisation for Scientific Research (NWO). The authors are grateful for computational resources provided by the LIGO Laboratory and supported by the National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459.

3.3.6 Appendix

Appendix A: Details on the architecture of the CNNs

In Section III we briefly described the networks, but in this section we provide with more details about the architecture (see Table 3.2), and the different hyper-parameters fine-tuned for heavy BNS category. The batch size of the training was 40 for networks 1 and 2, and 30 for network 3, due to memory issues. We employ the cross-entropy as loss-function. The optimizer is Adamax

[132] with a learning rate of 8×10^{-5} and a weight decay of 10^{-5}. We trained the networks over 40 epochs. Usually, the validation and training loss drop before epoch 5, as we show in Fig. 3.19, where plot the training and validation loss of the neural network 3, on heavy BNS category. As a consequence, and to avoid over-fitting, we generally use early stopping (around the 12^{th} epoch).

The training of the three networks was done with a dataset of 8000 frames. The data sets are balanced so that half of them correspond to noise and the other half are noise + waveform. The testing was performed with a testing set of 4000 frames, where again, half of them correspond to noise, the other half are noise + waveform.

Layers	Input	Output	Kernel size	Stride	Padding	Dilation	Activation
BatchNorm	1	1	-	-	-	-	-
Conv1D	1	32	128	1	0	1	ReLU
MaxPool1D	32	32	4	4	0	1	-
Conv1D	32	64	32	1	0	1	ReLU
MaxPool1D	64	64	4	4	0	1	-
Conv1D	64	128	16	1	0	1	ReLU
MaxPool1D	128	128	4	4	0	1	-
Conv1D	128	256	32	1	0	1	ReLU
MaxPool1D	256	256	4	4	0	1	-
Conv1D	256	612	128	1	0	1	ReLU
MaxPool1D	612	612	4	4	0	1	-
Dense	X	128	-	-	-	-	ReLU
Dense	128	2	-	-	-	-	SoftMax

Table 3.2: Complete architecture of our CNNs. Between the last MaxPool1D layers we flatten all the channels to obtain an output of dimension 1 and length X (the X depends of the OTW).

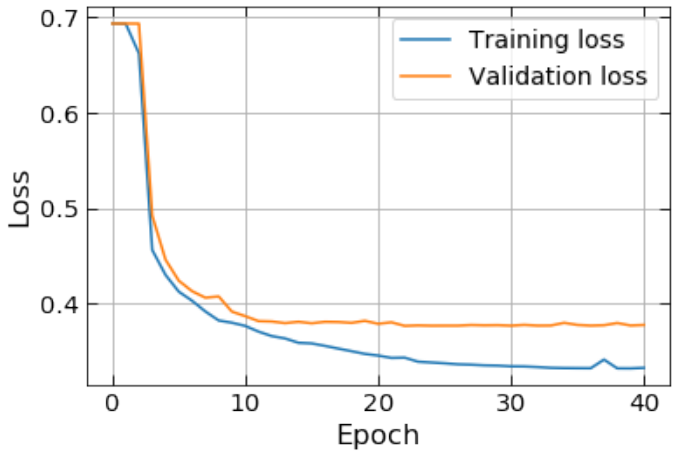

Figure 3.19: The loss variation for the training set and the validation set of CNN 3 as a function of the epoch.

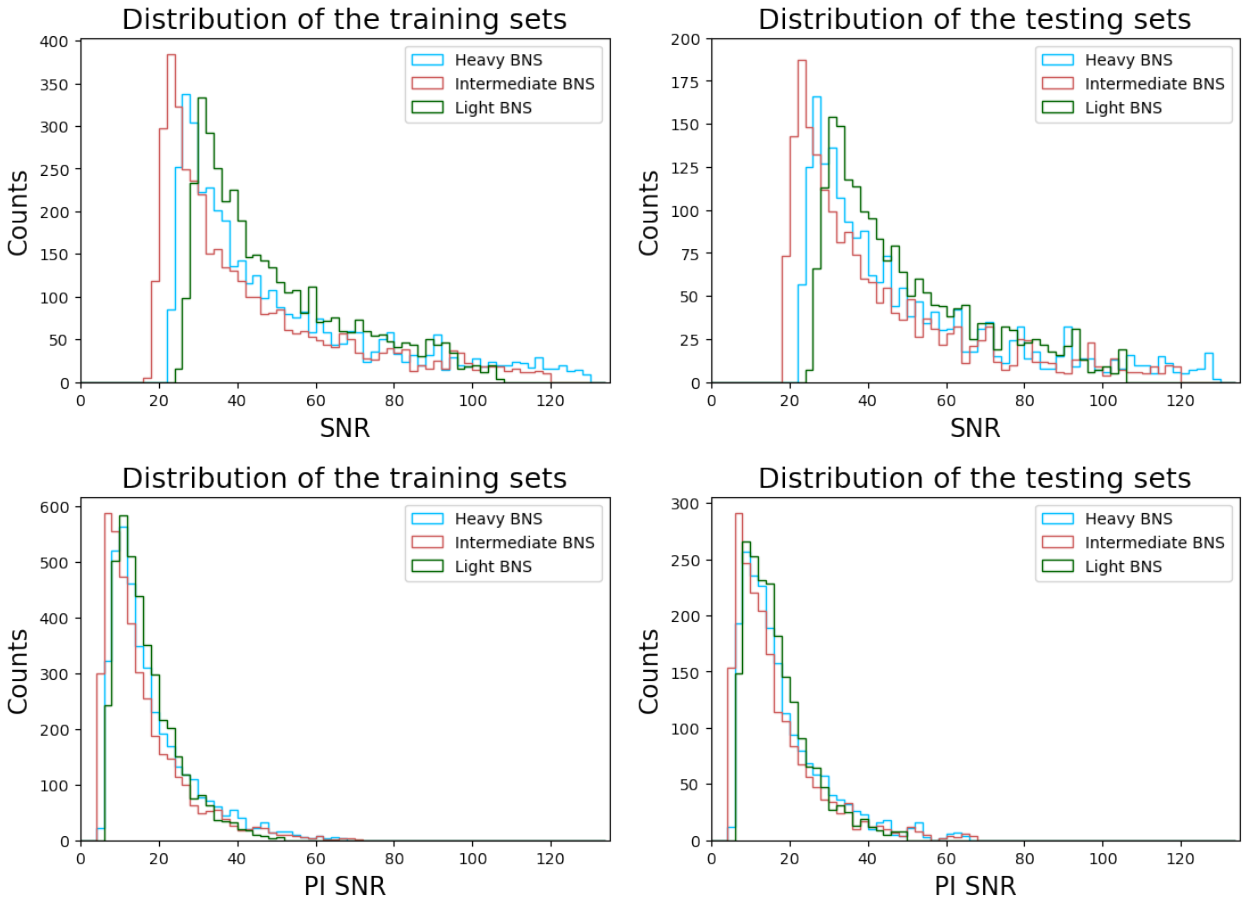

Figure 3.20: *Top left:* Distribution of SNR for the training sets. *Top right:* Distribution of SNR for the testing sets. *Bottom left:* Distribution of PI SNR for the training sets. *Bottom right:* Distribution of PI SNR for the testing sets.

Appendix B: Data distribution

In this section we represent the data distribution with respect to the to SNR and PI SNR. Each distribution employed for training contains 4000 frames, and each distribution employed for testing contains 2000 frames. In Fig. 3.20 we observe that the main difference between the data distributions against SNR or PI SNR is a shift and a decrease in the range of PI SNR due to the removal of the merger from the frames. Indeed, the SNR is in the range $\approx [20, 130]$, while the PI SNR is in the range $\approx [1, 70]$. Therefore, due to the smallness of the PI SNR, the classification task becomes more difficult.

3.4 Article 3

Convolutional neural network for gravitational-wave early alert: Going down in frequency

Grégory Baltus, Justin Janquart, Melissa Lopez, Harsh Narola, Jean-René Cudell

Abstract

We present here the latest development of a machine-learning pipeline for pre-merger alerts from gravitational waves coming from binary neutron stars (BNSs). This work starts from the convolutional neural networks introduced in [130] that searched for the early inspirals in simulated Gaussian noise colored with the design-sensitivity power-spectral density of LIGO. Our new network is able to search for any BNS with a chirp mass between 1 and 3 $M_\odot$, it can take into account all the detectors available, and it can see the events even earlier than the previous one. We study the performance of our method in three different scenarios: colored Gaussian noise based on the O3 sensitivity, real O3 noise, colored Gaussian noise based on the predicted O4 sensitivity. We show that our network performs almost as well in non-Gaussian noise as in Gaussian noise: our method is robust w.r.t. glitches and artifacts present in real noise. Although it would not have been able to trigger on the BNSs detected during O3 because their signal-to-noise ratio was too weak, we expect our network to find around 3 BNSs during O4 with a time before the merger between 3 and 88 s in advance.

3.4.1 Introduction

Multi-messenger astrophysics (MMA) makes use of messengers from different forces of the Universe to provide a wealth of information about various astrophysical processes. From previous investigations it is well-known that the combination of at least two of these signals gives qualitatively different and complementary types of information, capable of probing down to the densest and most energetic regions of cosmic objects, which were hidden from astronomers' sight up until now [183, 184, 185].

In the context of gravitational waves (GW) combined with other astrophysical signals, it has long been suggested that short gamma ray burst (GRB) might be related to binary neutron star mergers [186], which fall in the sensitivity band of second generation ground based-detectors [6, 50, 187]. Several studies investigated the expectations of electromagnetic (EM) follow-up efforts during the Advanced LIGO and Virgo era of compact binary coalescence (CBC) [78, 188]. On August 17, 2017 the Fermi Gamma-ray Burst Monitor [154] announced the detection of a GRB, later designated as GRB170817A [126].

Approximately 6 minutes later, a GW candidate, later re-labeled as GW170817, was registered in low latency based on a single-detector analysis of the Advanced Laser Interferometer Gravitational-wave Observatory (LIGO) Hanford data and disseminated through a Gamma-ray Coordinates Network (GCN) Notice. A rapid re-analysis of data from Hanford, Livingston, and Virgo confirmed a highly significant, coincident signal [127]. An extensive observing campaign was launched across the electromagnetic spectrum in response to the Fermi-GBM and LIGO–Virgo triggers, which led to the detection of the kilonova associated with GW170817, later called AT 2017gfo [189].

In recent times, there has been a sparkling interest in early warning (or pre-merger) alerts of BNS in the field of GW for EM and astro-particle follow-ups [18, 130, 144, 145, 152, 190, 191, 192, 193, 194], since the radiation emitted from these systems enters the sensitive region of the interferometers during the inspiral phase [195, 196]. Assuming all BNS's produce a short GRB with an X-ray, optical and radio afterglow, an LVK network, and a top-hat jet model, Ref. [197] predicts rates for the joint detections of $0.02 - 27$ per year for X-ray band, $0.01 - 190$per year for optical band, and $0.02 - 25$ per year for radio band, respectively, at design sensitivity for a three detector network [197]. It is relevant to note that the large uncertainty is due to the fact that BNS merger rate is not well constrained. The improving sensitivity of second-generation detectors and the even better sensitivity for the third-generation detectors, such as Cosmic Explorer and the Einstein Telescope [159, 160] will lead, via an increase in the signal-to-noise ratio (SNR), to a major improvement in the early-detection and sky-localization capabilities [191, 192, 198]. Another key element to develop MMA further is the design of low-latency pipelines for the production of real-time GW alerts, or even pre-merger alerts. The current state-of-the-art employs matched filtering techniques [78] to perform online analyses via the pipelines GstLAL [21, 22, 77, 79], PyCBC [18, 20, 128, 163, 164, 173, 199], MBTAOnline [165, 200], cWB [201, 202, 203], and SPIIR [129]. We refer the reader to [156] and [5, 8, 204] for a summary of the low-latency efforts carried out by the LIGO-Virgo collaboration during the second and third observing runs. Recent investigations in the GW field have focused on Machine Learning (ML) algorithms, due to their success in different tasks and

domains. The main advantage of ML techniques is their rapidity because most of the computations are made during the training stage. A widely used ML method for pattern recognition is based on convolutional neural networks (CNNs) [205], in the context of GW it has been applied to different tasks such as CBC identification [147, 148, 149, 150, 151], burst detection [182, 206, 207, 208], sky localization [168, 171, 209], glitch classification [210, 211] and synthetic data generation [212, 213]. See [170] for a review on this topic.

ML methods have also emerged as a new tool in the context of early warning [130, 152], allowing us to flag prompt triggers for GW candidates. The final goal of this work is to detect BNS signals before the merger. To do that, we have design a single CNN that takes as input the time-series data from all the online detectors and returns a classification between two classes: pure noise or noise plus inspiral. In this paper, we build on our previous work [130], improving on the techniques previously developed, and testing them on more realistic scenarios: we use real O3 noise, as well as the data from all available detectors. In addition, we retrained our network on predicted O4 noise and give expected efficiencies for this run.

The details of the differences with [130] are as follows:

- the addition of the spin effect to the BNS waveforms;
- a uniform sky location of the injections;
- the injection of simulated BNS signals in simulated O3 noise, real O3 noise and simulated O4 noise;
- a decrease of the minimal cutoff frequency from 20 Hz to 10 – 15 Hz;
- a fixed input-signal duration of 300 s with a sampling frequency of 512 Hz that allows to analyse any BNS signal for all allowed neutron star masses;
- the implementation of curriculum learning [214].

This paper is organized as follows: in Section 3.4.2, the method is explained. Subsection 3.4.2 introduces the definition of the SNR and the partial inspiral signal-to-noise ratio (PISNR) used in this work, as well as the relation between the frequency of a waveform and the time before the merger. The description of the data generation and the training strategy is made in Subsection 3.4.2. The last part of this section, 3.4.2, describes the architecture of the CNN used in this paper. Section 3.4.3 presents the results and the performance of our method in the three types of noise, as well as studying the number of BNS that are expected to be found in advance by our network in O4. Finally, we give our conclusions in section 3.4.4.

3.4.2 Method

Loudness and frequency evolution of the signal

In GW-searches, the matched-filtering SNR (ρ) [18, 175] is used to verify how well a template matches the data. The SNR definition follows that of the FINDCHIRP algorithm [20] as implemented in PyCBC [122]. One first transforms the signal $s(t)$ and templates $h(t)$ to frequency space:

$$\tilde{h}(f) = \int_{-\infty}^{\infty} h(t) e^{-2\pi i f t} dt \tag{3.19}$$

and similarly for $\tilde{s}(f)$. One can define the matched filtering output [175]

$$x(t) = 4\mathcal{R} \int_{0}^{\infty} \frac{\tilde{s}(f)^* \tilde{h}(f)}{S_n(f)} e^{2\pi i f t} df \tag{3.20}$$

where $S_n(f)$ is the one-sided noise strain power spectral density (PSD) of the detector and the * superscript denotes complex conjugation. One can show that this matched filtering output still depends on the phase at a reference time of the signal, for instance at the time when it enters the frequency band of the interferometer. It is possible to minimize $x(t)$ analytically w.r.t. that phase [19, 20] and express the result via the complex $z(t)$:

$$z(t) = 4 \int_{0}^{\infty} \frac{\tilde{s}(f)^* \tilde{h}(f)}{S_n(f)} e^{2\pi i f t} df \tag{3.21}$$

as

$$\min(x(t)) = |z(t)|$$

The variance of $|z(t)|$ is given by:

$$\sigma^2 = 4 \int_{0}^{\infty} \frac{\tilde{h}(f)^* \tilde{h}(f)}{S_n(f)} df. \tag{3.22}$$

The signal-to-noise ratio is then taken to be [19]

$$\rho(t) = \frac{|z(t)|}{\sigma}. \tag{3.23}$$

For a network of N detectors, identified by an index $i = 1...N$, one defines the network SNR as

$$\rho_{net}(t) = \sqrt{\sum_{1}^{N} \rho_i^2(t)}. \tag{3.24}$$

The SNR is a key quantity for the searches based on matched filtering, since it describes the amount of overlap between a template and an unknown signal. In these searches, the strategy is to create a template bank of pre-computed waveforms and use it to calculate the SNR over all the data strain. As a first step, a trigger is created when the SNR reaches a maximum value higher than a given threshold. After that, it undergoes a statistical treatment to be confirmed as a GW candidate [18]. As we are interested in searching for the early inspiral, a more meaningful indicator will be the partial-inspiral signal-to-noise ratio. It is defined as the SNR in which the template h is the partial template that contains only the fraction of the inspiral part of the waveform that our network tries to identify. For more details about the PISNR, and how it evolves depending on the length of the template, we refer to section II. A. of Ref. [130].

At the lowest order in velocity, the frequency f at a time t depends on the chirp mass $\mathcal{M}_c$ of

the system and the merger time t_m:

$$f(t) = \frac{1}{\pi}\left(\frac{G\mathcal{M}_c}{c^3}\right)^{-5/8}\left(\frac{5}{251}\frac{1}{(t_m - t)}\right)^{3/8}. \tag{3.25}$$

So for a given chirp mass, if we say that we can detect an event $(t_m - t)$ seconds before the merger, it is equivalent to say that we detect the signal when the maximum frequency is $f(t)$. Fig. 3.21 represents the time and frequency evolution for a GW.

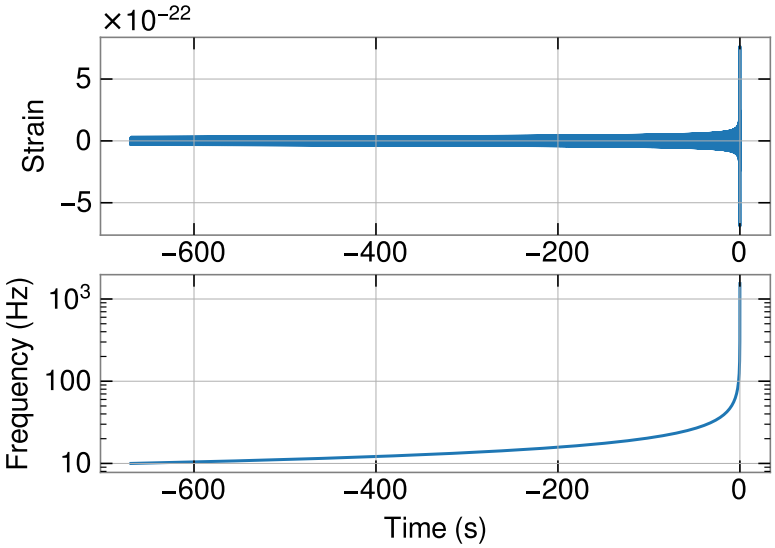

Figure 3.21: The top figure represents a GW signal corresponding to two objects of mass 1.8 $M_\odot$. The bottom figure represents the evolution in frequency for this binary.

Data and training strategies

In our previous work [130], we have shown the possibility to detect the early inspiral of a BNS injected in Gaussian noise. In this work, we want to turn to a more realistic scenario, using real O3 noise. To investigate the difference in performance between Gaussian and real noise, we also inject the signals in colored Gaussian noise generated from the O3 representative PSD. In addition, to assess the performance of our network in future observation runs, we consider colored Gaussian noise generated from the predicted O4 PSD.

The corresponding PSDs are represented in Fig. 3.22. To generate a frame of simulated O3 Gaussian noise, we use the PSDs from [181], provided by PyCBC [122][7]. To obtain data of O3, we directly download the strain of the detectors [5, 8, 204] using the *GWpy* package [215]. To be closer to a real time search, these downloaded strains are the ones recorded in low-latency, meaning that they are not filtered and cleaned as extensively as the final noise [8]. To generate the O4 Gaussian

[7]The PSD used for Gaussian O3 LIGO is *aLIGOaLIGO140MpcT1800545*, the one for Virgo is *aLIGOAdVO3LowT1800545*, both are provided by LIGO and Virgo and implemented via PyCBC [122].

[8]To download the real O3 data, we use the channels H1:GDS-CALIB_STRAIN, L1:GDS-CALIB_STRAIN, V1:Hrec_hoft_16384Hz, and the frame type: H1_llhoft, L1_llhoft, V1Online in *GWpy*.

noise, we use the predicted O4 PSD coming from the observing scenarios [181, 216][9].

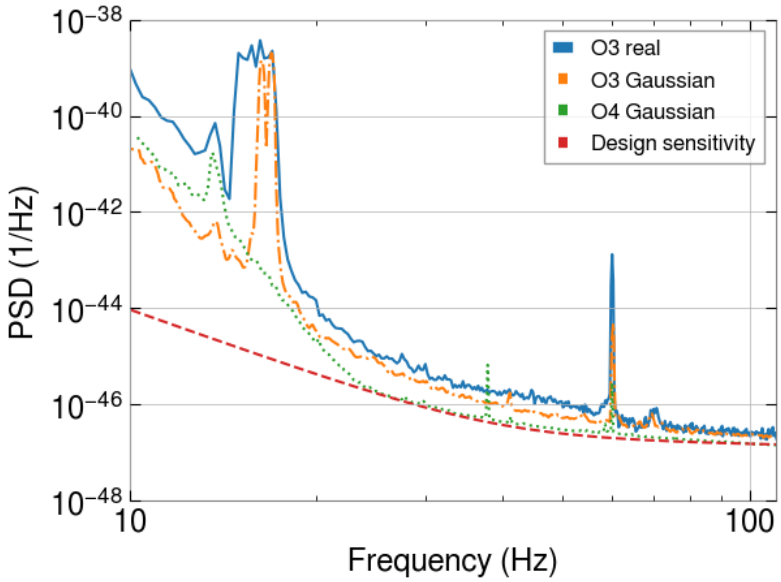

Figure 3.22: Representation of the different PSDs, for the Livingston detector, used to generate the different data sets. We also show the design sensitivity PSD provided by PyCBC [122] used in [130].

Since the problem at hand can be solved as a classification task, we need a data set containing two classes: noise and noise plus inspiral, also known as injections. For the injections, we generate waveforms using the approximant *SpinTaylorT4* [177]. We choose the component masses to be uniformly distributed between 1 and 3 solar masses to cover all the possible BNS systems [25]. The sources are uniformly distributed over the sky, and we also include the spin effects. With these parameters and a lower frequency of 10 Hz, the simulated signal is always longer than 300 seconds. In such a way, the inputs of the network contain only the early inspiral part (see Fig. 3.23 for an illustration). After injecting the simulated signal into the noise the frames are whitened, and we apply a low-pass filter at 100 Hz and a high-pass filter at 10 Hz. For O3 real noise, some significant peaks can appear in the whitened strain due to non-Gaussian effects. In our approach, these effects are vetoed by zeroing them out, see appendix 3.4.5. Afterwards, the final frame is renormalized, making all the values in the frame between -1 and 1. This will be the input data of the network, and we refer to a single sample as a *frame*.

For the training and testing, we choose a distribution in distance such that the distribution in PISNR is an inverse Gaussian with a mean of thirty-five and a scale of one hundred [10]. Despite having a large data set containing one million frames, we have observed a low performance when we decrease the maximum frequency to $\sim$ 25 Hz. This is because the CNN is good to detect a variation of frequency, and for earlier inspiral phase, the signal becomes more monochromatic.

[9]The LIGO and Virgo PSDs used for O4 correspond to the ones shown in Fig. 1 of [181], with the BNS detector horizon at 160 Mpc for the LIGO detectors and the horizon at 120 Mpc for the Virgo detector.

[10]We found that the inverse Gaussian (Wald) distribution fits better our goal. Indeed, this distribution gives a few very high PISNR events that enable the network to start its learning process.

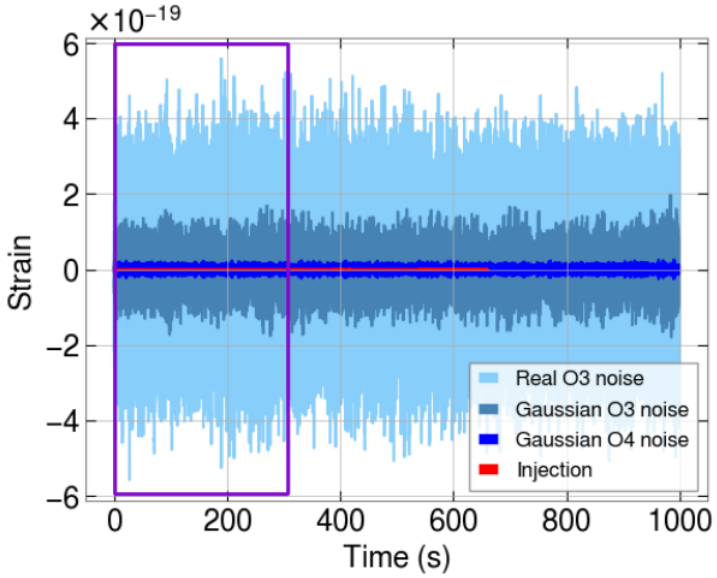

Figure 3.23: Representation of the different types of noises for the Hanford detector used together with an injection similar to GW170817, i.e. with neutron star masses of 1.46 and 1.27 $M_\odot$ [17]. Note that only the part in the rectangle is passed to the network.

To be able to detect events earlier, it is key to decrease the maximum frequency seen by our model. For this aim, we change the training strategy and use curriculum learning [214] as a function of the maximal frequency seen by the network, as it has shown an increase of the performance as a function of the SNR in previous works [182]. The principle of curriculum learning is to train the network first on easier data (on data with a high maximum frequency), and then gradually increasing the difficulty (on data with a lower maximum frequency). The network is then iteratively trained on each training set. To prevent the network from forgetting what it has learned, we keep all the data of the previous steps while adding the new ones. To that effect, we generate five different training sets. The parameter distributions for the injections stay the same, except for the maximal frequency seen by the network. This parameter is now chosen as a Gaussian distribution with a standard deviation of 2.5 Hz, and different mean depending on the data sets. More information about these data sets can be found in table 3.3. Each training set contains 20000 frames and half of them contain an injection. Note that 20 % of each training set was used for validation during the training. For each step, we train for six epochs as it was enough to make the loss converge without facing over-fitting. The use of curriculum learning allows to improve the performance on data set 3, 4, 5 with maximum frequency of, respectively, 30, 25, 20 Hz, while maintaining the performance at higher frequencies.

The training on the real noise data was done in a similar way. Note that we have done the training with noise coming only from O3a, meaning the first half of O3 [8]. We have vetoed the time of the real events from the GWTC-2.1 catalog [8] not to train on them, as most of them were BBH. For all the testing, we used noise coming from O3b, the second half of O3 [5, 204]. During O3 there are times when not all detectors are online. To take this fact into account, when a certain detector is offline, we fill the CNN entry corresponding to the detector with a vector of zeros. In

Data set	Max Freq	Min Freq	Min TBM	Max TBM
Data set 1	40 Hz	12.9 Hz	7 s	44 s
Data set 2	35 Hz	12.8 Hz	10 s	63 s
Data set 3	30 Hz	12.6 Hz	15 s	95 s
Data set 4	25 Hz	12.3 Hz	24 s	115 s
Data set 5	20 Hz	11.7 Hz	45 s	280 s

Table 3.3: Each data set corresponds to a value for the maximum frequency seen by the networks, which in turn leads to a minimum frequency and a Time Before the Merger (TBM). The numbers shown for the maximum and minimum frequency are the mean value in each data set. The maximum and minimum TBMs are the TBM for two objects of respectively 3 $M_\odot$ and 1 $M_\odot$.

this way, our network is able to perform the search regardless of the number of detectors available.

For the training parameters, we use a batch size of 50. For each step of curriculum learning we train for 6 epochs, it was enough to make the loss converge. The learning rate is 8×10^{-5} and the optimizer is ADAMAX with a weight decay of 10^{-5}. ADAMAX is a variant of ADAM, based on the infinity norm [132]. In our previous work [130], we have seen that using ADAMAX leads to a faster the convergence of the training.

We use the weighted cross-entropy loss [217]. At first, we employed the cross-entropy loss, which is standard for classification problems. However, this led to a large number of false positives. To remedy that we decided to weigh this loss [182] by a factor 0.4 for the frames with an injection. This reduces the chances that the network classifies a frame containing only noise as an event, so it reduces the number of false positives. We tried multiple values for the weight and found that for the task at hand a factor of 0.4 translates into a reduction of the number of false positive while maintaining the number of true positive. The duration of the training is about one day on a NVIDIA Tesla V100-PCIE-16Gb GPU.

Description of the network

The architecture of the network is similar to the one in Ref. [130], where multiple trial and errors were made to end up with the architecture. A representation of the neural network is given in Fig. 3.24, we use the *Pytorch* package to create the architecture [178]. The network takes 300 seconds of data for each available detector. In other words, it has three input channels, each corresponding to one of the three detectors (Hanford, Livingston, and Virgo)[11]. It is composed of a batch normalisation layer, followed by 5 blocks composed of a convolution layer, a ReLU activation, and a pool layer. For the convolution, the kernel sizes are successively 16, 8, 4, 8, 16. For the pool layers, the kernel size is always set to 4. The stride is set to 1 for the convolution layers and 4 for the pool layers. After these blocks, we add two linear layers with sizes of respectively 128 and 2 interspersed by a ReLU activation. The final layer is a softmax layer that returns a probability vector.

[11]The Conv1D layer as implemented in PyTorch allow us to give as input any number of channels, see `https://pytorch.org/docs/stable/generated/torch.nn.Conv1d.html` [178]

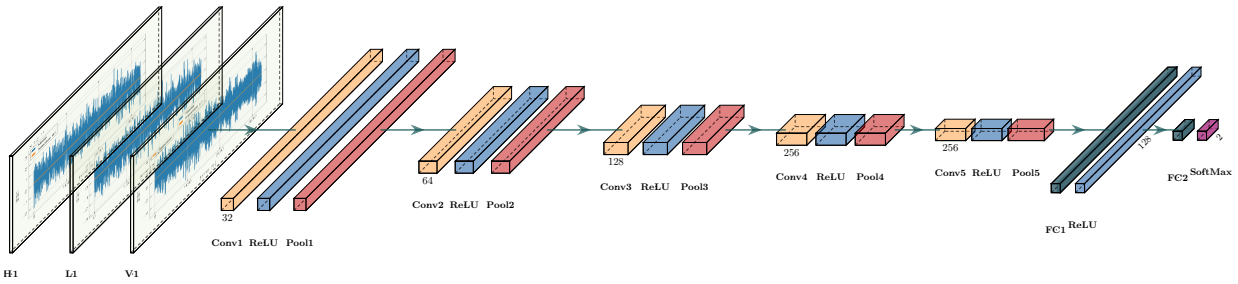

Figure 3.24: Representation of the CNN architecture, the yellow layers are the convolutions, the blue ones are the ReLU activation, the red ones are the pool layers, the purple ones are the dense layers, and the dark purple is the final softmax layer. The number under each layer represents the number of channels.

3.4.3 Results

Performance of the network

After the training, the testing sets come from the same distribution as the training sets, see table 3.3. The other parameter distributions are the same as for the training sets. Each of the test sets contains 4400 frames, half of which are pure noise and half noise plus injection. The total size of the test sets for a type of noise is then 22000 frames.

The efficiency of our network for the different steps of curriculum learning can be seen in Fig. 3.25. We define the True Alarm Probability (TAP) and the False Alarm Probability (FAP) as equation 7 in our previous work [130]. In Fig. 3.25, we represent the three lowest maximum frequencies data set of curriculum learning, as the higher maximum frequencies have performances similar to the 30 Hz data set. For the data sets with a maximum frequency > 25 Hz, an efficiency of 50% is obtained at ~ 15 PISNR, while the efficiency reaches 100% at 30 PISNR. This is not the case for the data set with a maximum frequency of 20 Hz, where the TAP is lower. This is expected since the sensitivity of the detectors becomes worse at lower frequencies, typically under 20 Hz, see Fig. 3.22. In all the figures shown in this work, the FAP is fixed at 1%.

Similarly, we have done the same test for the real O3 noise and the simulated O4 Gaussian noise. The different tests are summarised in Fig. 3.26, where each curve represents the results for the whole test set. In terms of PISNR, the efficiencies for O3 Gaussian noise and O4 Gaussian noise are very similar. However, since the noise floor is lower in the O4 case, the network can probe higher distances in this case. The performance for real noise is a bit worse than for the two Gaussian cases. The network needs a slightly larger PISRN to achieve the same performance. For example, the network needs a PISNR of 20 to have an efficiency of 50% in the case of real O3, whereas it only needs a PISNR of 17 to reach the same sensitivity in the two other cases. Even if some glitches and non-Gaussian features are present in the data, the network is still able to reach a high performance provided that the PISNR is high enough. To be more realistic with a real time search, the noise is downloaded from strains recorded at the time of low-latency. Therefore, it has a low quality, explaining the reduced performance.

After testing the network on independent 300 seconds-long frames, we generate longer frames of

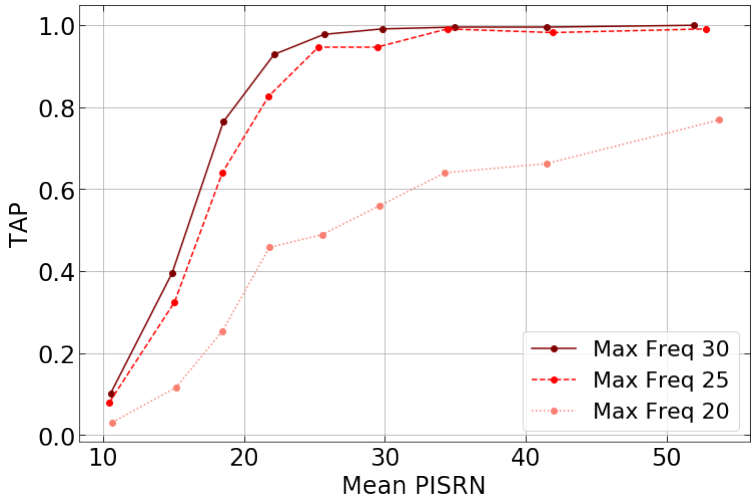

Figure 3.25: The True Alarm Probability as a function of the PISRN for the O3 Gaussian noise case. Each curve represents a different test set with a different maximum frequency seen by the CNN.

1000 seconds, and we inject a complete GW signal into them. Then, we slide a 300 seconds window over the frame, pass the data in the window to the CNN for each step and make a prediction. From one step to the next, the window is shifted by 5 seconds. This is repeated until the full 1000 seconds are covered. Note that the step of 5 seconds is arbitrary and can be reduced, since for a realistic early-alert pipeline the length of the minimum step should be equal to the time required to load 300 seconds of data, pre-process it, and predict it with our network. The deep-learning algorithm is fast and takes about 0.5 s on a CPU and 0.01 s on a *GeForce GTX 750 GPU*, the pre-processing is also fast: about 0.13 s to compute the PSD with Pycbc, 0.01 s to perform the whitening and 1 s to remove the peaks and do the renormalization. The limiting factor is to load 300 seconds of data for 3 detectors with *GWpy* [12], which takes around 2 s on an Intel Xeon E5-2650 v4 CPU. Note that the PSD used for the whitening is computed each time we load the 300s frame. To reduce further this time, one can compute the PSD at regular intervals and use the result for multiple steps.

Fig. 3.27 illustrates the time left before merger when our approach is able to detect the event for the different noise types. Each point contains 1000 frames with a duration of 1000 seconds and each frame has a different noise realisation. In each frame, we inject a BNS signal with component masses similar to those detected for GW170817 [17]. We choose fixed masses to keep the total duration of the signal fixed. The sky position of the signal is changed for each frame. We then slide a 300 seconds window over the 1000 seconds as described above. The process is then repeated for injections corresponding to a larger distance. Fig. 3.27 shows that, for a given distance, the events are detected the earliest in O4 Gaussian noise. It is also interesting to note that the time before merger for real O3 noise and Gaussian O3 noise are similar, even if the Gaussian case is slightly better. An event like GW170817 at a distance of 40 Mpc can be detected by our method 25 s in advance in real O3 noise, 35 s in advance in Gaussian O3 noise, and 50 s in advance for Gaussian O4 noise, showing quite good trigger capabilities in future observations runs.

[12] Using the built-in function *gwpy.timeseries.TimeSeries.get()*

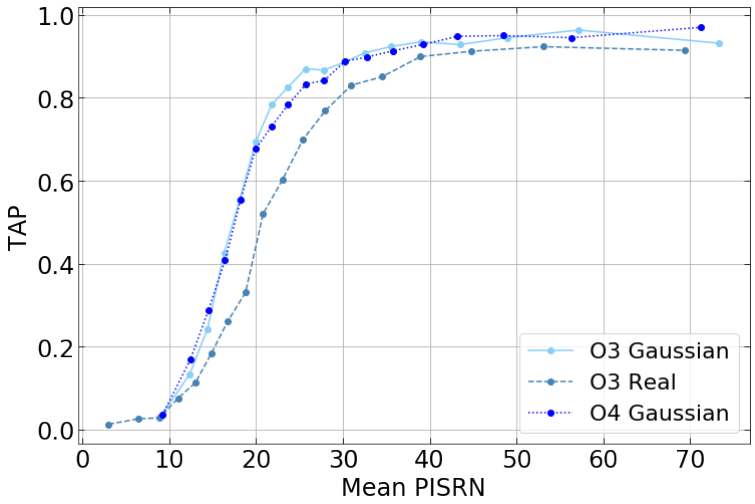

Figure 3.26: The True Alarm Probability as a function of the PISRN for the O3 Gaussian noise, real O3 noise, and O4 Gaussian noise.

For an online matched filtering search, the performance is often evaluated by a false alarm rate (FAR). It represents the probability that a trigger occurs because of the noise for a given period of time [165]. The matched filtering FAR is computed for each event and represents how often the noise is expected to produce a trigger with a ranking statistic value at least as high as the one of the event. With our method, we can not compute such a FAR, but it is possible to compute, what we call the False Positive Over Time (FP_t), which is defined as the number of false positives by a given period of time.

To compute the FP_t, we run our network over the entire O3b data using the same setup as the one described previously. We shift the observation windows by 5 seconds for each step, and veto the times corresponding to events reported in the GWTC-3 catalog [5] and assume that there are no other detectable events in the data[13]. The FP_t is then defined as the number of triggers divided by the total observation time. For O3b, we obtain a FP_t of 277.54 per day, which is too high to be used for online searches. To decrease its values, we can consider that an event is present when our network gives multiple triggers in a row, as shown in Fig. 3.28. If we keep detections with 5 consecutive triggers, the FP_t goes down to 12.31 per day, and it goes to 1.71 per day if we consider 10 triggers in a row. The use of multiple triggers implies larger waiting times before producing an alert, and reduces the time before merger for the detection. For example, considering 5 triggers leads to a delay of 20 seconds as we consider steps of 5 seconds when sliding the window. In the end, this shows that we would need to find a trade-off between the time before merger and the desired FP_t.

Another way to decrease the FP_t is to use coherent triggers between two or more detectors. The training strategy for the network does not make it favor coherent triggers. Indeed, since it is trained for one, two, or three detectors available, it learns to trigger even if only one detector is online. Furthermore, even if more than one detector is online, we do not use a minimum SNR in each

[13]This assumption is reasonable since our network needs relatively high SNRs to detect the inspiral, and the event would therefore have been detected.

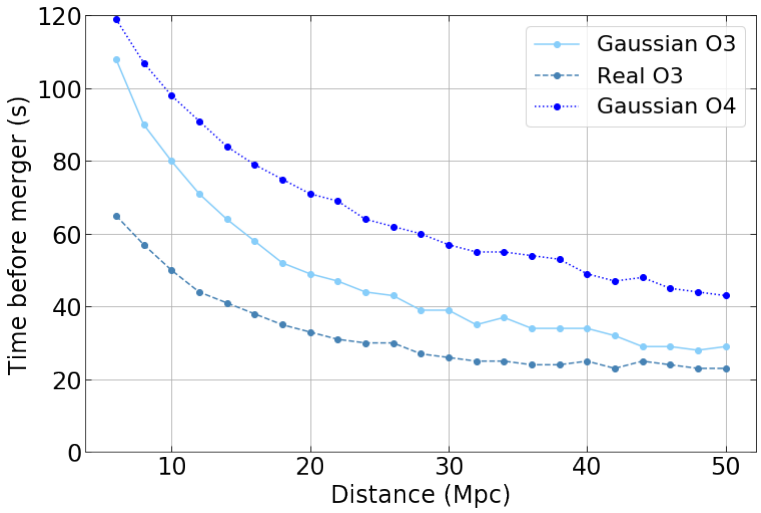

Figure 3.27: The time before merger at which the event is detected as a function of the distance. These curves are made for a BNS with component masses similar to those of GW170817.

detector for the training set. Hence, the network learns to trigger even if only one interferometer is picking up the signal. In the end, this means that as soon as the CNN sees something remotely close to a signal in one of the detectors, it triggers, leading to a relatively high FP_t.

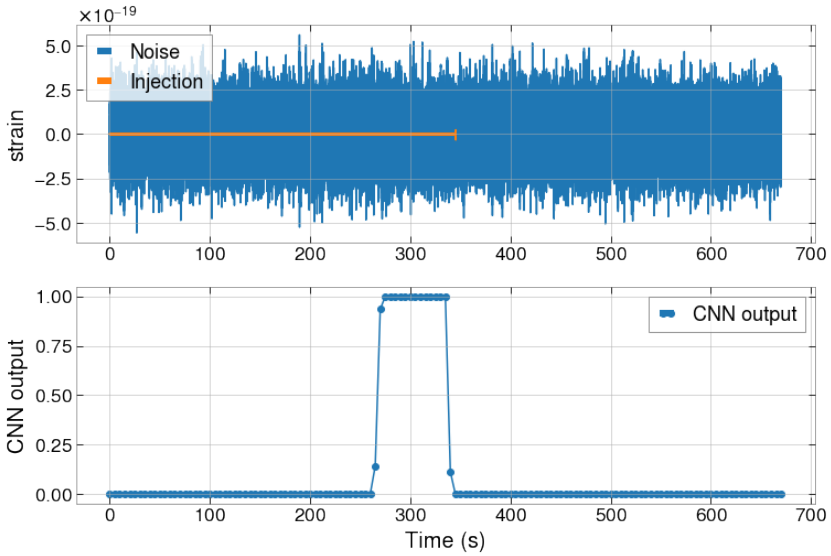

Figure 3.28: *Top:* Representation of a signal and the noise it is injected in. *Bottom:* Representation of the output of the CNN. Each point represents the probability to have an inspiral in the 300 seconds of data. By convention, the time of a point represents the end of the time window. The network does not trigger on the early inspiral because the PISNR is too low. When it becomes high enough, the networks produces a trigger until the injection leaves the frame, giving multiple points with a high probability in a row.

Time before merger	$\mathcal{M}_c$ ($M_\odot$)	network SNR	net PISNR at detection	Maximum frequency
88 s	1.19	71.87	15.32	25.45 Hz
59 s	1.08	63.77	23.43	31.35 Hz
58 s	1.26	53.01	16.63	28.75 Hz
25 s	1.16	28.72	16.31	41.39 Hz
22 s	1.95	64.07	19.8	31.45 Hz
22 s	2.06	54.88	18.01	30.42 Hz
19 s	2.15	30.55	10.37	31.29 Hz
14 s	1.69	31.26	14.42	40.75 Hz
11 s	1.98	27.68	13.63	40.43 Hz
10 s	2.0	28.95	16.28	41.58 Hz
10 s	1.79	25.04	12.9	44.52 Hz
7 s	2.01	20.47	11.87	47.48 Hz
7 s	1.72	34.21	25.71	52.2 Hz
3 s	2.12	28.28	20.68	63.01 Hz

Table 3.4: The time before merger, the maximum frequency seen by the network at detection, the chirp mass of the event, the network SNR, and the PISNR at the moment of the detection for all the detected BNSs in five years of simulated O4 data.

Number of BNS inspirals detectable in O4

To estimate the number of BNSs that our network could detect in O4, we simulate a population of BNSs. It is generated using the method described in [174] and the BNS merger rate is normalised so that the local rate is equal to the median rate given in [180]. The only difference with [174] is that we adapt the detection thresholds and the PSDs to our O4 scenario. We keep the BNS events with a network SNR higher than 13 and discard all the others. This threshold is chosen as we expect our network to find only BNSs that are clearly visible in the detector network and a global SNR of 13 corresponds approximately to an SNR of 8 in each detector.

To have more statistics, we compute the equivalent of 5 years of data, and we consider a duty cycle of 100% for all the detectors. Our simulations predict that, on average, around twenty BNSs per year will have a network SNR over 13 for O4 sensitivity. Our network can detect around three of those BNSs in advance. Fig. 3.29 represents the time before the merger of all the BNSs detected by our network for the 5 years of generated data. Even if our network is able to detected only three events out of twenty, it is important to note that these events are seen in advance and would therefore not be seen at that stage by the unmodified matched filtering searches. Nevertheless, matched filtering pipelines adapted to the early detection of long inspirals are also being developed [145, 144]. Those are also able to detect BNS mergers in advance. Even if the comparison between these works and ours is difficult (partially because of the difference in noise, but also in performance evalution), we can mention that times before merger of these algorithms are comparable to those obtained by our network, ranging from $\mathcal{O}(10)$ to $\mathcal{O}(100)$ seconds. An advantage of these early-warning matched filtering searches is that their FAR is lower than our FP_t (around one per month) but they require more computational ressources during the search as the highest cost for machine learning is moved to the training step. During the search, our method can run on a single GPU or even on a single CPU, while most of the standard matched filtering methods require parallelization on multiple CPUs.

Table 3.4 shows the different characteristics of the detected BNSs. The network can see an event when the net PISNR is between 10 and 25, which is expected according to Fig. 3.26. The time before the merger at which the CNN can detect a signal depends on two factors: a) the

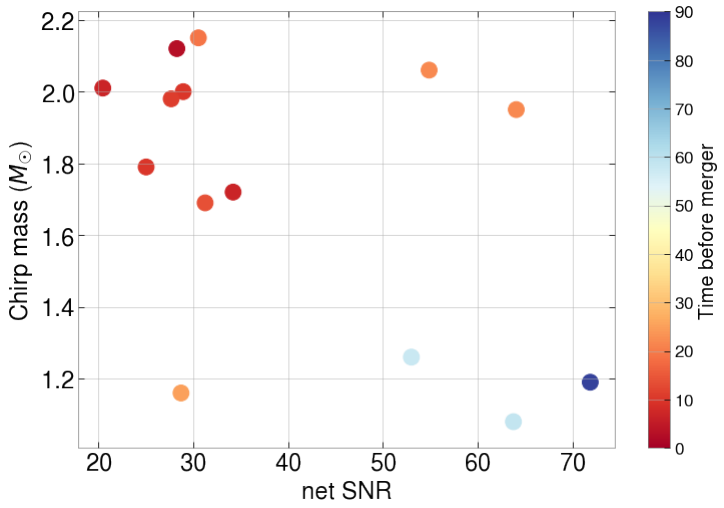

Figure 3.29: The number of BNSs detected in advance by our network for a simulated population of BNSs in five years of O4 data.

network PISNR and b) the length of the signal. The PISNR can be seen as a fraction of the SNR and its exact value depends on which part of the signal we are considering (hence the maximum frequency see by the CNN). Therefore, for a fixed signal duration, if the network SNR is high, the network can detect an event at a lower maximum frequency, corresponding to a longer duration before the merger. However, if we fix the SNR and the maximum frequency while increasing the duration of the signal (for example by decreasing both the chirp mass and the luminosity distance to compensate), the event will be detected with a larger duration before the merger. This behavior is well represented in Fig. 3.29, where events with a light chirp mass and a high SNR are detected the earliest. It also explains why some events with a lower chirp mass can be detected earlier, even if the network SNR is smaller than for other events.

3.4.4 Conclusion

This work builds upon the framework developed in [130]. We implement several upgrades and modifications to the CNN-based pipeline designed to detect the early inspiral phase of BNS events. A major upgrade is the increased duration of the frames passed to the network. That allows us to search for smaller frequencies and opens the door to earlier detections. Another benefit of this increased duration is that we can use a single network to look for all type of BNSs, which was not the case in our previous work. The detection of events for a smaller maximum frequency is not easy and required an adapted training methodology: curriculum learning. We consider realistic observation scenarios, including all the detectors of the LIGO-Virgo network and use realistic noise realisations: O3 and O4 Gaussian noises, as well as real O3 noise. We have also demonstrated that even in the real O3 noise our network is able to detect GW signals in advance. We expect our network to detect some BNSs in O4, up to minutes in advance if the SNR of the event is high enough. In future works, we will upgrade our method to search for neutron-star-black-hole mergers as well. As discussed in section 3.4.3, we will also develop methods to decrease the FP_t. Finally,

we will investigate a way to infer the sky position with only the early inspiral part.

Acknowledgments

The authors thank Thomas Dent and Srashti Goyal for their useful comments, as well as Maxime Fays, Vincent Boudart, Sarah Caudill and Chris Van Den Broeck for useful discussions. G.B. is supported by a FRIA grant from the Fonds de la Recherche Scientifique-FNRS, Belgium. J.R.C. acknowledges the support of the Fonds de la Recherche Scientifique-FNRS, Belgium, under grant No. 4.4501.19. M.L. H.N., and J.J are supported by the research program of the Netherlands Organisation for Scientific Research (NWO). The authors are grateful for computational resources provided by the LIGO Laboratory and supported by the National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation.

3.4.5 Appendix

Details on the veto of peaks for real noise

After the noise has been downloaded and whitened, large peaks can still be present in the data (see Fig. 3.30). This behavior only appears for real O3 noise, and leads to a problem for the normalisation. Indeed, before passing the data to the network, we normalise them to be between -1 and 1. To do so, we find the maximum absolute value of the strain and divided each point by that value. When a large peak is present, the maximum absolute value is the value of the peak and it makes the rest of the time series too small. That confuses the neural network, and we decided to veto these peaks. The vetoing is done according to the z-score, which is defined as:

$$Z_i = \frac{x_i - \mu}{\sigma} \tag{3.26}$$

where x_i is the value of a point i in the time series, μ and σ are respectively the mean and the variance of the time series. We then compute the standard deviation of the z-score and put all the points with a z-score larger than 5 times the standard deviation to zero, allowing to remove large peaks such as those seen in Fig. 3.30. The normalisation is then done on the vetoed frame.

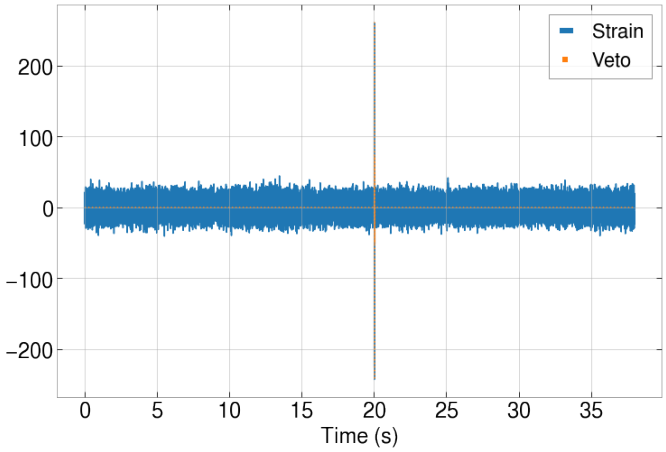

Figure 3.30: The blue curve represents O3a noise after application of the whitening, the low-pass filter, and the high-pass filter. The orange curve shows the part which will be vetoed.

Chapter 4

Sky localization with machine learning

4.1 Sky localization and importance sampling

The work presented in the two previous papers is made for the detection of the early inspiral of BNS but not for the sky localization of the detected events. As mentioned earlier, in order to build an early alert system useful for the MMA, one needs to detect the BNS before the merger, but also to produce a sky localization of the event before the merger is recorded into the detector. In this thesis, we investigate a pipeline for quick sky localization with machine learning. This project focuses on the full signal of BBH. Of course, it is just the first step and these methods can be adapted, in the future, in order to produce a machine learning algorithm that focuses on the early inspiral of BNS. Before presenting the paper, it is necessary to explain why such a pipeline using machine learning is needed and why the traditional tools for sky localization are not sufficient.

4.1.1 Parameter estimation

As mentioned in the chapter 2, GWs are usually detected via a template bank. However, one cannot use it to estimate the parameters of the event. The number of templates in the bank to obtain a very precise parameter estimation would be way too huge. In addition, some parameters could be degenerate and have the same effect on the waveform. The standard way to perform parameter estimation is to use a Bayesian method. The aim is to compute the posterior $p(\theta|d,\mathcal{H})$, which is the probability of having a set of parameters θ, with knowledge of the data d and under the hypothesis $\mathcal{H}$ (this hypothesis is, for example, the fact that there is a GW present in the data) [4, 218, 219, 220]. One can compute it with the Bayes formula:

$$p(\theta|d,\mathcal{H}) = \frac{p(d|\theta,\mathcal{H})p(\theta|\mathcal{H})}{p(d|\mathcal{H})} \tag{4.1}$$

This is a very important equation and this is the core of the parameter estimation. Let us take a moment to explain each factor. The first one, $p(d|\theta,\mathcal{H})$ is called the *likelihood* and represents the

probability of having some data d given the hypothesis $\mathcal{H}$ and the parameters θ. As mentioned previously, under our hypothesis $\mathcal{H}$ that there is a GW in the data, one expresses the data as $d = n + h(\theta)$. One can also show [218] that for Gaussian noise, one can express the probability of having a noise realization n as follows:

$$p[n] = \mathcal{N} e^{-\frac{1}{2}\langle n|n\rangle} \tag{4.2}$$

where $\mathcal{N}$ is a normalization constant, and the inner product $\langle n|n\rangle$ is defined in Equation (2.7). Hence, one can compute the likelihood:

$$p(d|\theta, \mathcal{H}) = \mathcal{N} e^{-\frac{1}{2}\langle d-h(\theta)|d-h(\theta)\rangle} \tag{4.3}$$

This likelihood implies the computation of the waveform $h(\theta)$ and the evaluation of this product $\langle d - h(\theta)|d - h(\theta)\rangle$. This computation can be done for a few points in a reasonable time.

The other factor of the numerator $p(\theta|\mathcal{H})$ is called the *prior*, it represents our belief about the distribution of the parameters before we make the observation. Once this probability distribution is chosen, it takes no time to compute.

It becomes more difficult when one tries to evaluate the denominator $p(d|\mathcal{H})$, called the *evidence*. One can express it as [4, 218]:

$$p(d|\mathcal{H}) = \int p(d|\theta, \mathcal{H}) p(\theta, \mathcal{H}) d^N\theta \tag{4.4}$$

where N is the number of different parameters. This integral could be very difficult to compute because of the high dimensionality of the parameters space. Indeed, one has to sample from the likelihood $p(d|\theta, \mathcal{H})$, which could be tricky. In practice, one needs to use some tricks and special algorithms, like nested sampling [219, 220], to be able to evaluate this evidence and compute the posterior.

There are 11 parameters in the case of CBC events, and they are:

- α: the right ascension
- δ: the declination
- r: the distance
- $t_\otimes$: the arrival time at geocenter
- ι: the inclination angle (see Equation (1.28))
- ψ: the polarization angle (see Fig. 2.1)
- ϕ_c: the coalescence phase

- m_1: the first component mass
- m_2: the second component mass
- $\mathbf{S}_1$: the first spin
- $\mathbf{S}_2$: the second spin two

If one wants to be unbiased, one needs to search for the posterior for all of these parameters (meaning $\theta = \{\alpha, \delta, r, t_\otimes, \iota, \psi, \phi_c, m_1, m_2, \mathbf{S}_1, \mathbf{S}_2\}$). Any subset of these parameters will have some correlations with the others, and evaluating only one subset will lead to some bias. One can split the parameters into two classes: the ones which are intrinsic to the CBC ($m_1, m_2, \mathbf{S}_1, \mathbf{S}_2$) and the others which are extrinsic. The major tools of the LIGO-Virgo-Kagra collaborations to perform the estimation of the parameters are lalinference [219], and more recently Bilby [221]. Depending on the signal, estimating precisely the parameters with these algorithms could take days or weeks. As always, the longer the signal, the longer the time to estimate accurately its parameters.

To be able to produce a rapid sky map, some other packages have been developed. BAYESTAR [222] is a package used by the collaboration to quickly produce sky maps. Unlike Bilby and other parameter estimation algorithms, BAYESTAR does not use as direct input the time series, but rather the output of the matched filtering, the times, the amplitudes, and the phases on arrival at the detectors. It triangulates the position in the sky from the difference of time of the GW which has been recorded in the different detectors. BAYESTAR is still a Bayesian analysis, but, contrary to Bilby, it does not infer all the parameters but only the sky localization. Consequently, it is way faster and takes only a couple of second (4 to 13 s if using 32 threads [222]).

4.1.2 Importance sampling

The goal of machine learning studies is to find a method that will be as fast as BAYESTAR, but also almost as precise as Bilby. The paper that we will present here will describe such an algorithm. The whole idea is to use a neural network to compute an approximate posterior of the sky map, and to feed it to an importance sampling algorithm, in order to improve the likelihood's approximation. Before presenting the paper, we still have to explain what importance sampling is.

Importance sampling is a technique to obtain information on a distribution while sampling another distribution. This is particularly useful when the distribution of interest is difficult or impossible to sample from. Like Monte Carlo methods, importance sampling can be used to estimate an expected value, which is defined with an integral of the type:

$$E_p[f(x)] = \int p(x) f(x) dx \tag{4.5}$$

where $p(x)$ is a probability density function of x and $f(x)$ is a scalar function that we are interested

in. The standard Monte Carlo estimates this expected value with a sum:

$$E_p[f(x)] \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i) \tag{4.6}$$

the x_i used in this sum are samples from $p(x)$. Importance sampling uses almost the same techniques but with an interesting trick that allows us to sample from another distribution $q(x)$, instead of $p(x)$. In the case of importance sampling, one can rewrite Equation (4.5):

$$E_p[f(x)] = \int q(x)\left[\frac{p(x)}{q(x)}f(x)\right]dx = E_q\left[\frac{p(x)}{q(x)}f(x)\right] \tag{4.7}$$

with $q(x)$ which is another distribution that one can choose. Let us write the importance weight distribution as $w(x) = p(x)/q(x)$. One can then rewrite the approximation of the expected value:

$$E_q[w(x)f(x)] \approx \frac{1}{N}\sum_{i=1}^{N} w(x_i)f(x_i) \tag{4.8}$$

this time, the x_i are samples from the distribution $q(x)$, so one can have an approximation of the expected value $E_p[f(x)]$ by sampling another distribution $q(x)$. This technique is particularly useful when it is not possible or very difficult to sample from the distribution $p(x)$. Nevertheless, in order to work, importance sampling needs to choose a $q(x)$ which respects some proprieties. First, one needs to be able to evaluate $p(x)$ at a specific x (at least up to a normalization constant Z), even if the parameter space is too large to sampled in its entirety. In other words, if $p(x)$ can be written as $p(x) = \frac{1}{Z}\theta(x)$, one needs to be able to evaluate $\theta(x)$. Then, the distribution $q(x)$ must be non-zero valued, where the distribution $p(x)f(x)$ is also non-zero. In other words, $q(x)$ needs to be large where $|p(x)f(x)|$ is large. That the tricky part of importance sampling: one needs to choose correctly $q(x)$, so that it respects those two proprieties and it is easy to sample from it.

Another nice propriety of importance sampling is the fact that the weight distribution $w(x)$ will converge towards the distribution of interest $p(x)$ given enough samples. In the article that we will now present, it is this specific propriety which is very important. The aim of this paper is to use importance sampling to approximate the likelihood with the distribution of weights and compute the posterior for the sky position. Yet, to be able to use importance sampling to approximate the likelihood, one needs to find a distribution $q(\lambda)$, in which λ are the parameters of the sky position, which respect the requirements of the previous paragraph. As mentioned, finding such a distribution is not an easy task. The idea is to train a neural network to approximate this likelihood and use it to start importance sampling.

4.2 Article 4

Swift sky localization of gravitational waves using deep learning seeded importance sampling

Alex Kolmus, Grégory Baltus, Justin Janquart, Twan van Laarhoven, Sarah Caudill, Tom Heskes

Abstract

Fast, highly accurate, and reliable inference of the sky origin of gravitational waves would enable real-time multi-messenger astronomy. Current Bayesian inference methodologies, although highly accurate and reliable, are slow. Deep learning models have shown themselves to be accurate and extremely fast for inference tasks on gravitational waves, but their output is inherently questionable due to the blackbox nature of neural networks. In this work, we merge Bayesian inference and deep learning by applying importance sampling on an approximate posterior generated by a multi-headed convolutional neural network. The neural network parametrizes Von Mises-Fisher and Gaussian distributions for the sky coordinates and two masses for given simulated gravitational wave injections in the LIGO and Virgo detectors. We generate skymaps for unseen gravitational-wave events that highly resemble predictions generated using Bayesian inference in a few minutes. Furthermore, we can detect poor predictions from the neural network, and quickly flag them.

4.2.1 Introduction

Gravitational waves (GWs) have immensely advanced our understanding of physics and astronomy since 2015 [46, 223, 224, 225]. These GWs are observed by the Hanford (H) and Livingston (L) interferometers of the Laser Interferometer Gravitational Wave Observatory (LIGO) [6] and the Advanced Virgo (V) interferometer [50]. The collaboration between these three detectors has enabled triple-detector observations of GWs [46], making it possible to do proper sky localisation of their astrophysical sources. This additional detector changes the sky distribution from a broad band to a more narrow distribution [46].

Better early sky localisation capabilities would allow for real-time multi-messenger astronomy (MMA), observing astrophysical events through multiple channels - electromagnetic transients, cosmic rays, neutrinos - only seconds after the GW is detected. MMA is limited to GWs originating from binary neutron star (BNS) and neutron star-black hole mergers. According to current literature, it is unlikely that binary black holes (BBHs) emit an electromagnetic counterpart during their merger [226, 227]. Currently, astrophysicists try to collect the non-GW channels in the weeks after the event. A notable example is GW170817 [17, 137]. This process takes an enormous amount of effort, while the obtained data quality is often sub-optimal. Having all channels observed for the full duration of the event would be a major leap forward. Real-time MMA would enable a plethora of new science, e.g. unravelling the nucleosynthesis of heavy elements using r- and s-processes, more accurate and novel tests of general relativity, and a deeper understanding of the cosmological evolution [139, 140, 228]. As aforementioned, real-time MMA relies on the generation of a skymap and it imposes two limits on the methodology used to obtain one. First, it needs to be swift in order to allow observatories to turn towards an event's origin, preferably only seconds after its observation. Second, the skymap needs to be as accurate as possible since telescopes have a limited area they can observe. Below we present current approaches in generating skymaps for GW events.

Most GW software libraries [122, 219] use Bayesian inference methods - in particular Markov chain Monte Carlo (MCMC) and nested sampling [229] - to construct the posterior over all GW parameters. These methods asymptotically approach the true distribution given a sufficient number of samples [230]. Although theoretically optimal, a chain with around 10^6 to 10^8 samples is required [219] to closely approximate the true posterior distribution for a GW event. Even when using Bilby [221] - a modern Bayesian inference library made for GW astronomy - to perform the inference for a single BBH event, takes hours to produce [167]; BNS events take even longer. Bayesian inference is the most accurate method available for GW posterior estimation, but its run-time is prohibitively long when it comes to MMA.

To overcome the speed limitations of the Bayesian approaches, Singer and Price developed BAYESTAR in 2016 [222], an algorithm that can output a robust skymap for a GW event within a minute. BAYESTAR realizes this speedup in two ways. First, it exploits the information provided by the matched filtering pipeline used in the detection of GWs. The inner product between time strain and matched filters contains nearly all of the information regarding arrival times, amplitudes and phases, which are critical for skymap estimation. Second, Singer and Price derive a likelihood function that is semi-independent from the mass estimation and does not rely on direct computation of GW waveforms, allowing for massive speedups and parallelization. Although BAYESTAR is fast,

its predictions tend to be broader and less precise than those made by Bilby [1].

Deep learning (DL) algorithms have shown themselves to be exceptionally quick and powerful when handling high-dimensional data [231, 232]. Therefore, they are an interesting alternative to the Bayesian methods. Several papers have proposed methods to estimate the GW posterior, including the skymap, using DL algorithms. Examples of such algorithms are Delaunoy et al. [233] and Green and Gair [169]. Delaunoy et al. [233] use a convolutional neural network (CNN) to model the likelihood-to-evidence ratio when given a strain-parameter pair. By evaluating a large amount of parameter options in parallel, they can generate confidence intervals within a minute. The reported confidence intervals are slightly wider than those made by Bilby. A completely different approach was taken by Green and Gair [169]. They showcase complete 15-parameter inference for GW150914 using normalizing flows. They apply a sequence of invertible functions to transform an elementary distribution into a complex distribution [234] which, in this case, is a BBH posterior. Within a single second, their method is able to generate 5,000 independent posterior samples that are in agreement with the reference posterior[2]. A Kolmogorov-Smirnov test confirms that these samples are very closely resemble the samples that are drawn from the exact posterior. Both DL methods are fast and seem to be accurate for the 100 - 1000 simulated GW events they have been evaluated on. However, these methods have a few issues: (1) they are both susceptible to changes in the power spectral density (PSD) and signal-to-noise ratio (SNR), (2) both are close in performance to Bilby but do not match it, (3) they can act unpredictably outside of the trained strain-parameters pairs and, even within this space, they can act unpredictably due to the blackbox nature of neural networks (NNs). Issues (1) and (2) have been addressed for the normalizing flow algorithm in a recent paper by Dax et al. [235], however the robustness guarantees remain behind those of traditional Bayesian inference.

Our method tries to bridge the gap between Bayesian inference and DL methods, allowing for fast inference while still guaranteeing optimal accuracy. It is to be noted that combining Bayesian inference and DL methods has recently gained traction in the GW community, see for example reference [171]. The goal of our algorithm is to restrict the parameter space such that, via sampling, one can quickly obtain an accurate skymap. We use a multi-headed CNN to parameterize an independent sky and mass distribution for a given BBH event. The model is trained on simulated precessing quasi-circular BBH signals resembling the ones observed by the HLV detectors. The parameterized sky and mass distributions are Gaussian-like and are assumed to approximate the sky and mass distributions generated by Bayesian inference. Using the parameterized sky and mass distributions, we construct a proposal posterior in which all other BBH parameters are uniformly distributed. By using importance sampling we can then sample from the exact reference posterior. This implies that we effectively match the performance of Bayesian inference in a short time span, without exploring the entire parameter space. We stress that this work is a proof of concept to show the promises of combining NNs and Bayesian inference. More flexible DL models and BNS events will be considered in future studies.

[1] The GWTC-2 catalog [7] data release provides skymaps made using Bayesian inference methods for recent events. Comparison with the skymaps made by BAYESTAR can be made by looking at skymaps on https://gracedb.ligo.org/latest/.

[2] Throughout this paper, reference posterior is used to imply a posterior that is generated using Bayesian inference.

This paper is organised as follows. Section 2 discusses the model architecture and importance sampling scheme. Section 3 details the performed experiments, including the model training. Section 4 covers the results of these experiments and subsequently assesses the performance of the model and importance sampling scheme by comparing it with skymaps generated using Bilby for a non-spinning BBH system. Conclusions and future endeavours are specified in Section 5.

4.2.2 Methodology

Our inference setup is a two-step method. In the initial step we infer simple distributions for the sky localization and the masses of the BBH by using a neural network. Subsequently, we apply importance sampling to these simple distributions to compute a more accurate posterior. The first subsection describes the role and implementation of importance sampling. The second subsection discusses the neural network setup and our method for distribution estimation.

Importance sampling

High-dimensional distributions in which the majority of the probability density is confined to a small volume of the space are hard to sample from, which results in long run times to get proper estimates when using MCMC methods. A well-known method to cope with this problem is importance sampling. By using a proposal distribution q that covers this high probability density region of the complex distribution p one can quickly obtain useful samples. There are two requirements when using importance sampling. First, the desired distribution p needs to be known up to the normalization constant Z: $p(\lambda) = \frac{1}{Z}\theta(\lambda)$, where $\theta(\lambda)$ is the non-normalized $p(\lambda)$. Second, the proposal distribution q needs to be non-zero for all λ where p is non-zero. Importance sampling can be understood as compensating for the difference between the distributions p and q by assigning an importance weight $w(\lambda)$ to the each sample λ,

$$w(\lambda) = \frac{\theta(\lambda)}{q(\lambda)}, \tag{4.9}$$

where the fraction is the likelihood ratio between the - not-normalized - p and q. The distribution created by the reweighted samples will converge to the p distribution given enough samples [236].

Generating accurate posteriors for GW observations using MCMC is very time consuming, and thus importance sampling is an interesting alternative. Importance sampling requires us to have a viable proposal distribution. Published posteriors for known gravitational waves show that the probability density in the posterior is relatively well confined for both the sky location and the two masses [7]. A Von Mises Fisher (VMF) and Multi Variate Gaussian (MVG) distribution are good first order approximations of the sky and mass distribution respectively, and thus suitable to use as a proposal distribution for importance sampling. We propose to construct this proposal distribution by assuming a uniform distribution over all non-spinning BBH parameters, except for the sky angles which will be represented by a VMF and a MVG distribution for the masses. Assuming that the BBH parameters, sky angles, and masses are independent, our proposal distribution becomes the product of these two distributions. In the next subsection we discuss how we create this proposal distribution using a neural network.

Importance sampling demands a likelihood function for the proposal distribution and the desired distribution. In the previous paragraph we have discussed how we want to create a proposal distribution, we will now focus on the desired distribution p. For the likelihood function of the GW posterior $p(s|\lambda)$ we take the definition given by Canizares et al. [237]:

$$p(s|\lambda) \propto \theta(s|\lambda) = \exp\left(-\frac{\langle s - h(\lambda)|s - h(\lambda)\rangle}{2}\right), \tag{4.10}$$

where s is the observed strain, $h(\lambda)$ is the GW template defined by parameters λ. The inner product is weighted by the PSD of the detector's noise. In practice we use the likelihood implementation provided by Bilby named *GravitationalWaveTransient*.

We now have all the parts needed to discuss how we utilize importance sampling for a given strain s. A trained neural network parameterizes the proposal distribution q for the given strain. The proposal distribution generates n samples, these samples represent possible GW parameter configurations. For each sample we calculate the logarithm of the importance weight,

$$\log w(\lambda) = \log \theta(s|\lambda) - \log q(\lambda) + C, \tag{4.11}$$

instead of the importance weight $w(\lambda)$ itself to prevent numeric under- and overflow. The constant C is added to set the highest $\log w(\lambda)$ to zero, to prevent very large negative values from becoming zero when we calculate the associated likelihood. Since we normalize the weights afterwards the correct importance weights are still obtained. The reweighted samples represent the desired distribution p.

If the proposal distribution does not cover the true distribution well enough, the importance samples will be dominated by only a single to a few weights if we restrict the run-time. We can use this as a gauge to check if the skymap produced by the neural network and importance sampling is to be trusted.

Model

Previous work done by George et al. [146] shows that convolutional neural networks (CNN) are able to extract the masses from a BBH event just as well as the currently-in-use matched filtering. Furthermore, work done by Fan et al. [238] indicates that 1D CNNs are able to locate GW origins. We therefore chose to use a 1D CNN to model both the distribution across the sky for the origin of the GWs and a multivariate normal distribution for the two masses of the BBH system.

The network architecture of this 1D CNN is presented in Figure 4.1 and consists of four parts: a convolutional feature extractor and three neural network heads. These heads are used to specify the two distributions. The following properties were tested or tuned for optimal performance: number of convolutional layers, kernel size, dilation, batch normalization, and dropout. The model shown in Figure 4.1 produced the best result on a validation set.

The convolutional feature extractor generates a set of features that characterize a given GW. This set of features is passed on to the neural heads. Each head is specialized to model a specific GW parameter. The first head determines the sky distribution, the second head the masses, and

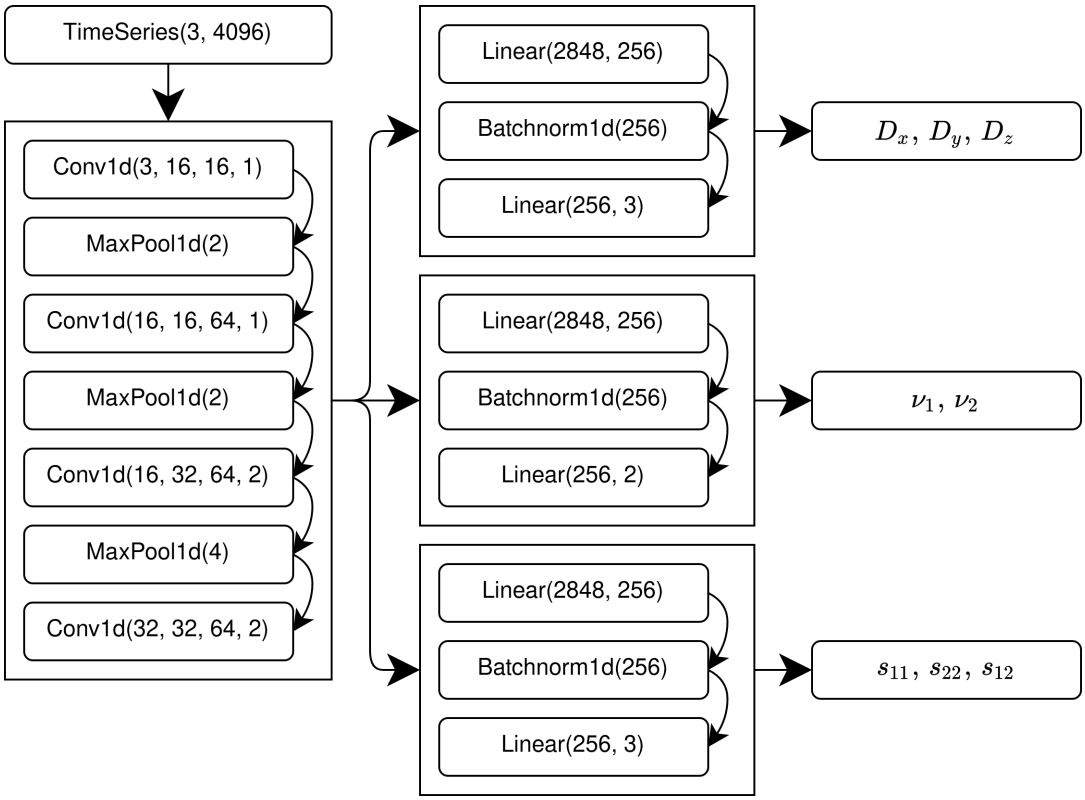

Figure 4.1: A graphical depiction of the convolutional neural network used in this work. After each MaxPool1d and Batchnorm1d layer a leaky ReLU activation function with an $\alpha = 0.1$ is applied. The convolutional part is shown on the left and takes as input a time series of 4096 elements with 3 channels. Conv1D(i, o, k , d) denotes a 1D convolution with i input channels, o output channels, kernel size k and dilation factor d. MaxPool1d(k) denotes a 1D max pooling layer with kernel size k. The output of the convolutions is given to three independent neural network heads. The first head predicts the sky location parameterized as $D = (D_x, D_y, D_z)$, the second head predicts the mean of the masses of the two black holes, and the last head predicts the uncertainty elements of the covariance matrix over the two masses. Linear(i, o) denotes a linear transformation with i input features and o output features. Lastly, Batchnorm1d(i) denotes a 1D batch normalization layer with i input features.

the third head the uncertainty over the two masses. Below we will elaborate on each of these heads and how they characterize these distributions.

The first head specifies the distribution of the GW origin. Since the sky is described by the surface of a 3D sphere, a 2D Gaussian distribution is an ill fit. A suitable alternative is the Von Mises-Fisher (VMF) distribution [239] which is the equivalent of a Gaussian distribution on the surface of a sphere. The probability density function and the associated negative log-likelihood (NLL) of the VMF distribution:

$$p(x|\mu,\kappa) = \frac{\kappa}{4\pi\sinh(\kappa)}\exp\left(\kappa x^T\mu\right) \tag{4.12}$$

$$\begin{aligned} NLL_{\text{VMF}}(x,\mu,\kappa) = &-\log(\kappa) - \log(1-\exp(-2\kappa)) \\ &-\kappa - \log(2\pi) + \kappa x^T\mu\,, \end{aligned} \tag{4.13}$$

where x and μ are normalized vectors in $\mathbb{R}^3$, with the former being the true direction and the latter being the predicted direction. κ is the concentration parameter, which determines the width of the distribution. It plays the same role as the inverse of the variance for a Gaussian distribution. We use this distribution by letting the first head output a three-dimensional vector $D = (D_x, D_y, D_z)$. The norm of D specifies the concentration parameter κ, and its projection onto the unit sphere gives the mean μ, $\kappa = |D|$, and $\mu = D/|D|$. These values together with the true direction x are used to calculate the negative log-likelihood, which is used as the loss function of the first head.

The second and third neural heads specify a 2D multivariate Gaussian (MVG), which describes the possible configurations of the masses. The means ν of the MVG are given by the second head and the covariance matrix Σ is specified by the third head. Given the true values of the masses $y = (m_1, m_2)$ the probability density function and associated negative log-likelihood of the MVG are:

$$\begin{aligned} p(y|\nu,\Sigma) = &\frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \\ &\exp\left(-\frac{1}{2}(y-\nu)^T\Sigma(y-\nu)\right) \end{aligned} \tag{4.14}$$

$$\begin{aligned} NLL_{\text{MVG}}(y,\nu,\Sigma) = &\frac{1}{2}(y-\nu)^T\Sigma^{-1}(y-\nu) + \\ &\frac{1}{2}\log\left(|\Sigma|\right) + \log\left(2\pi\right)\,. \end{aligned} \tag{4.15}$$

The inverse covariance term in the negative log-likelihood can contain imaginary numbers if the covariance matrix is not positive-definite. To ensure that the covariance matrix Σ remains positive-definite, it is parameterized through:

$$\Sigma_{11} = \exp(s_{11}) \tag{4.16}$$

$$\Sigma_{22} = \exp(s_{22}) \tag{4.17}$$

$$\Sigma_{21} = \Sigma_{12} = \tanh(s_{12})\sqrt{\Sigma_{11}\Sigma_{22}}\,. \tag{4.18}$$

The three variables s_{11}, s_{22}, s_{12} are predicted by the third neural head and define the covariance matrix completing the MVG prediction of the masses. The paramerization and implementation of the MVG is based on the work of Russell et al. [240].

By further assuming that the sky distribution is independent of the mass distribution, we obtain a first approximation of the posterior distribution, thereby satisfying the requirements for importance sampling.

4.2.3 Experiments

Experiments were performed on two different fronts: (1) training the neural network followed by the empirical evaluation of its performances on unseen test data, and (2) comparing the neural network model, importance sampling scheme, and Bilby based on several metrics and skymaps. Below we describe the experimental details and justify decisions we made. All experiments were performed on a computer with a 16-core AMD Ryzen 5950X CPU, NVIDIA 3090 RTX GPU, and 64 GB of RAM. Source code is available at `https://github.com/akolmus/swiftsky`.

Training and evaluating the neural model

To obtain strain-parameter pairs for training and validation, we sampled parameters from a BBH parameter prior (see Table 4.1) and generated the associated waveforms using the *IMRPhenomPv2* waveform model [241]. The waveforms were generated in the frequency domain in the frequency band of 20 to 2048 Hz. The duration of the signal is 2 seconds. Subsequently, these waveforms were projected onto the HLV interferometers. We sampled the SNR from a scaled and shifted Beta distribution with its peak set to 15 (see Figure 4.5). The luminosity distance in the prior was set to a 1000 Mpc and scaled afterwards to match the desired SNR. We generated Gaussian noise from the design sensitivity PSD for each detector. Finally, the signal was injected into the noise and an inverse Fourier transform was applied to obtain the strains as time series. This setup allowed us to generate an arbitrary amount of unique strain-parameter pairs, which resulted in every training epoch having a unique dataset.

We applied three preprocessing steps to the data. All time series were whitened with the aforementioned PSDs. Next, the time series were normalized. A normalizer was calculated such that noise-only strains have mean zero and a standard deviation of one. We found empirically that calculating a normalizer for the noise instead of noise plus signal allowed the neural network to converge faster and achieve lower losses. Lastly, to make the mass distribution easier to learn we calculated a shift and scaling factor for the target masses such that all target masses were between -1 and +1. The shifting and scaling were applied inversely to the neural network output during importance sampling to get the correct masses.

The model was trained for 300 epochs with a batch size of 128. During each epoch we drew 500 000 strain-parameter pairs for training and 100 000 strain-parameter pairs for validation. The Adam optimizer [242] was used to optimize weights of the model in conjunction with a cosine annealing scheme with warm restarts [243]. The learning rate oscillated between 10^{-3} and 10^{-5} with a period of 20 epochs; weight decay was set to 10^{-6}. Multiple hyperparameter configurations were tested; this configuration obtained the best performance.

In order to benchmark the trained model, an unseen test set was generated of 100 000 strain-parameter pairs at specific SNR values. The model was evaluated using the mean absolute angular error (maae) and the average 90% confidence area of the predicted VMF distributions.

Applying and evaluating importance sampling

To evaluate the importance sampling procedure, we constructed a slightly simpler test set in which we restricted the maximum spin magnitude to be zero. This was done to limit the Bilby run-time. The importance sampling procedure discussed in Section 2.2 was applied to the first 100 strain-parameter pairs of this test set at three different optimal SNR values: 10, 15, and 20. For each strain-parameter pair we generated 200 000 importance samples. In order to simulate multiple independent runs at various time points for the same strain-parameter pair, we subsampled from these 200 000 importance samples during the experiments.

We ran two experiments to test the convergence of the importance sampling method. In the first experiment, we used the importance sampling scheme as a maximum likelihood estimator. For a given set of importance samples we chose the sample with the highest likelihood and calculated the angle between this sample and the true sky coordinates. In the second experiment, we represented the probability density function of the importance samples by a kernel density estimator and tested how well the resulting density covered the true right ascension. Specifically, we used a Gaussian kernel density estimator[3] to fit the right ascension distribution proposed by the importance samples. The log-likelihood of the actual right ascension was used to measure the quality of the estimated density. We removed a few outliers from the second experiment, by restricting ourselves to only the right ascension the number of outliers was reduced. These outliers had densities that did not cover the true right ascension at all, resulting in extreme negative log-likelihoods which dominate the average log-likelihood. For both experiments we expect the metric to improve as the number of importance samples increases, and to level after a significant number of importance samples indicating convergence.

Generating skymaps

We use Bilby as a benchmark to generate skymaps for the first ten strain-parameter pairs of the test set and for each create a version at an SNR of 10, 15, and 20. To make a fair comparison, the prior given to the Bilby sampler has its spin components set to zero. Moreover, the posterior inference was performed with standard settings, and each run took between 2.5 and 7 hours to complete. During these runs the live points of the sampler were saved every 5 seconds and labelled by the total number of sampled points. These saved points were used to run the two importance sampling experiments for Bilby.

4.2.4 Results

In this section, we first discuss the performance of the CNN. Then, the importance sampling scheme is evaluated using the experimental setup discussed in the previous section. Lastly, we compare

[3]The *gaussian_kde* from the *scipy* python package.

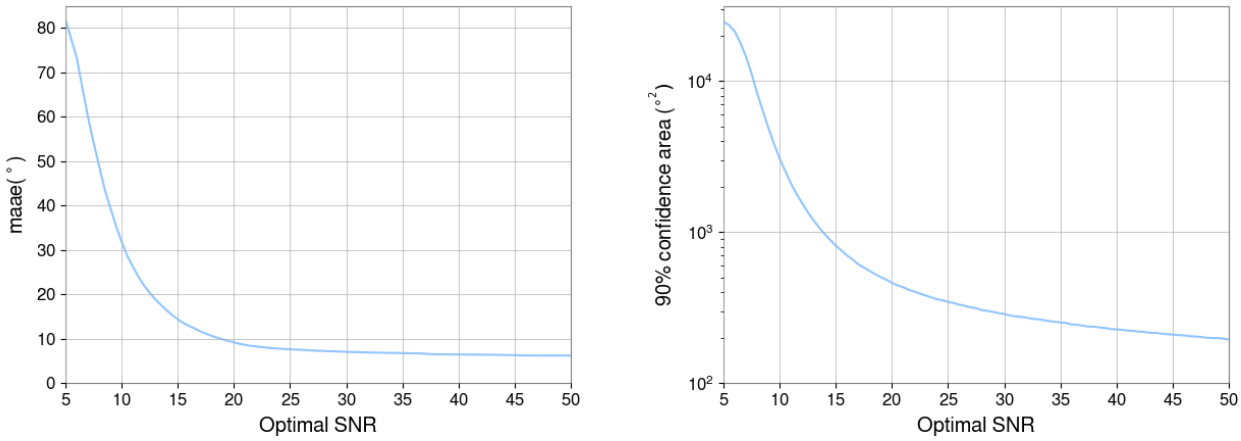

Figure 4.2: Characterization of the neural network in terms of accuracy and certainty over the test. Left: the maae (mean absolute angular error) between the sky angle predicted by the model and the actual sky location as a function of the SNR. Right: the average size of the 90% confidence area, expressed in degrees squared, of the predicted VMF distributions as a function of the SNR.

sky maps generated using only the neural network, importance sampling, and Bilby.

CNN

In Figure 4.2 we summarize the results for the first experiment: the left panel gives the mean absolute angular error (maae) in the sky location and in the right panel we plot the 90% confidence area of the VMF distribution. As expected, as the SNR increases the prediction error in the sky location decreases and the 90% confidence area becomes smaller. The error in the mass prediction is similar to those of other CNN approaches [146], see Figure 4.6, indicating that the setup works well. We do note that the error in the sky location seems to be quite high for SNR < 10 and that it does not converge to zero for high SNR. We can think of two possible explanations for the poor performance at low SNR. First, the detection rate using either CNNs or matched filtering pipelines at an SNR of 5 is less than 40% [146, 149]. At such a low SNR, it is difficult for the model to discern the differences in arrival time at each detector, which explains the slightly better than random predictions for SNR < 7. When we compare our angular error with other CNN approaches [238, 244], the average error seems to be similar. Furthermore, Chua and Vallisneri [245] reported that Gaussian approximations are only accurate for high SNR (SNR > 8) and even then multimodality might arise. Second, the sky distribution can be multimodal. This multi-modality is either due to strong noise or can be due to a sky reflection [219]. For three detectors, there are two viable solutions to the triangulation problem: the true sky location and its reflection. In most cases the amplitude information is sufficient to break the degeneracy between the location and its reflection. However, at certain sky angles this amplitude information does not lift the degeneracy and a multimodal distribution is required. For these angles the model has a 50% chance of guessing the wrong mode and thus having an average angular error of 90°.

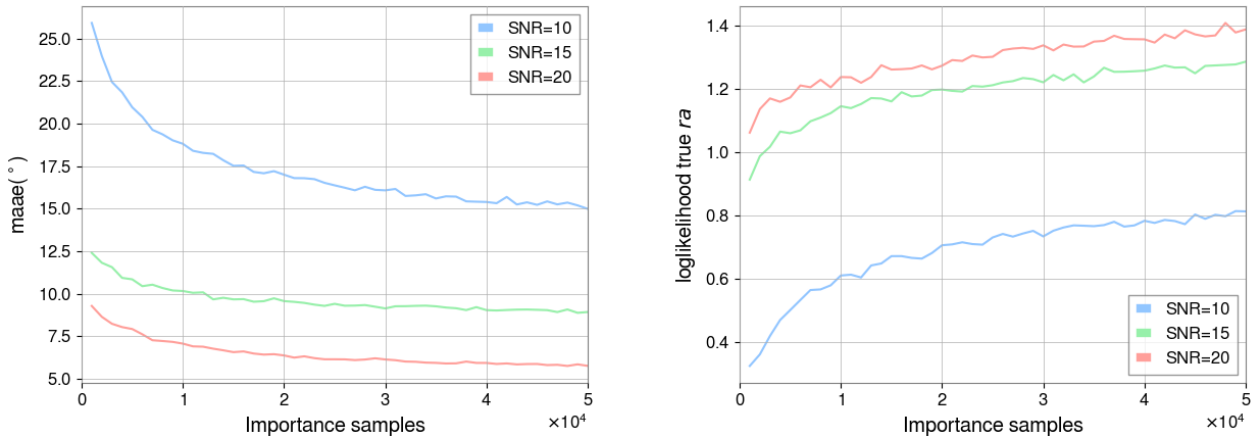

Figure 4.3: Characterization of the importance sampling, with the number of importance samples ranging from a 1 000 to 50 000. The colors represent different SNR values with blue, green, and red being 10, 15, and 20 respectively. Left: the maae of the importance sample with the highest likelihood as a function of the sample size. Right: the log-likelihood of the true right ascension according to the kernel density estimator created by importance samples as a function of sample size.

Importance sampling

The results of the importance sampling experiments are shown in Figure 4.3. The left panel shows the maae as the number of importance samples increases. The right panel shows the log-likelihood of the true right ascension given by kernel density based on a varying number of importance samples. The majority of the convergence in the maae seems to happen within the first 30 000 samples. The slow convergence can largely be attributed to strains for which the model predicted a wide sky distribution. When we compare this to results of Bilby, we see that the maae of the highest likelihood sample for all SNR is always between 1 and 8 degrees. Importance sampling is competitive for an SNR of 20 and is close for an SNR of 15, especially when we consider that in both cases 2 out of the 100 sky distributions were parameterized as the sky reflection. However, importance sampling is not competitive with Bilby in the second experiment. For all SNR values Bilby reports log-likelihoods between 2 and 3, see Figure 4.9, and importance sampling does not reach these values. If we consider runs that show good convergence, i.e. where 90% of the importance weight is not determined by less than ten importance samples, importance sampling also reports log-likelihoods between 2 and 3. In Figure 4.7 we have repeated the kernel density experiment, but only for the well-converged runs. These runs represent 30% of all runs, and almost no SNR < 10 runs.

Generating skymaps

As a final test, we generated skymaps using the neural network, importance sampling, and Bilby on the same signals. Three representative skymaps are shown in Figure 4.4. The skymaps generated by the neural network are significantly more spread out than those generated by importance sampling and Bilby. As we explained in the previous sections, this might be due to the neural network

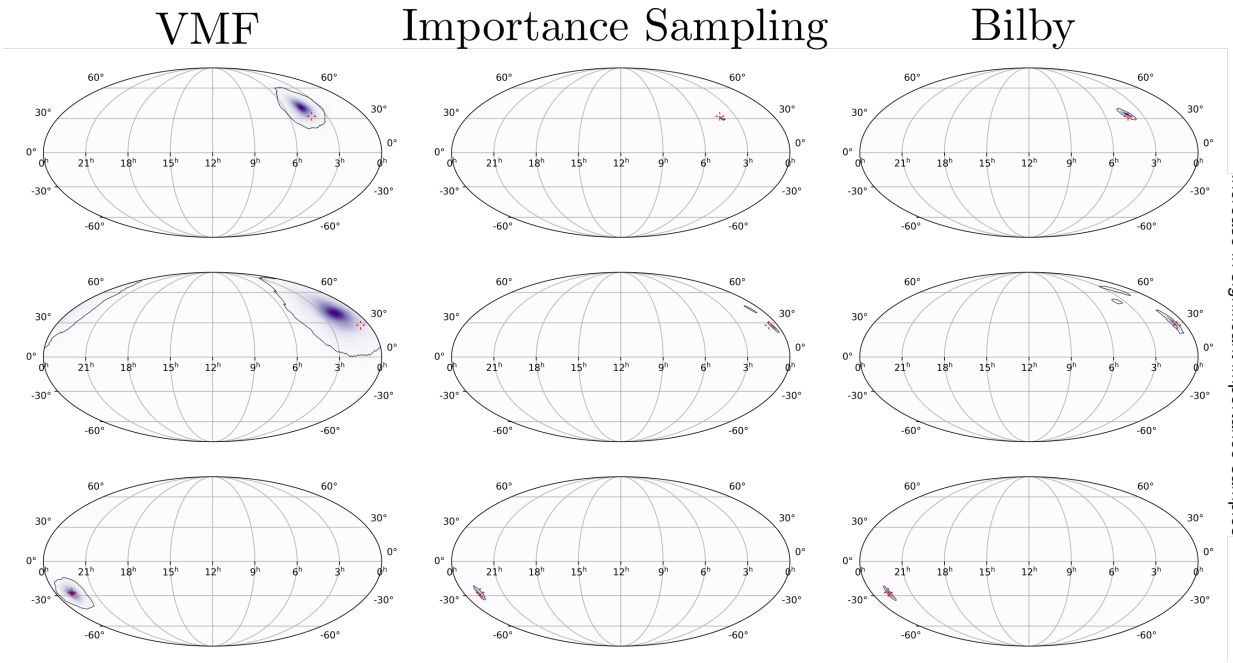

Figure 4.4: Examples of predicted skymaps by our neural network (left), importance sampling after 100 000 steps or roughly 5 minutes of computing time (middle), Bilby at convergence (right). The Bilby runs took at least 3 hours to complete. The true sky location is indicated in red. The shown skymaps were generated for signals with an SNR of 15. The number of significant importance samples, and hence the quality of the sky maps, increases as we go from the top row to the bottom row.

overestimating the uncertainty and having difficulty extracting the exact signal from the detector noise. The skymaps generated by importance sampling and Bilby resemble each other quite a lot, their peak intensities are in the same position and the sky distributions occupy roughly in the same area. However, the importance sampling skymaps are grainy and sometimes do not cover the complete area that Bilby does. As can be seen in the bottom row of Figure 4.4, when the predicted VMF distribution has its peak intensity on the correct position the importance sampling creates better looking sky maps. This improvement is due to the increased number of significant importance samples. These results indicate that a larger number of significant importance samples is needed, which is to be expected with only 5 minutes of run-time. Within only 1-4% of the Bilby run-time we are already able to recover the essentials of the skymaps.

4.2.5 Conclusion

In this paper, we produced skymaps for simulated BBH events using an importance sampling scheme that turns an approximate skymap made by a neural network into a skymap that represents the exact Bayesian posterior distribution. Experiments show that our method is competitive with Bilby and can produce the essentials of the skymap within 4% of the Bilby run-time. However, in some cases the proposal distributions made by the neural network are too crude, which hampers the efficiency of the importance sampling scheme. If the sampling efficiency is improved further, importance sampling could be used as a quick alternative to Bilby or LALInference for inferring

the GW posterior. Currently, the DL model has only been trained and tested on simulated noise with a given PSD. We expect that providing the DL model with various PSD representations as input into the model during training, as was done in [235], should allow the model to interpret the real world signals correctly regardless of the noise profile. In future work, we will also consider more advanced deep learning models such as normalizing flows to infer more accurate posterior distributions and apply the model to real measurements.

Acknowledgments

This work was (partially) funded by the NWO under the CORTEX project (NWA.1160.18.316). G.B. is supported by aFRIA grant from the Fonds de la Recherche Scientifique-FNRS, Belgium. J.J. is supported by the research program of the Netherlands Organisation for Scientific Research (NWO).

4.2.6 Appendices

Appendix A: Training details

Here we show the priors used for data generation (see Table 4.1) and the SNR distribution during training (see Figure 4.5).

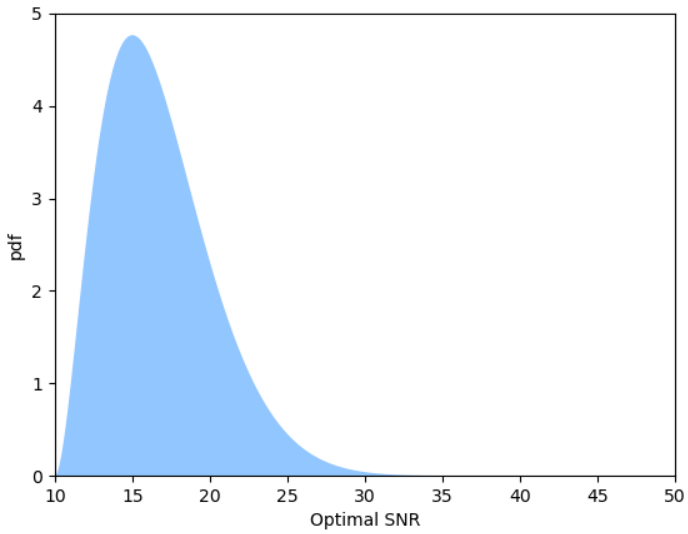

Figure 4.5: Scaled and shifted Beta distribution that acts as the SNR sampling distribution during training and validation. The vertical axis represents the probability density function of this Beta distribution, the horizontal axis represents the SNR value.

Appendix B: Mass estimation performance

In Figure 4.6, we show the mean relative error of the estimated masses over the test set. This figure closely resembles Figure 5 in [146]. Any differences are due to the difference in setup. The main differences are that our priors include spins and that we do not use a stationary sky origin.

Table 4.1: The priors used for the data generation. The luminosity distance in the prior was set to a 1000 Mpc and scaled afterwards to match the desired SNR.

parameter	prior	minimum	maximum	unit
Masses (constraint)	-	20	80	$M_\odot$
Chirp mass	Uniform	10	100	$M_\odot$
Mass ratio	Uniform	0.25	1.0	-
Spin magnitudes	Uniform	0	0.95	-
Spin polar angles	Sine	0	π	rad
Spin azimutal angles	Uniform	0	2π	rad
Right ascension	Uniform	0	2π	rad
Declination	Cosine	-0.5π	0.5π	rad
Binary inclination angle	Sine	0	π	rad
Coalescence phase angle	Uniform	0	2π	rad
Polarization angle	Uniform	0	2π	rad
Time Shift	Uniform	-0.1	0.1	s
Luminosity distance	-	1000	1000	Mpc

Appendix C: Importance sampling

We redid the importance experiment with only well converged runs, the shown log-likelihood values are close to those reported by Bilby.

Appendix D: Bilby run

For thirty Bilby runs, ten per SNR value, we repeated the experiments reported in Section 3.2. Below we show the results for one of the ten samples.

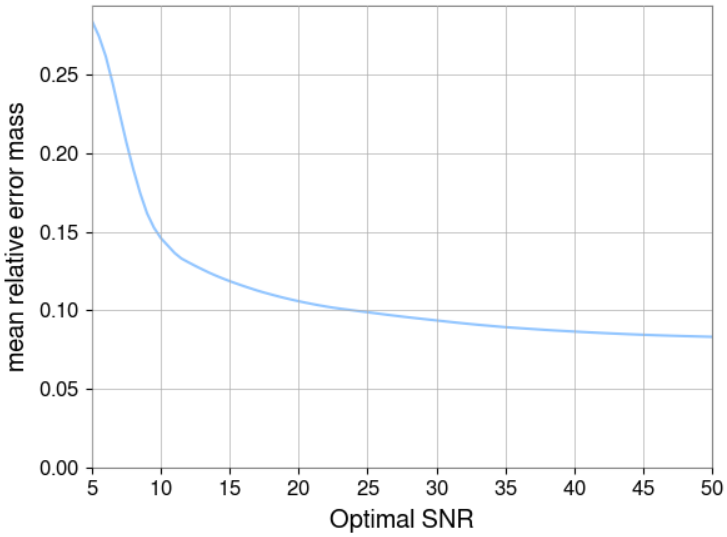

Figure 4.6: The mean relative error of the estimated masses by the neural network on the test set as a function of the optimal SNR. It is almost identical to the Figure 5 in [146].

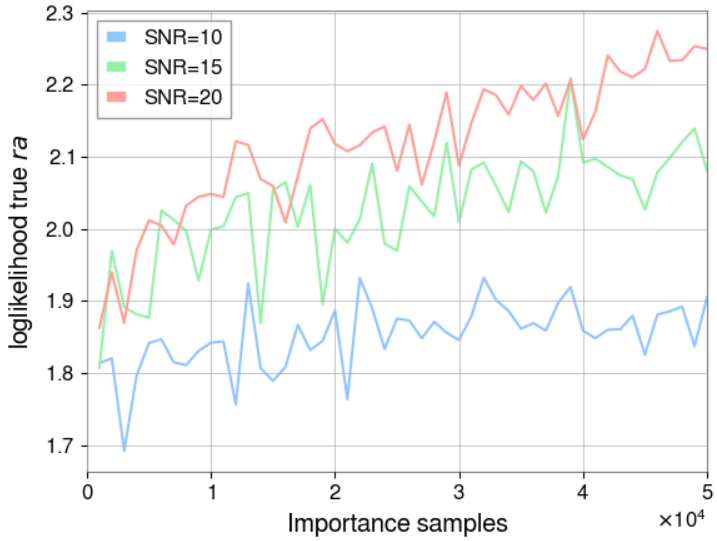

Figure 4.7: The log-likelihood of the true right ascension according to the kernel density estimator using only the importance samples of well converged runs. These values are more in line with those of Bilby, see Appendix D.

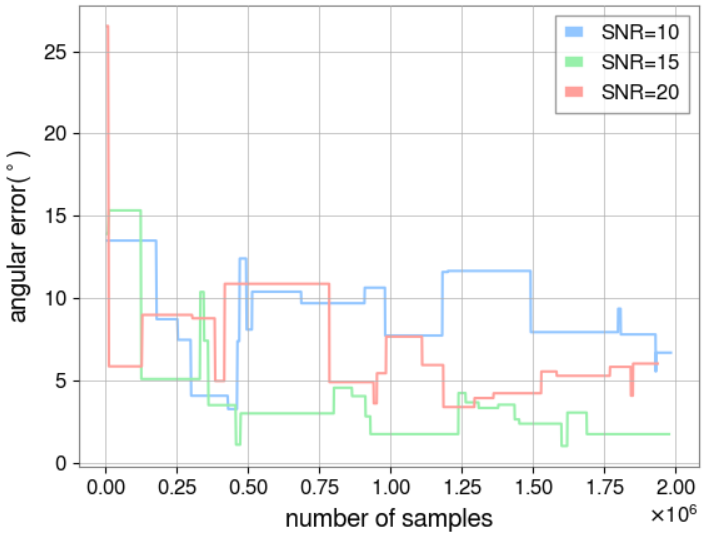

Figure 4.8: The angle between the sky location of the highest likelihood sample and the actual sky location. The vertical axis represents how many samples Bilby has generated (live and dead samples).

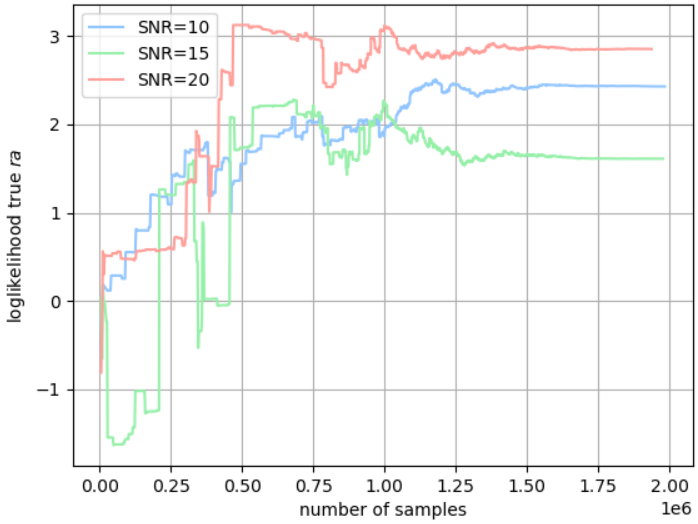

Figure 4.9: The loglikelihood of the true right ascension according to the kernel density estimator created by the Bilby samples. The vertical axis represents how many samples Bilby has generated (live and dead samples).

Chapter 5

Conclusion

The main purpose of this thesis was to build a pipeline for the early alert of binary neutron star mergers using machine learning. Before doing that, in order to understand better and have more feeling for the standard techniques, we helped to carry out a matched filtering search for sub-solar objects in the second run of observation of LIGO and Virgo, which lead to the paper that we have presented here. This search did not lead to any detection, but helped us to put some constraints on theories of primordial black holes. With this search, we have seen that we are getting closer and closer either to detecting some primordial black holes with sub-solar masses or rejecting some theories about them. The next run of observations O4 and O5 will be decisive.

As a second step, we started working on a deep learning pipeline for the detection of the early inspiral of binary neutron star mergers. In a first paper, we have shown that a CNN can recognize this partial signal even if we feed it directly with the 1D time-series data. This article is a proof of concept and comes with a lot of limitations: first, we used simulated Gaussian noise coming from the power spectral density at the design sensitivity of LIGO. Then, we limited ourselves to a minimum frequency of 20 Hz. Finally, we had to divide the BNSs into three categories depending on the masses, and used a different network for each category. In a second paper, we removed those limitations and only one network can now search for any BNS, whatever its masses. It can also search at lower frequency, which leads to a longer time before the merger, and we investigated the real noise of O3. The performance of the CNN is quite similar in real noise (coming from strain recorded at the time of low latency) and Gaussian noise, which means that the network is robust to the glitches and artifacts present in the real noise. We predicted that our network could see some BNSs in the next run of observation, with a warning time that can go beyond a minute before the merger. However, it was not able to see the one in O3 because the SNR was too low. Indeed, our method requires an SNR which is high enough to be able to find the early inspiral phase of a BNS. To be completely fair, the network still has too many false positives, but we still have some hints of techniques that we could try to improve it. Unfortunately, we did not have enough time to implement them in this thesis. A future improvement could be to train the network to trigger only if it sees a signal in multiple detectors, or if the SNR in multiple detectors is high enough.

Finally, with the purpose of having a full low latency pre-merger alert pipeline in mind, we decided to investigate how we can compute the sky map of a detected event with machine learning. We helped on a paper to localize the binary black hole with importance sampling and neural network. This paper introduces a method that can produce sky maps with a speed equivalent to BAYESTAR, but with a precision almost equivalent to Bilby. This project is the first step: improvement can be made so that the network could also produce the sky map for binary neutron star mergers, and then perform the sky localization of events only with the early inspiral. Combining this work and the previous one can result in a complete pipeline for the early warning of binary neutron stars.

Combining these last two work can result in a complete pipeline for the early warning of binary neutron stars. This could be particularly useful for multi-messenger astronomy, and could potentially lead to multiple detections of neutron star mergers with different channels of observation. In the next run of observation of LIGO-Virgo-Kagra the sensitivity band will go to smaller frequencies, then the signal will be longer, and it will be possible to detect them even earlier with this kind of pipeline. This type of research is also of interest for the next generation of interferometers like the Einstein Telescope and Cosmic Explorer. In these future instruments, the number of signals and their length will be a challenge for analyses. Machine learning will be useful, not only to detect events in advance, but also to perform parameter estimations or to denoise the strain data.

Bibliography

[1] B. P. Abbott *et al.*, "Observation of Gravitational Waves from a Binary Black Hole Merger," *Phys. Rev. Lett.*, vol. 116, no. 6, p. 061102, 2016.

[2] A. Einstein and N. Rosen, "On Gravitational waves," *J. Franklin Inst.*, vol. 223, pp. 43–54, 1937.

[3] S. Carroll, *Spacetime and Geometry: An Introduction to General Relativity*. Benjamin Cummings, 2003.

[4] C. Van Den Broeck, "Lecture notes on graviational waves." `https://www.nikhef.nl/~jo/gw2/index.html`, 2021.

[5] R. Abbott *et al.*, "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run," *arXiv:2111.03606*, 2021.

[6] J. Aasi *et al.*, "Advanced LIGO," *Class. Quant. Grav.*, vol. 32, p. 074001, 2015.

[7] 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*, vol. 11, no. 2, p. 021053, 2021.

[8] R. Abbott *et al.*, "GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run," *arXiv:2108.01045*, 2021.

[9] M. F. Bennett, C. A. van Eysden, and A. Melatos, "Continuous-wave gravitational radiation from pulsar glitch recovery," *Mon. Not. Roy. Astron. Soc.*, vol. 409, p. 1705, 2010.

[10] C. Palomba *et al.*, "Direct constraints on the ultralight boson mass from searches of continuous gravitational waves," *Phys. Rev. Lett.*, vol. 123, no. 17, p. 171101, 2019.

[11] B. Abbott *et al.*, "Searching for a Stochastic Background of Gravitational Waves with LIGO," *Astrophys. J.*, vol. 659, pp. 918–930, 2007.

[12] C. Ott, "The Gravitational Wave Signature of Core-Collapse Supernovae," *Class. Quant. Grav.*, vol. 26, p. 063001, 2009.

[13] M. H. P. M. van Putten, "Proposed source of gravitational radiation from a torus around a black hole," *Phys. Rev. Lett.*, vol. 87, p. 091101, 2001.

[14] S. Dall'Osso *et al.*, "Gravitational waves from massive magnetars formed in binary neutron star mergers," *Astrophys. J.*, vol. 798, no. 1, p. 25, 2015.

[15] A. Corsi and P. Mészáros, "GRB afterglow plateaus and Gravitational Waves: multi-messenger signature of a millisecond magnetar?," *Astrophys. J.*, vol. 702, pp. 1171–1178, 2009.

[16] K. Chatziioannou, "Neutron star tidal deformability and equation of state constraints," *Gen. Rel. Grav.*, vol. 52, no. 11, p. 109, 2020.

[17] B. P. Abbott *et al.*, "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral," *Phys. Rev. Lett.*, vol. 119, no. 16, p. 161101, 2017.

[18] A. H. Nitz *et al.*, "Rapid detection of gravitational waves from compact binary mergers with PyCBC Live," *Phys. Rev. D*, vol. 98, no. 2, p. 024050, 2018.

[19] W. G. Anderson and J. D. E. Creighton, "Searches for Gravitational Waves from Binary Neutron Stars: A Review," *Short-Period Binary Stars: Observations, Analyses, and Results*, pp. 23–52, 2008.

[20] B. Allen *et al.*, "FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries," *Phys. Rev. D*, vol. 85, p. 122006, 2012.

[21] S. Sachdev *et al.*, "The GstLAL Search Analysis Methods for Compact Binary Mergers in Advanced LIGO's Second and Advanced Virgo's First Observing Runs," *arXiv:1901.08580*, 2019.

[22] C. Hanna *et al.*, "Fast evaluation of multidetector consistency for real-time gravitational wave searches," *Phys. Rev. D*, vol. 101, no. 2, p. 022003, 2020.

[23] T. E. Riley *et al.*, "A NICER View of the Massive Pulsar PSR J0740+6620 Informed by Radio Timing and XMM-Newton Spectroscopy," *Astrophys. J. Lett.*, vol. 918, no. 2, p. L27, 2021.

[24] R. W. Romani, D. Kandel, A. V. Filippenko, T. G. Brink, and W. Zheng, "PSR J0952−0607: The Fastest and Heaviest Known Galactic Neutron Star," *Astrophys. J. Lett.*, vol. 934, no. 2, p. L18, 2022.

[25] B. Kiziltan *et al.*, "The neutron star mass distribution," *The Astrophysical Journal*, vol. 778, no. 1, p. 66, 2013.

[26] V. Kalogera and G. Baym, "The maximum mass of a neutron star," *Astrophys. J. Lett.*, vol. 470, pp. L61–L64, 1996.

[27] W. M. Farr, N. Sravan, A. Cantrell, L. Kreidberg, C. D. Bailyn, I. Mandel, and V. Kalogera, "The Mass Distribution of Stellar-Mass Black Holes," *Astrophys. J.*, vol. 741, p. 103, 2011.

[28] S. E. Woosley and A. Heger, "The Pair-Instability Mass Gap for Black Holes," *Astrophys. J. Lett.*, vol. 912, no. 2, p. L31, 2021.

[29] L. Kreidberg *et al.*, "Mass Measurements of Black Holes in X-Ray Transients: Is There a Mass Gap?," *Astrophys. J.*, vol. 757, p. 36, 2012.

[30] S. Bird *et al.*, "Did LIGO detect dark matter?," *Phys. Rev. Lett.*, vol. 116, no. 20, p. 201301, 2016.

[31] C. Kouvaris and P. Tinyakov, "Can Neutron stars constrain Dark Matter?," *Phys. Rev. D*, vol. 82, p. 063531, 2010.

[32] E. Cotner *et al.*, "Analytic Description of Primordial Black Hole Formation from Scalar Field Fragmentation," *JCAP*, vol. 10, p. 077, 2019.

[33] R. Bousso and S. W. Hawking, "The Probability for primordial black holes," *Phys. Rev. D*, vol. 52, pp. 5659–5664, 1995.

[34] B. Carr *et al.*, "Constraints on primordial black holes," *Rept. Prog. Phys.*, vol. 84, no. 11, p. 116902, 2021.

[35] S. Clesse and J. García-Bellido, "The clustering of massive Primordial Black Holes as Dark Matter: measuring their mass distribution with Advanced LIGO," *Phys. Dark Univ.*, vol. 15, pp. 142–147, 2017.

[36] T. Kawaguchi *et al.*, "Formation of intermediate-mass black holes as primordial black holes in the inflationary cosmology with running spectral index," *Mon. Not. Roy. Astron. Soc.*, vol. 388, pp. 1426–1432, 2008.

[37] C. T. Byrnes *et al.*, "Primordial black holes with an accurate QCD equation of state," *JCAP*, vol. 08, p. 041, 2018.

[38] 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*, vol. 9, no. 3, p. 031040, 2019.

[39] B. P. Abbott *et al.*, "Search for Subsolar-Mass Ultracompact Binaries in Advanced LIGO's First Observing Run," *Phys. Rev. Lett.*, vol. 121, no. 23, p. 231103, 2018.

[40] B. P. Abbott *et al.*, "Search for Subsolar Mass Ultracompact Binaries in Advanced LIGO's Second Observing Run," *Phys. Rev. Lett.*, vol. 123, no. 16, p. 161102, 2019.

[41] A. H. Nitz and Y.-F. Wang, "Search for Gravitational Waves from the Coalescence of Subsolar-Mass Binaries in the First Half of Advanced LIGO and Virgo's Third Observing Run," *Phys. Rev. Lett.*, vol. 127, no. 15, p. 151101, 2021.

[42] R. Biswas *et al.*, "The Loudest event statistic: General formulation, properties and applications," *Class. Quant. Grav.*, vol. 26, p. 175009, 2009. [Erratum: Class. Quant. Grav. 30, 079502 (2013)].

[43] R. Abbott *et al.*, "GW190521: A Binary Black Hole Merger with a Total Mass of $150M_{\odot}$," *Phys. Rev. Lett.*, vol. 125, no. 10, p. 101102, 2020.

[44] R. Abbott *et al.*, "GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object," *Astrophys. J. Lett.*, vol. 896, no. 2, p. L44, 2020.

[45] R. Abbott *et al.*, "GW190412: Observation of a Binary-Black-Hole Coalescence with Asymmetric Masses," *Phys. Rev. D*, vol. 102, no. 4, p. 043015, 2020.

[46] B. P. Abbott *et al.*, "GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence," *Phys. Rev. Lett.*, vol. 119, no. 14, p. 141101, 2017.

[47] B. P. Abbott *et al.*, "GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence," *Phys. Rev. Lett.*, vol. 116, no. 24, p. 241103, 2016.

[48] B. . P. . Abbott *et al.*, "GW170608: Observation of a 19-solar-mass Binary Black Hole Coalescence," *Astrophys. J.*, vol. 851, no. 2, p. L35, 2017.

[49] B. P. Abbott *et al.*, "GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2," *Phys. Rev. Lett.*, vol. 118, no. 22, p. 221101, 2017. [Erratum: Phys. Rev. Lett. 121, 129901 (2018)].

[50] F. Acernese *et al.*, "Advanced Virgo: a second-generation interferometric gravitational wave detector," *Class. Quant. Grav.*, vol. 32, no. 2, p. 024001, 2015.

[51] M. Sasaki *et al.*, "Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914," *Phys. Rev. Lett.*, vol. 117, no. 6, p. 061101, 2016. [Erratum: Phys. Rev. Lett. 121, 059901 (2018)].

[52] Y. B. Zel'dovich and I. D. Novikov, "The Hypothesis of Cores Retarded during Expansion and the Hot Cosmological Model," *Soviet Astronomy*, vol. 10, p. 602, 1967.

[53] S. Hawking, "Gravitationally collapsed objects of very low mass," *Mon. Not. Roy. Astron. Soc.*, vol. 152, p. 75, 1971.

[54] B. J. Carr and S. W. Hawking, "Black holes in the early Universe," *Mon. Not. Roy. Astron. Soc.*, vol. 168, pp. 399–415, 1974.

[55] G. F. Chapline, "Cosmological effects of primordial black holes," *Nature*, vol. 253, no. 5489, pp. 251–252, 1975.

[56] B. J. Carr and J. E. Lidsey, "Primordial black holes and generalized constraints on chaotic inflation," *Phys. Rev. D*, vol. 48, pp. 543–553, 1993.

[57] P. Ivanov, P. Naselsky, and I. Novikov, "Inflation and primordial black holes as dark matter," *Phys. Rev. D*, vol. 50, pp. 7173–7178, 1994.

[58] J. Garcia-Bellido, A. D. Linde, and D. Wands, "Density perturbations and black hole formation in hybrid inflation," *Phys. Rev. D*, vol. 54, pp. 6040–6058, 1996.

[59] H. I. Kim and C. H. Lee, "Constraints on the spectral index from primordial black holes," *Phys. Rev. D*, vol. 54, pp. 6001–6007, 1996.

[60] B. Carr, F. Kuhnel, and M. Sandstad, "Primordial Black Holes as Dark Matter," *Phys. Rev. D*, vol. 94, no. 8, p. 083504, 2016.

[61] B. Carr *et al.*, "New cosmological constraints on primordial black holes," *Phys. Rev. D*, vol. 81, p. 104019, 2010.

[62] B. Carr and F. Kuhnel, "Primordial Black Holes as Dark Matter: Recent Developments," *Ann. Rev. Nucl. Part. Sci.*, pp. 170:14.1–14.40, 6 2020.

[63] A. M. Green and B. J. Kavanagh, "Primordial Black Holes as a dark matter candidate," *J. Phys. G*, vol. 48, no. 4, p. 043001, 2021.

[64] S. Clesse and J. García-Bellido, "Seven Hints for Primordial Black Hole Dark Matter," *Phys. Dark Univ.*, vol. 22, pp. 137–146, 2018.

[65] B. Carr *et al.*, "Cosmic conundra explained by thermal history and primordial black holes," *Phys. Dark Univ.*, vol. 31, p. 100755, 2021.

[66] S. Chandrasekhar, "The Maximum Mass of Ideal White Dwarfs," *Astrophys. J.*, vol. 74, p. 81, July 1931.

[67] G. Rakavy and G. Shaviv, "Instabilities in Highly Evolved Stellar Models," *Astrophys. J.*, vol. 148, p. 803, June 1967.

[68] G. S. Fraley, "Supernovae Explosions Induced by Pair-Production Instability," *Astrophys. Space Sci.*, vol. 2, pp. 96–114, Aug. 1968.

[69] J. C. Niemeyer and K. Jedamzik, "Near-Critical Gravitational Collapse and the Initial Mass Function of Primordial Black Holes," *Phys. Rev. Lett.*, vol. 80, pp. 5481–5484, 1998.

[70] K. Jedamzik, "Primordial black hole formation during the QCD epoch," *Phys. Rev.*, vol. D55, pp. R5871–5875, 1997.

[71] S. Clesse and J. Garcia-Bellido, "GW190425, GW190521 and GW190814: Three candidate mergers of primordial black holes from the QCD epoch," *arXiv: 2007.06481*, 2020.

[72] L. Wyrzykowski and I. Mandel, "Constraining the masses of microlensing black holes and the mass gap with Gaia DR2," *Astron. Astrophys.*, vol. 636, p. A20, 2020.

[73] A. Kashlinsky, "LIGO gravitational wave detection, primordial black holes and the near-IR cosmic infrared background anisotropies," *Astrophys. J. Lett.*, vol. 823, no. 2, p. L25, 2016.

[74] A. H. Nitz and Y.-F. Wang, "Search for Gravitational Waves from High-Mass-Ratio Compact-Binary Mergers of Stellar Mass and Subsolar Mass Black Holes," *Phys. Rev. Lett.*, vol. 126, no. 2, p. 021103, 2021.

[75] A. H. Nitz and Y.-F. Wang, "Search for gravitational waves from the coalescence of sub-solar mass and eccentric compact binaries," *arXiv:2102.00868*, 2021.

[76] M. Raidal, C. Spethmann, V. Vaskonen, and H. Veermäe, "Formation and Evolution of Primordial Black Hole Binaries in the Early Universe," *JCAP*, vol. 02, p. 018, 2019.

[77] K. Cannon *et al.*, "GstLAL: A software framework for gravitational wave discovery," *SoftwareX*, vol. 14, p. 100680, 2021.

[78] K. Cannon *et al.*, "Toward Early-Warning Detection of Gravitational Waves from Compact Binary Coalescence," *Astrophys. J.*, vol. 748, p. 136, 2012.

[79] C. Messick *et al.*, "Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data," *Phys. Rev. D*, vol. 95, no. 4, p. 042001, 2017.

[80] LIGO Scientific Collaboration, "LIGO Algorithm Library - LALSuite." free software (GPL), 2018.

[81] S. V. Dhurandhar and B. S. Sathyaprakash, "Choice of filters for the detection of gravitational waves from coalescing binaries. 2. Detection in colored noise," *Phys. Rev. D*, vol. 49, pp. 1707–1722, 1994.

[82] R. Balasubramanian, B. S. Sathyaprakash, and S. V. Dhurandhar, "Gravitational waves from coalescing binaries: Detection strategies and Monte Carlo estimation of parameters," *Phys. Rev. D*, vol. 53, pp. 3033–3055, 1996. [Erratum: Phys. Rev. D 54, 1860 (1996)].

[83] B. J. Owen, "Search templates for gravitational waves from inspiraling binaries: Choice of template spacing," *Phys. Rev. D*, vol. 53, pp. 6749–6761, 1996.

[84] B. J. Owen and B. S. Sathyaprakash, "Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement," *Phys. Rev. D*, vol. 60, p. 022002, 1999.

[85] L. Blanchet *et al.*, "Gravitational radiation damping of compact binary systems to second postNewtonian order," *Phys. Rev. Lett.*, vol. 74, pp. 3515–3518, 1995.

[86] E. Poisson, "Gravitational waves from inspiraling compact binaries: The Quadrupole moment term," *Phys. Rev. D*, vol. 57, pp. 5287–5290, 1998.

[87] T. Damour, P. Jaranowski, and G. Schaefer, "Dimensional regularization of the gravitational interaction of point masses," *Phys. Lett. B*, vol. 513, pp. 147–155, 2001.

[88] T. Damour, B. R. Iyer, and B. S. Sathyaprakash, "A Comparison of search templates for gravitational waves from binary inspiral," *Phys. Rev. D*, vol. 63, p. 044023, 2001. [Erratum: Phys. Rev. D 72, 029902 (2005)].

[89] A. Buonanno *et al.*, "Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors," *Phys. Rev. D*, vol. 80, p. 084043, 2009.

[90] B. Mikoczi, M. Vasuth, and L. A. Gergely, "Self-interaction spin effects in inspiralling compact binaries," *Phys. Rev. D*, vol. 71, p. 124043, 2005.

[91] L. Blanchet, T. Damour, G. Esposito-Farese, and B. R. Iyer, "Dimensional regularization of the third post-Newtonian gravitational wave generation from two point masses," *Phys. Rev. D*, vol. 71, p. 124004, 2005.

[92] K. G. Arun, A. Buonanno, G. Faye, and E. 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*, vol. 79, p. 104023, 2009. [Erratum: Phys. Rev. D 84, 049901 (2011)].

[93] A. Bohé, S. Marsat, and L. Blanchet, "Next-to-next-to-leading order spin–orbit effects in the gravitational wave flux and orbital phasing of compact binaries," *Class. Quant. Grav.*, vol. 30, p. 135009, 2013.

[94] A. Bohé *et al.*, "Quadratic-in-spin effects in the orbital dynamics and gravitational-wave energy flux of compact binaries at the 3PN order," *Class. Quant. Grav.*, vol. 32, no. 19, p. 195010, 2015.

[95] C. K. Mishra *et al.*, "Ready-to-use post-Newtonian gravitational waveforms for binary black holes with nonprecessing spins: An update," *Phys. Rev. D*, vol. 93, no. 8, p. 084054, 2016.

[96] I. W. Harry, S. Fairhurst, and B. S. Sathyaprakash, "A Hierarchical search for gravitational waves from supermassive black hole binary mergers," *Class. Quant. Grav.*, vol. 25, p. 184027, 2008.

[97] S. Babak, "Building a stochastic template bank for detecting massive black hole binaries," *Class. Quant. Grav.*, vol. 25, p. 195011, 2008.

[98] I. W. Harry, B. Allen, and B. S. Sathyaprakash, "A Stochastic template placement algorithm for gravitational wave data analysis," *Phys. Rev. D*, vol. 80, p. 104014, 2009.

[99] G. M. Manca and M. Vallisneri, "Cover art: Issues in the metric-guided and metric-less placement of random and stochastic template banks," *Phys. Rev. D*, vol. 81, p. 024004, 2010.

[100] R. Magee *et al.*, "Methods for the detection of gravitational waves from subsolar mass ultracompact binaries," *Phys. Rev. D*, vol. 98, no. 10, p. 103024, 2018.

[101] K. Cannon, C. Hanna, and J. Peoples, "Likelihood-Ratio Ranking Statistic for Compact Binary Coalescence Candidates with Rate Estimation," *arXiv:1504.04632*, 2015.

[102] R. Essick *et al.*, "iDQ: Statistical Inference of Non-Gaussian Noise with Auxiliary Degrees of Freedom in Gravitational-Wave Detectors," *Machine Learning: Science and Technology*, vol. 2, no. 1, p. 015004, 2020.

[103] P. Godwin *et al.*, "Incorporation of Statistical Data Quality Information into the GstLAL Search Analysis," *arXiv:2010.15282*, 2020.

[104] K. Cannon, C. Hanna, and D. Keppel, "Method to estimate the significance of coincident gravitational-wave observations from compact binary coalescence," *Phys. Rev. D*, vol. 88, no. 2, p. 024025, 2013.

[105] W. M. Farr, J. R. Gair, I. Mandel, and C. Cutler, "Counting And Confusion: Bayesian Rate Estimation With Multiple Populations," *Phys. Rev. D*, vol. 91, no. 2, p. 023005, 2015.

[106] S. J. Kapadia *et al.*, "A self-consistent method to estimate the rate of compact binary coalescences with a Poisson mixture model," *Class. Quant. Grav.*, vol. 37, no. 4, p. 045007, 2020.

[107] V. De Luca, G. Franciolini, and A. Riotto, "NANOGrav Data Hints at Primordial Black Holes as Dark Matter," *Phys. Rev. Lett.*, vol. 126, no. 4, p. 041303, 2021.

[108] B. P. Abbott *et al.*, "GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4 M_{\odot}$," *Astrophys. J. Lett.*, vol. 892, no. 1, p. L3, 2020.

[109] Z.-C. Chen, C. Yuan, and Q.-G. Huang, "Pulsar Timing Array Constraints on Primordial Black Holes with NANOGrav 11-Year Dataset," *Phys. Rev. Lett.*, vol. 124, no. 25, p. 251101, 2020.

[110] V. Vaskonen and H. Veermäe, "Did NANOGrav see a signal from primordial black hole formation?," *Phys. Rev. Lett.*, vol. 126, no. 5, p. 051303, 2021.

[111] V. De Luca *et al.*, "GW190521 Mass Gap Event and the Primordial Black Hole Scenario," *Phys. Rev. Lett.*, vol. 126, no. 5, p. 051101, 2021.

[112] K. Kohri and T. Terada, "Solar-Mass Primordial Black Holes Explain NANOGrav Hint of Gravitational Waves," *Phys. Lett. B*, vol. 813, p. 136040, 2021.

[113] S. Clesse, J. García-Bellido, and S. Orani, "Detecting the Stochastic Gravitational Wave Background from Primordial Black Hole Formation," *arXiv:1812.11011*, 2018.

[114] Z. Arzoumanian *et al.*, "The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background," *Astrophys. J. Lett.*, vol. 905, no. 2, p. L34, 2020.

[115] Y. Ali-Haïmoud, E. D. Kovetz, and M. Kamionkowski, "Merger rate of primordial black-hole binaries," *Phys. Rev. D*, vol. 96, no. 12, p. 123523, 2017.

[116] A. D. Gow, C. T. Byrnes, A. Hall, and J. A. Peacock, "Primordial black hole merger rates: distributions for multiple LIGO observables," *JCAP*, vol. 01, p. 031, 2020.

[117] L. Liu, Z.-K. Guo, and R.-G. Cai, "Effects of the merger history on the merger rate density of primordial black hole binaries," *Eur. Phys. J. C*, vol. 79, no. 8, p. 717, 2019.

[118] B. Kocsis *et al.*, "Hidden universality in the merger rate distribution in the primordial black hole scenario," *Astrophys. J.*, vol. 854, no. 1, p. 41, 2018.

[119] G. Hütsi, M. Raidal, V. Vaskonen, and H. Veermäe, "Two populations of LIGO-Virgo black holes," *JCAP*, vol. 03, p. 068, 2021.

[120] V. Vaskonen and H. Veermäe, "Lower bound on the primordial black hole merger rate," *Phys. Rev. D*, vol. 101, no. 4, p. 043015, 2020.

[121] B. P. Abbott *et al.*, "Effects of data quality vetoes on a search for compact binary coalescences in Advanced LIGO's first observing run," *Class. Quant. Grav.*, vol. 35, no. 6, p. 065010, 2018.

[122] C. Biwer *et al.*, "PyCBC Inference: A Python-based parameter estimation toolkit for compact binary coalescence signals," *Publ. Astron. Soc. Pac.*, vol. 131, no. 996, p. 024503, 2019.

[123] C. Boehm *et al.*, "Eliminating the LIGO bounds on primordial black hole dark matter," *JCAP*, vol. 03, p. 078, 2021.

[124] V. De Luca, V. Desjacques, G. Franciolini, and A. Riotto, "The clustering evolution of primordial black holes," *JCAP*, vol. 11, p. 028, 2020.

[125] B. P. Abbott *et al.*, "Binary Black Hole Mergers in the first Advanced LIGO Observing Run," *Phys. Rev. X*, vol. 6, no. 4, p. 041015, 2016. [Erratum: Phys. Rev. X 8, 039903 (2018)].

[126] A. Goldstein *et al.*, "An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A," *Astrophys. J. Lett.*, vol. 848, no. 2, p. L14, 2017.

[127] B. P. Abbott *et al.*, "Multi-messenger Observations of a Binary Neutron Star Merger," *Astrophys. J. Lett.*, vol. 848, no. 2, p. L12, 2017.

[128] A. H. Nitz *et al.*, "Detecting binary compact-object mergers with gravitational waves: Understanding and Improving the sensitivity of the PyCBC search," *Astrophys. J.*, vol. 849, no. 2, p. 118, 2017.

[129] Q. Chu *et al.*, "SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences," *Phys. Rev. D*, vol. 105, no. 2, p. 024023, 2022.

[130] G. Baltus *et al.*, "Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal," *Phys. Rev. D*, vol. 103, p. 102003, 2021.

[131] G. Baltus *et al.*, "Convolutional neural network for gravitational-wave early alert: Going down in frequency," *Phys. Rev. D*, vol. 106, no. 4, p. 042002, 2022.

[132] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2014.

[133] A. v. d. Oord *et al.*, "WaveNet: A Generative Model for Raw Audio," *arXiv:1609.03499*, 2016.

[134] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv:1502.03167*, 2015.

[135] M. Spurio, *Particles and astrophysics : a multi-messenger approach.* Astronomy and Astrophysics Library. Springer., 2015.

[136] A. Burrows and J. M. Lattimer, "Neutrinos from SN 1987A," *Astrophys. J. Lett.*, vol. 318, pp. L63–L68, 1987.

[137] P. S. Cowperthwaite *et al.*, "The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. II. UV, Optical, and Near-infrared Light Curves and Comparison to Kilonova Models," *Astrophys. J. Lett.*, vol. 848, no. 2, p. L17, 2017.

[138] M. Soares-Santos *et al.*, "First Measurement of the Hubble Constant from a Dark Standard Siren using the Dark Energy Survey Galaxies and the LIGO/Virgo Binary–Black-hole Merger GW170814," *Astrophys. J. Lett.*, vol. 876, no. 1, p. L7, 2019.

[139] M. Fishbach *et al.*, "A Standard Siren Measurement of the Hubble Constant from GW170817 without the Electromagnetic Counterpart," *Astrophys. J. Lett.*, vol. 871, no. 1, p. L13, 2019.

[140] E. Berti, K. Yagi, and N. Yunes, "Extreme Gravity Tests with Gravitational Waves from Compact Binary Coalescences: (I) Inspiral-Merger," *Gen. Rel. Grav.*, vol. 50, no. 4, p. 46, 2018.

[141] B. P. Abbott *et al.*, "Tests of General Relativity with the Binary Black Hole Signals from the LIGO-Virgo Catalog GWTC-1," *Phys. Rev. D*, vol. 100, no. 10, p. 104036, 2019.

[142] B. P. Abbott *et al.*, "A gravitational-wave standard siren measurement of the Hubble constant," *Nature*, vol. 551, no. 7678, pp. 85–88, 2017.

[143] B. D. Metzger and A. L. Piro, "Optical and X-ray emission from stable millisecond magnetars formed from the merger of binary neutron stars," *Mon. Not. Roy. Astron. Soc.*, vol. 439, pp. 3916–3930, 2014.

[144] S. Sachdev *et al.*, "An Early-warning System for Electromagnetic Follow-up of Gravitational-wave Events," *Astrophys. J. Lett.*, vol. 905, no. 2, p. L25, 2020.

[145] A. H. Nitz, M. Schäfer, and T. Dal Canton, "Gravitational-wave Merger Forecasting: Scenarios for the Early Detection and Localization of Compact-binary Mergers with Ground-based Observatories," *Astrophys. J. Lett.*, vol. 902, p. L29, 2020.

[146] D. George and E. A. Huerta, "Deep Neural Networks to Enable Real-time Multimessenger Astrophysics," *Phys. Rev. D*, vol. 97, no. 4, p. 044039, 2018.

[147] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, "Matching matched filtering with deep networks for gravitational-wave astronomy," *Phys. Rev. Lett.*, vol. 120, no. 14, p. 141103, 2018.

[148] T. Gebhard *et al.*, "Convwave: Searching for gravitational waves with fully convolutional neural nets," in *Workshop on Deep Learning for Physical Sciences (DLPS) at the 31st Conference on Neural Information Processing Systems (NIPS)*, pp. 1–6, 2017.

[149] T. D. Gebhard, N. Kilbertus, I. Harry, and B. Schölkopf, "Convolutional neural networks: a magic bullet for gravitational-wave detection?," *Phys. Rev. D*, vol. 100, no. 6, p. 063015, 2019.

[150] P. G. Krastev, "Real-Time Detection of Gravitational Waves from Binary Neutron Stars using Artificial Neural Networks," *Phys. Lett. B*, vol. 803, p. 135330, 2020.

[151] M. B. Schäfer, F. Ohme, and A. H. Nitz, "Detection of gravitational-wave signals from binary neutron star mergers using machine learning," *Phys. Rev. D*, vol. 102, no. 6, p. 063015, 2020.

[152] H. Yu *et al.*, "Early warning of coalescing neutron-star and neutron-star-black-hole binaries from the nonstationary noise background using neural networks," *Phys. Rev. D*, vol. 104, no. 6, p. 062004, 2021.

[153] W. Wei and E. A. Huerta, "Deep learning for gravitational wave forecasting of neutron star mergers," *Phys. Lett. B*, vol. 816, p. 136185, 2021.

[154] C. Meegan *et al.*, "The Fermi Gamma-Ray Burst Monitor," *Astrophys. J.*, vol. 702, pp. 791–804, 2009.

[155] V. Savchenko *et al.*, "INTEGRAL Detection of the First Prompt Gamma-Ray Signal Coincident with the Gravitational-wave Event GW170817," *Astrophys. J. Lett.*, vol. 848, no. 2, p. L15, 2017.

[156] B. P. Abbott *et al.*, "Low-latency Gravitational-wave Alerts for Multimessenger Astronomy during the Second Advanced LIGO and Virgo Observing Run," *Astrophys. J.*, vol. 875, no. 2, p. 161, 2019.

[157] S. Akcay, "Forecasting Gamma-Ray Bursts Using Gravitational Waves," *Annalen Phys.*, vol. 531, no. 1, p. 1800365, 2019.

[158] D. N. Burrows *et al.*, "The Swift X-ray Telescope," *Space Sci. Rev.*, vol. 120, p. 165, 2005.

[159] D. Reitze *et al.*, "Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO," *Bull. Am. Astron. Soc.*, vol. 51, p. 035, 7 2019.

[160] M. Punturo *et al.*, "The Einstein Telescope: A third-generation gravitational wave observatory," *Class. Quant. Grav.*, vol. 27, p. 194002, 2010.

[161] B. Sathyaprakash *et al.*, "Scientific Objectives of Einstein Telescope," *Class. Quant. Grav.*, vol. 29, p. 124013, 2012. [Erratum: Class. Quant. Grav. 30, 079501 (2013)].

[162] T. Tsutsui, A. Nishizawa, and S. Morisaki, "Early warning of precessing compact binary merger with third-generation gravitational-wave detectors," *Phys. Rev. D*, vol. 104, no. 6, p. 064013, 2021.

[163] S. A. Usman *et al.*, "The PyCBC search for gravitational waves from compact binary coalescence," *Class. Quant. Grav.*, vol. 33, no. 21, p. 215004, 2016.

[164] T. 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*, vol. 90, no. 8, p. 082004, 2014.

[165] T. Adams *et al.*, "Low-latency analysis pipeline for compact binary coalescences in the advanced gravitational wave detector era," *Class. Quant. Grav.*, vol. 33, no. 17, p. 175012, 2016.

[166] D. George and E. Huerta, "Deep Learning for Real-time Gravitational Wave Detection and Parameter Estimation: Results with Advanced LIGO Data," *Phys. Lett. B*, vol. 778, pp. 64–70, 2018.

[167] H. Gabbard *et al.*, "Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy," *Nature Phys.*, vol. 18, no. 1, pp. 112–117, 2022.

[168] S. R. Green, C. Simpson, and J. Gair, "Gravitational-wave parameter estimation with autoregressive neural network flows," *Phys. Rev. D*, vol. 102, no. 10, p. 104057, 2020.

[169] S. R. Green and J. Gair, "Complete parameter inference for GW150914 using deep learning," *Mach. Learn. Sci. Tech.*, vol. 2, no. 3, p. 03LT01, 2021.

[170] E. Cuoco *et al.*, "Enhancing Gravitational-Wave Science with Machine Learning," *Mach. Learn. Sci. Tech.*, vol. 2, no. 1, p. 011002, 2021.

[171] M. J. Williams, J. Veitch, and C. Messenger, "Nested sampling with normalizing flows for gravitational-wave inference," *Phys. Rev. D*, vol. 103, no. 10, p. 103006, 2021.

[172] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv:1512.03385*, 2015.

[173] B. Allen, "χ^2 time-frequency discriminator for gravitational wave detection," *Phys. Rev. D*, vol. 71, p. 062001, 2005.

[174] A. Samajdar *et al.*, "Biases in parameter estimation from overlapping gravitational-wave signals in the third generation detector era," *arXiv:2102.07544*, 2021.

[175] C. Cutler and E. E. Flanagan, "Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral wave form?," *Phys. Rev. D*, vol. 49, pp. 2658–2697, 1994.

[176] B. F. Schutz, "Networks of gravitational wave detectors and three figures of merit," *Class. Quant. Grav.*, vol. 28, p. 125023, 2011.

[177] A. Buonanno, Y.-b. Chen, and M. Vallisneri, "Detecting gravitational waves from precessing binaries of spinning compact objects: Adiabatic limit," *Phys. Rev. D*, vol. 67, p. 104025, 2003. [Erratum: Phys. Rev. D 74, 029904 (2006)].

[178] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), pp. 8024–8035, Curran Associates, Inc., 2019.

[179] T. Fawcett, "An introduction to ROC Analysis," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.

[180] R. Abbott *et al.*, "Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog," *Astrophys. J. Lett.*, vol. 913, no. 1, p. L7, 2021.

[181] B. P. Abbott *et al.*, "Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA," *Living Rev. Rel.*, vol. 21, no. 1, p. 3, 2018.

[182] M. López *et al.*, "Deep learning for core-collapse supernova detection," *Phys. Rev. D*, vol. 103, no. 6, p. 063011, 2021.

[183] P. Mészáros *et al.*, "Multi-Messenger Astrophysics," *Nature Rev. Phys.*, vol. 1, pp. 585–599, 2019.

[184] W. D. Arnett, D. N. Schramm, and J. W. Truran, "On Relative Supernova Rates and Nucleosynthesis Roles," *Astrophys. J. Lett.*, vol. 339, p. L25, 1989.

[185] S. Ansoldi *et al.*, "The blazar TXS 0506+056 associated with a high-energy neutrino: insights into extragalactic jets and cosmic ray acceleration," *Astrophys. J. Lett.*, vol. 863, p. L10, 2018.

[186] J. Goodman, "Are gamma-ray bursts optically thick?," *Astrophys. J. Lett.*, vol. 308, pp. L47–L50, 1986.

[187] T. Akutsu *et al.*, "KAGRA: 2.5 Generation Interferometric Gravitational Wave Detector," *Nature Astron.*, vol. 3, no. 1, pp. 35–40, 2019.

[188] L. P. Singer *et al.*, "The First Two Years of Electromagnetic Follow-Up with Advanced LIGO and Virgo," *Astrophys. J.*, vol. 795, no. 2, p. 105, 2014.

[189] A. Perego *et al.*, "AT 2017gfo: An Anisotropic and Three-component Kilonova Counterpart of GW170817," *Astrophys. J. Lett.*, vol. 850, no. 2, p. L37, 2017.

[190] T. Tsutsui, A. Nishizawa, and S. Morisaki, "Early warning of precessing neutron-star black hole binary mergers with the near-future gravitational-wave detectors," *Mon. Not. Roy. Astron. Soc.*, vol. 512, no. 3, pp. 3878–3884, 2022.

[191] M. L. Chan, C. Messenger, I. S. Heng, and M. Hendry, "Binary Neutron Star Mergers and Third Generation Detectors: Localization and Early Warning," *Phys. Rev. D*, vol. 97, no. 12, p. 123014, 2018.

[192] Y. Li *et al.*, "Exploring the sky localization and early warning capabilities of third generation gravitational wave detectors in three-detector network configurations," *Phys. Rev. D*, vol. 105, no. 4, p. 043010, 2022.

[193] M. Kovalam *et al.*, "Early Warnings of Binary Neutron Star Coalescence Using the SPIIR Search," *Astrophys. J. Lett.*, vol. 927, no. 1, p. L9, 2022.

[194] R. Magee and S. Borhanian, "Realistic observing scenarios for the next decade of early warning detection of binary neutron stars," *arXiv: 2201.11841*, 2022.

[195] L. K. Nuttall and C. P. L. Berry, "Electromagnetic counterparts of gravitational-wave signals," *Astron. Geophys.*, vol. 62, no. 4, pp. 4.15–4.21, 2021.

[196] R. Magee *et al.*, "First demonstration of early warning gravitational wave alerts," *Astrophys. J. Lett.*, vol. 910, no. 2, p. L21, 2021.

[197] M. Saleem *et al.*, "Rates of Short-GRB afterglows in association with Binary Neutron Star mergers," *Mon. Not. Roy. Astron. Soc.*, vol. 475, no. 1, pp. 699–707, 2018.

[198] A. H. Nitz and T. Dal Canton, "Pre-merger Localization of Compact-binary Mergers with Third-generation Observatories," *Astrophys. J. Lett.*, vol. 917, no. 2, p. L27, 2021.

[199] G. S. Davies *et al.*, "Extending the PyCBC search for gravitational waves from compact binary mergers to a global network," *Phys. Rev. D*, vol. 102, no. 2, p. 022004, 2020.

[200] F. Aubin *et al.*, "The MBTA pipeline for detecting compact binary coalescences in the third LIGO–Virgo observing run," *Class. Quant. Grav.*, vol. 38, no. 9, p. 095004, 2021.

[201] S. Klimenko *et al.*, "Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors," *Phys. Rev. D*, vol. 93, no. 4, p. 042004, 2016.

[202] S. Klimenko and G. Mitselmakher, "A wavelet method for detection of gravitational wave bursts," *Class. Quant. Grav.*, vol. 21, pp. S1819–S1830, 2004.

[203] S. Klimenko *et al.*, "Localization of gravitational wave sources with networks of advanced detectors," *Phys. Rev. D*, vol. 83, p. 102001, 2011.

[204] D. Davis *et al.*, "LIGO detector characterization in the second and third observing runs," *Class. Quant. Grav.*, vol. 38, no. 13, p. 135014, 2021.

[205] I. Goodfellow *et al.*, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[206] P. Astone *et al.*, "New method to observe gravitational waves emitted by core collapse supernovae," *Phys. Rev. D*, vol. 98, no. 12, p. 122002, 2018.

[207] A. Iess *et al.*, "Core-collapse supernova gravitational-wave search and deep learning classification," *Machine Learning: Science and Technology*, vol. 1, 03 2020.

[208] V. Boudart and M. Fays, "Machine learning algorithm for minute-long burst searches," *Phys. Rev. D*, vol. 105, no. 8, p. 083007, 2022.

[209] A. Kolmus *et al.*, "Fast sky localization of gravitational waves using deep learning seeded importance sampling," *Phys. Rev. D*, vol. 106, no. 2, p. 023032, 2022.

[210] M. Zevin *et al.*, "Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science," *Class. Quant. Grav.*, vol. 34, no. 6, p. 064003, 2017.

[211] S. Soni *et al.*, "Discovering features in gravitational-wave data through detector characterization, citizen science and machine learning," *Class. Quant. Grav.*, vol. 38, no. 19, p. 195016, 2021.

[212] J. McGinn *et al.*, "Generalised gravitational wave burst generation with generative adversarial networks," *Class. Quant. Grav.*, vol. 38, no. 15, p. 155005, 2021.

[213] M. Lopez *et al.*, "Simulating transient noise bursts in LIGO with generative adversarial networks," *Phys. Rev. D*, vol. 106, no. 2, p. 023027, 2022.

[214] Y. Bengio *et al.*, "Curriculum learning," in *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48, 2009.

[215] D. M. Macleod *et al.*, "GWpy: A Python package for gravitational-wave astrophysics," *SoftwareX*, vol. 13, p. 100657, 2021.

[216] P. Petrov *et al.*, "Data-driven Expectations for Electromagnetic Counterpart Searches Based on LIGO/Virgo Public Alerts," *Astrophys. J.*, vol. 924, no. 2, p. 54, 2022.

[217] P. T. De Boer *et al.*, "A tutorial on the cross-entropy method," *Annals of operations research*, vol. 134, no. 1, pp. 19–67, 2005.

[218] C. Pankow, P. Brady, E. Ochsner, and R. O'Shaughnessy, "Novel scheme for rapid parallel parameter estimation of gravitational waves from compact binary coalescences," *Phys. Rev. D*, vol. 92, no. 2, p. 023002, 2015.

[219] J. Veitch *et al.*, "Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library," *Phys. Rev. D*, vol. 91, no. 4, p. 042003, 2015.

[220] U. Von Toussaint, "Bayesian inference in physics," *Reviews of Modern Physics*, vol. 83, no. 3, p. 943, 2011.

[221] G. Ashton *et al.*, "BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy," *Astrophys. J. Suppl.*, vol. 241, no. 2, p. 27, 2019.

[222] L. P. Singer and L. R. Price, "Rapid Bayesian position reconstruction for gravitational-wave transients," *Phys. Rev. D*, vol. 93, no. 2, p. 024013, 2016.

[223] B. P. Abbott *et al.*, "Tests of general relativity with GW150914," *Phys. Rev. Lett.*, vol. 116, no. 22, p. 221101, 2016. [Erratum: Phys. Rev. Lett. 121, 129902 (2018)].

[224] K. Yamada, T. Narikawa, and T. Tanaka, "Testing massive-field modifications of gravity via gravitational waves," *PTEP*, vol. 2019, no. 10, p. 103E01, 2019.

[225] L. Baiotti, "Gravitational waves from neutron star mergers and their relation to the nuclear equation of state," *Prog. Part. Nucl. Phys.*, vol. 109, p. 103714, 2019.

[226] Z. Doctor *et al.*, "A Search for Optical Emission from Binary-Black-Hole Merger GW170814 with the Dark Energy Camera," *Astrophys. J. Lett.*, vol. 873, no. 2, p. L24, 2019.

[227] R. Perna, D. Lazzati, and W. Farr, "Limits on electromagnetic counterparts of gravitational wave-detected binary black hole mergers," *Astrophys. J.*, vol. 875, no. 1, p. 49, 2019.

[228] L. Barack *et al.*, "Black holes, gravitational waves and fundamental physics: a roadmap," *Class. Quant. Grav.*, vol. 36, no. 14, p. 143001, 2019.

[229] J. Skilling, "Nested sampling for general Bayesian computation," *Bayesian Analysis*, vol. 1, no. 4, pp. 833–859, 2006.

[230] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[231] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[232] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[233] A. Delaunoy *et al.*, "Lightning-Fast Gravitational Wave Parameter Inference through Neural Amortization," *arXiv:2010.12931*, 2020.

[234] G. Papamakarios *et al.*, "Normalizing flows for probabilistic modeling and inference," *Journal of Machine Learning Research*, vol. 22, no. 57, pp. 1–64, 2021.

[235] M. Dax *et al.*, "Real-Time Gravitational Wave Science with Neural Posterior Estimation," *Phys. Rev. Lett.*, vol. 127, no. 24, p. 241103, 2021.

[236] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[237] P. Canizares *et al.*, "Gravitational wave parameter estimation with compressed likelihood evaluations," *Phys. Rev. D*, vol. 87, no. 12, p. 124005, 2013.

[238] X. Fan, J. Li, X. Li, Y. Zhong, and J. Cao, "Applying deep neural networks to the detection and space parameter estimation of compact binary coalescence with a network of gravitational wave detectors," *Sci. China Phys. Mech. Astron.*, vol. 62, no. 6, p. 969512, 2019.

[239] N. I. Fisher, T. Lewis, and B. J. J. Embleton, *Statistical analysis of spherical data*. Cambridge university press, 1993.

[240] R. L. Russell and C. Reale, "Multivariate uncertainty in deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[241] M. Hannam *et al.*, "Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms," *Phys. Rev. Lett.*, vol. 113, no. 15, p. 151101, 2014.

[242] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic gradient descent," in *ICLR: International Conference on Learning Representations*, 2015.

[243] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," in *ICLR: International Conference on Learning Representations*, 2017.

[244] C. Chatterjee *et al.*, "Using Deep Learning to Localize Gravitational Wave Sources," *Phys. Rev. D*, vol. 100, no. 10, p. 103025, 2019.

[245] A. J. K. Chua and M. Vallisneri, "Learning Bayesian posteriors with neural networks for gravitational-wave inference," *Phys. Rev. Lett.*, vol. 124, no. 4, p. 041102, 2020.