



Lensed Gravitational Waves :
Signal Processing and Machine Learning

Luyckx Adrien

Physics department
Sciences faculty
University of Liège

In fulfilment of the requirements for
the degree of Master of Science in Physics
Academic Year 2023-2024

Acknowledgements

First, I would like to thank my thesis advisor, Jean René Cudell, for the time he spent giving me advice and ideas to improve this master's thesis. Second, I want to thank the evaluation committee, Stéphane Dorbolo, Ngoc Duy Nguyen and Peter Schlagheck for the interest they gave in my master's thesis and the time they spent reading it. Third, I also want to underline the help of Srashti Goyal and Maxime Fays who helped me to solve programming issues and with whom I had interesting discussions. Special thanks also to Andrea Virtuoso for his help in understanding the Qp transform. Fourth, thanks to the Virgo Collaboration for the help in accessing the data and also for the opportunity they give me to be part of such an interesting project.

Moreover, I would like to thank my near and dear, family and friends, for the endless support without which I would not have come so far. Particularly, Sacha Peters for all the time we spent working together and discussing machine learning and gravitational wave physics.

Abstract

The focus of this master thesis is to use a new signal processing technique, the Qp transform [1] in a recent machine algorithm LensID [2] to enhance the detection rate of gravitational wave signals lensed by galaxies. First, we will introduce the concept associated with gravitational waves, signal processing and machine learning. Second, we will present one measured signal which is part of a lensing study, GW191105, where the Qp transform helps to distinguish a gravitational wave better. Third, we will show various examples of injected lensed signals which are better recognised with their characteristic chirp evolution. Finally, a discussion of the implementation of this technique with LensID and the associated improvement will be presented. Finally, we summarise our main findings and future prospects.

Contents

1	A Brief Introduction to Gravitational Waves	3
1.1	General Relativity: Basics	3
1.1.1	Notation	3
1.1.2	Metric	3
1.1.3	Christoffel Symbols	3
1.1.4	Geodesic Equation	4
1.1.5	Riemann Curvature Tensor, Ricci Tensor and Scalar	4
1.1.6	Geodesic Deviation Equation	4
1.1.7	Einstein Equation	4
1.2	Gravitational Waves	5
1.2.1	Linearised Equation	5
1.2.2	Vacuum Solution: Gravitational Waves	6
1.2.3	Geodesic Deviation in the Presence of Gravitational Waves	8
1.2.4	Sources	9
1.2.5	Detector: Working Principle and Noises	11
1.3	Lensing in Gravitational Waves	15
2	Signal Processing Technique	19
2.1	Fourier Transform	19
2.2	Whitening	20
2.3	Wavelet Transform	22
2.4	Q Transform	22
2.5	Qp Transform	23
3	Machine Learning	27
3.1	LensID : Spectrogram Analysis	32
4	GW191105	37
4.1	Original Problem	37
4.2	Towards an Origin for the Gap	38
4.2.1	Preprocessing Origin	38
4.2.2	Whitening Origin	39
4.2.3	Transform Origin	40
5	LensID with Qp Transforms	43
5.1	Injected Lensed Signals: Comparison between Transform	43
5.1.1	Comparison between GWpy and PyCBC	44
5.1.2	Others Findings: Q and Qp Transform	46
5.2	Training and Evaluation	50
5.3	Result	53
6	Conclusion	55
A	Access to C00 data	57
B	PyCBC vs GWpy	58
C	Q and Qp Transform	61

D Training : Loss and accuracy	66
E Other models : LensID with Q_p transform	68
References	70

List of Abbreviations

ASD Amplitude Spectral Density

CBC Compact-Binary Coalescences

CNN Convolutional Neural Network

FFT Fast Fourier Transform

FPR False-Positive Rate

GW Gravitational Wave

GWOSC Gravitational-Wave Open-Science Center

KAGRA Kamioka Gravitational-Wave Detector

LIGO Laser Interferometer Gravitational-Wave Observatory

LVK Laser Interferometer Gravitational-Wave Observatory (LIGO) Virgo Kamioka Gravitational-Wave Detector (KAGRA)

O3b Second part of the third observing run

PSD Power Spectral Density

ROC Receiver Operating Characteristic

SNR Signal to Noise Ratio

STFT Short-Time Fourier Transform

TPR True-Positive Rate

TT Transverse Traceless

Introduction

The captivating overlap of theoretical physics and advanced data analysis unfolds in the study of gravitational waves and signal processing techniques. The history of Gravitational Wave (GW) traces back to the 1915 formulation of the general theory of relativity of Albert Einstein [3]. Nevertheless, Einstein himself did not believe in gravitational waves at first [4], in June 1936, the year after he had sent a paper about the EPR paradox, he sent a paper with Rosen in which their calculations are wrong and stated that gravitational waves do not exist. Howard Percy Robertson was at the origin of the criticism against their paper and convinced Einstein that his calculation with Rosen could be modified to yield the correct solution. Predicted as ripples in the fabric of space-time, these waves result from the rate of change in the acceleration of quadrupole objects like merging black holes or neutron stars.

Despite the theoretical brilliance of Einstein, experimental technology took several decades to catch up. The 1970s saw indirect evidence through observations of a binary pulsar system, where the orbit decay of pulsar aligned with expectations of energy loss through gravitational radiation. Russell A. Hulse and Joseph H. Taylor, Jr. received the 1993 Nobel Prize in Physics for this discovery, providing strong indirect support for gravitational waves [5] [6].

Direct detection remained elusive until the Laser Interferometer Gravitational-Wave Observatory (LIGO) became operational in the early 2000s. In September 2015, LIGO made history by detecting gravitational waves from the merger of two black holes, confirming their existence and ushering in a new era in astrophysics [7].

The success of LIGO, along with detectors like Virgo and KAGRA, hinges on sophisticated signal processing techniques. Signal processing is pivotal for extracting meaningful information from the noisy data recorded by these detectors. Researchers utilise matched filtering, time-frequency analysis, and machine learning algorithms to identify and characterise gravitational wave signals within the data.

Matched filtering, inspired by radar signal processing, enhances the signal-to-noise ratio by cross-correlating observed data with theoretically predicted waveforms [8]. Time-frequency analysis, employing methods like spectrograms, studies the frequency evolution over time of the signal, offering insights into the astrophysical source. An interesting method has also emerged for gravitational wave physics, the Q transform, which offers a better representation of transient signals [9].

The history of signal processing intertwines with communication systems, radar technology, and the emerging field of gravitational wave astronomy. Mathematical foundations find roots in the work of mathematicians like Claude Shannon, the father of information theory. His 1948 paper "A Mathematical Theory of Communication" introduced concepts like entropy, setting the stage for signal processing formalisation [10].

Mathematical tools, especially Fourier analysis dating back to the 19th-century works by Jean-Baptiste Joseph Fourier [11], play a crucial role. The Fourier transform became

foundational, allowing the representation of signals in the frequency domain, and facilitating the analysis of complex signals by decomposing them into simpler sinusoidal components.

The advent of computers in the mid-20th century played a pivotal role, with the development of fast Fourier transform (FFT) algorithms of Cooley and Tukey in the 1960s revolutionising efficiency [12]. In the context of gravitational wave astronomy, signal processing techniques became indispensable for interferometric detectors like LIGO.

The rich history of signal processing includes luminaries like Alan V. Oppenheim and Ronald W. Schafer, whose book "Discrete-Time Signal Processing" significantly impacted the domain, introducing key concepts and algorithms [13].

As gravitational wave detectors push sensitivity and accuracy boundaries, signal processing remains at the forefront, extracting meaningful information. Machine learning automates detection and classification, boosting the efficiency of data analysis pipelines with the use of many algorithms, often called pipelines by the LIGO-Virgo-KAGRA (LVK) collaboration, which helps recover signals in the various noises present in the detectors.

Interdisciplinary collaboration between physicists, engineers, and mathematicians continues to drive innovations in both gravitational wave astronomy and signal processing, creating a symbiotic relationship that enhances our ability to explore the universe.

1 A Brief Introduction to Gravitational Waves

In this section, we give a brief introduction to General Relativity which helps to understand the gravitational waves we are going to study later. The developments presented here are taken from [14].

1.1 General Relativity: Basics

1.1.1 Notation

Usually, when one talks about General Relativity, one introduces the Einstein summation convention to simplify equations. The main idea is to hide summation symbols and understand that repeated indices (one index up and one down) imply a summation. For example,

$$\vec{v} = \sum_{i=1}^3 v^i e_i \equiv v^i e_i \qquad \mathbf{x} = \sum_{\mu=0}^3 x^\mu e_\mu \equiv x^\mu e_\mu \qquad (1.1)$$

In addition, we add that Latin letters will be associated with 3 components (1 to 3), space, while Greek letters are used for 4 (0 to 3), time and space. Moreover, letter with a vector, such as \vec{v} , has 3 components and bold ones, such as \mathbf{x} , 4. We will also use a comma to denote partial derivative of a quantity such that,

$$\frac{\partial}{\partial x_j} v_i \equiv v_{i,j} \qquad (1.2)$$

1.1.2 Metric

The metric or metric tensor is one key ingredient for General Relativity, it gives a notion of physical distances, more precisely it is the square of an infinitesimal distance between two points. Often one writes it as,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \qquad (1.3)$$

ds^2 is here the squared distance, $g_{\mu\nu}$ the metric tensor and dx^μ the dual basis of $\frac{\partial}{\partial x_\mu}$, a coordinate basis. In special relativity, one finds that $g_{\mu\nu} = \eta_{\mu\nu}$ with,

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Here we will use the first one.

1.1.3 Christoffel Symbols

The Christoffel symbols $\Gamma_{\mu\nu}^\rho$, also known as connection coefficients, play a key role in describing how vectors change as they are parallel transported around space-time. They are given by

$$\Gamma_{\mu\nu}^\rho = \frac{1}{2} g^{\rho\sigma} (\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}) \qquad (1.4)$$

1.1.4 Geodesic Equation

The geodesic equation describes the path of free-falling objects in space-time. It can be expressed as

$$\frac{d^2x^\rho}{d\tau^2} + \Gamma_{\mu\nu}^\rho \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0 \quad (1.5)$$

where τ is the proper time along the geodesic.

1.1.5 Riemann Curvature Tensor, Ricci Tensor and Scalar

The Riemann curvature tensor $R_{\sigma\mu\nu}^\rho$ is a measure of the curvature of space-time, defined as

$$R_{\sigma\mu\nu}^\rho = \partial_\mu \Gamma_{\nu\sigma}^\rho - \partial_\nu \Gamma_{\mu\sigma}^\rho + \Gamma_{\mu\lambda}^\rho \Gamma_{\nu\sigma}^\lambda - \Gamma_{\nu\lambda}^\rho \Gamma_{\mu\sigma}^\lambda \quad (1.6)$$

The Ricci tensor is obtained by contracting the Riemann tensor, $R_{\mu\nu} = R_{\mu\lambda\nu}^\lambda$, while the Ricci scalar is obtained by contracting the Ricci tensor, $R = R_{\mu}^\mu$. Note also that,

$$R_{\mu\nu} = \partial_\rho \Gamma_{\nu\mu}^\rho - \partial_\nu \Gamma_{\mu\rho}^\rho + \Gamma_{\mu\nu}^\sigma \Gamma_{\sigma\rho}^\rho - \Gamma_{\mu\rho}^\sigma \Gamma_{\sigma\nu}^\rho \quad (1.7)$$

It is well known [15] that the Riemann tensor has in general 14 independent components and the special case of $R_{\sigma\mu\nu}^\rho = 0$ imply a flat space-time. In addition to this, we note $R_{\mu\nu} = 0$, impose more conditions on the Riemann tensor, and we observe only 10 independent components. This last case is important because it holds for gravitational waves propagating in vacuum.

1.1.6 Geodesic Deviation Equation

We can now introduce the deviation of geodesics, this equation describes the relative motion of nearby particles moving along neighbouring geodesics in curved space-time. Roughly, the equation quantifies how the separation between two geodesics changes over time due to the curvature of space-time.

$$A^\nu = R_{\mu\sigma\lambda}^\nu T^\mu T^\sigma X^\lambda \quad (1.8)$$

where we have A^ν the relative acceleration of infinitesimally close geodesics, T^μ a tangent vector to the geodesics and X^λ an infinitesimal displacement between geodesics.

1.1.7 Einstein Equation

The Einstein field equations form the core of General Relativity, relating the geometry of space-time to the distribution of matter within it. They are given by

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \quad (1.9)$$

where $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ is the Einstein tensor, $R_{\mu\nu}$ is the Ricci tensor, R is the Ricci scalar, Λ is the cosmological constant, G is the gravitational constant, c is the speed of light, and $T_{\mu\nu}$ is the stress-energy tensor. Mathematically, the Einstein equation can be expressed, in a coordinate basis as a set of coupled non-linear partial differential equations for the metric.

1.2 Gravitational Waves

Knowing the basic principles of General Relativity, we can now focus on a peculiar solution, gravitational waves. Gravitational waves (GW) represent ripples in the fabric of space-time, a profound consequence of the general theory of relativity of Albert Einstein. As we are going to see, these waves are generated by the rate of change in the quadrupole of massive objects, leading to fluctuations in the geometry of space and time.

1.2.1 Linearised Equation

The mathematical formulation of gravitational waves begins with the field equations of Einstein. Even though, researchers have managed to show exact solution interpreted as gravitational wave [16], we will present here a derivation which involves perturbing the metric tensor and linearising the field equations, resulting in the wave equation for gravitational waves. Mathematically, this process explores the dynamics of the space-time metric under small perturbations.

We will assume that the propagation of gravitational waves happens as a small gravitational field on top of a flat metric. Let $g_{\mu\nu}$ represent the metric tensor and $h_{\mu\nu}$ denote the perturbation. For example, in the solar system we have $h_{\mu\nu} \lesssim \frac{GM_{\odot}}{R_{\odot}c^2} \approx 10^{-6}$. We thus have,

$$g_{\mu\nu} \approx \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1$$

When we want to linearise the Einstein field equation, we need to compute the Christoffel symbols which are here,

$$\Gamma_{\mu\nu}^{\alpha} = \frac{1}{2}\eta^{\alpha\beta}(h_{\mu\beta,\nu} + h_{\beta\nu,\mu} - h_{\mu\nu,\beta}) \quad (1.10)$$

We then note that the Ricci tensor becomes more simple as quadratic Christoffel symbols, see equation (1.4), are either zero or quadratic in $h_{\mu\nu}$, so that at first order in $h_{\mu\nu}$,

$$R_{\sigma\nu\rho}^{\mu} = \partial_{\nu}\Gamma_{\rho\sigma}^{\mu} - \partial_{\rho}\Gamma_{\nu\sigma}^{\mu} \Rightarrow R_{\mu\nu} = \partial_{\rho}\Gamma_{\nu\mu}^{\rho} - \partial_{\nu}\Gamma_{\mu\rho}^{\rho} \quad (1.11)$$

Assuming that here $\eta_{\mu\nu}$ lowers and $\eta^{\mu\nu}$ raises indices. We define $\square = \partial_{\lambda}\partial^{\lambda}$ and have at linear order,

$$R_{\mu\nu} = \frac{1}{2}(h_{\mu,\lambda\nu}^{\lambda} - \square h_{\mu\nu} - h_{\lambda,\mu\nu}^{\lambda} + h_{\nu,\lambda\mu}^{\lambda}) \quad (1.12)$$

If we take the approximation that there is no accelerated expansion of the Universe (i.e. $\Lambda = 0$) and compute the Ricci scalar, we can then write the linearised Einstein field equations as,

$$\square h_{\mu\nu} + h_{,\mu\nu} - h_{\mu,\lambda\nu}^{\lambda} - h_{\nu,\lambda\mu}^{\lambda} - \eta_{\mu\nu}\square h + \eta_{\mu\nu}h_{,\lambda\sigma}^{\lambda\sigma} = -\frac{16\pi G}{c^4}T_{\mu\nu} \quad (1.13)$$

We then define $\gamma_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h$ or equivalently $h_{\mu\nu} \equiv \gamma_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\gamma$ and obtain,

$$-\square\gamma_{\mu\nu} - \eta_{\mu\nu}\gamma_{\alpha\beta}^{\alpha,\beta} + \gamma_{\mu\alpha,\nu}^{\alpha} + \gamma_{\nu\alpha,\mu}^{\alpha} = -\frac{16\pi G}{c^4}T_{\mu\nu} \quad (1.14)$$

In addition, the theory, via the Einstein tensor, is invariant under coordinate transformation and we can pick the gauge transformation, see [14] for details ,

$$h_{\mu\nu} \rightarrow h'_{\mu\nu} + \xi_{\mu,\nu} + \xi_{\nu,\mu} \quad (1.15)$$

which can be interpreted as an *infinitesimal coordinate transformation* of the form,

$$x^\mu = x'^\mu - \xi^\mu(x) \quad (1.16)$$

Writing it in terms of $\gamma_{\mu\nu}$ and derivating one times, one gets,

$$\gamma'_{,\nu}{}^{\mu\nu} = \gamma_{,\nu}{}^{\mu\nu} + \square\xi^\mu + \xi_{,\nu}{}^{\nu,\mu} - \gamma'_{,\lambda}{}^{\lambda,\mu} = \gamma_{,\nu}{}^{\mu\nu} + \square\xi^\mu \quad (1.17)$$

We can always find a solution to $\gamma_{,\nu}{}^{\mu\nu} = -\square\xi^\mu$. In this gauge, often called Harmonic, Lorenz, Hilbert or even de Donder gauge, we can thus change $h_{\mu\nu}$ accordingly to equation 1.15 and also find,

$$\gamma_{,\beta}{}^{\alpha\beta} = 0 \quad (1.18)$$

It might remind you the electromagnetism, where the Lorenz gauge is given by $A_\mu{}^{,\mu} = 0$ in a way that we get the wave equation $\square A_\mu = 0$ for the four-potential in vacuum.

In contrast, we here find that the linearised Einstein field equations yield the wave equation,

$$\square\gamma_{\mu\nu} = -\frac{16\pi G}{c^4}T_{\mu\nu} \quad (1.19)$$

Here, \square is the d'Alembertian operator, G is the gravitational constant, c is the speed of light, and $T_{\mu\nu}$ is the energy-momentum tensor. A general solution of this equation takes the form,

$$\gamma_{\mu\nu}(x) = \frac{4G}{c^3} \int \frac{\delta(cx^0 - cy^0 - |\mathbf{x} - \mathbf{y}|)}{|\mathbf{x} - \mathbf{y}|} T_{\mu\nu}(y) d^4y = \frac{4G}{c^4} \int \frac{T_{\mu\nu}(x^0 - \frac{|\mathbf{x}-\mathbf{y}|}{c}, \mathbf{y})}{|\mathbf{x} - \mathbf{y}|} d^3y \quad (1.20)$$

We see in this solution that we have a retarded perturbation propagating at the speed of light. We will discuss vacuum solution in Section 1.2.2 and others in Section 1.2.4.

1.2.2 Vacuum Solution: Gravitational Waves

We will now focus on the vacuum solution which assumes $T_{\mu\nu} = 0$. In that case, the Einstein equations give,

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 0 \Leftrightarrow R_{\mu\nu} = 0 \quad (1.21)$$

We note here that all the physics lies inside the Riemann tensor which is contracted here to obtain the Ricci tensor.

When we pick the Hilbert gauge, we still have some degree of freedom and can pick $\gamma_\mu^\mu = 0$ (traceless), we are not going to detail this but explanations can be found in [14]. In this gauge, $\gamma_{\mu\nu} = h_{\mu\nu}$, thus our equation becomes,

$$\square h_{\mu\nu} = 0 \quad (1.22)$$

We can write the solution as a superposition of plane wave,

$$h_{\mu\nu} = \text{Re}(\epsilon_{\mu\nu} e^{ik_\sigma x^\sigma}) \quad (1.23)$$

From equation (1.18), (1.22) and the traceless condition, we find $k_\mu \epsilon_\nu^\mu = 0, k^2 = k_\mu k^\mu = 0$ and $\epsilon_\nu^\nu = 0$. We call $\epsilon_{\mu\nu}$ the polarisation vector of the gravitational wave.

Along with these conditions, in this Hilbert gauge, we still have the choice for $\xi^\mu(x)$ and pick $\xi^\mu(x) = \text{Re}(-i\epsilon^\mu e^{ik_\sigma x^\sigma})$ with a wave propagating in the third space direction, $k^\mu = (k, 0, 0, k)$. One can show that the resulting polarisation vector is,

$$\epsilon_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \epsilon_{11} & \epsilon_{12} & 0 \\ 0 & \epsilon_{12} & -\epsilon_{11} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (1.24)$$

We have here found an example of transverse, traceless tensor, often called Transverse Traceless (TT) gauge. In general, the TT gauge is thus a peculiar Hilbert gauge defined by,

$$h_{\mu 0} = 0 \quad \sum_k h_{kk} = 0 \quad h_{kj,j} = 0 \quad (1.25)$$

Note that here, as for light, we have two linearly independent polarisation states, the main difference between light and gravitational wave resides in their transformation under rotation. For example, if we make a spatial rotation by an angle ϕ about the third space dimension, we obtain,

$$\epsilon'_{\mu\nu} = R_\mu^\alpha(\phi) R_\nu^\beta(\phi) \epsilon_{\alpha\beta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \phi & \sin \phi & 0 \\ 0 & -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \phi & \sin \phi & 0 \\ 0 & -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \epsilon'_{\mu\nu} \quad (1.26)$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \epsilon_{11} \cos 2\phi + \epsilon_{12} \sin 2\phi & -\epsilon_{11} \sin 2\phi + \epsilon_{12} \cos 2\phi & 0 \\ 0 & -\epsilon_{11} \sin 2\phi + \epsilon_{12} \cos 2\phi & -\epsilon_{11} \cos 2\phi - \epsilon_{12} \sin 2\phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1.27)$$

We know that light is invariant under rotation of 360° , the objects we have discovered here, the gravitational waves, are different from light because they are also invariant under rotation of 180° .

1.2.3 Geodesic Deviation in the Presence of Gravitational Waves

Geodesic deviation provides a framework for understanding how gravitational waves affect the motion of particles. In the presence of a gravitational wave, nearby geodesics experience tidal forces, leading to the stretching and squeezing of space-time. We will look at equation (1.8) in the specific case of test particles at rest and follow them along ∂_t with separation vector X^μ , we then have,

$$\frac{d^2 X^i}{dt^2} = -R^i_{0j0} X^j \quad (1.28)$$

Remember, $R_{\mu\sigma\nu\rho} = g_{\mu\lambda} R^\lambda_{\sigma\nu\rho}$, so that here,

$$R_{i0j0} = \frac{1}{2}(h_{0j,0i} + h_{i0,j0} - h_{ij,00} + h_{00,ij}) = -\frac{1}{2}h_{ij,00}^{TT} \quad (1.29)$$

Where we use the TT gauge in the last equality so that finally, by lowering the i index,

$$\frac{d^2 X_i}{dt^2} = \frac{1}{2} \frac{\partial^2 h_{ij}^{TT}}{\partial t^2} X^j \quad (1.30)$$

Assuming that the test particles are initially at rest with coordinate $X_{i(0)}$ when $h_{ij} = 0$ and evolve along a geodesic $\zeta(t)$, we find a solution,

$$X_i(\tau) \approx X_{i(0)} + \frac{1}{2} h_{ij}(\zeta(t)) X_{(0)}^j \quad (1.31)$$

We can focus on particles aligned in the third direction, $\vec{X} = (0,0,a)$, with different a for different particles. In the case where we pick a perturbation of the form of equation (1.24) which is often written with $\epsilon_{11} = h_+$ and $\epsilon_{12} = h_\times$. In the case of a periodic plane wave propagating in the z -direction, we have,

$$h_+ = \text{Re}[A_+ e^{-i\omega(t-z)}] \quad \text{and} \quad h_\times = \text{Re}[A_\times e^{-i\omega(t-z)}] \quad (1.32)$$

We see that $h_{ij} X_{(0)}^j = 0$ meaning the perturbations are always transverse to the direction of the gravitational waves. We thus rewrite the equation (1.30) as,

$$\begin{pmatrix} \ddot{X}^1 \\ \ddot{X}^2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \ddot{h}_+ & \ddot{h}_\times \\ \ddot{h}_\times & -\ddot{h}_+ \end{pmatrix} \begin{pmatrix} X_{(0)}^1 \\ X_{(0)}^2 \end{pmatrix} \quad (1.33)$$

As h_{ij} is real symmetric, we can diagonalise with an orthogonal matrix R and write,

$$\begin{pmatrix} \ddot{X}^1 \\ \ddot{X}^2 \end{pmatrix} = \frac{1}{2} R \begin{pmatrix} \Omega & 0 \\ 0 & -\Omega \end{pmatrix} R^T \begin{pmatrix} X_{(0)}^1 \\ X_{(0)}^2 \end{pmatrix} \quad (1.34)$$

Defining the displacement along the principal direction,

$$\begin{pmatrix} \chi \\ \kappa \end{pmatrix} = R^T \begin{pmatrix} X^1 \\ X^2 \end{pmatrix}, \quad \begin{pmatrix} \chi_0 \\ \kappa_0 \end{pmatrix} = R^T \begin{pmatrix} X_{(0)}^1 \\ X_{(0)}^2 \end{pmatrix} \quad (1.35)$$

We find a "quadrupole" type oscillation, which can be imaged as in Figure 1 defined by,

$$\chi \approx \chi_0 + \Omega(t)\chi_0, \quad \kappa \approx \kappa_0 - \Omega(t)\kappa_0 \quad (1.36)$$

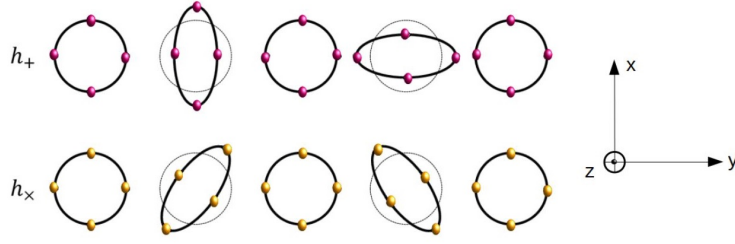


Figure 1: Gravitational effect on test particles with "quadrupole" type oscillation, from [17].

1.2.4 Sources

To have an idea of how gravitational waves are produced, we will develop a mathematical view of their generation process and then describe the known sources and present search. We now go back to the Hilbert gauge and take equation (1.20) for a localised and distant sources with $|\mathbf{x}| = r$,

$$\gamma_{\mu\nu}(t, \mathbf{x}) = \frac{4G}{c^4 r} \int T_{\mu\nu}(t - r, \mathbf{y}) d^3\mathbf{y} \quad (1.37)$$

From the linearised conservation law $T_{,\nu}^{\mu\nu} = 0$ and under the assumption of nearly Newtonian sources for which the energy density is dominated by the matter density ρ ($T^{00} \simeq \rho c^2$), one can find the spatial perturbation [14],

$$\gamma^{kj}(t, \mathbf{x}) = \frac{2G}{c^4 r} \frac{\partial^2}{\partial t^2} \int \rho(t - r, \mathbf{y}) y^k y^j d^3\mathbf{y} \quad (1.38)$$

We see with this equation that the first multipole associated with a gravitational wave is the quadrupole contrary to electrodynamics where one can have radiation with a dipole. For completeness, we give the quadrupole formula of Einstein which gives the luminosity observed for a given source,

$$L_{GW} = \frac{G}{5c^5} \langle \ddot{Q}^{ij} \ddot{Q}_{ij} \rangle \quad (1.39)$$

where angle brackets denote time averages over several characteristic periods of the source and Q is the traceless quadrupole tensor,

$$Q_{kj}(t) = \int (3x_k x_j - r^2 \delta_{kj}) \rho(t, \mathbf{x}) d^3x \quad (1.40)$$

So that in the TT gauge, one finds also,

$$h_{jk}^{TT}(t, r) = \frac{2G}{3c^4 r} \ddot{Q}_{jk}^{TT}(t - r) \quad (1.41)$$

To estimate the strain power associated with a gravitational wave, one can assume $\ddot{Q} \simeq Mv^2$ where M is the mass and v the orbital speed of binaries objects, the period under which these objects orbit each others is defined as T , we thus write those equations as,

$$h \simeq \frac{G}{c^4} \frac{Mv^2}{r} \quad (1.42)$$

$$L_{GW} \simeq \frac{G}{c^5} \left(\frac{Mv^2}{T} \right)^2 \quad (1.43)$$

We see in equation (1.42) that the perturbation measured in detectors is proportional the the energy of the sources and becomes smaller with the distance of the source. Near the merger, for black holes binaries of 10 solar masses travelling near the speed of light at 100 Megaparsec, we find a strain around 10^{-21} which matches with recent measurements. We see in equation (1.43) that the emitted power depends on the period on which the system oscillates meaning that quadrupole oscillating with a higher frequency would emit more energy per unit time.

On the one hand, the known and detected sources leading to gravitational waves are black hole binaries, neutron star binaries and black hole neutron-star binaries, lose energy by emitting gravitational radiation and thus their orbit shrinks with time until collision. One often called those type of object, the Compact-Binary Coalescences (CBC) as one can seen in Figure 2.

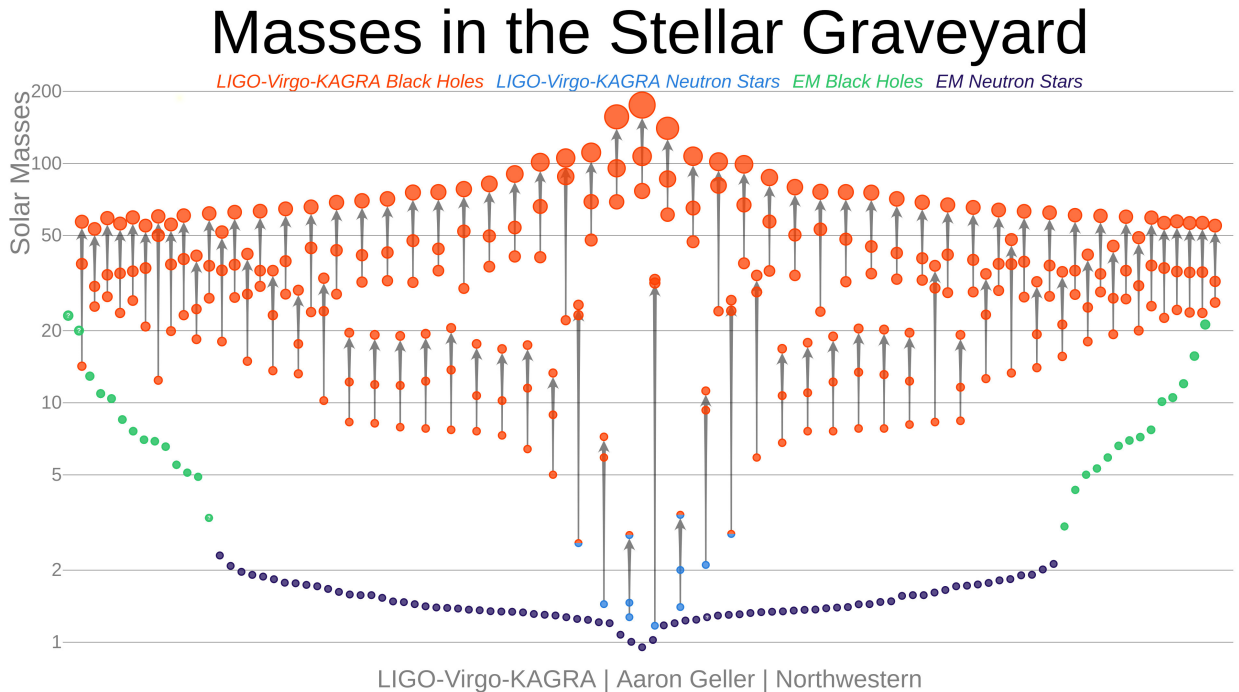


Figure 2: Detected CBC signals. Legend is included in the image to show the black hole/neutron star origin with an arrow explaining the merging scenario, the first 2 circles represent initial objects and the third one represents the final object, taken from the LIGO Virgo KAGRA (LVK) Collaboration made by Aaron Geller.

On the other hand, undetected and investigated sources are for example gravitational waves coming from white dwarf binaries, fast rotating neutron stars which are not symmetric due to their rotation. There should also exist a stochastic background of mixed gravitational waves from which we can only extract statistical properties [18], their origin depends on the frequency bands of the detectors, they can come from binaries or early universe sources for example [19].

1.2.5 Detector: Working Principle and Noises

Gravitational wave detectors, such as the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo, operate based on the principle of interferometry. A laser beam is split and sent along two perpendicular arms, each equipped with a highly reflective mirror suspended as pendulum. Gravitational waves cause tiny changes in the arm lengths in the sense discussed in Section 1.2.3, altering the interference pattern when the beams are recombined, see Figure 6. By precisely measuring these changes, detectors can identify and characterise gravitational wave signals.

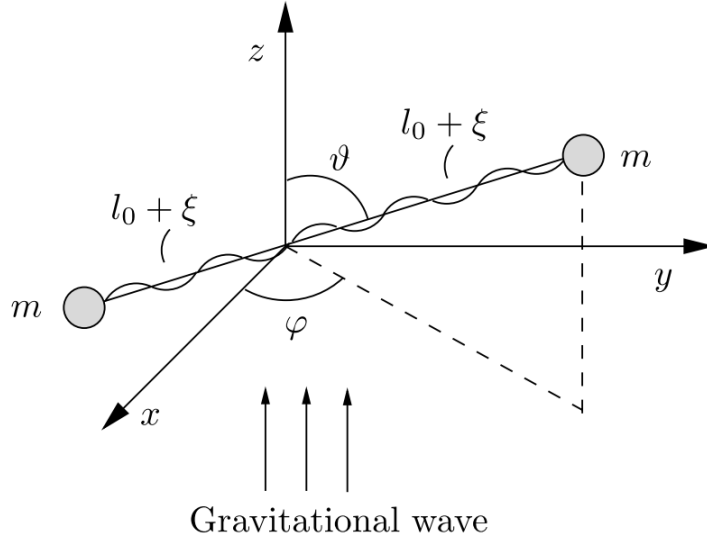


Figure 3: Two masses connected by a spring as an idealised gravitational wave detector, from [14].

To understand a bit more the physics behind the detection, let us describe a toy model, depicted in Figure 3 in which two masses m are connected by a massless spring of equilibrium length $2l_0$. The vector connecting the two masses forms an angle θ with the z -axis (direction of propagation of our gravitational wave) and ϕ with the x -axis. The gravitational wave produce a tidal acceleration T_{GW} on the test masses. Let ω_0 be the frequency of the oscillator, and we assume that the damping time $\tau_0 \ll 1/\omega_0$. The separation of the masses is $2(l_0 + \xi(t))$, where the relative separation $\xi(t)$ will satisfy the differential equation,

$$\ddot{\xi} + \frac{1}{\tau_0} \dot{\xi} + \omega_0^2 \xi = T_{GW} \quad (1.44)$$

In the case where we assume only polarisation along x and y ($A_x = 0$), from equation (1.32) and (1.33), when projected onto the detector (giving a factor $l_0 \sin^2 \theta \cos^2 2\phi$ due to the rotation about the z and y -axis), the tidal acceleration is,

$$T_{GW} = -\frac{1}{2} \omega^2 \text{Re}\{A_+ e^{-i\omega t}\} l_0 \sin^2 \theta \cos^2 2\phi \quad (1.45)$$

The driven solution of the equation is thus,

$$\xi = \text{Re}\{A e^{-i\omega t}\} \quad \text{with} \quad A = \frac{1}{2} A_+ l_0 \sin^2 \theta \cos^2 2\phi \frac{\omega^2}{\omega^2 - \omega_0^2 + i\omega/\tau_0} \quad (1.46)$$

We see here the transverse nature of gravitational wave in the term $\sin^2 \theta$ (perpendicular wave to the detector would not make perturbation). Moreover, we see that the amplitude of the wave is related to the relative displacement (i.e. $A_+ \simeq \frac{\xi}{l_0}$), this tends to explain why we usually define the strain of a gravitational wave to be,

$$h \equiv \frac{\Delta l}{l_0} \tag{1.47}$$

We understand that the gravitational wave through A_+ and ω is related to the change in length of the oscillator but also by its orientation with respect to the detector via θ and ϕ .

To give an idea of the typical strain observed now and in the future, we depicted in Figure 4 the typical sensitivity curve of certain detectors but also some astrophysical sources, using an online software developed within the LVK collaboration [20].

Unfortunately, gravitational waves are immersed in background noise, thus the success of gravitational wave detection relies on the comprehension of various sources of noise that can mask or mimic true signals. Understanding and minimising noise are crucial for enhancing the sensitivity of detectors.

Noise in gravitational wave physics is a wide subject and we will not take the time to detail this here but will simply go through some of the known noises [21], [22], [23], [24]. For example, we will look at some of the noises seen in the Advanced LIGO detector which consists in a Michelson interferometer with mirror being our test masses as depicted in Figure 5 and 6.

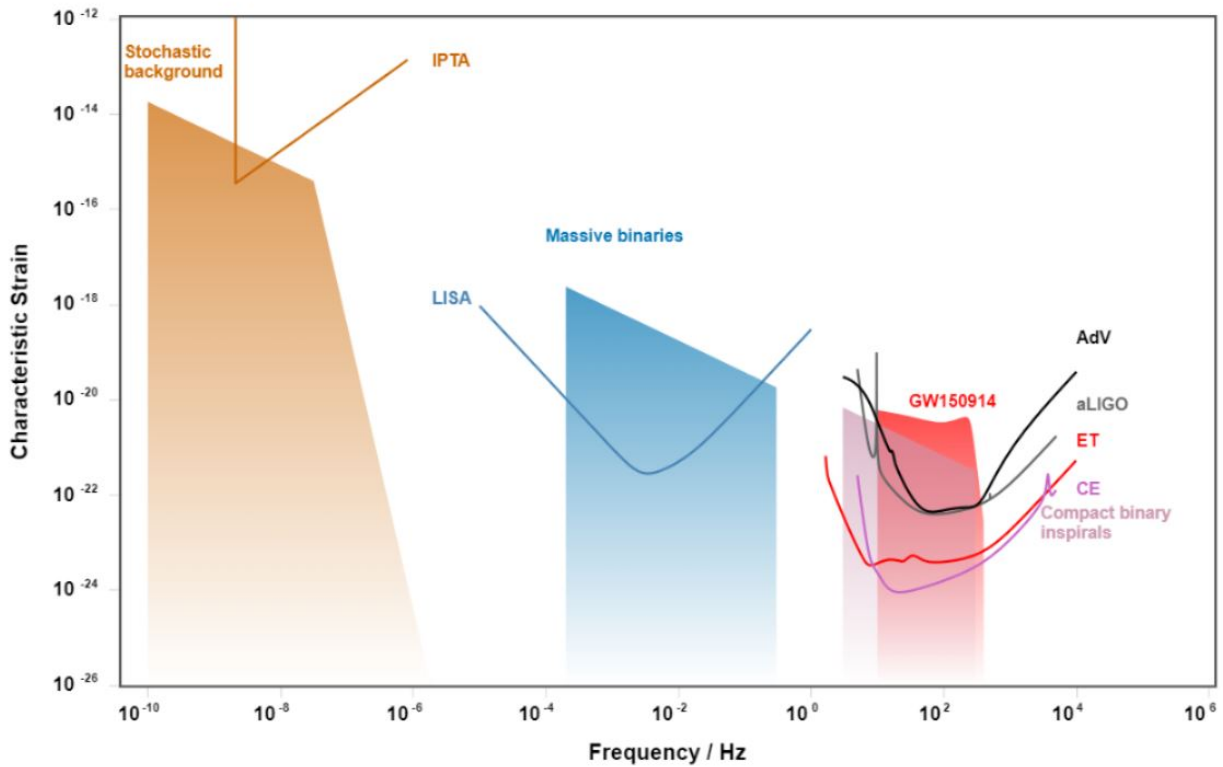


Figure 4: Sensitivity curve for different detectors and different frequency ranges. Current detectors, Advanced Virgo (AdV), Advanced LIGO(aLIGO) and International Pulsar Timing Array (IPTA) are shown but also next-generation detectors such as LISA, Einstein Telescope (ET) and Cosmic Explorer (CE). In addition, we have added typical strain observed for different classes of signals such as the stochastic background, massive binaries and compact binary inspirals but also the first signal detected GW150914.

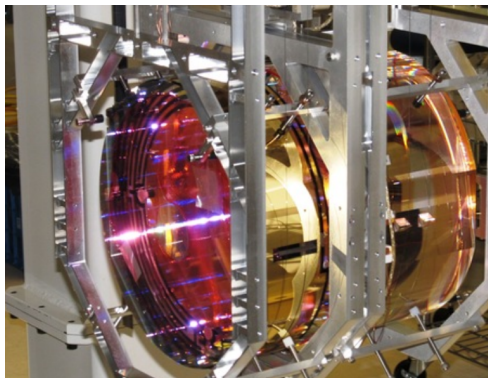


Figure 5: A 40 kg mirror from LIGO detectors, from [25].

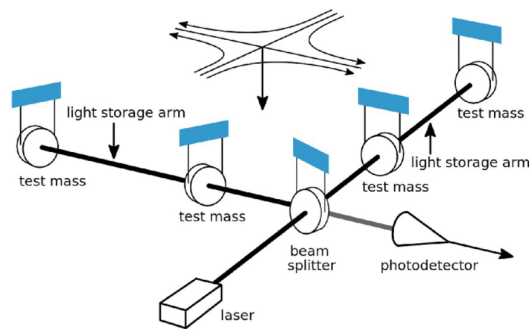


Figure 6: Sketch of the working principle of the LIGO interferometer. Light is sent into a beam splitter and then mirrors (test masses) reflect or let light go through to finish again in the beam splitter and form an interference pattern which changes with noise or with gravitational wave signals, from [25].

We have illustrated some of the associated noises in Figure 7 with the non-exhaustive list of the noise,

1. Quantum noise takes into account the effects of statistical fluctuations in detected photon arrival rate (shot noise) and radiation pressure due to photon number fluctuations.
2. Seismic noise comes from the transmission of vibrations to the suspension systems of the test masses or from seismic waves for example.
3. Newtonian noise sometimes called gravity gradient comes also from seismic waves which produce density perturbations in the earth close to the test masses, which in turn produce fluctuating gravitational forces on the masses.
4. Suspension thermal noise denotes the thermal fluctuation in the suspension which holds the test masses leading for example to the 510 Hz fundamental violin mode.
5. Coating Brownian, coating Thermo-optic and substrate Brownian noises arise from mechanical and elastic dissipation in the coatings.
6. Excess gas noise in the 4 km long beam tubes will lead to statistical variations in the column density of gas particles in the beam path. Indeed, even if the tubes contain one of the largest and cleanest vacuum system in the world, the remaining particles will make light scatter along its path and perturb the measurements.

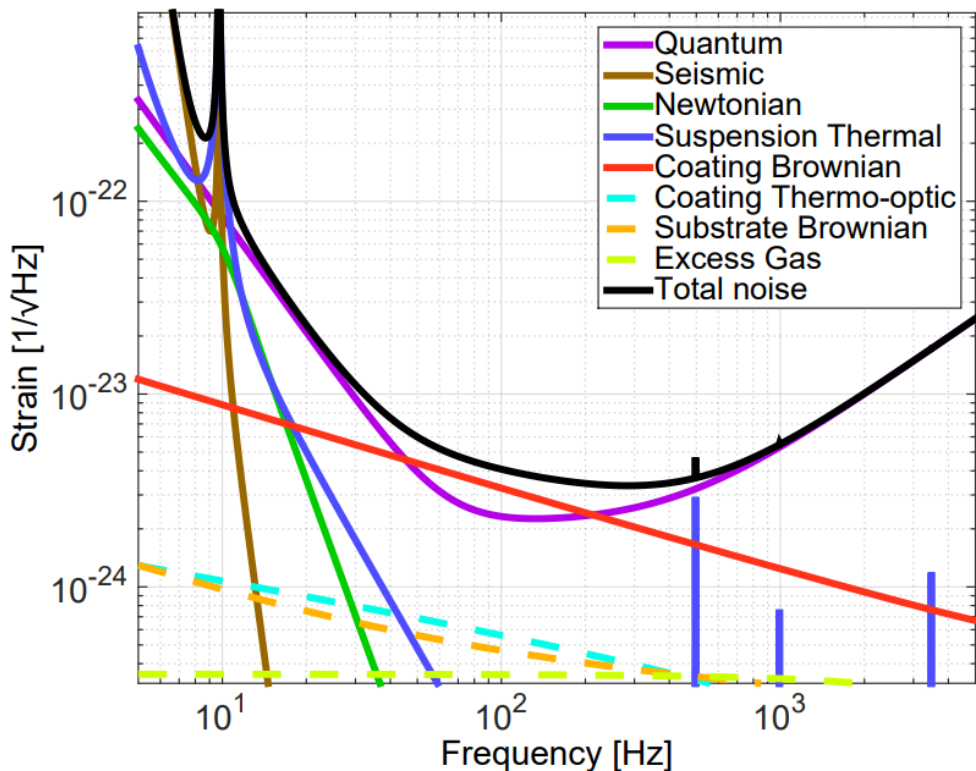


Figure 7: Sketch of the different noises for the Advanced LIGO detector, from [26].

Moreover, other noise sources include power lines, trucks passing close to the detector, and imperfections in the detector components. Advanced signal processing techniques are employed to distinguish true gravitational wave signals from the noise, this will be discussed in Section 2.

1.3 Lensing in Gravitational Waves

Now let us describe rapidly the properties of lensing as we are going to search gravitational wave lensing later. On the one hand, lensing in astrophysics is known to happen for light when going near large masses, we will use the development coming from [27], [28] and [29] without deriving them explicitly to give the physical intuition behind the phenomena. In Figure 8, one can observe the path taken by light when going through a galaxy cluster, explaining why we can see multiple images. In addition, we show in Figure 9, different measurements made for light deviation.

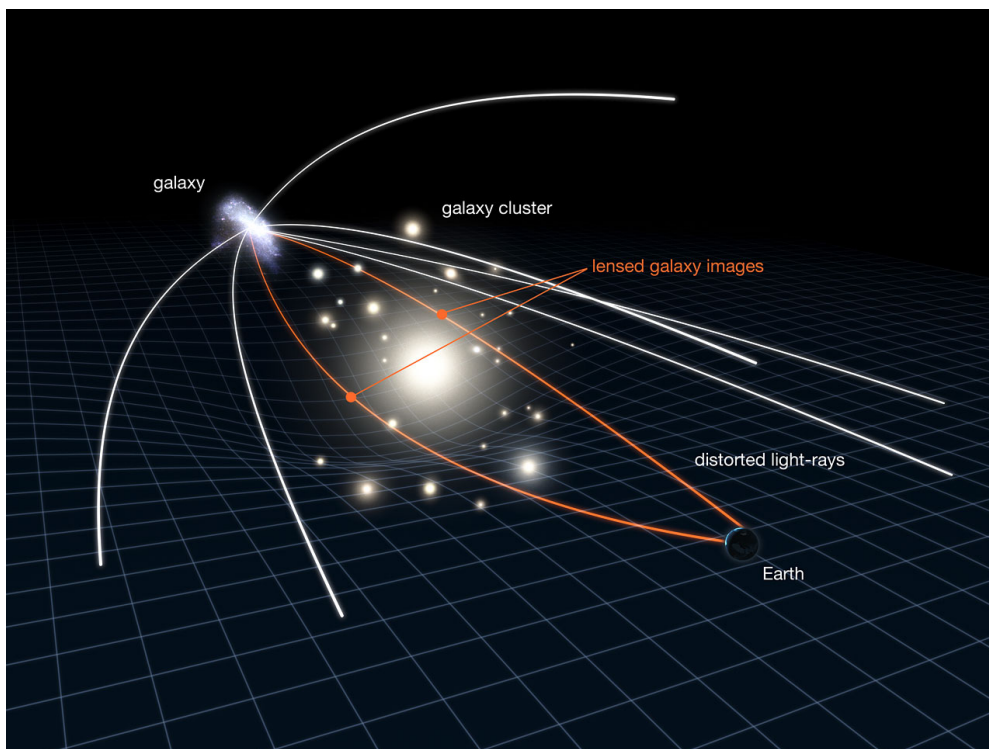


Figure 8: Sketch of the gravitational lensing observed for light when going through a galaxy cluster. Credit: NASA, ESA.

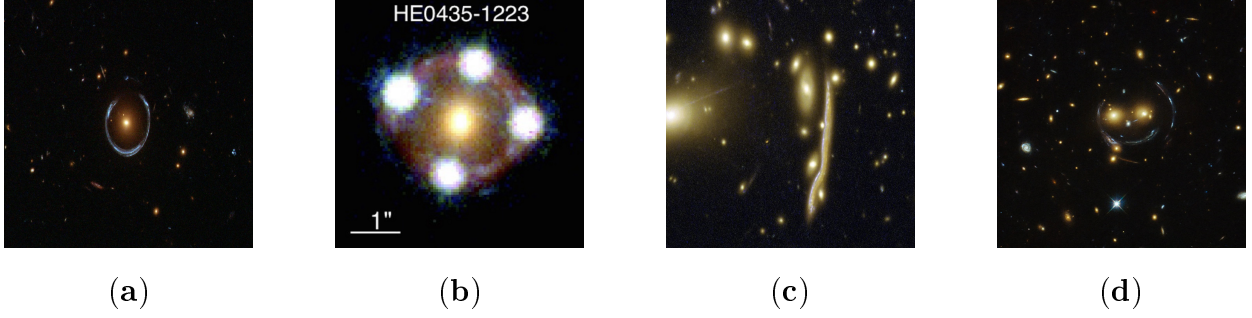


Figure 9: Examples of lensing observed for light. **(a)** Einstein ring, **(b)** Einstein cross, **(c)** Cosmic snake, **(d)** Smiley face. Credit: ESA/Hubble, NASA.

On the other hand, from now on, we will focus on lensing that happens on the gravitational waves. There exist different regimes depending on the size of the wavelength in comparison to that of the system in which it passes [30] and also depending on the density contrast [27],

$$\delta_L = \frac{\rho_L - \rho_{\text{univ}}^{\text{av}}}{\rho_{\text{univ}}^{\text{av}}} \quad (1.48)$$

where ρ_L and $\rho_{\text{univ}}^{\text{av}}$ are the density of the lens and that of the average universe. There exist two main categories, strong and weak lensing. Strong lensing occurs when the density of the object deforming the path of the wave, the lens, is much larger than the one of the average universe, i.e. $\delta \gg 1$, this lensing leads to images we can observe in Figure 9. People sometimes refer to milli and micro lensing as the strong lensing coming from star-size objects leading to images split by a milli or micro arcsecond. While weak lensing occurs when $\delta > 1$ and might sometimes come from $\delta < 0$, this leads to small deformation of the images, for example, a circle might become an ellipse. Those lensing types can also be described through the size of their wavelength such that micro lensing happens when the wavelength is greater or comparable to the size of the object (often described via their Schwarzschild radius $R_{Schw} = \frac{2GM}{c^2}$) while milli and strong lensing occurs when the geometric optics approximation holds (i.e. when the wavelength is much smaller than the object size), one can see how their strain changes in Figure 10. Micro lensing often presents changing amplitudes, milli lensing presents beating patterns and other kinds of strong lensing simply present time separation between their signals.

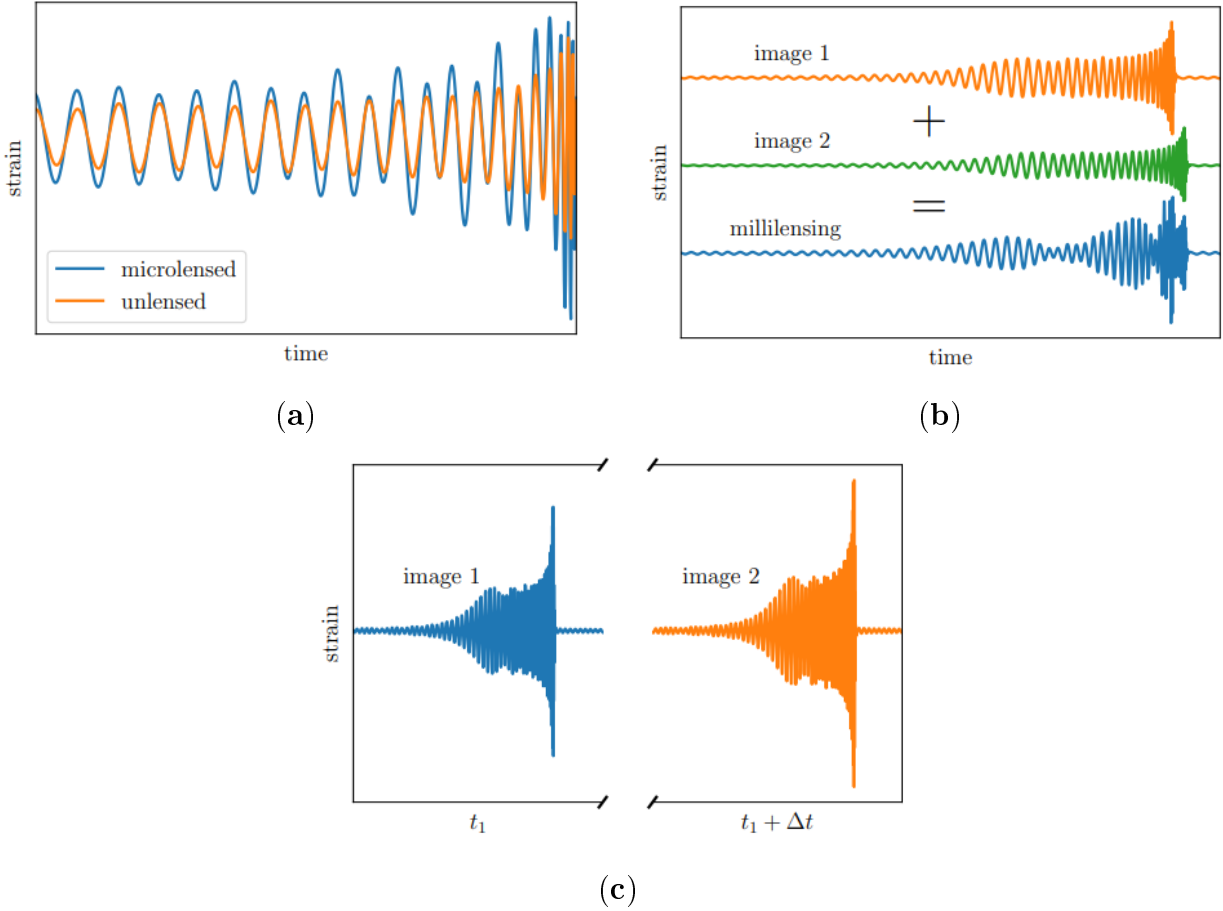


Figure 10: Examples of lensing observed for light. (a) micro lensing, (b) milli lensing, (c) strong lensing, from [30].

We have shown for light that multiple images could be observed coming from the same object at the same time. The same can happen for gravitational waves but signals will not always be observed at the same time, this can be explained in the context of a perturbed Minkowski background metric,

$$ds^2 = -(1 + 2\frac{U}{c^2})(cdt)^2 + (1 - 2\frac{U}{c^2})dl^2 \quad (1.49)$$

where $U \equiv U(\vec{r}) = -\frac{GM}{r}$ is the gravitational potential which we assume small ($U \ll c^2$), this is the weak field limit. This model gives us for null ray, the travel time being,

$$t = \int \sqrt{\frac{1 - 2\frac{U}{c^2}}{1 + 2\frac{U}{c^2}}} dl \simeq \int 1 - 2\frac{U}{c^2} dl = l - 2 \int \frac{U}{c^2} dl \quad (1.50)$$

where l is the Euclidean length of the path. We see that depending on the length of the path, the travel time changes. Moreover, if the gravitational wave goes through a more intense gravitational field, the travel time will become longer (remember that U is negative).

Other effects might also perturb the signals such as magnification, distortion and the phase shift known as the Morse phase, we will not discuss those technical points and focus on the characteristics of gravitational lensing made by galaxies.

Galaxies are much denser than the average universe, from this, we know that we are in a strong lensing regime. We will focus on the case where the wavelength is in the range of about $[3 \cdot 10^5, 3 \cdot 10^7]$ m (frequency in $[10, 10^3]$ Hz) and galaxy have typical masses around $10^{11} M_{\odot}$ corresponding to a Schwarzschild radius of $3 \cdot 10^{14}$ m, which tells us that we are not in the micro lensing regime. We will also suppose that signals are separated in time and not in the milli lensing regime.

During the third run of observation, O3, strong lensing searches were made using a set of algorithm LensID [31], GOLUM [32] [33] and HANABI [34]. LensID helps to identify potential lensing signals and will be discussed in Section 3.1, we will try to improve its results later. GOLUM then helps to filter the signals better and HANABI helps to get the final analysis. Such a separation between the different analysis tools is needed because for n detected gravitational wave signals, about n^2 pairs can be made and thus the number of analysis scales very fast. We understand that LensID can be viewed as a fast analysis algorithm so that its computation time should be small while the other two serve as finer filters.

Until now, no detection of lensed gravitational wave has been made but the effort to observe one continues in the LVK collaboration partially because measurement on light indicates its feasibility. Moreover, lensing via the multiple signals measurement coming from the same sources would help to better localise position of the signals in the sky [35], see Figure 11. We also know that lensing magnifies the signal coming from a gravitational wave events, this implies that the observed masses look greater than the actual masses of the objects. We thus understand that studying gravitational wave lensing would then also help to characterise which portion of our gravitational waves are lensed and thus understand the uncertainties on the observed masses better.

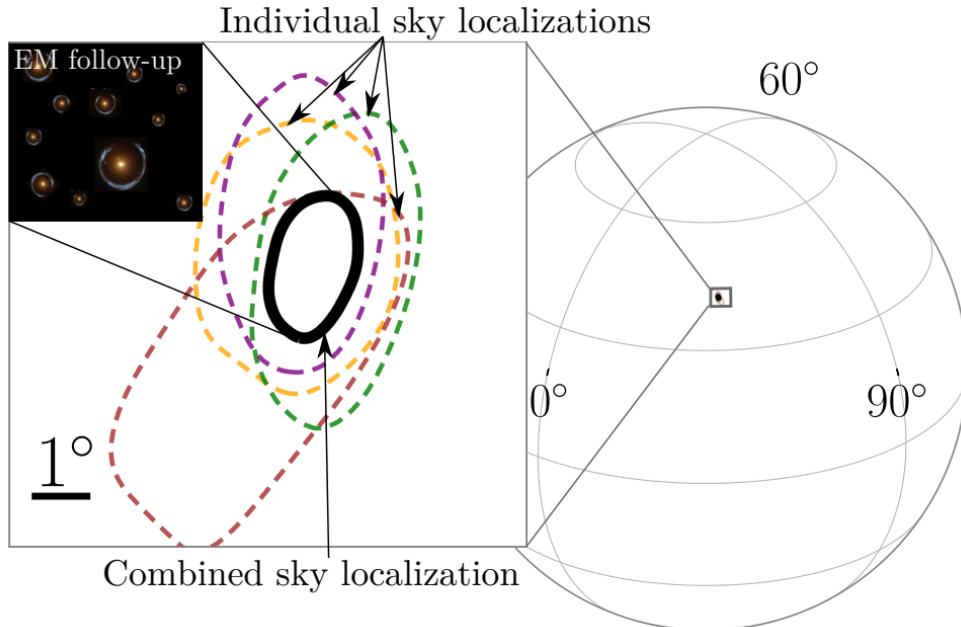


Figure 11: An illustration of a sky localisation of a quadruply lensed gravitational wave (4 signals). It shows both the individual (color) and the combined (black) sky localisations at 90% confidence. Each lensed gravitational wave essentially gives us a new set of detectors with which to localise the event in the sky, allowing for improved sky localisation, from [35].

2 Signal Processing Technique

To understand gravitational waves better, one often uses signal processing techniques to improve the representation of the data. These techniques will be presented in this section in a way such that one understands the general principles. We will sometimes refer to continuous definitions and sometimes to discrete ones but one should keep in mind that the data are discrete.

The fundamental object we are going to transform is the strain and we will represent it as,

$$s(t) \leftrightarrow [s(t_0), s(t_1), \dots, s(t_{N-1})] \quad (2.1)$$

One can observe an example strain of the gravitational wave event GW191105 in Figure 12. As its name indicates, it occurred the 05 November 2019. The time resolution, Δt , is given by the inverse of the sampling rate which is here 4096 Hz. Note that data are measured with a sampling rate of 16 384 Hz and usually down-sampled to 4096 Hz to reduce computation time.

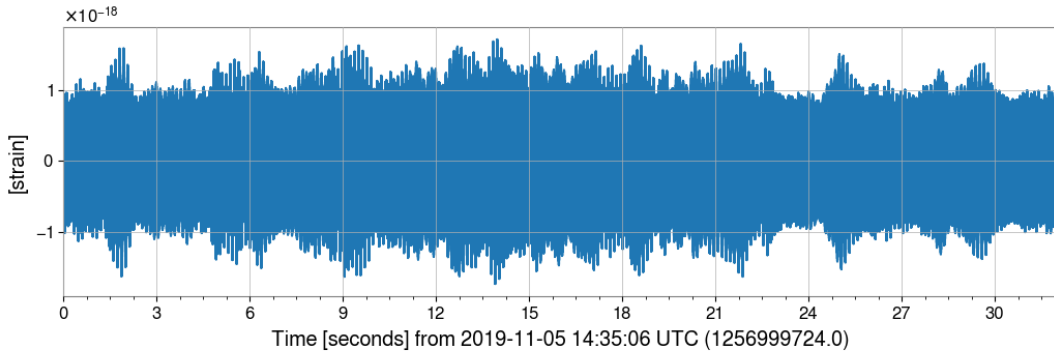


Figure 12: Example strain of GW191105 during 32 s at 4096 Hz.

2.1 Fourier Transform

First, let us introduce the key ingredient to discuss the time-frequency relation, the Fourier Transform. The Fourier Transform is a powerful mathematical tool used in signal processing, physics, and many other fields to analyse and represent functions in terms of their frequency components. It was developed by Joseph Fourier in the early 19th century [11].

Let us consider our function $s(t)$, which represents a signal in the time domain. The Fourier Transform of $s(t)$, denoted as $\tilde{s}(\nu)$, provides a representation of $s(t)$ in the frequency domain. The formula for the Fourier Transform is given by,

$$\tilde{s}(\nu) = \int_{-\infty}^{\infty} s(t) e^{i2\pi\nu t} dt \quad (2.2)$$

Here, ν represents the angular frequency (in radians per unit time), and i is the imaginary unit. The integral extends over the entire real line.

In terms of discrete definitions,

$$\tilde{s}(\nu) = \sum_{n=0}^{N-1} s(t_n) e^{i2\pi\nu n/N} \quad (2.3)$$

One important principle associated with this transform is the Gabor–Heisenberg uncertainty principle relating time and frequency uncertainties [36],

$$\sigma_t \cdot \sigma_\nu \geq \frac{1}{4\pi} \quad (2.4)$$

where for a given function satisfying $\psi \in L^2(\mathbb{R})$ and $\|\psi\|_{L^2(\mathbb{R})}^2 = 1$,

$$u = \int_{-\infty}^{\infty} t |\psi(t)|^2 dt \quad \sigma_t^2 = \int_{-\infty}^{\infty} (t - u)^2 |\psi(t)|^2 dt \quad (2.5)$$

$$\xi = \int_{-\infty}^{\infty} \nu |\tilde{\psi}(\nu)|^2 d\nu \quad \sigma_\nu^2 = \int_{-\infty}^{\infty} |\nu - \xi|^2 |\tilde{\psi}(\nu)|^2 d\nu \quad (2.6)$$

Here, σ_t represents the uncertainty in time, and σ_ν represents the uncertainty in frequency. This uncertainty principle implies that if we take a signal localised in time (i.e. one that has a short time resolution), then its frequency content becomes more spread out, and vice versa. It has implications in well-known subjects in physics. For example, in quantum mechanics, where the Heisenberg Uncertainty Principle states that there is a limit to the precision with which two conjugate observables, such as position and momentum, can be simultaneously known.

The mathematical framework of the Fourier Transform allows us to analyse and understand the frequency content of signals. By transforming a signal from the time domain to the frequency domain, we can gain insights into its spectral composition. This is particularly useful in applications such as audio processing, communication systems, and image analysis.

The Fourier Transform has many important properties, such as linearity, time-shifting, frequency-shifting, and convolution, which make it a versatile tool for analysing and manipulating signals in various domains where finding cyclic signals is needed.

2.2 Whitening

As discussed in Section 1.2.5, gravitational wave signals are hidden in various kinds of noise. An ingredient to distinguish those signals better is the whitening procedure. Given our set of data $s(t)$, one can compute its Power Spectral Density (PSD) via,

$$P(\nu) = \lim_{T \rightarrow \infty} \frac{1}{T} \left| \sum_{n=0}^{N-1} s(t_n) e^{i2\pi\nu t/N} \right|^2 \quad (2.7)$$

where T is the duration in which the PSD is evaluated. Note for completeness that one can define an Amplitude Spectral Density (ASD) via $\sqrt{\text{PSD}}$ which gives an order of magnitude of the amplitude of the signal at different frequencies, see discussion about strain in Section 1.2.5. Afterwards, one can whiten the data according to [37],

$$s_{\text{white}}(t) = \sum_{k=0}^{N-1} \frac{\tilde{s}(\nu_k)}{\sqrt{P(f_k)}} e^{-i2\pi tk/N} \quad (2.8)$$

This process helps to flatten the observed noise, an example of the change of the PSD can be observed in Figure 13. One can then more easily distinguish gravitational wave signals from noises.

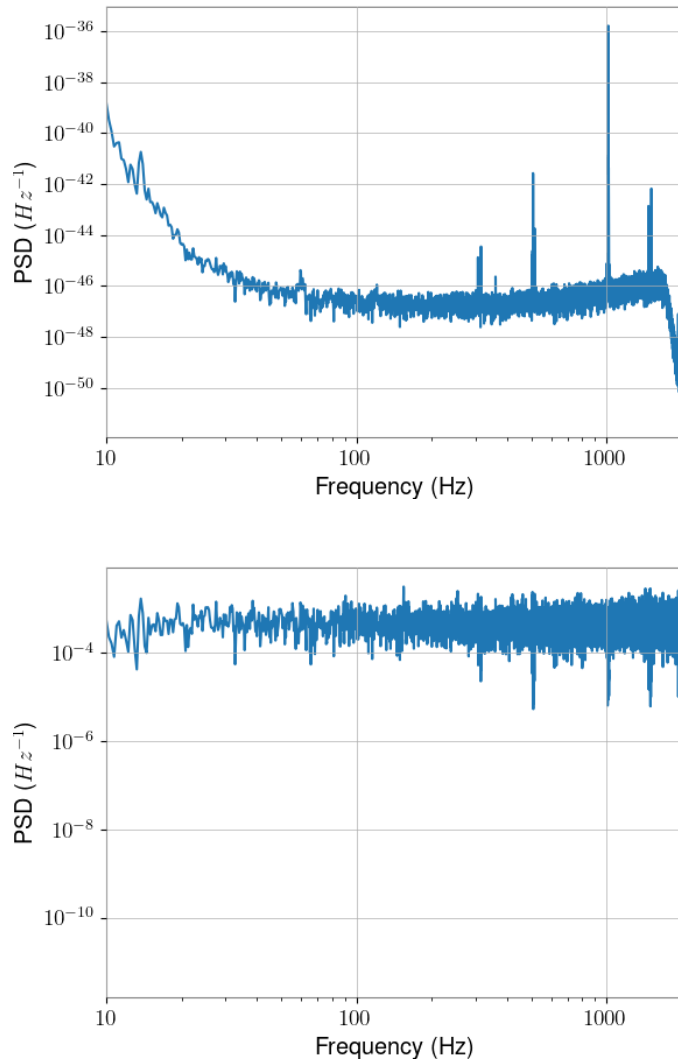


Figure 13: Comparison between PSD. The top image represent the PSD of GW191105 with $T=4s$. Bottom one, the whitened PSD of GW191105 with $T=4s$. The violin mode at 500 Hz takes less importance after whitening.

2.3 Wavelet Transform

Until now, signals were either presented in time or frequency domains, which is not suitable for gravitational wave analysis as their frequency evolves with time. To have both information about frequency and time, many other transforms have been developed. One central transform is the wavelet transform. General information about wavelet can be obtained in [38] and [39], we will focus on the needed pieces of information. We defined a wavelet transform as,

$$T_s(\nu, \tau) = \int_{-\infty}^{\infty} s(t)\psi_{\nu,\tau}^*(t) dt \quad (2.9)$$

There exists an inversion formula,

$$x(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_s(\nu, \tau)\psi_{\nu,\tau}(t) d\nu d\tau \quad (2.10)$$

in the case where one of the following *admissibility condition* holds,

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\tilde{\psi}(y)|^2}{|y|} dy < \infty \quad (2.11)$$

$$\int_{-\infty}^{\infty} \psi(x) dx = 0 \quad , \text{ i.e. } \tilde{\psi}(0) = 0 \quad \text{and} \quad \tilde{\psi}(x) \text{ is continuously differentiable} \quad (2.12)$$

Note that the Gabor-Heisenberg uncertainty in equation (2.4) holds for the ψ functions. It is known that the lower bound of the inequality can be obtained with the Gabor-Morlet wavelet [39],

$$\psi^*(t, \tau, \nu, \sigma) = \left(\frac{2}{\sigma^2 \pi} \right)^{\frac{1}{4}} e^{-\frac{(t-\tau)^2}{\sigma^2}} e^{-2\pi i \nu (t-\tau)}, \quad (2.13)$$

so that

$$\sigma_t = \frac{\sigma}{2} \quad \sigma_\nu = \frac{1}{2\pi\sigma} \quad (2.14)$$

In general, one cannot invert transform with the Gabor-Morlet wavelet, but thanks to the admissibility condition, we know that the transform can be inverted when $\nu\sigma \gg \frac{1}{\pi}$ because,

$$\int_{-\infty}^{\infty} \psi(x) dx = (2\pi\sigma^2)^{\frac{1}{4}} e^{-\pi^2 \nu^2 \sigma^2} \simeq 0 \quad (2.15)$$

2.4 Q Transform

Researchers have developed another transform which is known as the Q transform [40], [41], [42],

$$X(\tau, \nu, Q) = \int_{-\infty}^{\infty} s(t)w(t, \tau, \nu, Q)e^{-2\pi i \nu t} dt \quad (2.16)$$

where $Q = \frac{\nu}{\sigma_\nu} > 0$ and the function w is a time-domain window centred on time τ with a duration that is inversely proportional to the frequency under consideration. Note for example the following window which leads to a similar transform as the one associated with the Gabor-Morlet wavelet,

$$w(t, \tau, \nu, Q) = \left(\frac{8\pi\nu^2}{Q^2} \right)^{\frac{1}{4}} e^{-\left(\frac{2\pi\nu(t-\tau)}{Q} \right)^2} \quad (2.17)$$

The main difference with the Gabor-Morlet wavelet resides in the shift in time in the complex exponential but also in the definition of $\sigma = \frac{Q}{2\pi\nu}$. Nevertheless, one can still recover the signal through an inversion formula [1],[43] because of the connection between the Q transform and the Mellin transform.

To compute the Q transform one can use the package PyCBC [44] or GWpy [45], those packages compute the Q transform based on our formula. To know which Q parameter one should use, both algorithms have their technique to optimise the Q-value, this will be explained more in-depth in Section 5.1.1. You can see a Q transform based on GWpy in Figure 15 for the Livingston detector (L1) of the first GW event GW150914.

2.5 Qp Transform

Recently, a new paper [1] has suggested a wavelet Q transform where the only difference with the Morlet wavelet is the introduction of a parameter Q such that the wavelet from equation (2.13) is parametrised with $\sigma = \frac{Q}{2\pi\nu}$.

$$\psi^*(t, \tau, \nu, Q) = \left(\frac{8\pi\nu^2}{Q^2} \right)^{\frac{1}{4}} e^{-\left(\frac{2\pi\nu(t-\tau)}{Q} \right)^2} e^{-2\pi i\nu(t-\tau)} \quad (2.18)$$

We thus observe only a difference between the Q transform from equation (2.17) and the wavelet transform of equation (2.18) in the time-shift of the complex exponential. They both saturate the Gabor-Heisenberg uncertainty principle with,

$$\sigma_t = \frac{Q}{4\pi\nu} \quad \sigma_\nu = \frac{\nu}{Q} \quad (2.19)$$

Here we see that the parameter Q helps to define the time and frequency resolution. The higher the Q, the finer the frequency resolution and the wider the time resolution as a consequence of the equation (2.4).

In general, to study a given frequency, ν is a constant over time when used in the Q transform. We will here suppose that it might vary over time thus denoting it with $\nu(t)$ and denote $t - \tau \equiv \Delta t$. The next step towards the Qp transform is to make a Taylor expansion of the frequency around the time τ where the Q transform is evaluated,

$$\nu(t) = \nu(\tau) + \left. \frac{\partial \nu}{\partial t} \right|_{t=\tau} \Delta t \quad (2.20)$$

One then chooses,

$$\left. \frac{\partial \nu}{\partial t} \right|_{t=\tau} = \frac{(1+p)\nu(\tau) - (1-p)\nu(\tau)}{(\tau + \sigma_t) - (\tau - \sigma_t)} = 4\pi\nu^2(\tau) \frac{p}{Q} = \frac{1}{2\pi} \left(\frac{2\pi\nu(\tau)}{Q} \right)^2 2pQ \quad (2.21)$$

With this choice, the p parameter determines the fractional frequency change with respect to the central frequency over the $[\tau - \sigma_t, \tau + \sigma_t]$ time interval. The resulting non-linear

frequency change helps these wavelets in adapting to the CBC chirping gravitational wave signals.

We see here that the frequencies will evolve with the frequency of the local wavelet at time τ depending on the value of pQ . To find the new wavelet, we substitute only the oscillatory term and not the amplitude so that,

$$e^{-2\pi i\nu\Delta t} \rightarrow e^{-2\pi i\nu(\tau)\Delta t} e^{-\left(\frac{2\pi\nu(\tau)\Delta t}{Q}\right)^2 2ipQ} \quad (2.22)$$

With those changes, we find the following wavelet, where now ν is again a fixed value in the transform and not depending on the time.

$$\psi^*(t, \tau, \nu, Q, p) = \left(\frac{8\pi\nu^2}{Q^2}\right)^{\frac{1}{4}} e^{-\left(2\pi\nu\Delta t\frac{\sqrt{1+2ipQ}}{Q}\right)^2} e^{-2\pi i\nu\Delta t} \quad (2.23)$$

where here the time and frequency uncertainty does not saturate the Gabor-Heisenberg principle,

$$\sigma_t = \frac{Q}{4\pi\nu} \quad \sigma_\nu = \frac{\nu}{Q} \sqrt{1 + (2pQ)^2} \quad (2.24)$$

For example, when p is positive, the Gaussian envelope will have increasing frequency over time. This can be observed assuming Gaussian noise via the two-point correlation function ($\langle \cdot \rangle$ denote an average over time),

$$\langle T^*(\tau_0, \nu_0, Q, p) T(\tau, \nu, Q, p) \rangle \quad (2.25)$$

An explicit derivation can be found in [1], we will here only illustrate what it looks like for two cases when $p = 0$ and $p > 0$ in Figure 14. This Figure helps to understand why the Qp transform is better suited for chirp signals as it supposes rising frequency for $p > 0$ and thus takes into account the shape of the signal better. However, one should also pay attention to the fact that for fixed p this technique is more biased than the Q transform. For example, if one takes $p = 0.1$, all signals with decreasing frequency will be harder to recognise. To avoid this, one often lets p free and tries to find the best parameter according to a given criterion.

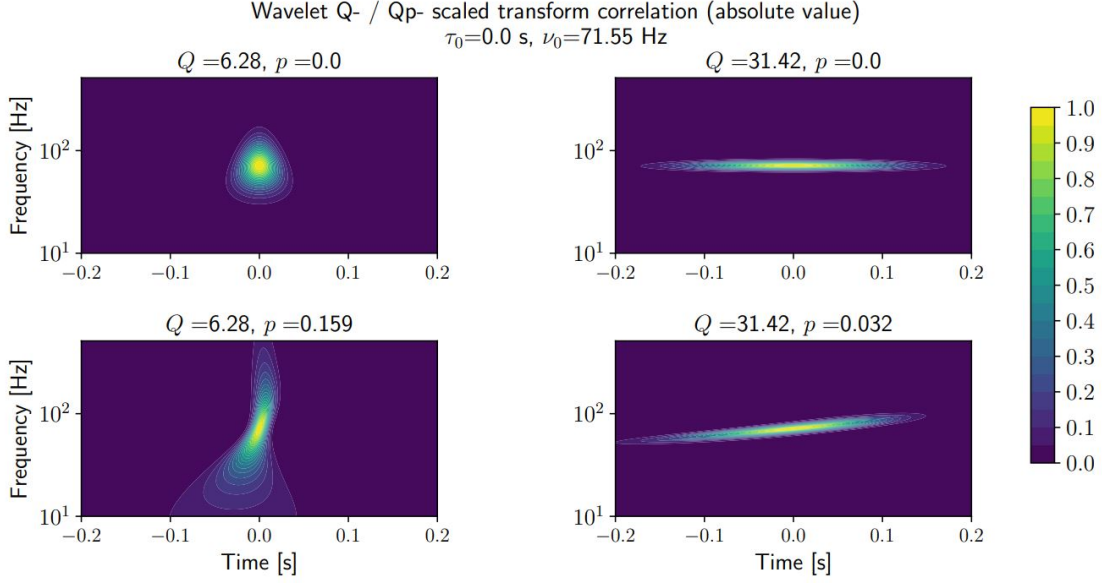


Figure 14: The module of the expected value for the correlation between $T_{\text{nd}}^*(\tau_0, \nu_0, Q, p)$ and $T_{\text{nd}}(\tau, \nu, Q, p)$, i.e. $|\langle T_{\text{nd}}^*(\tau_0, \nu_0, Q, p) T_{\text{nd}}(\tau, \nu, Q, p) \rangle|$, evaluated as a function of τ and ν , assuming Gaussian noise, from [1].

To implement this wavelet Qp-transform, the authors [1] followed the guidelines provided by the GWpy [45] module `qtransform.py`. Their code is available with the following link, <https://zenodo.org/records/10649073>.

You can see a Qp transform based on their algorithm in Figure 16 for the Livingston detector (L1) for the first GW event GW150914. Comparing Figures 15 and 16, one notices that the Qp transform helps to have a slightly better resolution and thus localise a chirp-like signal in this case better.

As mentioned earlier, the Qp transform is a first order Taylor expansion of the frequency around a given time. In Figure 14, we understand that it leads in a linear correlation between two point in the time-frequency plane. As a first approximation, this seems to be an interesting idea but maybe to limited. We think that developing a second order Taylor expansion would also help to take into account the typical pattern of gravitational wave frequency which looks like a parabola in the time frequency plane, in the same manner as [46]. Higher order Taylor expansion could also be interesting if one thinks that more complex pattern could be observed in spectrograms leading to a kind of polynomial transform.

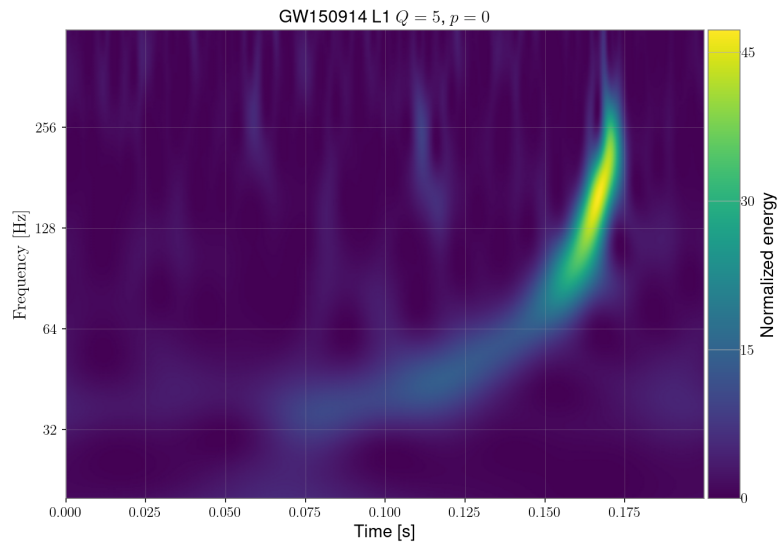


Figure 15: Q transform based on GWpy [1] for GW150914 in the L1 detector. Normalisation of the pixel is made with the median of their values.

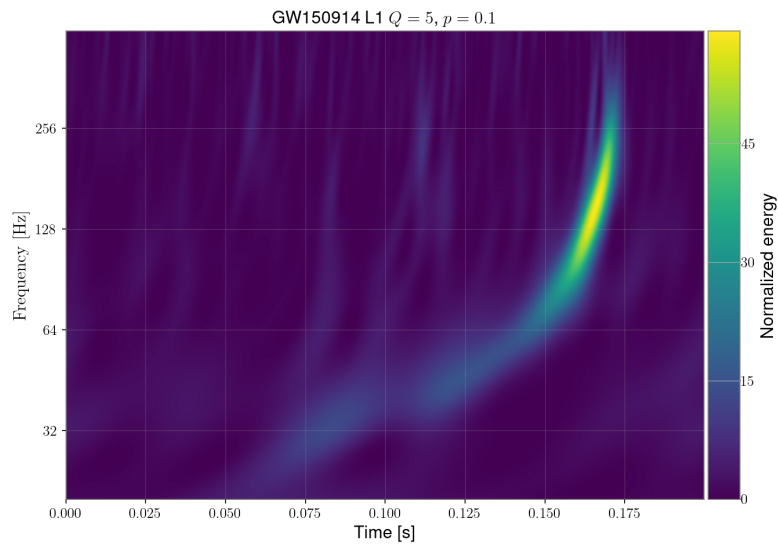


Figure 16: Qp transform based on GWpy [1] for GW150914 in the L1 detector. Normalisation of the pixel is made with the median of their values.

3 Machine Learning

Machine learning is a branch of artificial intelligence that focuses on the development of algorithms and statistical models that enable computers to learn and improve their performance on a specific task without being explicitly programmed. The goal of machine learning is to allow computers to discover patterns and relationships in data automatically, and use these insights to make predictions or decisions. We usually find three main types of machine learning.

The supervised learning where the algorithm learns from a labelled dataset, where each example in the dataset is accompanied by a corresponding label or outcome. The algorithm learns to map input data to the correct output based on the input-output pairs in the training data. Common supervised learning tasks include classification and regression.

The unsupervised learning where the algorithm learns from an unlabeled dataset, where the examples do not have corresponding labels. The goal of unsupervised learning is to discover the underlying structure or patterns in the data. Common unsupervised learning tasks include clustering and dimensionality reduction.

The reinforcement learning where the algorithm learns by interacting with an environment and receiving feedback in the form of rewards or penalties. The algorithm learns to take actions that maximise cumulative reward over time. Reinforcement learning is often used in areas such as game playing and robotics.

The typical machine learning methodology consists of several steps,

1. Data Collection: Gathering the data that will be used to train and evaluate the model.
2. Data Preprocessing: Cleaning, transforming, and normalising the data to make it suitable for training.
3. Model Selection: Choosing the appropriate machine learning algorithm and model architecture for the task at hand.
4. Training: Optimising the model parameters using the training data to minimise a predefined loss function.
5. Evaluation: Assessing the performance of the trained model on a separate validation or test dataset.
6. Deployment: Integrating the trained model into a real-world application or system.

Machine learning is a powerful tool that has been applied to a wide range of domains, including image recognition, natural language processing, and finance.

To understand the working principle of LensID, the machine learning model we are going to use, we will present different machine learning concepts.

As a starting point, one can see in Figure 17 an example of a fully connected neural network described mathematically by equation 3.1. In this case, we can thus view the machine learning algorithm as non linear function applied to linear combination of input neurons.

$$a_i^{(l+1)}(o) = f^{(l+1)} \left(w_{i,0}^{(l)} + \sum_{j=1}^{s_l} w_{i,j}^{(l)} a_j^{(l)}(o) \right) \quad l \in [0, L - 1] \quad (3.1)$$

More precisely, here, the neural networks is defined for an input object o with a list of neurons $a_i^{(l)}(o)$, where i define the index of a neurons in a layer l which belong to $[0, L]$ if the network has L layers. The inputs $a_i^{(0)}(o)$ passed through multiple layers of non linear function f^l where l again define the layer of the function, the importance of an input or neuron in the prediction of the algorithm is measured via the weight $w_{i,j}^l$, where i, j define that the weight connects a value of a neuron i in a layer $l + 1$ to a neuron j which is in layer l . This kind of neural network where every neurons are connected are thus referred as *fully connected*. In the end, one obtains outputs $a_i^L(o)$ for a neural network of size L . Note that each layer has s_l neurons so that i and j are in $[1, s_l]$ and that we also add biases $w_{i,0}^{(l)}$ for each layer that do not depend on an input value nor input neuron.

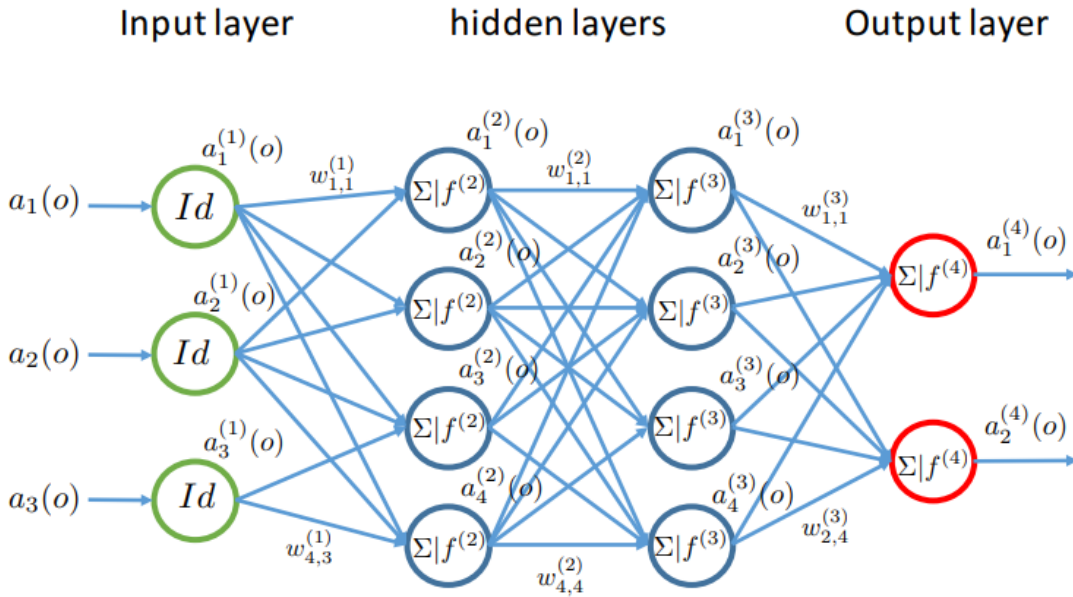


Figure 17: A neural network with 3 inputs($s_0 = 3$), 2 hidden layer with each 4 neurons ($s_1 = s_2 = 4$) and 2 outputs($s_{L=3} = 2$), from [47].

For example, non linear functions often used are the sigmoid $f(x) = \frac{1}{1+\exp(-x)}$ and ReLu $f(x) = \max(0, x)$. The advantage of those functions is their differentiability which enables the use of gradient descent technique to obtain one optimal value for the $w_{i,j}^l$. One other important aspect of machine learning is the loss function used for the gradient descent. The loss \mathcal{L} usually compares the output of the network with true known output $y(o)$ (e.g. measurement, labelling made by hand). It takes the form $\mathcal{L}(a^L(o), y(o))$, and the training of the neural network will then consist of minimising $\frac{1}{N} \sum_o \mathcal{L}(a^L(o), y(o))$ for N known outputs. Example of loss functions are the square error $(a^L(o) - y(o))^2$ and absolute

error $|a^L(o) - y(o)|$.

A better idea for classification problems is to normalise the output with a softmax function, for k classes,

$$p_i(o) = \frac{\exp(a_i^L(o))}{\sum_k \exp(a_k^L(o))} \quad (3.2)$$

Now the set of p_i define a probability and one can use the cross entropy loss, $-\sum_{i=1}^k y_i(o) \log p_i(o)$. Especially, in the case of a binary classification problem, one uses the binary cross entropy, $-[y \log(p) + (1 - y) \log(1 - p)]$ where we have used the fact that $y \equiv y_1 = 1 - y_2$ and $p \equiv p_1 = 1 - p_2$.

Afterwards, when one has a machine learning model and a loss function \mathcal{L} , one needs to optimise the weight w . Generally, to accomplish this, one represents the evolution of the loss versus the number of epochs. An epoch is a measure of how many times an algorithm has seen the whole training dataset. What we mean by training dataset is a dataset on which the model learns information and tries to reduce its loss while there also exist validation dataset on which one evaluates independently the quality of the model with data not present in the model during the training.

More precisely, often, one observe the pattern in Figure 18 for the training and validation loss. For example, if the model makes no error on the training data but always makes mistakes on the validation one, we say that it overfits the training data because it is good only at predicting the data on which it has been trained. This is observed generally for high number of epochs. While in the opposite case, with a low number of epochs, where it makes bad prediction on both data sets, we say that the model underfits, we want the model to be in between both situation with low loss on both the training and validation data. The region where this happens will determine when we stop training the algorithm.

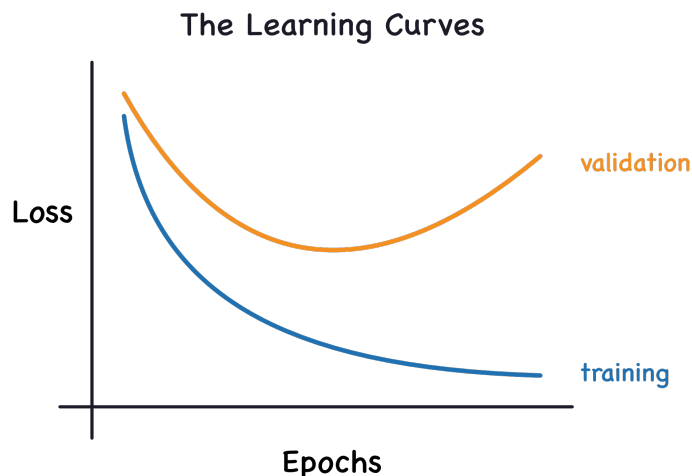


Figure 18: Training and validation curve example versus the number of epochs, from [48].

Now we will detail the working principle of a Convolutional Neural Network (CNN), as it is the basic idea behind LensID, the model we are going to use. In general, a CNN is used

to analyse 2D images but they can be generalised to 1D, 3D and so on, we will focus on 2D images. They are based on 3 concepts, fully connected layers (already presented in Figure 17), convolutional and pooling layers. The convolutional layer consists of a fixed number of connections/weights between neurons, for example for an image decomposed in RGB colour (1 image split into 3 colors), the convolutional layer might take a 5×5 grid for each RGB colour and optimise its weight to recognise features of the images like edges or shape, see Figure 19. The grid then slides onto the whole image to discern the features and get what is sometimes called an activation map. The way the sliding is made consists in a weighted sum of the input neurons along with a defined stride which tells what is the size of the pixel steps between each slide [49], see Figure 20 for example.

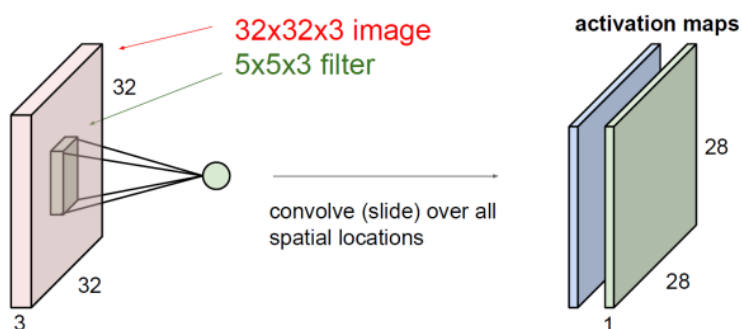


Figure 19: $5 \times 5 \times 3$ Convolutional layer applied on a 32×32 image split in a $32 \times 32 \times 3$ RGB image. Here the two resulting activation maps (output) are of size 28×28 . There are two activations maps if we use two different grids which leads to two different maps, from [47].

The other ingredient is the pooling layer, which consists again of a grid of fixed size which will again slide onto the whole image. This time the main idea is not to use weight which will connect many neurons but directly apply a function to the whole set of selected neurons. For example, an average or the maximum value of the neurons. LensID uses an average pooling layer [50].

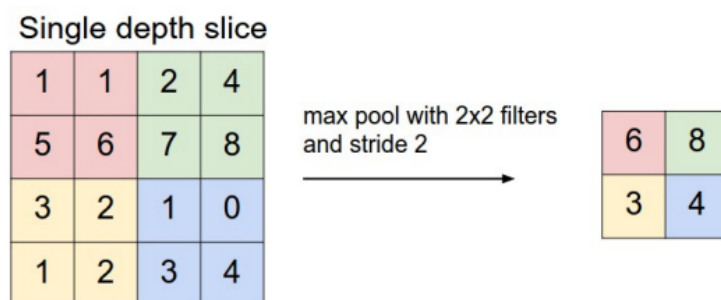


Figure 20: 2×2 Maximum pooling layer applied on a 4×4 images. The resulting output is of size 2×2 due to the stride = 2, from [47].

Finally, let us discuss how one can compare different binary classifiers such as those coming from LensID architecture. First, the nomenclature used here is summed up in Table 1, where for example, True Positive will be shorten to TP.

Actual Class	Predicted Class		Total
	Positive	Negative	
Positive	True Positive (TP)	False Negative (FN)	Positive (P)
Negative	False Positive (FP)	True Negative (TN)	Negative (N)

Table 1: A confusion matrix.

A common but limited way to compare different binary classifiers is the so-called contingency table or confusion matrix as defined in Table 1. Once it is defined, one can define several statistics to tell whether the classifier is good or not such as,

$$\text{Error rate} = \frac{FP + FN}{P + N} \qquad \text{Accuracy} = 1 - \text{Error rate} = \frac{TP + TN}{P + N} \qquad (3.3)$$

$$\text{Sensitivity (recall)} = \frac{TP}{P} \qquad \text{Specificity} = 1 - \frac{FP}{N} = \frac{TN}{N} \qquad (3.4)$$

On the one hand, the error rate counts how many objects are wrongly classified, while the accuracy counts how many objects are well classified. Those two measures are interesting to have an idea of the performance of the model but do not show how the error is distributed among the positive and negative cases. On the other hand, the sensitivity and the specificity give separately an idea of much positive and negative elements are well classified.

Along with those two quantities, one can construct what is called the Receiver Operating Characteristic (ROC) curve. This ROC curve displays the True-Positive Rate (TPR) (sensitivity) against the False-Positive Rate (FPR) (1 - specificity) as can be seen in Figure 21.

In general, one wants the highest TPR with the lowest FPR so the classifier D is the best one can achieve. All weighted random classifiers would lie on the dashed diagonal of Figure 21. What we mean by weighted random classifier is one that will predict a uniform probability between [0,1] for each data point, which is then labelled true or false based on a chosen threshold X between [0,1], predicting $100X\%$ of the time true and $100(1 - X)\%$ false. For example if $X = 1$, i.e. a random classifier that predicts always true, the classifier would be in the upper right corner while the classifier in the bottom left always predicts false. We thus understand that Classifier C tends to predict randomly about 60% as true and about 40% as false. One can say that all classifiers above (below) the dashed diagonal are better (worse) than a random classifier. Depending on the specific need to be accurate on the true or false prediction one might prefer classifier A or B. Moreover, note that the classifier E is worse than random but might become better if we swap its true and false predictions.

As stated before, classifier D is the best one can achieve. All classifiers with $TPR = 1$ are the best one can achieve for the different FPRs. In general, models do not predict true or false but rather a probability of being true or false. Accordingly, one usually then represents the ROC curve and not just a point. To accomplish that, different values of TPR versus FPR by changing the threshold at which one predicts true or false. In consequence, if one computes the area under the ROC curve, one will attempt to obtain 1. This area again loses information about the repartition of the error but is a good measure of the quality of a model. In general, LensID uses the ROC curve to compare the different results of the model.

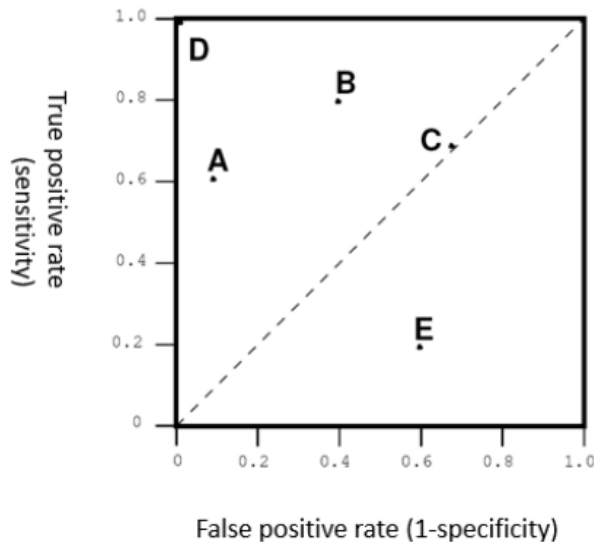


Figure 21: Example ROC plot, from [47]. A and B are good classifiers, C acts as a random classifier, D is a perfect classifier, E is worse than a random classifier and the dashed diagonal are all the possible random classifier.

3.1 LensID : Spectrogram Analysis

Here we present general information about the machine learning model we are using, namely, LensID [2] [51]. It consists of a machine learning algorithm able to solve a classification problem that determines if a pair of events is lensed or not. The main characteristic of LensID is that it shows results comparable to previous methods which can take up to several days, while taking much less time, about seconds or minutes.

LensID takes a superimposed Q transform spectrogram of two signals and sky maps as inputs, see Figure 22 for the superimposed Q transform of two unlensed and two lensed signals, each shade of colour associated with one signal. One sees easily with the white color (green+pink) that the two signals overlap for lensed signal.

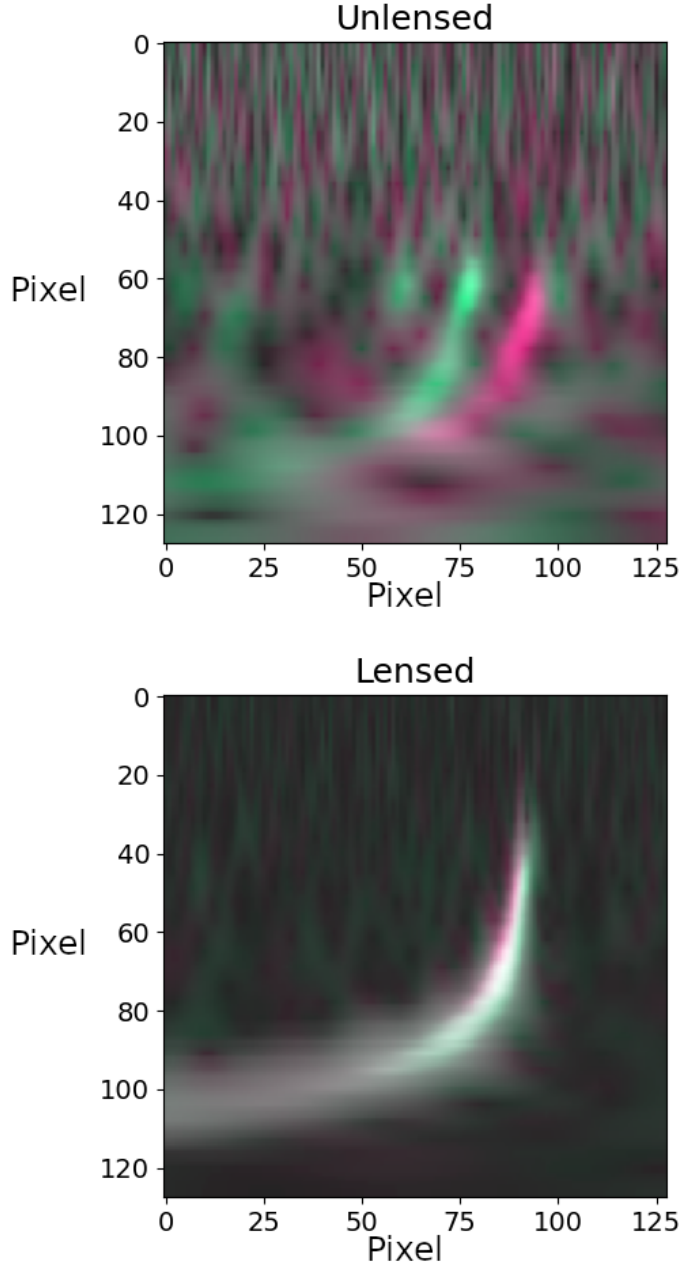


Figure 22: Superimposed Q transform taken as input by LensID, an unlensed pair (top) and an lensed one (bottom). Green represent one signal and pink the other one, overlapping signals thus imply white pixel. Remark also the difference in the background, here the background of the lensed signals is relatively smaller than the signals and relatively important for the unlensed pair. Axis represent image size with 128x128 pixels, all images have a y axis representing frequencies in $[15,1000]$ Hz and the x axis representing time around the merging of the black holes, i.e. $[t_{\text{merger}} - 0.2, t_{\text{merger}} + 0.1]$ s, normalisation of the pixel is made with the median of the pixels value and color scale is omitted.

The sky map gives information about the spatial location of the signal while the Q transform enables the machine learning model to compare the amplitude-frequency evolution of the signal in time, each detector (LIGO Hanford, LIGO Livingston and Virgo) has an associated machine learning network that is trained. We will not detail the sky map treatment

as this thesis focuses on the spectrograms and did not take into account sky maps. Simply note that LensID extracts features from the sky maps which are then passed through an XGBoost algorithm (combination of several decision trees) to have the probability that two events are lensed, see Figure 23 to understand the general structure better.

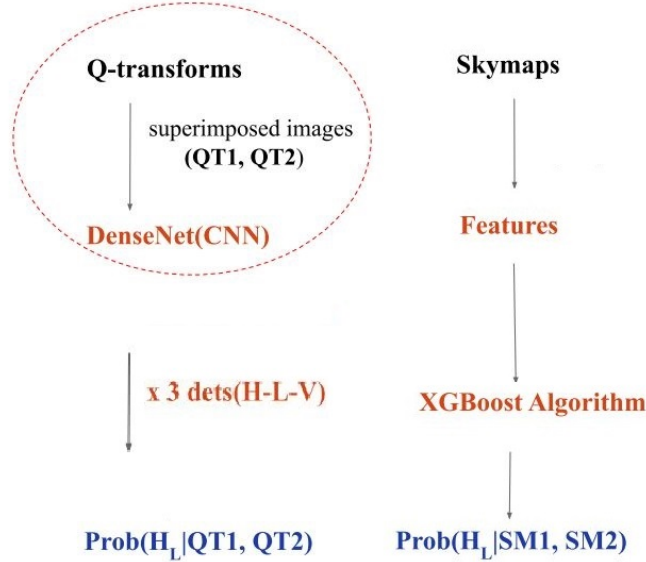


Figure 23: General structure of LensID, from [2].

At the end, the probability that an event is lensed is given by,

$$P(H_L|\{QT1, QT2\}; \{SM1, SM2\}) = P(H_L|QT1, QT2) \cdot P(H_L|SM1, SM2) \quad (3.5)$$

where each H_L , QT and SM denote respectively the hypothesis of lensed pair, Q transform and sky map.

Part of the architecture of LensID copies the one of DenseNet201 [52]. DenseNet201 is part of the supervised machine learning algorithms and more particularly part of the convolutional network, see Figure 24 for an example of its composition. CNNs are often used in image recognition which explains why this architecture was chosen. Especially, DenseNet201 helps to mitigate part of their defaults [52]. In the interest of understanding what are those defaults and how the model minimises them, let us describe its architecture.

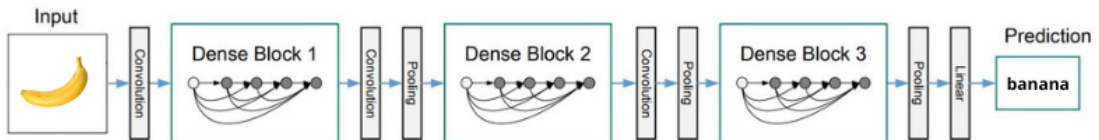


Figure 24: A DenseNet with three dense blocks. The layers between two adjacent blocks are referred to as transition layers consisting in convolution and pooling layer, from [52].

This model takes as inputs one image for each detector composed of a 3-dimensional array made by $128 \times 128 \times 3$ numbers, a square image of side 128 with RGB colours associated

with each pixel. Another characteristic of a machine learning algorithm is the loss function, commonly, for binary classification problems, one uses the binary cross entropy,

$$L_{CE} = -[y \log(p) + (1 - y) \log(1 - p)] \quad (3.6)$$

Here, y labels the data ($y = 1$ for lensed data and $y = 0$ for unlensed ones) and p is the predicted value by the algorithm.

The specificity of this CNN is that each layer in a dense block is connected to the preceding layers, leading for L layers to $\frac{L(L+1)}{2}$ connections. In general, it consists of *dense blocks* connected via *transition layers* where inputs are denoted with \mathbf{x}_i , this can be seen in Figure 25. Each connection in a dense block is associated with a function H_1 composed of a normalisation (*BN*) which simply normalise each data in a $[0,1]$ interval, a ReLu function (the activation function) and *Conv* which uses 3×3 convolution layer. The transition layer at the end of a dense block consists of normalisation, a 1×1 convolution layer and a 2×2 average pooling layer which have been described earlier in this Section. We thus understand that one chooses a DenseNet because it encourages feature reuse and strengthens the feature propagation through the network because of the connections between each layers [52].

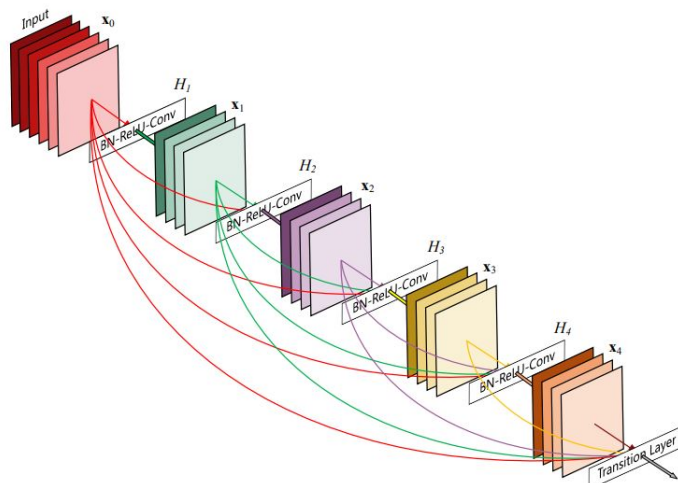


Figure 25: A 5 layer dense block, from [52].

The initial DenseNet201 model has already been trained on the Imagenet dataset and we load it via the Keras package [53]. Using a package already trained in this way is sometimes referred to as *transfer training*, which helps reduce the dataset size required for training. In addition, small modifications to the model are made to adapt it to our problem. The top layer is removed and a dense layer of 256 neurons with the ReLu activation function is added along with the final output layer of a single neuron with a sigmoid activation function with an Adam optimiser [54].

4 GW191105

4.1 Original Problem

Our search started with the gravitational event GW191105 which happened during the Second part of the third observing run (O3b). This event has two main characteristics, it has been studied as part of a pair of strongly lensed events with GW191103 [30] but its significance was not high enough to talk about the detection of a lensed pair.

In particular, this signal has high significance for both of the Bayesian inference method, GOLUM [32] and HANABI, [34] but not for LensID. The first thought was directly that there was an issue with the images associated with this event. Indeed, when one looks at the spectrogram for the GW191105 in Figure 26, one sees a gap inside the typical chirping pattern (e.g see Figure 15 with frequency increasing with time). We thus dived into the signal processing technique to understand what was wrong. Our first thought was that maybe something went wrong in the preprocessing of the LVK collaboration, we discuss this in Section 4.2.1. The second thought was that the whitening procedure was not done properly, see Section 4.2.2 but this was not the case.

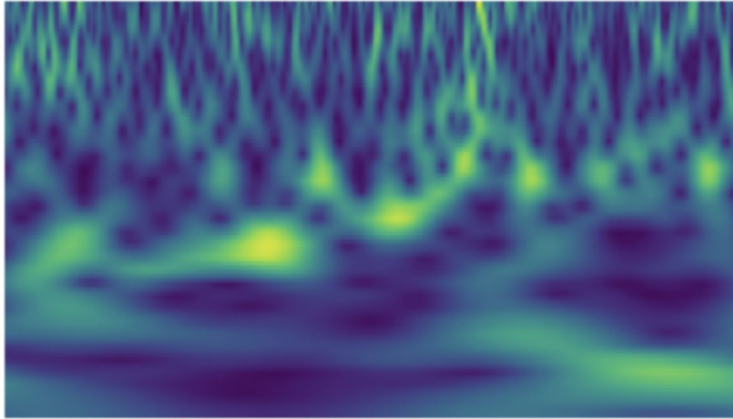


Figure 26: Image associated with the Q transform of GW191105, from [30].

Another idea was to look at the way spectrograms are made, meaning which transform one uses to create the images used as input by LensID. We observed that the typical transform used, the Q transform, was not the most suited one for chirping signals. In fact, the Qp transform seems to take into account the rising pattern of the frequency over time better, this will be discussed further in Section 4.2.3. For completeness, we present the parameter estimation of GW191103 and GW191105 [55] in Table 2. Note by comparing each side that both signals have overlapping parameters, this is roughly the way GOLUM [32] and HANABI [34] work to tell if signals are lensed.

Parameter	Value	Upper Error	Lower Error	Value	Upper Error	Lower Error
Mass 1 source (M_{\odot})	11.8	+6.2	-2.2	10.7	+3.7	-1.6
Mass 2 source (M_{\odot})	7.9	+1.7	-2.4	7.7	+1.4	-1.9
Final mass source (M_{\odot})	19.0	+3.8	-1.7	17.6	+2.1	-1.2
Final spin	0.75	+0.06	-0.05	0.67	+0.04	-0.05
Luminosity distance (Mpc)	990	+500	-470	1150	+430	-480
Network matched filter snr	8.9	+0.3	-0.5	9.7	+0.3	-0.5
Redshift	0.20	+0.09	-0.09	0.23	+0.07	-0.09
Sky area (deg ²)	2500			640		
Total mass source (M_{\odot})	20.0	+3.7	-1.8	18.5	+2.1	-1.3

Table 2: Parameter estimation for GW191103 (left side) and GW191105 (right side), taken from [55] (access date : 23 April 2023).

4.2 Towards an Origin for the Gap

4.2.1 Preprocessing Origin

As gravitational wave data are commonly given with quality qualified as C01 on websites such as Gravitational-Wave Open-Science Center (GWOSC) [55] under the format of GWF, HDF and TXT files with either 4 or 16 kHz sampling rates, we can use these files to observe the strain of a signal such as the one observed in Figure 27 for GW191105.

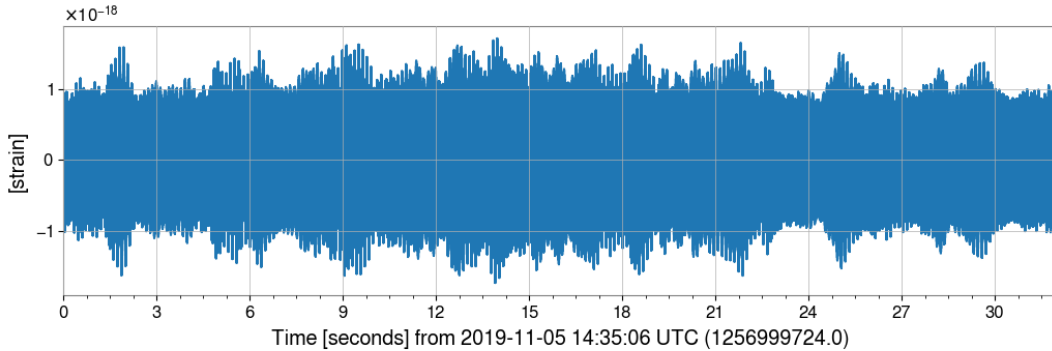


Figure 27: Example strain of GW191105 during 32 s at 4096 Hz.

Another level of data exists, the C00 one, consisting of raw data before the collaboration team performed noise subtraction [56], while data after noise subtraction are C01.

As the way to get those data is not trivial, we decided to detail it in Appendix A. Afterwards, to understand if this noise removal was at the origin of the gap, we bandpass the signals in [35 Hz, 400 Hz], the typical band where the signal extends. We then compute their Q transform, which is shown in Figure 28. We observe no major difference, except for the presence of an oscillating signal at about 60 Hz which is associated with a power line (frequency of the alternating current in the United States of America, where the Livingston detector is located is at 60 Hz). We thus conclude that the gap does not come from the preprocessing applied by the collaboration. As a side note, we could have normalised the spectrograms with the median value of the amplitude to obtain spectrograms similar in amplitude as the one of the following section, we did not because it made disappear the power line.

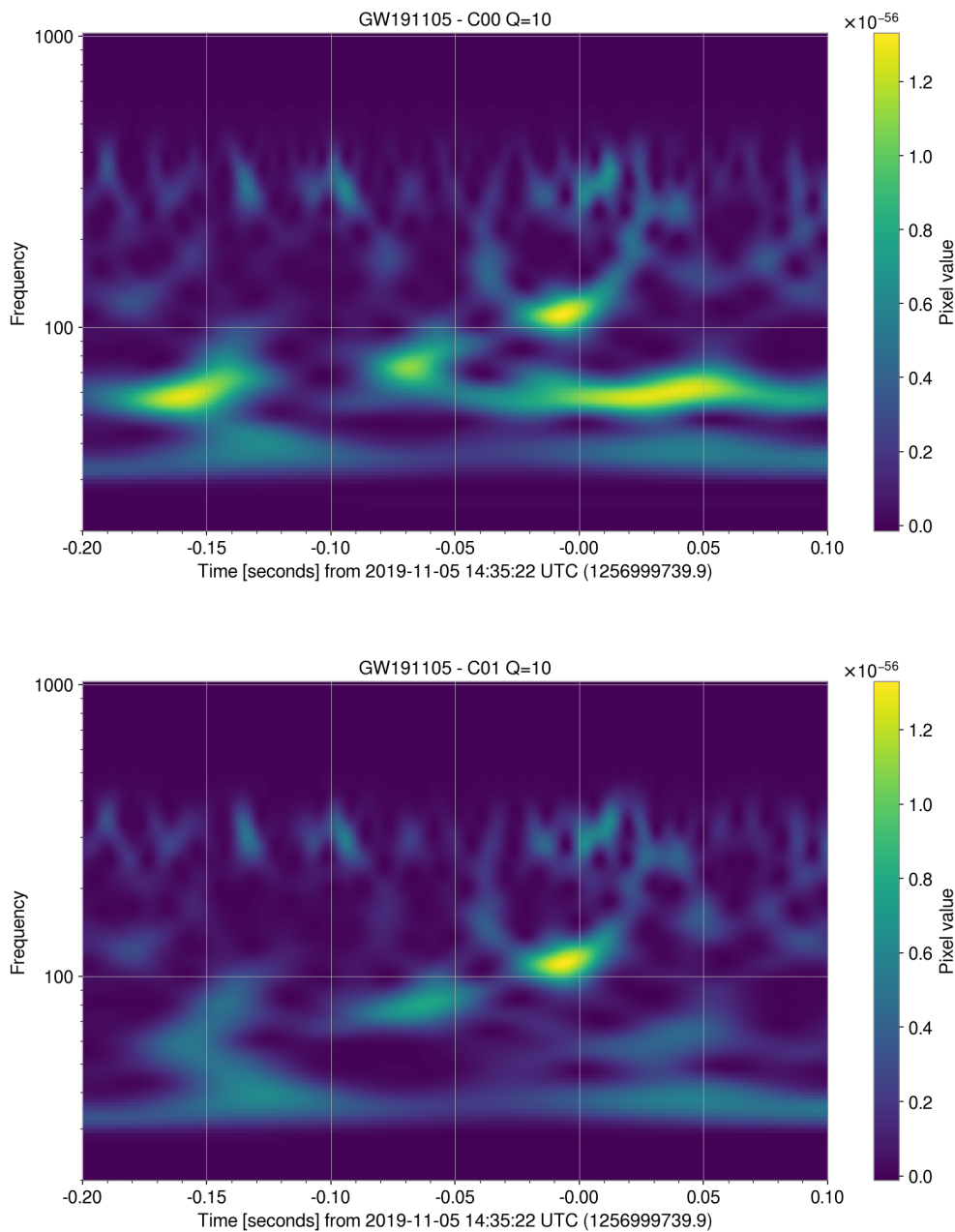


Figure 28: Q Transform of the two data type (C00 on the top and C01 on the bottom). We bandpass both signals in [35 Hz,400 Hz], where the signal resides.

4.2.2 Whitening Origin

Another hypothesis for the gap origin is the whitening of the signal. As described in Section 2.2, whitening consists in flattening the noise in which the signal lies and helps in the recognition of the typical chirp pattern. We have thus inspected various parameters in the whitening procedure via the GWpy module in python [45] and obtained at best the spectrograms of Figure 29. As this solution is not able to solve our issue with the gap, we will not detail it more.

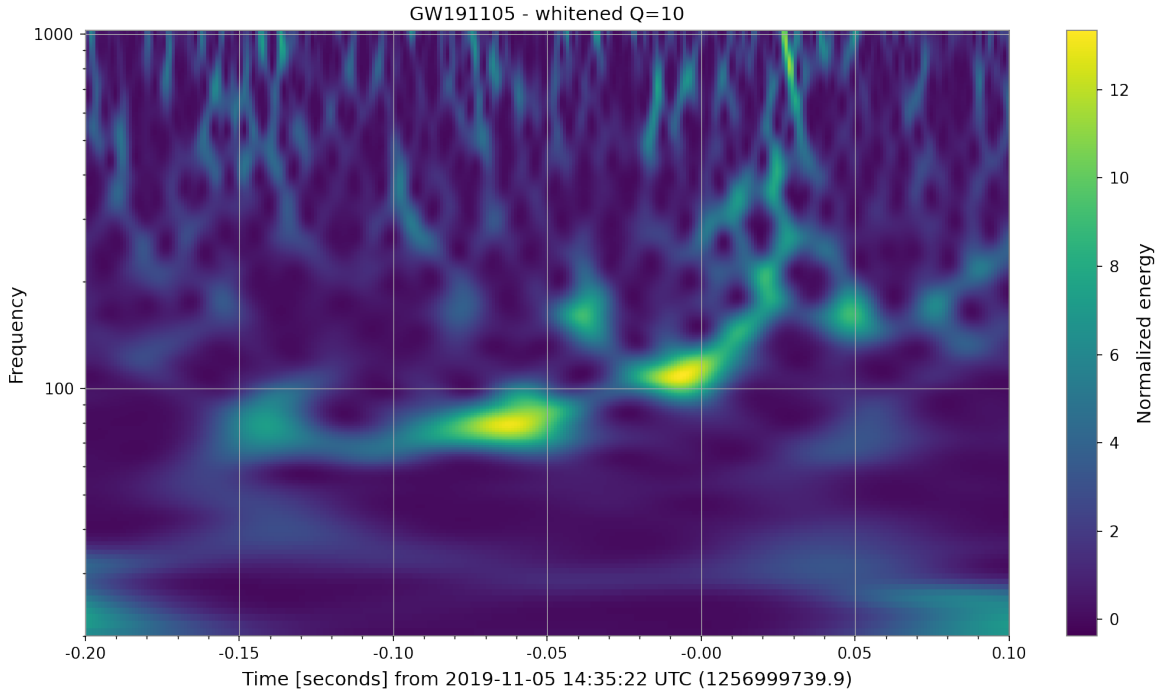


Figure 29: Normalised and whitened Q transform of GW191105 in L1. Normalisation is made with the median of the pixels value.

4.2.3 Transform Origin

Our last and best idea to treat the gap inside the GW191105 was to change the way the images are generated. Usually, one transforms the strain into spectrograms with the Fast Fourier Transform (FFT) or with the Q transform.

Our choice was to focus on another transform, the Q_p transform introduced in Section 2.5 which is an extension of the Q transform taking into account the typical rising pattern observed for the evolution of the frequency with time, the typical *chirp*. Along with the use of this Q_p transform, one needs to select values for Q and p . Those parameters are found by computing many Q_p transforms for different values of Q and p and then taking the spectrograms with the highest amplitude above a certain threshold.

We searched for Q and p parameters in the ranges $[0,100]$ and $[0,10]$ respectively during a million iterations which gave us the parameter $Q = 46.11$ and $p = 0.085$. To have the finest parameters, we made another search in the range $[40,50]$ and $[0.05,0.10]$ again during a million iterations and found $Q = 44.64$ and $p = 0.086$. Once we know the optimal Q_p parameters for our signal, we can evaluate it for the GW191105 and see the gap being filled in Figure 30 in such a way that we recover the chirping pattern. We underline that the Q_p transform supposes a chirping pattern in the data as discussed in Section 2.5 and one should be careful when one finds a chirping pattern in the spectrograms. In our case, the chirping pattern observed for a gravitational wave signal is supported by other pieces of evidence, e.g. the GWOSC [55] page for GW191105.

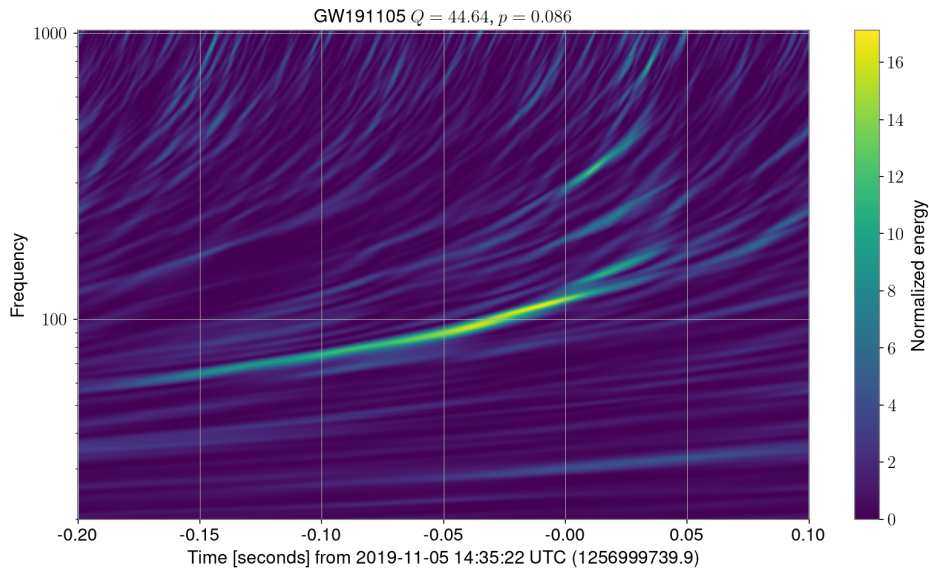


Figure 30: Normalise and whiten Qp transform of GW191105 in L1 detector. Normalisation is made with the median of the pixels value. The y axis is logarithmic and represents frequency in $[20,1000]$ Hz and the x axis represents time around the merging of the black holes, i.e. $[t_{\text{merger}} - 0.2, t_{\text{merger}} + 0.1]$ s.

5 LensID with Qp Transforms

Once we had find the Qp transform and its ability to discern gravitational wave signals, we had the will to evaluate it in a more concrete context, studying lensed gravitational waves seems promising and thus LensID becomes part of this thesis.

To train and evaluate the machine learning model, we have generated in total 6866 images with both Q transform and Qp transform. More information about the generation of those images can be found in Section 5.2. The machine learning model only observes images, so we decided not to include axes in the figures in this section. Simply note that all images sparse in a time-frequency interval of $[t_{\text{merger}} - 0.2, t_{\text{merger}} + 0.1]$ s by $[20, 1000]$ Hz where t_{merger} is the time where the black holes merge in the same way as Figure 30.

5.1 Injected Lensed Signals: Comparison between Transform

Once all the images have been generated, we inspected them, a thousand of them to be precise, to see whether the Qp transform performs better than the usual Q transform. We first notice that there exist two kinds of Q transform that have been developed, the one of PyCBC [44] and the one of GWpy [45]. Even though both transforms use the same formula, they differ in their implementation, thus also in the generated images and this is what we are going to focus on in Section 5.1.1.

Afterwards, in Section 5.1.2, we will discuss some interesting findings when comparing Q and Qp transform.

One important aspect in our analysis that we have not developed so far is the optimal matched filter Signal to Noise Ratio (SNR) [57], it is defined for a single detector via:

$$SNR_{\text{det}} = \int_{-\infty}^{\infty} \frac{\tilde{s}(f)\tilde{h}^*(f)}{P_{\text{det}}(f)} df \quad (5.1)$$

where s is the signal in the detector, h the waveform supposed to match the gravitational wave signal and $P_{\text{det}}(f)$ the noise PSD of the detector.

The SNR is thus a way to compare an effective gravitational wave waveform with the strain of data, containing both the signal and the noise. One often requires the network SNR $\sqrt{\sum_{\text{det}} (SNR_{\text{det}})^2}$ to be above 8 for 2 or 3 detectors to talk about a gravitational wave candidate. In this way, SNR in each detector should be equal to $\frac{8}{\sqrt{3}} \approx 4.62$ if all three detectors have the same sensitivity

The techniques associated with this matched filter perform in general better because they directly compare the strain with a template but they also take more time than machine learning models. The matched filtering technique takes more time because one has to search in a high hypothesis space for the parameters of the waveform associated with a gravitational event, while machine learning can simply have a look at the spectrogram and try to find the chirping pattern on which it has been trained. For example, in Figure 31, we represent a signal with SNR greater than 8 which is not distinguished in the image given to LensID. We also show the same injected signal but with a distance to the object divided by a factor of one hundred, which thus looks a hundred times more powerful in strain because $h \simeq 1/\text{distance}$, see equation (1.42).

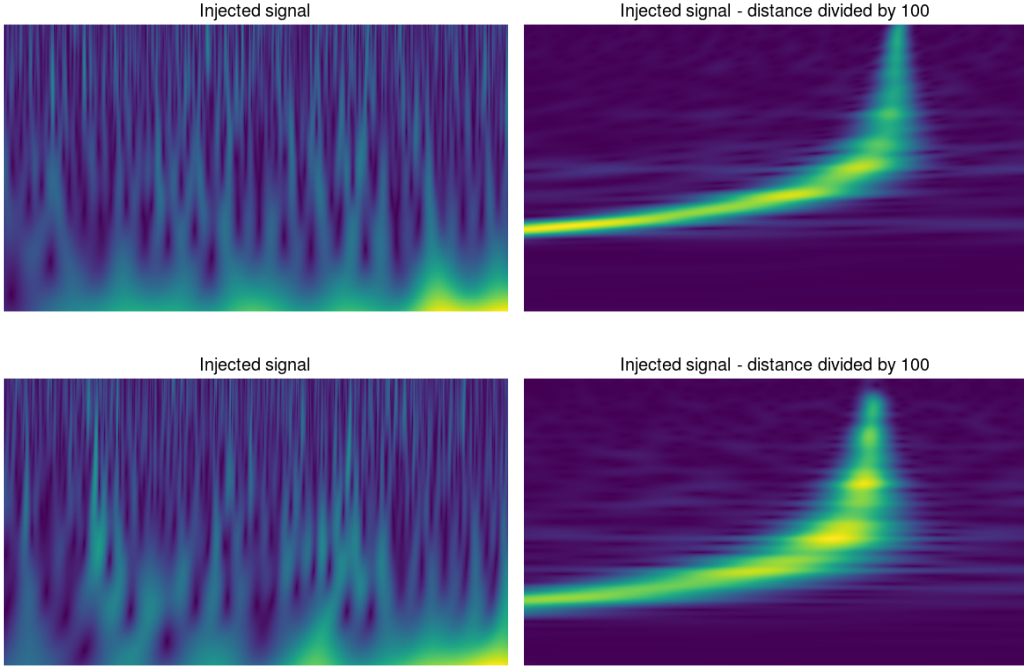


Figure 31: Spectrograms of two signals with $SNR > 8$ (detected via matched filtering). On the left, the one used for LensID with typical distances, on the right, the distance divided by a hundred to give an idea of its shape. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

5.1.1 Comparison between GWpy and PyCBC

We have discussed in Section 2.4 that two different implementations of the Q transform exist, the one of PyCBC [44] and the one of GWpy [45]. In our inspection, we noted that both perform as well, for example in Figure 32.

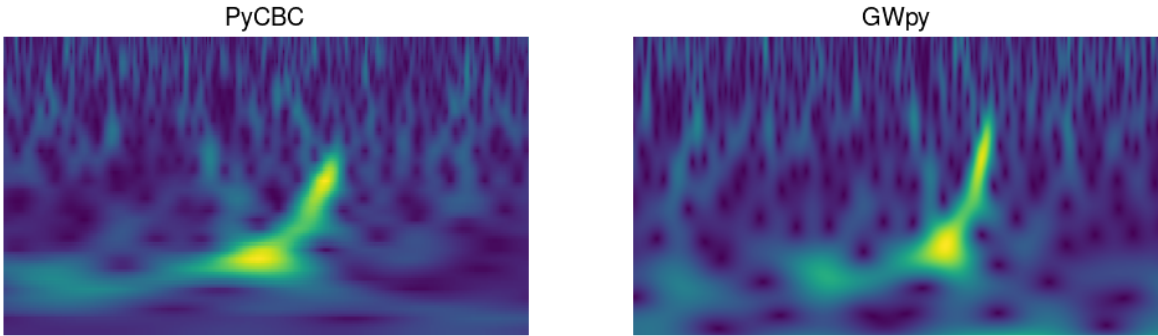


Figure 32: Comparison between PyCBC [44] and GWpy [45], here we see comparable representation. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

Nonetheless, there still exist certain signals where one transform helps to distinguish the signals better in comparison to the other one. For example in Figure 33, PyCBC represents

the chirping pattern better while in Figure 34, it is GWpy, for more comparison between PyCBC and GWpy see Appendix B.

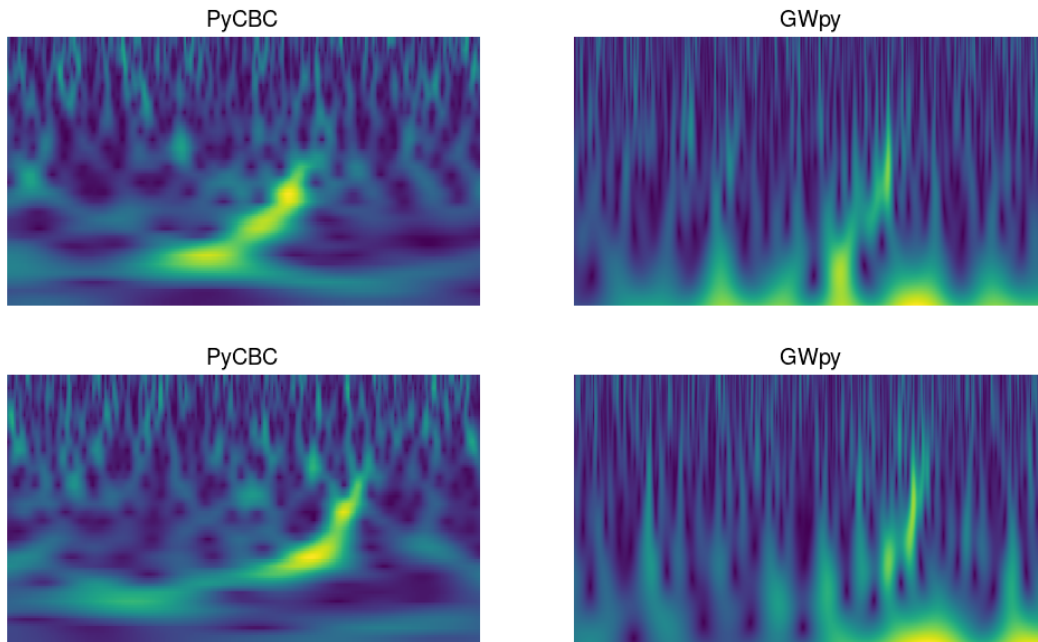


Figure 33: Comparison between PyCBC [44] and GWpy [45], here PyCBC represent the chirp better. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

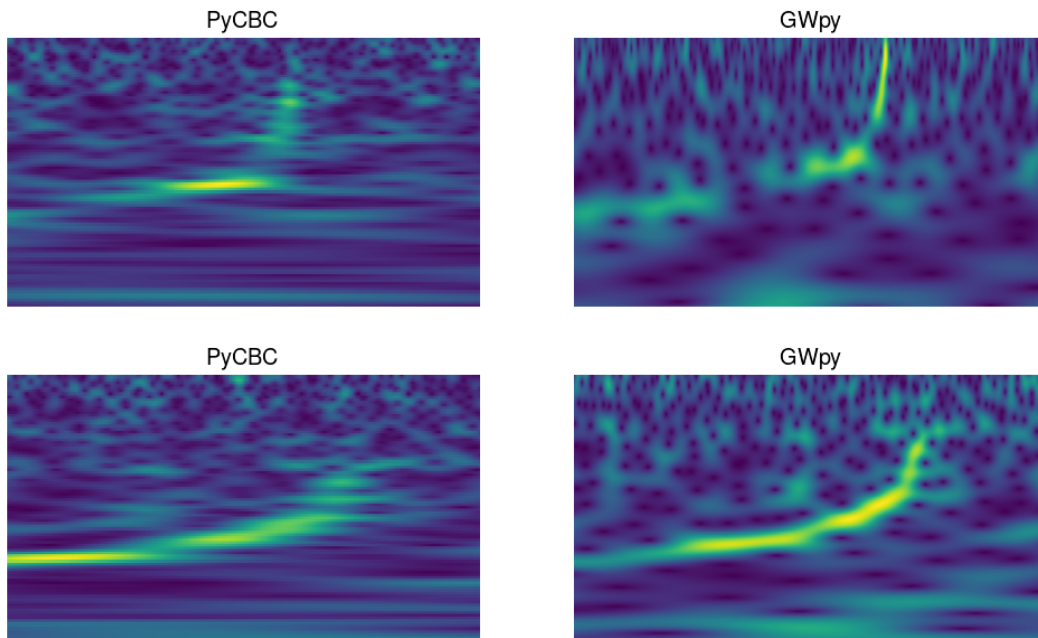


Figure 34: Comparison between PyCBC [44] and GWpy [45], here GWpy represent the chirp better. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

We tried to find some explanation for those differences in the implementation of the two transforms but did not manage to observe one. Nevertheless, it appears evident when looking at the images that the difference resides in the selection of the Q parameter. This hypothesis is supported for example by the Figure 33 where the Q parameter of GWpy is associated with a higher frequency extent. We can see that the signal is more spread out in the vertical orientation when compared to PyCBC. We thus understand that the representation is associated with a lower Q than the one of PyCBC because equation (2.19) gives $\sigma_\nu \approx \frac{1}{Q}$. The same kind of behaviour can be observed in Figure 34, where again PyCBC seems to have a lower frequency extent than GWpy. The same behaviour can be observed in many other examples in Appendix B but no obvious difference between the two implementations has been observed.

We here simply describe the working principle of the Q transform implementation. Both algorithms compute a windowed time Fourier transform which corresponds to the equation (2.16), they generally speak about the energy for the result of this transform. To accomplish the transform, they divide the time-frequency plane with a *tile* pattern, in this way, they get a list of energy for different time-frequency rectangles. Note also that both transforms do not use a Gaussian window but simply an approximation, the Bi-square windows $((1 - x^2)^2$ for $|x| < 1$, else 0) for computational reasons. Afterwards, one gets a list of spectrograms for different Q -values, the algorithm then selects the Q with the highest energy among all spectrograms.

5.1.2 Others Findings: Q and Qp Transform

Aside from the distinction between PyCBC and GWpy, we noticed some advantages in the use of Qp transform over the usual Q transform and we present them in this section. We introduce the notion of p_{\max} , it is defined via $p \in [0, p_{\max}]$. Remember we choose $p > 0$ because our signals have increasing frequency over time, $p < 0$ would imply decreasing frequency. We will always display four images, the Q transform of PyCBC, the Qp transform with $p_{\max} = 0.01$, $p_{\max} = 0.1$ which search for (Q, p) value during 1000 iteration and also Qp transform with $p_{\max} = 0.1$ searching during 1 million iteration, which we denote with long search.

First, we noticed that the value of p should not be too small, for example, when we have $p_{\max} = 0.01$ we see in Figure 35 that the signal is not clean because we are close to the usual Q transform of GWpy. Indeed, the Qp transform is based upon the algorithm of GWpy, which might sometimes be worse than the one of PyCBC as discussed in Section 5.1.1. While for $p_{\max} = 0.1$, we see a finer signal with better time-frequency resolution than the usual Q transform.

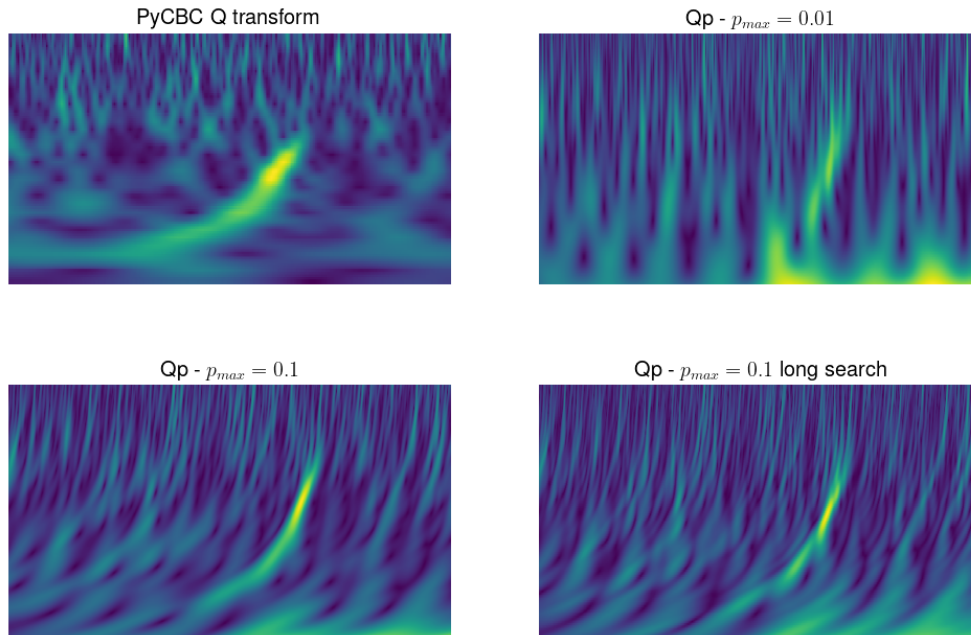


Figure 35: Comparison between transforms, this figure highlight that small p_{\max} could result in poorer signal representation. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

Second, we said that the Qp transform gives better resolution in time and frequency, this can also be observed in Figure 36. We also see that sometimes, using the long search makes the chirping pattern disappear which is a really strange behaviour.

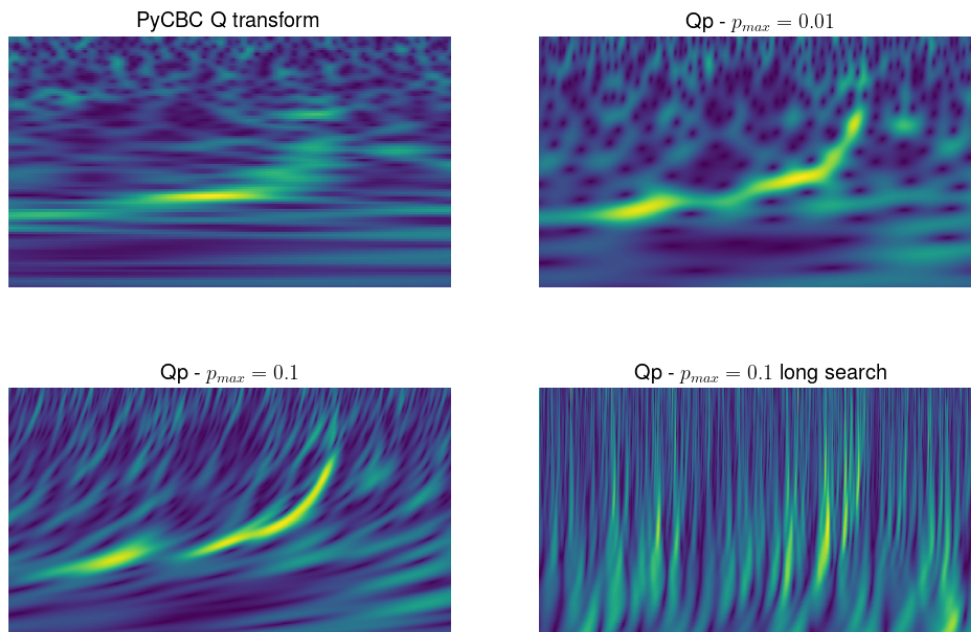


Figure 36: Comparison between transforms, this figure highlights that Qp long search could result in poorer signal representation. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

Third, the heart of Section 4 was to find a way to fill the gap inside the spectrogram of the event GW191105. During our inspection, we noticed other examples where the Qp transform helps to fill gaps within the spectrogram, we present one example in Figure 37.

Fourth, even though in Figure 35 we see that long searches might sometimes give worse results than shorter ones. We present an example in Figure 38 where it helps to recover a finer chirping pattern. In general, we think that long search should give better time-frequency resolution because it has more time to find the optimal parameters.

Afterwards, one interesting point among these transforms is the fact that the choice of the Q or p parameter is made for the whole image. An interesting idea may be to develop a transform which selects the parameters for some slice in the image, a local Qp transform following the same idea of the variable Q transform used in [58]. Our idea would be to slice either in time or frequency to obtain a finer choice of parameter. For example, in Figure 39, we see that a mixture between Qp transform with $p_{\max} = 0.01$ at the beginning and $p_{\max} = 0.1$ at the end would help to take the lower and higher frequency content both with highest resolution.

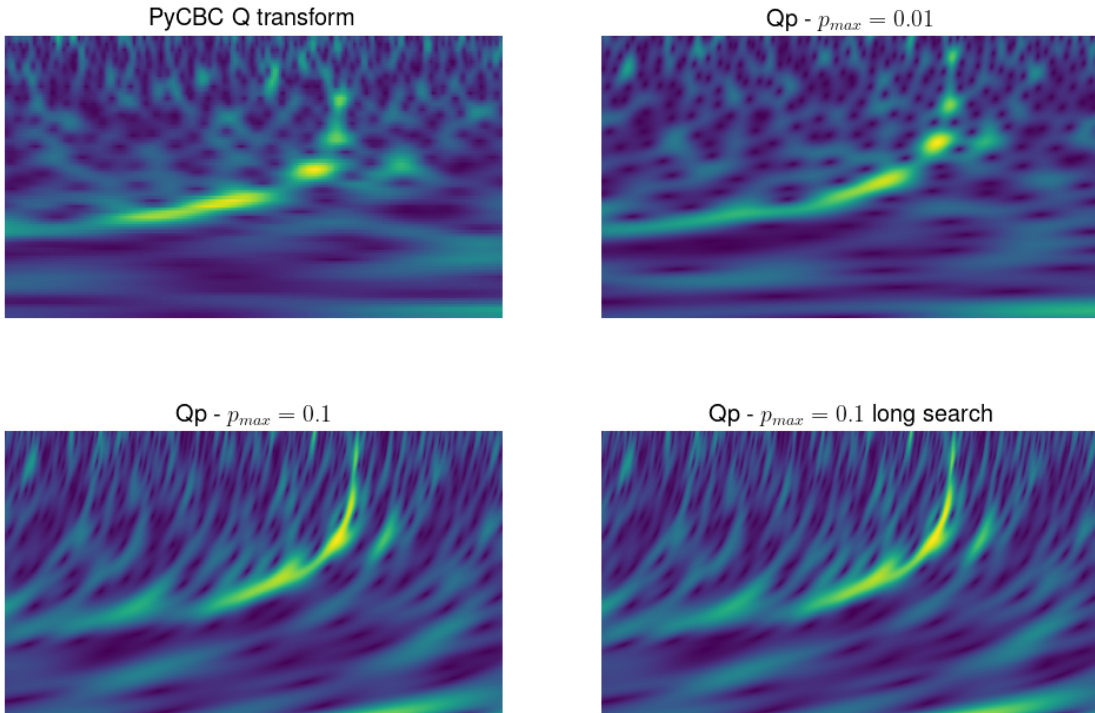


Figure 37: Comparison between transforms, this figure highlights that Qp helps to remove gap in signal representation. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

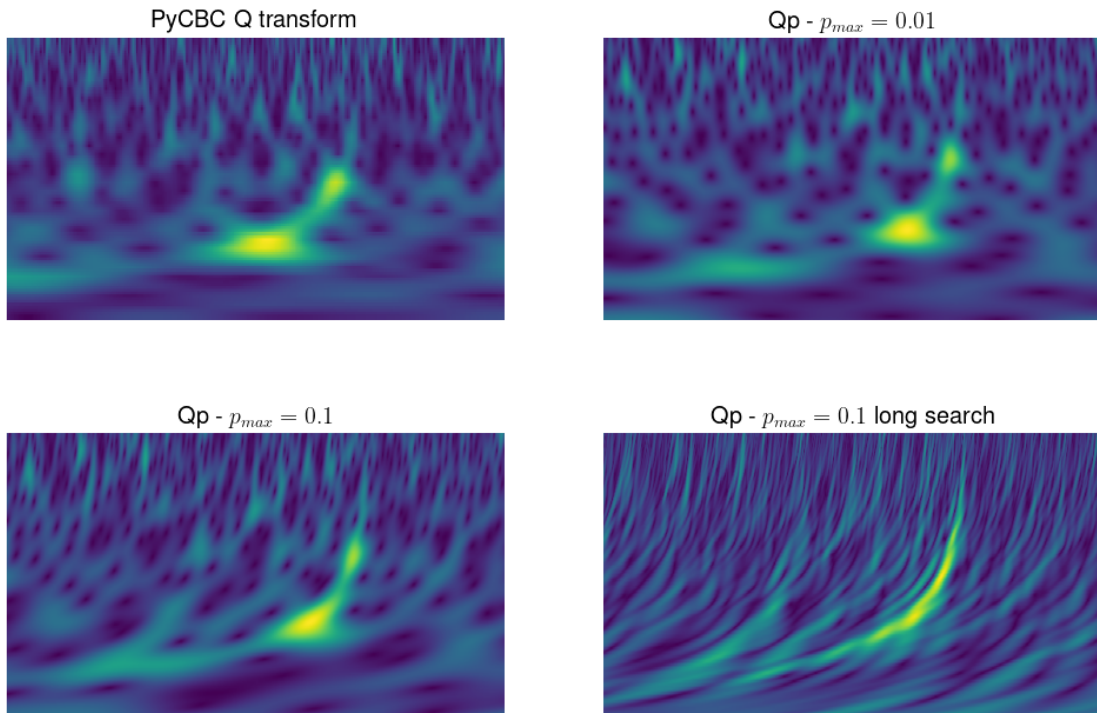


Figure 38: Comparison between transforms, this figure highlights that Qp long search could result in better signal representation. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

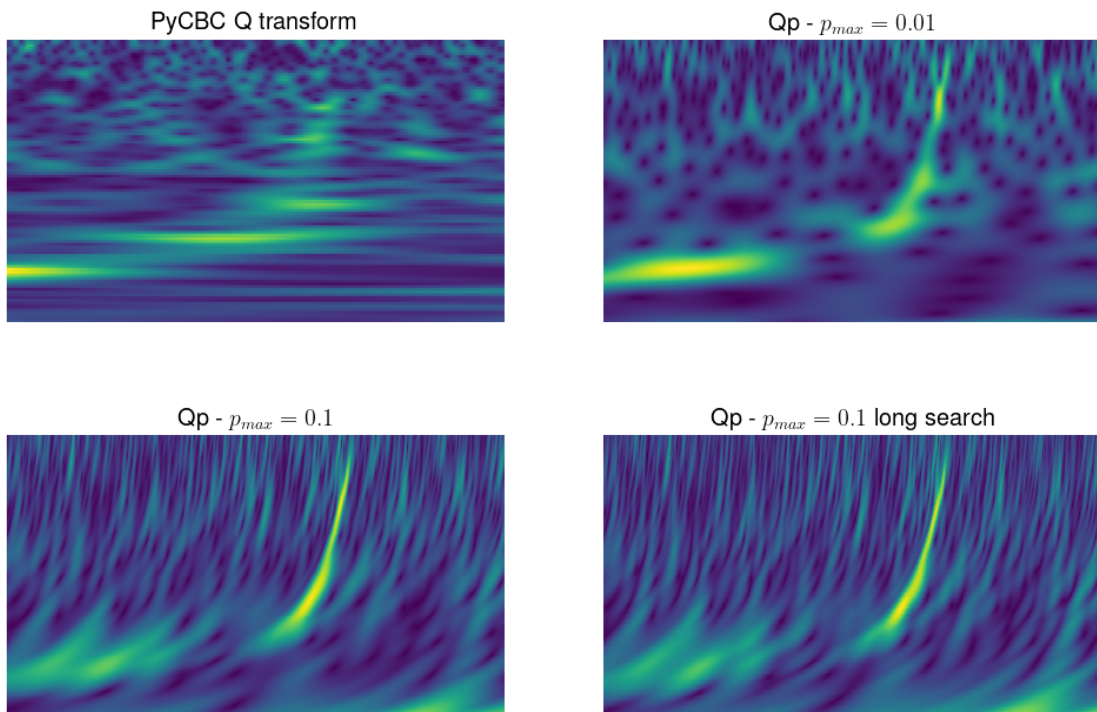


Figure 39: Comparison between transforms, this figure highlights that a mixture between local Qp transform could result in better signal representation. For example, by taking $p_{\max} = 0.01$ for the first half time and $p_{\max} = 0.1$ for the remaining time. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

More examples of the aspect presented in this section can be observed in Appendix C.

5.2 Training and Evaluation

In Section 5.1.2, we observe that the Qp transform seems to perform at the same level or even better in certain circumstances to represent the chirping patterns.

We will now discuss how we have generated the images used to train LensID. We emphasise once again that our focus is on the portion of LensID which searches within the spectrograms, we do not compare the sky maps associated with gravitational wave events to simplify the analysis.

The main principle behind this part of LensID is to compare two images and give a probability that both images come from the same event. In the case where they come from the same event, we will say we have a lensed pair, in the other case, an unlensed pair.

In all the 6866 images we have generated, we find 1000 unlensed images and 2933 lensed pairs, thus 2 times 2933 images. To make unlensed pairs, from n unlensed images we can associate them to get a maximum of $n(n-1)/2$ pairs because the solution of this problem can be mapped to $\sum_{i=1}^{n-1} i$. We thus separated our 1000 unlensed images into two groups of 500 images and were able to form 2 groups of 124,750 lensed pairs.

We follow [59] to generate a set of strongly lensed pairs of GW events. The way we generated all these images was to use the PyCBC module in Python [44]. First to generate gravitational wave signals we used `get_td_waveform` with the approximant `IMRPhenomXPHM` along with a set of parameters such as mass and distance. Second, we needed to generate noise in which the signals would lie. We thus picked a sample of detector data, coming from 15 October 2023, where we think there are no gravitational wave signals and evaluate its PSD. We then used this PSD to generate Gaussian noise using the function `noise_from_psd` of PyCBC [44]. At the end, the noise is added to our gravitational wave signal to form the strain of data which mimics a gravitational wave signal in a detector. Note that this Gaussian noise generation is based on a stochastic process thus when comparing different transforms (Q and Qp for example), one should keep track of the noise with a seed to generate the noise.

Once we had all those images, we distributed them in a training, validation and test set. The validation set is here used to determine whether the model has a good capability of generalisation. Afterwards, we choose to test the obtained model on completely different data, a test set, to obtain statistics about our model.

We thus need to choose the size of the training, validation and test set to evaluate our method. Here we choose to use 8000 unlensed and 1933 lensed pairs for the training and validation with 80 percent of this dataset for the training and the remaining for the validation. We used a batch of 124,750 unlensed pairs along with the remaining 1000 lensed pairs to test the performance once the model is trained.

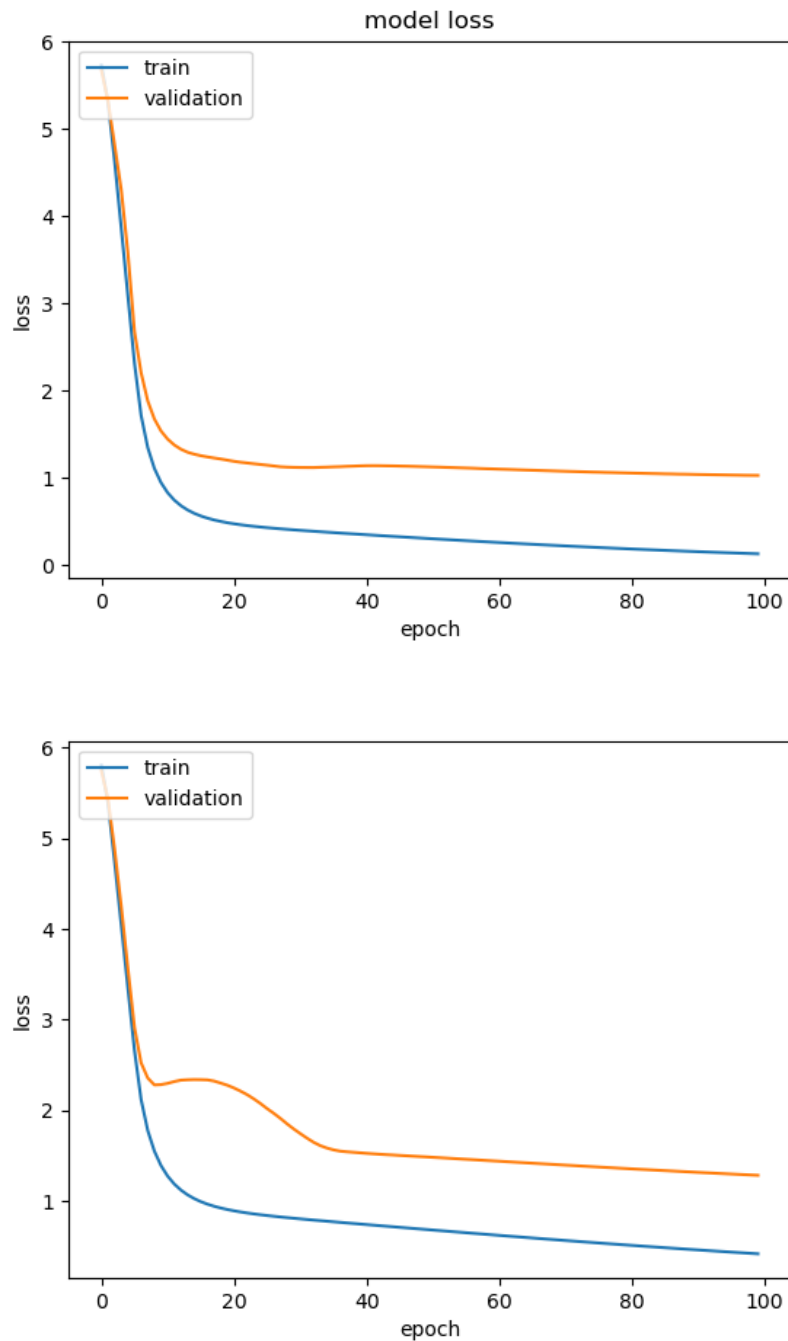


Figure 40: Loss from Q transform (top) and Q_p_{max=0.1} transform (bottom).

We present in Figure 40, an example of a training and validation curve for the Q transform and Q_p transform with $p_{\max} = 0.1$, they represent the evolution of the loss, here the binary cross-entropy as described in Section 3.1. Usually, LensID is trained during 25 epochs, which means that the algorithm sees 25 times the training data set. We see in our case that the loss does not increase after 25 epochs but is quite constant meaning that the models could be trained more to obtain better results.

One other way to check if the model could be trained more is via the accuracy of the

model over different epochs, it describes how many pairs are classified in the right categories, we represent this curve in Figure 41. We again see that the model needs more than 25 epochs to be at maximum efficiency, we would recommend training the model between 40 and 60 epochs to have a stable solution in this case. In our case, as we needed to train several models, Q transform, Qp transform with different values of p_{\max} , we choose to keep a hundred epochs for all models and simply check whether the loss seems to increase or the accuracy seems to decrease. For all models, a hundred epochs work fine and we thus kept the final model to evaluate the ROC plot. All the associated plots can be found in Appendix D if one wants to check our findings.

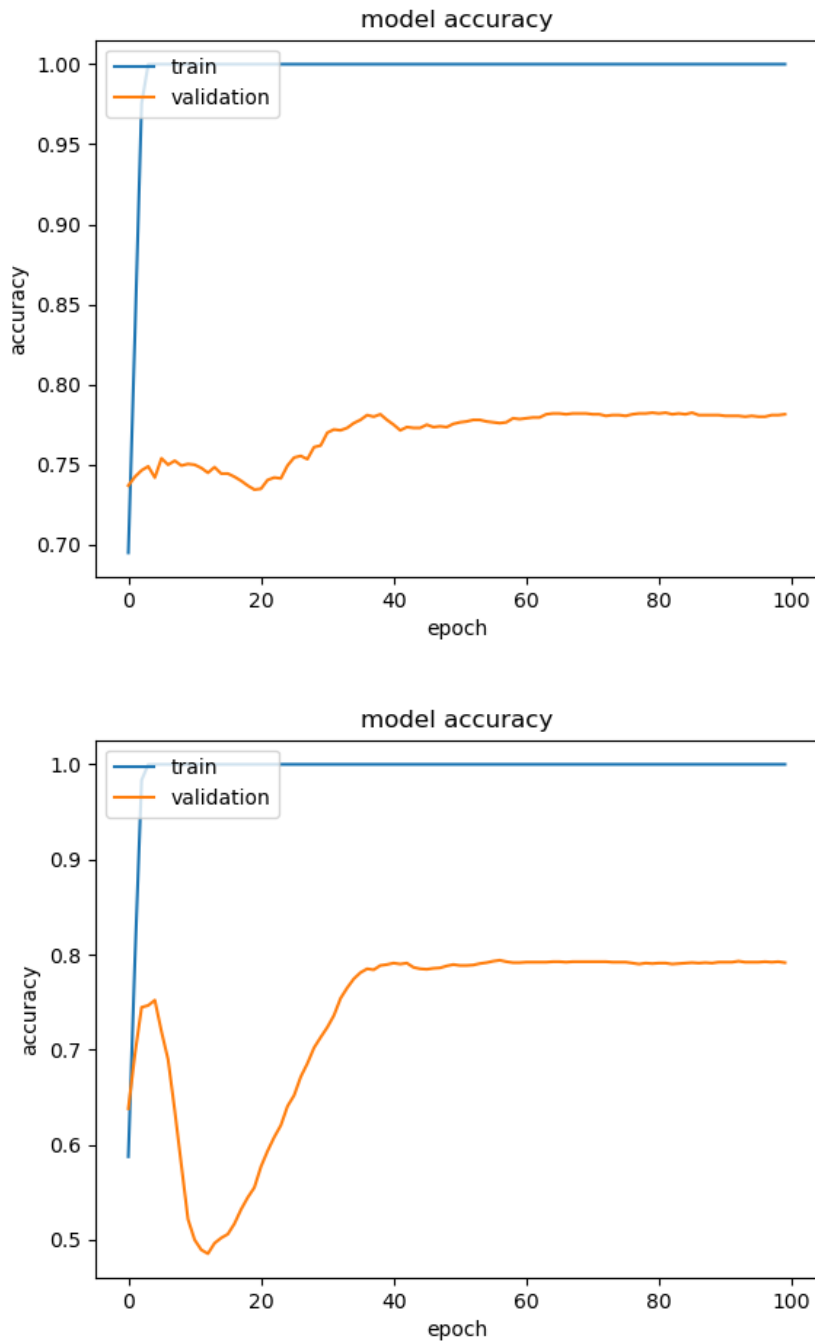


Figure 41: Accuracy from Q transform (top) and Qpmax=0.1 transform (bottom).

As a side note, we want to underline the peculiar behaviour observed in Figure 40 and 41 for the Qp Transform with $p_{\max} = 0.1$. We can observe a local increase of the loss and decrease of the efficiency of the model between epochs 10 and 25, where usually we only observe monotonic behaviour. We hypothesise that the models momentarily forget certain features learned from the ImageNet database on which it was pre-trained and start to learn the one of chirping pattern.

5.3 Result

After the training of our models during one hundred epochs, we obtained several models depending on the transform used. We will here discuss which models outperform others and how we think they manage to reach such sensitivity.

As discussed in Section 3.1, LensID consist in a fast recognition algorithm of lensed signals. In this objective, LensID fixes a threshold under which a couple of signals is said to be potentially lensed and then passed to the other algorithm (GOLUM [32] and HANABI [34]), this threshold is generally taken at a FPR of 10^{-2} . With this in mind, we are going to evaluate the performance of our models by comparing the ROC curve with FPR between $[10^{-5}, 10^{-2}]$.

First, let us compare how the Q transform of PyCBC performs in comparison to the one of GWpy (computed via the Qp transform with $p=0$). We see in Figure 42, that GWpy seems to perform way better than PyCBC while both are way better than random models as they are above the dashed diagonal.

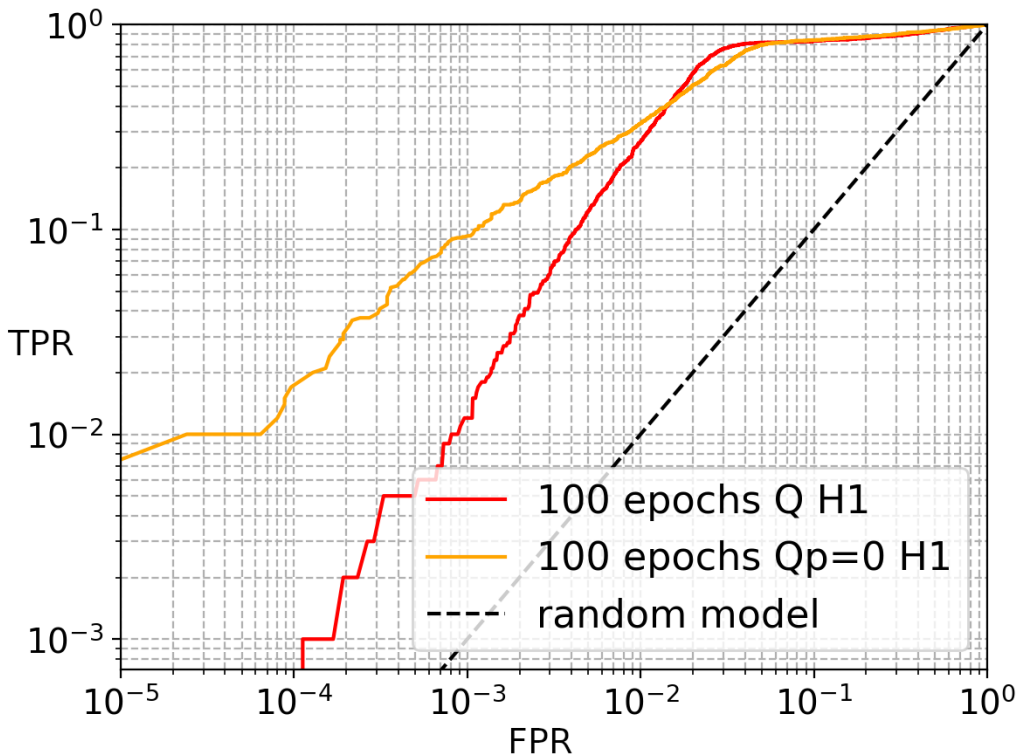


Figure 42: Comparison between different ROC curve for the Q transform of PyCBC and GWpy.

Next, we wanted to compare how the Qp transform performs in comparison to the Q transform of GWpy, we tried to compare the Q transform and the Qp transform with p_{\max} equal to 0.01 and 0.05 during 10^3 iterations to find the optimal Q and p parameter, see Figure 43 for their ROC curve. Here we see that below a FPR of 10^{-2} , both Qp transform perform better than the Q transform.

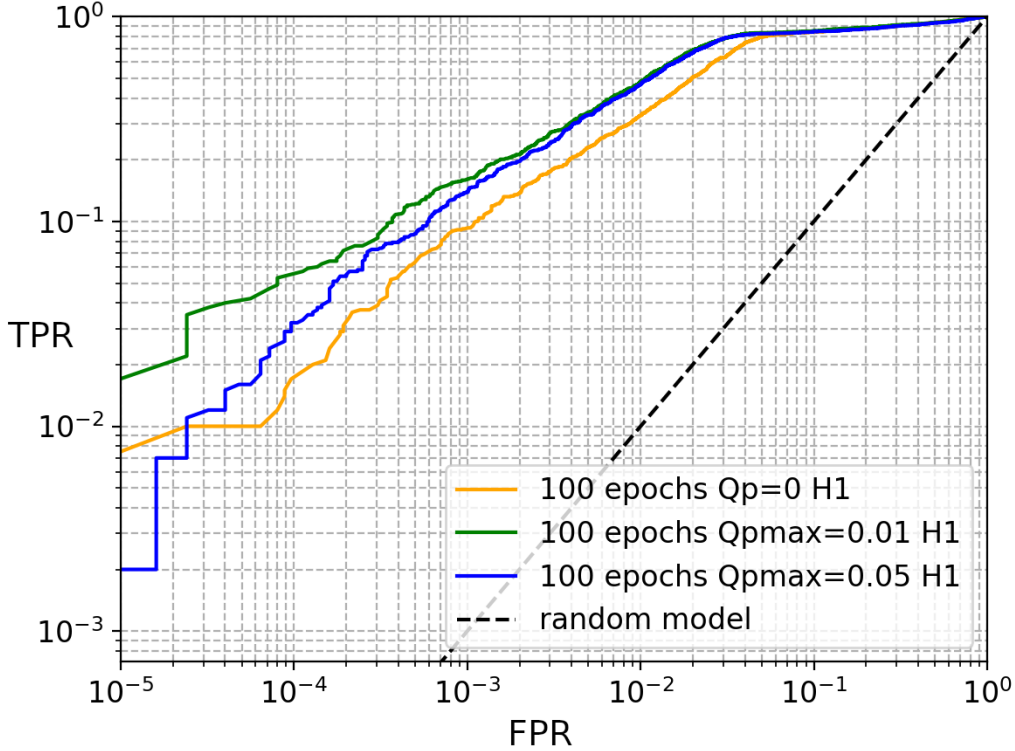


Figure 43: Comparison between different ROC curve for the Q transform and the Qp transform with $p_{\max} = 0.01$ and 0.05 during 10^3 iterations.

These results make sense as the Qp transform helps to recognise the shape of the chirp pattern observed for gravitational waves better.

Choosing a p_{\max} seems to be a real challenge to take the form of the chirp into account, this is why we think that extension of the Qp transform either at second order or making it a local transform should be an interesting point to make this task less arduous as they would take the shape of chirp into account better. In addition, improving the optimisation algorithm which selects the Q and p value is of great importance to enhance the rate of image analysed in a given time. This could be done by studying a large sample of gravitational wave signal and then evaluating the Q and p parameter distribution obtained in order to reduce the hypothesis space. Other models are present in Appendix E.

6 Conclusion

In our search, we studied the Qp transform [1] in various cases. First, in Section 4, the consideration of the Qp transform was made due to the need to fill a gap in the spectrogram representation of the event GW191105, in this manner we hoped that the recognition of lensed gravitational wave could be more efficient. This idea emerged because LensID [51], a machine learning algorithm partially consists in a comparison between spectrograms to tell whether two gravitational wave events come from the same object. The main problem came from GW191105 which did not match the standard representation of a gravitational wave signal with the usual Q transform while it does with the Qp transform. Second, our objective was to check if the Qp transform increases LensID sensitivity. We thus dived into the generation of spectrograms in Section 5.1, where we observed that the Qp transform helps to improve the spectrogram representation of the gravitational wave signal. During this inspection, we also found that two implementations, PyCBC [44] and GWpy [45], of the well-known Q transform used in gravitational wave astronomy did not lead to the same spectrograms representation because of the Q -parameter selection. Third, we trained the spectrogram analysis of LensID with the different transforms where we indeed observed that the Qp transform, when limiting the hypothesis space of the p parameter, performs better than the usual Q transform. During the training of the different algorithms, we also observed that the usual training made during about 25 epochs was not sufficient to reach the maximal sensitivity but rather needed about 40 epochs.

During this thesis, several ideas came to our mind to enhance gravitational wave representation with spectrograms. As discussed in Section 2.5, the Qp transform consists in a first-order Taylor expansion in the frequency of the Q transform around a given time leading to a supposed linear shape for signal in spectrograms. We think it might be interesting to go up to second order to take into account the gravitational wave pattern in spectrograms which looks like parabolas, in the same manner as [46]. A generalisation of this idea would be to make some kind of polynomial expansion of the Q transform for more complicated patterns observed in spectrograms. Later on, in Section 5.1.2, we have considered the idea of making a local transform to enhance the spectrogram representation, this transform could look like the variable-Q transform introduced in [58] while the current Q and Qp transform used in gravitational wave physics choose their parameter Q and p based on the amplitude of the whole spectrogram, see also Section 5.1.1 for more detail. Our idea would be to choose different parameters for different slices in the spectrograms, slices which could be along time or frequency bands. This idea, in a different manner than the polynomial transform would help to represent signals in spectrograms because it would help to represent with an increased resolution different parts of the spectrograms. We want to highlight the fact that mixing those ideas might be also interesting, with for example a local second order Q transform aiming with the local aspect in a high precision of representation and with the second order with a suited parabola shape which represents finely a gravitational wave signal. Other interesting questions were highlighted such as the fact that we do not know the Q and p parameter distribution for different gravitational wave signals in Section 5.3 or the increase in the loss while training LensID observed in Section 5.2.

Finally, let us mention other areas of research where spectrograms also contribute to understanding our world better such automatic detection of whistle made by dolphin [60] or more generally all areas where frequency needs to be extracted while keeping time information. As a consequence, all the observations and questions highlighted in this thesis do not only take part in the gravitational wave analysis but also more generally in the fields of signal processing and machine learning methods.

A Access to C00 data

In order to access C00 data, one needs access to the server of the LIGO collaboration with LIGO credentials via SSH connexion, it will grant you access to the needed C00 data. Then one should use the following lines of code :

```
#set server for token
export HTGETTOKENOPTS="--vault server vault.ligo.org --issuer igwn"

#use token to read data
htgettoken -a vault.ligo.org --audience ANY --scopes gwdatafind.read
#then login on the web page shared in Powershell with the Ligo email adress

#look for data in a specific detector & time (here L1 around GW191105) in OSDF
python3 -m gwdatafind -r datafind.igwn.org -o L -t L1_HOFT_C00 -s 1256999700 -e
1256999800 -u osdf

#Then one gets the name of the file one is looking for such as
osdf:///igwn/ligo/frames/03/hoft/L1/L-L1_HOFT_C00-12569/
L-L1_HOFT_C00-1256996864-4096.gwf

#get a token to read
htgettoken -a vault.ligo.org --audience ANY --scopes read:/ligo
#download the data
stashcp osdf:///user/ligo/frames/ER8/hoft/L1/L-L1_HOFT_C00-11262/
L-L1_HOFT_C00-1126256640-4096.gwf ./

#Often one needs to get the channel to load data and use it, use
FrChannels L-L1_HOFT_C00-1126256640-4096.gwf
#First output string is the channel name and the second is the sampling rate
#Then use common technique to read a gwf file
```

B PyCBC vs GWpy

We present here images which help to compare PyCBC [44] and GWpy [45], two different implementations of the Q transform. Two columns are shown each containing on the left, PyCBC and on the right, GWpy implementation.

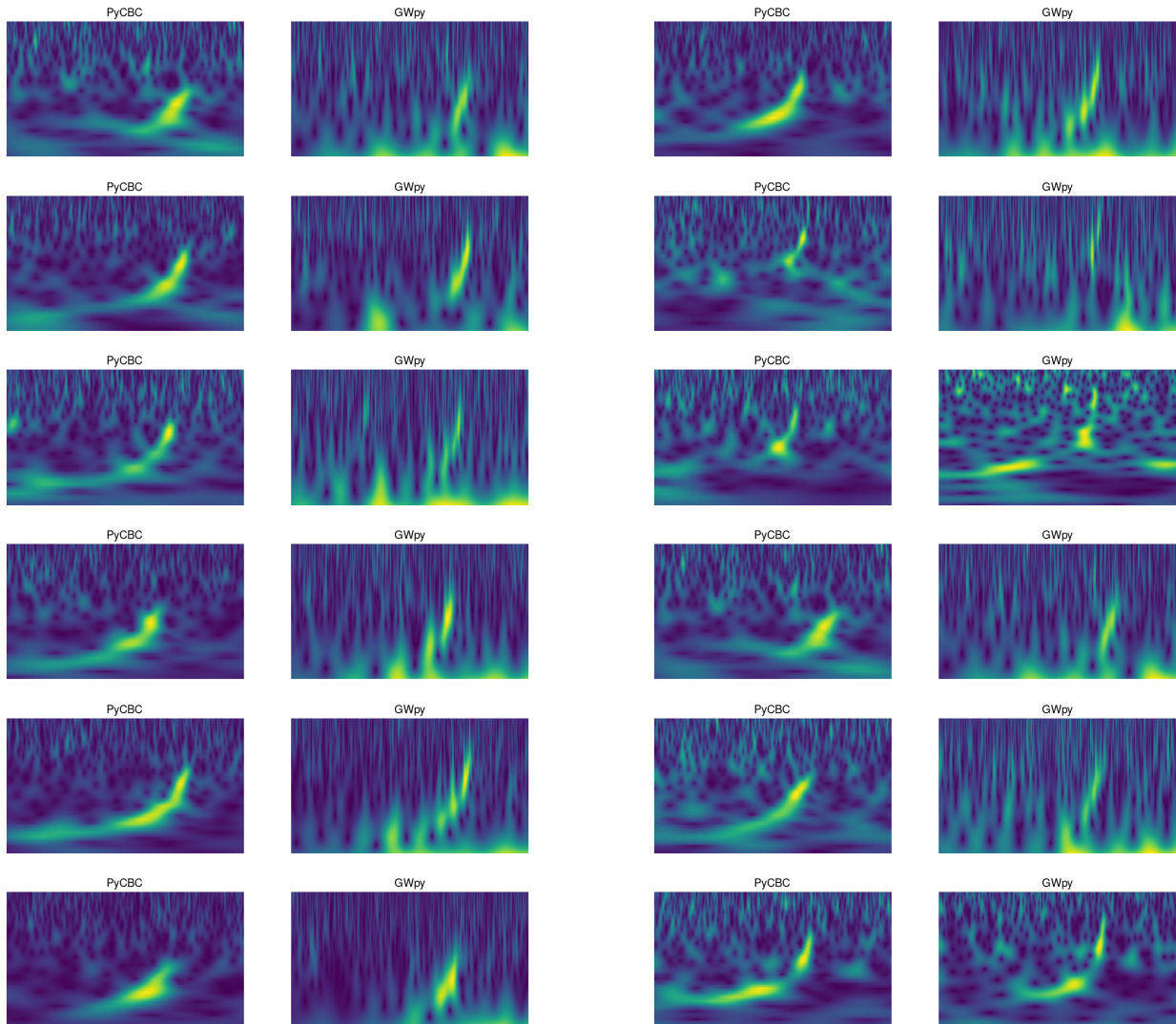


Figure 44: Two columns are shown each containing on the left, PyCBC and on the right, GWpy implementation. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

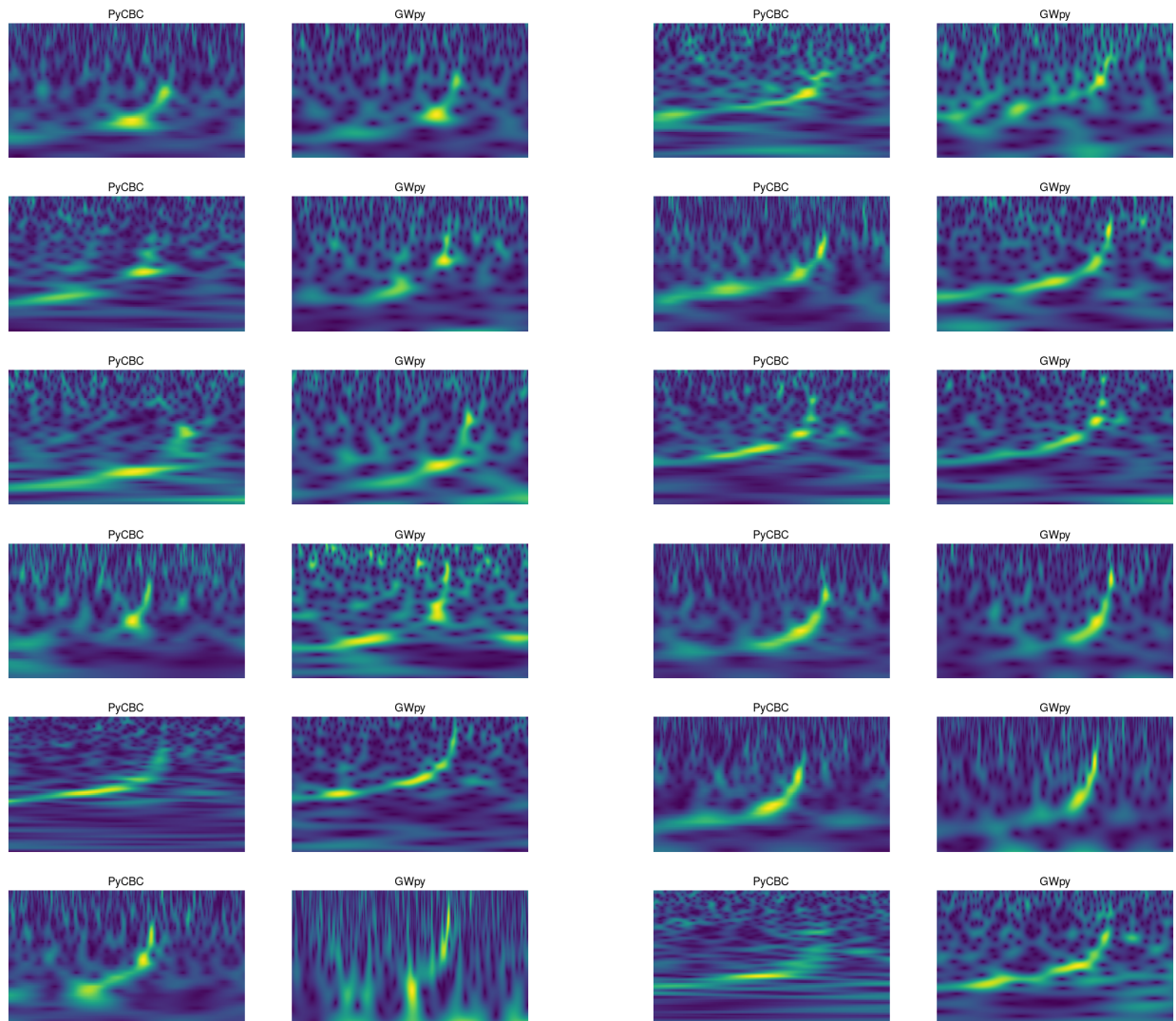


Figure 45: Two columns are shown each containing on the left, PyCBC and on the right, GWpy implementation. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

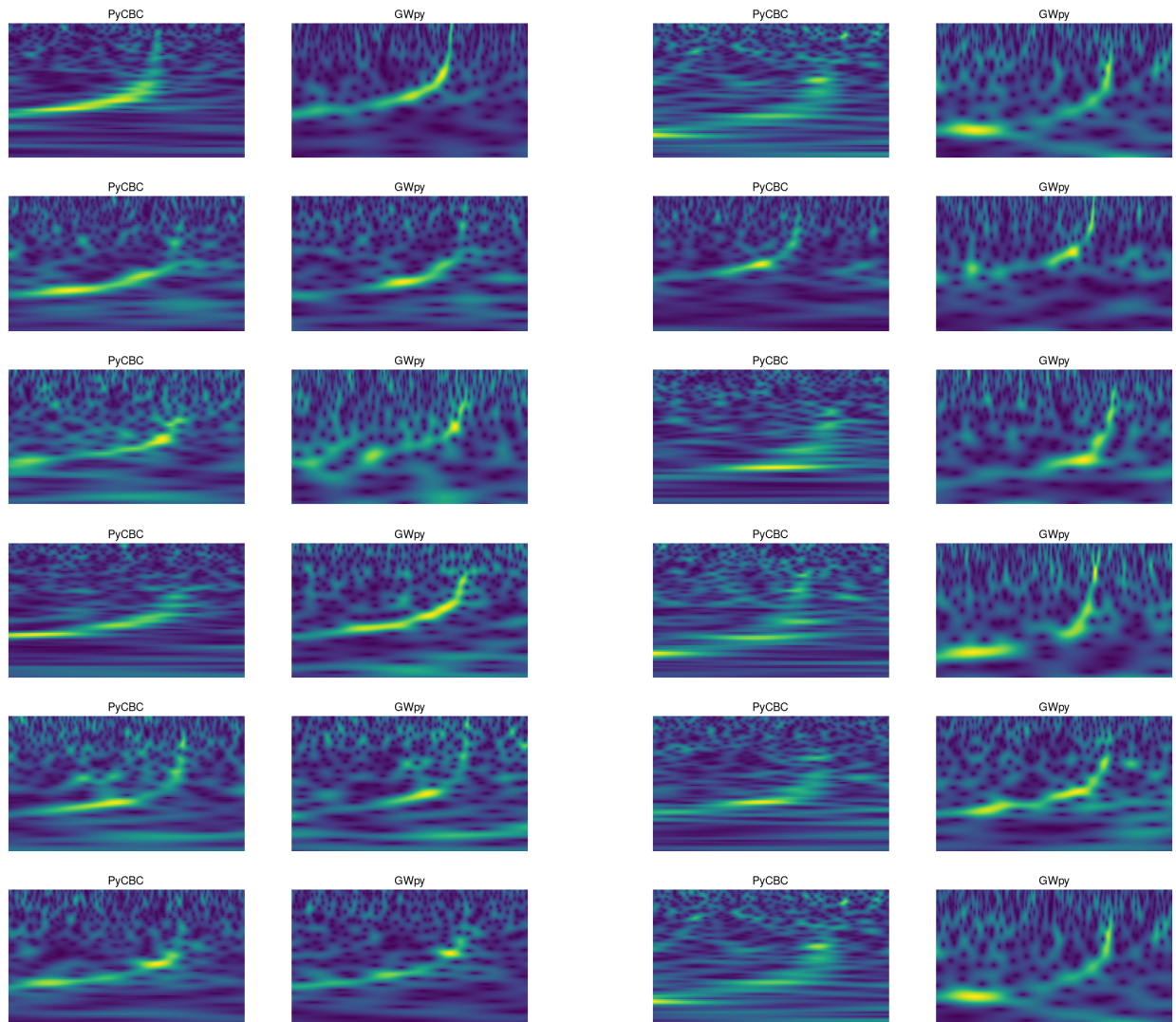


Figure 46: Two columns are shown each containing on the left, PyCBC and on the right, GWpy implementation. Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

C Q and Qp Transform

The objective of this Appendix is to show more comparison of Q and Qp transform to mark the differences. We remind the notion of p_{\max} which is defined via $p \in [0, p_{\max}]$. We will always display four images, Q transform of PyCBC, Qp transform with $p_{\max} = 0.01$, $p_{\max} = 0.1$ which search for (Q, p) value during 1000 iteration and also Qp transform with $p_{\max} = 0.1$ searching during 1 million iteration, which we denote with long search.

First, other examples of images which shows that p_{\max} should not be too small in some cases because the chirping pattern is harder to distinguish. We also observe that the Qp transform helps to have better resolution.

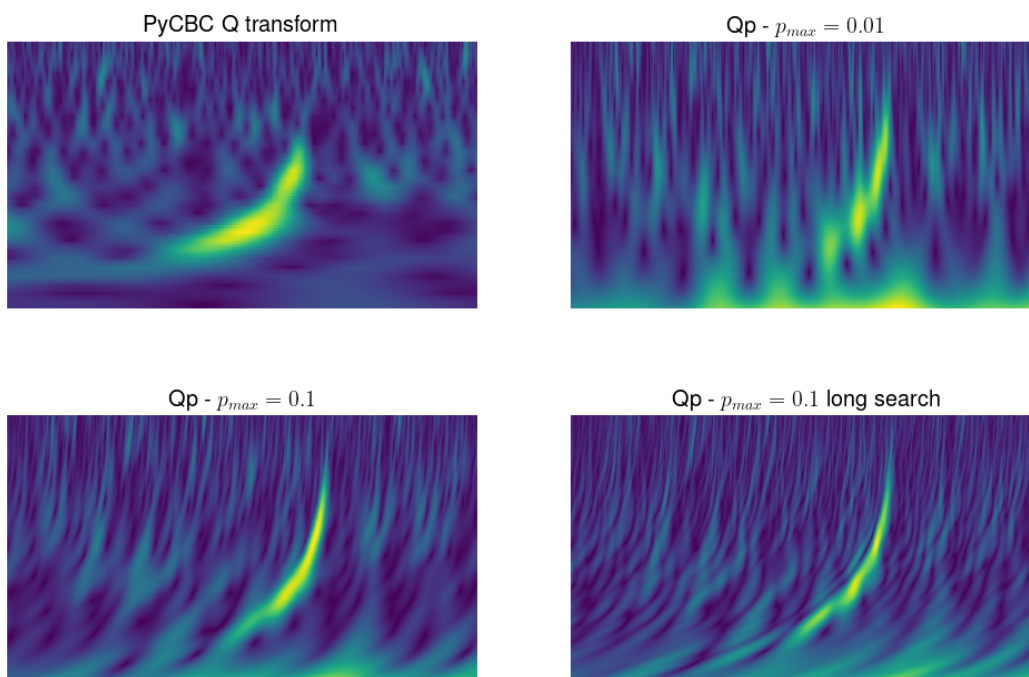


Figure 47: Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

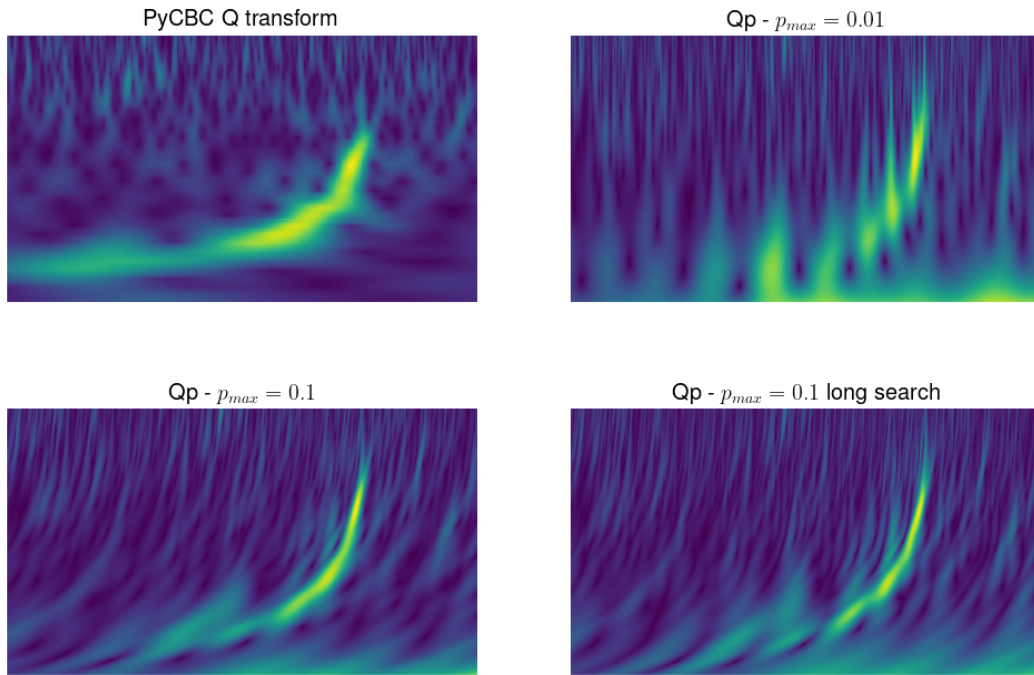


Figure 48: Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

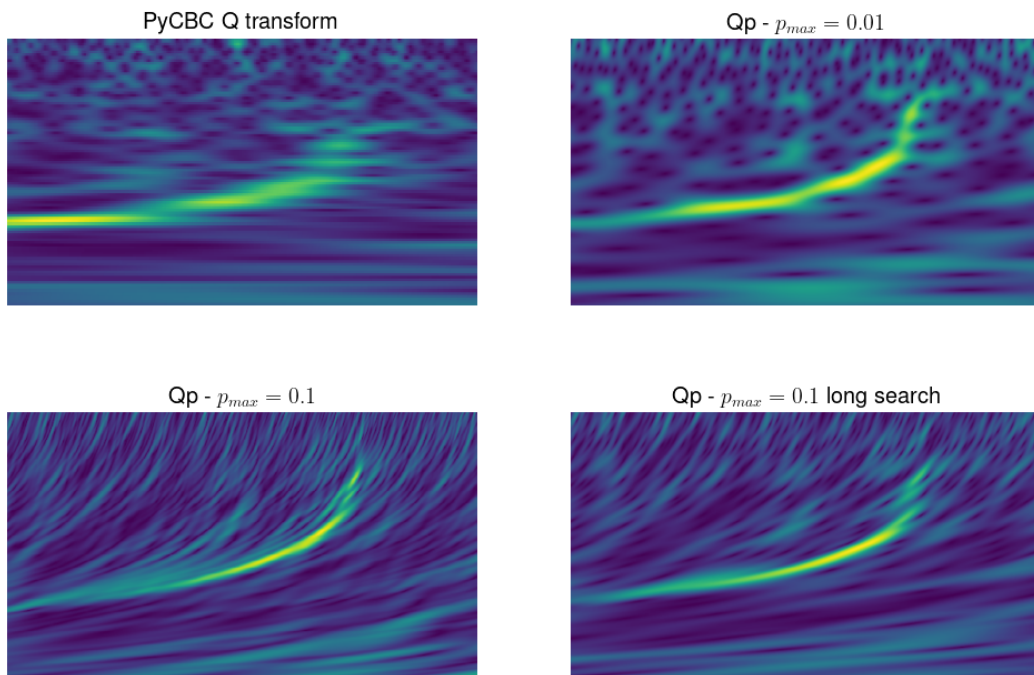


Figure 49: Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

Second, other examples of images where Qp transform helps to fill gaps inside the spectrograms.

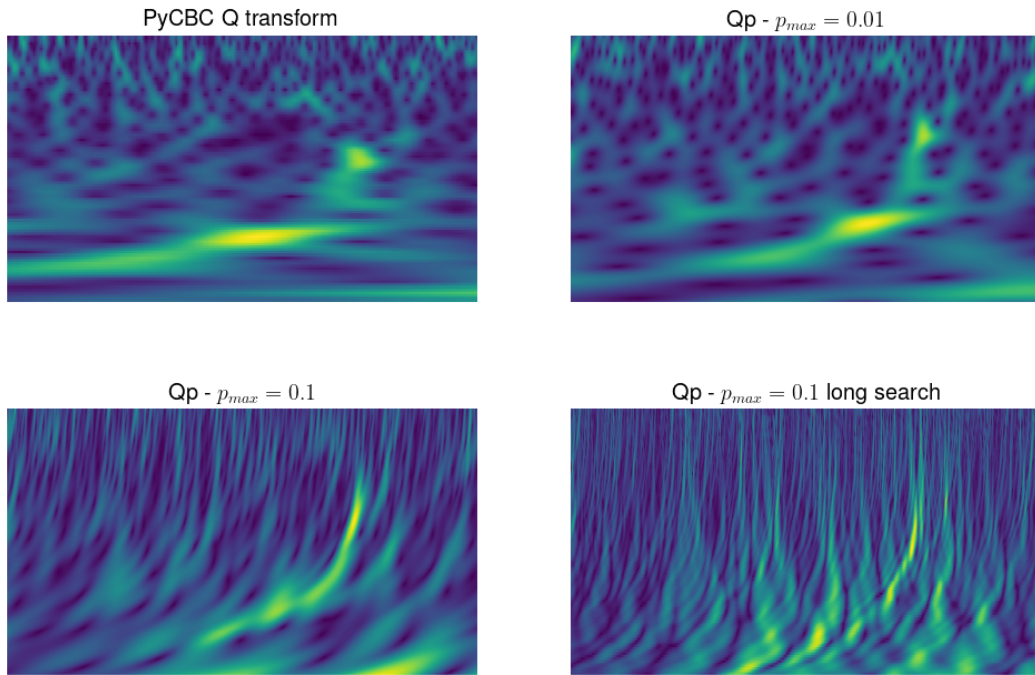


Figure 50: Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

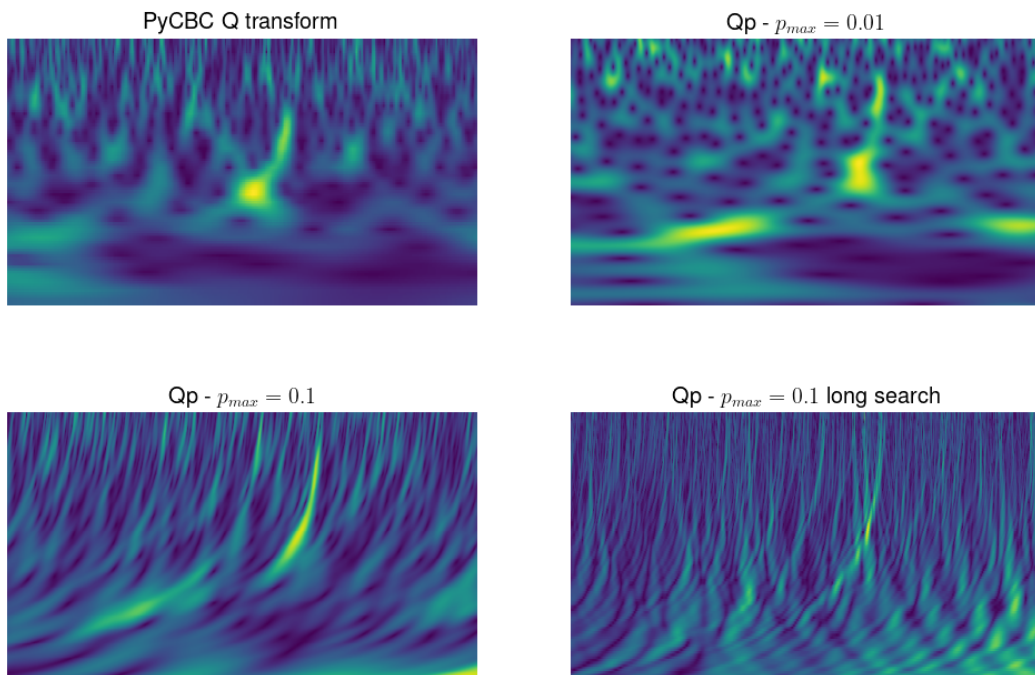


Figure 51: Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

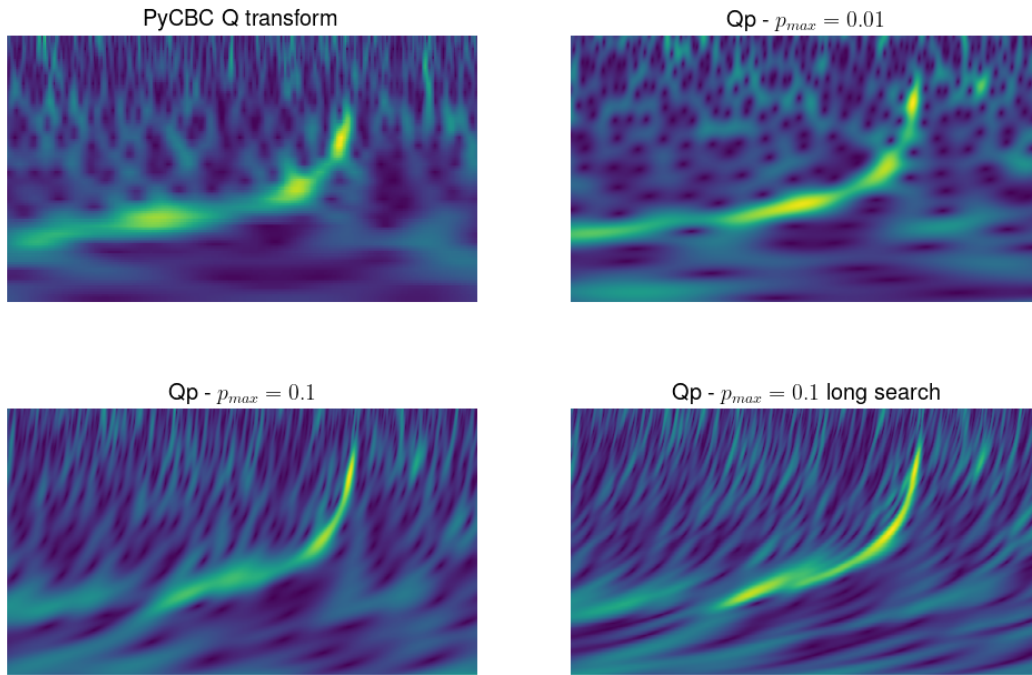


Figure 52: Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

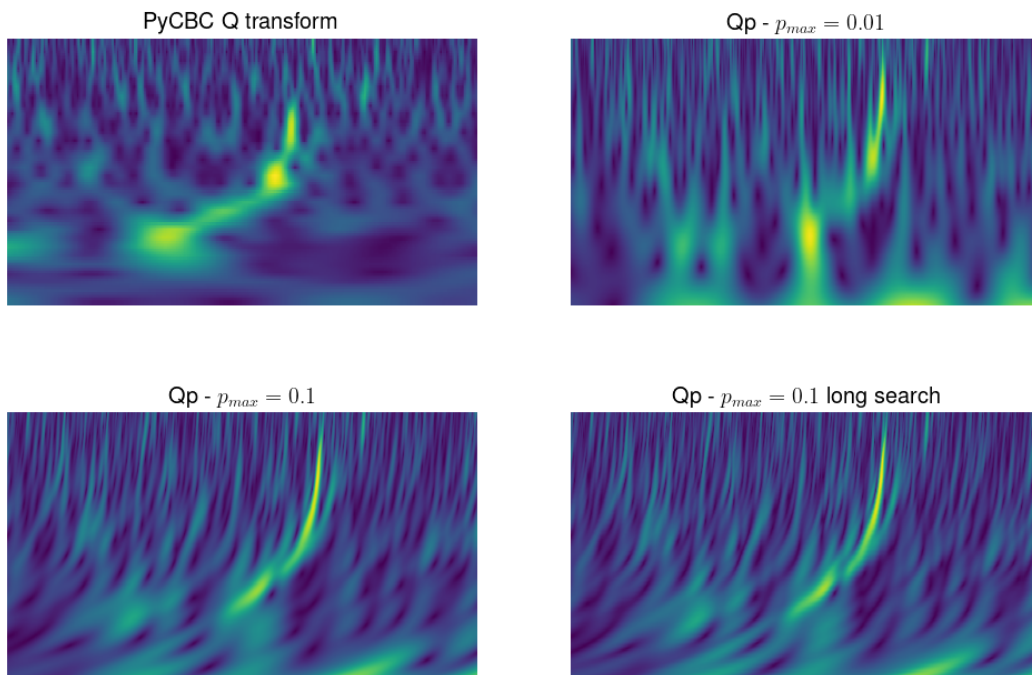


Figure 53: Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges.

Third, another example of images indicating that developing a local Qp transform might be interesting to make mixture of increased resolution for both low and high frequency.

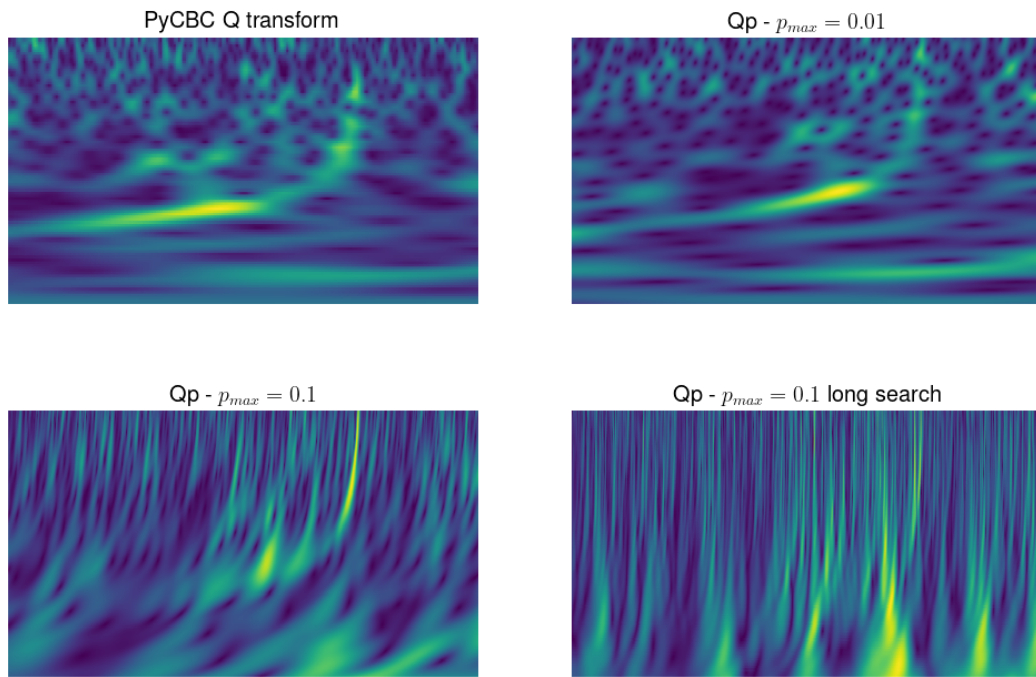


Figure 54: Axis are removed to represent what the machine learning model sees, see Figure 30 for ranges. Low frequency is well represented by Q transform and high frequency by Qp transform with $p_{\max} = 0.1$.

D Training : Loss and accuracy

We present here the other losses of our machine learning model. For discussion, see Section 5.2.

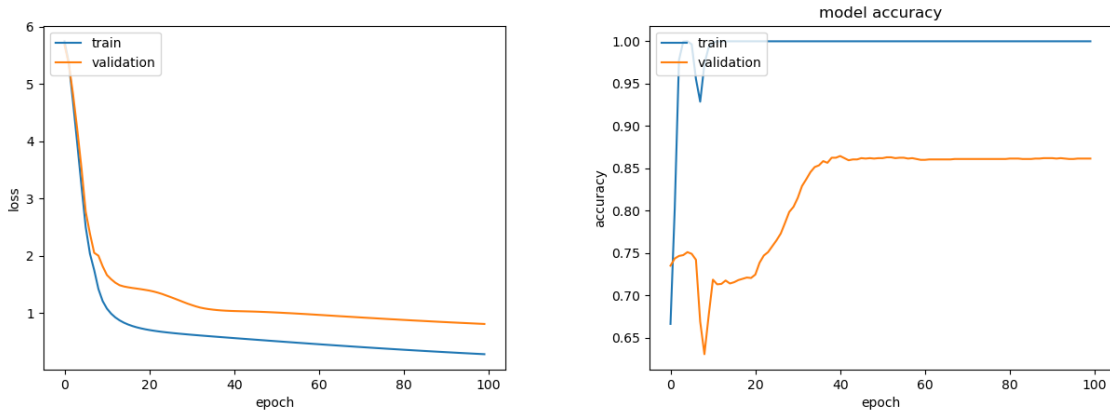


Figure 55: Loss and accuracy from $Q_{pmax}=0.01$ transform.

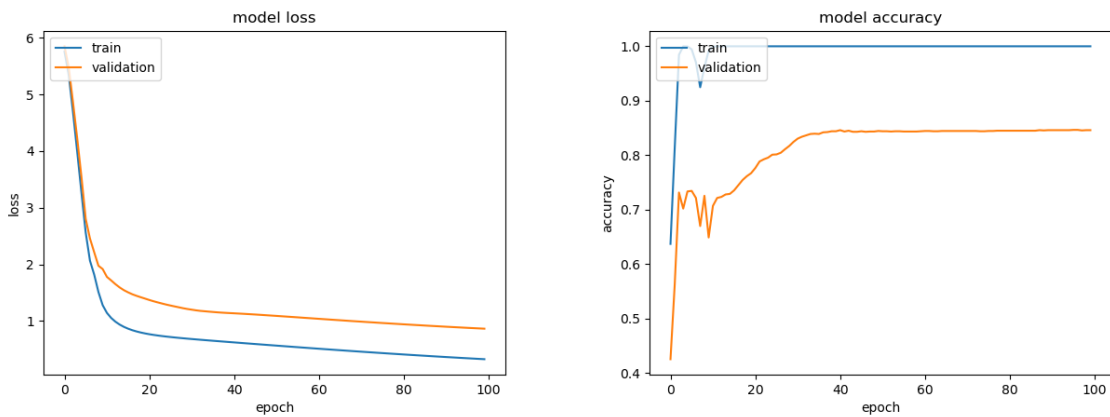


Figure 56: Loss and accuracy from $Q_{pmax}=0.05$ transform.

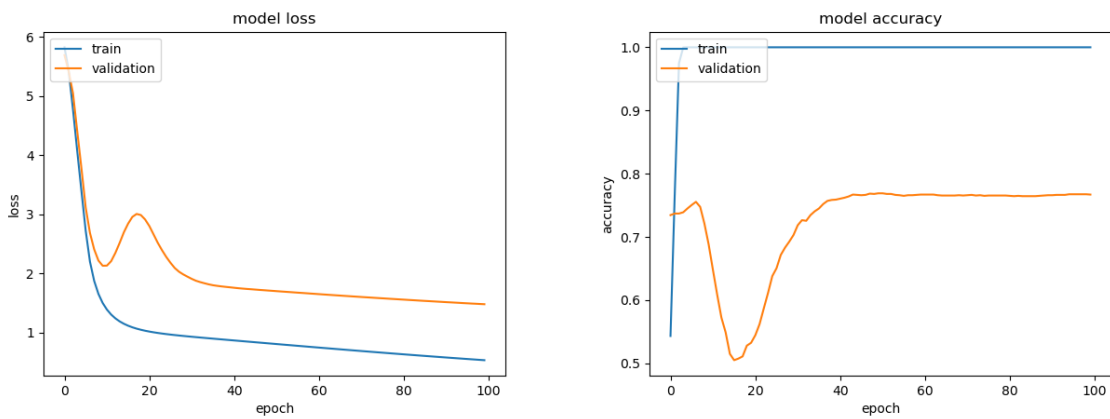


Figure 57: Loss and accuracy from Q_p long search transform.

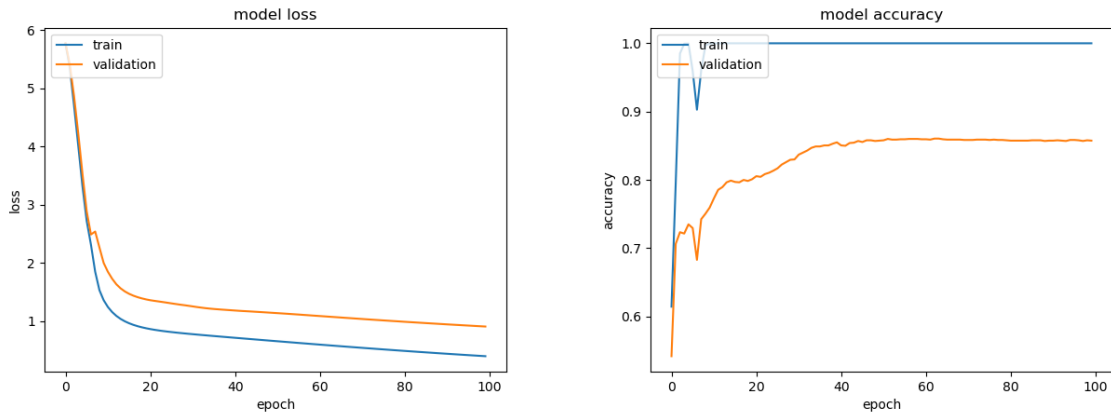


Figure 58: Loss and accuracy from Qp=0 transform.

E Other models : LensID with Qp transform

Here, we wanted again to compare how the Qp transform performs in comparison to the Q transform of GWpy, but now for the Qp transform with $p_{\max} = 0.1$ during 10^3 and 10^6 iterations (long search) to find the optimal Q and p parameter, we present the ROC curve in Figure 59. In this case, we observe that the Q transform outperforms the Qp transform at FPR below 10^{-3} .

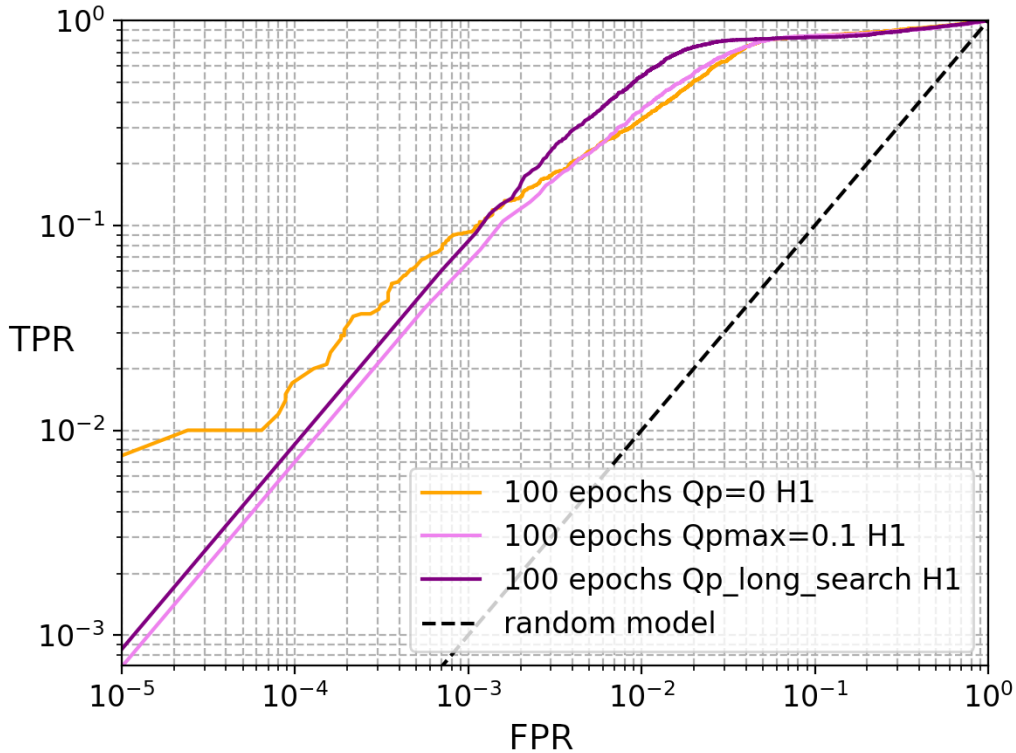


Figure 59: Comparison between different ROC curve for the Q transform and the Qp transform with $p_{\max} = 0.1$ during 10^3 and 10^6 iterations.

Moreover, in Figure 60, we compare the result of the different Qp transform. At the lowest FPR, it is clear that Qp transform with $p_{\max} = 0.01$ performs the best but it is also important to note that above 10^{-2} , all algorithm have nearly the same TPR.

This result is surprising because one should expect that taking a value of $p_{\max} = 0.1$ with a high number of iterations to find the optimal Q and p would help to find the optimal parameter and thus long search $p_{\max} = 0.1$ should be better than $p_{\max} = 0.01$ which is clearly not the case here. This was discovered at the end of our search, we decided to present it here for future investigation.

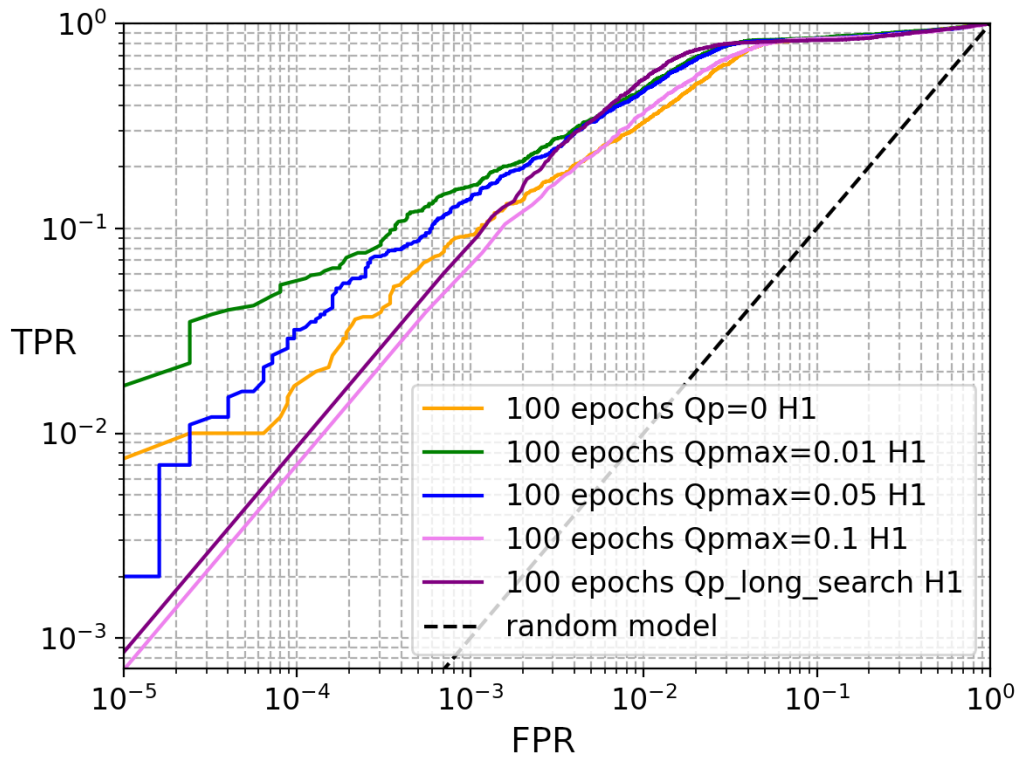


Figure 60: Comparison between different ROC curve for the Q transform and the Q_p transform with $p_{max} = 0.01, 0.05$ during 10^3 iterations and for 0.1 during 10^3 and 10^6 iterations.

References

- [1] A. Virtuoso and E. Milotti, *Wavelet-based tools to analyze, filter, and reconstruct transient gravitational-wave signals*, 2024.
- [2] S. Goyal, *Strong lensing of gravitational waves and tests of general relativity*, PhD thesis (International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bengaluru, India, 2023).
- [3] A. Einstein, *Die feldgleichungen der gravitation*, Sitzungsberichte der Koniglich Preussischen Akademie der Wissenschaften, 844–847 (1915).
- [4] D. Kennefick, *History of gravitational wave emission*, in 6th francis bacon conference (2016).
- [5] R. A. Hulse and J. H. Taylor, *Discovery of a pulsar in a binary system*, The Astrophysical Journal **195**, L51 (1975).
- [6] N. P. Organization, *Discovery of a new type of pulsar, a discovery that has opened up new possibilities for the study of gravitation*, Accessed : 20/02/2024, 1993.
- [7] B. P. Abbott et al., *Observation of gravitational waves from a binary black hole merger*, Physical Review Letters **116**, 10.1103/physrevlett.116.061102 (2016).
- [8] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), *Gw150914: first results from the search for binary black hole coalescence with advanced ligo*, Phys. Rev. D **93**, 122003 (2016).
- [9] S. Soni, E. Marx, E. Katsavounidis, R. Essick, G. C. Davies, P. Brockill, M. W. Coughlin, S. Ghosh, and P. Godwin, *Qoq: a q-transform based test for gravitational wave transient events*, Classical and Quantum Gravity **41**, 015012 (2023).
- [10] C. E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal **27**, 379–423 (1948).
- [11] L. Debnath, *A short biography of joseph fourier and historical development of fourier series and fourier transforms*, International Journal of Mathematical Education in Science and Technology **43**, 589–612 (2012).
- [12] J. W. Cooley and J. W. Tukey, *An algorithm for the machine calculation of complex fourier series*, Mathematics of computation **19**, 297–301 (1965).
- [13] A. Oppenheim, *Discrete-time signal processing*, Prentice-Hall **2**, 878–882 (1999).
- [14] N. Straumann, *General relativity*, 2nd, Graduate Texts in Physics (Springer Verlag, Berlin, 2013).
- [15] L. D. Landau and E. M. Lifšic, *Physique théorique: tome ii: théorie des champs* (Mir, 1970).
- [16] H. Stephani, D. Kramer, M. MacCallum, C. Hoenselaers, and E. Herlt, *Exact solutions of einstein's field equations*, 2nd ed., Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2003).
- [17] I. Belahcene, *Searching for gravitational waves produced by cosmic strings in ligo-virgo data*, Theses (Université Paris Saclay (COmUE), Oct. 2019).
- [18] N. van Remortel, K. Janssens, and K. Turbang, *Stochastic gravitational wave background: methods and implications*, Progress in Particle and Nuclear Physics **128**, 104003 (2023).
- [19] C. Caprini and D. G. Figueroa, *Cosmological backgrounds of gravitational waves*, Classical and Quantum Gravity **35**, 163001 (2018).

- [20] C. J. Moore, R. H. Cole, and C. P. L. Berry, *Gravitational-wave sensitivity curves*, Classical and Quantum Gravity **32**, 015014 (2014).
- [21] T. L. S. Collaboration and J. Aasi, *Advanced ligo*, Classical and Quantum Gravity **32**, 074001 (2015).
- [22] F. a. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, G. Balestri, G. Ballardin, et al., *Advanced virgo: a second-generation interferometric gravitational wave detector*, Classical and Quantum Gravity **32**, 024001 (2014).
- [23] B. P. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, M. Adamo, C. Adams, T. Adams, P. Addesso, et al., *Characterization of transient noise in advanced ligo relevant to gravitational wave signal gw150914*, Classical and Quantum Gravity **33**, 134001 (2016).
- [24] P. Nguyen, R. Schofield, A. Effler, C. Austin, V. Adya, M. Ball, S. Banagiri, K. Banowetz, C. Billman, C. Blair, et al., *Environmental noise in advanced ligo detectors*, Classical and Quantum Gravity **38**, 145001 (2021).
- [25] *LIGO - laser interferometer gravitational-wave observatory*, Accessed: April 9, 2024.
- [26] J. Marx et al., *The gravitational wave international committee roadmap: the future of gravitational wave astronomy*, 2011.
- [27] A. O. Petters, H. Levine, and J. Wambsganss, *Singularity theory and gravitational lensing*, Vol. 21 (Springer Science & Business Media, 2012).
- [28] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational lenses as astrophysical tools* (Springer, 1992).
- [29] M. Meneghetti, *Introduction to gravitational lensing: with python examples*, Vol. 956 (Springer Nature, 2021).
- [30] J. Janquart et al., *Follow-up analyses to the o3 ligo–virgo–kagra lensing searches*, Monthly Notices of the Royal Astronomical Society **526**, 3832–3860 (2023).
- [31] S. Goyal, H. D., S. J. Kapadia, and P. Ajith, *Rapid identification of strongly lensed gravitational-wave events with machine learning*, Phys. Rev. D **104**, 124057 (2021).
- [32] J. Janquart, O. A. Hannuksela, K. Haris, and C. Van Den Broeck, *A fast and precise methodology to search for and analyse strongly lensed gravitational-wave events*, Monthly Notices of the Royal Astronomical Society **506**, 5430–5438 (2021).
- [33] J. Janquart, K. Haris, O. A. Hannuksela, and C. Van Den Broeck, *The return of golum: improving distributed joint parameter estimation for strongly lensed gravitational waves*, Monthly Notices of the Royal Astronomical Society **526**, 3088–3098 (2023).
- [34] R. K. L. Lo and I. Magaña Hernandez, *Bayesian statistical framework for identifying strongly lensed gravitational-wave signals*, Physical Review D **107**, 10.1103/physrevd.107.123015 (2023).
- [35] O. A. Hannuksela, T. E. Collett, M. Çalıřkan, and T. G. Li, *Localizing merging black holes with sub-arcsecond precision using gravitational-wave lensing*, Monthly Notices of the Royal Astronomical Society **498**, 3395–3402 (2020).
- [36] D. Gabor, *Theory of communication. part 1: the analysis of information*, Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering **93**, 429–441 (1946).

- [37] M. D. Morales, J. M. Antelis, C. Moreno, and A. I. Nesterov, *Deep learning for gravitational-wave data analysis: a resampling white-box approach*, *Sensors* **21**, 10.3390/s21093174 (2021).
- [38] S. Mallat, *A wavelet tour of signal processing* (Elsevier, 1999).
- [39] I. Daubechies, *Ten lectures on wavelets* (SIAM, 1992).
- [40] J. C. Brown, *Calculation of a constant q spectral transform*, *The Journal of the Acoustical Society of America* **89**, 425–434 (1991).
- [41] S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis, *Multiresolution techniques for the detection of gravitational-wave bursts*, *Classical and Quantum Gravity* **21**, S1809 (2004).
- [42] S. K. Chatterji, *The search for gravitational wave bursts in data from the second ligo science run*, PhD thesis (Massachusetts Institute of Technology, 2005).
- [43] G. Gambardella, *The mellin transforms and constant- q spectral analysis*, *The Journal of the Acoustical Society of America* **66**, 913–915 (1979).
- [44] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, *Pycbc inference: a python-based parameter estimation toolkit for compact binary coalescence signals*, *Publ. Astron. Soc. Pac.* **131**, 024503 (2019).
- [45] D. M. Macleod, J. S. Areeda, S. B. Coughlin, T. J. Massinger, and A. L. Urban, *Gwpy: a python package for gravitational-wave astrophysics*, *SoftwareX* **13**, 100657 (2021).
- [46] V. Bruni, M. Tartaglione, and D. Vitulano, *A fast and robust spectrogram reassignment method*, *Mathematics* **7**, 10.3390/math7040358 (2019).
- [47] L. Wehenkel and P. Geurts, *Introduction to machine learning - lecture 5 (deep) neural networks*, Lecture slides, Accessed 15/04/2024, 2022.
- [48] R. Holbrook, *Overfitting and underfitting*, Kaggle Notebook, (2024) <https://www.kaggle.com/code/ryanholbrook/overfitting-and-underfitting> (visited on 05/13/2024).
- [49] PyTorch, *Pytorch documentation : conv2d*, Accessed: April 16, 2024.
- [50] PyTorch, *Pytorch documentation : avgpool2d*, Accessed: April 16, 2024.
- [51] S. Goyal, D. Harikrishnan, S. J. Kapadia, and P. Ajith, *Rapid identification of strongly lensed gravitational-wave events with machine learning*, *Physical Review D* **104**, 124057 (2021).
- [52] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, *Densely connected convolutional networks*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), 4700–4708.
- [53] C. François, *Keras*, 2015.
- [54] D. P. Kingma and J. Ba, *Adam: a method for stochastic optimization*, 2017.
- [55] R. Abbott et al., *Open data from the third observing run of ligo, virgo, kagra, and geo*, *The Astrophysical Journal Supplement Series* **267**, 29 (2023).
- [56] L. Collaboration, *Gravitational wave open science center(gwosc)*, Accessed: 13/02/2024, (2024) <https://gwosc.org/techdetails/>.
- [57] D. G. Keppel, *Signatures and dynamics of compact binary coalescences and a search in ligo’s s5 data*, PhD thesis (California Institute of Technology, 2009).
- [58] C. Schörkhuber, A. Klapuri, N. Holighaus, and M. Doerfler, *A matlab toolbox for efficient perfect reconstruction time-frequency transforms with log-frequency resolution*, in (Jan. 2014).

- [59] K. Haris, A. K. Mehta, S. Kumar, T. Venumadhav, and P. Ajith, *Identifying strongly lensed gravitational wave signals from binary black hole mergers*, arXiv preprint arXiv:1807.07062 (2018).
- [60] O. Serra, F. Martins, and L. Padovese, *Active contour-based detection of estuarine dolphin whistles in spectrogram images*, *Ecological Informatics* **55**, 101036 (2020).