

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Physics of Data

Final Dissertation

Mitigation of instrumental glitches in gravitational wave data analysis

External Supervisors

Prof. Chris Van Den Broeck

Dr. Tom Dooney

Candidate

Giulia Doda

Internal Supervisor

Prof. Jean-Pierre Zendri

Academic Year 2025/2026

Abstract

The development of a deep learning-based methodology for the accurate and fast classification and reconstruction of noise transients, called glitches, in gravitational-wave detectors is underway. Glitches, which originate from environmental disturbances or instrumental artifacts, appear in different unmodelled morphologies and represent one of the major challenges in gravitational-wave data analysis: beyond reducing detection confidence, they can significantly compromise the extraction of astrophysical information from observed signals. To date, approximately 25% of detected events require some form of glitch mitigation due to an overlap between the gravitational-wave signal and transient noise contamination. Improving the identification and subtraction of glitches is therefore essential for maximising the scientific output of present and future observing runs of the LIGO–Virgo–KAGRA collaboration and next generation detectors. The goal of this work is first to validate the methodology based on the recently developed tool called DeepExtractor on real interferometer data, ensuring physically consistent reconstructions and compatibility with existing gravitational-wave data-analysis pipelines. Results of the validation study mainly highlight great reconstruction capabilities of DeepExtractor, especially for certain glitch classes, and limitations of the `glitchfind` method employed in the analysis. In addition, the work includes the development of a dedicated Deep Extractor model specifically tailored to the Virgo detector. Such a model will account for the unique instrumental configuration, environmental conditions, and detector-specific noise characteristics of Virgo.

Acknowledgements

I would like to express my sincere gratitude to the members of the GRASP group at Utrecht University and the Nikhef GW group in Amsterdam for their support, guidance, and for providing such a stimulating research environment throughout my thesis work.

First and foremost, I would like to thank my supervisor, Prof. Dr. Chris Van Den Broeck, for giving me the opportunity to carry out my thesis within the GRASP group, as well as for his guidance and scientific vision.

I am also deeply grateful to Dr. Tom Dooney, who supervised my work on a daily basis. Thank you for your constant guidance, patience, and willingness to discuss ideas, problems, and results with me. I have learned a great deal from you, both scientifically and personally, and I truly appreciate all the time and effort you dedicated to my work. Thank you also for your trust in me.

I would like to thank all the members of the group, both on the particle and GW sides, for making me feel welcome from the very beginning and for the many interesting discussions and moments shared throughout these months.

Thanks to Suzanne, Luca, and Nick for being such good friends and for all the time we spent together. Thanks to Thibeau for the cookies at Journal Club, and to Melissa for her help with some computational issues and for being such an inspiring researcher. Thanks to Thomas, Fabian, and Mick for the coffee breaks, somewhere between academic advice and philosophical debates.

I am also very grateful to Mees, with whom I had the pleasure of sharing an office as fellow Master's students. It was great to have someone with whom to exchange ideas, discuss the ups and downs of our theses, and learn from each other's experiences. I particularly enjoyed our many discussions about our projects, as well as the Dutch lessons along the way.

I would like to thank everyone else in the group for contributing to such a friendly and supportive atmosphere. I am very grateful for the opportunity to have been part of this research community and for everything I have learned during my time here.

Finally, I am also grateful to my supervisor at the University of Padova, Prof. Dr. Jean-Pierre Zendri, for his guidance, for pointing me toward the Nikhef GW research group and for holding me in such high esteem.

Contents

Introduction **1**

1 Gravitational Waves and How to Detect Them **3**
1.1 Einstein's equations in the weak field limit 4
1.2 Gravitational waves sources 7
1.3 Gravitational waves detection 9
1.3.1 Noise in optical interferometers 13
1.4 Multimessenger astrophysics 23

2 Gravitational Waves Data Analysis **27**
2.1 Gravitational waves data 28
2.2 Detector characterization 30
2.2.1 Data Quality 31
2.2.2 Glitch mitigation 33
2.3 Searches 34
2.4 Machine learning 35

3 Deep learning for glitch mitigation **39**
3.1 DeepExtractor 40
3.1.1 Training strategy 41
3.1.2 Architecture of the model 42
3.1.3 Comparison with Bayes Wave and reconstruction of actual GW signals 43
3.2 DeepExtractor for Virgo 45
3.2.1 Data and training strategy 45
3.2.2 Validation data and tools 49
3.2.3 Validation results 53

Conclusions **65**

Introduction

The direct detection of gravitational waves (GWs) has opened up a whole new world of physics, especially regarding observations of compact object mergers and multimessenger astrophysics. Currently operating GW detectors are continuing to improve in sensitivity and next generation detectors are being designed and will be operating in the next decades. In this perspective, the quality and the reliability of GW data has become increasingly critical. One of the main challenges in GW data analysis is the presence of non-Gaussian noise artifacts, named glitches. They can mimic astrophysical GW signals or overlap with them, strongly reducing data quality. Glitches are unpredictable; they are generated by instrumental or environmental disturbances, but the lack of known instrumental couplings makes it difficult to identify their source. Also, if a glitch overlaps with a GW signal, which happens in about 25% of the cases, glitch mitigation is required, but these techniques inevitably have a lot of consequences on the physical information that can be extracted from the signal. The main example is the only detected signal from a binary neutron star merger: a glitch overlapped with the signal in one detector, and the impact of the glitch on the extracted physical information about neutron star physics is still being debated. Developing accurate and robust glitch mitigation strategies is therefore essential to maximize the scientific output of current and future observing runs, in particular stratiegies must be prepared for next generation detectors.

In this context, deep learning methodologies have demonstrated great potential for glitch classification and reconstruction. However, these models often offer limited interpretability and their integration in GW analysis pipelines requires proper and careful validation. A very promising deep learning tool is DeepExtractor [20]: the model has been designed to extract excess power, no matter the source, from detector data and it can be therefore useful both for glitch mitigation and actual GW signals reconstruction. It has been proved to have potential to complement and surpass the state of the art glitch mitigation tools used in official analyses. The DeepExtractor model was based only on LIGO detectors, therefore this work focuses on the development of a tailored model for Virgo detector that could account for its peculiarities given by the instrumental design and the environmental conditions, leading to better performances. However, the crucial step in the applcation of a machine learning tool in physics is testing its reliability in producing scientific results. The developed DeepExtractor model for Virgo has been therefore validated with `glitchfind` [36] method to check for excess power in DeepExtractor background noise reconstructions. This work underlines the importance of data quality in GW detectors, that will be increasingly crucial with next generation detectors, and also how essential it is to ensure the interpretability of machine learning methods especially in physics. Before being deployed, the great potential of DeepExtractor to be incorporated into official data analysis pipelines must be indeed thoroughly reviewed.

This study is organized as follows: Chapter 1 contains a brief introduction on GW theory and detection, from optical interferometers, with a specific focus on the noise, to the astrophysical processes emitting GW signals. In Chapter 2 is presented an overview of the data obtained from GW interferometers, with a particular attention on data quality, and the different pipelines of GW data analysis are illustrated. Finally, the contributions of machine learning to the different stages of GW data analysis are briefly described. Chapter 3 presents DeepExtractor, the developement of a tailored model for the Virgo detector and the results of the validation study carried out on DeepExtractor's reconstructions to assess whether the model produces reliable data from a physics perspective.

Chapter 1

Gravitational Waves and How to Detect Them

A brief review of the theoretical background of Einstein's theory of General Relativity is essential for understanding how gravitational waves (GWs) arise in this theory and how they can be detected. General Relativity assumes that gravity is universal, meaning that gravity acts on all particles in the same way. Therefore, what we experience as gravity should be a fundamental feature of the common background tissue permeating the universe. In this context gravity should appear as a manifestation of the spacetime curvature. Indeed, General Relativity describes the relationship between matter and spacetime curvature. It shows how matter shapes spacetime and how spacetime curvature acts as gravity on matter.

GWs are predicted by General Relativity and they can be seen as ripples in spacetime, alterations of the gravitational field propagating at the speed of light. Many astrophysical objects are predicted to emit GWs in different processes, from coalescences between compact objects, such as black holes and neutron stars, to supernovae explosions. As of now, the most common way to detect GWs are optical interferometers. They rely on the fact that distances between two test masses should change as the GW pass through the system. As a consequence, we should observe a corresponding change in the time spent by the light travelling between the test masses. However, the trace left by a GW crossing a ground-based optical interferometer is extremely faint and it is necessary to deal with a huge amount of background noise to confidently identify a GW signal and consequently estimate physical parameters relative to the signal source. Noise modelling is therefore of paramount importance in this context and data quality is essential, as reliable scientific results ultimately depend on the ability to identify and remove instrumental artifacts before any physical interpretation can be made.

Since the first detection in 2015, GWs have provided us with a new tool to investigate the universe and a new field of physics has opened up. One of the most significant contributions offered by GWs is multimessenger astrophysics, allowing to study cosmic events using combined data also from gravitational radiation. By merging complementary signals, we can obtain a more comprehensive understanding of some phenomena, like neutron star mergers. GWs physics is a very young field and many discoveries can therefore be made, as well as a lot of benefits can be gained by other fields using GWs, such as cosmology. Additionally, new ground- and space-based GWs detectors are being built

and designed to explore new GW sources and improve sensitivity.

1.1 Einstein's equations in the weak field limit

The core idea of Albert Einstein's theory of General Relativity is condensed in the corresponding field equations. These equations characterize gravity in a compact and innovative way with respect to Newtonian gravitation. Gravity is not seen as a force anymore: General Relativity describes it as a geometrical property of the spacetime itself. The Einstein's field equations of General Relativity are:

$$R_{\mu\nu} - \frac{1}{2}\, g_{\mu\nu}\, \mathrm{R} \;=\; \frac{8\pi \mathrm{G}}{c^4}\, T_{\mu\nu}\,, \tag{1.1}$$

where:

- $R_{\mu\nu}$ is Ricci's tensor, a contraction of Riemann's tensor, expressing the spacetime curvature,
- $g_{\mu\nu}$ is the metric tensor, defining the spacetime geometry,
- R is the Ricci scalar, the trace of Ricci's tensor,
- G is the universal gravitation constant and c is the light speed in vacuum,
- $T_{\mu\nu}$ is the energy-momentum tensor, describing the distribution of energy and matter in spacetime, and it represents the source of gravitational field.

Additionally, the left hand side of 1.1 is also the definition for the Einstein tensor $G_{\mu\nu}$. Therefore, the Einstein's field equations express the relationship between the geometry of spacetime, represented by $G_{\mu\nu}$, and the matter content of spacetime, encoded in $T_{\mu\nu}$. In particular, these equations tell us how the spacetime curvature reacts to the presence of energy and matter, so how matter determines the gravitational field. At the same time, they also show how the gravitational field influences the behavior of matter, how spacetime curvature manifests itself as gravity. The game played by spacetime geometry and energy, whose rules are set by Einstein's equations, shapes the universe and manages its evolution.

Solving Einstein's equations would allow to understand how spacetime behaves depending on whether or not there is an energy source present and the dynamics in presence of gravity. Einstein's field equations are non-linear differential tensorial equations and the first, natural step towards solving them is turning to a simpler regime, the *weak-field limit*. It corresponds to the linear expansion of the equations about the flat-space metric:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}\,, \tag{1.2}$$

where $\eta_{\mu\nu}$ is Minkowski's metric tensor[1] and $h_{\mu\nu}$ is the tensor that introduces a small curvature perturbation on flat space, with the assumption that $|h_{\mu\nu}| \ll |\eta_{\mu\nu}| = 1$. Linearizing means therefore considering small disturbances on Minkowski's spacetime, where geometry is simple and linear. The linearized version of General Relativity therefore describes a symmetric tensor field $h_{\mu\nu}$ propagating on a flat background spacetime. The goal is therefore obtaining the equations of motion obeyed by

[1] The signature adopted here is (-,+,+,+).

the perturbations $h_{\mu\nu}$. Then, after expanding up to first order in $h_{\mu\nu}$ the Ricci tensor and the Ricci scalar, and introducing the reverse-trace metric as

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\,\eta_{\mu\nu}\,h \quad \text{where} \quad h \equiv h^{\sigma}_{\sigma}\,, \tag{1.3}$$

the linearized field equations follows:

$$\Box\,\bar{h}_{\mu\nu} = -\frac{16\pi \mathrm{G}}{c^4}\,T_{\mu\nu}\,. \tag{1.4}$$

Note that the $\Box = \partial_\mu\partial^\mu$ operator corresponds to the d'Alambertian, describing the propagation of the perturbation $h_{\mu\nu}$ both in time and through space. The expansion in weak-field regime allows to treat $h_{\mu\nu}$ itself as a scalar field on a flat Minksowski space. Mathematically, this is a set of equations with source terms very similar to the situation occurring in classical eletromagnetism.

The linearized equations can be solved *in vacuum*, where they yield to the usual wave-like solutions for each component of the $h_{\mu\nu}$ tensor. However, not all the components of $h_{\mu\nu}$ are independent because the Equation 1.2 is true in more than one coordinate system: it does not represent a unique decomposition and it leaves some extra freedom. Therefore, the non-physical degrees of freedom in $h_{\mu\nu}$ can be removed by imposing a gauge condition. This denotes a particular and convenient choice of coordinates that simplifies the equations and leaves only the true dynamical components, which represent the actual gravitational perturbations. The most useful gauge choice is the *transverse-traceless (TT) gauge*. The linearized equations in vacuum expressed in this gauge become

$$\Box h^{TT}_{\mu\nu} = 0\,, \tag{1.5}$$

where $h^{TT}_{\mu\nu}$ is the perturbation tensor expressed in TT gauge, enjoying the following properties:

- it is *purely spatial*, $h^{TT}_{0\nu} = 0$, meaning that, in this coordinate system, distorsions involving time can be transformed away, thus leaving only the measurable changes in spatial displacements,
- it is *transverse*, $\partial_\mu h^{\mu\nu}_{TT} = 0$, expressing the fact that spatial distorsions are othogonal to the perturbation's propagation direction,
- and it is *traceless*, $\eta^{\mu\nu}h^{TT}_{\mu\nu} = 0$, implying that only (spatial) shape distorsions are observable, and if there is expansion in one direction, then there will be contraction in the other one, recalling that the third spatial direction is the perturbation's propagation direction and the perturbation is transverse.

Since the Equation 1.5 is a wave equation, the immediate solution is a plane wave:

$$h^{TT}_{\mu\nu} = C_{\mu\nu}e^{ik_\sigma x^\sigma} \tag{1.6}$$

where $C_{\mu\nu}$ is a constant and symmetric tensor, traceless and purely spatial, and k_μ is the wave vector. Here the solution is for a single Fourier mode only, however a complete general solution will be a superposition of all possible modes. Note that since $h^{TT}_{\mu\nu}$ is real, only the real part of 1.6 must be considered. By imposing the equation of motion 1.5, the dispersion relation for the considered Fourier

mode is thus obtained:

$$k_\mu k^\mu = 0\,. \tag{1.7}$$

This condition states the wave vector must be null, meaning that the GW propagates at the speed of light. Additionally, exploiting the transversality condition leads to

$$k_\mu C^{\mu\nu} = 0 \tag{1.8}$$

meaning that the wave vector is ortohogonal to the $C_{\mu\nu}$ tensor. To better understand the physical content of a passing GW, by choosing spatial coordinates such that the wave is travelling along z-axis, an explicit form for the $C_{\mu\nu}$ can be obtained, recalling also that:

- $k^\mu = (\omega, 0, 0, k^3) = (\omega, 0, 0, \omega)$ because $k_{\mu\nu}$ is null,
- $C_{0\nu} = 0$ and $k_\mu C^{\mu\nu} = 0$ imply that $C_{3\nu} = 0$,
- $C_{\mu\nu}$ is traceless and symmetric.

Therefore in general

$$C_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & C_{11} & C_{12} & 0 \\ 0 & C_{12} & -C_{11} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.9}$$

It is then clear that, in the TT-gauge, after getting rid of non-physical redundancies in the description, the GW solution is completely characterized only by two degrees of freedom and by the frequency ω. The measurable effects occurr only in the plane orthogonal with respect to the propagation direction, in this case the x-y plane. To get some insight about the spacetime distorsions induced by a passing GW, it is convenient to start with rewriting the solution as follows:

$$h^{TT}_{\mu\nu}(t) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} e^{i\omega(z-t)} = (C^+_{\mu\nu} + C^\times_{\mu\nu}) e^{i\omega(z-t)} = h^+_{\mu\nu}(t) + h^\times_{\mu\nu}(t)\,, \tag{1.10}$$

where $+$ and $\times$ denote the two linear polarization states of the GW, respectively *plus* and *cross*, and in general a GW is a superposition of the two. Then, considering a ring of massive stationary particles in spacetime, $h^+_{\mu\nu}$ compresses it in one direction while stretching it in the other, as it is clear from Figure 1.1. Meanwhile $h^\times_{\mu\nu}$ produces the same distortion as $h^+_{\mu\nu}$ rotated by 45°. It is important to note that $h_{\mu\nu}$, called also GW *strain*, represents a relative change in the proper distance between the particles over time. The particles' coordinates remain indeed fixed, since TT coordinates stretch themselves to follow the GW. However, their proper distance ds^2 change over time while the GW passes and it includes the coordinates difference. Moreover, there is no "center" of distortion. Particles can only move closer or farther away; a force does not act on a specific point to cause distortion. The effects of the change in spacetime curvature highlight how curvature is an intrinsic characteristic of spacetime.

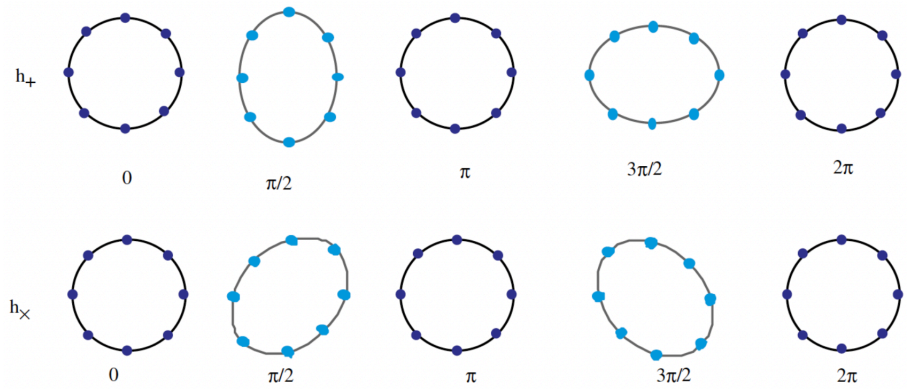


Figure 1.1: Deformation of a ring of test particles at rest on the x-y plane due to plus (top) and cross (bottom) polarization states of a GW propagating along the z-axis. From left to right, the phase of the wave increases as time passes highlighting the deformation patterns induced by the crossing GW.

The TT-gauge solution to Einstein's equation in a vacuum allows the effects of a passing GW through a system of massive particles to be seen concretely. This is essential to understanding GWs detection later on because it clearly points out that the core effect of a GW is the change in relative distances over time. However, this solution is only valid in vacuum, meaning far enough away from massive bodies, and up to linear order in the perturbation. Additionally, since linearity implies the superposition principle, stating that the fields of two particles can be summed with no effect on the particles themselves, the linearized theory predicts no effect of gravity on matter. Linearized theory also disregards the energy-momentum of the gravitational field as a source of the gravitational field itself. Neglecting nonlinearity means overlooking the focal point of General Relativity, "gravity gravitates".

1.2 Gravitational waves sources

The linearized theory can be useful also to find a general solution of linearized equations (1.4), taking into account $T_{\mu\nu}$ too. Since the equations are linear, an expression for the GW strain can be found through a Green's function, representing the solution in the case of a delta-like source:

$$\Box_x G(x-y) = \delta^4(x-y)\,. \tag{1.11}$$

Note that the Green's function depends only on the coordinate difference: x and y are four-vectors representing respectively the coordinate at which the solution is evaluated and the source coordinate. Any generic source can be therefore built out of an infinite sum of deltas using the superposition principle[2] and recalling that the background spacetime is flat. Therefore the integral is evaluated simply on flat spacetime:

$$\bar{h}_{\mu\nu}(x) = -\frac{16\pi G}{c^4}\int d^4y\, G(x-y)\, T_{\mu\nu}(y)\,. \tag{1.12}$$

The Green's function should however account for the accumulated effect of signals to the past of the point under consideration only, and integrating over the time variable, meaning over the "history" of

[2]The solution of the homogeneous equation will be neglected from now on.

the source, it follows that

$$\bar{h}_{\mu\nu}(t,\mathbf{x}) = \frac{4G}{c^4}\int d^3y\,\frac{T_{\mu\nu}(t-|\mathbf{x}-\mathbf{y}|,\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}\,, \tag{1.13}$$

where $t-|\mathbf{x}-\mathbf{y}| \equiv t_r$ indicates the *retarded time*, ensuring that the perturbation in the spacetime curvature at $(t,\mathbf{x})$ comes from the energy and momentum sources from the past light cone. The general GW solution to the linearized field equations can be expressed also in TT gauge and is valid outside the source only, because of the weak-field limit. At distances much bigger than the spatial extension of the source the GW strain is indeed inversely proportional to the distance from the source[3]. A multipole expansion for the spatial coordiate difference $r = |\mathbf{x}-\mathbf{y}|$, considering non-relativistic sources[4], shows indeed that the spatial part of the strain can be expressed through the *quadrupole formula*:

$$h_{ij}^{TT}(t,\mathbf{x}) = \frac{2G}{c^4}\frac{\ddot{Q}_{ij}^{TT}(t_r)}{r}\,, \tag{1.14}$$

where $\ddot{Q}_{ij}^{TT}$ is the second derivative of the traceless (spatial) quadrupole moment tensor of the energy density of the source, defined as

$$Q_{ij}(t) = \int d^3y\,T^{00}\,y_i\,y_j\,, \tag{1.15}$$

where T^{00}, the total energy density, is dominated by the mass density for non-relativistic sources in the weak-field regime. The quadrupole moment Q_{ij} tracks the spatial shape of the source and is a constant tensor on each surface of constant time.

The linearized theory that accounts for the source term too is key to understanding that changes in the mass shape distribution only produce GWs: a spherical body or system does not emit gravitational radiation. The GW strain produced by an isolated non-relativistic source is therefore proportional to the second derivative of the quadrupole moment of the energy density at the point where the past light cone of the observer intersects the source. The leading contribution to gravitational radiation comes from the change in the quadrupole moment of the mass density, in contrast with the electromagnetic case in which the leading contribution comes from the change in the dipole moment of the charge density. A change in the dipole moment corresponds to a motion of the center of density, but motion of the center of mass of an isolated system would violate conservation of momentum. Also the monopole term is not included, expressing mass conservation. However, the linear theory neglects back-action: predicting mass and momentum conservation during GW emission means that the gravitiational radiation emitted by the system has no effect on the system itself. General Relativity predicts energy loss through gravitational radiation and GWs actually carry away energy and momentum, as has been extensively observed[5]. This limitation highlights that self-gravitating systems cannot be described in the weak-field regime.

Concretely, GW sources should be massive and with significant bulk motions. The simplest example are binary system of compact bodies, such as black holes and neutron stars. Two compact bodies can orbit around their center of mass and the system emits GWs because there is no spher-

[3]This can be compared to the classical Newton's law for universal gravitation, stating that the gravitational force between two bodies is inversely proportional to the square of the distance between them.

[4]Non-relativistic is referred to the internal motion of the source.

[5]Hulse and Taylor (1974) discovered a binary star system where one component is a pulsar and observed changes in the orbital period that are consistent with energy loss due to GWs emission.

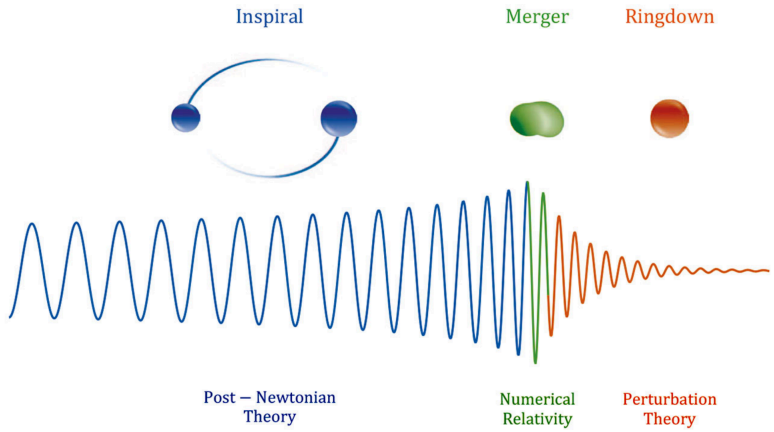


Figure 1.2: Simplified scheme of a CBC highlighting the main phases of the astrophysical process along with the typical waveform of a GW emitted during the evolution and the tools to describe it.

ical symmetry in the orbital motion. GWs emission comes at the expense of potential energy: the two objects spiral towards each other, speeding up to higher and higher angular frequencies. The inspiral phase continues until the two bodies coalesce into a single compact object. After the merger, the newly formed object vibrates, but these oscillations are dumped by GW emission. The main phases of a *compact binary coalescence* (CBC) are illustrated in Figure 1.2. The typical waveform from a CBC increases both in frequency and amplitude during the inspiral until the merger, then during the ringdown it shows dumped oscillations. Predicting the waveform is not trivial and requires many mathematical tools, such as *Post-Newtonian* (PT) theory, an approximation method for solving Einstein's field equations considering weak gravitational field and slow-motion systems, numerical solutions of Einstein's equations, useful in strong relativistic regimes, and black hole perturbation theory to analyze the remnant's behavior during ringdown. CBCs can involve a binary system composed by two black holes (BBH), a black hole and a neutron star (BH-NS) or two neutron stars (BNS).

Other examples of GW sources are continuous waves, which are long-lasting and nearly constant signals predicted to be emitted by rapidly spinning neutron stars with tiny surface bumps or irregularities. Core collapse supernovae (CCSN) are also expected to emit GW signals, but there are no clear models for this process, which consists of the violent explosion of a massive star following the rapid collapse of its core and is predicted to emit bursts of GW signals. Additionally, processes in the early universe are theoretically capable of generating GWs, but they are not directly related to the concept of asymmetry in mass distribution. Modelling GWs in this case is extremely difficult, particularly due to incomplete theories.

1.3 Gravitational waves detection

In order to detect and measure GWs, the starting point is recalling the physical effect of a passing GW: a change in the relative distance between freely-falling masses. The key ingredients for a GW detectors are therefore freely-falling masses and a way to measure their relative distance. Light travel time is quite a simple and precise tool to measure distances, meanwhile mirrors can both reflect light and be used as test masses. This is precisely the basic scheme of an *optical interferometer*. Ground-based laser-interferometric detectors are the currently active and performing GW detectors: they are

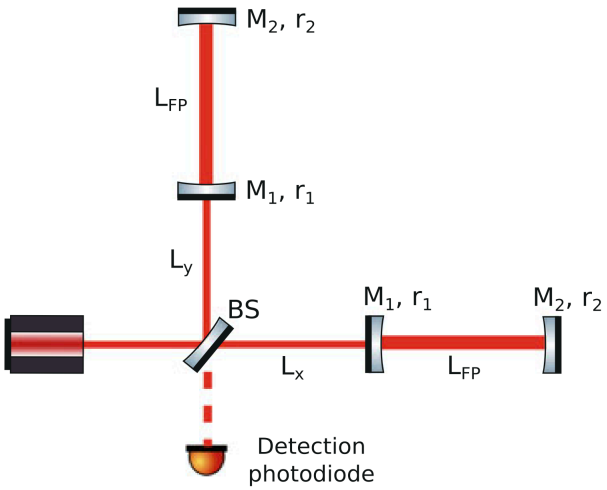


Figure 1.3: Simplified scheme of a GW optical interferometer. The laser beam is split along the arms (respectively along the x-axis and y-axis of a Cartesian reference frame) by a beam splitter (BS), then the light reaches the mirrors at the ends of the arms (marked as M_2). After many reflections within the cavities, the beam is recombined, and a photodiode collects the light. The two pairs of dielectric mirrors M_1 and M_2 make up two Fabry-Perot (FP) resonant optical cavities that enhance the effect of the passing GWs.

modified Michelson interferometers and detect GWs relying on the fact that the distance, and therefore the light travel time, between two mirrors changes as the GW crosses the detector.

Building a feasible ground-based laser-interferometric detector is a great challenge from various perspectives. First, it is necessary to understand the amount of relative change in distance that needs to be measured. For the current purposes, the detailed calculations can be left to one side. Instead, only rough estimates will be provided to give an idea of the required scale for the measurement. The relative change in the distance between two test masses is approximately proportional to the GW strain amplitude. For instance, considering a binary system in our galaxy composed of two stars with the same mass $M \sim 10\, M_\odot$ and orbital radius $R \sim 10^5\,\mathrm{m}$, the frequency of the emitted GW is approximately $f \sim 100\,\mathrm{Hz}$, consequently

$$\frac{\Delta L}{L} \sim h \sim 10^{-21}\,, \tag{1.16}$$

where L is the distance between the two test masses in the detector. If L is in the order of kilometers, this GW strain amplitude would require a sensitivity to changes in the differential distance $\Delta L \sim 10^{-18}\,\mathrm{m}$. It is much smaller than the size of the constituent atoms out of which any test masses would have to be made, since the typical size of a nucleus is $10^{-15}\,\mathrm{m}$. Essentially, sensitivity is the challenge at the core of GWs detection. However, from this argument it is also clear that the distance between the two test masses in a ground-based detector, that is in the order of kilometers, is strictly related to the frequency of the GW signal that can be detected. Indeed, ground-based GW detectors can observe signal only in a restricted frequency range, that is approximately from above 10Hz to 1kHz. This frequency band corresponds to GW signals typically emitted for example by CBCs.

More in detail, to face the challenge of GWs detection, optical interferometers (Figure 1.3) consist of two arms orthogonal to each other, with two mirrors hanging at either end. Light from a monochromatic laser is injected into the interferometer, and through a beam splitter half is reflected and goes into one arm, while the other half is transmitted and goes on the other. The light then travels in the

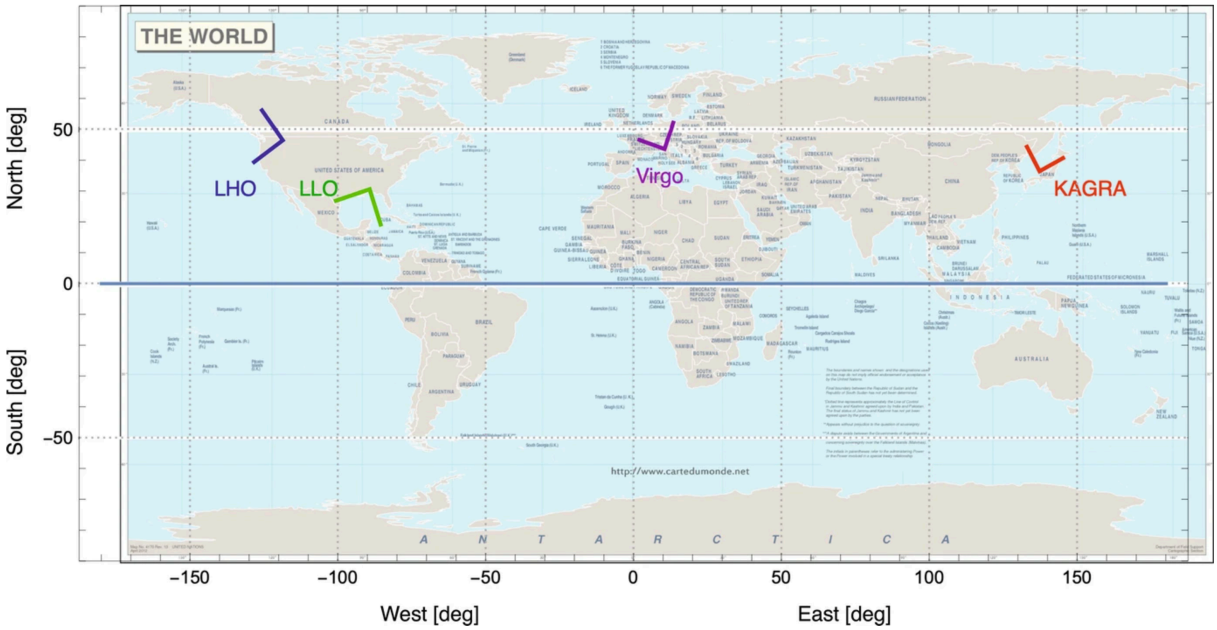


Figure 1.4: Localization and orientation of the currently operating GW interferometers [8]. They are LIGO Hanford Observatory (LHO, H1) and LIGO Livingston Observatory (LLO, L1) in the United States, Virgo in Italy and KAGRA in Japan.

resonant optical cavities of each arm and reflects on the mirrors a very high number of times before the two beams coming from the two arms rejoin. The system is arranged such that, if the test masses are perfectly stationary, meaning no GW is passing, the returning beams interfere destructively, sending no signal to the detection photodiode. More specifically, the optical cavities are called *Fabry-Perot* cavities and are resonant, meaning that the reflected light returns in phase and add coherently. This aspect of cavities is key for detecting GWs: a GW crossing the detector perturbs arm lengths in opposite senses, consequently leading to a phase shift in the beam reaching the photodiode that will break the destructive interference. Thanks to many round trips through the cavity arms, the effective length of the arm increases, the phase shift accumulates and becomes measurable by the photodiode. The detection photodiode indeed measures the interfering beam power, which depends on the phase shift, which, in turn, contains physical information about the passing GW. From the phase shift, the differential length change in the light path in the two arms can be calculated and consequently the GW strain amplitude can be extracted.

Nowadays, the LIGO-Virgo-KAGRA (LVK) scientific collaboration includes the three active interferometers, operating in unison as a single observatory (Figure 1.4), although during a scientific observation period, it is possible that one or more interferometers may not be fully operational for a variable period of time due to technical issues or necessary upgrades. The interferometric network is now composed of four detectors. LIGO, Laser Interferometer Gravitational-wave Observatory, refers to two detectors located in the United States, specifically in Hanford, Washington, named *H1*, and in Livingston, Louisiana, named *L1*. They are separated by more than 3000 km, they basically share the same design and structure, with the arm length being 4 km, and they started operating in early 2000s. Virgo is an interferometer located near Pisa, Italy: the arm length is 3 km and it started observing in early 2000s. The other now active detector is KAGRA, Japan: it is the only one located underground, while other interferometers lie on the Earth's surface, the arm length is 3 km and it started operating in February 2020. Multiple interferometers are required because, in order to establish detection

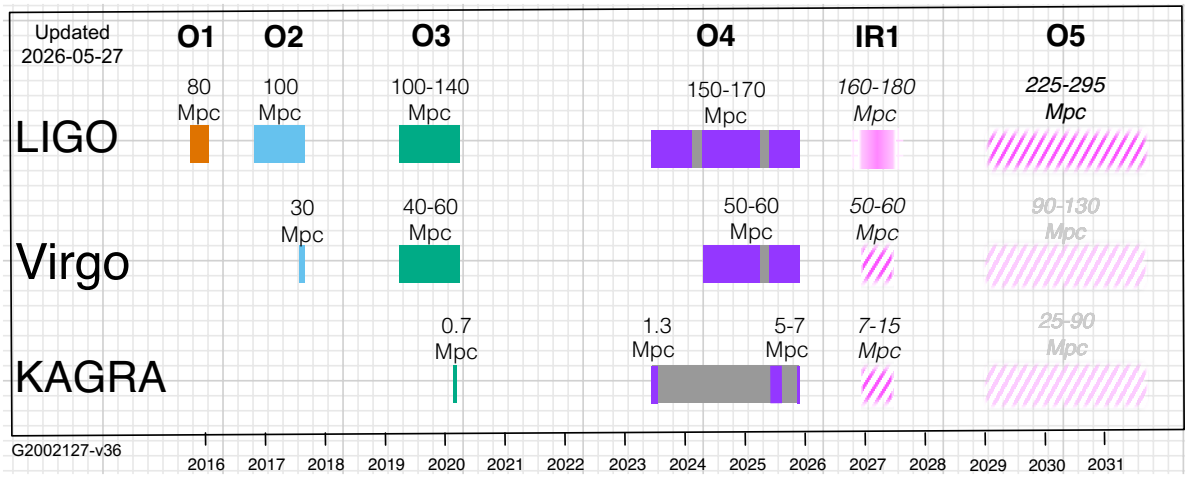


Figure 1.5: Observation runs plan, updated May 2026, from public updates of LVK planning. For each run the target sensitivity is specified in terms of BNS mergers range for a single-detector, since these signals are a well-studied case. Both LIGO detectors will be observing in IR1, and Virgo and KAGRA will join as they are available, accordingly to their upgrade plans. Plans and timeline for the fifth observing run (O5) are being reassessed.

confidence, observations in more than one detector are required, since the sensitivity of a detector is so small that even the slightest vibration can mimic a GW signal [8], and only observations in coincidence can be taken into account. The separation of the detectors is essential for confirming a GW detection. Any signal that appears in more than one detector with a time separation greater or smaller than the light travel time could not have been caused by passing GWs. Moreover, there is no possible shield against the effect of GWs, so it is not feasible to precisely estimate the background for an interferometer. Lastly, multiple interferometers are required to localize the source in the sky, and to measure the polarization of GW signals. At least three detectors are necessary for triangulating the signal and uniquely identifying the source localization, comprising direction, orientation and distance.

The scientific observations of the detector networks are organized in runs interspersed with down time for upgrade and commissioning, and transitional engineering runs. Commissioning refers to the set of activities required to prepare the detector for scientific observations, including achieving the specified sensitivity and the operational stability necessary for the observations. After commissioning, engineering runs aim to intensively test the whole functioning of the detector, from instrumentations to data analysis pipeline, before switching to full observing mode. Figure 1.5 illustrates the observing schedule since the first observational run, named O1, during which the first detection occurred in September 2015. Only the two LIGO detectors participated in O1, Virgo joined only at the end of the second observational run, O2, after major upgrades. KAGRA actively joined the network at the end of O3. The O3 run is divided into two parts, named respectively O3a and O3b. The fourth observing run is divided into three parts and Virgo joined only from O4b. The next upcoming run is the first Intermediate Run (IR1), which aimes to provide a bridge between O4 and O5 maintaining data collection and test upgrades.

During the observation periods, the GW detectors network has been collecting an increasing number of detections, all of which being signals from CBCs, mainly from BBH mergers. GW170817 (named after its detection date during O2) was the only confirmed BNS with a detected EM couterpart. Up to O4a, the confidentially detected GW events are in total 218 [11], and more than half of them were

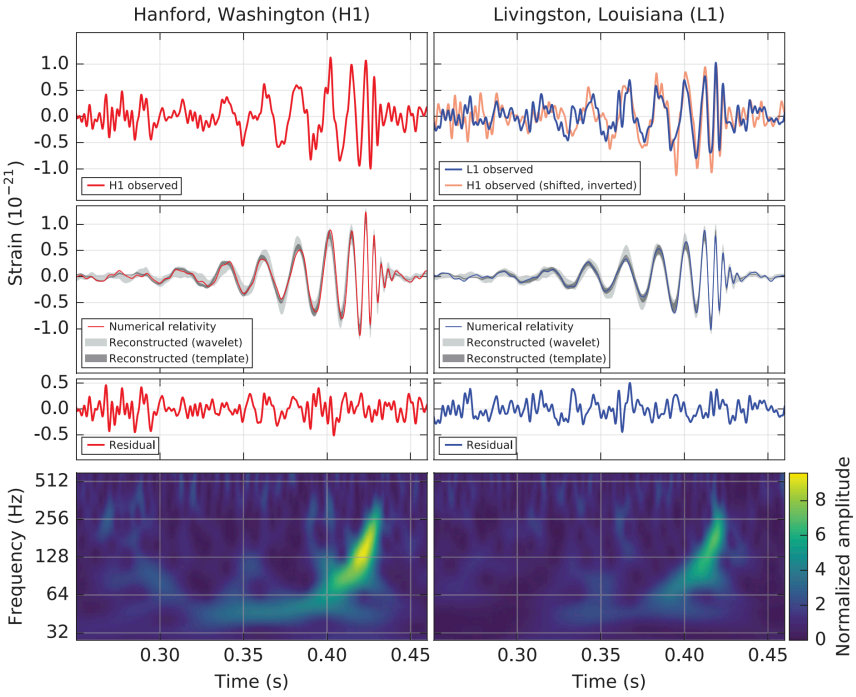


Figure 1.6: GW150914, the first detection [3]. Top: time series for the detector's strain, both for H1 and L1. The GW signal arrived first at L1, then at H1, but here the H1 data are shifted in time and inverted, accounting for the detectors' relative orientations. Middle: numerical relativity waveform for a system with parameters consistent with those recovered from the detected event and residuals after subtracting it from the detector time series; shaded areas show 90% credible regions for two independent waveform reconstruction, one (dark grey) from an astrophysical BBH model and the other (light rey) built through a linear comination of sine-Gaussian wavelets. Bottom: time-frequency representation of the strain data.

discovered only during O4a. The rest of the data collected from the last run is still being analyzed, but the number of detected events continues to increase with each run thanks to numerous upgrades in instrument sensitivity. In Figure 1.6 is depicted the first GW event confidentially detected by LIGO interferometers on 15th September 2014, consequently named GW150914 [3]. The extracted GW strain is here visualized both as a time series (first row) and through a spectrogram (bottom row). The easiest way to observe the detector's output is indeed a *time-frequency representation*, allowing to observe how the GW signal evolves both in time and frequency. The spectrograms clearly shows the frequency increase over time typical of a CBC, called *chirp*. The corresponding astrophysical event is a BBH merger approximately 410 Mpc far from Earth. The masses of the two starting BHs were approximately $m_1 = 36\,\mathrm{M}_\odot$ and $m_1 = 29\,\mathrm{M}_\odot$, the energy radiated in the form of GWs during the process was around $3\,\mathrm{M}_\odot c^2$ and the mass of the BH resulted from the merger is about $62\,\mathrm{M}_\odot$.

1.3.1 Noise in optical interferometers

To achieve the required sensitivity, the detection system must be isolated from the external world since any vibration could cause disturbances. For example, the mirrors must be carefully suspended to be as close as possible to the freely-falling condition, as well as optical benches; also, laser beams need to travel in vacuum to avoid air-induced disturbances. Providing the full isolation and control system that allows the detector to operate at the required sensitivity is a huge technological challenge.

Although many strategies have been developed to shield GW detectors, detection remains strongly noise-dominated, meaning signals are buried in a huge amount of noise. Therefore, it is essential to

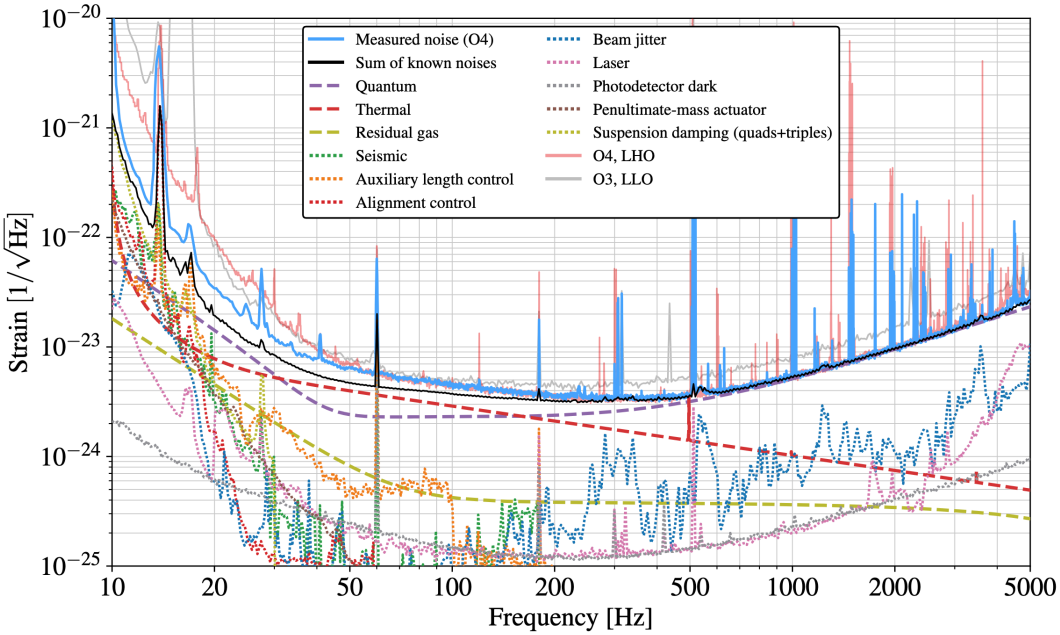


Figure 1.7: Noise budget for LLO at the start of O4 scientific run expressed as the amplitude spectral densitiy of the strain noise [10]. Dashed lines represent the predicted noise from models or measurements from detector components. Dotted lines indicates noise measurements from auxiliary channels. At low frequency there is a discrepancy between the measured total noise and the expected one and the dominating noise source is electronic control noise. Between 50Hz and 500Hz quantum and thermal noise are dominant, meanwhile at high frequencies shot noise is prevailing. The noise budget of LLO is very similar to the LHO one for the start of O4 run.

study how noise affects detection from both a detector characterisation perspective and a data analysis point of view. Detector characterisation involves studies of instrument performance compared to predictions, tracking noise sources, and validating data quality, since environmental or instrumental artifacts can hinder detection. Data quality assessments are indeed crucial for subsequent data analysis that will extract astrphysical information. In particular noise budget (Figure 1.7), a quantitative analysis of the various sources of noise that affect the detectors' sensitivity, is a fundamental tool that allows to investigate how different know noise sources contribute to the total noise amount since it allows the comparison between measured detector sensitivity and incoherent sum of all known noise contributions. Noise can therefore partially be modelled and consequently subtracted, also thanks to additional sensors to monitor specific disturbances and auxiliary channels (that comprise environmental sensor and instrument controls). However there are environmental or instrumental artifacts, also due to couplings between the different subsystems that compose the whole detector, that cannot be predicted. These transient disturbances are called *glitches* and can spoil detection by reducing the sensitivity, mimicking an astrophysical signal or overlapping with it.

Noise sources

Considering the frequency range in which ground-based optical interferometers can listen to GW signals, at lower frequencies, above 10Hz, the main issue is to prevent the mirrors, as well as the optical benches, from moving. Meanwhile, at higher frequencies, the limit is the ability to measure small displacement caused by a signal instead of a high frequency noise source. The main noise sources are specified in details below.

- *Seismic* noise: due to geological, sea, seismic and anthropic activities, it affects mainly mirrors and optical benches, limiting the sensitivity below about 10Hz. Ground motion is of order 10^{-6}m and can mimic GW signals: it can be suppressed thanks to an active control system over the mirrors' and optical benches' motion and also with a passive strategy involving a cascade of inverse pendula, that composes the suspension system after the active stage.

- *Newtonian* noise: low frequency noise associated with gravity gradients acting directly on the mirrors, and due to stochastic fluctuations in the distribution of the masses surrounding the detector, that is the local gravitational field. This noise comprises also disturbances caused by variations in the atmospheric pressure. Newtonian noise is therefore not fully predictable and controllable, and in particular it cannot be shielded otherwise GWs would be shielded too. The adopted solution to mitigate it is a network of sensors to measure ground displacement and atmospheric pressure variations, that allows to model the effect of the measured disturbances on the mirrors and subtract from the output channel of the detector.

- *Thermal* noise: caused by random motion of atoms due to thermal excitation acting mainly in mirror suspensions and mirror coating because of the high power laser in the cavities. Thermal vibrations of the suspension fibers are typically higher than residual seismic noise and cause the mirrors to vibrate. Thermal fluctuations of the optical coatings result in coating Brownian noise, limiting the sensitivity around 10Hz. Possible solutions to thermal noise include using low-loss materials for mirrors and monolithic mirror suspensions that are essentially fused to the mirrors. Furthermore, the reflection phase of the laser beam is particularly affected by Brownian thermal noise in mirror coating. Solution options include new coating materials and new layer structures.

- *Quantum* noise: it is the sum of quantum radiation pressure noise and shot noise, that are essentially fluctuations due to the quantum nature of light limiting sensitivity mainly between 100Hz and 1.4kHz. Quantum radiation pressure is due to the high laser power since a fraction is transferred to the mirrors' mechanical modes. Shot noise refers to the fluctuations in the number of photons reaching the photodiode. The main implemented solution is quantum squeezing, consisting in the injection of a squeezed vacuum state (i.e. with no photons) to the output port of the interferometers. The choice of the vacuum state is because fluctuations of the electromagnetic field come from the vacuum state. A squeezed vacuum is a nonclassical state with asymmetric fluctuations, and the amount of fluctuations can be controlled. The considered fluctuations are respectively relative to the shot noise and to the quantum back action noise (disturbance caused in a quantum system by measuring it). In the last implementation in the interferometers for O4 run, squeezing is frequency-dependent, meaning that fluctuations are controlled depending on frequency, allowing to obtain a great improvement in sensitivity around 1kHz.

- *Gas* noise: residual gas particles in the cavities interact with the laser resulting in thermal motion of the gas molecules that collide with mirrors exchanging momentum. Also, scattered photons by the gas molecules modulates the optical phase of the beam, that is the core measured quantity in the interferometers.

- *Scattered light* noise: due to mirror imperfections that cause a fraction of the laser light to scatter around the cavities and when it mixes when the main beam it can create a spurious signal, i.e. a glitch. Other minor sources of scattered light are optical baffles too, that in principle should absorb and mitigate stray laser light reflecting off the walls of the vacuum tubes. However stray light can bounce off the vibrating baffles and scatter back into the main laser beam.

- *Electronic* noise: generated by the analog and digital electronics used for detection, control and noise reduction, like for example electronic sensors monitoring the mirrors' displacement that inject their noise back into the strain data channel of the interferometer in the low frequency band.

Additional to noise sources, since the detectors are built on the Earth's surface, they are susceptible to transient weather conditions, especially variations in external temperatures and wind. Finally, many instrumental subsystems couple into the strain channel, producing narrow spectral lines that can mimic or obscure continuous GW signals.

Detector sensitivity

The noise in the optical interferometers is approximately stationary over shotìrt time scales, and therefore its characterisation in the frequency domain is most straightforward [1]. The sensitivity for a GW interferometer is therefore represented in the frequency domain by the amplitude spectral density (ASD) of the background noise relative displacement or strain, that is the square root of the correspondent power spectral density (PSD). The PSD provides a statistical description of how the noise power is distributed across frequencies and for a real[6] signal $n(t)$ is defined as follows:

$$S(f) = \lim_{T\to+\infty} \frac{2}{T} \left| \int_0^{\frac{T}{2}} dt\, n(t)\, e^{-2\pi i f t} \right|^2 , \tag{1.17}$$

where the noise time series is taken over a time interval T. The PSD unit of measure is 1/Hz and the PSD is the Fourier transform of the time-domain autocorrelation function.

It is common practice to refer to detector sensitivities also in terms of the range over which a typical BNS can be detected. This provides a more practical, observation-related way of comparing sensitivities. This range is computed assuming an homogeneous BNS population and averaging over the volume and the orientation [12]. BNS are chosen as the reference signal because it is a well known phenomenon that allows for a simple calculation in terms of detector sensitivity to CBCs. Figure 1.8 shows sensitivity curves for the two LIGO detectors and for Virgo during O3 and the first and the second part of O4 runs: LLO and LHO clearly improved their sensitivities and BNS ranges thanks to various successful upgrades, especially frequency-dependent squeezing; meanwhile Virgo faced some obstacles during the upgrade phase and could not take part in O4a. In the sunburst charts are also displayed the fractions of the run duration in which the correspondent detectors were actually operating. During a scientific run an interferometer can indeed face some technical problems preventing it from operating correctly.

[6]For a real signal it holds $S(f) = S(-f)$ and the PSD is consequently called *single-sided*.

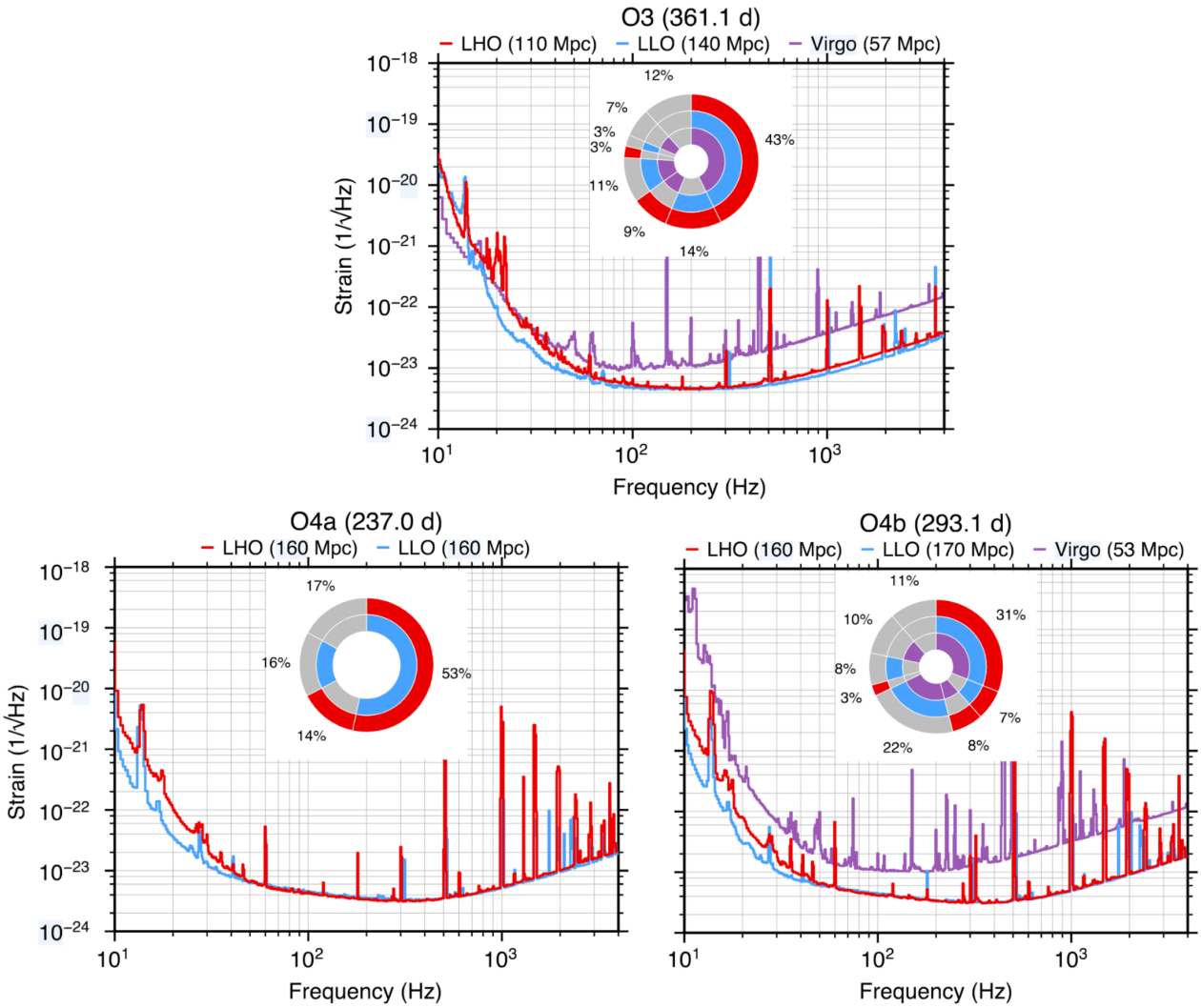


Figure 1.8: Amplitude spectral densities of strain noise for LLO, LHO and Virgo during O3 (first row) and first and second part of O4 runs (second row). The run name is accompanied by the whole run duration expressed in days. BNS ranges are specified too and in the sunburst charts there are fractions of the run duration during which different combinations of detectors were observing (clockwise, all the possible combinations, grey shows that a detector is not actually operating according to requirements) [12].

Glitches

As anticipated before, glitches are transient disturbances that can hide or mimic a GW signal, but in most cases, their causes remain unknown. Their duration is usually comparable with GW signals duration and their occurrance rate can vary through time and typically increases if the detector sensitivity improves. Glitches are in particular non Gaussian and non stationary, and they reduce the detector's sensitivity, becoming data quality's main issue. They can appear in many morphologies with different duration, amplitude and frequency. Glitches occur frequently enough that they often can be coincident in also two detectors and can mimic astrophysical signals [37], also, glitches are dangerous because they can cause data loss, especially when they overlap with a signal.

In Figure 1.10 it is shown the rate of glitches during the first and the second part of the O3 scientific run. The glitch rate is approximately slightly less than 1 glitch per minute, but depends on the specific detector and can vary during an observing run due for example to weather conditions related to the seasons. It is increased with respect to O2 in both LIGO detectors where the main glitch

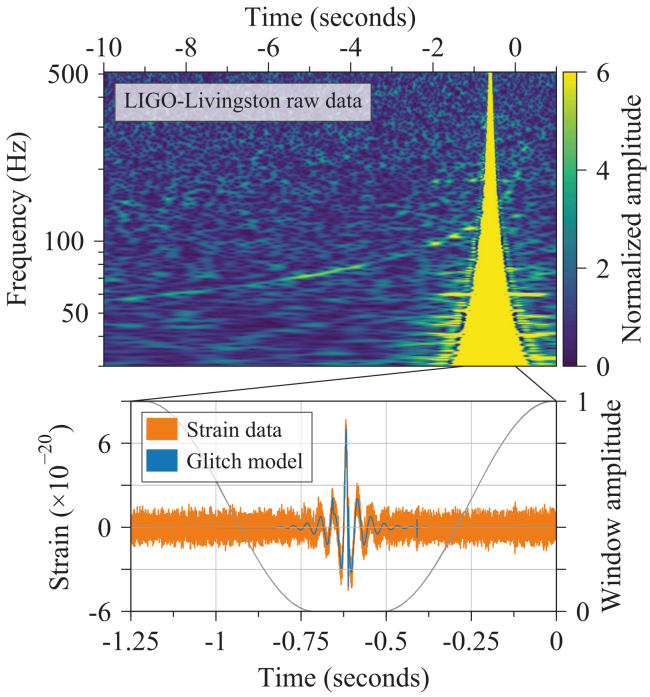


Figure 1.9: Glitch appeared in LLO during GW170817 (top, raw data from strain channel) and its mitigation (bottom) [2]. On the top, it is shown a time-frequency representation of the glitch overlapping with the BNS signal just before the coalescence time. Below the plot shows the glitch mitigation strategy in time domain: for a first analysis, the raw strain data were multiplied by an inverse Tukey window (right axis) that zeroed out the data around the glitch. Meanwhile, to extract astrophysical parameters an approximated model for the glitch has been subtracted from the data. The amplitude of the glitch in this case clearly buries the signal, which amplitude is instead of order 10^{-22}.

source was scattered light. Meanwhile in Virgo low frequency glitches from sea activity dominates, but the rate is lower compared to O2 thanks to many stability upgrades [4] [5]. Glitch rates therefore vary significantly across different interferometers and observing runs, and are generally expected to increase with sensitivity improvements. This makes it necessary to strike a balance between mitigating noise sources and stabilising the interferometer's control. Moreover, glitch morphologies can vary over time depending on instrumentation upgrades, and are specific to each detector. New morphologies can also appear and disappear, making glitches difficult to find and track over time. From this perspective, the need for continuous detector characterisation is even more clear.

The interplay between glitches and astrophysical signals became evident with GW170817 [2]: a short instrumental noise transient, caused by a brief saturation in a digital-to-analog converter, appeared in L1 detector just before the coalescence time of the signal. Because the glitch was short compared to the long, several-seconds-long BNS signal and was well separated from the merger itself, it could be identified and subtracted from the data [30]. This episode is illustrative rather than exceptional: a systematic study of the third observing run found that roughly a quarter of all GW candidates occurred close in time to a glitch [27], highlighting the practical relevance of glitch mitigation for an increasing fraction of detections as detector sensitivity improves.

Omicron

Omicron is the primary tool used within LVK collaboration to identify and characterize transient noise in the detector's data stream [33], both in the main strain channel $h(t)$ and in the auxiliary

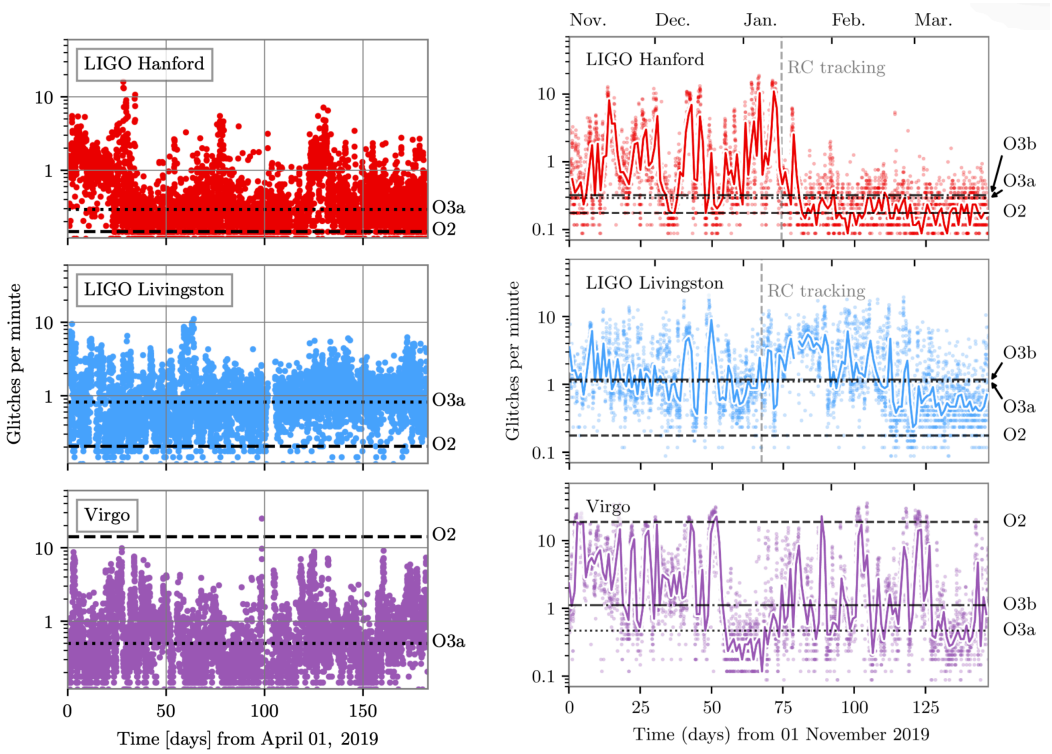


Figure 1.10: Glitch rate during O3a [4] and O3b [5]. Each point represents the average rate per minute estimated over a 2048-s interval and black lines show the median rate over entire runs. Only glitches with with SNR > 6.5 and frequency between 10 Hz and 2048 Hz identified by Omicron are included. On the left, for O4a, the increased rate of glitches in the LIGO detectors limited the overall sensitivity of searches for GW signals, meanwhile the glitch rate in Virgo decreased significantly thanks to various upgrades both in noise sources mitigation and in the control of the operational point of the detector. On the right, for O4b, continuous curves represent the daily median of the rates. The vertical dashed lines for LIGO interferometers corresponds to the implementation of reaction-chain (RC, that is parallel suspension system used to stabilize the detector's mirrors) tracking, which reduced the rate of scattered light glitches, especially for LHO.

channels essentially in real time. For example, in LIGO detectors, Omicron is run on the gravitational strain channel and on 800–900 auxiliary channels representing the subsystems, and results are stored in their on-site computing clusters [16]. Monitoring the detector this way is of paramount importance to identify correlations that can lead to understand coupling mechanisms between different subsystems that can generate glitches. When a detector is active, a continuous Omicron analysis is running over thousands of LIGO and Virgo channels with a low latency regardless whether the detector is operating correctly or not. The triggers are saved to disk a few seconds after detector's data are recorded along with some specifications characterizing the trigger itself.

Omicron is an algorithm designed to search for excess power in time series thanks to the Q-transform, a version of the standard Fourier transform in which the time window duration varies inversely with the frequency, allowing for a multi-resolution time-frequency analysis. The standard way to analyze signals in time-frequency domain would be the Short Time Fourier Transform (STFT), that applies a fixed window to the data and computes the Fourier trasform within it producing a uniform time-frequency grid. The STFT resolution is constrained by the window size: improving time resolution worsens the frequency resolution and vice versa, introducing some limitations in case of signas that vary across multiple time-frequency scales. The Q-transform is more flexible and sensitive because the time resolution is adapted to the frequency content since high frequency components are analyzed with short windows, while low-frequency components are analyzed with longer windows,

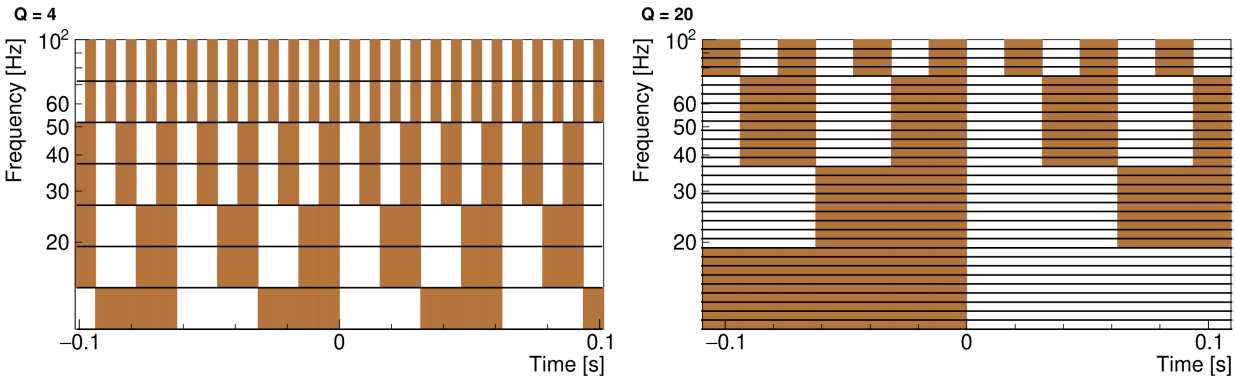


Figure 1.11: Examples of time–frequency planes highlighting the difference in resolution for two values of Q, $Q = 4$ (left) and $Q = 20$ (right) [33]. Each Q-plane is divided in tiles, the light and dark rectangles, along frequency rows. The shape of the tiles is mainly determined by the value of Q, allowing to analyze signals characterized by great variation in frequency over time in a more complete way, by appropriately weighting the components at different frequencies in order to better examine their evolution over time.

allowing for a multi-resolution signal analysis technique. In particular the time window duration is controlled by the quality factor Q as follows:

$$\sigma_t = \frac{Q}{\sqrt{8}\pi f}\,. \tag{1.18}$$

Consequently, following also the uncertainty principle, the time resolution increases (while the frequency resolution decreases) when the Q value decreases. Therefore the time-frequency decomposition of a time series $x(t)$ becomes

$$X(\tau, f, Q) = \int_{-\infty}^{+\infty} dt\, x(t)\, w(t-\tau, f, Q)\, e^{-2\pi i f t}\,, \tag{1.19}$$

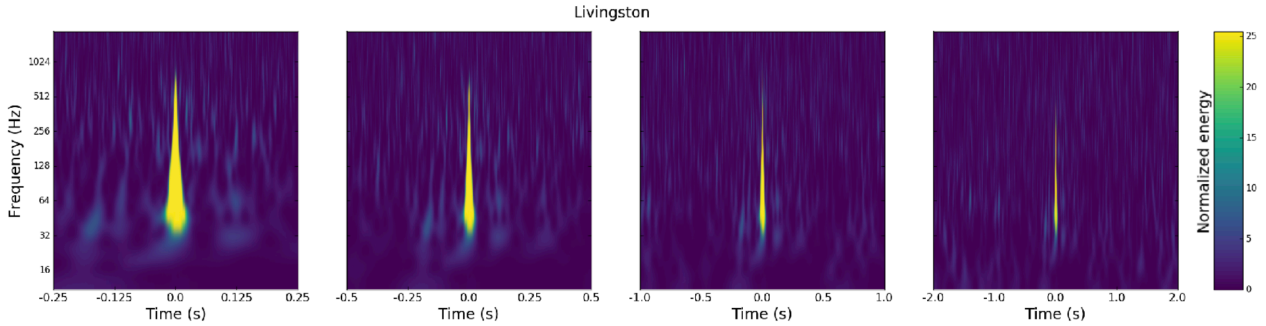
meaning that it is being projected a basis of sinusoidal Gaussian functions in the complex field, also called wavelets. The transformed time series therefore indicates the average signal amplitude and phase in a time-frequency region called *tile*, centered on time τ and with frequency f. The quality factor Q determines the shape and the area of each tile, along with the Gaussian window w, which width is given by σ_t. The Q-transform thus live in the parameter space given by (τ, f, Q) where Q-planes are logarithmically spaced.

The input time series $x(t)$ is first whitened (i.e. normalized in frequency) in order to obtain for the PSD a frequency independent expectation value that is equal to 2 in case of stationary noise. Consequently it is possible to associate to the Q-transform a signal-to-noise ratio (SNR) $\hat{\rho}$ as follows:

$$\hat{\rho}(\tau, f, Q) = \sqrt{|X^{th}(\tau, f, Q)|^2 - 2}\,, \tag{1.20}$$

with the obvious requirement that $|X^{th}(\tau, f, Q)|^2 - 2 \geq 0$, otherwise $\hat{\rho}(\tau, f, Q) = 0$. Consequently, to produce spectrograms, the Q planes are filled with the tile SNR. Then, tiles with a SNR above a previously established threshold are collected and clustered over time resulting in a so called *trigger*. A trigger is accompanied by some characterizing parameters, like time, frequency, SNR, and Q, that is given by the tile with the highest SNR in the cluster. Figure 1.11 shows the difference between different Q-planes, clearly indicating that the Q-transform is a multi-resolution technique in the time-frequency

(a) Blip glitch



(b) Whistle glitch

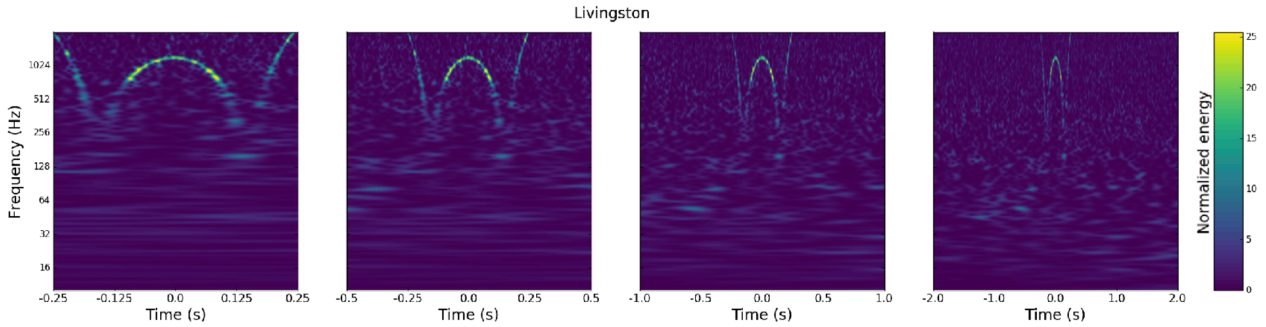


Figure 1.12: Omega scans, i.e. time-frequency-energy spectrograms, for different time resolutions, that allows to visualize the glitch morphology. Both machine learning and human classification needs all these time resolutions to help with both long and short duration glitches, since their shape could be similar but with a different time extension resulting in two different glitch classes. In the first row (a) we can observe a Blip glitch and in the second one (b) a Whistle glitch, both of them appeared in LLO during O1 [37].

domain.

Omicron software plays a key role in characterising transient noise associated with the validation of GW events. Omicron combines high processing speed, high detection efficiency and high resolution in order to reconstruct the parameters of transient noise. Omicron continuously processes thousands of channels from the interferometers. The resulting triggers provide a valuable starting point for noise studies. Omicron can help at many levels, from noise mitigation to improving data quality and characterising GW events.

Gravity Spy

Glitches are the main issue affecting the quality of the data. Due to their varying occurrence rates and morphologies, they present a significant challenge to detector characterization when it comes to studying them systematically. To achieve GW signal detection rates corresponding to the ones predicted by the interferometers' design sensitivity, a robust glitch characterization is necessary. This is precisely the goal of the Gravity Spy tool, that aim to find a flexible and adaptive trade-off between a purely manual approach by detector characterization expert, infeasible due to the huge amount of data to process, and a fully automated pipeline to classify glitches since it would require continuous updates to keep up with the pace at which the glitches landscape varies [37]. The Gravity Spy project merges citizen science with machine learning into a single glitch classification framework resulting in the first (and up to now unique) project to classify in a single dataset all the glitches occurred during

observation runs in LIGO and Virgo detectors.

Gravity Spy was first deployed in LIGO detectors right before the start of O2, and has been extended to Virgo afterwards. Citizens can volunteer to label time-frequency representations of glitches providing training data for machine learning classification algorithms that are then applied to real interferometer's data. Combining the two classification methods improves efficiency and accuracy compared to using only one of the two methods, helping thus researches to identify glitches and as a consequence, if possible, the related cause and remove it, improving therefore GW observations. Moreover, considering data quality, it is of great importance to associate data quality flags with specific categories of glitches. This could also help in identifying potential correlations with auxiliary channels and consequently make hypothesis on the glitch source.

Starting from Omicron triggers, a machine learning classifier predicts a known glitch class for each glitch sample. If the glitch class is unknow, human classification boost the pipeline potentially defining new glitch classes with wich the machine learning classifier is updated. Note that glitches with a SNR (given by Omicron) below 7.5 are neglected, otherwise they would be too difficult to classify by eye. The whole workflow results in a dataset where each glitch is assigned to a specific class. A glitch class is characterized by the characteristics of the glitch morphology, that usually corresponds to similar properties in frequency, amplitude, duration and SNR. Figure 1.12 shows two examples of glitches displayed by the ensemble of Omega scans, that represents the classification model input. Omega scans are essentially spectrograms in the time-frequency-energy domain, representing signals as combinations of sine-Gaussian functions. Each sine-Gaussian function is associated with a quality factor Q and Omega scans identify the Q value providing the loudest SNR for each time-frequency tile computed thanks to Q-transforms (Equation 1.20), and the normalized energy (corresponding to the color scale) is directly related to the SNR of a tile.

Up to O3, glitch classes used for classification are 23 [24] and are showed in Figure 1.13. The most common ones and those that require special attention are described in detail below (in alphabetical order).

- *Blip*: broadband glitches with very short duration, approximately $40ms$. In the time-frequency domain they exhibit a teardrop shape and their origin is yet unkonwn. Moreover, they can hinder searches for high-mass binary black hole GW signals.

- *Chirp*: in the time-frequency plane they resemble the characteristic signal of a CBC and originally included hardware injections.

- *Extremely loud*: they present a very wide broadband and a very high SNR, corresponding to the loudest disturbances appearing in the detectors and may be accompanied by a reduction in the astrophysical observational range since they also have a quite large duration.

- *Koi fish*: they are characterized by a large bandwidth and high SNR and could be misclassified as Extremely loud, but generally they present a shorter duration and reduced amplitude.

- *Low-frequency burst*: short duration (approximately 0.25s) glitches with quite large amplitude and frequency strictly peaked and constant slightly above 10Hz.

- *Low-frequency lines*: glitches with quite large amplitude, very low SNR and frequency strictly peaked and constant slightly above 10Hz, duration approximately 1.5s.
- *No glitch*: glitch class identifying Omicron triggers without a visible excess power in Omega scans. This class actually contains glitches that could be very different from each other, but all being quite difficult to resolve in the Omega scans.
- *Repeating blips*: blip glitches repeating with a period of approximately 0.250.50s.
- *Scattered Light*: long duration and low frequency glitches that are correlated with ground motion, it was one of the glitch types with the highest occurrence during O3 in all interferometers.
- *Scratchy*: glitches related to light scattering from baffles that consist in a series (about 10-30 per second) of sharp peaks at intermediate frequencies.
- *Tomte*: very similar to Blip glitches, but they exhibit a more triangular shape in the Omega scans and a lower frequency support. As for the Blips, they can influence the search for high-mass binary black hole signals.
- *Whistle*: these transients are typically U-, V- or W-shaped, with most of the noise content above 128Hz and most of the support above 500Hz. The most probable cause of these glitches are radio-frequency signals interacting with voltage-controlled oscillators.

If a glitch does not fit into any of the predefined categories, a user can classify it as "None of the Above" and will be asked to describe the morphology of the glitch in terms of its duration, frequency, and time evolution. By this process, new classes of glitches can be identified and integrated allowing Gravity Spy glitch classes to evolve and follow changes in the glitch types that occur in the detectors. "None of the Above" is therefore the category which includes all glitches that do not fit in the other classes.

1.4 Multimessenger astrophysics

Many of the astrophysical sources capable of producing detectable GW signals, particularly binary neutron star and neutron star–black hole mergers, are expected to have EM counterparts, potentially spanning a broad range of wavelengths. These events may also produce high-energy neutrinos and cosmic rays. Combining these different messengers for the same astrophysical event gives rise to the field of multimessenger astrophysics, allowing for the extraction of a far richer physical picture than any single messenger could provide on its own.

The GW170817 event, the first and only BNS merger confidently detected so far, observed on 17 August 2017 during O2, was the first event for multimessenger astrophysics. About two seconds after the merger time inferred from the GW signal, the Fermi[7] and INTEGRAL[8] satellites independently detected a short gamma-ray burst, consistent in sky location with the GW event [2]. This prompted an extensive, rapid electromagnetic follow-up campaign by observatories worldwide, guided by the sky localization region reconstructed from the GW data. This event therefore represents the first

[7]Fermi Gamma-ray Large Area Space Telescope
[8]INTErnational Gamma-Ray Astrophysics Laboratory

joint detection of gravitational and electromagnetic radiation from the same astrophysical source, and it illustrates why low latency analisys of GW signals is essential to multimessenger astrophysics: identifying the transient electromagnetic counterpart critically depends on promptly distributing a GW alert, together with an as-accurate-as-possible sky localization, so that telescopes can begin the search before the counterpart fades.

The joint observation of this single event yielded, among other results, an independent confirmation that short gamma-ray bursts can be produced by BNS mergers, and an independent measurement of the Hubble constant obtained by combining the GW-inferred luminosity distance with the electromagnetic redshift of the host galaxy [2]. More broadly, multimessenger observations of compact binary mergers offer a route to constrain the neutron star equation of state, to probe the physics of relativistic jets, and to test fundamental physics, such as the equivalence principle and the speed of gravity relative to the speed of light, by comparing the arrival times of the different messengers. In this perspective, low latency data quality assessments for GW strain data play a key role.

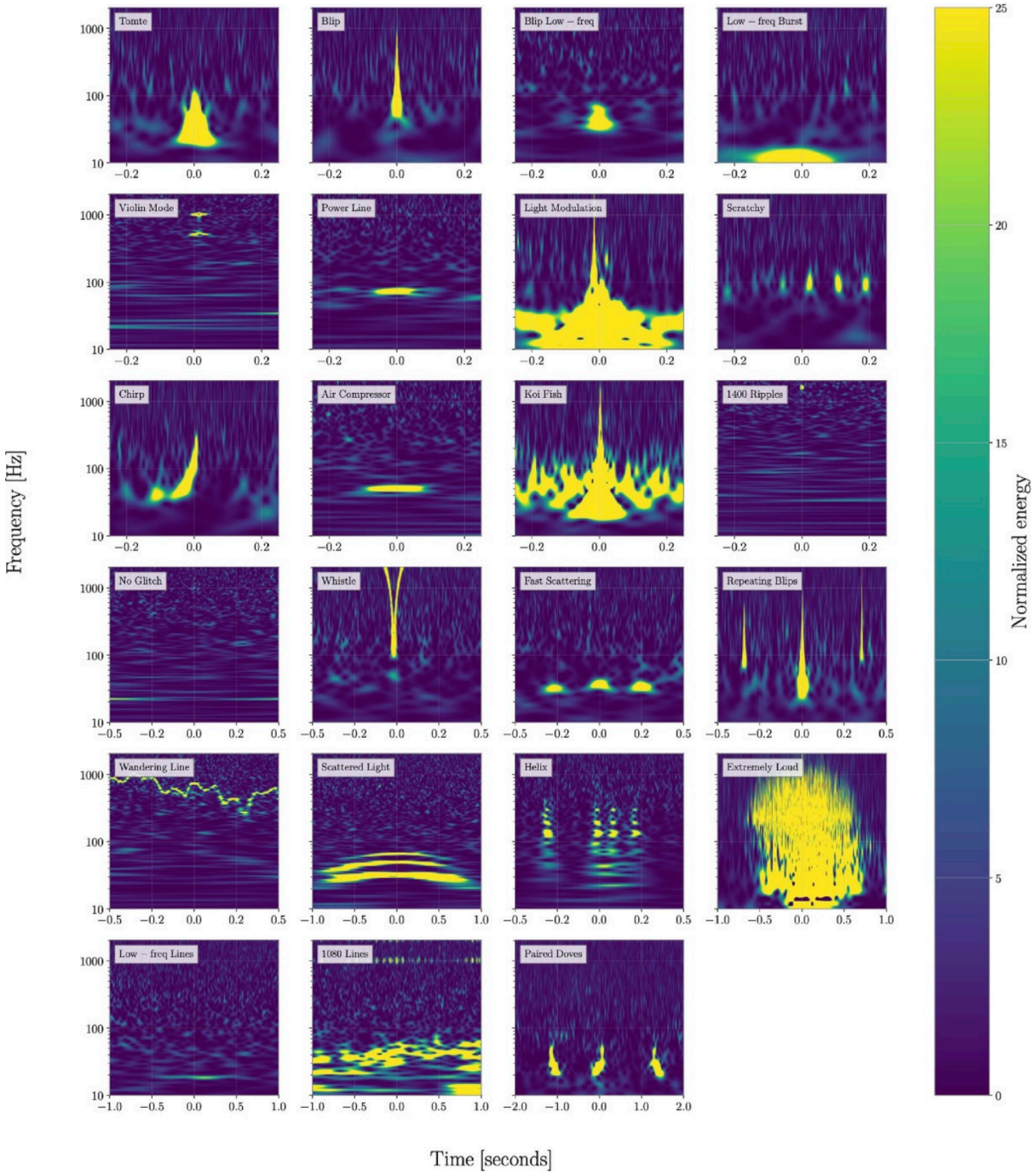


Figure 1.13: Omega scans for Gravity Spy classes up to O3 [24]. Classes are grouped by different time duration (0.5s, 1s, 2s, 4s) to best illustrate their morphologies. Glitches displayed here has been recorded in LIGO detectors during O3 run.

Chapter 2

Gravitational Waves Data Analysis

GW interferometers collect a huge amount of data, both from the main strain channel and from hundreds of thousands of auxiliary channels monitorning the different subsystems that compose the detector. Each detector produces indeed several PB of data per year, and only a few TB of them contain strain data [13], that needs to be managed and analyzed. The main goal is, of course, to find GW signals and extract astrophysical information from them, ideally as fast as possible to potentially generate multimessenger astrophysics follow ups as mentioned in the previous Chapter.

However, the path from data acquisition to the extraction of physically confident information requires to follow different steps. Data needs indeed to be first calibrated to produce actual time series, then data must pass quality assessments to verify that it is suitable for subsequent analyses aimed at extracting physical information. The data quality stage involves analyzing the detector status, primarily from auxiliary channels, to confirm that the interferometer is operating correctly and that the data is free of disruptive noise and safe for further analysis. Detector characterization studies plays a key role at this stage because they produce information about how the detector is working and, most importantly, the presence of glitches. Since a glitch can hide an actual signal or partially overlap with it, glitch mitigation takes care of neutralizing the presence of the glitch in order to minimize their adverse effects on data analyses. At the same time, information about glitches is fundamental to gather information on the detector status. After data quality, the search for GW signals can start: data needs to be preprocessed to be analyzed in frequency domain, then searches can be either based on waveforms already known or agnostic on the process that generated the signals. After a GW event is found it has to be validated by ensuring the search is confident the signal is not noise. Finally astrophysical parameters can be extracted by parameter estimation pipelines. The whole process actually takes place at two different paces: online/low-latency analyses are necessary to provide fast and roughly accurate information for a potential multimessenger follow-up, then a full and complete analysis can be performed offline.

Machine learning (ML) methods for GW data analysis are growing more and more, mostly within noise studies and parameter estimation. ML algorithms can indeed offer some key benefits: regarding investigation about the detector, they could predict and help in mitigating environmental disturbances in real time, ensuring detectors operate at peak performance [15], exploiting also the enormous amount of data coming from diagnostic channels. Meanwhile, on the parameter estimation side, ML

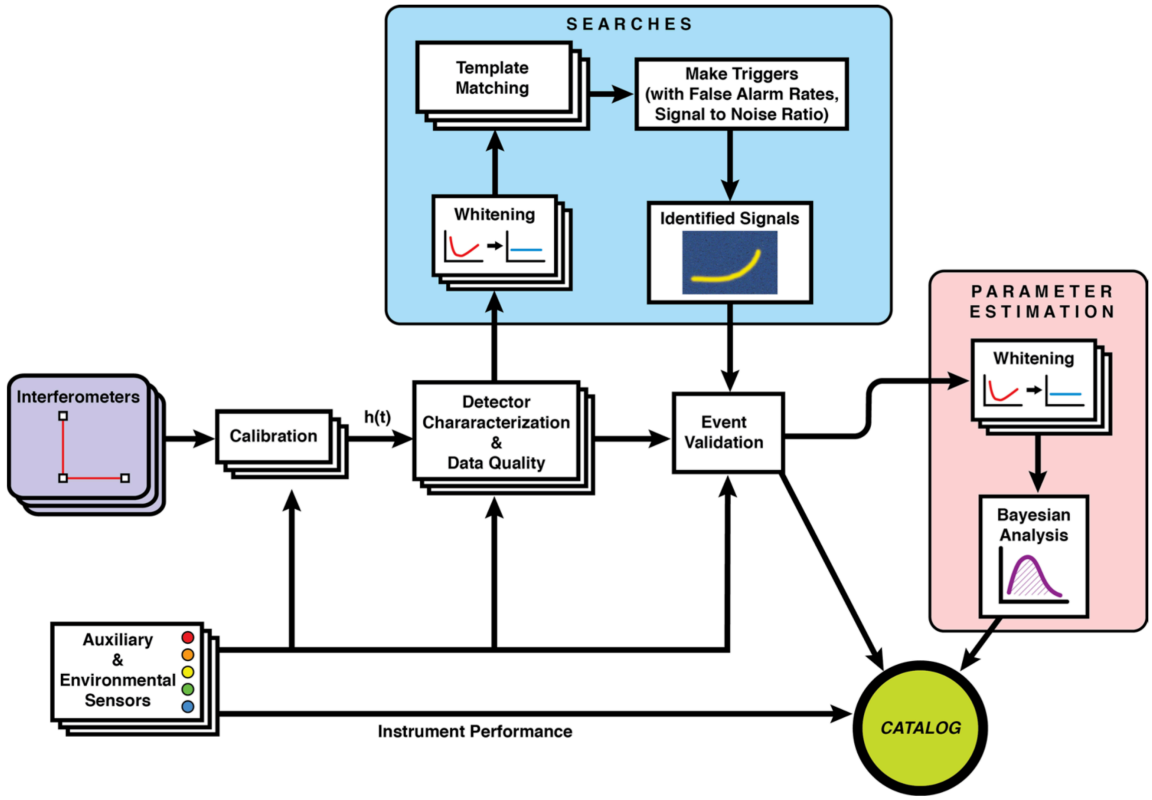


Figure 2.1: Simplified scheme of data processing system within LVK collaboration [1], from raw data produced by the interferometers to the formation of the GW transient catalog such as GWTC 5.0, the last published one [12].

models can provide a speed up in analyses and exploit the amount of data more efficiently than traditional methods. These advantages could be particularly beneficial for next generation gravitational interferometers such as LISA (Laser Interferometric Space Antenna, a space-based GW interferometer approved by the European Space Agency) and ET (Einstein Telescope, a proposed underground-based interferometer) since the amount of data will increase, the sensitivity will be higher and GW data analysis will deal with more complex challenges, requiring scalable methods that can operate at low latency. Some ML tools have already become the state of the art, as seen for Gravity Spy in the previous Chapter. However, in many areas traditional methods remain dominant since ML algorithms needs to be robust in order to produce scientific results and ML interpretability is not trivial. Efforts in interpretability and robustness of ML methods in GW data analysis are needed more than ever, particularly with regards to the introduction of next-generation detectors in the coming years.

2.1 Gravitational waves data

The first step after the electronic acquisition system has collected the raw data, for example from the photodiode that detects the main strain channel, is *data calibration*. This essential process translates raw electronic measurements into accurate physical units of relative displacement, i.e. gravitational strain, as time series. Strain calibration therefore requires models and measurements of all readout electronics, as well as the electronics and transfer functions of the hardware acting on the mirrors through multiple points in the suspension systems. Time series are typically sampled at 16384 Hz and calibration is valid above 10 Hz and below 5 kHz. Calibration errors and uncertainties arise

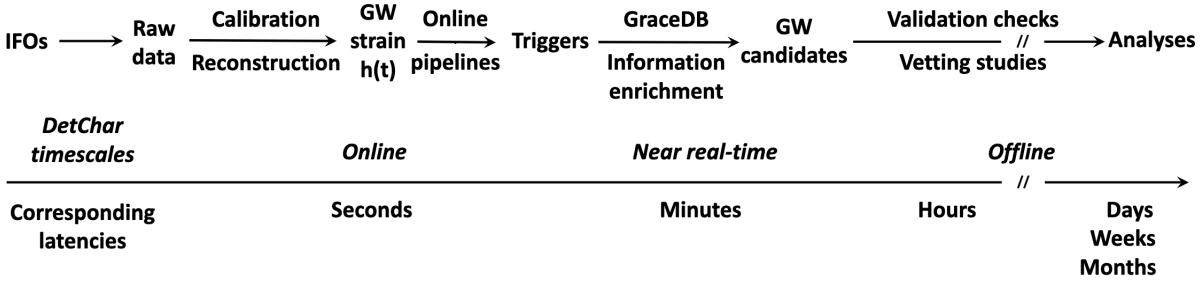


Figure 2.2: Dataflow from detector data acquisition to offline validation of GW events showing the three main timescales at which the Virgo detector characterization contribute (online, low-latency, offline) [6]. GraceDB (Gravitational-wave Candidate Event Database) is the LVK collaboration online archive and management system used to record, coordinate, and communicate potential detections in real time.

from differences between models and measurements (including measurement errors) of actuation and sensing functions. Calibrated strain data are created online for use in low-latency searches. Calibration measurements are taken periodically during each observing run and once observing runs have finished, final time-dependent calibrations are generated for each detector. Please note that calibrated strain data may be non-stationary and non-Gaussian at certain times and frequencies [1].

All the collected data from the different interferometers are available through the Gravitational Wave Open Science Center (GWOSC), a service provided by LVK collaboration to access GW data products [1] like calibrated strain. LVK collaboration manages also internal databases and archives, divided in low-latency products and data for offline analyses. Figure 2.1 shows the general data processing system adopted by LVK collaboration: the most crucial step involves the interplay between detector characterization studies and data quality assessments. Data quality involves identifying and characterising noise features in order to exclude data of low quality or establish the impact of remaining artefacts on searches for GW signals. After calibration and data quality, the *data preprocessing* involves *filtering* and *whitening* to suppress non-astrophysical features and to standardize the noise properties. First, for typical applications, it is often sufficient to analyze the downsampled data at 4 096Hz or 2 048Hz. Data analysis pipelines usually operate in the frequency domain by means of the *discrete Fourier transform* (DFT), since data are sampled in a discrete way. The DFT is a series in the complex field evaluated on a frequency grid with constant spacing and is given by

$$\tilde{h}_k = \sum_{n=0}^{N-1} h_n \, e^{-2\pi i \frac{n}{N} k} \,, \tag{2.1}$$

where k is the frequency index and N is the number of points, related to the time series length T and the discrete sampling time interval Δt[1] through

$$N = \frac{T}{\Delta t} \,. \tag{2.2}$$

Moreover, the inverse DFT is given by

$$h_n = \sum_{k=0}^{N-1} \tilde{h}_k \, e^{2\pi i \frac{k}{N}} \,. \tag{2.3}$$

[1]The inverse of the sampling frequency.

For a discrete signal, the maximum measurable frequency that can be represented in the frequency domain is the Nyquist frequency, exactly half of the sampling frequency.

In particular, the PSD is estimated using Welch's method, which provides a more stable estimate by reducing the effect of noise. The time series considered for a PSD estimate is divided into overlapping segments, and each segment is windowed and transformed into the frequency domain using the DFT. The resulting estimates are then averaged to obtain the final PSD. This procedure reduces statistical fluctuations, but comes at the cost of lower frequency resolution, especially at low frequencies, where fewer frequency bins are available. Although Welch's method is widely used because of its simplicity and robustness, its parameters need to be carefully selected to achieve a good balance between frequency resolution and variance. In particular, an inappropriate choice of segment length can significantly affect the quality of the PSD estimate. Similarly, the choice of the windowing function is important to reduce spectral leakage and obtain an accurate PSD estimate.

In the frequency domain, the interferometer data is filtered to prevent aliasing and isolate the frequency bands of interest. This makes subsequent analyses simpler and improves the accuracy of the results. Moreover, discrete sampling over time typically involves correlations between samples. Eliminating these correlations from gravitational strain data is of paramount importance since it allows astrophysical signals to emerge more clearly. The standard preprocessing technique that accomplishes this task is data whitening, which rescales each frequency component so that the variance of the power spectrum is uniform across frequency bins. Thus, whitening normalizes the strain data across the frequency spectrum, enhancing excess power relative to background noise. Although glitches can introduce residual correlations, whitening helps identify and mitigate them. A whitened strain time series can be expressed as follows:

$$h^W(t) = \int_{-\infty}^{+\infty} df \, \frac{\tilde{n}(f)}{\sqrt{S(f)}} \, e^{2\pi i f t} \,, \tag{2.4}$$

where $S(f)$ is the PSD as introducted in Equation 1.17.

2.2 Detector characterization

As mentioned earlier, the ensemble of studies and activites aimed at understanding, monitoring, and diagnosing the instrumental behavior of GW interferometers, with the ultimate goal of ensuring that the data quality is high enough for scientific analysis, undergoes the name of detector characterization. Being the bridge between instrument science and data analysis, it requires knowledge of the detector functioning, combined with analytical tools to process the enormous amount of auxiliary data. It is a continuous process involving commissioning periods, observing runs, and offline reprocessing. Figure 2.2 illustrates the different Virgo detector characterization pipelines with the corresponding time scales towards GW events validation and astrophysical analyses, highlighting the crucial role of online and low-latency analyses. Figure 2.3, more specifically, shows Virgo detector characterization tools and data quality products and how alla the pipelines operates in parallel to deliver information to search and parameter estimation pipelines.

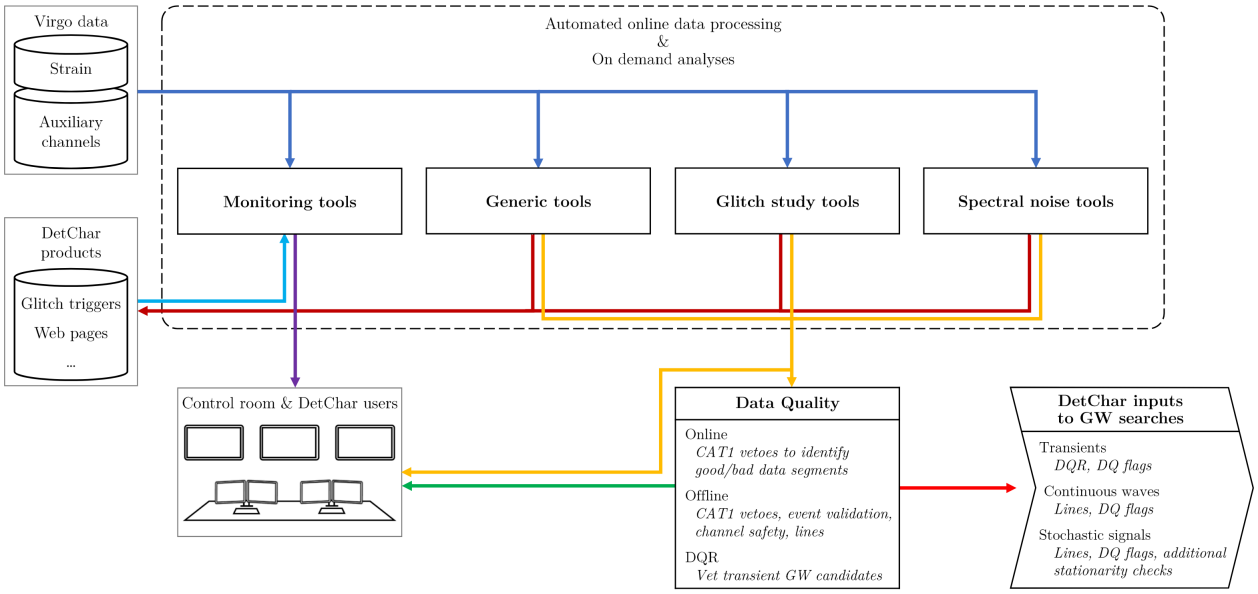


Figure 2.3: Virgo detector characterization tools and data quality (DQ) products [6]. Collected data are analyzed in parallel by different online pipelines to investigate noise sources that produce most importantly glitch triggers. DQ products, like vetoes, event validation and data quality reports (DQRs), are necessary for subsequent data analysis pipelines, most especially searches.

2.2.1 Data Quality

More in detail, identified glitches and periods of poor data quality are flagged. Periods of data are removed, or more technically *vetoed*, at various levels or categories depending on the severity of the problems [1]. Additionally, channels indicating different levels of data quality issues are also available: data quality vetoes are indeed labelled according to severity. Critical hardware or calibration issues result in exclusion from all analyses, whereas minor correlated disturbances are selectively removed from specific search pipelines [6]. In particular, times when noise sources with a known physical connection to the detector's strain channel are likely to be active and cause glitches are identified. Candidates at or around these times may then be excluded from the search results. The key feature helping this process is an array of environmental monitors that detect disturbances in the physical environment and respond to them more sensitively than the detector's gravitational strain channel. This array includes seismometers, microphones, accelerometers, radio receivers, and magnetometers for monitoring ambient noise. The sensitivity of the environmental sensors is verified via a set of noise injections performed at the beginning and end of each observing run. Acoustic, magnetic, radio frequency and vibrational tests are conducted to quantify the coupling of ambient noise to gravitational strain data. These injections are conducted at multiple locations around the detector to verify how the sensors couple with $h(t)$ via multiple potential coupling paths and to improve understanding of this process. Figure 2.4 summarizes the joint activities involving data quality for Virgo: Omicron is the main tool to identify transients, as already seen, but each tool complements each other.

For an interferometer to operate in a nominal way and produce strain data that can be scientifically analyzed, it is required to be in *lock*. The locking condition technically refers to the optical cavities being in simultaneous resonance, and all control systems in their nominal operating point. Consequently, a key part of data quality processes is temporal segmentation of the data stream into intervals during which the detector status has been precidely defined. More in detail, the interferometer

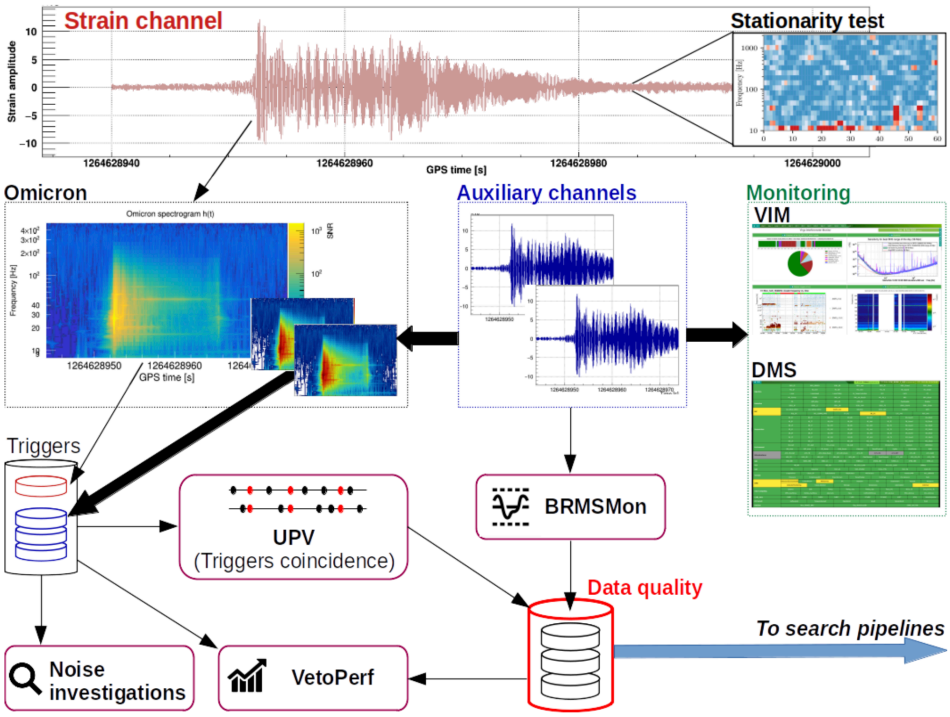


Figure 2.4: Typical workflow for transient studies in Virgo with the full set of detector characterization monitoring tools [6]. Data from the strain channel is analyzed both to search for excess power and to ensure properties like stationarity and Gaussianity, which are the basic assumptions for GW events search. Moreover, UPV (Use-Percentage Veto) tool triggers glitch coincidence thanks to information coming from auxiliary channels. VetoPerf tool has been implemented to find a balance between vetoes and removed data. BRMSMon (Band-limited Root Mean Square Monitor) looks for noise patterns, VIM (Virgo Interferometer Monitor) and DMS (Detector Monitoring System) collect the status off all detector components, from hardware to online software processes.

is said to be in *science mode*, the nominal data taking state [6], when it is locked, fully calibrated and operating within its nominal noise budget, with all subsystems correctly configurated and no ongoing activities that could compromise data integrity. Only data that has been collected while the detector is in science-mode is considered valid for astrophysical searches. The transition of the interferometer into science mode must be cross-checked against auxiliary indicators (e.g., stable laser power, absence of ongoing hardware injections or maintenance activity) before data is certified as scientifically usable [16].

The locked state is inherently fragile: environmental disturbances (seismic transients, strong winds, anthropogenic activity), instrumental faults (like laser instabilities, electronics failures), or even large glitches themselves can cause the interferometer to fall out of resonance, an event known as lock loss. Following a lock loss, the detector enters an automated or semi-automated re-acquisition sequence that can take from a few minutes to several hours depending on the severity of the disturbance and the complexity of the locking sequence, during which no science data are produced. These intervals, together with scheduled maintenance periods, commissioning breaks, and unplanned outages, constitute the down-time of the detector. Down-time directly determines the detector's duty cycle and the effective observing time available for searches, but monitoring it could be useful for diagnostic purposes too, as recurring lock-loss patterns often point to specific instrumental weaknesses that need to be addressed.

The building blocks of GW data are thus *data segments*, meaning stretches of time series data with

common properties (data quality level, particular detector condition, etc.) [6]. Segments are identified by a starting and an ending integer GPS times and associated with a specific flag and version that allows to query time intervals satisfying the data quality criteria needed for a considered analysis. Data quality information is encoded in flags that label data segment and is stored in the Data Quality Segment Database (DQSEGDB) [23], that stores approximately 600 million metadata entries per day during an observational run.

2.2.2 Glitch mitigation

Other than detector monitoring, diagnostics, and data quality assessments, as already seen, another pillar of detector characterization is noise mitigation, that involves both searching for a root cause and performing mitigation at a data level. Glitch mitigation consists of identifying, removing glitches and neutralizing their effect in detector data to prevent biased parameter estimation. Parameter estimation pipelines generally assume stationary Gaussian noise, conditions violated in presence of a glitch that is overlapping with a signal or very close to it, because in the latter case it strongly impacts PSD estimation for example.

Depending on their morphology and their time-frequency overlap with the astrophysical signal, glitches have been shown to introduce statistically significant biases in the recovered astrophysical parameters [30]. Sky localisation is particularly sensitive to this effect, since already in regular conditions the posterior distributions over sky position can exhibit several distinct, well-separated modes due to degeneracies in the response of the detector network, which makes it harder to disentangle an actual secondary mode of the signal from one artificially induced by a nearby noise transient. For events with an expected electromagnetic counterpart, such spurious features in the sky map can seriously hinder the follow-up effort [27]. The presence of glitches is indeed generally correlated with multimodality in parameter estimation [5]. Moreover, for example in GW170817 event, the glitch occurred in L1 detector slightly before the merger, was removed by simply gating data, a strategy that is obviously not applicable when the glitch and the signal lengths are comparable or if the glitch covers extensively the signal. Gating glitches near to signals can also introduce negative effects on the PSD evaluation, meanwhile reconstructing the glitch and the signal simultaneously may reduce the errors in signal parameter estimation [32].

All these issues emphasize the importance of thorough glitch removal and mitigation. Nowadays, the state of the art for glitch mitigation in LVK collaboration is *BayesWave* algorithm [14]. BayesWave is part of glitch mitigation routine during scientific runs, and from O3 also *gwsubtract* has been introduced [17]. It models and subtract glitches by exploiting information from auxiliary channels that are independent from the gravitational strain one. Limitations of this algorithm reside in its assumptions: noise sensors are not perfect and their transfer functions may not be linear, resulting in the introduction of systematic errors [17].

Bayes Wave

BayesWave is a Bayesian inference-based algorithm designed for the joint detection and characterization of glitches and GW transients. Excess power is modelled using a frame of Morlet-Gabor

wavelets[2] and the entire analysis can be implemented in the frequency domain. The wavelet representation is a natural one since glitches and signals are placed in a compact region of the time-frequency plane [14]. The signal or glitch model can contain a variable number of wavelet components, whose number and parameters are sampled using trans-dimensional Reversible Jump Markov Chain Monte Carlo (RJMCMC), giving it great flexibility at identifying generic transients [28].

The Bayesian approach allows to introduce parametrized models that explicitly account for non-stationarity and non-Gaussianity in data. The model degree of complexity is not fixed in advance, but rather, determined by Bayesian model selection, that allows to find the appropriate balance between model fidelity and model complexity, avoiding the risk of overfitting the data. Moreover, the trans-dimensional RJMCMC sampler employs a Parallel Tempering strategy: multiple chains are run in parallel at different "temperatures". The likelihood function is divided by the temperature parameter allowing "hot" chains to explore the entire prior volume, while the "cool" ones map the posterior distribution peak [14].

BayesWave presents however some limitations, like the introduction of systematic errors in the produced models [14] and issues arise in particular at low frequency and low SNR, especially for Scattered Light glitches. BayesWave is therefore not able to fully model all glitch feature for all glitch classes and the glitch reconstruction is more accurate if the glitch duration is short [17]. Moreover, on the computational side, the RJMCMC is an expensive process that does not guarantee convergence and can cause problems in online and low-latency analyses.

Another example that points out BayesWave modeling implications in the astrophysical interpretation is the GW200129 event, a CBC between two black holes. The difference between a spin-precessing and a nonprecessing interpretation for this event is smaller than the statistical and systematic uncertainty of the glitch subtraction, finding that the support for spin-precession depends sensitively on the glitch modeling. In particular a glitch with low SNR in the LIGO L1 detector overlapped with the signal: it was removed both with gwsubtract and BayesWave, with the latter predicting a quite different model for the glitch, in particular with greater amplitude. However, both reconstructions carried an uncertainty larger than uncertainty induced by detector calibration. Moreover, also the choice for the prior distribution for the BayesWave glitch model influences and interferes in understanding physical processes [31].

2.3 Searches

Search methods are used to identify candidate signals, which are broadly divided into *modeled* CBC searches and *unmodeled* searches. Unmodeled searches identify coherent excess power from poorly modeled or unexpected sources across detectors. Although unmodeled searches are less sensitive to CBC signals, they are capable of detecting a wider variety of transient phenomena.

Modeled searches are based on *matched filtering*, a strategy that compares detector data with a bank of waveforms obtained by modelling CBCs mainly through Post-Newtonian (PN) expansions or effective-one-body (EOB) formalism. The matched filter signal to noise ratio statistics is defined by

[2] A Morlet-Gabor wave function is obtained by multiplying a complex sine wave, the carrier, by a bell-shaped Gaussian curve, the window. A frame generalizes the concept of a basis to allow for sets that are linearly dependent.

the following scalar product between data $s(t)$ and a template waveform $h(t)$:

$$\langle s(t), h(t) \rangle = 4\,\mathrm{Re} \int_0^{+\infty} df\, \frac{\tilde{s}(f)\tilde{h} * (f)}{S(f)} \tag{2.5}$$

where $S(f)$ is the previously introduced noise PSD. It can be shown that matched filtering is the optimal linear method for detecting a known signal hidden in stationary Gaussian noise. In practice, the SNR is computed for each template's time-shifted versions. An alert is triggered when a local maximum exceeds a threshold. Template waveforms are collected in a so called template bank that covers, in a discretized way, the parameter space of possible CBC signals. To evaluate how well a signal has been identified through matched filtering, the *match* represents the overlap between the signal and the template:

$$M(\theta_1, \theta_2) = \max_t \left| 4 \int_{f_{\min}}^{f_{\max}} df\, \frac{\tilde{h}^*(f;\theta_1)\tilde{h}(f;\theta_2)}{S(f)}\, e^{2\pi i f t} \right|^2 . \tag{2.6}$$

A value for the match lower than unity represents a loss of recovered SNR, however a trade-off between computational costs and SNR loss must be faced. Matched filtering results however have to be assessed by signal consistency tests and a detection statistics has to be produced along with the computation of the false alarm rate probabiity, computed against background rates using time-shifting, and the probability that the candidate is of astrophysical origin.

2.4 Machine learning

As already seen with Gravity Spy, ML methods are becoming increasingly part of the GW data analysis world. ML methods can be indeed useful at various levels as described below [15].

- **Noise mitigation strategies**
  ML methods includes contributions to noise characterization from strain data and with auxiliary channels, interferometer control and denoising. ML methods can usefully complement conventional signal processing methods to enhance data quality performances. In particular, self-contained algorithms that can be incorporated into already-existing searches or noise investigation processes are especially useful. Some examples of ML methods currently implemented in the detector pipelines at this levels are GSpyNetTree (Gravity Spy Convolutional Neural Network Decision Tree) [7] and iDQ [22]. GSpyNetTree is a classifier fed with spectrograms (as Gravity Spy) currently used in the validation of GW event candidates. iDQ is a supervised learning algorithm to identify noise transients in real time by leveraging data from auxiliary channels, it is a wavelet-based extractor that generates feature tables. Other than supervised ML methods, also unsupervised ones have been explored, especially for learning the underlying distribution of glitches and creating for example glitch catalogs. As for interferometer control, ML algorithms can be useful since they can exploit nonlinearity: although the detector is intrinsically nonlinear, most detector control techniques rely on linear controllers. This is where ML methods can help by approximating instrumental nonlinear behavior around the planned mode of operation or substitute linear control systems. For example, Deep Loop Shaping [9] is a reinforcement learning system that is capable of strongly suprressing control noise from mirror

stabilization control as it has been successfully tested on LLO during the last scientific run. ML will thus also be able to help us with increasingly complex detector architectures, in identifying and reducing nonlinear couplings in various noise sources. Finally, denoising concerns the removal of excess noise after noise identification. Denoising ML methods are divided in two classes: models to remove persistent noise (like spectral lines) and models to remove glitches. Some examples are DeepClean [29] and NonSENS (NON-Stationary Estimation of Noise Subtraction) [35]. DeepClean is based on a one-dimensional Convolutional Neural Network (CNN) to estimate excess noise in gravitational strain data from auxiliary channels data; it was successfully tested in LIGO detectors during O2 and used during O3. NonSENS is a deep learning method designed to characterize non-stationary noise couplings in auxiliary witness data streams and perform time-domain noise subtraction in the strain channel; it has been used in LIGO detectors during O3.

- **Modeling and data generation**
  ML algorithms can also contribute to modeling of astrophysical sources and GW signals, and noise generation. Traditional approaches have been very limited in their ability to realistically simulate detector noise, therefore ML solutions are already considered the state of the art, especially methods implementing Generative Adversarial Networks (GANs). In particular, generative noise modeling is closely related to noise mitigation and data characterization tasks. Having simulated noise datasets allows to identify and address systematic errors or biases in data analysis pipelines. Moreover, generating synthetic data that mimics real observations allows us to stress-test our detection or parameter estimation pipelines.

- **Signal detection**
  Methods based on ML have been developed for modeled and unmodeled searches, continuous-wave and stochastic searches. For example, coherent Wave Burst (cWB), the main pipeline for unmodeled searches, has been upgraded with XGBoost[3] to automate the signal-noise classification of cWB events.

- **Parameter estimation and physical interpretation**
  ML algorithms can target source classification and parameter estimation, applications for multi-messenger astrophysics, population interence and cosmology, in particular by making parameter estimation easier by enabling faster determination of key parameters such as the masses and spins of CBCs. To date, the main application of ML to Bayesian parameter estimation is implemented with neural posterior inference, especially normalizing flow as in the DINGO pipeline [25] [18].

- **Next generation GW detectors**
  In this scenario, where great improvement in sensitivity, huge event rate, and much higher achievable SNRs are expected, ML models can contribute by for example generating more accurate waveform models and making processing techniques more efficient overall. In particular, ML can help facing new challenges, like longer and overlapping signals in LISA, and mitigation of different noise sources.

[3]Extreme Gradient Boosting, a ML library implementing gradient boosting on decisional trees.

The ML revolution is quickly changing how data analysis is done. ML promises to be faster and more flexible than standard methods, and more robust when models of the signal and noise are incomplete or imperfect. ML algorithms have already been shown to be faster and more efficient for some tasks. However, in other areas they are simply another option, not necessarily a better one. Also, any new algorithm proposed to a large collaboration like the LVK must go through a careful review process before it can become an official, production-level analysis tool. As mentioned earlier, ML methods are already used in production within LVK for removing noise and for classifying CBC alerts at low-latency. For search pipelines and parameter estimation, several ML methods are getting ready for production use, and this trend is likely to grow before the next observing run. More importantly, there are open questions about whether ML-based searches for GW signals can produce reliable significance estimates on their own, or at least about how well these estimates carry over from standard training/testing setups to real-world use. Another common criticism is the lack of reproducibility and explainability in many ML results, along with the risk of unexpected failure modes [15].

Chapter 3

Deep learning for glitch mitigation

One of the recently developed and promising ML-based methods for GW data analysis is DeepExtractor [19] [20]. DeepExtractor is a deep learning model designed specifically for glitch mitigation, that has been however demonstrated to be capable of accurately reconstruct any excess power in the data no matter the source, including GW signals. DeepExtractor exploits the fact that, from a machine learning algorithmic point of view, the whitened underlying background strain noise is much easier to model (or to "learn") compared to all the possible morphologies of excess power. The statistical distribution of signal and glitches, other than not being constant in time and mostly unknown, in principle can also overlap, and this is the bottleneck faced by all GW signal ML classification methods that has been designed in a physics-driven way, other than problems regarding unbalanced labelling between classes. The whitened strain noise distribution can be approximated as locally stationary and Gaussian: it is then much easier to characterize and it distinctly stands out from the distribution of glitches and actual GW signals, characterized by non-Gaussianities. The DeepExtractor target is therefore the mapping of time series containing excess power to the underlying background, allowing to subsequently extract the excess power time series by subtraction.

DeepExtractor's effectiveness has been assessed through various experiments, including a direct comparison against BayesWave. Moreover, DeepExtractor is capable of reconstructing actual O3 GW events without being trained on GW waveforms. However, DeepExtractor has only been tested on LIGO data. Since the LIGO detectors are very similar, with only some conditions varying, it is also important to test this method on Virgo in the framework of GW data analysis for the LVK collaboration. Virgo's design is quite different from the LIGO detectors, and it suffers from different noise sources and instabilities that strongly affect its strain noise. This work therefore includes the developement of a DeepExtractor model customized for Virgo and its validation in order to verify whether the model output is physically consistent and compatible inside the detector data quality workflow. The validation has been carried out on Gravity Spy's O3 Virgo dataset and the tool used for the validation process is *glitchfind* [36], an algorithm that computes the probability of any excess power in the data.

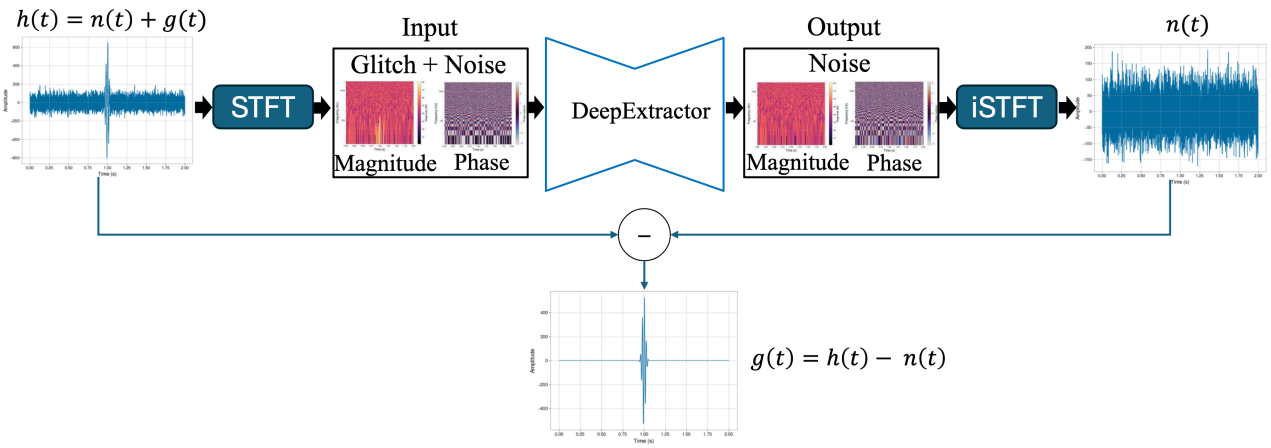
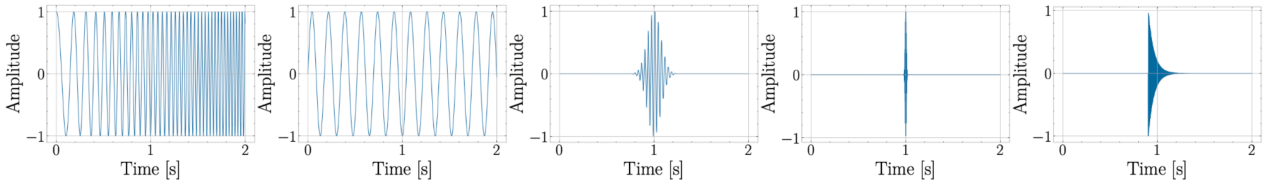


Figure 3.1: DeepExtractor's reconstruction framework [20]. STFT is applied to a 2 s strain data segment $s(t)$ containing some excess power $g(t)$, in order to obtain magnitude and phase spectrograms that serves as input for the network. The network produces the corresponding noise magnitude and phase spectrograms. By applying the inverse STFT, the reconstructed background noise time series $\hat{n}(t)$ can be obtained and the reconstructed excess power $\hat{g}(t)$ is extracted by subtraction of the output from the input.



(a) Chirp (b) Sine (c) Sine-Gaussian (d) Gausian Pulse (e) Ringdown

Figure 3.2: The glitch classes used for injections in training and validation samples, they are displayed with random parameters. They have been selected since they are typically used in unmodelled searches and could therefore be useful in unmodelled reconstruction as in the DeepExtractor framework.

3.1 DeepExtractor

DeepExtractor employs a U-Net, a deep learning network first developed in [34], to perform excess power extraction from gravitational strain time series data. Figure 3.1 illustrates the DeepExtractor framework: the input is a whitened 2s strain data segment containing some excess power, and the output is the reconstructed excess power time series. More in detail, the input data are transformed in frequency domain through the STFT, as introduced in Section 1.3.1, obtaining magnitude and phase spectrograms. The network produces reconstructed background noise spectrograms, and to obtain the corresponding time series an inverse STFT is applied. DeepExtractor is built on `PyTorch`[1] and is publicly available as a package[2]. The framework is based on the assumption that excess power, whatever its source may be, whether it is a GW signal or a glitch, adds linearly to the underlying background noise in the gravitational strain channel as follows:

$$s(t) = n(t) + g(t)\,. \tag{3.1}$$

Consequently, the extracted excess power $\hat{g}(t)$ can be obtained by subtraction:

$$\hat{g}(t) = s(t) - \hat{n}(t). \tag{3.2}$$

[1]Python deep learning library.
[2]https://github.com/tomdooney95/deepextractor

3.1.1 Training strategy

Following the linear assumption described above, synthetic training data are carefully generated. The target is a 2s strain noise data segment, with a sampling frequency $f_s = 4096$. The network is therefore fed by the simulated strain noise in which multiple simulated glitches are linearly injected. The number, the type and the glitch parameters are randomly selected, as well as the injection time for each glitch. In Figure 3.2 are illustrated the different glitch classes used for building the training dataset. The maximum number of glitches that can be injected is 30, follows a similar approach to BayesWave modelling glitches with basis wavelets. By learning to isolate background noise from diverse combinations of waveforms from these classes, DeepExtractor is capable of interpolating to arbitrary signals and glitches. The whitened noise time series containing glitches are then standardized using `StandardScaler` from `sklearn` library, obtaining data with zero mean and unit variance. The standardization method assumes Gaussian distribution of the data, making it appropriate for Gaussian background noise. It can also help with training by creating a more balanced dataset. The scaler is fitted to the input data and the output data undergoes an inverse scaling process with the same parameters as the input data. This ensures that both inputs and outputs occupy the same standardized space. Input data are then transformed using the STFT[3] to train the model, that is therefore fed with a two-dimensional input, the magnitude and the phase spectrogram, each one with dimensions 257 x 257. The STFT implements a Hann window to reduce spectral leakage and STFT parameters, like Fourier transform size (that determines frequency resolution), window size and the overlap between window frames, are set to ensure compatibility with the Constant Overlap-Add condition (COLA), that ensures continuous signal reconstruction from overlapped segments. The simulated dataset consists of 250,000 training samples and 25,000 validation samples.

The objective of the network is minimizing a mean squared error (MSE) loss function:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \left[\hat{n}_i(f) - n_i(f)\right]^2 , \tag{3.3}$$

where N is the batch size, fixed to 32. Moreover, DeepExtractor makes use of an Adam optimizer and to optimize the learning process, the `ReduceLROnPlateau` scheduler by `PyTorch` is implemented, allowing to reduce the learning rate after 4 training epochs without gain in the validation loss. Also, early stopping is applied after 10 epochs without improvement ensuring an efficient and flexible training duration that can adapt based on the model's performance on the validation set.

A model trained exclusively on simulated data could suppress spectral lines when applied to real data, leading to unreliable reconstructions, because the simulated noise does not include spectral lines from instrumental and environmental characteristics. Therefore, incorporating real data into the training of DeepExtractor is crucial for adapting to the non-Gaussian, non-stationary nature of GW detector noise and ensuring accurate modeling of real events outside the training distribution. The transfer learning strategy can overcome this limitation by leveraging knowledge learned from simulated and more controlled data, the easy task, to enhance performance or reduce training effort on real data, the difficult task. Only real data from O3 has been considered, starting from the selection of background noise samples avoiding GW events and Omicron triggers. After backgrounds have been

[3] Already implemented in the PyTorch library.

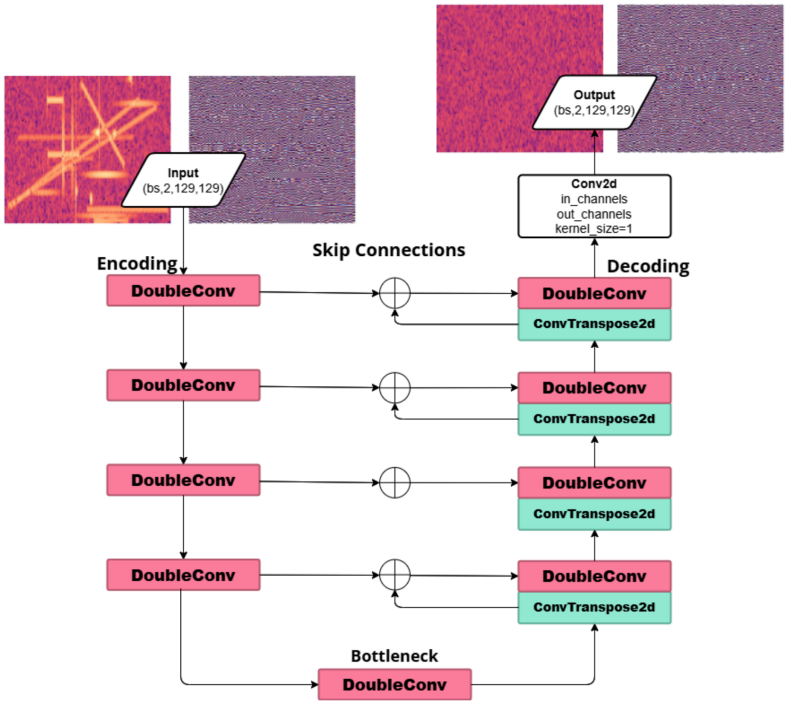


Figure 3.3: DeepExtractor U-Net architecture [20]. The network is fed in batches containing magnitude and phase spectrograms, that are processed simultaneously. The default DeepExtractor model is a 4-level U-Net that is fed with 257x257 spectrograms (and not 129x129 as indicated here). The spectrograms displayed here belong to the simulated training phase.

collected, synthetic glitches can be injected as described above. The dataset obtained consists of 45,400 samples.

All models were trained on an NVIDIA A100 GPU[4]: the DeepExtractor model trained on the simulated data in approximately 12 hours, corresponding to 36 epochs. Transfer learning on real LIGO O3 data completed in under 20 minutes with 3 additional epochs.

3.1.2 Architecture of the model

The U-Net architecture is based on the autoencoder neural network, which aims to learn efficient, low-dimensional representations of data in an unsupervised framework. An autoencoder consists of an encoder, that maps the input to a latent representation, and a decoder, which reconstructs the input from the latent representation. The MSE loss drives the training to compress in latent space the essential data features. Convolutional layers are widely used within autoencoders and are particularly effective for spatial or time-frequency data since they can exploit local correlations through shared convolutional filters. U-Net, originally developed for biomedical image segmentation, redesigns the autoencoder by adding skip connections between corresponding levels of encoder and decoder layers. Skip connections allow fine-grained details to be retained while maintaining a large receptive field. The detailed information recorded in this manner during compression is useful for more accurate and faithful reconstruction while decoding. The U-Net is therefore capable of capturing both high and low-level features through skip connections. U-Nets are particularly well-suited for spectrogram-based representations of signals and glitches that span wide time–frequency ranges, where conventional deep autoencoders may struggle due to information loss from aggressive compression. U-

[4]Featuring 80 GB of RAM. Thanks to LIGO computing infrastructures.

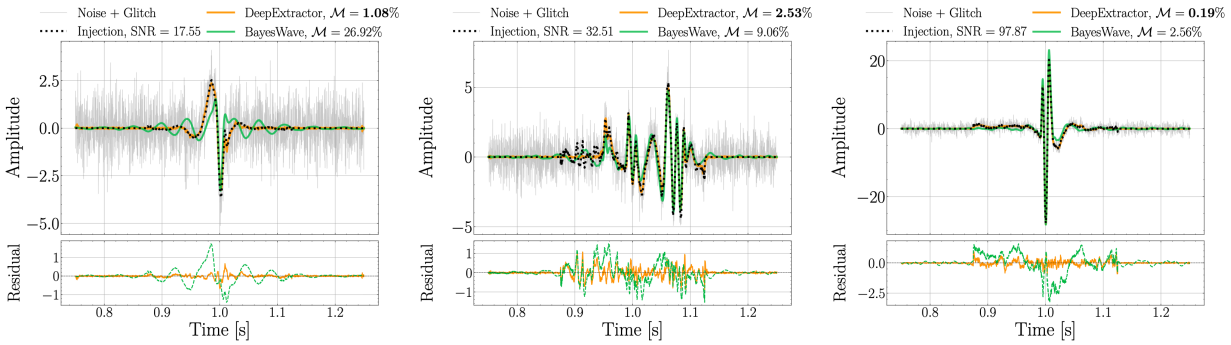


Figure 3.4: Examples of the comparison between DeepExtractor and BayesWave reconstructions [20]. The black dotted curves show the injected glitches, which SNR is specified. The evaluation metric is the mismatch, which for DeepExtractor is significantly lower than BayesWave. The bottom panels show residuals both for DeepExtractor and BayesWave, and the latter is clearly missing some structure of the injected waveform.

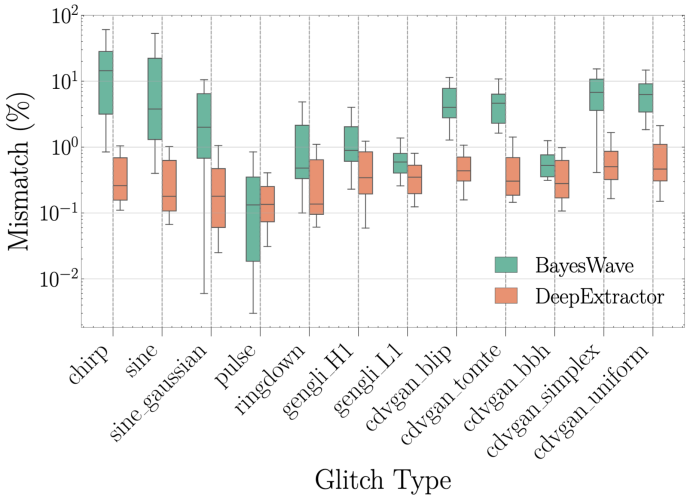


Figure 3.5: Mismatches between injected and reconstructed samples per simulated glitch class [20]. Other than the 5 glitch classes used also for simulated training data generation (where pulse stands for Gaussian pulse), glitches generated with GANs are included too. DeepExtractor clearly shows better reconstruction capabilities in terms of mismatch.

Nets are widely used in in acoustic source separation and denoising. Figure 3.3 illustrates the U-Net architecture on which DeepExtractor is based on. The encoder branch extracts hierarchical features thanks to double (since the input is two-dimensional) convolution blocks with kernel size 3, each followed by batch normalization and ReLU activation function. The bottleneck captures the most abstract representations of the input with a double convolution block. The decoder path reconstructs the original input while integrating features saved at each encoder level: each layer consists of upsample and double convolution blocks. Finally, to produce the output, a single two-dimensional convolutional layer with kernel size 1 and linear activation function is implemented.

3.1.3 Comparison with Bayes Wave and reconstruction of actual GW signals

DeepExtractor has been compared against Bayes Wave, the state of the art for glitch mitigation. The two methods are compared by evaluating their performance on simulated glitch injections. The performance is quantified by the mismatch, computed between the injected and the reconstructed waveform, and defined as

$$\mathcal{M} = 1 - M\,, \tag{3.4}$$

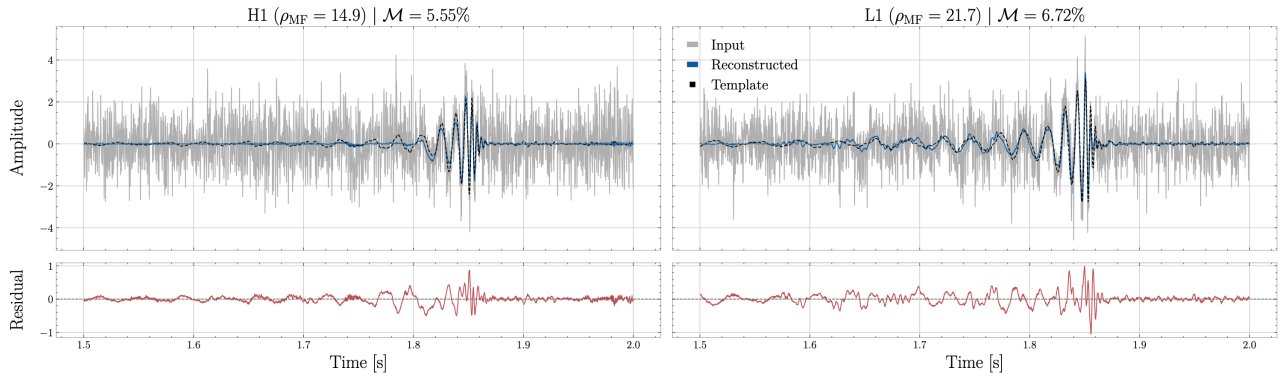


Figure 3.6: DeepExtractor reconstructions for actual O3 GW signal GW200129_065458, both for H1 (right) and L1 (left) strain data [20]. Other than strain data and DeepExtractor reconstructions, it is also shown the template waveform along with the matched filter SNR (ρ_{MF}). DeepExtractor achieves mismatches way lower than 10%, however in the residuals on the bottom some structure is left.

where M denotes the match defined in Equation 2.6. The dataset on which the two methods are compared consists of 30 samples per simulated glitch class. Each 2 s segment contains only one synthetic glitch. The considered glitch types are the 5 already seen in Figure 3.2 and glitches generated by deep learning methods too, to extend also to unseen glitches. These deep learning methods are based on GANs and can generate synthetic glitches. More in detail, `gengli` [26] and `cDVGAN` [21] are two independent realistic glitch generators: `gengli` can generate Blip glitches from both LIGO detectors, while `cDVGAN` can produce Blip and Tomte glitches, and signals from CBCs between binary black holes too, but also hybrid samples (labelled as *simplex*) mixing features from different training classes, allowing to introduce diversity. The lower SNR threshold is set to 15, to prevent potential convergence issues with BayesWave. Figure 3.4 shows different examples of reconstruction with DeepExtractor and Bayes Wave: residuals from DeepExtractor reconstruction generally display lower amplitude compared to Bayes Wave ones, which clearly still contain low-frequency components. DeepExtractor consistently outperforms BayesWave in terms of mismatch, even on previously unseen test classes, as can be observed in Figure 3.5.

As for computational costs, DeepExtractor reconstruct a single glitch sample in an average of 0.14 s on a CPU or 0.05 s when processing batches of 256 samples. This represents a gain of a factor of 10,000 compared to Bayes Wave and makes DeepExtractor well-suited for online and low-latency large-scale applications in GW data analysis.

Since DeepExtractor is actually agnostic on the source of excess power, it bridges glitch mitigation and signal reconstruction. DeepExtractor has also been applied to the reconstruction of actual GW signals from O3, and an example can be found in Figure 3.6. Selected GW events are among the loudest that have been confidently detected by both LIGO detectors during O3. In this case DeepExtractor performance is compared to the maximum likelihood template obtained from parameter estimation and achieves less than 10% mismatch. The template is given according to colored detector noise, so it has been whitened with the same PSD used for the signal, which has been computed from a 14 s stretch of data before the 14 s segment containing the signal. Working on the alignment of DeepExtractor's preprocessing and training strategy with parameter estimation pipeline could improve reconstruction accuracy. This application of DeepExtractor shows the capabilities of this framework for reconstruction of actual GW signals.

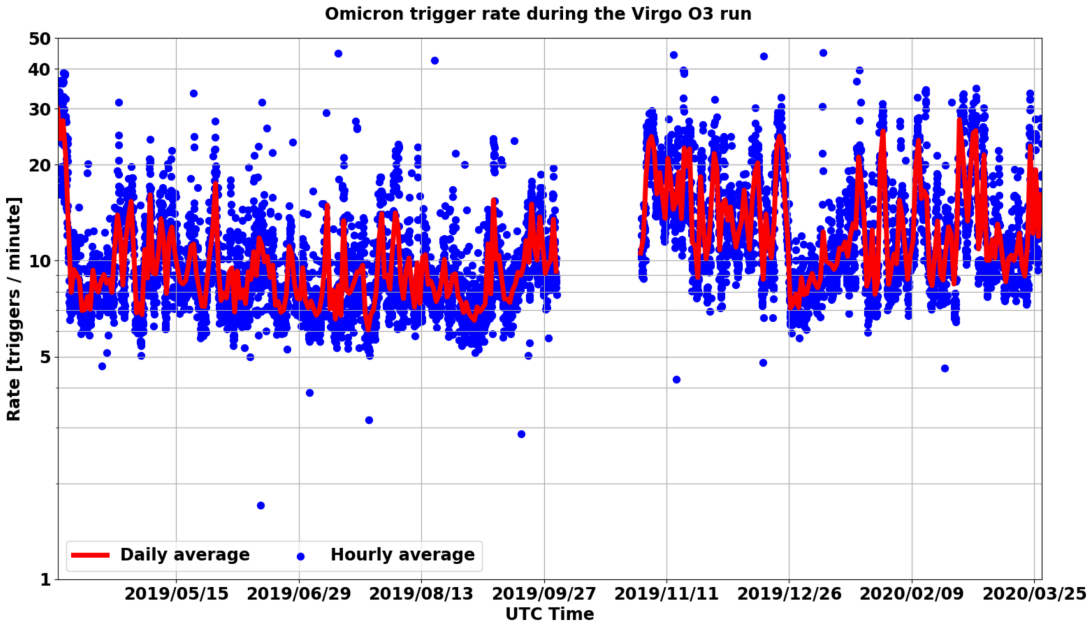


Figure 3.7: Virgo glitch rate from Omicron triggers during O3 run excluding CAT1-vetoed segments (meaning no serious detector issues included) [6]. There is no filter on glitch frequency and all flags for the detector operating mode are included. Blue dots indicates hourly rate average and the red curve shows daily rate average. The gap separates O3a and O3b with a 1-month commissioning break.

3.2 DeepExtractor for Virgo

As previously mentioned, a DeepExtractor model tailored for Virgo is necessary due to the different instrumental condition that reflects on the data. Also, the default DeepExtractor model trained on simulated data considers only simulated LIGO noise. Following the transfer learning strategy used for LIGO, the default DeepExtractor model previously trained on simulated data as described above, has been fine tuned on real Virgo data. The obtained model contains about 30 million trainable parameters.

Before considering real data, a simulation study has been completed in order to get an understanding of what to expect and obtain a benchmark which can then be used to compare performance against real data. The default model has been trained on simulated Virgo noise and evaluated on simulated data.

3.2.1 Data and training strategy

The data considered for building the DeepExtractor Virgo model is taken from O3 run. Whether to obtain background noise data for training or to obtain glitches to reconstruct for testing, it is necessary to start with the Omicron triggers. From Figure 3.7 it can be observed that the trigger rate is on average of 10 triggers per minute, rougly one trigger every 6 seconds. It is therefore not trivial to find clean stretches of data that can be selected as background noise for the training and validation datasets. Moreover, from Figure 3.8 it is clear that triggers with low SNR dominate and triggers with growing SNR are more and more rare. Also, considering frequency ranges, triggers with highest rate have frequency support below 45 Hz or above 460 Hz.

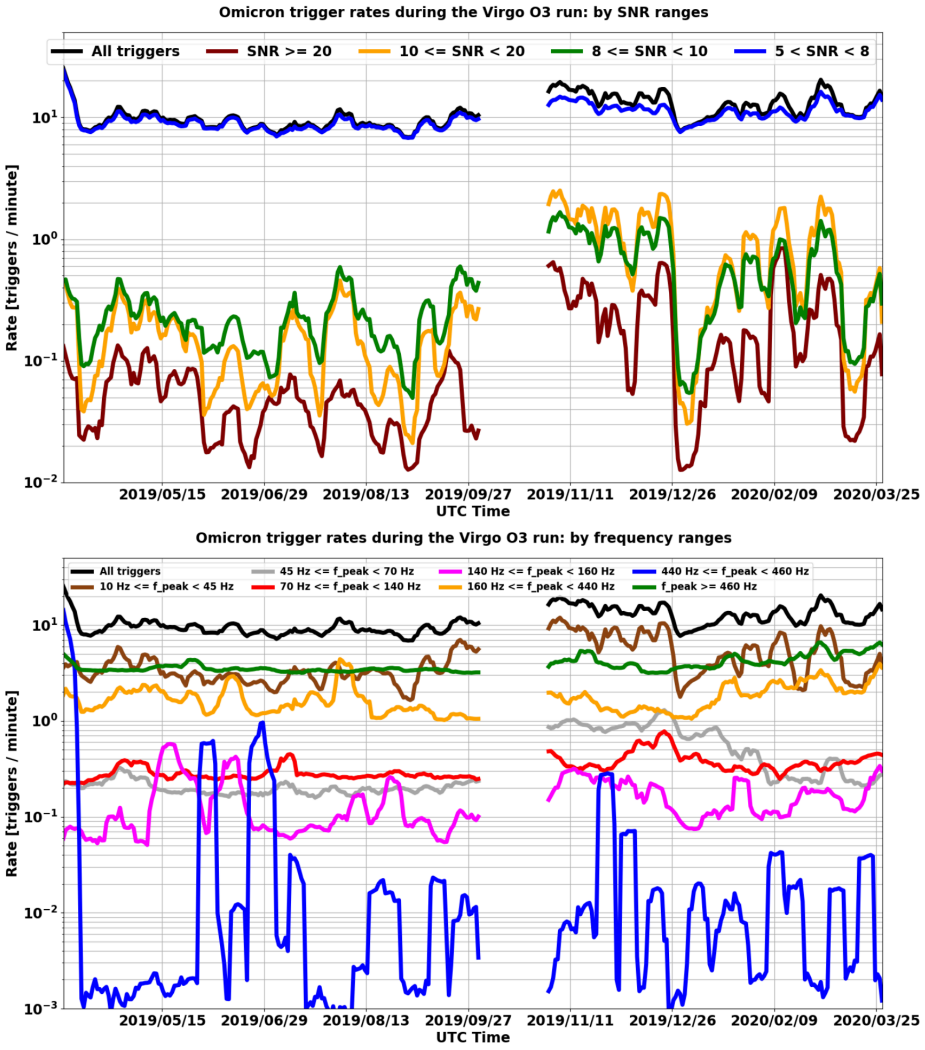


Figure 3.8: Virgo glitch rate from Omicron triggers during O3 run excluding CAT1-vetoed segments and highlighting glitch SNR ranges (top) and frequency ranges (bottom) [6]. Curves show daily averages. The gap separates O3a and O3b with a 1-month commissioning break.

Analyzing Gravity Spy dataset during O3[5], it is possible to explore correlations between glitch classes and information from Omicron triggers (3.9). The most common glitches are Scattered Light, Low Frequency Busts and Lines, and Whistle, characterized by low SNR. Duration of these glitches is about a few seconds. Whistle is the only glitch class with an high occurrence rate exhibiting peak frequency at around 500 Hz, but also lower amplitude compared to low-frequency ones. Bandwidth is not so informative in order to characterize and distinguish between differen glitch types, but it could depend on how Omicron runs specifically on Virgo. Overall, there are no significant differences in glitch characteristics between O3a and O3b, apart from the No Glitch class that always shows no clear behavior because of its definition.

Comparing with LIGO detectors (Figures 3.10 and 3.11), it is clear that differences in instrumentation, design and environmental condition strongly affect noise behavior in the strain channel. Hanford detector suffers from low-frequency and low-SNR glitches, but also from Koi Fish glitches, that can exhibit an SNR close to 100. Blip, Whistle and Tomte presence is also non negligible. There are also some difference between the first and the second part of O3: 1080 Lines glitch compares only

[5]Publicly available here: https://zenodo.org/records/5649212

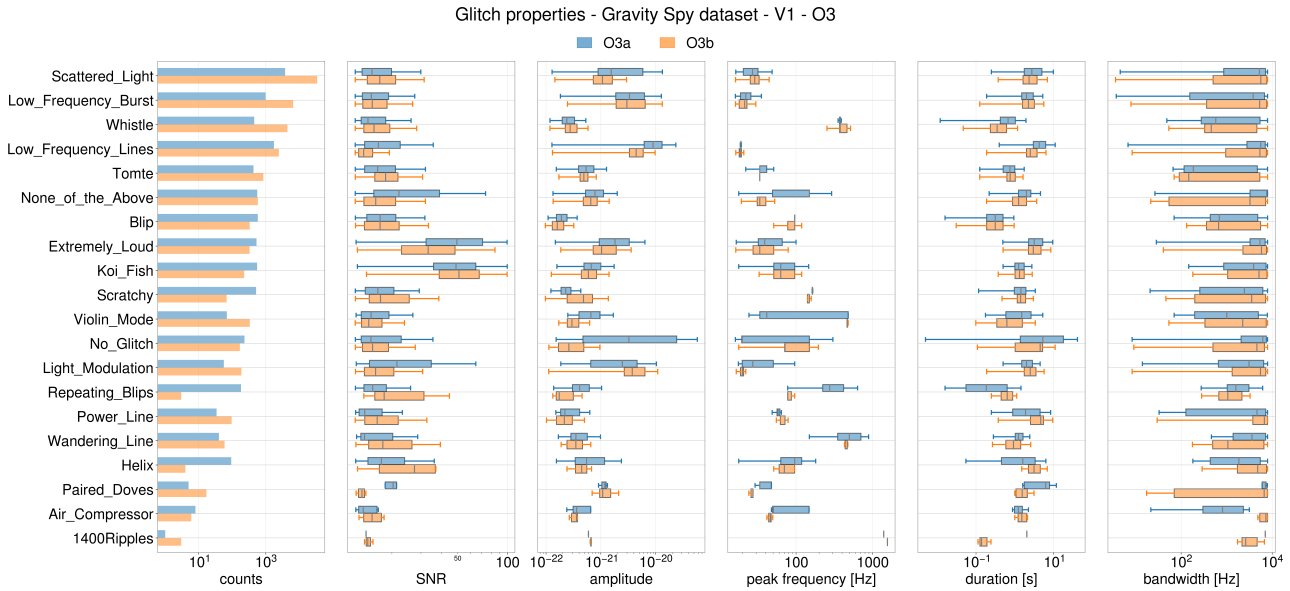


Figure 3.9: Box plot for Virgo (V1) O3 glitches properties using Omicron triggers additional information. Glitches data are taken from Gravity Spy and glitches are filtered by SNR in $[12, 100]$.

in O3b while Wandering Line only in O3a, Power Line glitch SNR range in O3a is lower compared to O3b, Scratchy glitch SNR range becomes wider and higher in O3b, meanwhile Fast Scattering SNR becomes lower in O3b. In Livingston detector, a lot of glitch types appear more than 1,000 times, differently from Virgo and Hanford. Overall, it is noticeable that the same glitch class as labelled by Gravity Spy show different characteristics depending on which detector: for example, Whistle glitches, in Virgo and H1 have an SNR on average less than 20, in L1 an SNR on 30 on average and covering a larger SNR range. Whistle peak frequency is at around 1 kHz in LIGO detectors, differently from Virgo, where they usually last longer too.

The implemented strategy for building the training dataset follows therefore the schema described in Figure 3.12, that slightly differs from the one used for LIGO data in [20], since here all the information provided by Omicron is exploited to retrieve cleaner data. Virgo is noisier compared to LIGO detectors, therefore, during the training phase, the model must be helped in distinguishing between a noise background with the detector's characteristics and a noise background contaminated by for example glitch tails, which could quite impact a 2 s data segment. First of all, after retrieving Omicron triggers for O3 from the Virgo archive, they must be filtered considering the operational mode of the detector. Only triggers inside data segments flagged with the `V1:ITF_SCIENCE` flag are retained to ensure data quality from nominal operational conditions. Then, to extract background data samples, considering two subsequent Omicron trigger, if the time stretch between the end GPS time of the first trigger and the start GPS time of the second is longer than 16 s and shoter than 500 s, then 14 s segments are extracted. The number of 14 s segments depends on the duration of the time stretch between triggers. Each 14 s segment is whitened using the method implemented in the `PyCBC` library with the PSD estimated over the same segment, and then cropped to avoid spectral artifacts at both ends by 2 s. The 10 s segment obtained in this way is divided in 2 s segments, each of which is rescaled to the mean and discarded if the Q-transform contains values greater than 20 to filter out any residual excess power that could bias the background noise distribution. The background samples are selected almost equally from the first and the second part of O3 run to avoid possible biases in the training

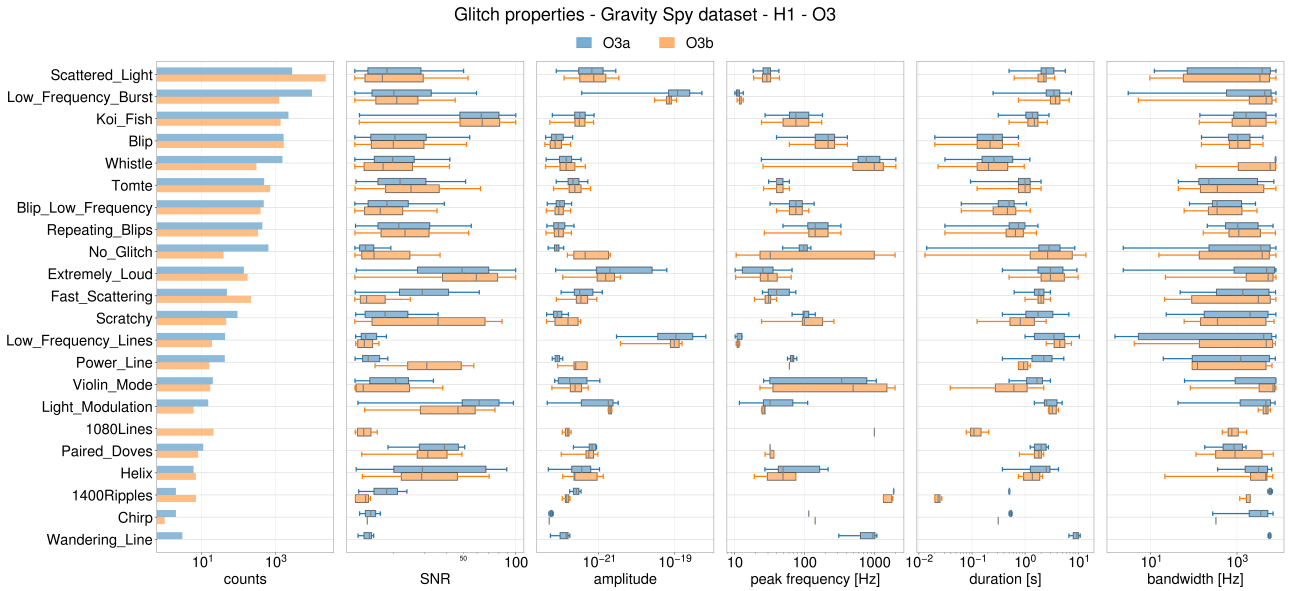


Figure 3.10: Box plot for H1 O3 glitches properties using Omicron triggers additional information. Glitches data are taken from Gravity Spy and glitches are filtered by SNR in $[12, 100]$.

distribution.

Having the background samples, glitch injections can be performed. The considered glitch classes are reported in Figure 3.2, moreover, `gengli` glitches are included too. Each glitch duration is randomly selected from a uniform distribution in $[0.125, 2]$ s, the number of glitch injections is randomly selected from 1 to 30, the glitch SNR is uniformly selected between 1 and 250 and each glitch is randomly injected in the 2 s segment. Details for glitch parameters are given below.

- Chirp: the chirp glitch is generated using the signal package of `SciPy` library, starting and ending frequency are randomly selected from a uniform distribution in the frequency range of $[1, 2048]$ Hz, where 2048 Hz is the Nyquist frequency since the sampling rate is 4096 Hz.

- Sine: the sine frequency is selected from a uniform distribution in the frequency range of $[1, 2048]$ Hz.

- Sine-Gaussian: the sine frequency is selected from a uniform distribution in the frequency range of $[1, 2048]$ Hz, while the Gaussian curve width τ is uniformly selected in the range $[(glitch\,duration)/200, (glitch\,duration)/4]$ s.

- Gaussian pulse: the minimum center frequency is uniformly selected in $[1, 2048]$ Hz, other parameters are uniformly selected in the default range given by the function from signal package of `SciPy` library.

- Ringdown: the analytic expression is the following

$$h(t) = A \exp^{-\frac{t-t_0}{\tau}} \sin\left[2\pi f_0 (t - t_0) + \phi\right] \tag{3.5}$$

  where $A = 1$ is the amplitude, f_0 is the starting frequency chosen uniformly in $[10, 2048]$ Hz, t_0 is the starting time chosen uniformly in $[(glitch\,duration)/4, 3(glitch\,duration)/4]$ s, and the phase ϕ is uniformly selected in $[0, 2\pi]$. Moreover the waveform has a 50% chance to flip, making

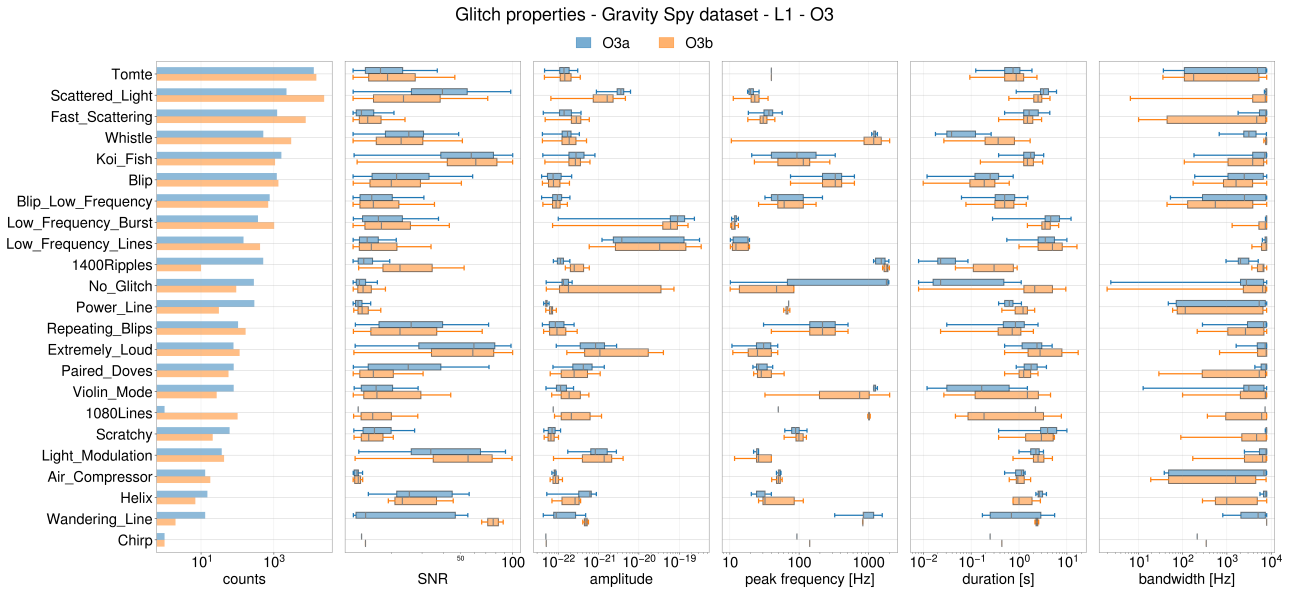


Figure 3.11: Box plot for L1 O3 glitches properties using Omicron triggers additional information. Glitches data are taken from Gravity Spy and glitches are filtered by SNR in $[12, 100]$.

a so-called "ringup".

- `gengli` glitch: L1 or H1 generated Blip glitches with given duration.

After the glitch injections, the dataset with synthetic glitches injected in real noise background samples is split in training and validation sets, with the proportion training:validation being 5:1. The `StandardScaler` is then fitted to the training set only to avoid information leakage in the validation set. Next, a time reversal data augmentation strategy has been adopted for the training set only. Finally, magnitude and phase spectrograms are generated and DeepExtractor input training and validation data are ready. The obtained training dataset size is 49,188, meanwhile the validation dataset size is 6,148. To complete the transfer learning phase, it took 9 additional epochs.

3.2.2 Validation data and tools

A validation study is necessary to verify that the background samples reconstructed by DeepExtractor are physically consistent, meaning there are no artifacts and no excess power is introduced arbitrarily. The study will also investigate DeepExtractor's limitations and ensure that the results of deep learning can be interpreted from a physics perspective. These two factors are essential for developing a data analysis tool that can be incorporated into the official production pipelines of the LVK collaboration. The goal of the sudy is indeed characterize the DeepExtractor for Virgo model and to validate it as a tool that can be used in physics analysis.

To test the generalization capabilities of the Virgo DeepExtractor model, real Virgo glitch samples have been retrieved by the Gravity Spy O3 dataset. Figure 3.13 how validation data is retrieved: starting from Gravity Spy Virgo O3 dataset entries, they have been filtered by SNR in $[12, 100]$ to have a defined reference SNR range and to be able to assess how well in those boundaries DeepExtractor can perform. Considering then the glitch GPS times, that correspond to the peak GPS times identified by Omicron, a 14 s segment centered on the glitch GPS time is extracted and whitened (with `PyCBC` method) with the PSD calculated using the 14 s segment before the one with the glitch. The segment

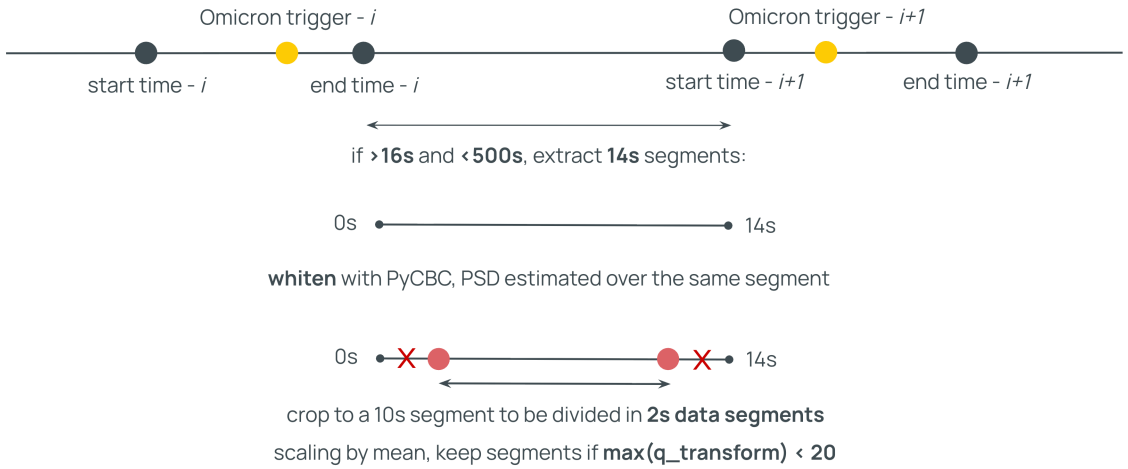


Figure 3.12: Scheme illustrating how to retrieve real background samples to later obtain training and validation datasets. Raw data are taken from GWOSC and are already preconditioned. The stretch between the end of an Omicron trigger and the start of the following one should be greater that 16 s and smaller than 500 s. From this stretch, 14 s segments are extracted, whitened with `PyCBC` library method with the PSD estimated over the same segment. The segment is cropped at both ends to avoid whitening artifacts obtaining a 10 s segment that is then divided in 2 s segments. Finally, the mean value of the segments was subtracted from each segment. Only 2 s segments were kept for the training set if their Q-transform did not contain excess power (normalized energy under 20).

containing the glitch is then cropped to a 2 s segment centered on the glitch GPS time and scaled accordingly to ths `StandardScaler` parameters found with the fit on training data. Finally, magnitude and phase spectrograms are produced. Notice that, in order to avoid glitch tails in the segment used to compute the PSD, the glitch duration must be smaller than 7 s and the distance between subsequent glitches must be greater than 28 s. The obtained dataset size is 11,475.

Examples of reconstructions using Gravity Spy visualization standards for Omega Scans are displayed in Figure 3.14. The glitches included here mostly belong to the most common glitch classes and, from both the time domain reconstructions and the Omega scans, DeepExtractor seems to perform very well. However, visual inspection is not enough for statistical assessments on the quality of DeepExtractor reconstructions.

Glitchfind

The tool included in this study to quantify DeepExtractor performance from a physics perspective is `glitchfind` [36]: it is a method developed to identify glitches using the Q-transform and measuring the significance of any excess power in the data. `glitchfind` is therefore useful to statistically determine whether or not the noise is Gaussian: under the assumption that the usual noise background is Gaussian, any excess power that does not follow the expected distribution can be assigned a p-value and be identified as excess power. Nowadays `glitchfind` is widely use as a production tool within LVK collaboration in data quality reports to look for glitches around a candidate event. In this case, `glitchfind` is employed to verify whether the reconstructed background samples by DeepExtractor contain excess power or not.

The `glitchfind` method first computes the Q-transform as implemented in `gwpy` library for a discretized sample: a map of tiles is thus obtained and the value of interest is the energy at each tile. The set of tiles is referred to as a Q-gram within `gwpy` framework. Instead of selecting the Q value

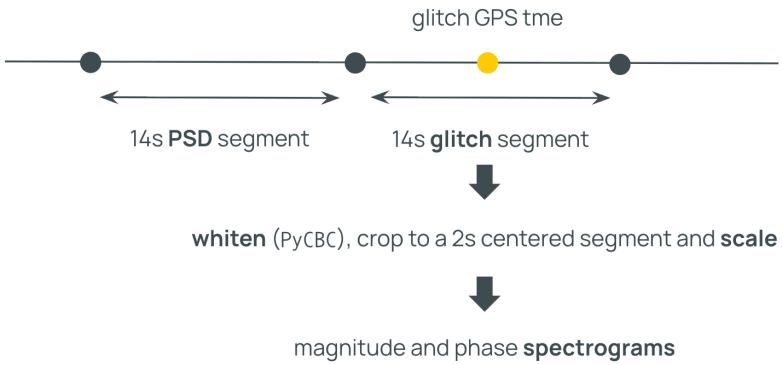


Figure 3.13: Simplified scheme illustrating how to retrieve test samples containing actual glitches. Raw data are taken from GWOSC and are already preconditioned. Glitch duration must be smaller than 7 s and distance between peak GPS times should be greater than 28 s to conservatively avoid the presence of glitches in the segment used to compute the PSD. Two 14 s segments are selected: one is centered on the peak glitch GPS time, the other covers the 14 s before the start of the glitch segment and is used to compute the PSD included in the glitch segment whitening to ensure that excess power is not suppressed. The glitch segment is then cropped to a 2 s segment centered on the peak glitch GPS time and scaled. Finally, magnitude and phase spectrogram are produced.

leading to the largest individual tile energy as naturally done with Q-transform, a value of $Q = 20$ is chosen since it is claimed to provide best structure to the plots and to allow the window function to perform in an optimal way in this purpose. To determine whether the data contains excess power, each tile is then assigned a significance under the assumption of stationary Gaussian white noise. The energies of the Q-gram are indeed expected to follow the exponential distribution given by

$$P(\epsilon > \epsilon[m,k]) \propto \exp\left(-\frac{\epsilon[m,k]}{\mu_k}\right), \tag{3.6}$$

where k corresponds to the frequency, μ_k is the mean tile energy at that frequency, and m indicates the time from the discretized version of the Q-transform. The distribution of Q energies could be fitted following the exponential distribution according to

$$P(\epsilon) = Ae^{-\epsilon t}, \tag{3.7}$$

where ϵ is the tile energy and A and t are the fit parameters, normalization factor and exponential rate. For each tile, the probability that its energy is greater than a given threshold energy ϵ_0 is given by

$$\tilde{P}(\epsilon > \epsilon_0) = \int_{\epsilon_0}^{\infty} P(\epsilon')d\epsilon' = \int_{\epsilon_0}^{\infty} \frac{Ae^{-\epsilon' t}}{N} d\epsilon' = \frac{Ae^{-\epsilon_0 t}}{tN}, \tag{3.8}$$

where N is the number of tiles in the Q-gram. Assuming then that the probability distribution of observing a tile energy above a certain threshold ϵ_0 is a Poisson process, the probability of observing ϵ_0 as the loudest tile in the whole Q-gram is

$$P = 1 - e^{-\lambda\tau} = 1 - \exp\left[-\left(N\frac{A}{tN}e^{-\epsilon_0 t}\right)N\right] = 1 - \exp\left[-\left(N\frac{A}{tN}e^{-\epsilon_0 t}\right)N\right], \tag{3.9}$$

where λ is the rate of occurrence (that is the expected number of tiles with energy above the threshold, assuming that the probability distribution of the number of tiles with energy above the threshold follows a Gaussian distribution) and τ is the length of data considered, meaning the number of tiles

in the Q-gram. This expression defines the p-value associated to a Q-gram.

To validate DeepExtractor reconstruction, the Q-energy distribution of reconstructed background samples is fitted following Equation 3.7 and the corresponding p-value is computed following Equation 3.9. At first, the Q-value for evaluating the Q-grams was fixed to 20, but then it was left free to vary in a range between 4 and 40 since it would be more sensible in identifying excess noise, leading to a more conservative approach from a physics point of view although less robust from a statistical perspective. Leaving Q free to vary would be more conservative since Q-transoforms are designed to point out excess power, so the algorithm can choose which Q would best highlight excess power. On the other hand, using different Q values would be less robust statistically speaking since in principle it would be necessary to evaluate all the samples with the same Q. The p-values have been computed both for DeepExtractor input and output data, keeping the same Q-value, more specifically the one obtained from the Q-gram on input data to ensure consistency in the results. Finally, to ensure even more consistency between different methods, the Q-value given by the glitch Omicron trigger has been used to evaluate the Q-gram both for the segment containing the glitch and the reconstructed background segment given by DeepExtractor. Using the same Q-value as the one given by Omicron appeared to be the most consistent approach and the results that will be referred to are obtained in this way. Another parameter that can influence the p-value results is the number of bins of the Q-energy distribution considered in the fit: `glitchfind` fits the distribution until the 15th bin, with the binning spanning from 2 to 40 in energy values. The same approach has ben followed in this study. In the Q-gram computation the SNR lower threshold can be chosen too: `glitchfind` set it to 2 and here the same approach is adopted since it is necessary to consider the "non-glitch" parts too. Finally, since the frequency range selected in the Q-gram computation could influence p-value evaluation, exploiting Omicron information, the validation analysis was tailored according to the glitch peak frequency: if the peak frequency is smaller than 50 Hz, the Q-gram is evaluated at low frequency bands (more specifically [10, 128] Hz, [10, 256] Hz, [10, 512] Hz), meanwhile if the peak frequency is greater than 50 Hz, the Q-gram is evaluated in the frequency range given by [10, 1024] Hz. The frequency threshold was chosen to distinguish the low-frequency glitches according to Figures 3.8 and 3.9.

The `glitchfind` method has been applied also on Virgo O3 background training data in four different frequency ranges to obtain a benchmark performance of `glitchfind` on pure noise samples and to observe how much the frequency range has an impact. The Q-value for evaluating the Q-gram was free to vary in the range [4,40] and the different frequency ranges considered are 10- 128 Hz, 10 - 256 Hz, 10 - 512 Hz and 10 - 1024 Hz.

Finally, `glitchfind` has also been applied to evaluate how much transfer learning on actual Virgo data was effective. The DeepExtractor model trained on simulated data was applied on real Virgo O3 glitches and the p-value was computed in order to compare the results with The DeepExtractor model for Virgo. In this case, two frequency bands are considered: if the peak frequency of the glitch is greater than 50 Hz, the Q-gram is evaluated in the [10, 1024] Hz frequency range, otherwise in the [10, 128] Hz.

Table 3.1: Results obtained by applying the `glitchfind` method on Virgo O3 background training data. Q-grams are evaluated in different frequency bands with Q-values free to vary in the range [4,40]. Here are reported the percentages of samples leading to a p-value lower than the threshold, meaning samples that according to `glitchfind` method contain non-Gaussian features.

Frequency band	% of samples with $p < 0.05$
10 - 128 Hz	38.9%
10 - 256 Hz	26.1%
10 - 512 Hz	15.6%
10 - 1024 Hz	10.1%

3.2.3 Validation results

Here follows the results from the validation study that hightlights both strenghts and limitations of DeepExtractor and also limitations of the tools used to evaluate DeepExtractor performance.

Glitchfind results

Analyzing first Table 3.1, the `glitchfind` method clearly shows some limitations: also with the largest frequency range, there is a 10% of background samples, that by construction do not contain any excess power, that obtain a p-value lower than the threshold. If a sample obtains a p-value lower than the threshold it means that it cannot be considered compatible with Gaussian noise. Moreover, if the frequency range is then restricted to a lower band, ideally for better capturing low-frequencies, approximately 40% of the background samples are flagged as non-Gaussian by `glitchfind`. These results suggest that subsequent results, from the application of `glitchfind` to DeepExtractor reconstructions in cases of actual glitches, should be considered with caution because there is the possibility of obtaining false-positives results. However, it is important to remember that DeepExtractor's 2-second time window is limited, especially when considering low frequencies, and that `glitchfind` is typically applied to larger time windows (for example, in DQRs `glitchfind` uses a 4-second window), therefore it may be not perfectly suitable for this time window.

Figure 3.15 shows the fitted Q-energy distributions corresponding to the samples in Figure 3.14. All the input samples obtained a p-value equal to zero, indicating the presence of non-Gaussianities, meanwhile all DeepExtractor reconstructions present a p-value greater than the 0.05 threshold, suggesting compatibility with the hypothesis of Gaussian noise. With the aim of better highlighting and analyzing the frequency range in which the glitch is most prominent, a lower frequency band has been used to evaluate the Q-gram for Scattered Light and Low Frequency Burst glitches since their peak frequency is lower than 50 Hz.

Table 3.2: Fraction of DeepExtractor input and output samples obtaining $p < 0.05$. If the glitch peak frequency is greater than 50 Hz, the Q-gram is evaluated in the [10, 1024] Hz frequency range, in lower frequency bands otherwise to highlight the different behavior at low frequencies. The total number of samples for each glitch class N_{tot} is reported too.

Glitch type	Freq. range [Hz]	N_{tot}	Input p<0.05 [%]	Output p<0.05 [%]
Blip	10-1024	393	100	5.60
	10-128	2	100	0.0
	10-256	2	100	0.0
	10-512	2	100	0.0
Extremely Loud	10-1024	102	100	59.8
	10-128	232	97.8	59.1
	10-256	232	97.8	57.8
	10-512	232	97.4	54.3
Helix	10-1024	32	100	37.5
	10-128	7	100	71.4
	10-256	7	100	71.4
	10-512	7	100	71.4
Koi Fish	10-1024	285	100.0	12.6
	10-128	64	98.4	31.3
	10-256	64	98.4	26.6
	10-512	64	98.4	23.4
Light Modulation	10-1024	18	94.4	11.1
	10-128	89	91.0	62.9
	10-256	89	92.1	50.6
	10-512	89	93.3	49.4
Low Frequency Burst	10-1024	4	75.0	25.0
	10-128	491	96.3	55.0
	10-256	491	95.7	44.6
	10-512	491	94.7	41.1
Low Frequency Lines	10-1024	2	50.0	0.0
	10-128	437	72.8	53.1
	10-256	437	68.2	45.5
	10-512	437	59.7	35.5
No Glitch	10-1024	107	24.3	13.1
	10-128	24	25.0	29.2
	10-256	24	16.7	37.5
	10-512	24	16.7	29.2
None of the Above	10-1024	237	85.2	46.4
	10-128	234	100	39.3
	10-256	234	100	34.6
	10-512	234	100	28.2

continues to the next page

Table 3.2 – continued

Glitch type	Freq. range [Hz]	N_{tot}	Input p<0.05 [%]	Output p<0.05 [%]
Paired Doves	10-128	9	100	44.4
	10-256	9	100	44.4
	10-512	9	100	44.4
Power Line	10-1024	60	51.7	23.3
Repeating Blips	10-1024	72	100	22.2
Scattered Light	10-1024	3	100	0.0
	10-128	442	89.8	57.9
	10-256	442	87.6	47.3
	10-512	442	85.5	42.8
Scratchy	10-1024	260	33.5	33.5
Tomte	10-1024	56	100	8.93
	10-128	414	100	20.8
	10-256	414	100	13.8
	10-512	414	100	6.28
Violin Mode	10-1024	177	92.7	36.7
	10-128	24	100	50.0
	10-256	24	100	50.0
	10-512	24	100	41.7
Wandering Line	10-1024	41	90.2	53.6
Whistle	10-1024	451	86.0	24.0

Table 3.2 reportts the fractions of both samples containing glitches (DeepExtractor input) and reconstructed background samples (DeepExtractor output) obtaining a p-value lower than the threshold of 0.05, indicating the presence of non-Gaussianities according to `glitchfind`. Notice that here glitch classes with a very high frequancy support, like 1400 Ripples and Air Compressor, have been excluded both because of their very low occurrance rate (the amount of samples would not have yielded statistically significant results) and since the frequency ranges condsidered in the application of `glitchfind` method would have excluded most of the glitch parts. Starting from the input samples, it is clear that for glitch classess exhibiting low SNR and low frequencies, the p-value test struggles to identify glitches: glitches like Low Frequency Lines, No Glitch and Scratchy are clearly problematic to be identified no matter the frequency range. This result emphasizes once again how difficult it is to detect glitches under low SNR and low frequency conditions. Another limiting factor that certainly contributes to this result is DeepExtractor's analysis framework, which considers only 2-second time windows. Considering then results regarding reconstructed backgrounds, DeepExtractor performance is highly satisfactory overall, taking also into account the already mentioned limitations. According to `glitchfind`, the glitch categories whose backgrounds are best reconstructed are Blip, Tomte, Koi Fish and Whistle too. Meanwhile, results for glitch classes characterized by low frequency features,

such as Low Fequency Burst, Low Frequency Line and Scattered Light, do not seem good and this adds to the already and well-known difficulty of identifying and reconstructing low-frequency glitches. Moreover, it is important to keep in mind that `glitchfind` can produce false positives, especially at these frequencies. Generally, restricting the frequency band leads to a greater fraction of samples flagged as not Gaussian: the original intent behind this strategy was to better focus the glitch within its frequency band, but the downside is that the Q-gram's energy distribution does not account for the Gaussian features of the noise found at high frequencies in correspondence of the glitch. This ends up in considering too few data for the evaluation of Gaussian noise. In addition, DeepExtractor seems to struggle in recostructing Extremely Loud glitches and this is probably due to the extremely high SNR and large bandwidth, combined with a quite long duration, characterizing this class: these features cause the time-frequency window to be heavily dominated and contaminated by the glitch, preventing DeepExtractor from reconstructing the background realistically. Figure 3.16 shows an exceptionally bad reconstruction of an Extremely Loud glitch. It is clearly evident that there is no Gaussian features to reconstruct because the glitch is strongly dominating the segment. Moreover, magnitude and phase output spectrograms are smoothed and averaged out, meaning that in those cases the model finds a sample very distant from the distribution it has learnt and the only thing it can do is an average, that in this case is also equal to zero. However, a bad reconstruction like this one is very rare and DeepExtractor is capable of good reconstructions (in terms of `glitchfind` result) also in case of high SNR and very large bandwidth. The bad reconstruction example is shown only to highlight the peculiar behavior of the model in this case. It is then important to highlight that if both the input and the output samples obtain a p-value greater than the 0.05 thresold, glitch mitigation with DeepExtractor still happens, as it is clear from Figure 3.17. Here, the Q-scan for the input sample shows excess power in correspondence of the peak frequency given by Omicron, but the associated p-value does not point out non-Gaussianities. It is another clear evidence that the `glitchfind` method could produce non reliable results, also in case of glitches without strong low frequency support. However, DeepExtractor is capable of glitch mitigation, as it is visible both from the Q-scan for the reconstructed background and from the output magnitude spectrogram: excess power that was present in the input is no longer there. Moreover, DeepExtractor could particularly benefit the characterization of glitches belonging to the No Glitch class. Figure 3.18 shows an example in which DeepExtractor performs glitch mitigation in case of a No Glitch. Unlike Q-scans, the magnitude input spectrograms shows some excess power corresponding to the peak frequency given by Omicron. The Q-energy distribution of the reconstructed background does not include any bins at higher energies, which are present in the distribution for the input: this is evidence of glitch mitigation. DeepExtractor application to No Glitch class would allow to better characterize those glitches by filling the gap left by Gravity Spy when Q-scans do not show evidence of excess power.

Finally, examining Table 3.3, that reports the `glitchfind` evaluation of the effectiveness of transfer learning, it is clear that fine tuning on real data greatly improved DeepExtractor performances overall. For all classes the number of reconstructed backgrounds with DeepExtractor for Virgo obtaining $p < 0.05$ is generally lower compared to the base DeepExtractor case, demonstrating the effectiveness of transfer learning in capturing real data features that leads to better reconstructions. However, glitch classes characterized in general by low-frequency features show an improvement only in the largest frequency band and not in the [10, 128] Hz range. This result may be due to the Q-gram's energy

distribution that in this frequency range does not account for the Gaussian features of the noise, that may be instead present at high frequencies allowing to evaluate the distribution as Gaussian. Shortly, using a too restricted low frequency range may weight too much low frequency features for a very short time window as the 2 s segment. The only glitch class that shows a worsening with transfer learning is Paired Doves: this type of glitch is characterized in Virgo by extremely low SNR and low frequencies, therefore it could be very difficult to characterize. Moreover, Paired Doves glitches appear very rarely compared to other glitch types, and the number of samples in this class considered in this study is so low that it cannot yield statistically robust results. Overall, a larger frequency band like [10, 1024] Hz seems more robust since it allows for a better capture of what is happening in strain data, providing a bigger picture. Also, transfer learning on real data has been demonstrated effective once again: this strategy potentially allows to develop a DeepExtractor model for each detector and for each scientific run (or part of it) in order to better capture the specific features determined by the conditions of that detector at that moment. This is of paramount importance since, just as glitch types and features vary over time, background noise also changes over time depending on environmental conditions and upgrades carried out on the detector, and DeepExtractor could keep up this pace.

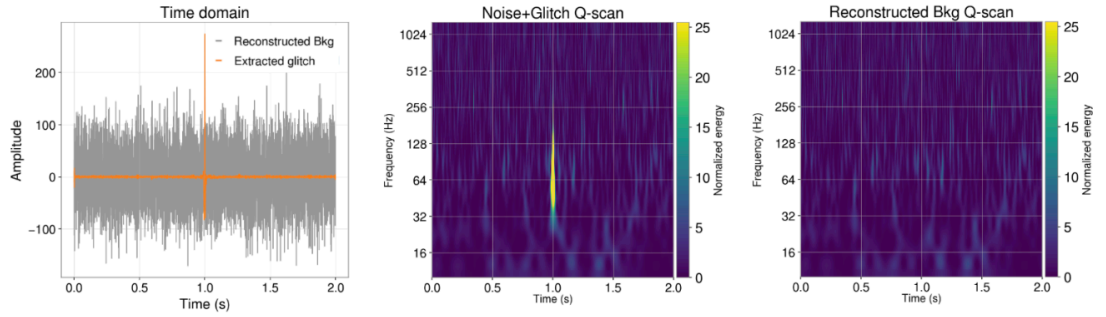
Table 3.3: Fraction of DeepExtractor output samples obtaining $p < 0.05$ with the `glitchfind` test, both in case of the base model trained on simulated data and of the fine-tuned model on real Virgo O3 data. If the glitch peak frequency is greater than 50 Hz, the Q-gram is evaluated in the [10, 1024] Hz frequency range, in lower frequency band [10, 128] Hz otherwise.

Glitch type	Freq. range [Hz]	DE base output p<0.05 [%]	DE Virgo output p<0.05 [%]
Blip	10-1024	25.6	7.08
	10-128	0.0	0.0
Extremely Loud	10-1024	77.3	56.4
	10-128	61.4	54.2
Helix	10-1024	46.9	31.3
	10-128	57.1	28.6
Koi Fish	10-1024	42.2	13.0
	10-128	33.8	28.2
Light Modulation	10-1024	50.0	16.7
	10-128	38.2	60.7
Low Frequency Burst	10-1024	55.6	22.2
	10-128	41.5	55.8
Low Frequency Lines	10-1024	90.0	20.0
	10-128	43.9	55.5
No Glitch	10-1024	36.5	18.7
	10-128	33.3	20.8
None of the Above	10-1024	81.5	38.1
	10-128	31.7	37.7
Paired Doves	10-128	55.6	88.9

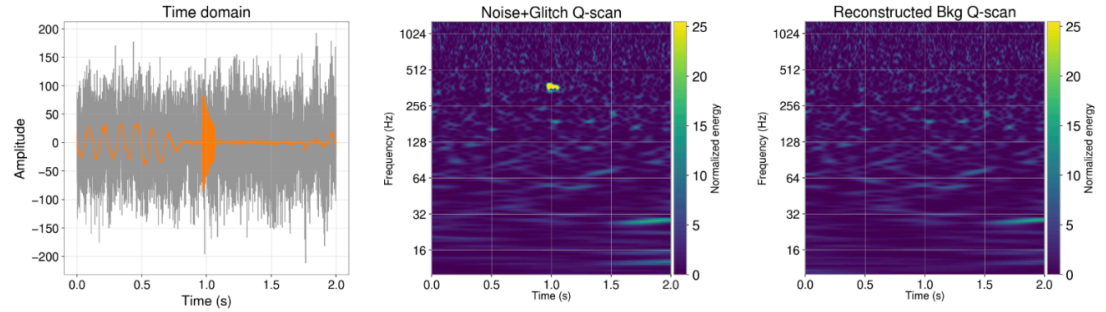
continues to the next page

Table 3.3 – continued

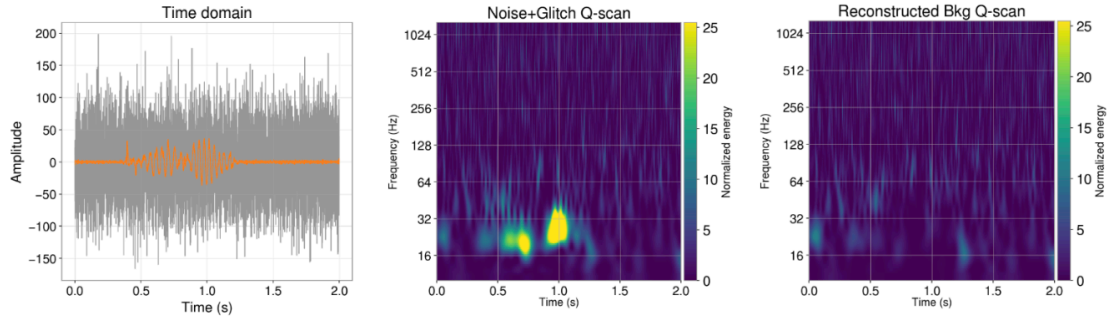
Glitch type	Freq. range [Hz]	DE base output p<0.05 [%]	DE Virgo output p<0.05 [%]
Power Line	10-1024	54.1	21.3
Repeating Blips	10-1024	37.5	20.8
Scattered Light	10-1024	66.7	41.7
	10-128	47.0	60.8
Scratchy	10-1024	60.4	34.1
Tomte	10-1024	29.5	13.1
	10-128	27.8	18.8
Violin Mode	10-1024	65.5	36.7
	10-128	37.5	37.5
Wandering Line	10-1024	85.4	43.9
Whistle	10-1024	71.83	23.6



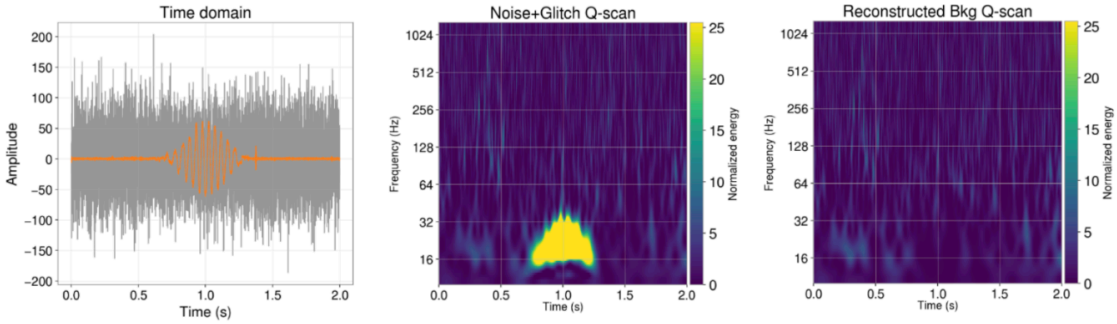
(a) Blip



(a) Whistle

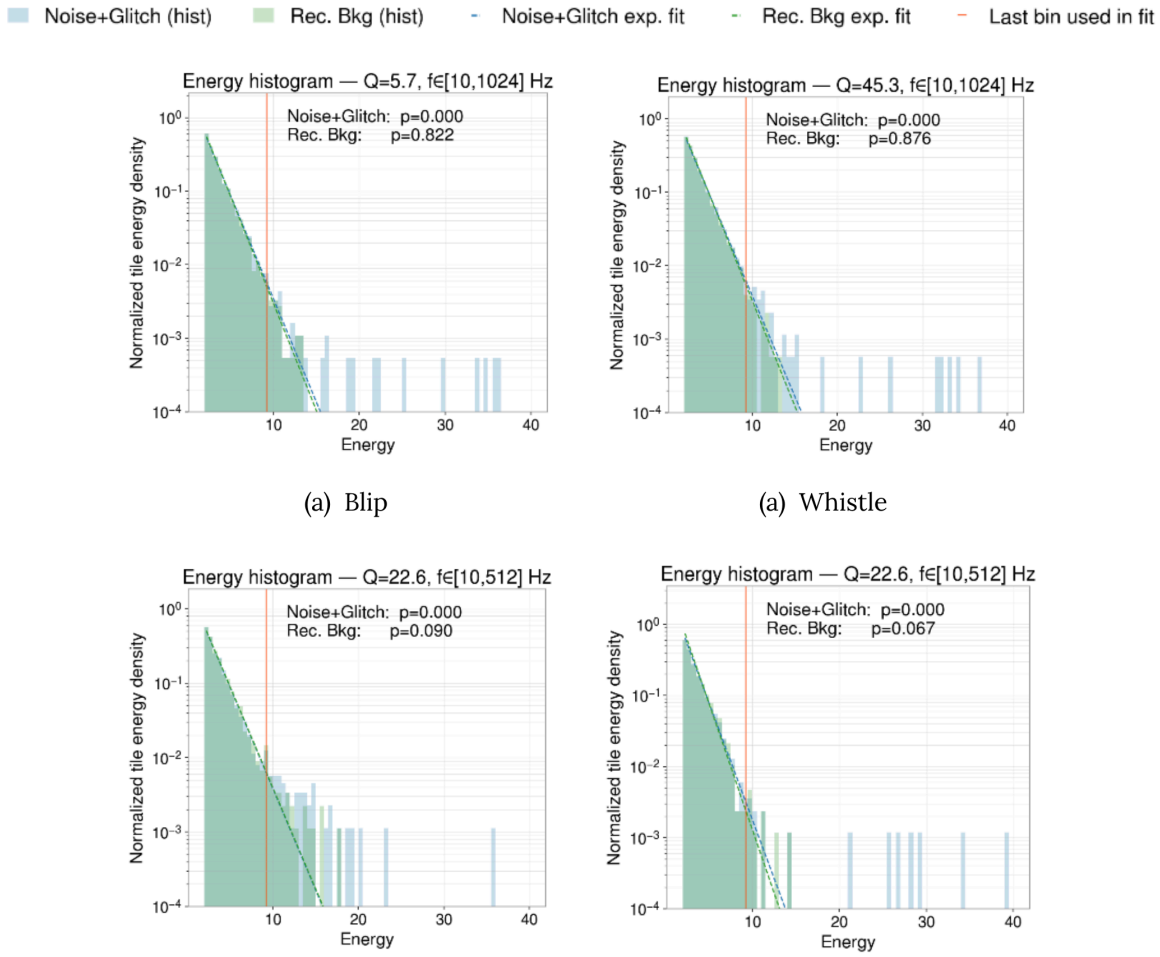


(c) Scattered Light



(d) Low Frequency Burst

Figure 3.14: Examples of glitch reconstructions from real Virgo O3 glitches. On top (a) the Blip glitch appeared during O3a with a duration of 0.250 s, a peak frequency of 96 Hz and an SNR of 16. In the second row (b) the Whistle glitch appeared during O3b with a duration of 0.141 s, a peak frequency of 370 Hz and an SNR of 15. The Scattered Ligth glitch (c) appeared during O3b with a duration of 4.125 s, a peak frequency of 25.7 Hz and an SNR of 15. On the bottom (d), the Low Frequency Burst glitch appeared during O3b with a duration of 2 s, a peak frequency of 21.2 Hz and an SNR of 25. The spectrograms (Q-scans or Omega scans) are displayed here according to Gravity Spy standards: the Q-value is allowed to vary in a range from 4 to 64, the frequency range is set to [10, 1290] Hz, the frequency resolution is 0.5 Hz and the time resolution is 2 ms. Moreover, the Q-transform for visualizing the spectrogram is computed over the 10 s segment from which the glitch was subtracted in order to avoid distorsions and the spectrogram was then cropped to a 2 s duration.



(a) Blip (a) Whistle

(c) Scattered Light (d) Low Frequency Burst

Figure 3.15: Q-energy distributions for samples in Figure 3.14 fitted following `glitchfind` method. Both DeepExtractor input (real Virgo O3 glitch) and output (reconstructed background) Q-energy distributions are showed and fitted. The Q-value used for evaluating the Q-gram is the one given with the Omicron trigger of the glitch. For Scattered Light (c) and Low Frequency Burst (d) glitches a lower frequency band has been used to evaluate the Q-gram since their peak frequency is lower than 50 Hz.

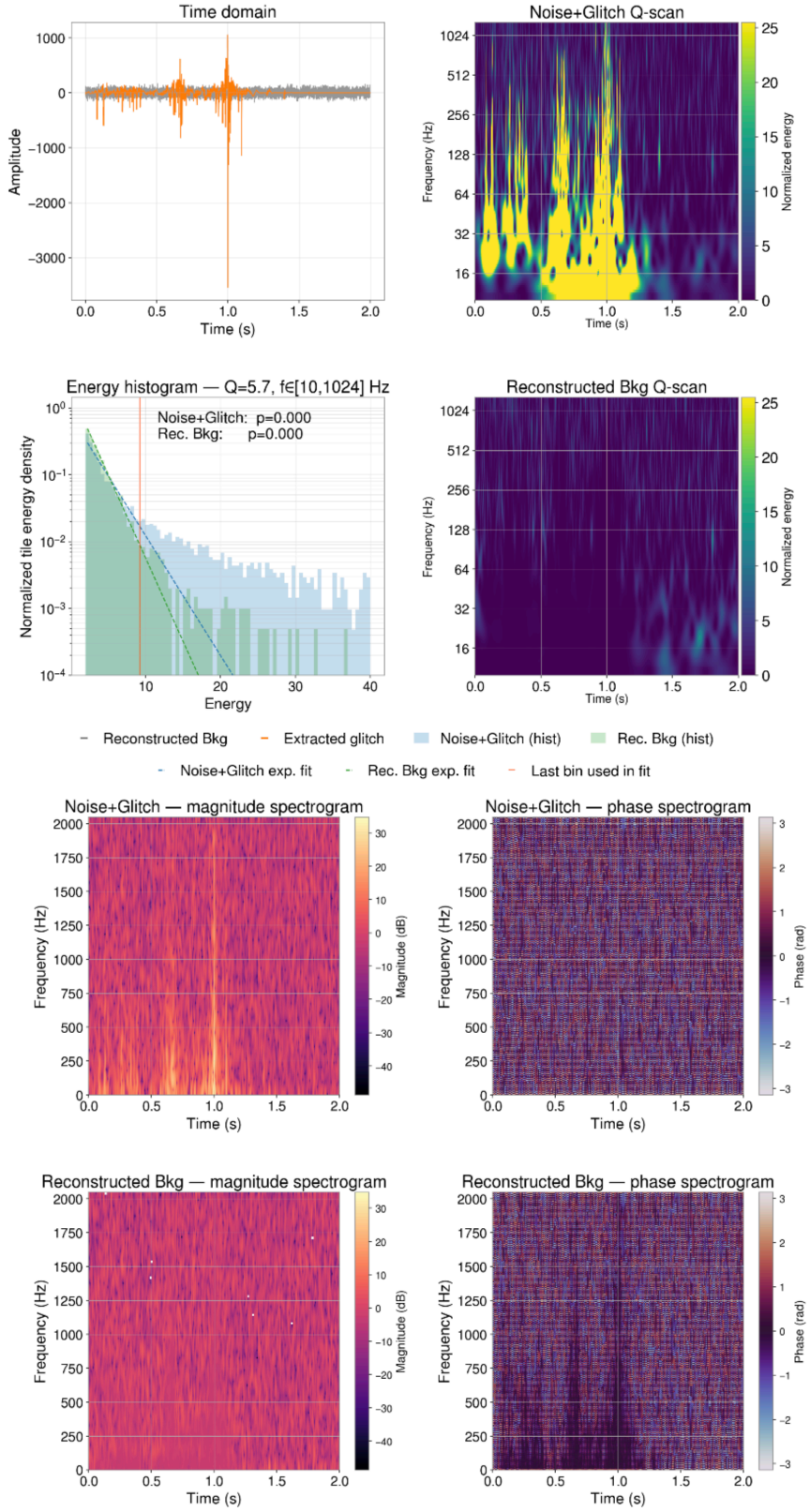


Figure 3.16: Example of failed glitch mitigation in case of an Extremely Loud glitch. It appeared during O3a in Virgo with an SNR of more than 94, a duration of 3 s, and a peak frequency of 223 Hz. In the time domain reconstruction, the glitch shows an extremely large amplitude and both from the Q-scan and the Q-energy distribution is clear how much the glitch is permeating the time-frequency plane. From output phase and magnitude spectrograms it is evident that DeepExtractor is trying to average in the reconstruction, compensating also with the phase, that in input showed no clear issues.

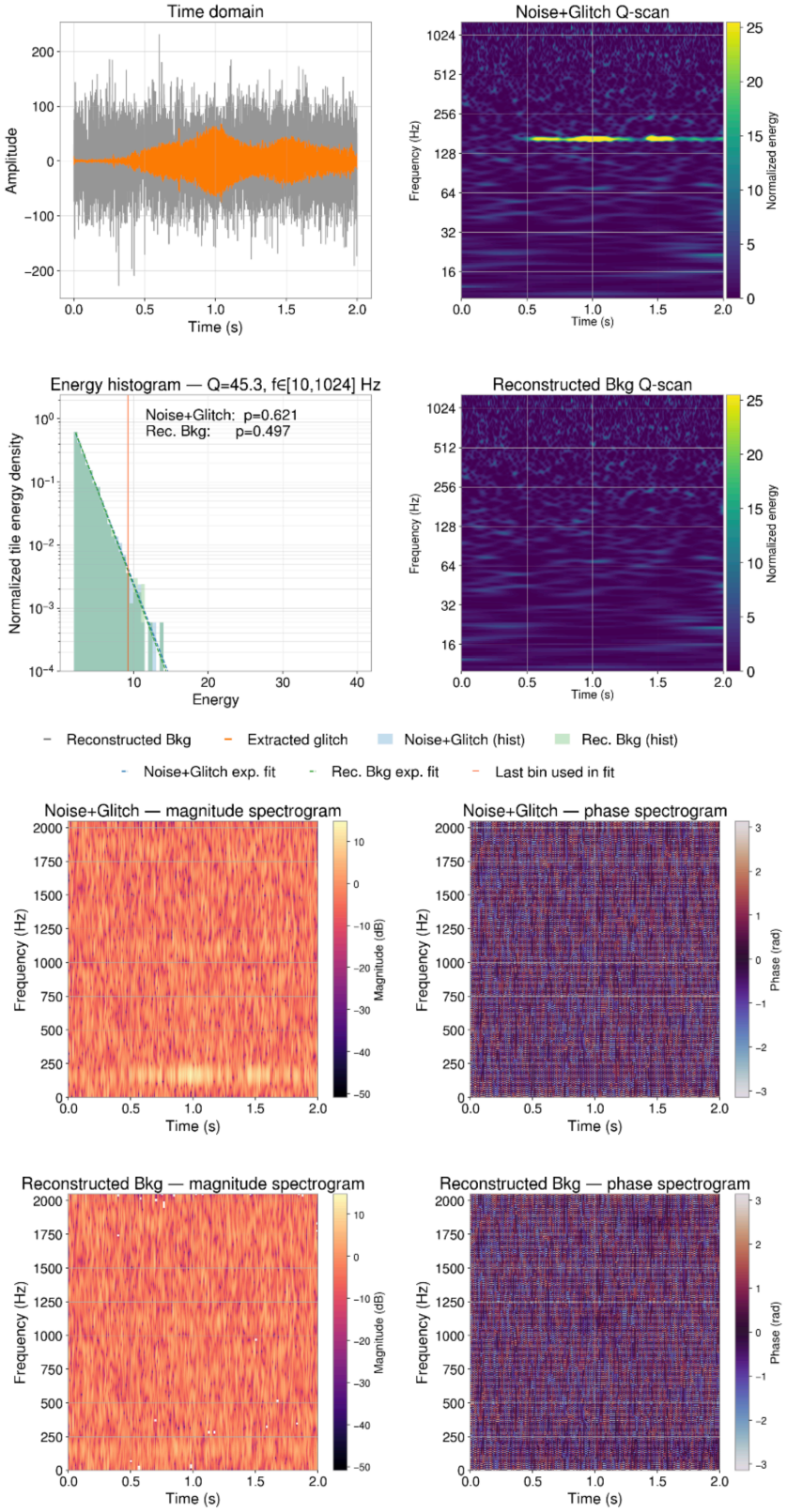


Figure 3.17: Example of successful glitch mitigation in case of a Scratchy glitch. This glitch appeared in Virgo during O3a and lasted approximately 2 s, exhibiting a peak frequency of 164 Hz and an SNR of 19. In this case the p-values obtained both for the input and the output samples are above the thresold of 0.05, even though the Q-scan for the input sample is showing evidence of excess power in correspondence of the peak frequency. DeepExtractor is however capable of glitch mitigation.

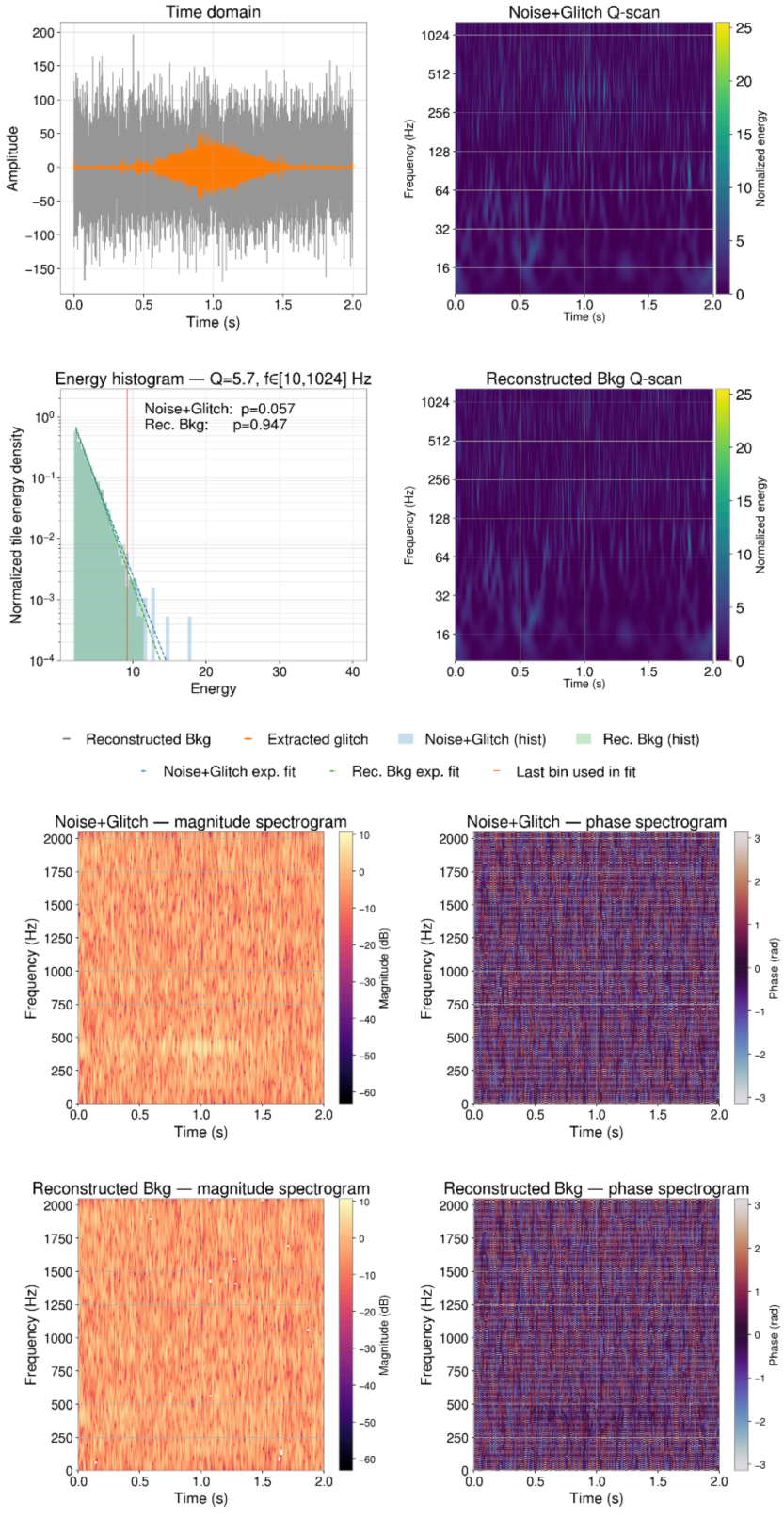


Figure 3.18: Example of successful glitch mitigation in case of a No Glitch glitch. It appeared in Virgo during O3a, lasted 0.6 s with a peak frequency of 419 Hz and an SNR of 12. In this case the Q-scan for the input sample seems clear, showing no evident excess power, but in the magnitude spectrogram some persistent noise around the peak frequency is clearly evident. The p-values obtained both for the input and the output samples are above the thresold of 0.05, but the Q-energy distribution of the reconstructed background does not include any bins at higher energies, which are present in the distribution for the input. This is evidence of glitch mitigation. DeepExtractor application to No Glitch class would allow to better characterize those glitches by filling the gap left by Gravity Spy when Q-scans do not show evidence of excess power.

Conclusions

GWs are a very powerful tool to investigate the Cosmos and to explore its still-dark corners, as well as offering a testing ground for General Relativity and other theories of gravity. In this perspective, data analysis of GW signals is essential: ML tools can offer a great potential and many benefits for GW data analysis, especially for next generation detectors, however, their non trivial interpretability and reliability for scientific results must be carefully assessed. This study focuses exactly on this point, by covering the development of DeepExtractor, a promising deep learning tool, and by checking its consistency with the physics operating behind the data.

The DeepExtractor model tailored for Virgo detector was successfully obtained by applying transfer learning on the model trained on simulated data. DeepExtractor for Virgo was therefore tested on real data containing glitches from Virgo O3 run and its reconstructions were evaluated with `glitchfind`. The validation study mainly highlighted great reconstruction capabilities of DeepExtractor and limitations of the `glitchfind` method. Results also pointed out that identifying and mitigating glitches at low frequency and low SNR could be very challenging, and Bayes Wave, the state of the art algorithm, already struggles a lot in these conditions. It can only be stated that DeepExtractor overall performs proper glitch mitigation, however the quality of its results cannot be assessed conclusively through `glitchfind`, especially at low frequency, given the limitations of the methods and DeepExtractor framework.

Given these outcomes, in the future it should be explored also the application of the `glitchfind` method to a longer stretch of data. A further step would be to check DeepExtractor reconstructions with Omicron algorithm and analyze the produced triggers. In particular, it would be particularly useful to analyze both DeepExtractor input and output with Omicron and verify that Omicron produces a trigger for the input glitch that should be removed in the reconstruction. This analysis could be helpful also in testing whether DeepExtractor arbitrarily introduces excess power in its reconstructions. However, the analysis should take into account the different signal processing strategies of DeepExtractor and Omicron: for example, a 2 s segment is too short to be processed by Omicron. Working on the alignment of DeepExtractor's preprocessing and training strategy with other pipelines used by LVK collaboration could certainly improve reconstruction accuracy.

Bibliography

[1] B P Abbott et al. "A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals". In: *Classical and Quantum Gravity* 37.5 (Mar. 2020), p. 055002. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/ab685e. URL: https://iopscience.iop.org/article/10.1088/1361-6382/ab685e (visited on 12/15/2025).

[2] B. P. Abbott et al. "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral". en. In: *Physical Review Letters* 119.16 (Oct. 2017), p. 161101. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.119.161101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.119.161101 (visited on 12/15/2025).

[3] B. P. Abbott et al. "Observation of Gravitational Waves from a Binary Black Hole Merger". en. In: *Physical Review Letters* 116.6 (Feb. 2016), p. 061102. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.116.061102. URL: https://link.aps.org/doi/10.1103/PhysRevLett.116.061102 (visited on 04/24/2026).

[4] R. Abbott et al. "GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo during the First Half of the Third Observing Run". en. In: *Physical Review X* 11.2 (June 2021), p. 021053. ISSN: 2160-3308. DOI: 10.1103/PhysRevX.11.021053. URL: https://link.aps.org/doi/10.1103/PhysRevX.11.021053 (visited on 09/13/2026).

[5] R. Abbott et al. "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run". en. In: *Physical Review X* 13.4 (Dec. 2023), p. 041039. ISSN: 2160-3308. DOI: 10.1103/PhysRevX.13.041039. URL: https://link.aps.org/doi/10.1103/PhysRevX.13.041039 (visited on 12/15/2025).

[6] F. Acernese et al. "Virgo Detector Characterization and Data Quality: results from the O3 run". In: *Classical and Quantum Gravity* 40.18 (Sept. 2023). arXiv:2210.15633 [gr-qc], p. 185006. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/acd92d. URL: http://arxiv.org/abs/2210.15633 (visited on 12/15/2025).

[7] Sofia Alvarez-Lopez et al. *GSpyNetTree: A signal-vs-glitch classifier for gravitational-wave event candidates*. arXiv:2304.09977 [gr-qc]. Apr. 2023. DOI: 10.48550/arXiv.2304.09977. URL: http://arxiv.org/abs/2304.09977 (visited on 09/17/2026).

[8] Patrick Brady, Giovanni Losurdo, and Hisaaki Shinkai. "LIGO, VIRGO, and KAGRA as the International Gravitational Wave Network". en. In: *Handbook of Gravitational Wave Astronomy*. Ed. by Cosimo Bambi, Stavros Katsanevas, and Konstantinos D. Kokkotas. Singapore: Springer Singapore, 2021, pp. 1–21. ISBN: 978-981-15-4702-7. DOI: 10.1007/978-981-15-4702-7_51-1. URL: https://link.springer.com/10.1007/978-981-15-4702-7_51-1 (visited on 04/24/2026).

[9] Jonas Buchli et al. "Improving cosmological reach of a gravitational wave observatory using Deep Loop Shaping". In: *Science* 389.6764 (Sept. 2025). arXiv:2509.14016 [astro-ph.IM], pp. 1012–1015. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.adw1291. URL: http://arxiv.org/abs/2509.14016 (visited on 09/17/2026).

[10] E. Capote et al. "Advanced LIGO detector performance in the fourth observing run". In: *Physical Review D* 111.6 (Mar. 2025). arXiv:2411.14607 [gr-qc], p. 062002. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.111.062002. URL: http://arxiv.org/abs/2411.14607 (visited on 09/11/2026).

[11] The LIGO Scientific Collaboration et al. *GWTC-4.0: Updating the Gravitational-Wave Transient Catalog with Observations from the First Part of the Fourth LIGO-Virgo-KAGRA Observing Run*. arXiv:2508.18082 [gr-qc]. Sept. 2025. DOI: 10.48550/arXiv.2508.18082. URL: http://arxiv.org/abs/2508.18082 (visited on 12/15/2025).

[12] The LIGO Scientific Collaboration et al. *GWTC-5.0: An Introduction to Version 5.0 of the Gravitational-Wave Transient Catalog*. arXiv:2605.27223 [gr-qc]. June 2026. DOI: 10.48550/arXiv.2605.27223. URL: http://arxiv.org/abs/2605.27223 (visited on 09/11/2026).

[13] The LIGO Scientific Collaboration et al. "Open Data from LIGO, Virgo, and KAGRA through the First Part of the Fourth Observing Run". In: *The Astrophysical Journal* 1004.2 (June 2026). arXiv:2508.18079 [gr-qc], p. 232. ISSN: 0004-637X, 1538-4357. DOI: 10.3847/1538-4357/ae211e. URL: http://arxiv.org/abs/2508.18079 (visited on 09/14/2026).

[14] Neil J. Cornish and Tyson B. Littenberg. "BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches". en. In: *Classical and Quantum Gravity* 32.13 (July 2015). arXiv:1410.3835 [gr-qc], p. 135012. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/0264-9381/32/13/135012. URL: http://arxiv.org/abs/1410.3835 (visited on 12/15/2025).

[15] Elena Cuoco et al. "Applications of machine learning in gravitational-wave research with current interferometric detectors". en. In: *Living Reviews in Relativity* 28.1 (Feb. 2025), p. 2. ISSN: 1433-8351. DOI: 10.1007/s41114-024-00055-8. URL: https://link.springer.com/10.1007/s41114-024-00055-8 (visited on 01/08/2026).

[16] D Davis et al. "LIGO detector characterization in the second and third observing runs". en. In: *Classical and Quantum Gravity* 38.13 (July 2021), p. 135014. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/abfd85. URL: https://iopscience.iop.org/article/10.1088/1361-6382/abfd85 (visited on 12/15/2025).

[17] D Davis et al. "Subtracting glitches from gravitational-wave detector data during the third LIGO-Virgo observing run". en. In: *Classical and Quantum Gravity* 39.24 (Dec. 2022), p. 245013. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/aca238. URL: https://iopscience.iop.org/article/10.1088/1361-6382/aca238 (visited on 12/15/2025).

[18] Maximilian Dax et al. "Real-Time Gravitational Wave Science with Neural Posterior Estimation". en. In: *Physical Review Letters* 127.24 (Dec. 2021), p. 241103. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.127.241103. URL: https://link.aps.org/doi/10.1103/PhysRevLett.127.241103 (visited on 01/08/2026).

[19] T. Dooney. "Deep Learning for Modelling and Separating Gravitational Wave Signals, Glitches, and Noise". English. PhD thesis. Open Universiteit (faculties), Feb. 2026. DOI: 10.71583/20260226td.

[20] Tom Dooney et al. "DeepExtractor: Time-domain reconstruction of signals and glitches in gravitational wave data with deep learning". en. In: *Physical Review D* 112.4 (Aug. 2025). arXiv:2501.18423 [gr-qc], p. 044022. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/s91m-c2jw. URL: http://arxiv.org/abs/2501.18423 (visited on 12/15/2025).

[21] Tom Dooney et al. "One flexible model for multiclass gravitational wave signal and glitch generation". In: *Physical Review D* 110.2 (2024). ISSN: 2470-0029. DOI: 10.1103/physrevd.110.022004. URL: http://dx.doi.org/10.1103/PhysRevD.110.022004.

[22] Reed Clasey Essick. *Detectability of dynamical tidal effects and the detection of gravitational-wave transients with LIGO*. 2017. URL: http://hdl.handle.net/1721.1/115024.

[23] Ryan P. Fisher et al. "DQSEGDB: A time-interval database for storing gravitational wave observatory metadata". en. In: *SoftwareX* 14 (June 2021), p. 100677. ISSN: 23527110. DOI: 10.1016/j.softx.2021.100677. URL: https://linkinghub.elsevier.com/retrieve/pii/S2352711021000224 (visited on 09/15/2026).

[24] J Glanzer et al. "Data quality up to the third observing run of advanced LIGO: Gravity Spy glitch classifications". In: *Classical and Quantum Gravity* 40.6 (Mar. 2023), p. 065004. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/acb633. URL: https://iopscience.iop.org/article/10.1088/1361-6382/acb633 (visited on 12/16/2025).

[25] Stephen R. Green, Christine Simpson, and Jonathan Gair. "Gravitational-wave parameter estimation with autoregressive neural network flows". In: *Physical Review D* 102.10 (Nov. 2020). ISSN: 2470-0029. DOI: 10.1103/physrevd.102.104057. URL: http://dx.doi.org/10.1103/PhysRevD.102.104057.

[26] Melissa Lopez et al. *Simulating Transient Noise Bursts in LIGO with gengli*. 2022. arXiv: 2205.09204 [astro-ph.IM]. URL: https://arxiv.org/abs/2205.09204.

[27] Ronaldas Macas et al. "Impact of noise transients on low latency gravitational-wave event localization". In: *Physical Review D* 105.10 (May 2022). ISSN: 2470-0029. DOI: 10.1103/physrevd.105.103021. URL: http://dx.doi.org/10.1103/PhysRevD.105.103021.

[28] Ann-Kristin Malz and John Veitch. "Joint inference for gravitational wave signals and glitches using a data-informed glitch model". In: *Physical Review D* 112.2 (July 2025). arXiv:2505.00657 [gr-qc], p. 024071. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/fp4b-mvzx. URL: http://arxiv.org/abs/2505.00657 (visited on 09/16/2026).

[29] Rich Ormiston et al. "Noise Reduction in Gravitational-wave Data via Deep Learning". In: *Physical Review Research* 2.3 (July 2020). arXiv:2005.06534 [astro-ph.IM], p. 033066. ISSN: 2643-1564. DOI: 10.1103/PhysRevResearch.2.033066. URL: http://arxiv.org/abs/2005.06534 (visited on 09/17/2026).

[30] Chris Pankow et al. "Mitigation of the instrumental noise transient in gravitational-wave data surrounding GW170817". en. In: *Physical Review D* 98.8 (Oct. 2018). arXiv:1808.03619 [gr-qc], p. 084016. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.98.084016. URL: http://arxiv.org/abs/1808.03619 (visited on 12/15/2025).

[31] Ethan Payne et al. "Curious case of GW200129: Interplay between spin-precession inference and data-quality issues". en. In: *Physical Review D* 106.10 (Nov. 2022), p. 104017. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.106.104017. URL: https://link.aps.org/doi/10.1103/PhysRevD.106.104017 (visited on 12/15/2025).

[32] Jade Powell. "Parameter Estimation and Model Selection of Gravitational Wave Signals Contaminated by Transient Detector Noise Glitches". In: *Classical and Quantum Gravity* 35.15 (Aug. 2018). arXiv:1803.11346 [astro-ph.IM], p. 155017. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/aacf18. URL: http://arxiv.org/abs/1803.11346 (visited on 09/15/2026).

[33] Florent Robinet et al. "Omicron: A tool to characterize transient noise in gravitational-wave detectors". en. In: *SoftwareX* 12 (July 2020), p. 100620. ISSN: 23527110. DOI: 10.1016/j.softx.2020.100620. URL: https://linkinghub.elsevier.com/retrieve/pii/S2352711020303332 (visited on 12/15/2025).

[34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. en. arXiv:1505.04597 [cs]. May 2015. DOI: 10.48550/arXiv.1505.04597. URL: http://arxiv.org/abs/1505.04597 (visited on 12/15/2025).

[35] G. Vajente et al. "Machine-learning nonstationary noise out of gravitational-wave detectors". en. In: *Physical Review D* 101.4 (Feb. 2020), p. 042003. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.101.042003. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.042003 (visited on 12/16/2025).

[36] Leah Vazsonyi and Derek Davis. "Identifying glitches near gravitational-wave signals from compact binary coalescences using the Q-transform". en. In: *Classical and Quantum Gravity* 40.3 (Feb. 2023), p. 035008. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/acafd2. URL: https://iopscience.iop.org/article/10.1088/1361-6382/acafd2 (visited on 01/20/2026).

[37] Michael Zevin et al. "Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science". en. In: *Classical and Quantum Gravity* 34.6 (Mar. 2017). arXiv:1611.04596 [gr-qc], p. 064003. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/aa5cea. URL: http://arxiv.org/abs/1611.04596 (visited on 01/07/2026).