

School of Physics and Astronomy



Gravity Spy and X-Pipeline: A multidisciplinary
approach to characterizing and understanding
non-astrophysical gravitational wave data and its impact
on searches for unmodelled signals

Scott Benjamin Coughlin

Submitted for the degree of Doctor of Philosophy
School of Physics and Astronomy
Cardiff University

October 7, 2019

Summary of thesis

With the first direct detection of gravitational waves, the Advanced Laser Interferometer Gravitational-wave Observatory (aLIGO) has initiated a new field of astronomy by providing an alternate means of sensing the Universe. The extreme sensitivity required to make such detections is achieved through exquisite isolation of all sensitive components of aLIGO from non-gravitational-wave disturbances. Nonetheless, aLIGO is still susceptible to a variety of instrumental and environmental sources of noise that contaminate the data. Of particular concern are noise features known as *glitches*, which are transient and non-Gaussian in their nature, and occur at a high enough rate that the possibility of accidental coincidence between the two aLIGO detectors is non-negligible. Glitches come in a wide range of time-frequency-amplitude morphologies, with new morphologies appearing as the detector evolves. Since they can obscure or mimic true gravitational-wave signals, a robust characterization of glitches is paramount in the effort to achieve the gravitational-wave detection rates that are allowed by the design sensitivity of aLIGO. For this reason, over the past few years, glitch classification techniques have been developed to help make this task easier. Specifically, I explore the effect of glitches, and their suppression, on key gravitational-wave searches such as that for a Galactic supernova. Moreover, I explore the impact of including machine learning techniques in the post-processing stage of the gravitational-wave search algorithm, “X-Pipeline”. When performing a two detector network search for a gravitational wave from a Galactic supernova, this thesis finds that including information about glitch families and using machine learning techniques in the post-processing stages of the analysis can improve the sensitive range of the search by 10-15 percent over the standard post-processing method.

Contents

1	Introduction	1
2	Gravity Spy	6
2.1	The problem of non-Gaussian noise features	6
2.2	Characterization of transient noise in LIGO	8
2.2.1	Impact of Glitches on Gravitational-Wave Data Analysis	8
2.2.2	Identifying Glitches	8
2.2.3	Mitigating Glitches	10
2.3	Gravity Spy Project	11
2.3.1	Data Preparation	13
2.3.2	Citizen Science	22
2.3.3	Machine Learning	27
2.3.4	Socio-Computational Research Support	32
2.4	Preliminary Results	33
2.4.1	Initial Machine Learning Performance	33
2.4.2	Gravity Spy System Beta Testing Results	34
2.5	Conclusions and Future Prospects	36
3	Classifying the unknown: discovering novel gravitational-wave detector glitches using similarity learning	38
3.1	Transfer Learning	40
3.2	Identifying Novel Glitches	41
3.3	Results	43
3.3.1	Different Configurations	44
3.4	Conclusions	46
4	X-Pypeline	48
4.1	Unmodelled Gravitational Wave Searches	49
4.2	Overview of X-Pypeline Data Analysis	50
4.2.1	Characterizing the Data	50
4.2.2	Time-Frequency Representation of the data	52
4.2.3	Characterizing a single time-frequency pixel	56
4.2.4	Building a Detection Statistic for Gravitational Wave Bursts	58
4.2.5	Building a Bayesian Detection Statistic for Gravitational Wave Bursts	62
4.2.6	Coherent Consistency Checks	65
4.3	Tuning the Analysis in X-Pypeline	70
4.3.1	Standard Tuning	70
4.3.2	Using Random Forests to Tune	74

4.3.3	Using Convolutional Neural Networks to Tune Coherent Cuts	78
4.3.4	Using Gaussian Processes to Tune Coherent Cuts	81
4.4	Incorporating Gravity Spy	85
4.5	Conclusion	87
5	A Galactic Core-Collapse Supernova	89
5.1	Neutrino Detection of a Galactic Supernova	91
5.2	Discussion of Gravitational Waves from Core-Collapse Supernovae .	93
5.2.1	Rapidly Rotating Core-Collapse Supernovae	94
5.2.2	Non-Rapidly Rotating Core-Collapse Supernovae	95
5.2.3	Numerical Relativity Simulations	98
6	Results	100
6.1	SNEWS Triggered All-Sky vs. Optically Triggered Search	101
6.2	Analysis Set Up	102
6.3	Results	106
6.4	Conclusions	113
7	Conclusions	114

List of Figures

1.1	Illustration of the optical layout of Advanced Laser Interferometer Gravitational Wave Observatory [3].	2
1.2	Example of an isolated Electrostatic Drive Overflow similar to that which occurred in the Livingston detector at the time of the BNS signal [13]. Top panel: Gravitational wave strain data that has been high and low pass filtered at frequencies of 50 and 290, respectively. Bottom panel: A special type of spectrogram that is made by performing a wavelet transformation on the timeseries data. The resulting image is referred to as a q-scan [14].	3
2.1	Spectrogram representation of three example glitches, with color representing the ‘loudness’ of the signal. Blips (a) are short glitches that usually appear in LIGO’s gravitational-wave channel with a symmetric ‘teardrop’ shape in time-frequency. Blips are the single most important class of glitches in LIGO [28], as they appear in both Hanford and Livingston detectors and are the most stringent limit on LIGO’s ability to detect binary black hole merger signals [39]. No clear correlation to any auxiliary channel has yet been identified. Whistles (b), also known as radio frequency beat notes, usually appear in time-frequency plots with a characteristic ‘W’ or ‘V’ shape. Whistles are caused by radio signals at megahertz frequencies that beat with the LIGO Voltage Controlled Oscillators [42]. Scratchy glitches (c) are believed to be related to the baffle in the input arm of the interferometer which has several openings used to intercept stray laser light and is referred to as the “Swiss Cheese Baffle” [43]. Prior to May 2017 most of the Hanford scratchy glitches seemed to have 12-15 nodes (bright dots) per second, but after the May 2017 work at Hanford, in which the “Swiss Cheese Baffle” was damped with rubber corks in order to limit its movement, most of the scratchy glitches began to appear with about 25 nodes per second [44]. After the dampers were added, the drop in the sensitive range of the detector related to this glitch disappeared [43]. These types of images are what volunteers in the Gravity Spy project classify, and what the associated machine learning algorithms use for training.	9
2.2	Gravity Spy system architecture, and overall data flow through the interconnected, interdisciplinary components of the project.	12

2.3	LIGO data processing and rendering workflow. The glitch triggers (red) are filtered through three separate criteria (blue). Analysis ready refers to time segments when the detector is in <i>lock</i> and in observing mode, meaning the state of the detector was adequate enough to be searching for gravitational waves and ready for data analysis. Those glitch triggers that survive are rendered into Omega Scans (green) of four durations and added to the unlabeled test set (yellow).	14
2.4	Omega Scan images for example members of each class within the Gravity Spy dataset. From top left to bottom right; row one: 1080Lines, 1400Ripples, Air Compressor, Blip, row two: Chirp, Extremely Loud, Helix, Koi Fish, row three: Light Modulation, Low Frequency Burst, Low Frequency Lines, No Glitch, row four: None of the Above, Paired Doves, Power Line, Repeating Blips, row five: Scattered Light, Scratchy, Tomte, Violin Mode, row six: Wandering Line, Whistle. Chirp is not strictly a glitch but is an important category as real gravitational waves can appear in our data stream and the example of None of the Above is only one example as this class can have various forms.	17
2.5	Gravity Spy user interface. This image shows the <i>Black Hole Merger</i> workflow (see Section 2.3.2), with all 22 currently designated categories as options.	23
2.6	Movement of images and volunteers through the Gravity Spy project. Yellow boxes represent the multiple workflows within the project (including the images which are forwarded to experts within the LSC), blue boxes represent the machine learning and crowdsourcing image classifiers, and orange boxes represent the full sets of images, which are designated either as training or testing images (the ‘golden set’ is the subset of the training set which is used to train volunteers). Note that there are multiple beginner workflows with an increasing number of glitch classes which volunteers progress through as they proceed through the training regimen.	26
2.7	Relationship between machine learning confidence in glitch classification (x-axis) and proportion of images from that class assessed by human volunteers at different skill levels. Example glitches classified as a single class (“Power Line” glitches) with differing machine learning confidence scores are shown above.	27
2.8	Deep CNN used for glitch image classification. The network has been introduced on top of the four merged glitch durations. Dimensions of the kernels and feature maps are in units of pixels.	28
2.9	Confusion matrix for the 22 glitch classes in the testing set classified using CNNs, with recall and precision values appended below for reference. The x and y axes represent the predicted and true classes, respectively, and the confusion matrix is normalized by the total number of glitches in each class in the training set. Due to the normalization chosen, the diagonal elements are identical to the recall values for each class. Closer to unity in precision and recall values corresponds to a more accurate classification for a particular class.	34

2.10 Two new O1 glitch classes uncovered during Gravity Spy beta testing: “Paired Doves” (left) and “Helix” (right). “Paired Doves” [78] resemble chirps, but alternate between increasing frequency and decreasing frequency. These glitches are potentially related to 0.4 Hz motion of the beamsplitter at the Hanford detector. “Helix” [90] are possibly related to glitches in the auxiliary lasers (called photon calibrators) that are used to push the LIGO mirrors and calibrate the detectors.	35
3.1 Visual representation of the training set in the DIRECT feature space using the t -distributed Stochastic Neighbor Embedding (t -SNE) statistic. This metric is purely designed to project groups of samples in the N -dimensional feature space into 3 dimensions and has no physical meaning.	40
3.2 New infrastructure proposal for Gravity Spy. This design differs from that described in [15] by facilitating the direct follow-up of single examples of unknown transients through the similarity search algorithm. This is in contrast to the reliance on the None-of-the-Above classifications for filtering out novel glitches from the data set. . . .	42
3.3 <i>Top</i> : Nominal examples of the Raven Peck (<i>left</i>) and Water Jet (<i>right</i>) glitches. <i>Bottom left</i> : The fraction of known Raven Peck samples that have a higher similarity score than a given percentage of other data set samples when calculating similarity to a single known Raven Peck glitch. For example, while retaining 50.0% of known Raven Peck glitches, we can disregard about 99.0% percent of the other data set samples, increasing the purity of the set to be examined by the user. <i>Bottom right</i> : Same for Water Jet glitches. Similarly, while retaining 50.0% of known Water Jet glitches, we can disregard about 99.0% percent of the other data set samples.	45
4.1 Coordinate frame in which the antenna pattern functions are described. An example sky location, given by (ϕ, θ) and reference polarization angle ψ as measured by local south are shown above. The (X, Y, Z) coordinates are such that X and Y are along the arms of the interferometer. (x', y', k) coordinates define the propagation and polarization of the gravitational wave.	51
4.2 Three examples of discrete window functions over a span of two segments of 512 samples. In blue, three fifty percent overlapping Tukey windows. In green, two boxcar windows with no overlap. In purple, three Hann windows with fifty percent overlap. Each Tukey and boxcar window have an area of 1, and each Hann an area of 0.5. We use these examples to illustrate the many different ways the same number of samples could be windowed.	53
4.3 A blip glitch visualized using a STFT with two different resolutions, (top) 0.125 seconds and (bottom) 1.0 seconds. The top represents pixels which have a better time resolution and the bottom represents pixels that have better frequency resolution. For this event, it is easier to resolve the event with the shorter duration, better time resolution STFT. Note that in each case the pixels have the same area $\delta t \delta f = 1$	55

4.4 Illustration of the representation of data \mathbf{d} in the space of detector strains for the three detector case. The red ellipse is the sensitivity of the detector network to linearly polarized gravitational waves. We mark the plane spanned with the unit detector response vectors \mathbf{f}^+ and \mathbf{f}^\times and the orthogonal vector to the plane as \mathbf{e}^n which forms the null space. See Equation 4.14 for more details. Here we see the projection of data \mathbf{d} onto the plane spanned by \mathbf{f}^+ and \mathbf{f}^\times . This (bold) line represents the standard likelihood, the (dashed) line parallel to \mathbf{e}^n represents the null energy of \mathbf{d}	66
4.5 Example of E_+ versus I_+ and E_\times versus I_\times for clusters produced by background noise (x) and by simulated GWBs (□). The color scale is the base-10 logarithm of the detection statistic, in this case the standard likelihood, associated with each cluster. We can see that this is an example of a predominantly + polarized GW signal as the coherent and incoherent plus energies fall to the right of the $E - I$ line, and, conversely the coherent and incoherent cross energies fall to the left of the $E - I$ line.	69
4.6 Example of decision tree utilizing the coherent, incoherent and detection statistics associated with background and injection triggers to predict whether the cluster is a signal or background. Samples are the number of events from injection trails and off source trials. Samples can be one of two classes, background or signal. Values are the number of samples with true label background or signal that fell into a given node based on the question asked in the node above. Nodes coloured orange indicate events with features consistent with background and those coloured blue indicate events with features consistent with signal. Due to the bootstrap sampling (i.e. replacement sampling), the numbers in the values do not match up with the number reported by samples. This decision tree is limited to only two layers and makes decisions based on the incoherent plus and cross energies and the ratio of incoherent to coherent plus energy. As expected, even in such a small decision tree, a threshold on the ratio between the coherent and incoherent energies proves a valuable way to distinguish signals from the noise.	75
4.7 Illustration of the one dimensional convolutional neural network used to classify clusters from injection and background trials as either signal or noise. This selection is motivated by a desire to balance concerns of overfitting the data by providing too many tuneable parameters while limiting the chance the neural network cannot infer the desired relationship between the coherent and incoherent energies. Each convolutional layer utilizes the hyperbolic arc tangent activation function.	79

4.8 Simple example of the use of the Matern kernel with a value of $\nu = \frac{1}{2}$ applied to an example of points $[x_i, \dots, x_M]$ where $x_i \in [0, 5]$ and $y_i = 0$ when $x_i < 2.5$ and $y_i = 1$ when $x_i > 2.5$. Although they are similar due to the simplicity of the example, the optimized GP (purple) provides a closer prediction of the real distribution of data points than the prior (magenta). Specifically, we can see here that after fitting to the training set, the values of the hyperparameters between the prior and the optimized kernel are different. Specifically, the hyperparameters of the prior were $1^2 * K(l = 1, \nu = 0.5)$ and $39.9^2 * K(l = 6.43, \nu = 0.5)$ were the optimized hyper parameters. As we fixed $\nu = \frac{1}{2}$ it makes sense that this did not change through optimization.

82

4.9 The time-frequency representation of two glitches as they would appear in “X-Pypeline” using a STFT with a duration of $\frac{1}{128}s$, with color representing the total energy, see Equation 4.55, of the pixel. Similarly to Figure 2.1, we again show an example of the Blip glitch (a) and the Scratchy glitch (b). In terms of coherent and incoherent energies, it will depend on what is happening in the other detector, which should typically be Gaussian noise or, in any case, should be uncorrelated. Therefore, these glitches should average to zero correlation ($I = E$) if one sums over enough pixels. Therefore, the Scratchy glitch will typically be closer to $I = E$, while the Blip glitch might have random fluctuations such that to $I \neq E$ due to small-number statistics.

86

5.1 Two examples of GW emission from rapidly rotating CCSN. Here we present a case of “mildly” rapidly rotating progenitor with a stiffer EOS and a slightly more rapidly rotating progenitor with a softer assumed EOS [145]. As is expected, the amplitude of the GW emitted from the more rapidly rotating progenitor and softer EOS is larger than that of the less rapidly rotating progenitor. Both amplitudes are scaled to a distance of one kiloparsec.

94

5.2 Two examples of GW emission from non-rapidly rotating CCSN. The three-dimensional Powell waveform is dominated by convection occurring at the surface of the PNS and does not expect a strong impact from SASI [137]. The three-dimensional Couch model, however, shows a significant contribution of the GW emission from SASI [144]. Both amplitudes are scaled to a distance of one kiloparsec.

96

6.7 Efficiency curves of four waveform families utilizing the random forest method trained with all Gravity Spy glitches except for Blip, Koi Fish, Scratchy and Extremely Loud in the training set, and with only those families in the training set. We demonstrate the results of using the two training sets when applied to waveforms from non-rapidly rotating CCSN (Couch and Powell) and two waveforms from rapidly rotating CCSN (Richers and Scheiddeger), see Chapter 5 for more information on the waveforms. Across the different waveform families, there is a consistent improvement in the detectable volume of the search of 10 percent at a FAP of 1% when the random forest is trained with a training set containing clusters associated with the Blip family of glitches. 112

List of Tables

2.1	Breakdown of morphological categories in the Gravity Spy training set, indicating the number of training set samples of each class that comes from Livingston detector data and Hanford detector data. Note that some of the glitches are detector dependent.	21
2.2	Model Specification	29
3.1	The fraction of the original data set with similar scores lower than the similarity score of 50.0% of other known Raven Peck (left) and Water Jet glitches (right). Columns refer to different choices in the activation layer used in the dense layer of the model and the number of training rounds where each round draws a new set of X number of similar and dissimilar pairs. In bold is the configuration(s) that yielded the best reduction versus retention rate for both glitches. . .	44
5.1	From left to right it provides the physical mechanism, literature reference, model name, the root sum square gravitational wave strain (h_{rss}), the characteristic frequency of gravitational wave emission (f_{peak}), and polarizations for the numerical waveform injections utilized in Chapter 6.	98
6.1	Information about the detector network sensitivity to the sky location used in the search at the time of the dummy neutrino trigger. The magnitude of the detector response is given by $\sqrt{(f^+ ^2 + f^\times ^2)}$. We can see that the network is fairly sensitive to the plus and cross polarization from this sky location, and therefore, it presents a good opportunity to benchmark the algorithm and the tuning methods. At the same time, we did not choose a sky location that maximized the network sensitivity at this trigger time as this presented too optimistic a scenario.	106

I am grateful to many for this work. First of all, I thank my advisers Patrick Sutton and Vicky Kalogera, without whom, this PhD would not have been possible. Their joint support of me not only now, but through my undergraduate and master’s research, laid the foundation for this PhD. I thank the Cardiff School of Physics and Astronomy and CIERA, the Center for Interdisciplinary Research and Exploration in Astrophysics, for their support. In addition, I would like to thank the NSF for funding the Gravity Spy project without which this research and joint project would not be possible. In this same vein, I would like to thank the entire Gravity Spy team for all of their efforts to make the machine that is Gravity Spy not only work, but work so well. I thank my fiancée, Jessica, for her support throughout this program. Last but not least, I thank the rest of my family for their support over this year, especially my Dad and Mom for encouraging me to take a chance and apply for this joint supervised PhD.

Chapter 1

Introduction

Einstein's general theory of relativity (GR) predicts that all accelerating objects with non-symmetric mass distributions produce gravitational waves (GW), which are oscillations in the space-time metric [1]. In the same way as light, gravitational effects do not propagate with infinite speed. Whenever the distribution of mass in a given system changes (for example, when one drops a basketball), the gravitational field adapts to this new mass distribution. The speed at which this change propagates is equal to the speed of light, and the resulting changes in the curvature of space-time are GWs. GWs expand and contract space-time orthogonal to the direction of their motion. Therefore, an instrument designed to detect the displacement of two objects relative to each other could sense a gravitational wave passing through it. The problem, however, lies in the amplitude of GWs, and, consequently how much they displace objects. Even from the most dramatic birthplaces of GWs, the moments before, during and after the merger of two large compact binary objects such as binary black holes (BBH) and binary neutron stars (BNS), the amplitude of the resulting waves at cosmological distances from the event are on the order of 10^{-18} m when sensed by a 4 kilometer interferometer. For context, this displacement of space-time is 1000 times smaller than the nucleus of an atom.

Despite these challenges, GWs are detectable through the combined use of multiple instruments called interferometers. These Michelson interferometers with Fabry-Perot cavities utilize laser light that is sent to a beam splitter which causes half the light to go down each of two orthogonal arms. At the end of each of these arms is a reflective mirror. The beams reflect off these mirrors and recombine at the beam splitter, thereby sending a portion of the light toward a photodiode, and the remaining light back towards the laser. This photodiode outputs a current proportional to the average photon flux at the detector [2]. Any differential variation in the lengths of the arms will change the power seen at the photodetector. For example, if a GW passes through an interferometer perpendicular to its arms (i.e. incident from directly below or above it), the mirror of one arm will be expanded away from the photodiode as the other mirror is contracted towards the photodiode. As a result,

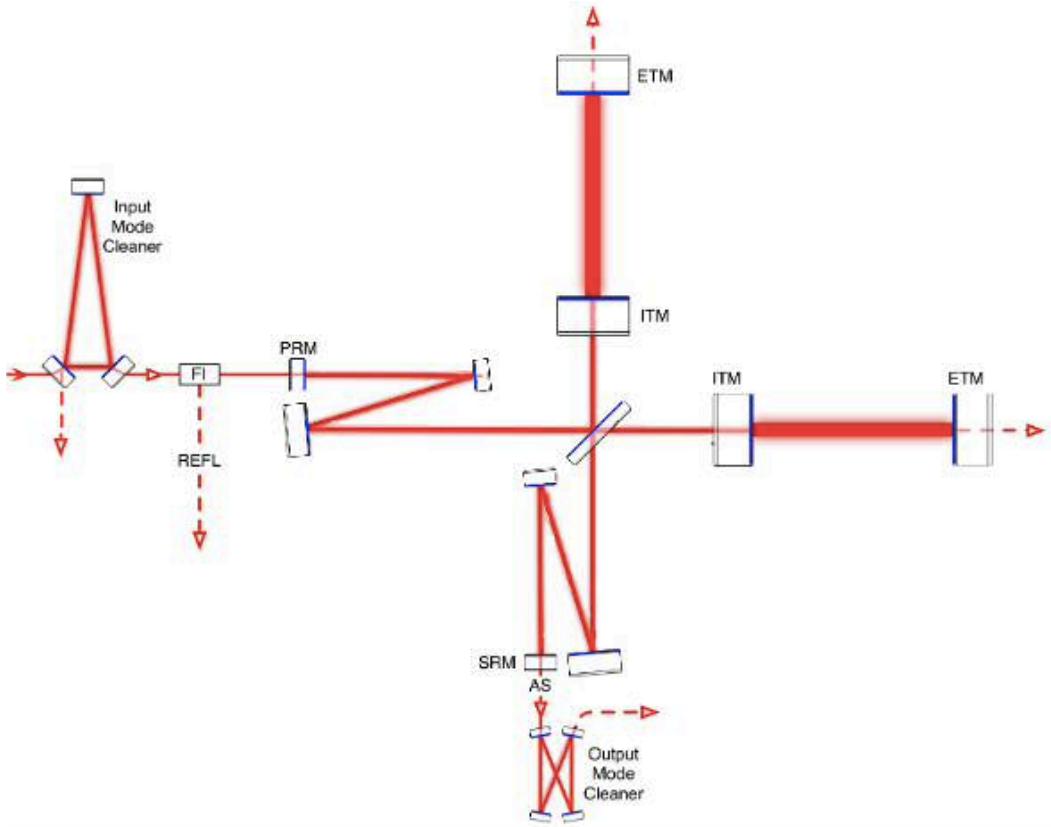


Figure 1.1: Illustration of the optical layout of Advanced Laser Interferometer Gravitational Wave Observatory [3].

the power seen at the photodiode modulates as the length of both arms change as a function of time. In this way, the length of the cavity affects the sensitivity of an interferometer as longer cavities yield larger phase delays between the light in the arms in the cavities.

Current ground-based detectors include the Advanced Laser Interferometer Gravitational Wave Observatory (aLIGO) [3], a diagram of which can be seen in 1.1, Advanced Virgo (AdVirgo) [4], and GEO600 [5]. Future ground-based detectors include KAGRA [6], which will be located in Japan, and LIGO India [7]. aLIGO consists of two 4 kilometer interferometers at Hanford, WA (H1) and Livingston, LA (L1) in the United States. Virgo consists of one 3 kilometer interferometer located in Pisa, Italy (V1). GEO consists of one 600 meter interferometer located in Hanover, Germany (G1). Despite the interferometers inherit abilities to sense the passing of a GW, many environmental (i.e. earthquakes, magnetic field, etc) and instrumental-sources of noise can move the mirrors of the interferometer or affect the light measured at the photo-diode in such a way as to either mimic or mask a GW. Therefore, great care has been taken to isolate the mirrors from as many known sources of transient non-gravitational wave noise, such as the addition of pendulums to reduce the motion of the mirrors and transducers which use seismometers to de-

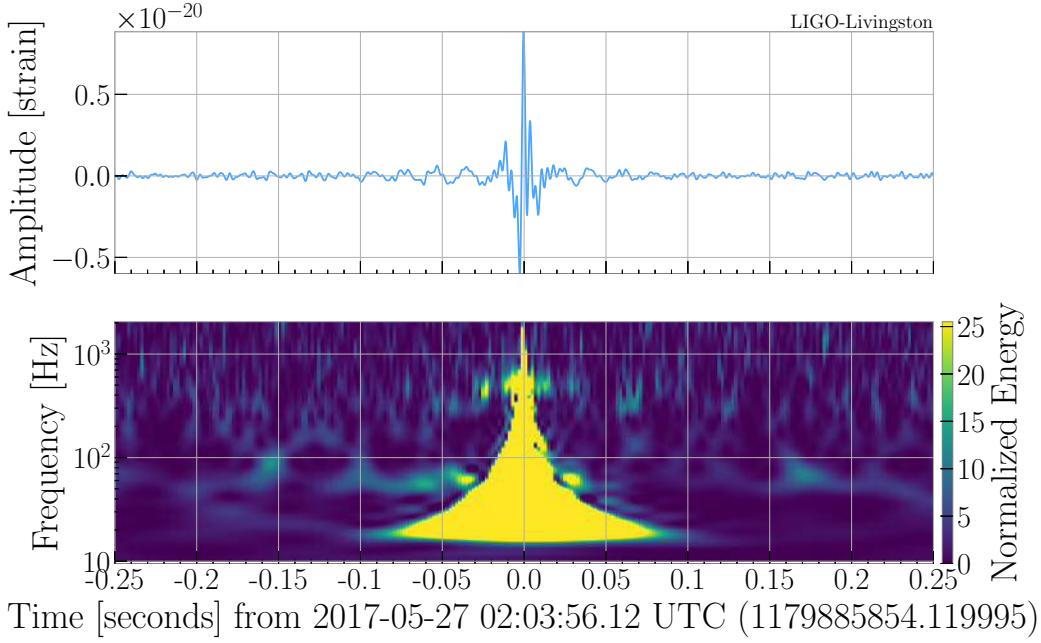


Figure 1.2: Example of an isolated Electrostatic Drive Overflow similar to that which occurred in the Livingston detector at the time of the BNS signal [13]. Top panel: Gravitational wave strain data that has been high and low pass filtered at frequencies of 50 and 290, respectively. Bottom panel: A special type of spectrogram that is made by performing a wavelet transformation on the timeseries data. The resulting image is referred to as a q-scan [14].

tect and counteract earthquakes. Starting in September 2015 and continuing to this day, the success of this work has been seen in the detection of GWs from compact binary coalescences of binary black holes [8–12] and binary neutron stars [13] by aLIGO and AdVirgo. These detections came while the aLIGO detectors collected data as part of their first observing run from September 2015 to January 2016 (O1) and second observing run from November 2016 to August 2017 (O2) and AdVirgo collected data between August 01, 2017 until August 25, 2017.

Despite these successes and efforts to isolate the interferometers from sources of transient non-gravitational wave noise, aLIGO and AdVirgo data is still contaminated with these transient artifacts. In the 51.5 days of O1 alone, approximately 10^6 glitches over a minimum signal-to-noise ratio (SNR) threshold of 6 were recorded [15]. Many of these noise features appear similar when viewed in a time-frequency space known as a spectrogram. Figure 1.2 shows the strain and spectrogram of one such excess noise occurrence at the Livingston detector that is similar to the noise feature seen in coincidence with the BNS detection [13]. Due to the sheer quantity of these excess noise features, however, the ability to comprehensively group or understand these morphological classes has been challenging for LIGO and Virgo scientists. Clean data is desirable for a number of reasons including, but not

