A STUDY ON THE CHARACTERIZATION AND IMPLEMENTATION OF TOOLS FOR ADVANCED LIGO

by

Hunter Gabbard

A thesis submitted to the University of Mississippi in partial fulfillment of the requirements of the Sally McDonnel Barksdale Honors College

Oxford

May 2016

Approved by

_____

Advisor: Associate Professor Marco Cavaglia

_____

Reader: Assistant Professor Kate Dooley

_____

Reader: Associate Professor Joseph Gladden

ABSTRACT

The Laser Interferometer Gravitational Wave Observatory (LIGO) is aimed at directly detecting gravitational waves, small perturbations or ripples in the fabric of space-time. Because of their extreme sensitivity, the LIGO detectors are affected by many sources of non-astrophysical noise. In the first part of this thesis we test a pipeline designed for the identification of short-duration noise transients, called Omicron. We first inject simulated noise waveforms in engineering run data from the LIGO detector in Livingston, Louisiana and then determine Omicron efficiency by attempting to recover these injections. In the second part of this thesis, we present a novel method for the characterization of signals in LIGO data. Using data from LIGO's sixth science run, we develop an algorithm to classify noise transients by their morphology, as well as other parameters such as signal-to-noise ratio, duration, and bandwidth. Two methods, the Kohonen self organizing feature maps and the discrete wavelet transform coefficients, are used to reduce the multidimensional trigger set into an easily readable two-dimensional format.

TABLE OF CONTENTS

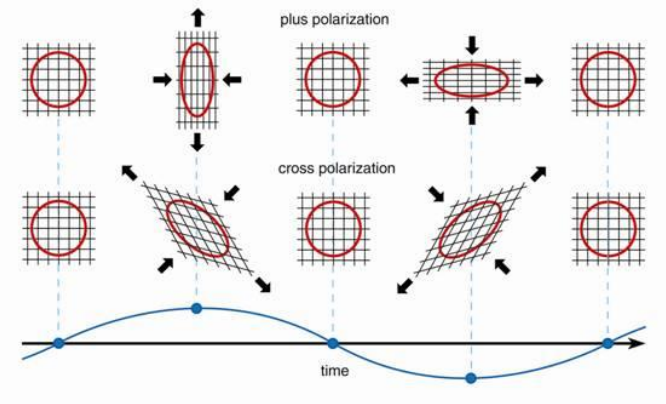
I. Introduction

Einstein's Theory of General Relativity implies the existence of gravitational waves (GWs), small wavelike perturbations of the flat space-time metric [1]

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

where $\eta_{\mu\nu}$ is the metric for flat space-time and $h_{\mu\nu}$ is the tensor describing the perturbation. Possible astrophysical sources of GWs are supernova explosions, spinning neutron stars, mergers of binary neutron stars [2] or black holes [1]. Energy is released from these systems in the form of GWs which propagate outward from the source at the speed of light. As GWs propagate they perturb space-time as shown in figure 1.

Figure 1: Effect of a GW propagating orthogonally to the page. GWs are transverse waves with two polarizations. The top row represents the effect of a GW with "plus" polarization while the bottom row represents the effect of a GW with "cross" polarization [3].

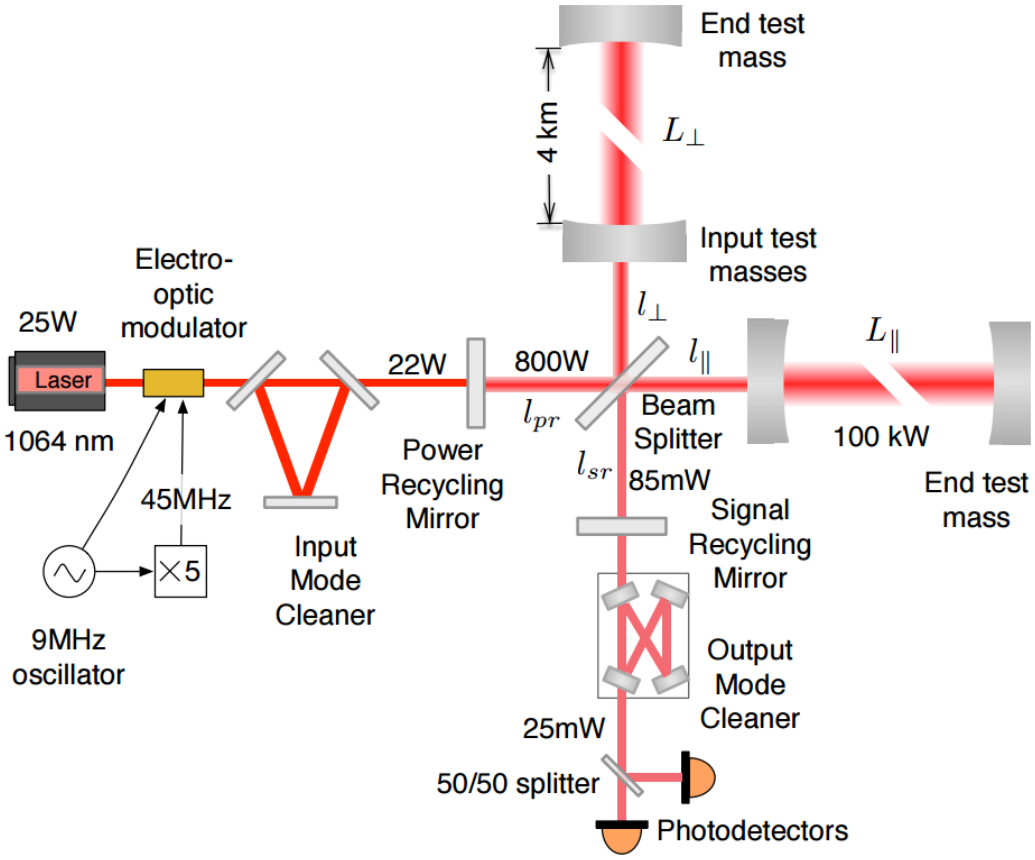


The amplitude of a gravitational wave scales as $\frac{1}{r}$, where r is the distance from the source. GWs coming to Earth from an astronomical distance produce a strain on free-falling test masses that is minuscule in amplitude. A pair of 10 solar mass black holes orbiting around each

other at a distance of 200 Mpc is expected to produce a typical GW strain amplitude by the time the waves reach earth of $\frac{\Delta L}{L} = h \sim 1 \times 10^{-24}$ [4]. The Laser Interferometer Gravitational Wave Observatory (LIGO) detectors have length 4km. Thus, they must be sensitive to changes in distance on the order of $\frac{1}{10000}$ the diameter of a proton.

LIGO consists of two large interferometric GW detectors, each with arm lengths of 4km. One is located in Hanford, Washington and the other in Livingston, Louisiana [5]. These two detectors are a part of a larger worldwide detector network with sites near Hannover, Germany (GEO 600) [6], Cascina near Pisa, Italy (Virgo) [7], and in the Kamioka mine in Japan (KAGRA) [8].

Figure 2: Simplified diagram of an Advanced LIGO detector (not to scale). A "plus" polarized GW along the x and y axes of the figure propagating orthogonally to the detector plane will lengthen one 4-km arm and shorten the other during one half-cycle of the wave; these length changes are reversed during the other half-cycle. The output photodetector records these differential cavity length variations. While a detector's directional response is maximal for this case, the detector is sensitive to GWs for most other angles of incidence or polarizations. [4, 8].



The basic configuration of the LIGO detector is that of a Michelson-Morley Interferometer with added Fabry-Perot cavities (see fig. 2). A laser beam of wavelength 1064 nm is first sent through an optical input mode cleaner (IMC). The IMC "cleans" the laser beam by minimizing its directional and geometric fluctuations and provides frequency stabilization [9]. The laser beam then travels through a beam splitter which splits the laser light into two beams of equal power. Each beam travels down a separate arm which forms a Fabry-Perot cavity. The Fabry-Perot cavities act to increase the power stored in the arms as well as the effective arm lengths. The two beams are reflected by mirrors at the end of the two arms (end test masses)

and recombine at the beam splitter. The interferometer design is such that, if the arms remain unperturbed, the two beams combine destructively.

When a GW impinges on the detectors it produces a relative difference in arm length between the X and Y arms causing the beams in the arms to become out of phase with one another. The output photodetector is used to record these differential cavity length variations and detect the GW signal.

The strain amplitude data is digitized and recorded in a GW channel. Due to the extreme sensitivity of the detectors, external forces can cause unwanted noise in the GW data. Noise disturbances appear both in the time and frequency domains. Localized disturbances in time produce short-duration noise transients, or “glitches”. Localized disturbances in frequency produce “continuous” noise, or frequency lines.

II. The Detector Characterization Group

The aim of the LIGO Scientific Collaboration (LSC) Detector Characterization (DetChar) group is to identify non-astrophysical instrumental and/or environmental disturbances in the interferometer and improve data quality.

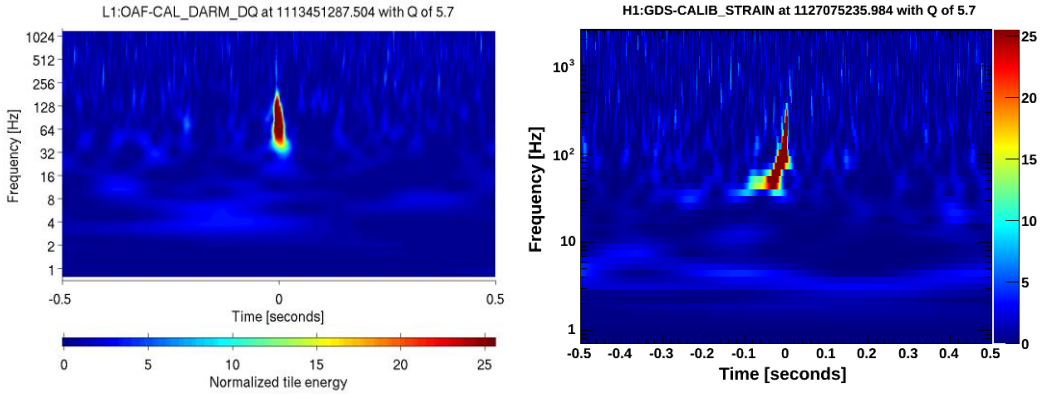
To achieve this goal, LIGO scientists analyze data from ~ 200,000 different sensors whose readout is digitized and recorded in auxiliary channels. These sensors measure external or instrumental disturbances and are used to monitor the constant status of the interferometer. Examples of sensors include accelerometers, microphones, seismometers, and voltage monitors [11].

Auxiliary channel data are analyzed with sophisticated algorithms to determine possible correlations in noise among various components of the detector and/or environmental factors. These correlations may give an indication of the source of non-astrophysical noise, which can then be reduced or mitigated. If the noise source is not fully understood and characterized, data may be discarded [11]. Several pipelines are used to identify and classify noise transients, such as Kleine Welle [12], PCAT [13], dmt_Omega [14] and Omicron [15].

III. Characterization of the Omicron Trigger Generator

Omicron is an algorithm that searches for short-duration noise transients with a signal-to-noise ratio (SNR) above a given threshold. The first stage of the Omicron pipeline reads raw data from a set of channels which is specified by the user.

Figure 3: Glitches can mimic a GW signal. This can make it harder for GW search algorithms to distinguish a true signal from noise [16]. The plot on the left is the time-frequency map of a glitch, while the plot on the right is the time-frequency map of a simulated GW signal, or "injection".



The pipeline then loads data by chunks and breaks these chunks into segments to be analyzed. The power spectrum density (PSD) for each chunk of data is computed, Fourier-transformed and normalized using the PSD. Data is projected onto a parameter space which is tiled in two dimensions, time and frequency. The amplitude SNR is computed for each tile in the time-frequency map. Omicron produces *triggers* which are defined when a tile has SNR above a given threshold [15].

During the “noise hunting” process, short-duration transients found by Omicron are grouped into discrete types, i.e., they are classified in *glitch families* based on similar characteristics of their parameters. Omicron’s efficiency in identifying and characterizing transients can be tested by simulating and then attempting to recover simulated glitches. We developed a code to find coincidences between simulated glitches, or injections, randomly inserted into the data set and triggers recovered by running Omicron on the data. Different injection types allow us to test Omicron’s ability to retrieve different types of noise transients. The procedure was tested over an eight-hour long stretch of science data taken by the Hanford detector between Sep. 30, 2010 00:00:00 UTC and Sep. 30, 2010 08:00:00 UTC.

The code randomly distributes injection parameters for each injection type over a pre-determined range. A C++ Coincidence Finder program reads the injection files representing different waveform morphologies (sine-Gaussian, white noise burst, or string cusp) and injection parameters. The code then stores the relevant parameters from the injection file in vectors. Checks are performed to verify that the injection parameters (amplitude, frequency, time) are within the Omicron search range. After Omicron is run on the data, another code reads and clusters the Omicron triggers in a ROOT format [17] and stores the trigger parameters. A coincident event is defined when an Omicron trigger and an injection *overlap* in time. Time coincidence testing is performed between Omicron triggers and injections based upon a pre-defined coincidence window. If there are several matches within the window, the first injection is taken to prevent bias.

To quantify the performance of Omicron, the program computes several figures of merit. The detection efficiency

$$\varepsilon_{hrss} = \frac{N_d}{N_t}$$

is defined as the ratio between the number of injection events found to be coincident by Coincident Finder, N_d, and the total amount of injection events, N_t. The quantity

$$\varepsilon_f = \frac{N_{df}}{N_{tf}}$$

represents the detection efficiency as a function of injection frequency. It is defined as the ratio between the number of detected injection events at a given frequency, N_{df}, and the total amount of injection events with given frequency values, N_{tf}. The peak time difference

$$\Delta_p = (t_t - t_i)$$

is defined as the time difference between Omicron trigger time, t_t, and injection time, t_i. Finally,

$$\Lambda_a = \frac{\log f_t - \log f_i}{\log f_t + \log f_i}$$

is the "accuracy of the frequency reconstruction", where f_t is the trigger frequency and f_i is the injection frequency. As random coincidences of low SNR triggers increase the background noise and may artificially decrease the injection retrieval efficiency of Omicron, only coincidences with SNR > 8 are considered in the analysis.

Sine-Gaussian waveforms depend on three parameters; the "Q-factor", the frequency f, and amplitude h_{rss}:

$$h = h_{rss} e^{-Qt} \sin(2\pi f t).$$

The "Q-factor" is related to the width of the waveform by

$$\lambda_w = \frac{2\pi}{Q^2}.$$

The above equation shows that the injected amplitude $hrss$ decrease as f decreases and Q increases.

Figure 4: Example of sine-Gaussian waveform.

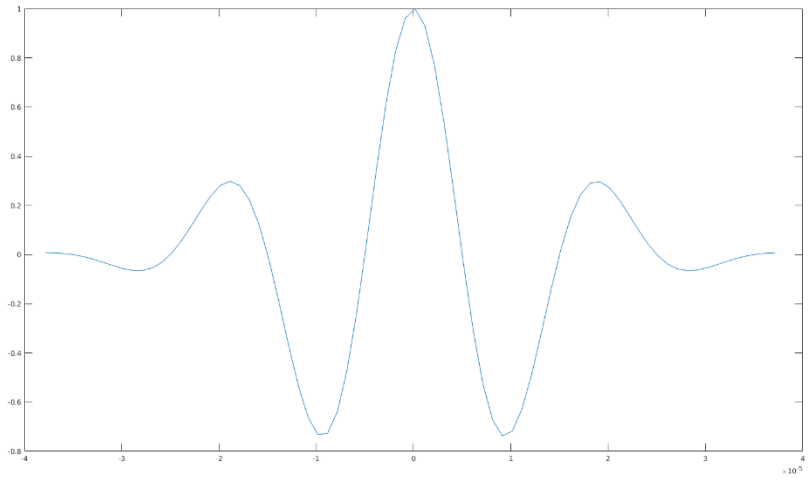
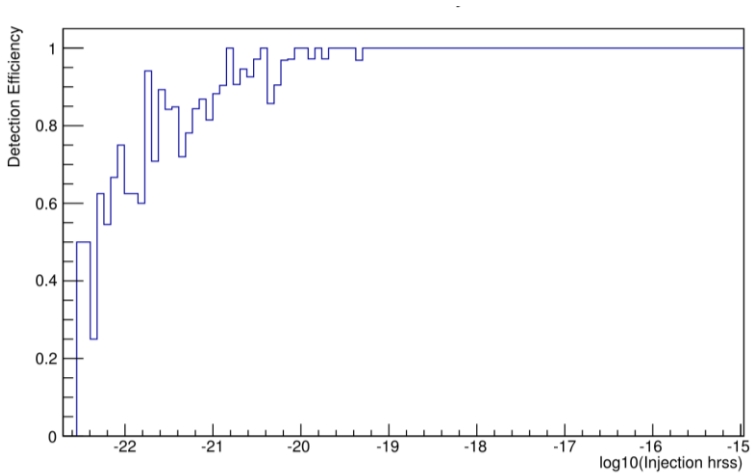


Figure 5: Efficiency curve for sine-Gaussian injections with Q = 75 as a function of the amplitude hrss (log scale). Omicron's efficiency is above 60% for triggers with amplitude hrss $> 10^{-22}$.



Omicron's detection efficiency for sine-Gaussian transients with Q=75 is shown in figure 5. Figures 5 and 6 show that the efficiency does not noticeably vary at high Q values, however at low Q the detection efficiency of Omicron decreases. This decrease is due to the e^{-Qt}

exponential decay factor which is dependent on Q. Additionally, as Q decreases, the duration of the waveform also increases.

Figure 6: Efficiency curve for sine-Gaussian injections with Q = 30 as a function of the amplitude. Omicron's efficiency is above 60% for triggers with amplitude hrss $> 10^{-22}$.

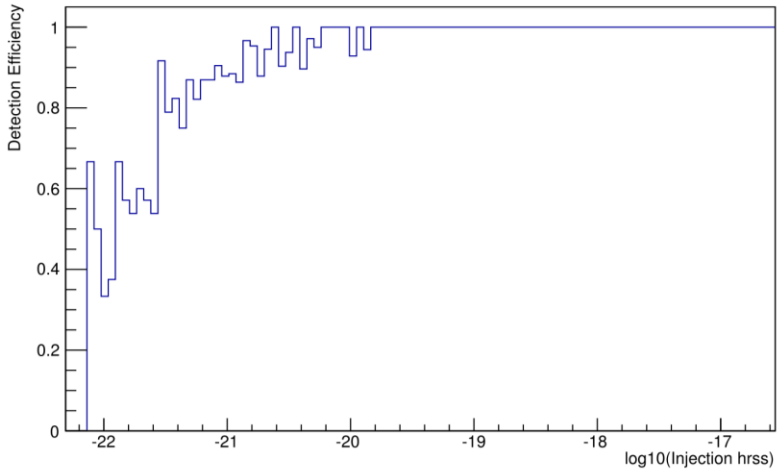
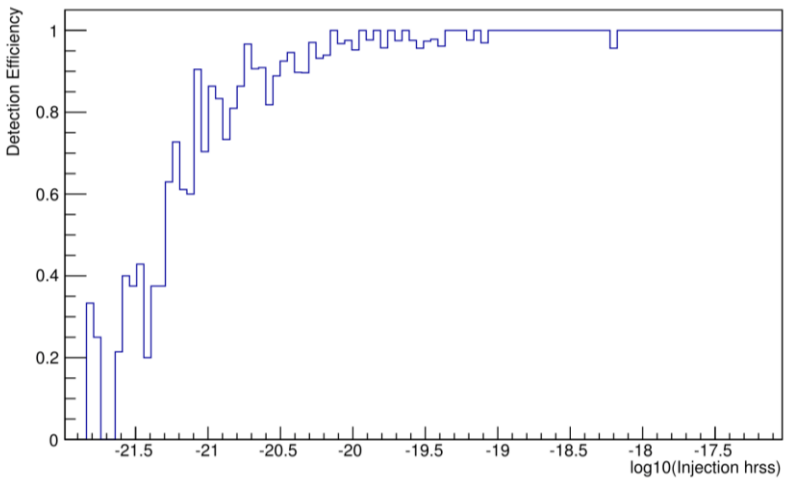
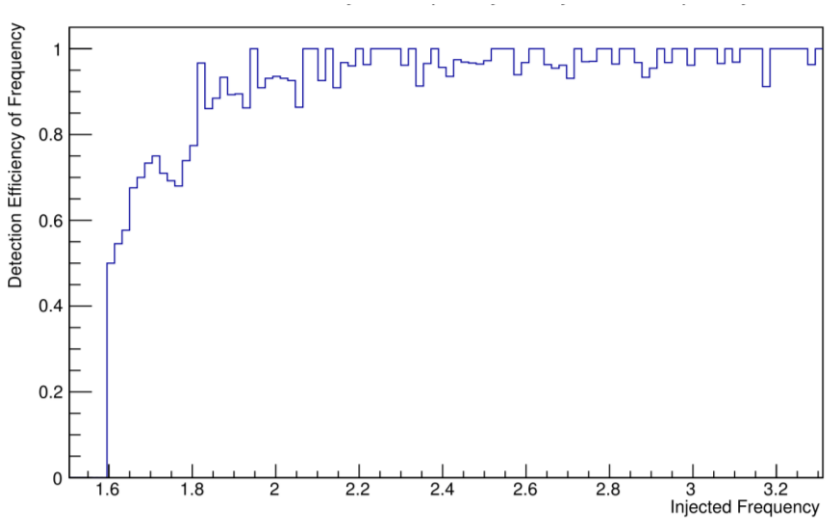


Figure 7: Efficiency curve for sine-Gaussian injections with Q = 5 as a function of amplitude. Omicron's efficiency is above 60% for triggers with amplitude hrss $> 10^{-21}$.



Because Omicron identifies triggers based on an SNR threshold, there should be no coupling between the detection efficiency of Omicron and the injection frequency. Figure 8 shows the detection efficiency as a function of the injection frequency.

Figure 8: Sine-Gaussian efficiency curve as a function of the log of the injection frequency for Q = 5.

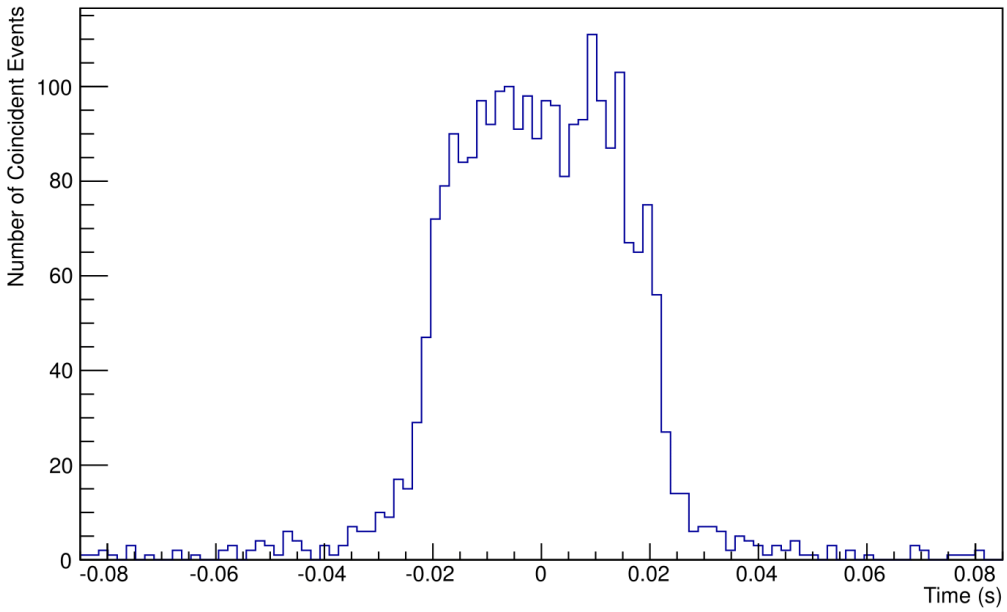


The unexpected decrease in the detection efficiency at low frequencies can be explained by a redefinition of the injected amplitude implemented in the code [18]

$$hrss_t = \sqrt{\frac{2\pi f(hrss_i)}{Q}},$$

where $hrss_t$ is the redefined amplitude, f the injection frequency, and Q the quality factor.

Figure 9: Plot showing peak time difference for sine-Gaussians with Q = 30. Data indicates that the difference between the injection time and the recovered Omicron trigger time is in the range +- 20 ms.



The peak time is useful to understand Omicron injection time retrieval accuracy. It is computed by determining the difference between trigger time and injection time. Figure 9 shows that the distribution is centered at $t = 0$. The uncertainty is +- 20ms. This uncertainty is due to the difference in the geocentric (time referenced from the center of the earth) injection time and the local detector time when triggers are retrieved. When an injection is made, it is given a randomized orientation in the sky and the injection time is referenced by the geocentric time. The time it takes for a gravitational wave to propagate from the center of the earth to its surface is on the order of 20ms. Thus, depending upon the sky location, the time at which Omicron detects the injection may vary by up to +- 20ms.

White noise burst (wnb) waveforms depend on three parameters: the amplitude equivalent isotropic radiated energy (solar mass per parsec squared), duration of injection (seconds), and the bandwidth of the injection (Hertz).

Figure 10: Example of a wnb waveform

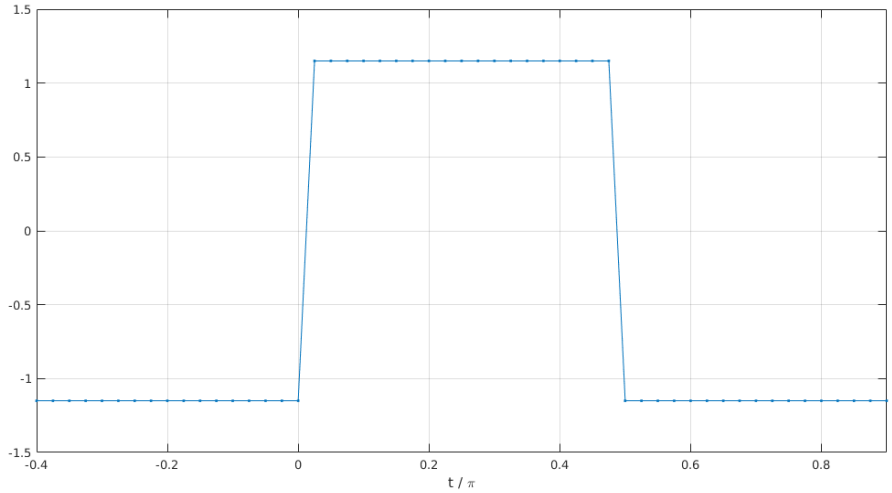
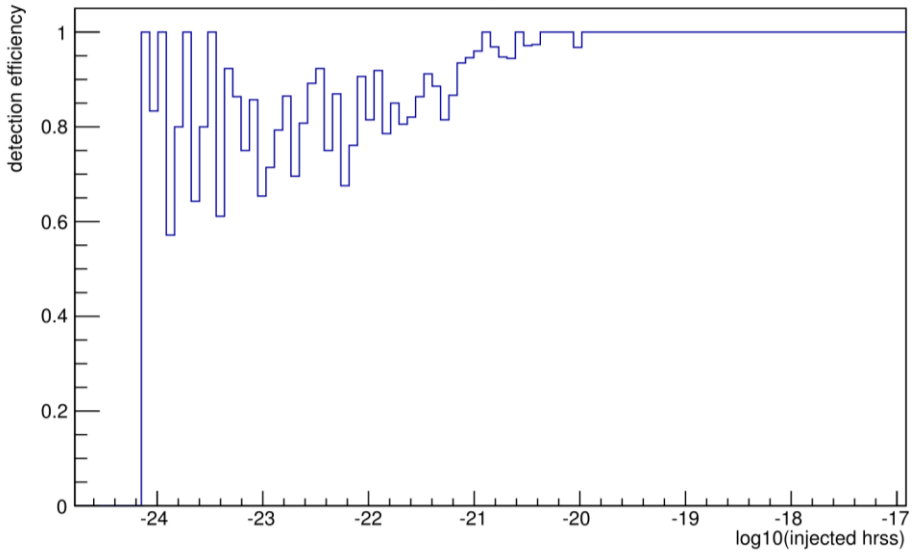


Figure 11: Efficiency vs. hrss. WNB injections are retrieved by Omicron with an 80% detection efficiency starting at $\approx 10^{-23}$.



At low injection amplitudes, $hrss \sim 10^{-24}$, the detection efficiency is ~100%. This can be explained by random coincidences between the low number of random injections with this amplitude (one or two) and the high number of background noise triggers at low amplitude.

Figure 12: Peak time difference of wnb injection. Omicron retrieves wnb injection times with high precision, but with higher random coincidence with respect to sine-Gaussian injections which can be seen in the elevated coincident background.

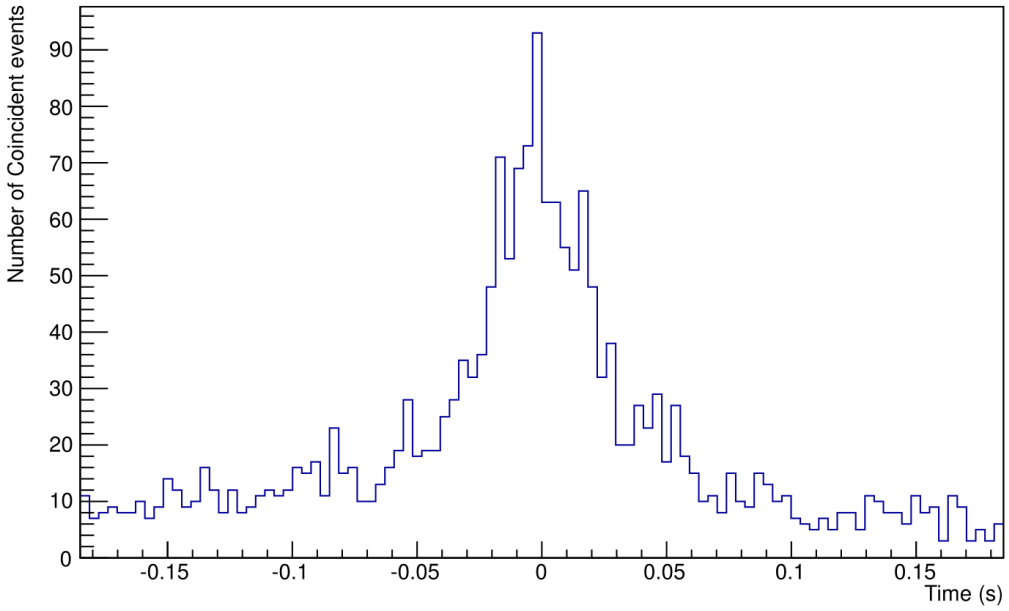


Figure 12 shows that the recovered Omicron trigger time matches the injection time with an uncertainty of +-5 ms. However, wnb injections show an overall increase in the number of random coincidences by a factor of 10 compared to sine-Gaussian. Since wnb injections can have a duration of up to 2 seconds, the duration coincidence window is larger than for sine-Gaussian injections. Therefore, there is a higher probability for random coincidences.

Figure 13: Accuracy of frequency reconstruction for wnb injections. The analysis indicates that not all injections are detected with frequency equal to the injected frequency. The asymmetry in Omicron

injection frequency retrieval is due to the presence of a high volume of background noise triggers at high frequency and low SNR which are randomly coincident with wnb injections.

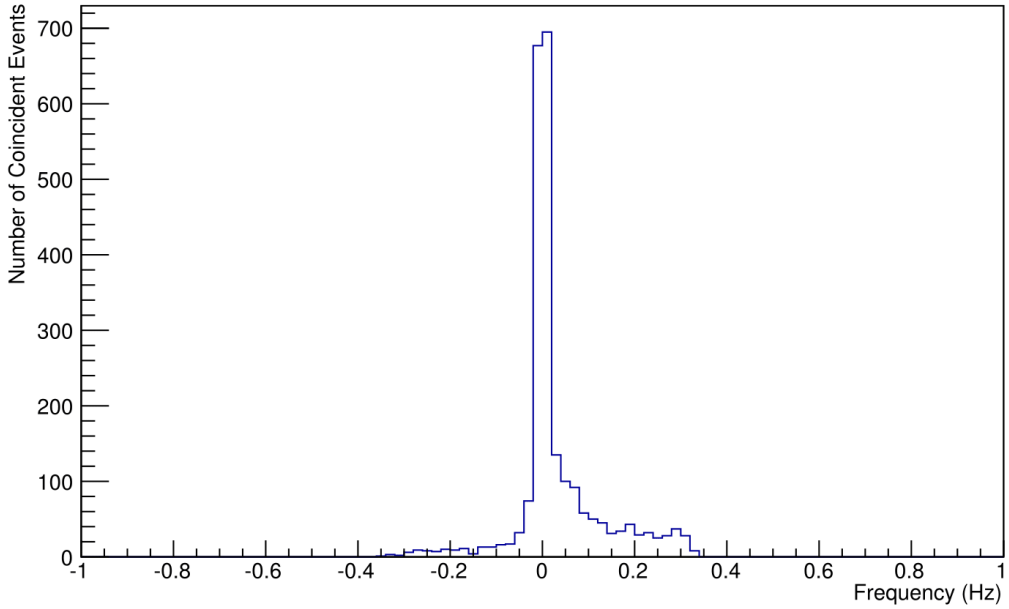
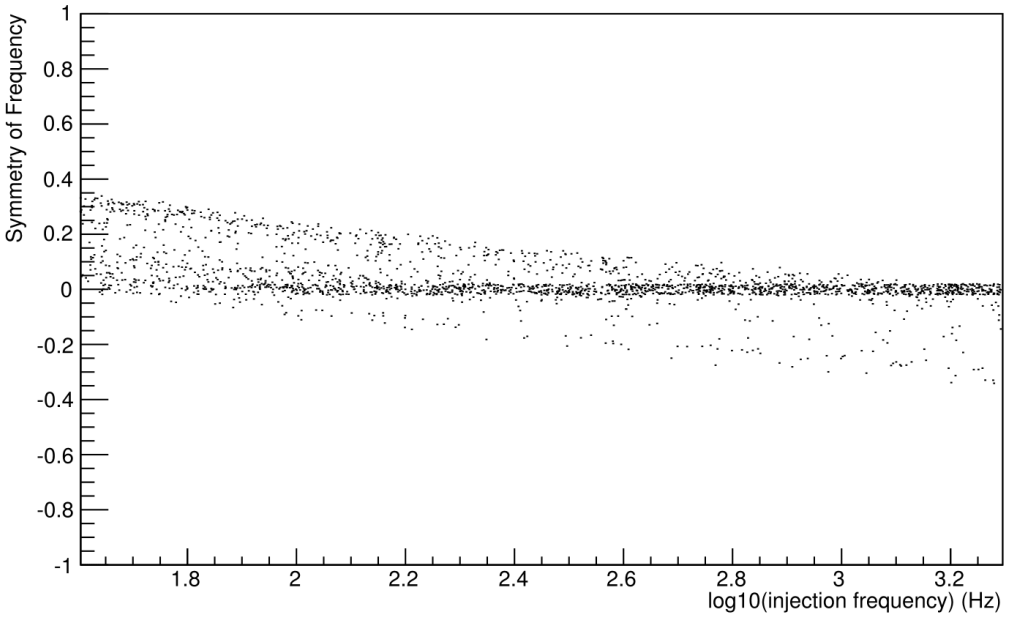


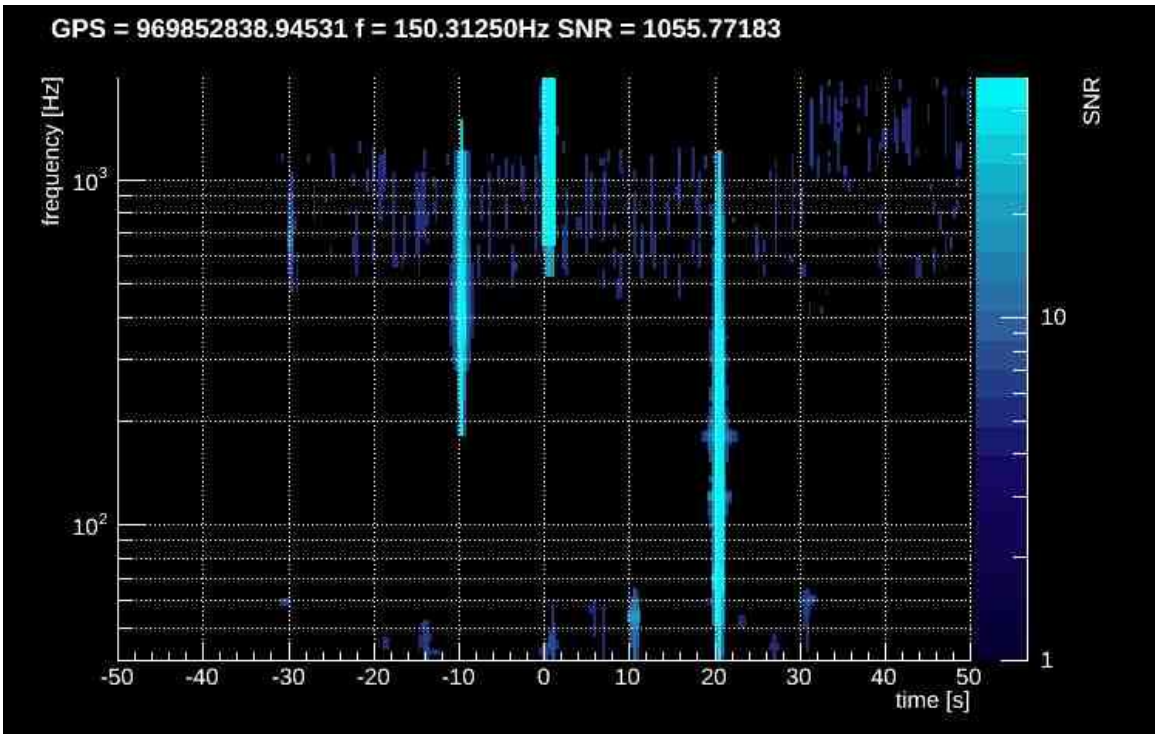
Figure 14: Accuracy of frequency reconstruction as a function of injection frequency. As the central injection frequency increases the accuracy of frequency reconstruction varies monotonically from approximately +0.2 to -0.2. The injected frequency is recovered exactly at around 2 Hz.



The frequency reconstruction in fig. 13 indicates that not all injections are detected at the injection frequency defined by the injection generator program. Several injections are found at a higher/lower frequency than their injected frequency. Omicron does not determine the significance of a trigger based upon its central frequency value so one should not expect a significant error in the frequency reconstruction of Omicron.

As the central injection frequency increases the accuracy of the frequency reconstruction decreases (see fig. 14). This is due to an artifact of the data set. Figure 15 shows that the GW channel used for this study has excess noise at around 1kHz. Therefore, because there is a *higher* number of *low* SNR noise artifacts with frequency values larger than many injections, occasionally random coincidences may occur between noise artifacts. Figure 15 shows the faint structure of this 1kHz noise as well as the large, broadband injections.

Figure 15: Frequency-time map of the GW channel (color denotes SNR). Because of the high volume of background noise triggers at ~1 kHz, injections with a central frequency value higher or lower than the central frequency value of the background noise triggers may be randomly coincident with background noise triggers.



String cusp injections are described by the waveform

$$h(f) = Af^{-\frac{5}{3}}\theta * (f - f_{inj}),$$

where A is the amplitude, f is the frequency, θ is the sign function, and f_{inj} is the cutoff frequency.

Figure 16: Example of a string cusp waveform [19].

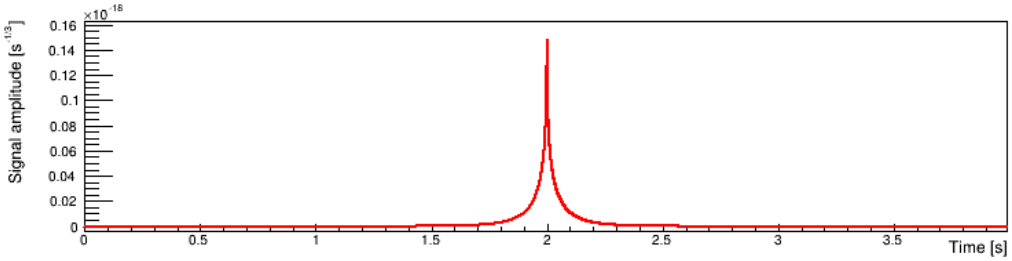


Figure 17: Detection efficiency of string cusp injections as a function of amplitude.

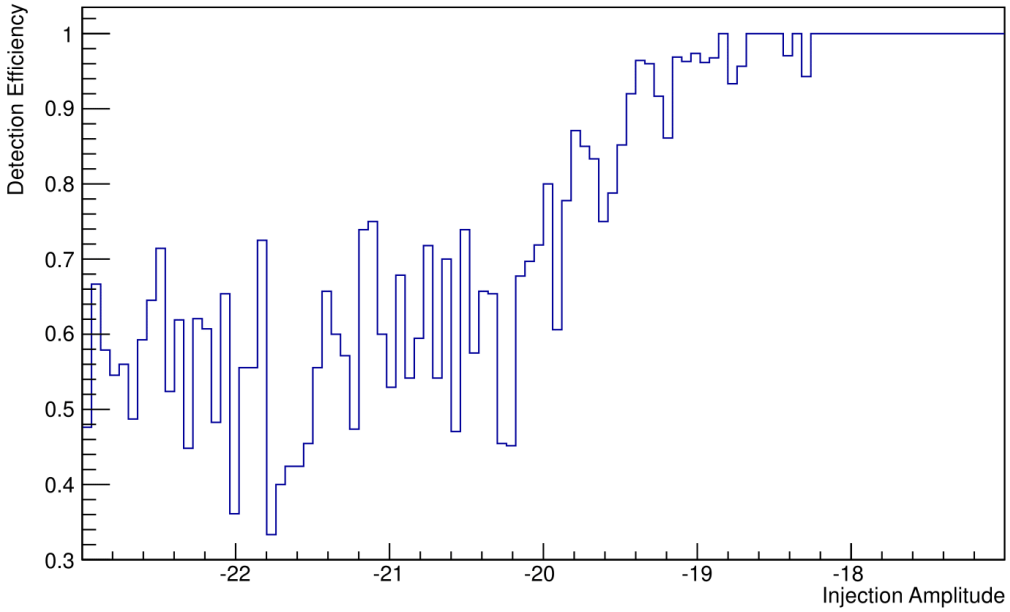


Figure 17 shows that Omicron is only able to detect injections with an efficiency of 60% at $\sim 10^{-20}$, performing much worse than for wnb and sine-Gaussian injections. The cause for the poor detection efficiency is not fully understood, but may partially be due to the short-duration of string cusp injections as well as the non-Gaussian shape of their waveform. The efficiency curve flattens out at $\sim 10^{-21}$ due to random coincidences.

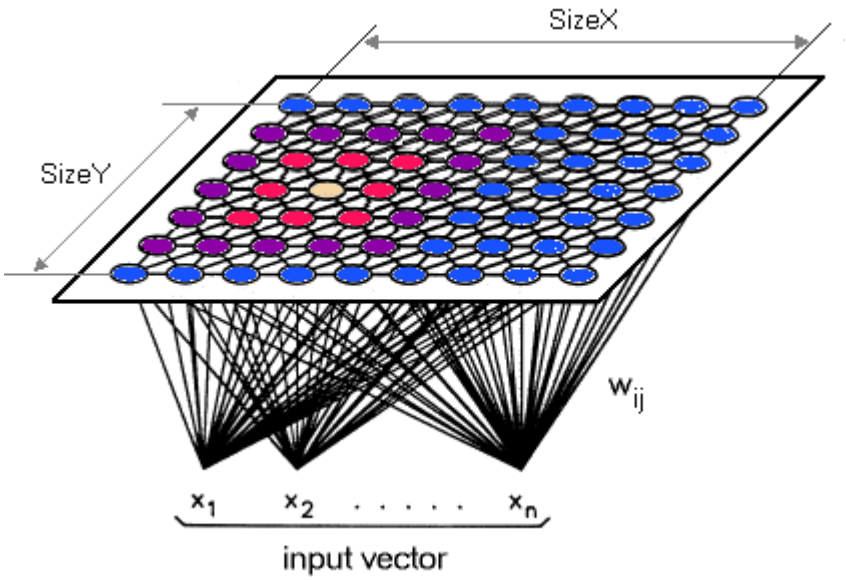
IV. A low-latency Glitch Classifier Based in Waveform Morphology for Advanced LIGO

In order to determine the source of short-duration noise transients in low-latency, it is advantageous to develop tools that classify triggers in an unsupervised fashion. Quick characterization of noise sources can significantly help increase the duty cycle of the detectors. In this part of the thesis we will discuss a new approach to unsupervised glitch classification based on Kohonen Self Organizing Feature Maps.

Developed by Teuvo Kohonen in 1982, Self-Organizing Maps (SOMs) provide a unique way of representing multidimensional data in low dimensional spaces. An SOM is able to convert complex non-linear statistical relationships between a high-dimensional set of samples into simple, human-readable geometric relationships [20]. An SOM is realized as an unsupervised learning neural network. In an unsupervised machine-learning classification algorithm the characteristics of classes, based on the training set, are unknown before the algorithm is applied. In order to characterize these classes, the SOM uncovers patterns in the training set without any human interaction.

Figure 18 is a visual representation of the structure of an SOM neural network. In this example there is a set of N input vectors, $\overline{V_n}$, where $n = 1 \ldots N$. One can think of these input vectors as triggers from an auxiliary or GW channel with a given set of parameters. These triggers will eventually be stored in a 2-D lattice of nodes where SizeX and SizeY represent the dimensions of the lattice.

Figure 18: Representation of a simple Kohonen Network [20].



Every node i has a topological position and a corresponding weight described by a vector $\overline{W_i}$ with the same dimension as the input vectors $\overline{V_n}$. During the first step of the algorithm, each node's weights are initialized. One vector is then chosen at random from the initial sample set and introduced to the network lattice. Each node is analyzed in order to determine which node's weight is most similar to the input vector. We call the node that is most similar the best matching unit (BMU). The best matching unit is computed by running through all the nodes and calculating the Euclidean distance between the node's weight vector and current input vector

$$D_{ni}{}^2 = |\overline{V_n} - \overline{W_i}|^2,$$

where $\overline{V_n}$ is the given input vector, $\overline{W_l}$ is the node's weight vector. The node with the shortest Euclidean distance is the BMU. The radius of the neighborhood of the BMU is calculated. The radius is initially about the size of the lattice, but then decreases over many iterations by applying the following functions

$$\sigma(t) = \sigma(0)e^{-t/\lambda},$$

$$\lambda = \frac{q_t}{R_m},$$

where t is the index of the current iteration, λ is the time constant, $\sigma(0)$ is the initial neighborhood radius of the map, q_t is the predetermined total number of iterations to be performed, and R_m is the map radius. As the number of iterations, or time increases, the radius approaches a minimum equivalent to $\sigma(0)e^{-R_m}$. Each node found to be in the BMU neighborhood during each iteration is adjusted. The closer a node is to the BMU, the more weight is given to its weight vector, $\overline{W}$.

According to the formula

$$\overline{W}(t+1) = \overline{W}(t) + \Theta(t)L(t)(\overline{V}(t) - \overline{W}(t)),$$

$$L(t) = L_0 e^{-t/\lambda},$$

$$\Theta(t) = e^{-\sigma(0)^2/2\sigma^2},$$

where $\overline{W}(t+1)$ is the weight adjustment, $L(t)$ is the learning rate, $\Theta(t)$ is the distance from the BMU, and $\sigma(0)$ is the neighborhood radius. The SOM algorithm then picks another random vector and repeats this process for N iterations [20].

LIGO pipelines such as Omicron [15], Kleine Welle [12], and omega [14], attempt to clarify glitches by exploiting many different trigger properties (duration, signal-to-noise ratio, and central frequency). However, an important aspect of triggers, the morphology, has largely been overlooked. The waveform morphology of triggers often contains rich information [21]. Coefficients produced from discrete wavelet transforms (dwt) can be used in order to properly estimate the shape of a trigger waveform.

The continuous wavelet transform of a signal s(t) takes the form

$$W(a,b) = \frac{1}{\sqrt{a}} \int \bar{\psi}\left(\frac{t-b}{a}\right) s(t)dt,$$

where $\bar{\psi}$ is the complex conjugate of the analyzing wavelet, a is a time dilation and b is a time translation. With regards to signal processing, the significance of the continuous wavelet transform is best illustrated by comparing it to the short-time Fourier transform:

$$F(\omega, b) = \int h(t-b)e^{i\omega t}s(t)dt.$$

The equation above is an expansion of the signal in terms of a family of functions $h(t-b)e^{i\omega t}$, which are generated from a single function $h(t)$ through translations b in time and translations ω in frequency. On the other hand, the wavelet transform is an expansion in functions $\bar{\psi}(\frac{t-b}{a})$ generated by translations b in time and dilations a in time. So, the continuous wavelet transform resembles a (continuous) bank of short-time Fourier transforms with a different window for each frequency. It is important to note that while the basis functions in $F(\omega, b)$ have the same time and frequency resolution at all points of the transform plane, those of $W(a, b)$ have time resolution which decreases with a and frequency resolution which increases with a. This property can be a great advantage in the estimation of trigger waveform morphologies since high frequency signal characteristics are generally localized in time whereas slowly varying signals require accurate low frequency resolution [22].

The discrete wavelet transform is defined as

$$W(j,m) = \sum_{n=1}^{N} \frac{1}{\sqrt{2j}} \psi\left[(\frac{n}{2j} - m)\Delta t\right],$$

where n is the index of the n-th point in the time series, j is the dilation index of the wavelet, m Is the translation index, N is the number of points in the series and Δt is the sampling rate. The wavelet coefficients are calculated iteratively by sliding the wavelet function over the time series and doubling the support of the function at each iteration.

The Haar wavelet is fast and works well when applied to a non-stationary time-series. Even within a small window, GW and auxiliary channel time-series cannot be described well by a combination of sine and cosine terms. Other wavelets such as Daubechies, Mexican hat

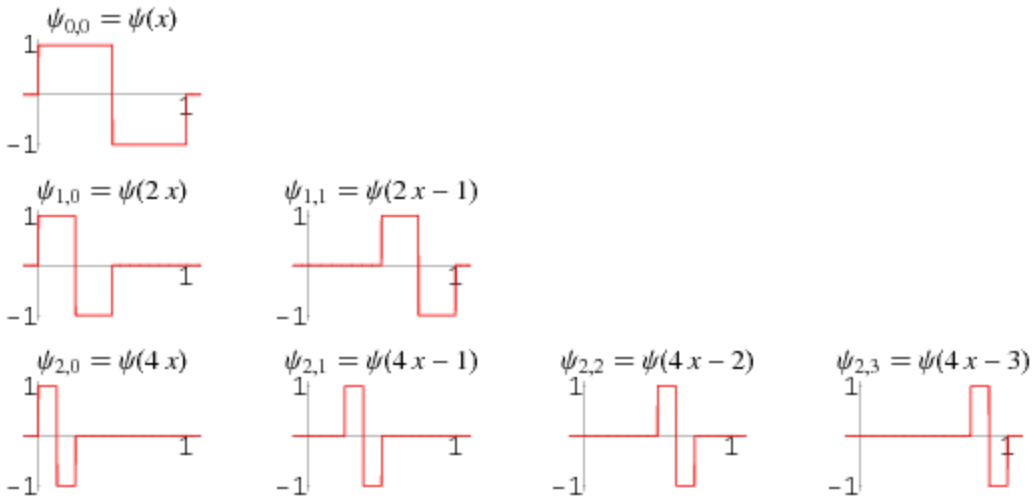
wavelets, and Morlet wavelets have the advantage of better resolution for a smoothly changing time-series, but they have the disadvantage of being expensive to calculate relative to Haar wavelets.

Haar wavelets use a rectangular window defined as

$$\psi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \end{cases}.$$

The first few elements of the discrete Haar functions are shown in fig. 19. below.

Figure 19: Example of discrete Haar wave function [23].



For the purpose of this study, triggers are produced using omega [14]. Omega works by first whitening the data time series and then performing a filtering process. After the time series is filtered, the code then normalizes this time series. The data is tiled logarithmically in frequency and linearly in time. Each tile has a corresponding number of Q-planes, where Q represents the quality factor of the waveform. Omega then calculates a SNR ratio for each tile. The algorithm

defines a trigger based upon a predetermined threshold SNR value set by the user (typically about 8).

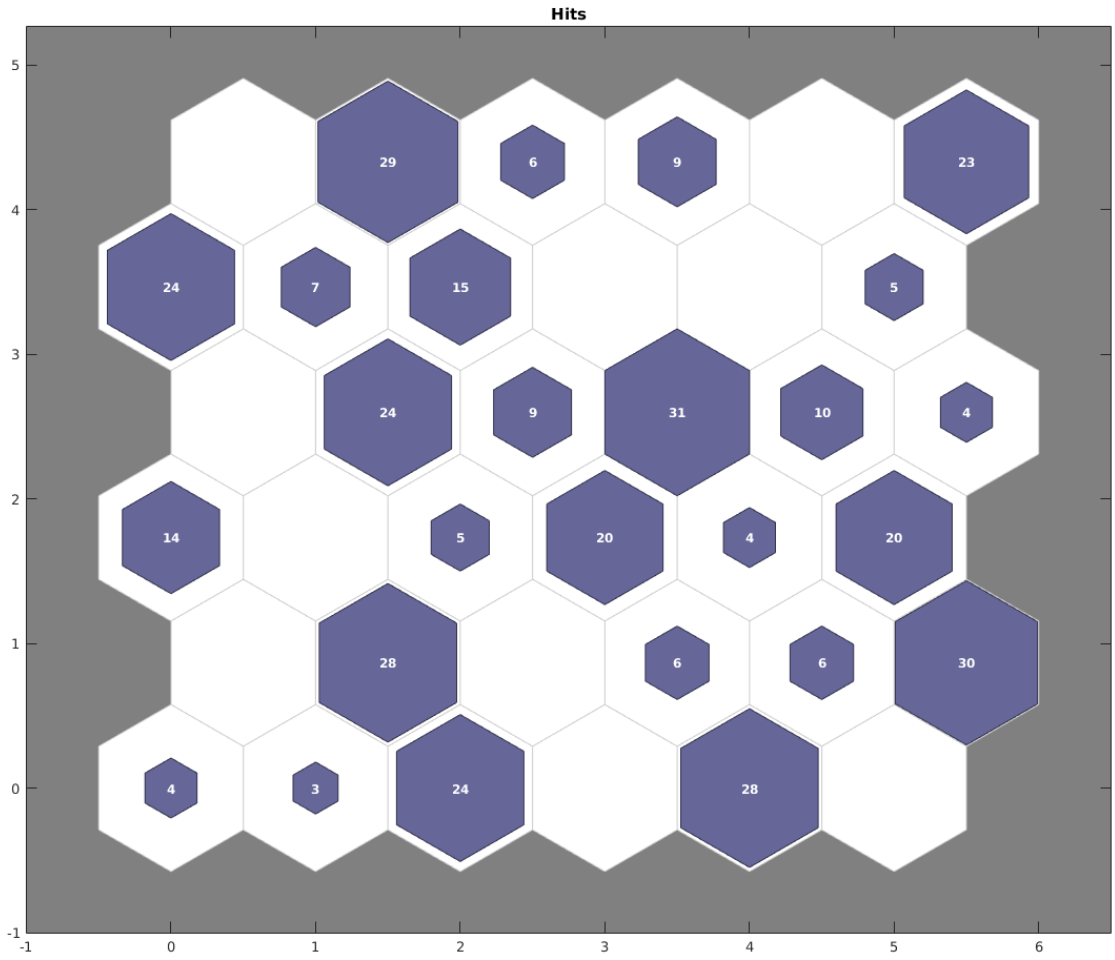
Each trigger is described by five parameters: GPS time of the event, peak frequency (Hz), trigger duration (s), bandwidth (Hz), and normalized energy. The algorithm computes the SNR of each trigger as

$$SNR = \sqrt{2E_n},$$

where E_n is the normalized trigger energy. Triggers with SNR > 20 are extracted in order to obtain waveforms that are well defined. Time, frequency, bandwidth, duration and SNR from both the differential arm lock (DARM) channel (i.e. GW channel) and auxiliary channels are considered as trigger parameters. This is done in order to determine the overall coherence between the instrumental sub-systems and the GW channel. This serves as a sanity check for the classification algorithm. Triggers are then band-passed and whitening is performed to prevent frequency bias.

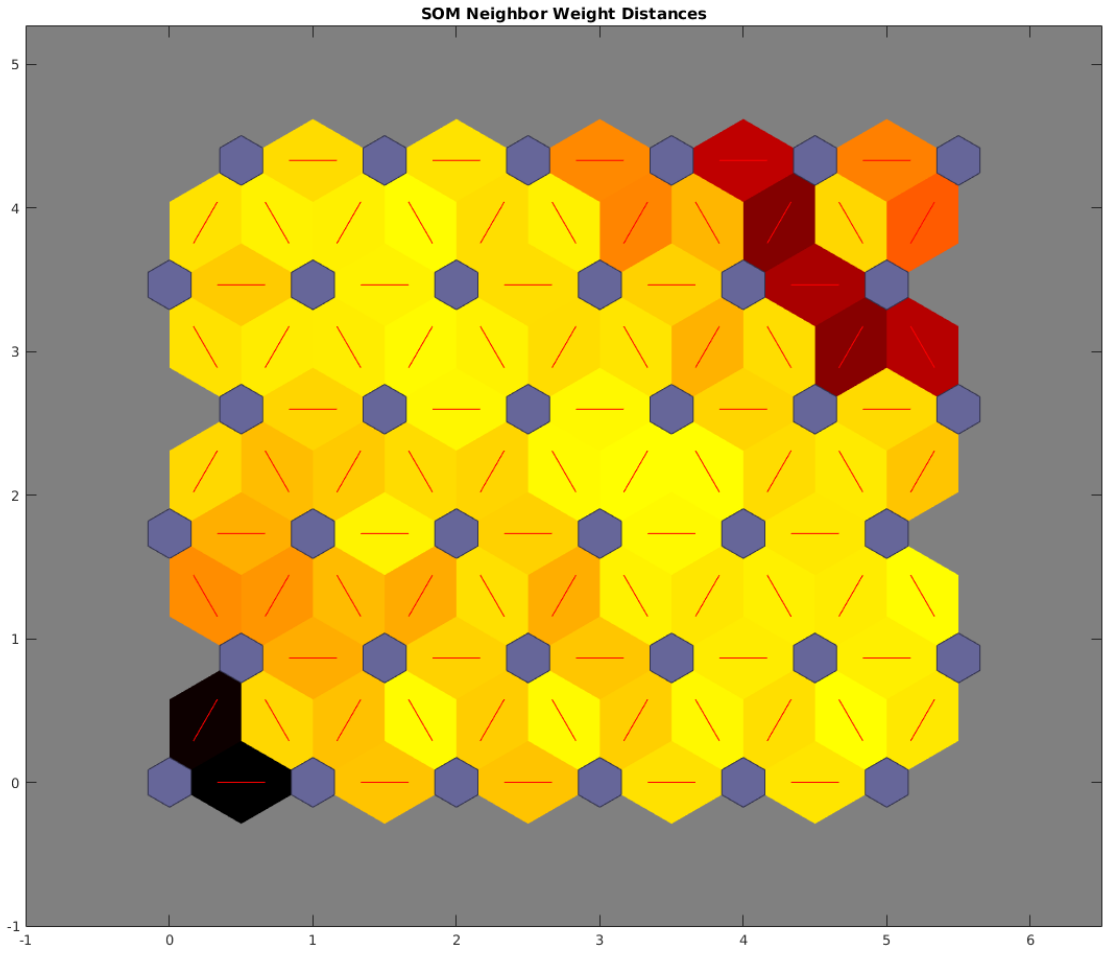
The above procedure was tested on one week of data from LIGO's sixth science run (S6). Triggers from both auxiliary and GW channels were considered. After extraction, data conditioning and normalization of the omega triggers, the SOM glitch classification pipeline was run over three days' worth of LIGO S6 data (July 2, 6, and 7, 2010). The SOM algorithm was then applied to both the DARM and auxiliary channels. Results are plotted in a 6x6 hexagonal SOM hits map, figure 20. This study focused on the DARM GW channel.

Figure 20: A 6x6 hexagonal hits map of all auxiliary channels during the day July 2, 2010 using 16 discrete wavelet transform coefficients. Each neuron contains a discrete number of triggers (white numbers in the hexagons) as well as information about the sub system associated with the highest number of triggers present in that neuron (stored in MATLAB vectors not pictured).



The relationship between each neuron is shown in the SOM neighbor distances map, figure 21. This map illustrates the euclidean distance between each neuron's weight vector, otherwise known as the "similarity". Similarity is qualitatively expressed with lighter colors (yellow) representing a high degree of coherence and darker colors (black) representing a low degree of coherence between neighboring neurons.

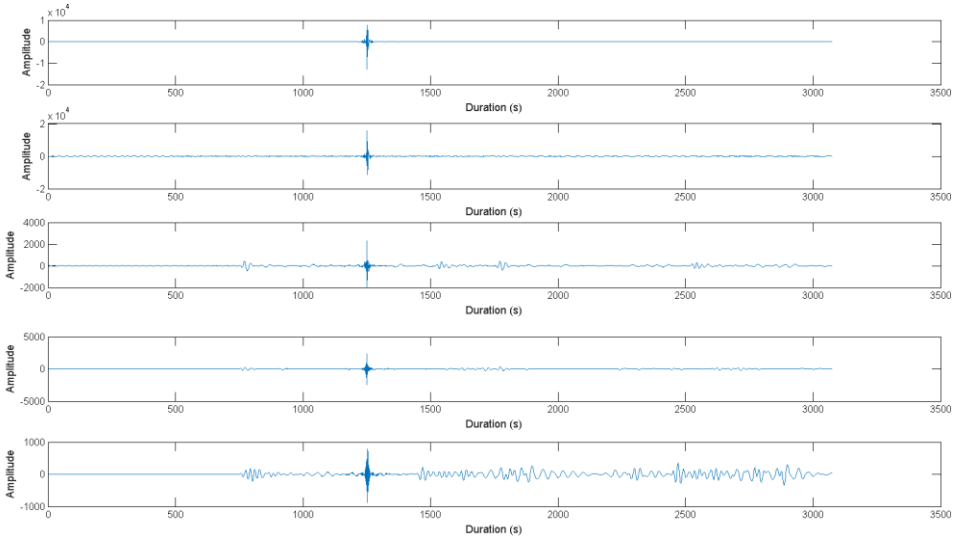
Figure 21: SOM neighbor distances map for auxiliary channels during July 2, 2010. Lighter colors (yellow) represent a high degree of coherence and darker colors (black) represent a low degree of coherence between neighboring neurons.

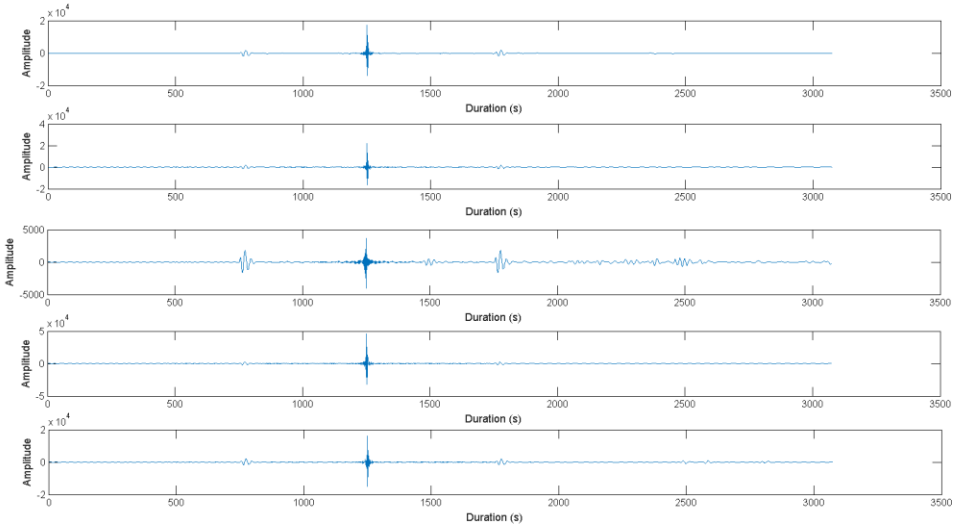


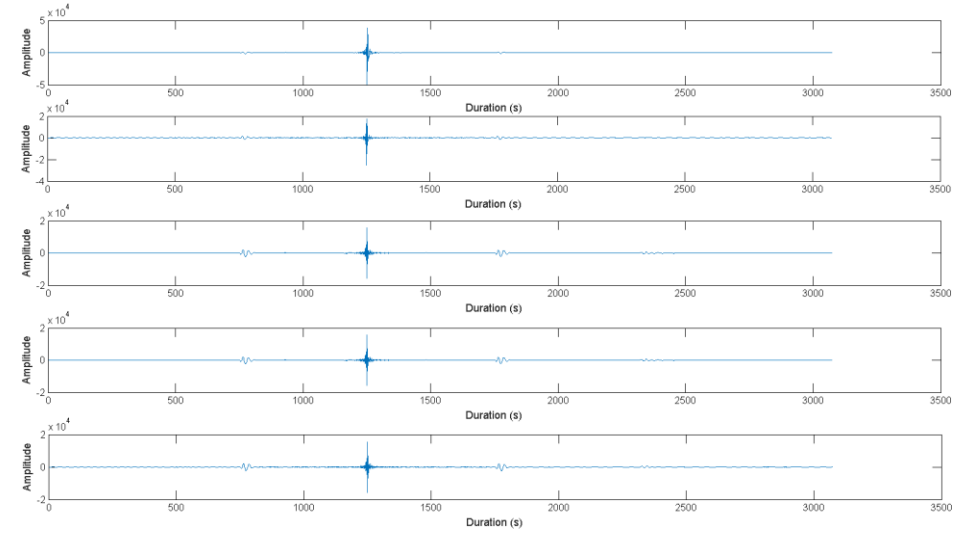
By comparing SOM hits and neighbor maps one can classify the nodes. In this study, four classes are chosen based on the neighbor weight distance maps. Results are shown below for one day of data.

On July 7th there were approximately 1,997 triggers with SNR $>$ 20 that were classified in four classes.

Figure 22: Representatives of four classes of trigger waveforms from the GW channel (DARM) for the July 7th, 2010 run. Triggers in each class are characterized by a similar waveform.


Amplitude
Duration (s)


Amplitude
Duration (s)


Amplitude
Duration (s)

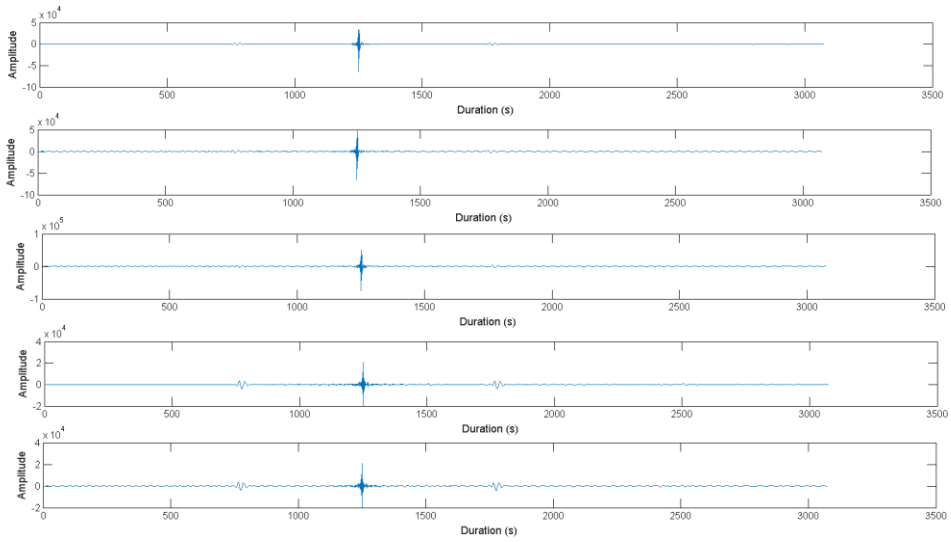


Figure 22 illustrates the SOM glitch classification pipeline's ability to classify triggers based on the morphology of the trigger waveform. Class 1 waveforms have medium to high amplitudes (~1000-15000) with occasional ringing after the peak of the waveform. Class 2 consists of waveforms with a large spectrum of amplitudes (~5000-50000). Class 3 have large amplitude (~20000-50000). Class 4 triggers have variable amplitude between ~8000 and ~50000. The similarity in waveforms by class is further illustrated in figure 23 below.

Figure 23: Trigger waveforms from class 1, 2, 3, and 4 of the DARM channel during July 7, 2010. The amplitudes of the waveforms are plotted as a function of time. The four classes are distinctly different.

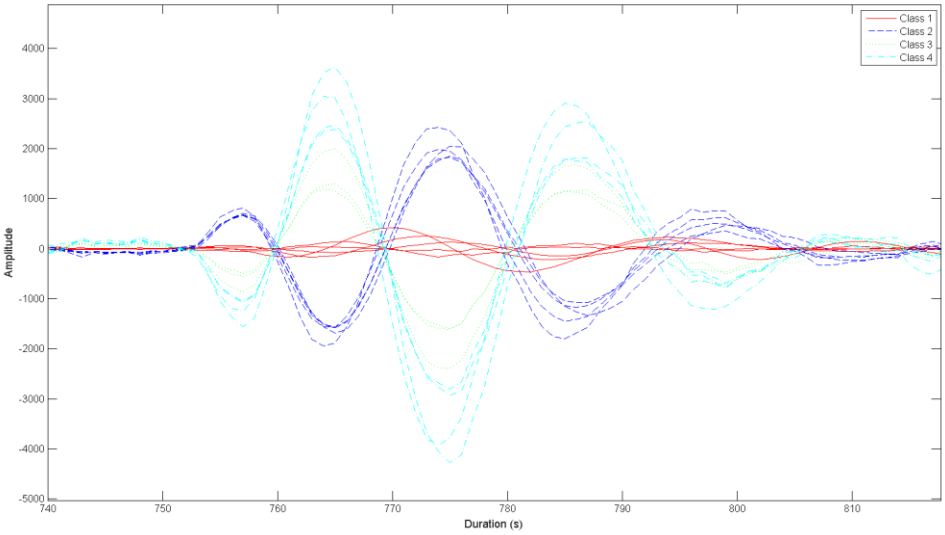
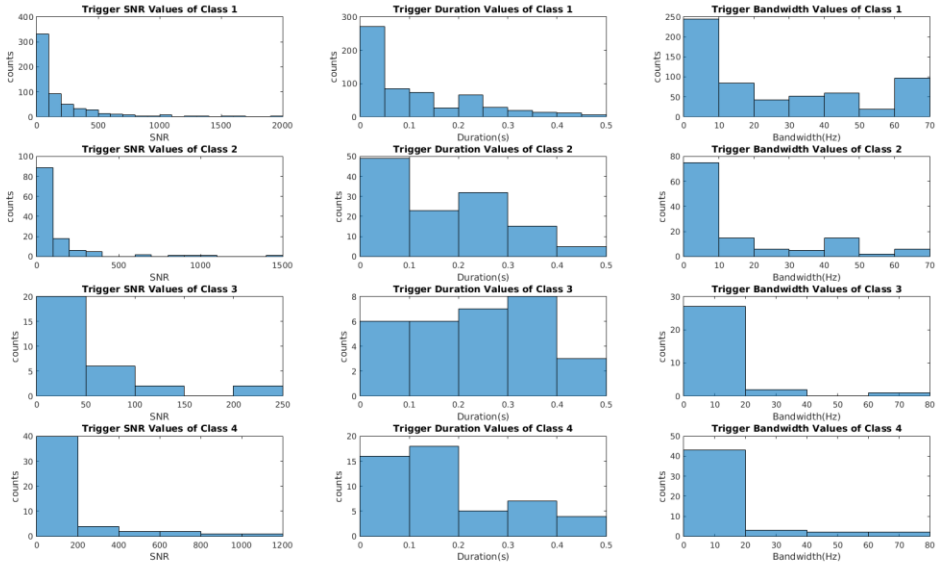


Figure 24 shows histograms of the SNR, duration, and bandwidth values for the four trigger classes. Class 1 consists of triggers with a large spectrum in SNR (20-2000), mostly short duration, and broadband. Class 2 consists of relatively low SNR, largely low duration, short bandwidth triggers with a handful of broadband triggers. Class 3 has medium SNR values (30-250), large duration, short bandwidth triggers. Class 4 contains a large range of SNR, medium to low duration (~0.1s), short bandwidth triggers.

Figure 24: SNR, duration, and bandwidth values for classes 1-4 from the DARM channel during July 7, 2010.

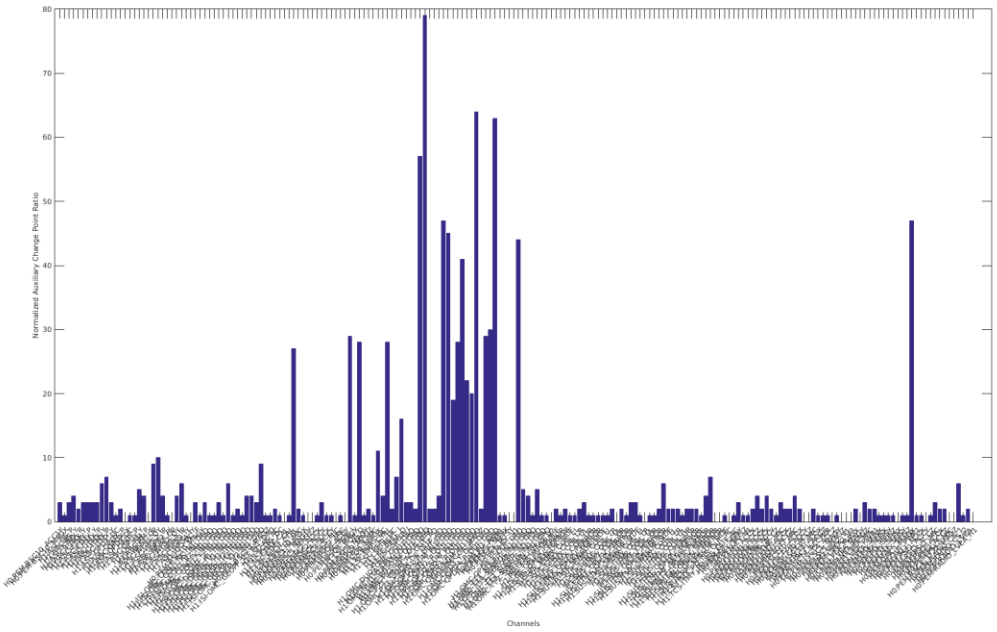


The source of the noise associated to a trigger class can be determined by the highest SNR trigger. The number of triggers in all auxiliary and environmental channels within a 2 second window around the trigger are used to produce trend plots with a time window of +-5 minutes around the high SNR trigger time. If there is a high number of triggers around the high SNR trigger time for a given auxiliary channel, it is likely that that channel identifies the source of the noise for the class. The difference between the standard deviation calculated from all the trigger times found within the time window, σ, and the number of triggers in each auxiliary channel during the two second time window, α, is calculated. The top ten channels with the maximum change are recorded. The probability of the source originating from one of these auxiliary/environmental channels is then determined by

$$\gamma = \frac{\alpha}{\sigma}.$$

The highest SNR trigger found in class 1 of the DARM channel occurred at GPS time 962562049 with an SNR of 1484. The number of triggers in all auxiliary and environmental channels within the 2 second time window around this trigger can be seen in fig. 25.

Figure 25: 196 auxiliary channels with their corresponding number of coincident triggers within the two second time window around the central trigger time. The top three performing channels are the output mode cleaner channels H1:OMC-PD_SUM_OUT_DAQ, H1:OMC-QPD3_SUM_IN1_DAQ, and H1:OMC-QPD_SUM_IN1_DAQ.



Trend plots are made for each auxiliary and environmental channel and serve as a sanity check for the most probable channel source. The auxiliary or environmental channel source probability for the top 10 channels can be seen in fig. 26.

Figure 26: Top ten channels in class 1 with the maximum α are plotted with their corresponding normalized auxiliary change point ratio (γ) values. The most likely source of noise for class 1 is identified by the H1:OMC-QPD3_SUM_IN1_DAQ channel.

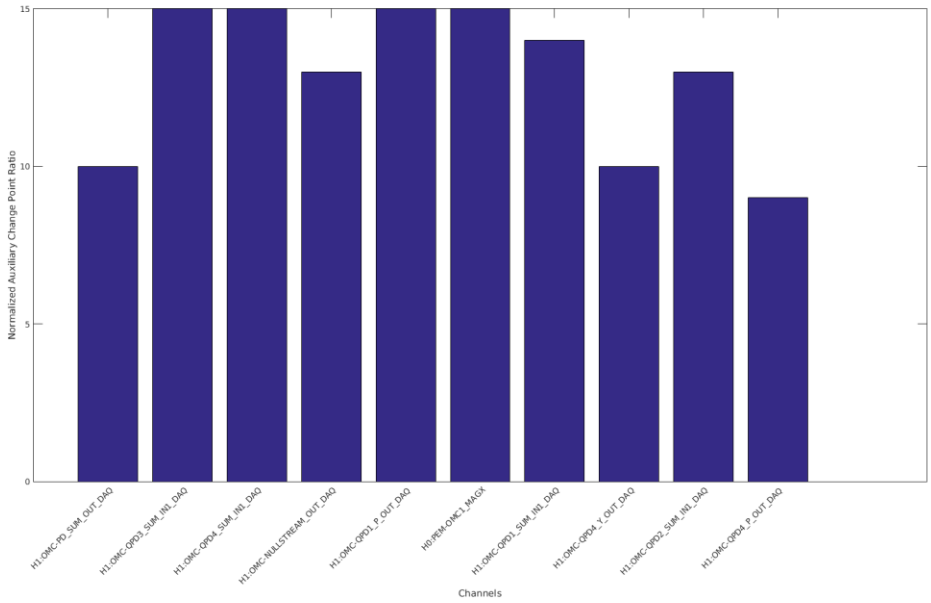
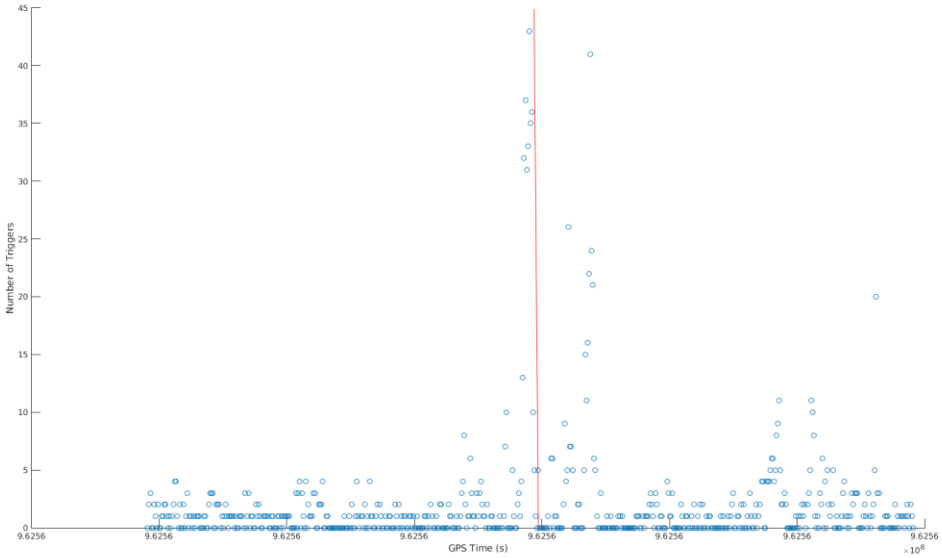


Figure 27: Trend plot of H1:OMC-QPD3_SUM_IN1_DAQ plus/minus 5 minutes around the central trigger time. The number of triggers in the auxiliary channel are plotted as a function of time. The red vertical line represents the central trigger time. A trend up in trigger production occurs around the central trigger time.


Number of Triggers
GPS Time (s)
45
40
35
30
25
20
15
10
5
0
9.6256
9.6256
9.6256
9.6256
9.6256
9.6256
9.6256
9.6256
×10^8

V. Conclusions

In the first part of this thesis, it was found that Omicron detects transients with good efficiency at amplitudes as low as 10^{-22}-10^{-21} well above the strain sensitivity of the LIGO detectors. Omicron also reconstructs the peak time and frequency of injections with good accuracy. Because of its good performance and computational speed, Omicron has been used to characterize Advanced LIGO sub-systems. The Coincidence Finder code we wrote is currently used to perform sanity checks of updated versions of Omicron. The results from this project validate the robustness of Omicron as a trigger generator for use in future observational runs at LIGO and Virgo.

The glitch classification pipeline described in section IV presents strong evidence, or proof-of-principle, that self-organizing maps, when applied to detector characterization, are a useful tool to do low-latency detector characterization over a large multidimensional trigger set. Future studies may focus on optimizing the glitch classification pipeline by adding clickable feature maps, as well as a segment data base. Once fully developed, this tool may be useful to LSC scientists in order to characterize signals in LIGO data.

VI. Acknowledgements

This author would like to thank Dr. Soma Mukherjee and Robert Stone at the University of Texas Rio Grande Valley (UTRGV) for their expertise and guidance. In addition, this author would like to thank the UTRGV REU principle investigator Dr. Joey Key for making the program available to myself and others. Thank you to the University of Florida, Dr. Bernard Whitting, Dr. Guido Muller, and Dr. Andonis Mitidis for allowing me the wonderful opportunity to experience a new and exciting research environment. A special thanks to Dr. Florent Robinet for his constant help and support during the University of Florida IREU (I sometimes wondered how he ever got any work done). Thank you to all the members of the LAL group at Orsay who were extremely warm and welcoming. This author would like to thank his home institution research advisor Dr. Marco Cavaglia for his support and guidance over the past four years. This material is based upon work supported by the National Science Foundation under Grant Number PHY-1461237 and PHY-1005036. This author is also partially supported by his research advisor Dr. Marco Cavaglia and the National Science Foundation through award PHY-1404139 and PHY-1067985.

List of References

[1] A. Einstein, Sitzungsber. K. Preuss. Akad. Wiss. 1, 154 (1918).

[2] The LIGO Scientific Collaboration: B.Abbott, et al, Analysis of LIGO data for gravitational waves from binary neutron stars, Phys. Rev. D 69 (2004) 122001.

[3] Huwyler, Cédric. "Gravitational Waves." *Welcome to My Space at Itp.uzh.ch*. N.p., n.d. Web. 21 Apr. 2016.

[4] Saulson, Peter R. Interferometric Gravitational Wave Detectors. Singapore: World Scientific Publising, 1994. Print.

[5] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari et al., Observation of Gravitational Waves from a Binary Black Hole Merger, Phys. Rev. Lett. 116, 061102 (2016).

[6] Dooley, K. L. et al, "GEO 600 and the GEO-HF Upgrade Program: Successes and Challenges." *Class. Quantum Grav. Classical and Quantum Gravity* 33.7 (2016): 075009. Web.

[7] Accadia, T., Acernese, F., Alshourbagy, M., et al. 2012, Journal of Instrumentation, 7, 3012.

[8] Blair, David et al. "ArXiv.org Physics ArXiv:1602.02872." *[1602.02872] Gravitational Wave Astronomy: The Current Status*. ArXiv, 2015. Web. 02 Apr. 2016.

[9] Dooley, K. et al. Rev. Sci. Instrum. 83, 033109 (2012).

[10] Martynov, D.V. et al, *The Sensitivity of the Advanced LIGO Detectors at the Beginning of Gravitational Wave Astronomy*, LIGO dcc P1500260

[11] Effler, A. *et al.* “Environmental Influences on the LIGO Gravitational Wave Detectors during the 6th Science Run” Class.Quant.Grav. 32 (2015) no.3, 035017 arXiv:1409.5160 [astro-ph.IM]

[12] S. Chatterjee et al, Class. Quant. Grav. 21, S1809 (2004).

[13] M. Cavaglia, D. Trifiro, https://dcc.ligo.org/LIGO-T1300001

[14] The Omega Pipeline, https://geco.phys.columbia.edu/omegar

[15] Robinet, Florent. "GWOLLUM and Friends." GWOLLUM. LAL/Virgo, n.d. Web. <http3A2F2Fvirgo.in2p3.fr2FGWOLLUM2Findex.html3FFriends2Fomicron.html>.

[16] L. Nuttall, LIGO dcc G1501580 (2016).

[17] "Mathematical Libraries | ROOT a Data Analysis Framework." *Mathematical Libraries | ROOT a Data Analysis Framework*. N.p., n.d. Web. 02 Apr. 2016.

[18] Brown, Duncan. "LALApps” LSC Algorithm Library Applications." University of Wisconsin-Milwaukee (2005): 1-182. LSC Algorithm Library Applications. Web. June-July 2014. <https://www.lsc-group.phys.uwm.edu/daswg/projects/lalapps/lalapps-5.0.pdf>.

[19] "Do Cosmic Strings Exist?" *LIGO Scientific Collaboration*. N.p., n.d. Web. 06 May 2016.

[20] Kohonen, Teuvo. ”The Self-organizing Map.” ScienceDirect (1998): 1-6. Web. 18 July 2015.

[21] S. Mukherjee et al, LIGO dcc P1100061 (2011)

[22] M. Shensa, “The Discrete Wavelet Transform: Wedding the A Trous and Mallat Algorithms”, IEE Transactions on Signal Processing, Vol 40, No. 10, 1992. Print.

[23] ”Haar Function.” – from Wolfram MathWorld. N.p., n.d. Web. 19 July 2015.