Optimal vetoing methods for removing glitches that can be modelled from interferometric detector data

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Prasanna Mohan Joshi

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2020

Supervisor: Prof. Sanjeev Dhurandhar

Certificate

This is to certify that this dissertation entitled *Optimal vetoing methods for removing glitches that can be modelled from interferometric detector data* towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Prasanna Mohan Joshi at Inter-University Centre for Astronomy and Astrophysics under the supervision of Prof. Sanjeev Dhurandhar, Emeritus Professor, during the academic year 2019-2020.

S. V. Dhurandhar

Prof. Sanjeev Dhurandhar

P Joshi

Prasanna Mohan Joshi

Committee:

Prof. Sanjeev Dhurandhar

Dr. Suneeta Varadarajan

Declaration

I hereby declare that the matter embodied in the report entitled *Optimal vetoing methods for removing glitches that can be modelled from interferometric detector data*, is the result of the work carried out by me at the Inter-University Centre for Astronomy and Astrophysics, Pune, under the supervision of Prof. Sanjeev Dhurandhar and the same has not been submitted elsewhere for any other degree.

S. V. Dhurandhar

Prof. Sanjeev Dhurandhar

P Joshi

Prasanna Mohan Joshi

Acknowledgment

I would like to thank my project advisor Prof. Sanjeev Dhurandhar for this opportunity to work with him. Under his guidance, I have been able to learn a lot about my field through interesting discussions. His conceptual clarity and his approach to understanding and solving a problem is something that I aspire to gain. I would also like to thank Prof. Sukanta Bose for his advice throughout the duration of my project. He always made time to indulge me in discussion and helped me with any of my problems. I would also like to thank Dr. Suneeta Vardarajan for guiding me as my TAC mentor.

I would like to thank the DST-INSPIRE Scholarship by the Government of India for the financial support during the entire span of my degree. I would like to thank IISER for providing a high quality of undergraduate education and also for creating an environment in which I was able to grow not just as a scientist, but also as a person. I would also like to thank the IUCAA for all their resources including the IUCAA LIGO Data Grid (LDG) cluster, Sarathi.

I would like to thank my friends Palash Singh, Rahul Poddar, Shomik Adhicary, Raj Patil and Prasham Jain for engaging in fruitful discussions about several topics in physics related to our projects. I would like to thank the members of my band - Rahul, Palash, Simran, Akshay and Sunil for all the good times making music and making me realize my passion for it. I would also like to thank all of my friends, for making my time at IISER the best five years of my life.

Finally, I would like to thank my family for their unwavering support ever since I first expressed a desire to become a scientist.

Abstract

The LIGO and VIRGO Detectors have conducted two observation runs, in which a total of ten binary blackhole mergers and neutron star mergers have been detected. The third observation run has already seen several potential candidates for CBCs. Search algorithms involve cross-correlation of data from detectors with a template and searching with a template bank in parameter space. This method works very well, if noise in the detector data is Gaussian and stationary which is not the case. The data is extremely noisy and contains a lot of glitches which produce false triggers. Tests that have been introduced to act as vetos to discriminate between real signals and glitches are called the χ^2 tests. The most successful among them is called the power χ^2. But these tests have also been unable to perform well against all the types of glitches. Thus, development of better χ^2 tests is a very important problem. The unified χ^2 formalism is a rigorous mathematical framework which can be used to study and construct new χ^2 tests. I will use this formalism to construct an optimal χ^2 test specifically to discriminate between signals and short, transient glitches that can be modelled using sine-gaussian functions. I will then demonstrate its performance as compared to the power χ^2 test using simulated data.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General Relativity

The Newton's Law of Gravitations works well for describing the motion of planets and stars far away from each other. But in the region where gravity is strong, it fails to do so. His theory also predicts that a local change in the mass distribution will instantaneously affect the masses in the entire universe. This is in direct conflict with the axioms of Special Relativity that states that there is an upper limit on the speed of information transfer in the universe. To make the law of gravitation compatible with Special Relativity, Einstein introduced General Relativity (GR).

In the theory of GR, space and time are treated together on the same footing, as a four dimensional *spacetime* The presence of matter/energy makes the spacetime curved and this curvature is responsible for the motion of matter/energy in the spacetime. Over the last century several experiments were performed to test GR and it was found that GR is consistent with all the experiments. The first experiment to test GR was done by Eddington in 1919, in which he observed bending of light by gravitational field of a star. GR also explained the precession in the orbit of mercury to a very high accuracy. A previously unknown prediction of GR is the existence of Gravitational Waves (GW) which are ripples in the spacetime. GW were indirectly observed in 1974 by Hulse and Taylor. The decrease in the time period of a binary pulsar system was found to be exactly consistent with energy lost due to GW as predicted in GR. In the more rececnt times, GW from a binary blackhole merger were directly observed by the Laser Interferometer Gravitational-wave Observatory (LIGO) on 14^{th} September, 2015 [1] [2]. Since then, nine additional BBH mergers

and one binary neutron star merger has been observed by LIGO.

1.2 Gravitational waves in weak-field approximation

The interdependence of matter/energy and curvature of spacetime is described using Einsteins equations. Einstein's equations are coupled, non-linear differential equations whose solutions are difficult to find. In regions of spacetime far away any matter or energy source, the gravitational field is weak. The metric of the spacetime $g_{\mu\nu}$), can be written as a perturbation to the Minkowski metric ($\eta_{\mu\nu}$).

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1.1}$$

Then, the Einstein's equations reduce to weak field equations in vacuum, as follows.

$$\Box \bar{h}_{\mu\nu} = 0 \tag{1.2}$$

where $\Box$ is the D'Alembertian operator and $\bar{h}$ is defined as follows.

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h^{\rho}_{\rho} \tag{1.3}$$

This form is analogous to 3-dimensional wave equation without a source. Therefore, above equation has plane wave solutions which we can write as follows.

$$\bar{h}_{\mu\nu} = A_{\mu\nu} e^{ik_{\lambda}x^{\lambda}} \tag{1.4}$$

where $A_{\mu\nu}$ is the tensor amplitude of the wave and k_{λ} is the wave-vector.

The form of weak-field Einstein's equations in Eq. (1.2) is valid only in Lorenz gauge. Therefore, the solutions of the equation should also satisfy the Lorentz gauge condition which is as follows.

$$\partial_{\mu} \bar{h}^{\mu\nu} = 0 \tag{1.5}$$

Chapter 2

Detection of gravitational-waves and Data analysis

When GW pass through a region, the spacetime undergoes stretching and squeezing. This leads to fluctuations in the spacetime interval between two masses close to each other. The strain that is created has order 10^{-22}. In order to detect GW, we need to be able to measure spacetime interval with such sensitivity. Resonant bar detectors were initially used to measure this strain. But it was not sensitive enough to detect the strain of this order. Then, people suggested using ground-based interferometric detectors to perform this measurement. And now after several years of research on developing such intererometers, LIGO has successfully detected 11 sources of GW.

2.1 Ground-based Interferometric gravitational-wave detectors

To detect GW, the interferometers need to be built such that they are sensitive to strain ($\Delta L/L$) of order 10^{-22}. The schematic diagram of the LIGO detectors is given in Fig. (2.1).

As it can be seen in the figure, the interferometer has two perpendicular arms of length 4 km. There is a beam splitter at the intersection of the two arms. A laser which emits coherent beam of light is opposite to one of the arms and a very sensitive photo-diode is placed opposite to the other arm of the interferometer. There are heavy mirrors at the end of both the arms. This is the basic setup of an interferometer. The laser beam incident on the interferometer splits and enters

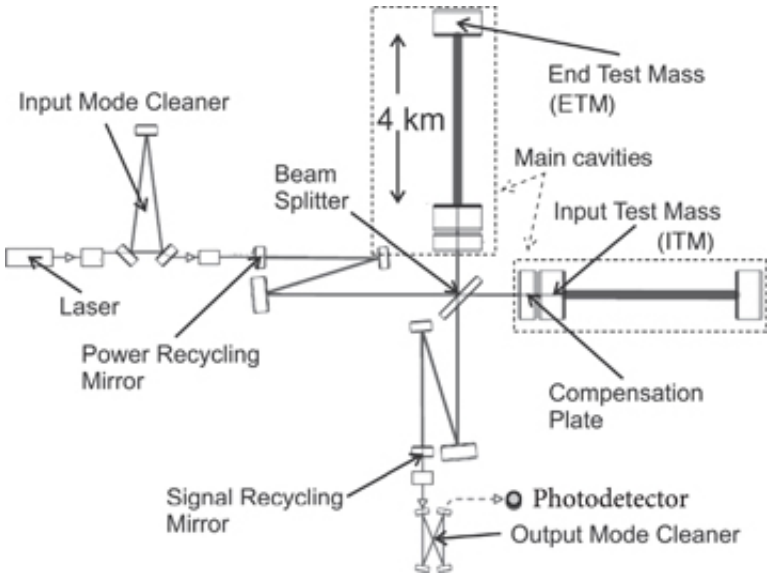


Figure 2.1: A schematic diagram of LIGO detector setup.[3]

both the arms of the interferometer. The light reflects from the mirrors at the end and reflects toward the beamsiplitter. The light passes through the beam splitter again and is incident on the photo-diode. If the arms of the interferometer have exactly equal length, then the light undergoes a perfect destructive interference and intensity of light on the photo diode is zero. But when a Gravitational wave passes through the interferometer, it changes the lengths of the two arms. Then we are able to measure strain using the interference pattern

But in order to achieve the sensitivity, the design has been modified by adding more components. The arms of the interferometers have been converted to Fabry-Perot cavities. This increases the amount of time spent by photons in the arms of the interferometer and thus increases sensitivity. There several other technologies that have made the working of such a sensitive detector possible.

2.2 Sources of noise

One of the major problems in constructing the detector was to reduce the noise in the data. Since the magnitude of the strain that we want to measure is 10^{-22}, all the different types of noise give rise to more strain than that. Several new technologies are being used in the interferometer to reduce the level of noise from different sources. These sources can be classified into two categories: environmental noise and fundamental noise

The environmental noise sources include seismic noise, thermal noise, residual gas noise and coating noise. The seismic noise is caused by motion of the Earths crust, vehilcles moving close enough, etc. The vibrations due to these kinds of sources are very high compared to GW strain. Thermal fluctuations cause damped motions in mechanical system giving rise to vibrations that can lead to strain in the detector. The arms of the interferometer are vacuum sealed. But vacuum is not perfect. The residual gas moelcules scatter the light and cause fluctuations in phase of the light.

The fundamental noise sources include shot noise and radiation pressure noise. Shot noise refers to the quantum fluctuations in the number of photons reaching the photo diode. Radiation pressure noise refers to the noise produced due to dispacement of the mirrors by pressure of high intensity radiation.

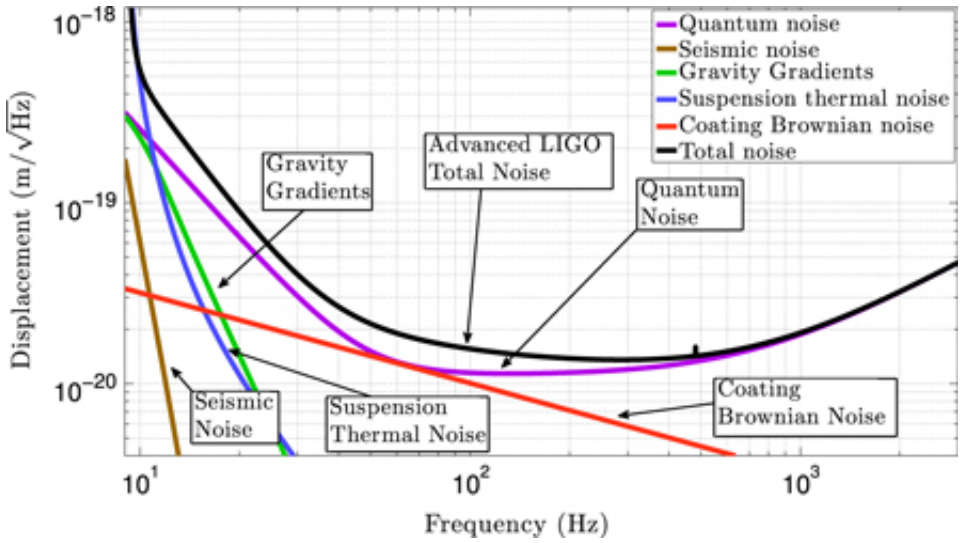


Figure 2.2: Different sources of noise that determine the sensitivity of advanced LIGO [3]

2.3 Data analysis for search of gravitational waves

The LIGO detectors measure the strain in the arms of the detector caused when a GW passes through the detector. The continuous signal is converted to a discrete time series by sampling points at a constant frequency. The sampling rate is decided based on the frequency of the GW to be detected. According to the Nyquist theorem[4], a signal containing highest frequency f can be represented in discrete form if data is sampled at a frequency $2f$ or higher. The maximum

frequency of GW that we want to detect is of the order of a few kHz. Therefore, a sampling frequency $f_s = 16384$ Hz or $f_s = 8192$ Hz is used in LIGO.

The data in the form of discrete time-series is dominated by noise. The signal if it is there, is expected to be weak compared to the noise and therefore, it is buried in the noise. In order to extract the signal from this noise, knowledge of properties of noise is important. But before we do that, convention for the Fourier transform and Inverse Fourier transform that we will follow in our analysis is given below. Let $x(t)$ be a time series and $\tilde{x}(f)$ be the Fourier transform of $x(t)$.

$$\tilde{x}(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} dt \tag{2.1}$$

$$x(t) = \int_{-\infty}^{\infty} \tilde{x}(f) e^{i2\pi ft} df \tag{2.2}$$

2.3.1 Modelling noise in the detector

The noise $n(t)$ in the detector, can be modelled as a random variable. Each point in the discrete time-series is sampled from some distribution. Thus, detection of signal in the data becomes a statistical problem. We make an assumption that the noise has zero mean at every point if time t. Thus, $\langle n(t) \rangle = 0$, where $\langle \rangle$ denotes ensemble average. The autocorrelation function $K(t,t')$ of the noise is another important function that characterizes the noise.

$$K(t,t') = \langle n(t) n(t') \rangle \tag{2.3}$$

In general, the distribution of noise can change over a period of time. If it doesnt, then the noise is said to be stationary. It also implies that all the moments of the noise will remain constant over time. But, we assume a weaker condition that the noise is the detector is Wide Sense Stationary (WSS). This means that only the first and the second moments of the distribution of noise do not change over time.

$$\langle n(t) \rangle = \langle n(t+\tau) \qquad \langle n(t) n(t') \rangle = \langle n(t+\tau) n(t'+\tau) \rangle \tag{2.4}$$

Under this assumption for LIGO noise, the mean of noise remains zero at all times and the auto-

correlation function depends only on the time-difference and not on absolute time.

$$K(t,t') = K(t-t') \tag{2.5}$$

Due to this symmetry of time translation, the auto-correlation in frequency domain can be written as follows.

$$\begin{aligned}
\langle \tilde{n}(f)\tilde{n}^*(f')\rangle &= \int\int dt dt' \langle n(t)n(t')\rangle e^{-i2\pi ft} e^{i2\pi f't'} && (2.6)\\
&= \int\int dt dt' K(t-t') e^{-i2\pi ft} e^{i2\pi f't'} && (2.7)\\
&= \int\int dt d\tau K(\tau) e^{-i2\pi ft} e^{i2\pi f'(t+\tau)} && (2.8)\\
&= \int d\tau K(\tau) e^{i2\pi f'\tau} \times \delta(f-f') && (2.9)\\
\langle \tilde{n}(f)\tilde{n}^*(f')\rangle &= S_h(f)\delta(f-f') && (2.10)
\end{aligned}$$

where $S_h(f)$ is the Fourier transform of the auto-correlation function K. It is known as the two-sided Power Spectral Density (PSD) of the noise in the detector. Since, $K(\tau)$ is an even function by its definition, $S_h(f)$ is a purely real function. $S_h(f)$ is also a positive definite function ($S_h(f) \geq 0$). $S_h(f)$ is also an even function. It can be folded over itself and be defined only for $f \geq 0$. Then it is called a one-sided PSD. The amplitude of one-side PSD is twice that of a two-sided PSD. If the PSD is a constant function of frequency f, then the noise is said to be white. Otherwise, it is said to be coloured.

Uptil now, we have not specified a distribution for the noise random variable. If the distribution of noise is Gaussian, then we can define a noise-weighted inner product between two time-series in the frequency domain.

$$(\mathbf{x},\mathbf{y}) = 4\,\mathrm{Re}\int_0^\infty \frac{\tilde{x}^*(f)\tilde{y}(f)}{S_h(f)} df \tag{2.11}$$

Here, the PSD that is used is the one-sided PSD.

2.4 Matched Filter

Consider that a segment of data ($x(t)$) in the time interval $[0,T]$ contains a signal of the form $h(t)$ and noise ($n(t)$).

$$x(t) = n(t) + h(t) \tag{2.12}$$

If the $K(t,t')$ is the auto-correlation function of the noise in the detector, then the matched filter ($q(t)$) is defined as the solution to the following integral equation [4].

$$\int_0^T K(t,t')q(t')dt' = h(t) \tag{2.13}$$

We make an assumption that the noise is WSS, then the equation can be written as follows.

$$\int_0^T K(t-t')q(t')dt' = h(t) \tag{2.14}$$

The above equation is a convolution of $K(\tau)$ with $q(t)$. We can convert this to the Fourier domain in which it is easier to solve. If $\tilde{h}(f)$ is the Fourier transform of the signal waveform, the solution of the equation i.e. the matched filter in the Fourier domain ($\tilde{q}(f)$) can be written as follows.

$$\tilde{q}(f) = \frac{\tilde{h}(f)}{S_h(f)} \tag{2.15}$$

where $S_h(f)$ is the two-sided PSD of the noise. We define the statistic c which is the correlation of the data ($x(t)$) with the matched filter.

$$c = \int_{-\infty}^{\infty} \tilde{q}^*(f)\tilde{x}(f)df \tag{2.16}$$

The statistic c is a random variable as it is a function of the data that contains noise. The signal-to-noise ratio of data with the matched filter is defined as the ratio of mean of c to the standard deviation of c. Since, the noise has mean zero, the mean of c can be shown to be the following.

$$\langle c \rangle = \int \tilde{q}^*(f)\tilde{h}(f)df \tag{2.17}$$

The standard deviation of noise (σ_c) is derived with the use of the definition of two-sided PSD in Eq. (2.10).

$$\sigma_c^2 = \int |\tilde{q}(f)|^2 S_h(f) df \tag{2.18}$$

Then the expression for SNR (ρ) is as follows.

$$\rho = \frac{\int \tilde{q}^*(f)\tilde{h}(f) df}{\int |\tilde{q}(f)|^2 S_h(f) df} \tag{2.19}$$

From Eq. (2.19), ρ is invariant under scaling of $\tilde{q}$. Using this property we can simplify the expression for the SNR by imposing the condition that the denominator should be 1.

$$\int |\tilde{q}(f)|^2 S_h(f) df = 1 \tag{2.20}$$

This fixes the normalisation factor or scale factor of the matched filter. Thus the SNR can now be defined as follows, subject to the condition in the above.

$$\rho = \int \tilde{q}^*(f)\tilde{h}(f) df \tag{2.21}$$

It can be proved that if data contains signal of form $h(t)$, then out of all the possible filters, the matched filter gives the highest SNR. Thus, we have the maximum probability of detecting the signal if we use a matched filter that corresponds to that signal waveform.

Thus, in order to search for GW from different sources, we need to know the waveform of the GW emitted by all the different possible sources. Such waveforms or templates can be generated by solving the Einstein's equations for that particular system. Due to the non-linear and coupled nature of the equations, exact solutions with a source term cannot be found analytically. But we can make approximations to the Einstein's equations and obtain an approximate analytical solution. For the particlar case of a compact binary coalescence, we solve the Einstein's equations in the Newtonian approximation and and write down the analytical expression of the waveform. Analytical expression of the Newtonian template is as follows.

$$\tilde{h}(f) = h_0 f^{-7/6} e^{-i\psi(f)} \tag{2.22}$$

The frequency dependence of phase of the waveform, $\psi(f)$ is as follows.

$$\psi(f) = 2\pi f t_c - \phi_c - \frac{\pi}{4} + \frac{3}{128}(\pi f \mathscr{M})^{-5/3}$$

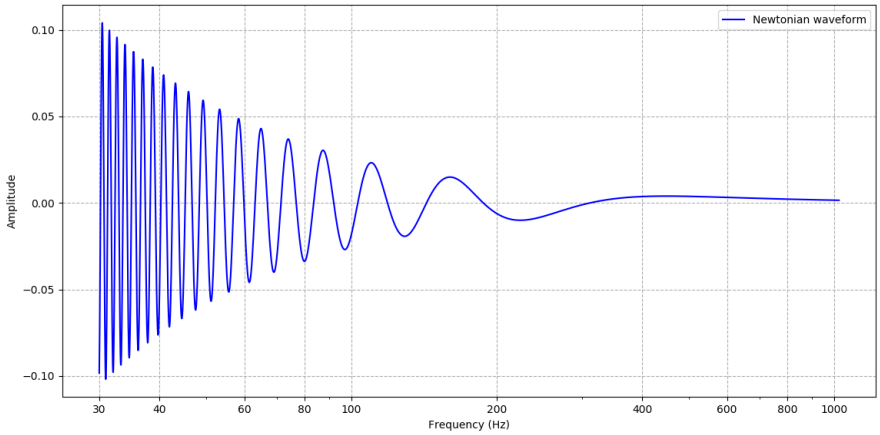


Figure 2.3: Zero phase waveform of a Newtonian template corresponding to a binary with component masses $15M_{\odot} - 15M_{\odot}$

where t_c is the time of coalescence or time of trigger of the binary, ϕ_c is the phase of the waveform at t_c and $\mathscr{M}$ is the chirp mass of the binary system in units of time. It is related to chirp mass in its natural units by the equation $\mathscr{M} = GM_{chirp}/c^3$. The chirp mass of a binary system is the following function of its component masses.

$$M_{chirp} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \tag{2.23}$$

h_0 is the normalization factor of the of the template. We can express $\psi(f)$ in terms of a more suitable parameter which is the chirp time τ_0 instead of $\mathscr{M}$. It is the time taken by the binary system to coalesce, starting from a fiducial frequency f_s We can consider this to be the seismic cutoff frequency $f_a = 10$ Hz. The expression for τ_0 is as follows.

$$\tau_0 = \frac{5}{256\pi f_s} (\pi \mathscr{M} f_s)^{-5/3} \tag{2.24}$$

We can write ψ in terms of τ_0 as shown in the following equation.

$$\psi(f) = 2\pi f t_c - \phi_c - \frac{\pi}{4} + \frac{6\pi f_s \tau_0}{5} \left(\frac{f}{f_s}\right)^{-5/3}$$

Better approximations to the waveform which can be calculated using higher order terms in

perturbation theory. Templates can also be generated by solving the Einstein's equations using numerical techniques. Particularly in case of the CBC templates, a very good approximation to the waveform is obtained by considering three phases of the waveform, Inspiral, Merger and Ringdown. In the inspiral part of the system, the components of the binary revolve around each other in an orbit that keeps shrinking and the frequency of revolution keeps increasing. This phase of the waveform is modelled by using Post-Newtonian theory which is basically the use of perturbation theory to calculate higher order terms in the expansion in the powers of v/c. In merger part of the system, the two bodies come in contact with each other and merge to form a single entity. This part of the system is the least understood and the waveform in this region is generated by running simulation of Numerical relativity. In the ringdown part of the waveform, the entity created in the merger loses its asymmetry as it emits GW. This part of the waveform is modelled by using black-hole perturbation theory.

As seen in case of the Newtonian template, these templates depend on several parameters which can be classified as intrinsic and extrinsic parameters. Intrinsic parameters are the parameters of source itself like component masses (m_1 and m_2) and spins of the bodies (S_1 and S_2). Extrinsic parameters are the parameters of source with respect to the detector. It includes time of coalescence (t_0), phase at the time of coalescence (ϕ_c), Distance from binary to Earth, Sky position, etc. Thus a search for GW signals with all possible values of parameters is required.

If $\tilde{h}(f;\lambda^{\alpha})$ is a general template with parameters λ^{α}, then we can write the expression for the detection statistic c in Eq. (2.16) in terms of this template. We will also use the property that if data is real in time domain, $\tilde{x}(-f) = \tilde{x}^*(f)$ and redefine the integral over only the positive range of frequencies.

$$c = 4Re\int_0^{\infty} \frac{\tilde{h}^*(f;\lambda^{\alpha})\tilde{x}(f)}{S_h(f)}df \tag{2.25}$$

Here, the $S_h(f)$ is the one-sided PSD. The normalisation condition for the matched filter in Eq. (2.20)translates to the normalisation condition for the template as follows.

$$\int \frac{|\tilde{h}(f)|^2}{S_h(f)}df = 1 \tag{2.26}$$

We can perform a search over parameter t_c by shifting the template $h(t)$ with $t_c = 0$ to $h(t-t_0)$ where $t_c = t_0$ The corresponding change in the Fourier domain is from $\tilde{h}(f)$ to $\tilde{h}(f)e^{i2\pi f t_0}$. Thus,

we get the detection statistic as a function of t_0

$$c(t_0) = 4Re \int_0^\infty \frac{\tilde{h}^*(f;\lambda^\alpha)\tilde{x}(f)}{S_h(f)} e^{i2\pi f t_0} df \tag{2.27}$$

This calculation if done by computing the integral over both positive and negative frequencies, using the two-sided PSD, it attains the form of an Inverse Fourier transform. Using this property, searching over t_c becomes a very simple job. It has been shown in Ref. [5], that we do not need to define templates over the extrinsic parameters and only over the intrinsic parameters - masses (m_1 and m_2) and spins (S_1 and S_2). The search over extrinsic parameters like t_c can be done by scanning over the time domain as shown above. For searching over phase ϕ_c, we can decompose the template at an arbitrary phase into a linear combination of two templates with phases 0 and $\pi/2$ respectively. We can compute the detection statistic individually for the two templates and then the net detection statistic is the square root of the sum of the squares of the two individual detection statistics.

$$c(t_0) = \sqrt{c_0(t_0)^2 + c_{\pi/2}(t_0)^2} \tag{2.28}$$

For the intrinsic parameters, templates need to be defined at regular intervals in the parameter space [5][6][7]. The value of c is computed for each of these templates for all values of the extrinsic parameters. Once the values of the c are calculated covering the entire parameter space of the template, its maximum value is computed. This value is compared with a preset threshold. If this value is above the threshold, then we claim that there is a trigger. This threshold is set by considering the value of the detection statistic for data containing only noise. The parameter values that correspond to the maximum value of c are reported as the parameters of the source of the GW. This is the mechanism of the search for GW using the matched filter test.

All the above analysis was done assuming the stationary and gaussian nature of noise in the detector. But the actual noise in the detector is not entirely stationary. Instead, it is stationary over small periods of time. The distribution of noise has a small non-gaussian component. These approximations affect the performance of the matched filter test. These effects can produce false triggers also. Therefore, it is necessary to devise a test to discriminate between real signals and false triggers.

Chapter 3

χ^2 discriminator tests

The matched filtering procedure using a template bank provides a very good way to search for signals in data. However, all the triggers that are generated by the procedure may not be from real GW signals. It is possible that the noise in the detector can manifest in some way to produce false triggers with the matched filter test. To ensure that we have a real GW signal, we can construct tests to check consistency of the trigger with the modelled waveform. Such tests involve comparison of the distribution of power in different frequency ranges of the power spectrum which will be very different for a real signal and a false trigger. The χ^2 test is a good example of such a consistency check. It can effectively distinguish between a real signal and a false trigger.

The most widely used χ^2 test for such a discrimination is the *power* χ^2 test proposed by Bruce Allen in 2005 [8] [9]. I will briefly describe the construction of this test in the following section.

3.1 The Power χ^2

3.1.1 Construction of the power χ^2 test

The main idea of this test is to compare the power spectrum of data to the power spectrum of the template in different frequency bins. If f_{low} and f_{up} are the lower and upper frequency cutoffs considered for the matched filter test, it is possible to divide the frequency interval into p number of non-overlapping bins by choosing $f_1, f_2, ..., f_{p-1}$ such that, $f_{low} < f_1 < f_2 < .... < f_{p-1} < f_{up}$.

We can label the frequency bins in the following way.

$$\begin{aligned}
\Delta f_1 &= \{f | f_{low} < f < f_1\} \\
\Delta f_2 &= \{f | f_1 < f < f_2\} \\
\Delta f_3 &= \{f | f_2 < f < f_3\} \\
&\cdot \\
&\cdot \\
\Delta f_p &= \{f | f_{p-1} < f < f_{up}\}
\end{aligned} \tag{3.1}$$

Using the definition of the matched filter, it is possible to define an inner product on each of these bins in the following way.

$$(\mathbf{x}, \mathbf{y})_j = 4\,\mathrm{Re} \int_{\Delta f_j} \frac{\tilde{x}^*(f)\tilde{y}(f)}{S_h(f)} df \tag{3.2}$$

where $j = 1, 2, ..., p$. As the bins are non-overlapping, the sum of the inner products of two vectors in each of the bins is equal to the matched filter.

$$(\mathbf{x}, \mathbf{y}) = \sum_j (\mathbf{x}, \mathbf{y})_j \tag{3.3}$$

Using these inner products, we choose the boundaries of bins such that the power of the template in each of the bins is equal. Thus the power in each of the bins for a normalised template ($(\mathbf{h}, \mathbf{h}) = 1$) is $1/p$.

$$(\mathbf{h}, \mathbf{h})_j = \frac{1}{p} \tag{3.4}$$

If $\mathbf{x}$ is the data segment, the we can compute the correlation of the data with the template using the restricted inner product in Eq. (3.2) in each of the bins. We denote the correlation, or the SNR in the j^{th} frequency bin as c_j.

$$c_j = (\mathbf{x}, \mathbf{h})_j \tag{3.5}$$

The total SNR of the data segment is equal to the sum of the SNR in the frequency bins. This follows from the property of the restricted inner products in Eq. (3.3) Now, the aim of the power χ^2 is to capture the difference between the observed SNR (c_j) and the expected power (c/p) in the each bin. If we denote this difference by $\Delta c_j = c_j - c/p$, then we can define the power χ^2 statistic

as follows:

$$\chi^2 = p \sum_{j=1}^{p} \left(\Delta c_j\right)^2 \tag{3.6}$$

3.1.2 The statistics of power χ^2

The ability of the power χ^2 to act as a good discriminator comes from its statistical properties. In order to study these properties, we first need to study the statistics of the c and c_j. We assume that the noise in the detector is zero-mean Gaussian noise. Then, the expextation value of c can be calculated as follows.

$$\begin{aligned}
\langle c \rangle &= \langle (\mathbf{x}, \mathbf{h}) \rangle \\
&= \langle (\mathbf{n}, \mathbf{h}) \rangle + \langle (\mathbf{h}, \mathbf{h}) \rangle \\
&= (\mathbf{h}, \mathbf{h}) \\
\langle c \rangle &= 1
\end{aligned} \tag{3.7}$$

The second moment of c can be calculated as shown below.

$$\begin{aligned}
\langle c^2 \rangle &= \langle (\mathbf{x}, \mathbf{h})^2 \rangle \\
&= \langle [(\mathbf{n}, \mathbf{h}) + (\mathbf{h}, \mathbf{h})]^2 \rangle \\
&= \langle (\mathbf{n}, \mathbf{h})^2 \rangle + 2 \langle (\mathbf{n}, \mathbf{h}) \rangle (\mathbf{h}, \mathbf{h}) + (\mathbf{h}, \mathbf{h})^2 \\
\langle c^2 \rangle &= \langle (\mathbf{n}, \mathbf{h})^2 \rangle + 1
\end{aligned} \tag{3.8}$$

The expectation value of the quantity $(\mathbf{n}, \mathbf{h})^2$ can be calculated by writing the explicit form of the inner product and using the property $\langle \tilde{n}^*(f) \tilde{n}(f') \rangle = S_h(f) \delta(f - f')$ of stationary noise. Substituting the above equation in Eq. (3.8), we get the following.

$$\langle c^2 \rangle = 2 \tag{3.9}$$

The expectation value of c_j can be calculated in a way similar to the derivation of $\langle c_j \rangle$

$$\begin{aligned}
\langle c_j \rangle &= (\mathbf{h}, \mathbf{h})_j \\
\langle c_j \rangle &= \frac{1}{p}
\end{aligned} \tag{3.10}$$

The second moment of c_j can also be calculatd by a proecedure similar to the derivation of $\langle c^2 \rangle$.

$$\begin{aligned} \langle c_j^2 \rangle &= \langle (\mathbf{n},\mathbf{h})_j^2 \rangle + (\mathbf{h},\mathbf{h})_j^2 \\ &= (\mathbf{h},\mathbf{h})_j + (\mathbf{h},\mathbf{h})_j^2 \\ \langle c_j^2 \rangle &= \frac{1}{p} + \frac{1}{p^2} \end{aligned} \tag{3.11}$$

From above equations it is very clear that $\langle \Delta c_j \rangle = 0$. The expression for $\langle \left(\Delta c_j\right)^2 \rangle$ can be derived as follows.

$$\begin{aligned} \langle \left(\Delta c_j\right)^2 \rangle &= \left\langle \left(c_j - \frac{c}{p} \right)^2 \right\rangle \\ &= \langle c_j^2 \rangle + \frac{\langle c^2 \rangle}{p^2} - 2\frac{\langle c_j c \rangle}{p} \\ \langle \left(\Delta c_j\right)^2 \rangle &= \frac{1}{p}\left(1 - \frac{1}{p}\right) \end{aligned} \tag{3.12}$$

Using the expressions derived above, we can compute that first and the second moment of the power χ^2. Their expressions are as follows.

$$\langle \chi^2 \rangle = p - 1 \qquad\qquad \langle [\chi^2]^2 \rangle = p^2 - 1 \tag{3.13}$$

3.1.3 Properties of power χ^2

From the definition, we can see that if the data contains only signal, then we see that the $\chi^2 = 0$. If the data contains pure Gaussian noise, then it follows a χ^2 distribution with $p-1$ degrees of freedom is a more non-trivial claim. The proof of this statement is given in [8]. The first and the second moments calculated in Eq. (3.13) are the same as that of a χ^2 distribution with $p-1$ d.o.f. These quantities determine the width of the χ^2 distribution which is $\sqrt{2(p-1)}$. If the data contains signal and Gaussian noise, it also follows a χ^2 distribution with $p-1$ degrees of freedom. Therefore, we can claim that if data contains signal and Gaussian noise, there is a very high probability that the χ^2 value will lie in the interval $[p-1-\sqrt{2(p-1)}, p-1+\sqrt{2(p-1)}]$. But in case of glitches, due to the mismatch between the power in each of the frequency bins, the χ^2 will be very high. Thus, this test can act as a good discriminator.

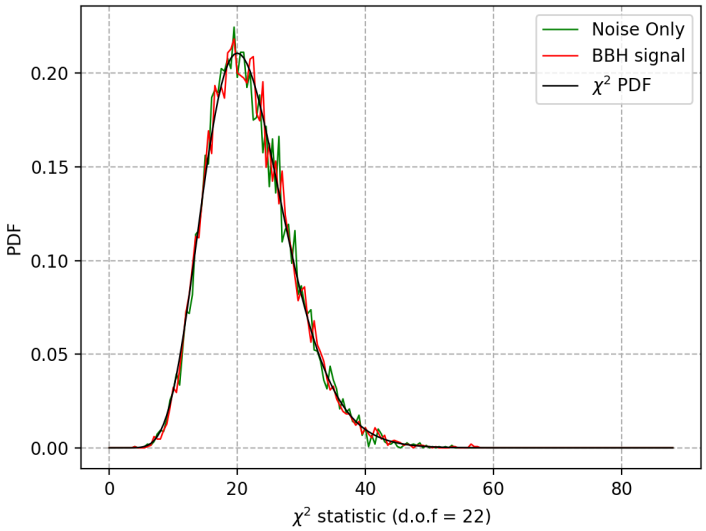


Figure 3.1: The above plot contain the distribution of values of χ^2 for 10000 data segments along with a plot of χ^2 distribution. Each of the data segments contained a different realisation of noise. The BBH template with component masses of $25M_\odot - 25M_\odot$ was used as a signal with perfect match

The power χ^2 test is not the only test that has been introduced to function as a discriminator between signals and glitches. Several other χ^2 tests have been suggested in [10]. The performance of these tests varies for different values of parameters of the template and glitches. A general formalism that encompasses all the different possible χ^2 tests will explain these differences in the performance of the χ^2 tests. We will study such a formalism in the next section.

3.2 The Unified χ^2

The unified χ^2 formalism [11] is a mathematical formalism which provides a rigid framework to describe any χ^2 discriminator test. As mentioned in the previous section, this formalism has the ability to explain the performance of different χ^2 tests for different ranges of parameters of templates and glitches. Moreover, it also gives a general mechanism for constructing all such possible χ^2 tests. Thus, it is possible to use this mechanism to construct an optimal χ^2 discriminator which can perform better than all other χ^2 tests.

3.2.1 Mathematical Formalism

Consider a segment of data in the time interval $[0, T]$ which is sampled at a certain constant rate f_s. This segment belongs to the space of all the possible data trains in the time interval $[0, T]$ sampled at the frequency f_s. This space is a Hilbert space, denoted by $\mathscr{D}$, with an inner product derived from the definition of the matched filter. If $\mathbf{x}$ and $\mathbf{y}$ are two vectors in $\mathscr{D}$, then the inner product $(,)$ is defined as follows.

$$(\mathbf{x},\mathbf{y}) = 4\,\mathrm{Re}\int_{f_{low}}^{f_{up}} \frac{\tilde{x}^*(f)\tilde{y}(f)}{S_h(f)}df \tag{3.14}$$

where $S_h(f)$ is the one-sided power spectral density (PSD) of the noise in the detector. With this definition of inner product, we can define an L_2-norm on this space as follows.

$$\|\mathbf{x}\|^2 = (\mathbf{x},\mathbf{x}) \tag{3.15}$$

A binary black-hole (BBH) template is a vector in this Hilbert space and is denoted by $\mathbf{h}(\lambda^a)$ where λ^a $(a = 1,2,...,m)$ are the parameters of the template.

We will first define the unified χ^2 for a single template vector $\mathbf{h}$, with a fixed set of parameters. Consider the subspace of $\mathscr{D}$ which is orthogonal to the vector $\mathbf{h}$ denoted by $\mathscr{N}_{\chi^2}(\mathbf{h})$.

$$\mathscr{N}_{\chi^2}(\mathbf{h}) = \{\mathbf{x} \in \mathscr{D} | \, (\mathbf{x},\mathbf{h}) = 0\} \tag{3.16}$$

We define the χ^2 using a p-dimensional subspace $\mathscr{S}$ of $\mathscr{N}_{\chi^2}(\mathbf{h})$. Let $\mathbf{x}_{\mathscr{S}}$ be the projection of data vector $(\mathbf{x})$ onto this subspace $\mathscr{S}$. Then, the χ^2 is defined as the square of the norm of the projected data vector.

$$\chi^2 = \|\mathbf{x}_{\mathscr{S}}\|^2 \tag{3.17}$$

Let $\mathbf{e}_\alpha$ $(\alpha = 1,2,...,p)$ be an orthonormal basis on $\mathscr{S}$. Therefore, $\left(\mathbf{e}_\alpha,\mathbf{e}_\beta\right) = \delta_{\alpha\beta}$. We can infer the follwing properties of the χ^2.

1. The χ^2 can be defined in terms of the basis vectors as given in the following equation.

$$\chi^2(\mathbf{x}) = \sum_{\alpha=1}^{p} |\,(\mathbf{x},\mathbf{e}_\alpha)\,|^2 \tag{3.18}$$

2. $\mathscr{S}$ is orthogonal to the template vector $\mathbf{h}$. Therefore, if data contains only signal ($\mathbf{x} = \mathbf{h}$), $\chi^2(\mathbf{x}) = 0$

3. If data contains only zero-mean Gaussian noise with PSD $S_h(f)$ ($\mathbf{x} = \mathbf{n}$), then each term in the sum in Eq. (3.18) is the square of a Gaussian random variable with unit variance. Therefore, $\chi^2(\mathbf{n})$ will follow a χ^2 distribution with p degrees of freedom.

$$\chi^2(\mathbf{n}) = \sum_{\alpha=1}^{p} |\,(\mathbf{n}, \mathbf{e}_\alpha)\,|^2 \tag{3.19}$$

The properties (2) and (3), which are also true for the power χ^2, are very important properties for a χ^2 test. These properties set the interval which contains 99% of the χ^2 values for a data segment containing a signal and Gaussian noise. If p is the number of degrees of freedom of the χ^2, then this interval is $[p - 3\sqrt{2p}, p + 3\sqrt{2p}]$. Moreover, they are independent of the choice of orthonormal basis on the subspace $\mathscr{S}$ and also the choice of the subspace $\mathscr{S}$ itself. This freedom in the choice of subspace is the key feature of this formalism which can be exploited to create different χ^2 tests. Also, given a χ^2 test, it is possible to find a subspace of $\mathscr{N}_{\chi^2}(\mathbf{h})$ which corresponds to that particular χ^2 test. This has been done for the power χ^2 test in [11]

The above definition is in context of a template with fixed values of parameters. For a general template $\mathbf{h}(\lambda^a)$, the parameter space is m-dimensional. Each point in this parameter space maps to a vector in $\mathscr{D}$. The space that is spanned by these vectors is an m-dimensional subspace of $\mathscr{D}$. For each of the template vectors, we can define a p-dimensional subspace $\mathscr{S}$ orthogonal to the template vector and define χ^2 in the same fashion as described above for the case of a single template. Such a mathematical structure, if constructed smoothly, is a vector bundle of $m + p$ dimensions.

3.2.2 Constructing a generic χ^2 test

From the definiton of the χ^2, we see the choice of subspace $\mathscr{S}$ defines the χ^2 statistic. This choice determines the performance of the test. In order for it to function as a good discriminator, the χ^2 value for a glitch should very high compared to the mean of the χ^2 distribution. Thus, $\mathscr{S}$ must be chosen such that the projection of a glitch on $\mathscr{S}$ is as high as possible. We can use the known glitch vectors themselves or a suitable model in order to build this subspace. Let $\mathbf{y}_\alpha$ be a set of suitable chosen, linearly independent vectors. Then the χ^2 will be calculated using the difference between the observed and expected correlation of the data with each of the vectors $\mathbf{y}_\alpha$

Consider template vector $\mathbf{h}$ with fixed values of parameters. It can be decomposed into two template vectors with phases 0 and $\pi/2$, denoted respectively by $\mathbf{h}_0$ and $\mathbf{h}_{\pi/2}$. In order to compute total SNR, we need to compute the correlation of the data vector $\mathbf{x}$ with both of these template vectors. In case of a perfect match between signal and template, signal vector ($\mathbf{s}$) will have the same set of values of the parameters as the template. Then we can write the signal vector as follows.

$$\mathbf{s} = (\mathbf{s}, \mathbf{h}_0)\, \mathbf{h}_0 + \left(\mathbf{s}, \mathbf{h}_{\pi/2}\right) \mathbf{h}_{\pi/2} \tag{3.20}$$

The correlation of this signal vector with the $\mathbf{y}_\alpha$ is as follows.

$$(\mathbf{s}, \mathbf{y}_\alpha) = (\mathbf{s}, \mathbf{h}_0)\, (\mathbf{h}_0, \mathbf{y}_\alpha) + \left(\mathbf{s}, \mathbf{h}_{\pi/2}\right) \left(\mathbf{h}_{\pi/2}, \mathbf{y}_\alpha\right) \tag{3.21}$$

The values of the correlations of data vector with the templates, denoted by c_0 and $c_{\pi/2}$, and their mean values are as follows.

$$c_0 = (\mathbf{x}, \mathbf{h}_0) \qquad c_{\pi/2} = \left(\mathbf{x}, \mathbf{h}_{\pi/2}\right) \tag{3.22}$$

$$\langle c_0 \rangle = (\mathbf{s}, \mathbf{h}_0) \qquad \langle c_{\pi/2} \rangle = \left(\mathbf{x}, \mathbf{h}_{\pi/2}\right) \tag{3.23}$$

Using the equations (3.21), (3.22) and (3.23), we can write the expected correlation (c^e_α) of the data vector along $\mathbf{y}_\alpha$ as follows.

$$c^e_\alpha = c_0\, (\mathbf{h}_0, \mathbf{y}_\alpha) + c_{\pi/2} \left(\mathbf{h}_{\pi/2}, \mathbf{y}_\alpha\right) \tag{3.24}$$

The observed correlation (c^o_α) of data vector with $\mathbf{y}_\alpha$ is as follows.

$$c^o_\alpha = (\mathbf{x}, \mathbf{y}_\alpha) \tag{3.25}$$

The difference between the observed and expected correlation $\Delta c_\alpha(\mathbf{x})$ for each $\mathbf{y}_\alpha$ can be computed to be as follows.

$$\Delta c_\alpha(\mathbf{x}) = (\mathbf{x}, \Delta\mathbf{y}_\alpha) \tag{3.26}$$

where we define $\Delta\mathbf{y}_\alpha$ as follows.

$$\Delta\mathbf{y}_\alpha = \mathbf{y}_\alpha - (\mathbf{y}_\alpha, \mathbf{h}_0)\, \mathbf{h}_0 - \left(\mathbf{y}_\alpha, \mathbf{h}_{\pi/2}\right) \mathbf{h}_{\pi/2} \tag{3.27}$$

These vectors are orthogonal to the templates $\mathbf{h}_0$ and $\mathbf{h}_{\pi/2}$, hence the span of these vectors forms the subspace $\mathscr{S}$. We have constructed a χ^2 test by using an arbitrary set of vectors $\mathbf{y}_\alpha$. The square of the norm of the data vector projected on this subspace is the value of the χ^2 statistic. But we cannot compute the χ^2 by taking the sum of the squares of Δc_α because Δc_α are correlated Gaussian random variables. This is because $\Delta \mathbf{y}_\alpha$ are not orthogonal to each other. We can orthonormalize this set of vectors and hence generate an orthonormal basis on $\mathscr{S}$. Then, we can compute χ^2 using the formula in Eq. (3.18)

$$\chi^2(\mathbf{x}) = \sum_{\alpha=1}^{p} |\,(\mathbf{x}, \mathbf{e}_\alpha)\,|^2 \tag{3.28}$$

With an appropriate selection of vectors $\mathbf{y}_\alpha$, we can thus construct a χ^2 test using the unified formalism.

Chapter 4

Unified χ^2 for Sine-gaussian glitches

Several short-duration, high-amplitude noise transients have been found in interferometer data. Due to their high amplitude, these transients can produce a high SNR upon using the matched filter test. If the SNR is above the threshold, then these noise artifacts can be be falsely classified as gravitational-wave signals. Such noise transients are called glitches. Glitches can be classified in several different types based on their structure in a time-frequency plot as shown in Fig. (4.1). These glitches are either environmental or intrumental in origin and are known to occur in the data as frequently as once every hour. Therefore, it is important to discriminate between glitches and real signals.

The χ^2 tests for consistency of the waveform can act as effective vetos against these glitches. The effectiveness of a χ^2 veto test can change based on the type of glitch. In every observation run, new types of glitches have been discovered. Therefore, it is important to develop new and effective veto procedures to eliminate the triggers caused by these glitches.

4.1 Sine-gaussian glitches

Short, transient noise bursts can be effectively modelled using a sinusoidal function with a gaussian window [13]. Several types of the glitches shown in Fig. (**??**) can be modelled using such sine-gaussian functions. We can use this model for the glitches and construct a χ^2 test using the unified χ^2 formalism. We will use our control over the construction of this χ^2 to make an optimal statistic.

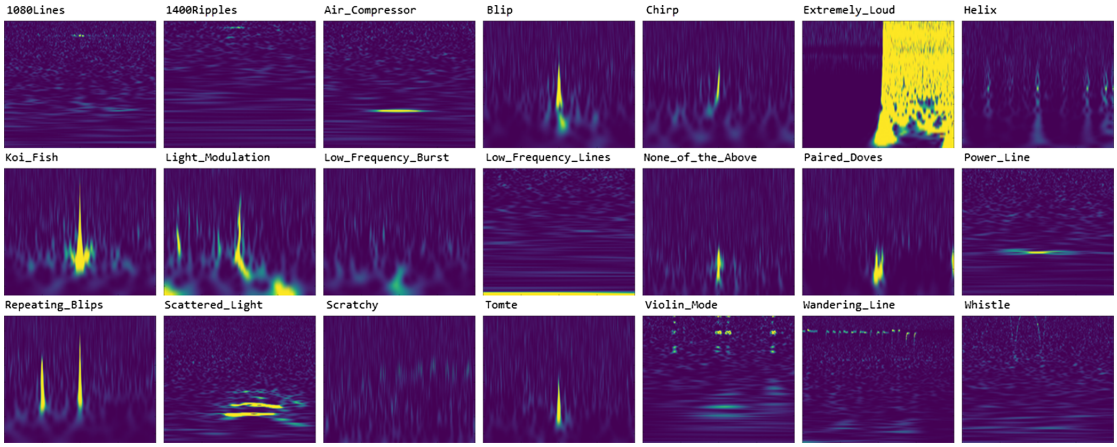


Figure 4.1: Different types of glitches found in the first observation run of LIGO [12]

The sine-gaussian function is parametrized by three quantities, central time (t_0), central frequency (f_0) and the quality factor (Q). t_0 is the maxima of the gaussian window in time domain and label for the the position of the sine-gaussian in time. f_0 is the maxima of the gaussian window in the frequency domain. Q is the number of oscillations of the sinusoidal function in the gaussian window. As seen in the equation below, it is proportional to the width of the gaussian window in the time-domain. The expression for the sine-gaussian function in the time-domain, is as follows:

$$s(t) = s_0 e^{-(t-t_0)^2/\tau^2} sin(2\pi f_0(t-t_0)) \tag{4.1}$$

where τ is the width of the gaussian window function and is related to Q by the expression $Q = 2\pi f_0 \tau$. The Fourier domain expression, that can be obtained by taking a Fourier transform of the above expression, is as follows:

$$\tilde{s}(f) = \frac{1}{2i}\left(\frac{2}{\pi}\right)^{1/4}\sqrt{\frac{Q}{f_0}} e^{-(f-f_0)^2 Q^2/4f_0^2} e^{-2\pi i f t_0} \tag{4.2}$$

In this definition, the sine-gaussian function is normalized such that, its norm under the inner product in Eq. (3.14) with a white power spectral density is 1. This can be expressed as follows.

$$(\mathbf{s},\mathbf{s}) = 4\,\mathrm{Re}\int_{f_{low}}^{f_{up}} \frac{|\tilde{s}(f)|^2}{S_h(f)} df = 1 \tag{4.3}$$

where $S_h(f) = 1$ is the power spectral density.

In order to effectively use this model, we will study the effect of the matched filter test on the

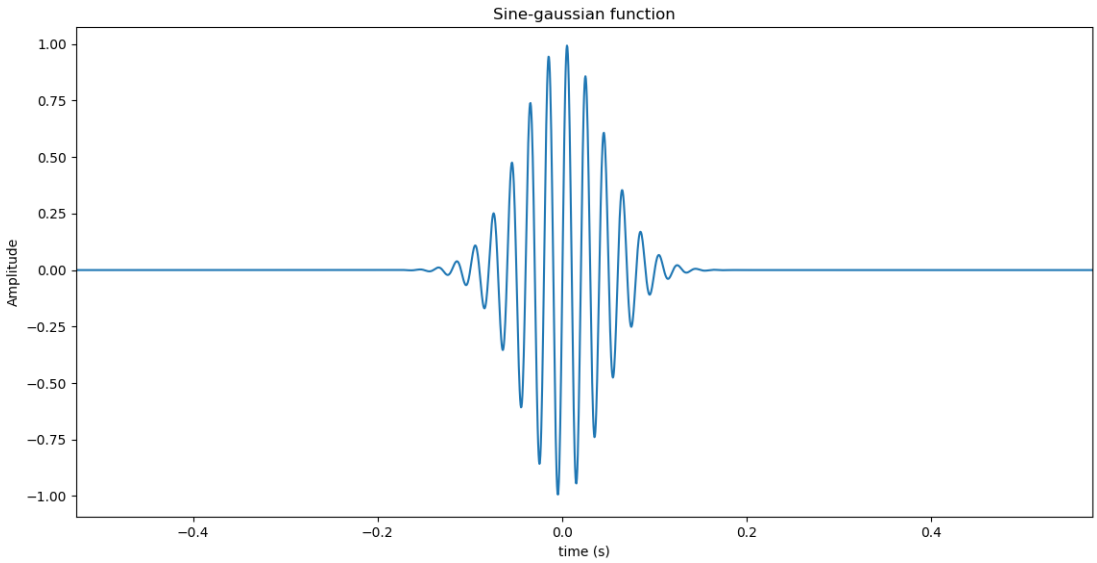


Figure 4.2: Plot of a Sine-gaussian function in time-domain. The parameters of the sine-gaussian are $t_0 = 0$, $f_0 = 50$ Hz and $Q = 20$

sine-gaussian function [14]. This can be studied analytically if the template used for the matched filter test is the Newtonian template as given in Eq. (2.22). The most important feature of the action of matched filter test on a sine-gaussian function is the difference between the coalescenece time t_c (time of trigger) and the central time of the sine-gaussian. This time-delay has also been observed for real glitches in interferometer data. If this time-delay is long, it is difficult to correlate the trigger and the glitch, which increases the probability of false alarm for that trigger. The time-delay can be calculated analytically for the Newtonian template in Eq. (2.22), by finding the output of matched filter (SNR) as a function of time-delay and maximizing it. This gives the follwing analytical expression for the time-delay.

$$t_d = \tau_0 \left(1 - \frac{16}{3Q^2}\left(\zeta + \frac{2}{3}\right)\right) \tag{4.4}$$

where τ_0 is the chirp time of the template as defined in Eq. (2.24) with the fiducial frequency $f_s = f_0$. In the above equation, ζ is the derivative of logarithm of Power Spectral Density of noise evaluated at f_0. If the detector noise has a white PSD, then $\zeta = 0$. Then, the expression for

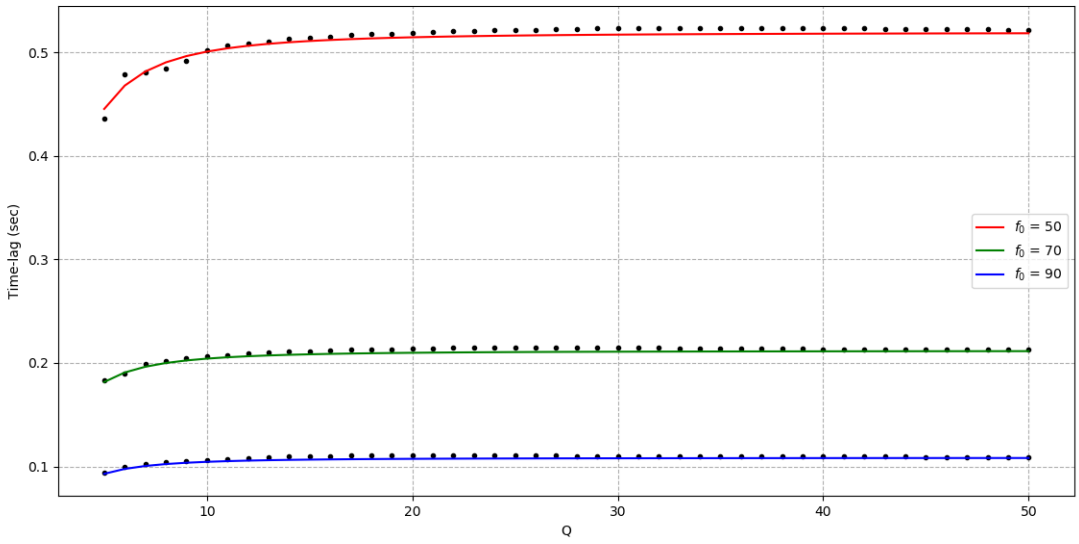


Figure 4.3: This is a plot of time-delay v/s Q at constant f_0 of 50 Hz, 70 Hz and 90 Hz. Plot shows excellent agreement between the computational data and the analytical formula. The computational data was also gathered using the Newtonian template.

time-delay is as follows.

$$t_d = \tau_0 \left(1 - \frac{32}{9Q^2}\right) \tag{4.5}$$

With this background, we are now ready to use the sine-gaussian functions to construct an optimal χ^2 test.

4.2 Parameter space of sine-gaussian functions

As we have seen, sine-gaussian function is described by three parameters, central time (t_0), central frequency (f_0) and quality factor (Q). Therefore, the parameter space of sine-gaussians is a continuous, real three-dimensional space. In order to define an optimal χ^2 test, we need to construct a subspace of the space of all possible data trains ($\mathscr{D}$) which is orthogonal to the template vector (**h**), such that, the glitch vector will have a high projection on this subspace. We can do this by constructing this space using several sine-gaussian vectors with different parameter values. Thus

the problem is essentially to sample enough number of uniformly spaced points on the parameter space which will adequately represent the entire space.

The best way to do this is to define a metric on this parameter space [15] [7]. If $s(\lambda^\alpha)$ and $s(\lambda^\alpha + d\lambda^\alpha)$ are two normalized sine-gaussian vectors with parameter values $\lambda^\alpha = (t_0, f_0, Q)$ and $\lambda^\alpha + d\lambda^\alpha = (t_0 + dt_0, f_0 + df_0, Q + dQ)$, then $\Delta s = s(\lambda^\alpha + d\lambda^\alpha) - s(\lambda^\alpha)$. The metric is defined as follows:

$$\begin{aligned} ds^2 &= (\Delta s, \Delta s) \\ &= (s(\lambda^\alpha + d\lambda^\alpha) - s(\lambda^\alpha), s(\lambda^\alpha + d\lambda^\alpha) - s(\lambda^\alpha)) \\ &= (s(\lambda^\alpha + d\lambda^\alpha), s(\lambda^\alpha + d\lambda^\alpha)) + (s(\lambda^\alpha), s(\lambda^\alpha)) - 2\,(s(\lambda^\alpha + d\lambda^\alpha), s(\lambda^\alpha)) \\ &= 2(1 - (s(\lambda^\alpha + d\lambda^\alpha), s(\lambda^\alpha))) \end{aligned} \tag{4.6}$$

$$ds^2 = 2(1 - p) \tag{4.7}$$

where $p = (s(\lambda^\alpha + d\lambda^\alpha), s(\lambda^\alpha))$ is the component of $s(\lambda^\alpha + d\lambda^\alpha)$ along $s(\lambda^\alpha)$. If $d\lambda^\alpha << \lambda^\alpha$, then we can expand $s(\lambda^\alpha + d\lambda^\alpha)$ in a Taylor series and ignore terms that are higher than linear order in $d\lambda^\alpha$.

$$s(\lambda^\alpha + d\lambda^\alpha) \approx s(\lambda^\alpha) + \left(\frac{ds}{d\lambda^\beta}(\lambda^\alpha)\right) d\lambda^\beta \tag{4.8}$$

In this approximation, we can write the metric in Eq. (4.6) as follows:

$$\begin{aligned} ds^2 &= \left(\left(\frac{ds}{d\lambda^\alpha}\right) d\lambda^\alpha, \left(\frac{ds}{d\lambda^\beta}\right) d\lambda^\beta\right) \\ &= \left(\frac{ds}{d\lambda^\alpha}, \frac{ds}{d\lambda^\beta}\right) d\lambda^\alpha d\lambda^\beta \\ ds^2 &= g_{\alpha\beta} d\lambda^\alpha d\lambda^\beta \end{aligned} \tag{4.9}$$

where $g_{\alpha\beta} = \left(\frac{ds}{d\lambda^\alpha}, \frac{ds}{d\lambda^\beta}\right)$ are the components of the metric. The distance between any two points in this metric is the mismatch between the sine-gaussian functions with the corresponding parameter values. The mismatch is defined as a function of $p = (s(\lambda^\alpha + d\lambda^\alpha), s(\lambda^\alpha))$ in Eq. (4.7).

Using the normalized sine-gaussian function in Eq. (4.2), we can compute all the components of this metric. In order to write these functions in a simple form in which the metric is a diagonal

matrix, we will perform the following coordinate transformations.

$$\omega_0 = 2\pi f_0 \qquad\qquad \nu = 2\pi f_0/Q \tag{4.10}$$

In these coordinates, the metric can be written as follows:

$$ds^2 = (\nu^2 + \omega_0^2)dt_0^2 + \frac{1}{4\nu^2}d\omega_0^2 + \frac{1}{2\nu^2}d\nu^2 \tag{4.11}$$

The region of parameter space under consideration in both the coordinate systems is as follows. These values have been chosen with knowledge of the parameters of real glitches that occur in the detectors.

$$\begin{aligned} f_0(\text{in Hz}) &\in [40, 120] & \qquad Q &\in [5, 50] \\ \omega_0(\text{in Hz}) &\in [80\pi, 240\pi] & \qquad \nu(\text{in Hz}) &\in [8\pi/5, 48\pi] \end{aligned}$$

Now before we try to sample points using this metric, we need to consider the effect of matched filter test on a sine-gaussian function. We have seen that there is a time-difference between the central time of the sine-gaussian function and the time of trigger. For a fixed trigger time, the time-delay of sine-gaussians is a function of Q, f_0 and the component masses of the binary (m_1, m_2). Sine-gaussian functions with the appropriate time-lag defined by Eq. (4.5) are the only sine-gaussians that will contribute to the high SNR at the fixed trigger time. Therefore, we only need to sample points from a two-dimensional surface in the parameter space of sine-gaussian functions which is defined by the following equation:

$$t_d = \tau_0 \left(1 - \frac{32}{9Q^2}\right) \tag{4.12}$$

In the chosen range of parameters, the term $1/Q^2$ is very small compared to 1. Therefore, we can ignore that term in order to make calculation simpler. Thus, the surface is approximately defined by the equation:

$$t_d = \tau_0 = \frac{5}{256\pi f_0} \left(\pi \mathscr{M} f_0\right)^{-5/3} \tag{4.13}$$

In order to sample points on this surface, we will find the metric on this surface. We will restrict the metric given by Eq. (4.11) to the surface defined by Eq. (4.13). This can be done by considering the Taylor expansion of t_d as a function of Q and f_0 and keeping terms upto linear order in dQ and

df_0.

$$t_d(f_0+df_0,Q+dQ) \approx t_d(f_0,Q)+\left(\frac{dt_d}{df_0}\right)df_0+\left(\frac{dt_d}{dQ}\right)dQ$$
$$t_d(f_0+df_0,Q+dQ)-t_d(f_0,Q)=\left(\frac{dt_d}{df_0}\right)df_0+\left(\frac{dt_d}{dQ}\right)dQ$$
$$dt_d=\left(\frac{dt_d}{df_0}\right)df_0+\left(\frac{dt_d}{dQ}\right)dQ \tag{4.14}$$

We will substitute Eq.(4.14) instead of dt_0 is the metric in Eq. (4.11) and we get the metric on the chosen surface.

$$ds^2=\left[C^2\omega_0^{-16/3}+C^2\omega_0^{-22/3}v^2+\frac{1}{4v^2}\right]d\omega_0^2+\frac{dv^2}{2v^2} \tag{4.15}$$

where $C=(5/48)2^{5/3}\mathscr{M}^{-5/3}$. The terms $C^2\omega_0^{-22/3}v^2$ and $1/4v^2$ are very small compared to $C^2\omega_0^{-16/3}$ in our chosen region of the parameter space. Therefore, we can ignore those terms. We can perform another coordinate transformation to simplify the functional form of metric. The definitions of the new suitable coordinates are as follows.

$$z=(\mathscr{M}\omega_0)^{-5/3} \qquad\qquad y=ln(v) \tag{4.16}$$

The metric with the above coordinates takes the following form.

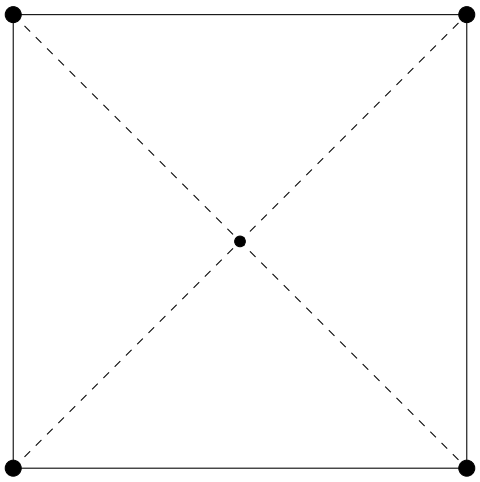
$$ds^2=\frac{9}{25}C_2dz^2+\frac{1}{2}dy^2 \tag{4.17}$$

where $C_2=(5/48)^2\times 2^{10/3}$. The metric, thus takes the form a standard Euclidean metric on a flat space with scaled coordinates. Using this metric, it is possible to place a rectangular grid on points in the chosen region of the parameter space with a constant spacing. We can then use the sine-gaussian vectors corresponding to the grid-points to build the subspace. The spacing between the two grid points is the mismatch (ε) between the two points. This mismatch decides the number of grid-points in the chosen region of parameter space and thus, how well the space represented by the points. But instead of mismatch between two points, the projection of sine-gaussian corresponding to one point on the sine-gaussian corresponding to the other point(p) is a more relevant parameter to describe the spacing between two points. It is related to ε by the following relation.

$$\varepsilon=\sqrt{2(1-p)} \tag{4.18}$$

p provides a better intuition about how well the parameter space is represented by the points. This is a parameter whose value can be tuned in order to construct an optimal χ^2 test.

Consider one cell bound by four grid-points as shown in the figure below. In a standard Eu-



clidean metric, the point in the centre of the square is the farthest from each of the grid-points. Any other point in the square will be closer to one of the grid-points than the rest and this closest distance will be a less than half of the diagonal of the square. Thus we can decide on a minimum projection p which describes the distance equal to half of the diagonal. Any other point in the square will have a projection greater than p on one of the grid-points. Thus, we can claim that parameter space is represented with a projection greater than or equal to p and thus, quantify the representation by a number. If the minimum projection corresponds to a maximum mismatch ε, then the distance between two grid points should be $\Delta l = \sqrt{2}\varepsilon$

After deciding the spacing between grid-points, the next step of the procedure is to calculate the coordinates of the points. Instead of putting a perfectly rectangular grid, we will put a slightly twisted rectangular grid on the space. This is because the boundaries of the region of the parameter space in $y-z$ coordinates are not straight lines. Therefore, it is more convenient to choose points along a line with constant value of Q in the space. Equation of line with constant Q in $y-z$

coordinates can be derived as follows.

$$\begin{aligned} y &= \ln(\nu) \\ &= \ln\left(\frac{\omega_0}{Q}\right) \\ &= \ln\left(\frac{z^{-3/5}}{Q\mathscr{M}}\right) \\ y &= -\frac{3}{5}\ln z - \ln Q - \ln \mathscr{M} \end{aligned} \tag{4.19}$$

The second axis of the grid is the line constant z-coordinate.

We will now calculate the number of points along the two axes - constant Q lines and constant z lines. To find number of points along constant Q line, we will first calculate the distance between the boundaries of the parameter space along the z-coordinate. Since, the boundaries along the z-coordinate are straight lines, the end-points of the parameter space do not depend on the y-coordinate. The distance can thus be computed using the follwing formula.

$$\begin{aligned} S_Q &= \int_{z_{min}}^{z_{max}} \sqrt{\frac{9}{25}C_2}dz \\ S_Q &= \sqrt{\frac{9}{25}C_2}(z_{max} - z_{min}) \end{aligned} \tag{4.20}$$

The number of points along the constant Q line can then be calculated by the following formula.

$$N_Q = \left\lceil \frac{S_Q}{\Delta l} \right\rceil \tag{4.21}$$

Similarly, in order to find the number of points along the constant z line, we need to compute distance between the end-points of the parameter space along a constant z line. Since, the end-points of the parameter space are different for different values of z coordinate, we will pick $z = z_{min}$. Then the distance between the endpoints can be calculated as follows.

$$\begin{aligned} S_y &= \int_{y_{min}}^{y_{max}} \frac{1}{\sqrt{2}}dy \\ S_y &= \frac{y_{max} - y_{min}}{\sqrt{2}} \end{aligned} \tag{4.22}$$

The number of points along the constant z line can be found by the formula below.

$$N_y = \left\lceil \frac{S_y}{\Delta l} \right\rceil \tag{4.23}$$

Thus, the total number of the points sampled in the parameter space is $N = N_y \times N_Q$

Now, in order to calculate the coordinates of the points, we will first find the values of constant Q along which the points are to be placed. Consider the line of constant z coordinate at $z = z_{min}$. The values of y-coordinate of the points on the required constant Q line can be found by using the following formula.

$$y_i = y_{min} + \frac{(0.5+i)}{N_y}(y_{max} - y_{min}) \tag{4.24}$$

where $i = 1, ..., N_y$. The required values of Q corresponding to the values y_i are as follows.

$$Q_i = \frac{z^{-3/5}}{\mathscr{M}} e^{-y_i} \tag{4.25}$$

We get the z-coordinates of the points similarly by dividing the distance between the end-points by N_Q.

$$z_j = z_{min} + \frac{(0.5+j)}{N_Q}(z_{max} - z_{min}) \tag{4.26}$$

where $j = 1, ..., N_Q$. For each of the above values of z-coordinate, we find the corresponding values of y-coordinate for each of the N_y values of Q using the following formula.

$$y_{ij} = -\frac{3}{5}\ln z_j - \ln Q_i - \ln \mathscr{M} \tag{4.27}$$

Thus the coordinates of the sampled points are (z_j, y_{ij}) as described in the above equations and can be calculated in a step-by-step fashion as described above.

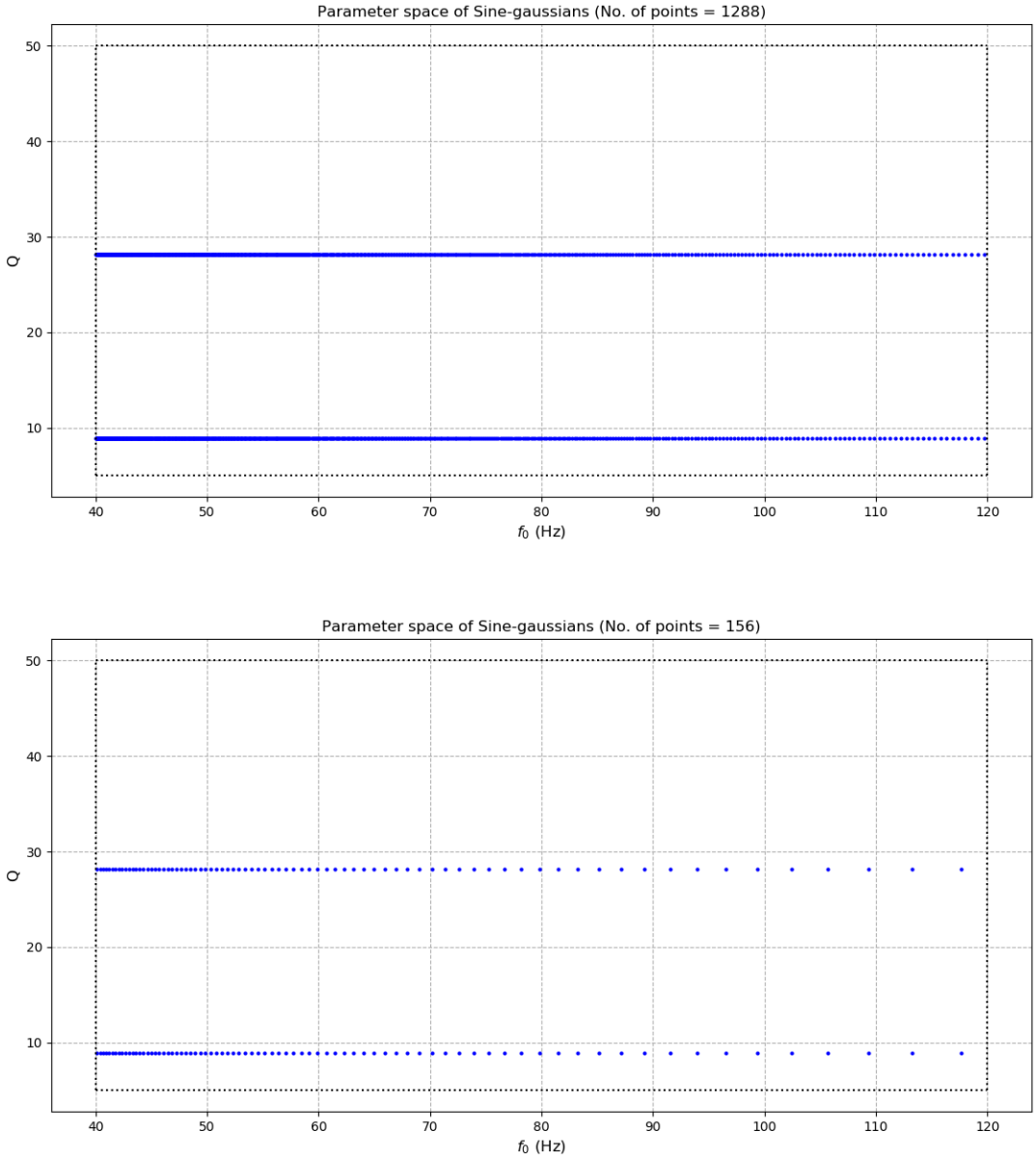


Figure 4.4: The above plots show the points chosen to represent the parameter space with 80% minimum projection. The upper plot corresponds to binary of mass $7M_{\odot} - 7M_{\odot}$ with 1288 points. The lower plot corresponds to binary of mass $25M_{\odot} - 25M_{\odot}$ with 156 points

4.3 Construction of basis on the orthogonal subspace

Each of the points sampled in the parameter space corresponds to a sine-gaussian vector ($\mathbf{s}_\alpha$) in the $\mathscr{D}$. These sine-gaussian vectors have been shifted appropriately back in time with respect to the template vectors $\mathbf{h}_0$ and $\mathbf{h}_{\pi/2}$. h_0 and $h_{\pi/2}$ are the templates with plus and cross polarisations of gravitational-wave respectively. In order to construct the orthogonal subspace using these shifted sine-gaussian vectors, we will subtract the components of these sine-gaussian vectors that are parallel to $\mathbf{h}_0$ and $\mathbf{h}_{\pi/2}$ as shown below.

$$\mathbf{s}_\alpha^\perp = \mathbf{s}_\alpha - (\mathbf{h_0}, \mathbf{s}_\alpha)\,\mathbf{h}_0 - \left(\mathbf{h}_{\pi/\mathbf{2}}, \mathbf{s}_\alpha\right)\mathbf{h}_{\pi/2} \tag{4.28}$$

The space that is spanned by these 'clipped' sine-gaussian vectors is a subspace of $\mathscr{D}$ and is orthogonal to the templates. This is the optimal subspace $\mathscr{S}$ that will be used to define the optimal χ^2 statistic. The dimension of this subspace is equal to the number of sine-gaussian vectors. As we have seen before, this number can be of the order of 100 or more. This will be very expensive in terms of computational cost. Therefore, we will use a procedure called singular value decomposition (SVD) to approximate the subspace using a less number of vectors.

4.3.1 Singular Value Decomposition

Consider a $m \times n$ matrix, A_{mn} such that all the entries are complex numbers. There is a theorem that states it is possible to find matrices U_{mm}, Σ_{mn} and V_{nn} such that the following is true.

- $A_{mn} = U_{mm}\Sigma_{mn}V_{nn}^\dagger$
- U_{mm} is a unitary matrix. The columns of U are the orthonormal eigenvectors of $AA^\dagger$ and are called the left-singular vectors of A.
- V_{nn} is a unitary matrix. The columns of V are the orthonormal eigenvectors of $A^\dagger A$ and are called the right-singular vectors of A.
- Σ_{mn} is a diagonal matrix whose entries are positive real numbers called the singular values of A. They are arranged in a descending order i.e $\sigma_1 > \sigma_2 > ....$ and so on.

Such a decomposition of matrix A is called a Singular Value Decomposition (SVD).

Consider the rows of matrix A to be vectors. These are m vectors in an n-dimensional space. The SVD has been constructed such that if we consider the first k right-singular vectors ($1 \leq k \leq m$), then the Young-Eckart-Mirsky theorem states the space spanned by these k vectors is the best k-dimensional subspace that approximates the m-dimensional space spanned by the rows of A [16]. In fact for $k = 1$, this is the same as the problem of finding the *best least-squares fit* line in which we minimize the perpendicular distance between the given points and the line. This is equivalent to maximizing the projection of the vectors or rows of A onto the line. For $k > 1$, the projection of the vectors on the space spanned by the k right singular vectors is maximized. This yields a best k-dimensional subspace that approximates the space spanned by the rows of matrix A.

4.3.2 Construction of sine-gaussian matrix

As discussed above, we can approximate the orthogonal subspace of a large dimension by a small number of vectors using SVD. We will construct the matrix required for SVD such that its rows are the normalized 'clipped' sine-gaussian vectors. These vectors belong to the Hilert space $\mathscr{D}$ with the inner product in Eq. (3.14). But the algorihtms to compute SVD use the standard Euclidean inner product on a complex vector space. We will convert the sine-gaussian vectors to a form which is suitable for use with a Euclidean inner product. Consider $\mathbf{x}$ and $\mathbf{y}$ are two vectors in $\mathscr{D}$. The inner product of these two vectors is as follows.

$$(\mathbf{x},\mathbf{y}) = 4\,\mathrm{Re}\int_{f_{low}}^{f_{up}} \frac{\tilde{x}^*(f)\tilde{y}(f)}{S_h(f)}df \tag{4.29}$$

Define new vectors $\mathbf{x}'$ and $\mathbf{y}'$ as follows.

$$\tilde{x}'(f) = \sqrt{\frac{2}{S_h(f)}}\tilde{x}(f) \qquad\qquad \tilde{y}'(f) = \sqrt{\frac{2}{S_h(f)}}\tilde{y}(f) \tag{4.30}$$

Then we write the inner product as follows.

$$(\mathbf{x},\mathbf{y}) = 2\,\mathrm{Re}\int_{f_{low}}^{f_{up}} \tilde{x}'^*(f)\tilde{y}'(f)df \tag{4.31}$$

The vectors $\mathbf{x}'$ and $\mathbf{y}'$ are called whitened vectors and the operation of dividing the vectors by the the square root of the PSD is called whitening Using the identity of complex numbers that

$2\,\mathrm{Re}[z] = z + z^*$, we get the following.

$$\begin{aligned}
(\mathbf{x},\mathbf{y}) &= \int_{f_{low}}^{f_{up}} \tilde{x}'^*(f)\tilde{y}'(f)df + \int_{f_{low}}^{f_{up}} \tilde{x}'(f)\tilde{y}'^*(f)df \\
&= \int_{f_{low}}^{f_{up}} \tilde{x}'^*(f)\tilde{y}'(f)df + \int_{f_{low}}^{f_{up}} \tilde{x}'^*(-f)\tilde{y}'(-f)df \\
&= \int_{f_{low}}^{f_{up}} \tilde{x}'^*(f)\tilde{y}'(f)df + \int_{-f_{up}}^{-f_{low}} \tilde{x}'^*(f)\tilde{y}'(f)df \\
&= \int_{-f_{up}}^{f_{up}} \tilde{X}^*(f)\tilde{Y}(f)df \\
(\mathbf{x},\mathbf{y}) &= (\mathbf{X},\mathbf{Y})_e
\end{aligned} \tag{4.32}$$

where $(,)_e$ is the standard Euclidean inner product on a complex vector space. Here, we have use the property of data $\mathbf{x}$ that states that if $x(t)$ is purely real function, then $\tilde{x}(-f) = \tilde{x}^*(f)$. The new vectors $\tilde{X}$ and $\tilde{Y}$ under a complex Euclidean inner product give the same value as the vectors $\mathbf{x}$ and $\mathbf{y}$ under the inner product in Eq. (3.14). These new vectors are defined as follows.

$$\begin{aligned}
\tilde{X}(f) &= \tilde{x}'(f) && f_{low} \leq f \leq f_{up} \\
&= 0 && -f_{low} \leq f \leq f_{low} \\
&= \tilde{x}'^*(f) && -f_{up} \leq f \leq -f_{low}
\end{aligned} \tag{4.33}$$

We perform the above transformation on the clipped sine-gaussian vectors and then we can construct the matrix A using these transformed clipped sine-gaussian vectors as rows. The transformations in Eq. (4.33) and Eq. (4.30) are bijective and invertible. After computing the SVD of A, we take the first k right-singular vectors of A which are orthonormal to each other under the standard Euclidean inner product. If we now apply the inverse transformations to the singular vectors, they will remain orthonormal to each other under the inner product in Eq. (3.14). Moreover, the singular vectors after the inverse transformation will still span the k-dimensional subspace that best approximates the space $\mathscr{S}$. Thus, by using this procedure, we get a k-dimensional space with an orthonormal basis which is orthogonal to the template vectors.

The choice of the number k is important because it decides the degree of approximation to the actual subspace $\mathscr{S}$ that we make. A good way of deciding this number is to set a cutoff on the percentage of total singular values contained in k number of vectors. Since, singular values of matrix A are the projections of the rows of A on the corresponding singular vector, such a cutoff can very accurately describe the degree of approximation. This is a parameter that can be tuned to

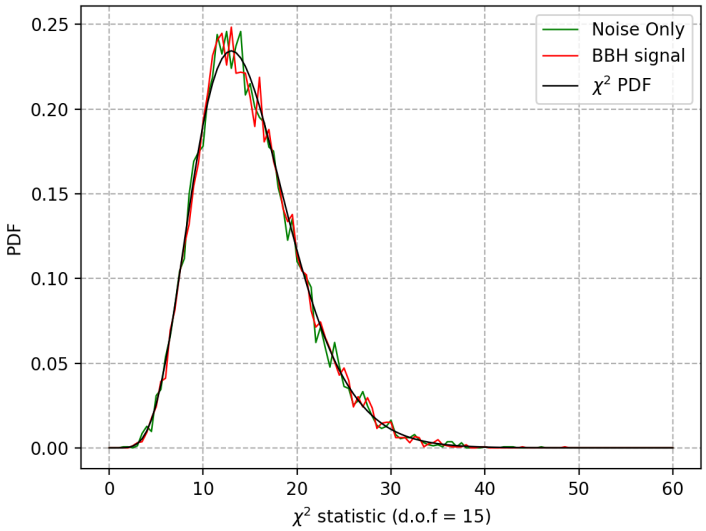


Figure 4.5: The above plot contain the distribution of values of unified χ^2 for 10000 data segments along with a plot of χ^2 distribution. Each of the data segments contained a different realisation of noise. The BBH template with component masses of $25M_{\odot} - 25M_{\odot}$ was used as a signal with perfect match

construct an optimal χ^2 test.

Let the first k right-singular vectors of A, after applying the inverse transformations, be denoted by $\mathbf{e}_{\alpha}(\alpha = 1, 2, ..., k)$. Then, the χ^2 statistic is defined as follows.

$$\chi^2 = \sum_{\alpha=1}^{k} |(\mathbf{x}, \mathbf{e}_{\alpha})|^2 \tag{4.34}$$

If data contains only gaussian noise or signal and gaussian noise, then each $(\mathbf{x}, \mathbf{e}_{\alpha})$ is an independent Gaussian random variable. Therefore, this statistic follows a χ^2 distribution of k degrees of freedom. But if the data contains any glitches that can be modelled by using sine-gaussian functions, its projections on the basis vectors will be large. Therefore, it will give a high value of χ^2. For an optimal χ^2 the difference between the χ^2 values for signals and glitches will be the maximum, even at lower SNRs.

Chapter 5

Results and Discussion

5.1 Optimal χ^2

In the previous section, we have constructed the unfied χ^2 statistic for sine-gaussian glitches. In order to make it an optimal statistic, we need to fix the values of parameters of the unified χ^2, minimum projection (p) and the percentage of singular values such that the projection of glitches on the χ^2 is maximum. The parameter values need to be high enough that the parameter space of sine-gaussians is adequately represented. But at the same time, the computational cost should not be very high. This choice is also affected by the different approximations that were made while constructing this test. I will discuss the most important approximations that have the potential of affecting the performance of the χ^2 statistic in next few paragraphs.

We made a few analytical approximations in order to get the simple metric on the parameter space of sine-gaussians. The most important among them is the use of Newtonian template for defining the analytic expression for the time-lag and then the metric. There exist much better approximations to the waveform of binary black-hole merger, like IMR waveforms, which are used for searches of GW by LIGO. The IMR waveforms are modelled by using three phases of the waveform - inspiral, merger and ringdown and hence are very high degree of accuracy. In comparison, the Newtonian waveform is not a good approximation to the waveform of a binary black-hole merger as seen in Fig (5.1). In particular, the difference between the two waveforms is more apparent at frequencies close to the merger. There is a difference in the time-lag calculated using Newtonian waveform and using the IMR waveform as seen in Fig (5.2). This discrepency

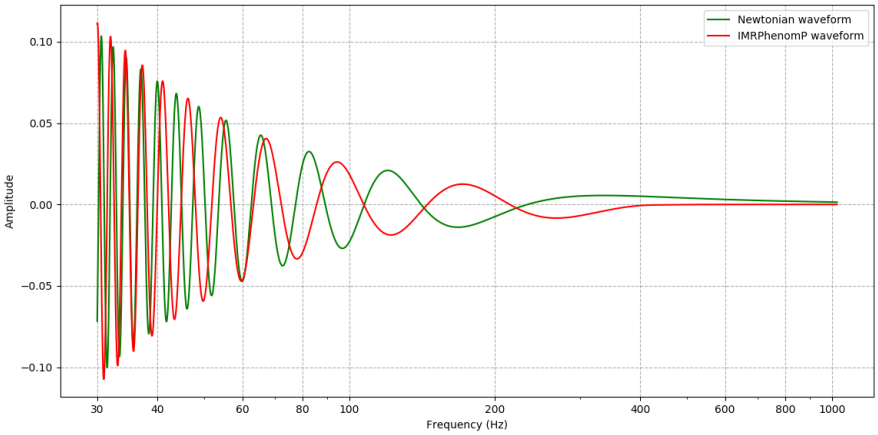


Figure 5.1: A plot showing comparison of the Newtonian waveform with IMRPhenomP waveform for a binary with component masses of $20M_{\odot}$ each

has introduced errors which reduce performance of the χ^2 statistic, especially against glitches with a value of f_0 close to the frequency of merger of the binary.

The metric has been constructed with a fixed normalization of the frequency-domain expression of the sine-gaussian function. This normalization has been done by assuming a white PSD. But the noise in LIGO detectors is coloured with as shown in Fig (5.3). We will use the same algorithm as the case of white PSD to sample points in the parameter space for the above PSD which is called the 'advanced LIGO Zero Detuned High Power' PSD. This approximation can also reduce the performance of the χ^2 statistic.

Keepin the above conditions in mind, the only way to decide the parameter values if to compute the value of the statistic for different values of the parameters, compare the results and choose the values of the parameters that give the best performance. Upon doing that, we found that the percentage of singular vectors that gives the best performance is 90%. This is a precentage that is high enough that the chosen orthogonal vectors after performing SVD are a good approximation to the original space. Any percentage higher than this adds more number of vectors which have a small projection of the glitch along their direction unneccessarily increasing the computational cost.

The choice of minimum projection (p) is more complicated. The performance of the χ^2 reduces in case of glitches with a high value of f_0 with templates of high mass. In case of high mass

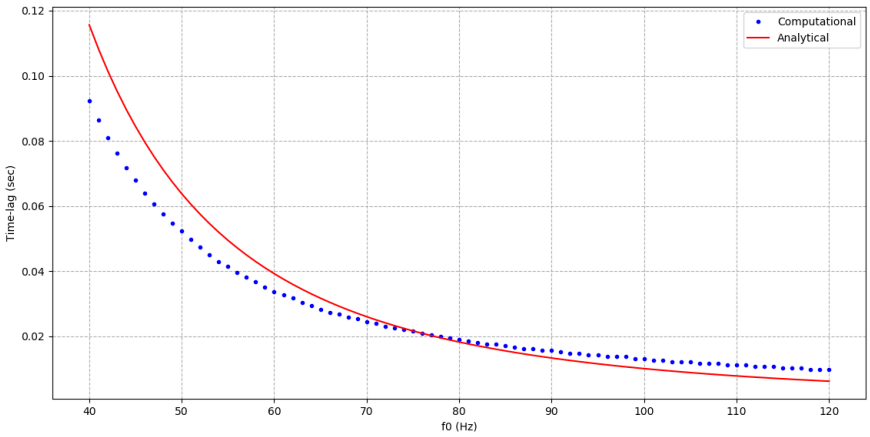


Figure 5.2: This plot shows the deviation of the analytically calculated time-delay for a Newtonian template from the computationally calculated value using the IMRPhenomP waveform. The values are for a constant value of $Q = 20$

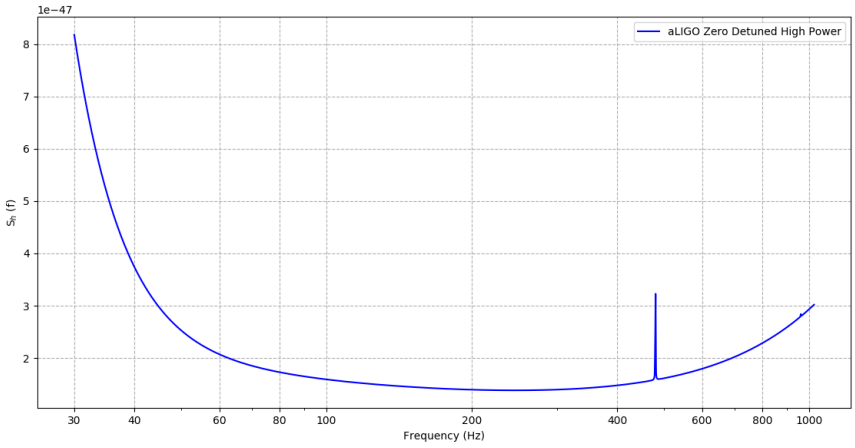


Figure 5.3: This plot shows the aLIGO Zero Detuned High Power PSD [17] [3]

Range of the Total mass of binary	**Minimum projection** (p)
$10M_{\odot} - 70M_{\odot}$	0.80
$70M_{\odot} - 90M_{\odot}$	0.85
$90M_{\odot} - 100M_{\odot}$	0.90
$100M_{\odot} - 120M_{\odot}$	0.95
$120M_{\odot} - 130M_{\odot}$	0.975
$130M_{\odot} - 160M_{\odot}$	0.99

Table 5.1: This table contains the values of minimum projection (p) that will be used to construct the orthogonal subspace, if a template with the total mass in the corresponding range has been triggered in the matched filter test.

templates, the frequency of merger is close to the high frequency end of the parameter space. The mismatch between the Newtonian and the IMR waveform is high at frequencies close to the merger. Thus, in the high f_0 region of the parameter space performance goes down. In order to counter that, consider a higher value of p for templates with high masses and thus sample points closer to each other. This does not increase the computational cost by an uncomfortable margin. The value of p for different ranges of the total mass of the binary is as shown in the Table (5.1). We have thus set the values of the parameters to get an optimal χ^2. We can now run performance tests with Gaussian noise and also compare our optimal χ^2 with the power χ^2.

5.2 χ^2 v/s SNR plots

In order to compare the performance of the two χ^2 tests, we need to test their abilty to distinguish between signals and glitches in simulated data. A simulated data segment can be generated by creating a segment containing pure noise of the appropriate PSD and adding an appropriately scaled known signal or a glitch. In our analysis, we generated simulated data segment of 32 seconds and a sampling frequency of 2048 Hz with three kinds of injections.

1. Signal injection - BBH template as a signal and Gaussian noise having the appropriate PSD.

2. Glitch injection - Sine-gaussian function as a glitch and Gaussian noise with appropriate PSD.

3. Pure noise injection - Gaussian noise with appropriate PSD.

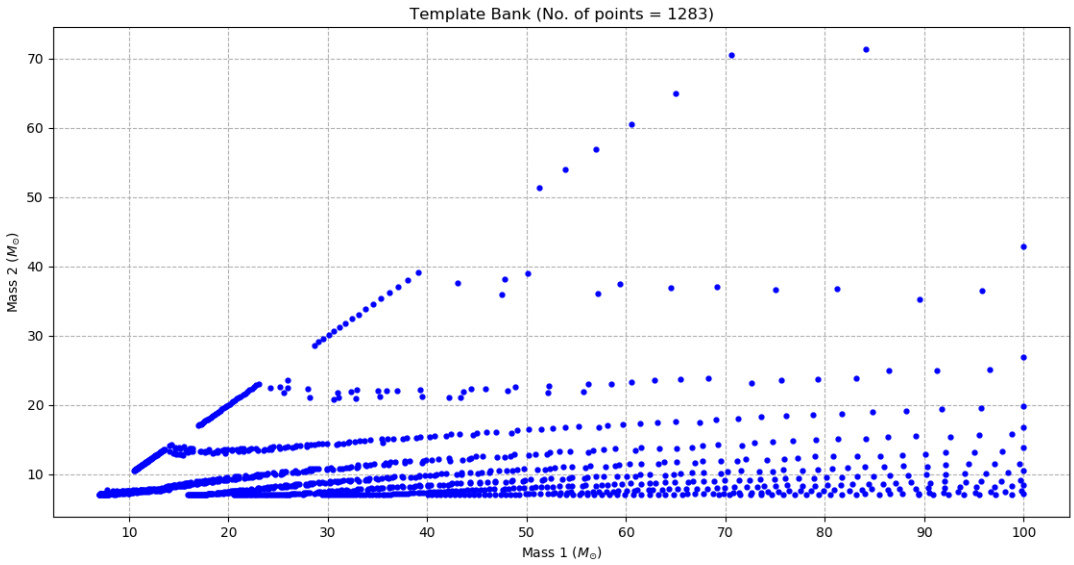


Figure 5.4: Template bank generated for data containing noise with white PSD.

On each of these data segments, we ran a matched filter search for GW signal using a template bank. The template bank that we used in the search (Fig. (5.4) and (5.5))had the following properties.

- Non-spin templates - The waveform of the templates assume that the component black-holes in the binary are not spinning.
- Approximant of the template - IMRPhenomP
- $f_{low} = 30$ Hz.
- PN order = 2PN.
- Maximum mismatch = 1%.
- Minimum component mass = $7M_\odot$.
- Maximum component mass = $100M_\odot$.
- Maximum total mass = $160M_\odot$.

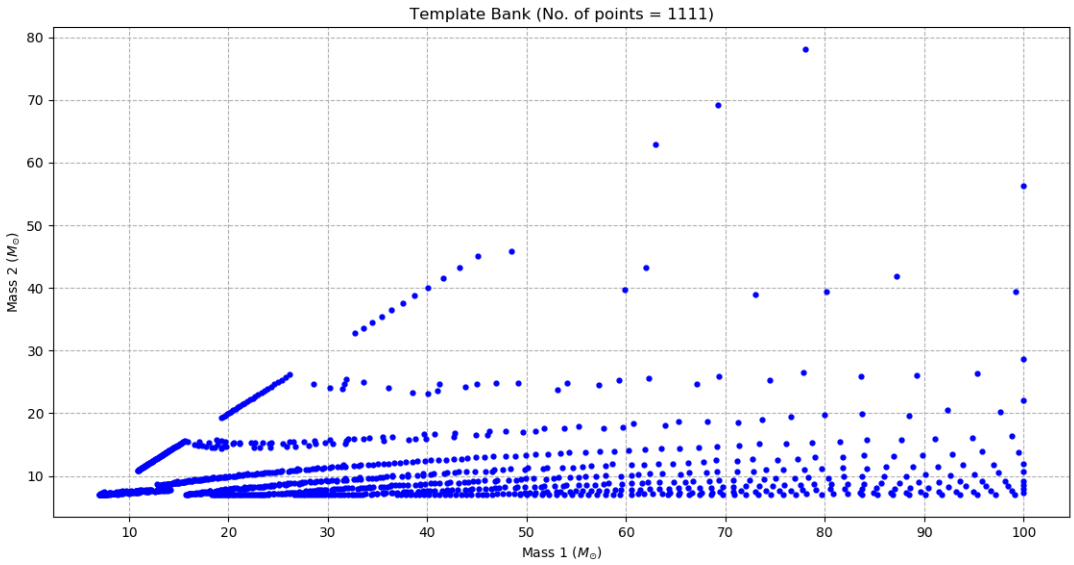


Figure 5.5: Template bank generated for data containing noise with aLIGO Zero Detuned High Power PSD.

If the SNR at any data point was higher than the threhsold (ρ_0) equal to 5, we treated it as a trigger. We computed the value of the power χ^2 and the unified χ^2 at that data point. But, the number of degrees of freedom of the two χ^2s must be equal for a valid comparison. By the nature of the construction of unified χ^2, we cannot fix the number of degrees of freedom beforehand. Therefore, we calculated the unified χ^2 first and then the power χ^2 with the same d.o.f. But, in order to see trends, we need to compare the the two values among a large number of data segments. There is no way to make the d.o.f equal for all the data segments which we use for simulations. Therefore, we will compare the values of χ^2 per d.o.f. or reduced χ^2 (χ^2_r) among all the simulated data segments. This is done so that the mean of the χ^2 distribution is scaled down to 1. Thus, we expect a trend of the χ^2_r for all the signal injections to be close to 1. With all the simulated data, we made plots of χ^2_r v/s SNR. A separation between the values of χ^2_r for glitches and χ^2_r for signals is an indicator of how well the χ^2 test acts as a discriminator. The test is more effective if this separation is higher. This is a very basic way to compare the two statistics. But it can still provide important insights on the overall trends followed by the statistics.

The performance of the unified χ^2 is different in different regions of the parameter space. We want to compare the performance of the two χ^2 tests in broadly different regions. In order to do that, we ran simulations with sets of glitch injections such that in each set, the glitches are sampled

from only one of these regions. For each set, we will plot the χ^2_r v/s SNR using the same signal injections and pure noise injections. We have considered three such broad regions of that parameter space. For each of the regions, we did simulations with noise having a White PSD and also with noise having aLIGO Zero Detuned High Power PSD. Each plot is contains data from 1000 glitch injections, 1200 signal injections and 600 pure noise injections. The ranges of parameters defining the different regions are as follows.

- High Q and low f_0: $25 \leq Q \leq 50$, $40\text{Hz} \leq f_0 \leq 80\text{Hz}$
- High Q and high f_0: $25 \leq Q \leq 50$, $80\text{Hz} \leq f_0 \leq 120\text{Hz}$
- Low Q and low f_0: $5 \leq Q \leq 15$, $40\text{Hz} \leq f_0 \leq 70\text{Hz}$

The χ^2_r v/s SNR plots are in Fiq. (5.6), (5.7) and (5.8).

In these plots, we can see a few clear trends follows by the values of the χ^2_r. The values of both the χ^2_r for signal injections are close to 1 at all SNRs as is required. The values of both the χ^2_r for glitch injections follow an approximate power law. The separation between the χ^2_r values of signals and glitches is quite clear for SNRs greater than 9. Thus, both the χ^2 tests can effectively discriminate between glitches and signals for those SNRs. But for SNRs lower than 9 the χ^2_r values of signals and glitches start overlapping. There is also the possibility that pure gaussian noise is also mistaken for a signal of a low SNR. Therefore, this plot is not a very effective way to test the χ^2_r statistics for low SNRs.

Comparison between the two statistics can be made based on the how high the value of χ^2_r is for a glitch injection. In the region of high Q and low f_0, we can clearly see that the trend of values of unified χ^2_r is higher than the trend of the values of the power χ^2_r. Thus, in that region, the unified χ^2 performs better than the power χ^2. In the other two regions, the values of the two χ^2_rs are overlapping and there is no clear distinction between the trends. Therefore, we will plot ROC curves for the statistics to compare performance.

5.3 Receiver Operator Characteristic (ROC) Curves

The plots of χ^2_r v/s SNR show us a reasonable comparison between the two statistics. But this comparison is only useful if the difference between the two is high. In most of the cases seen

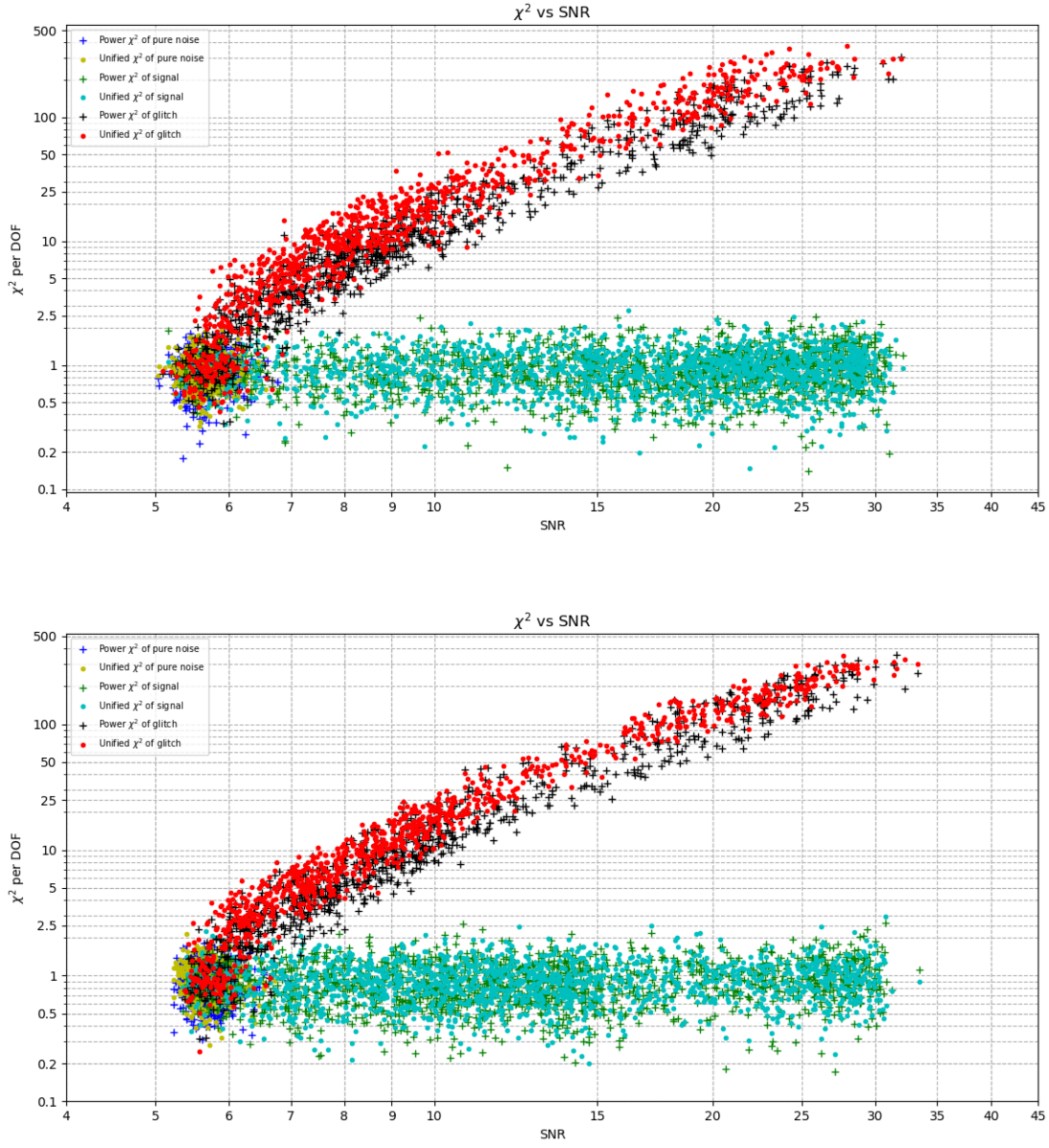


Figure 5.6: χ_r^2 v/s SNR plots for the region 'high Q and low f_0'. The upper and lower plots contain data from injections with a white and aLIGO Zero Detuned High Power PSD respectively

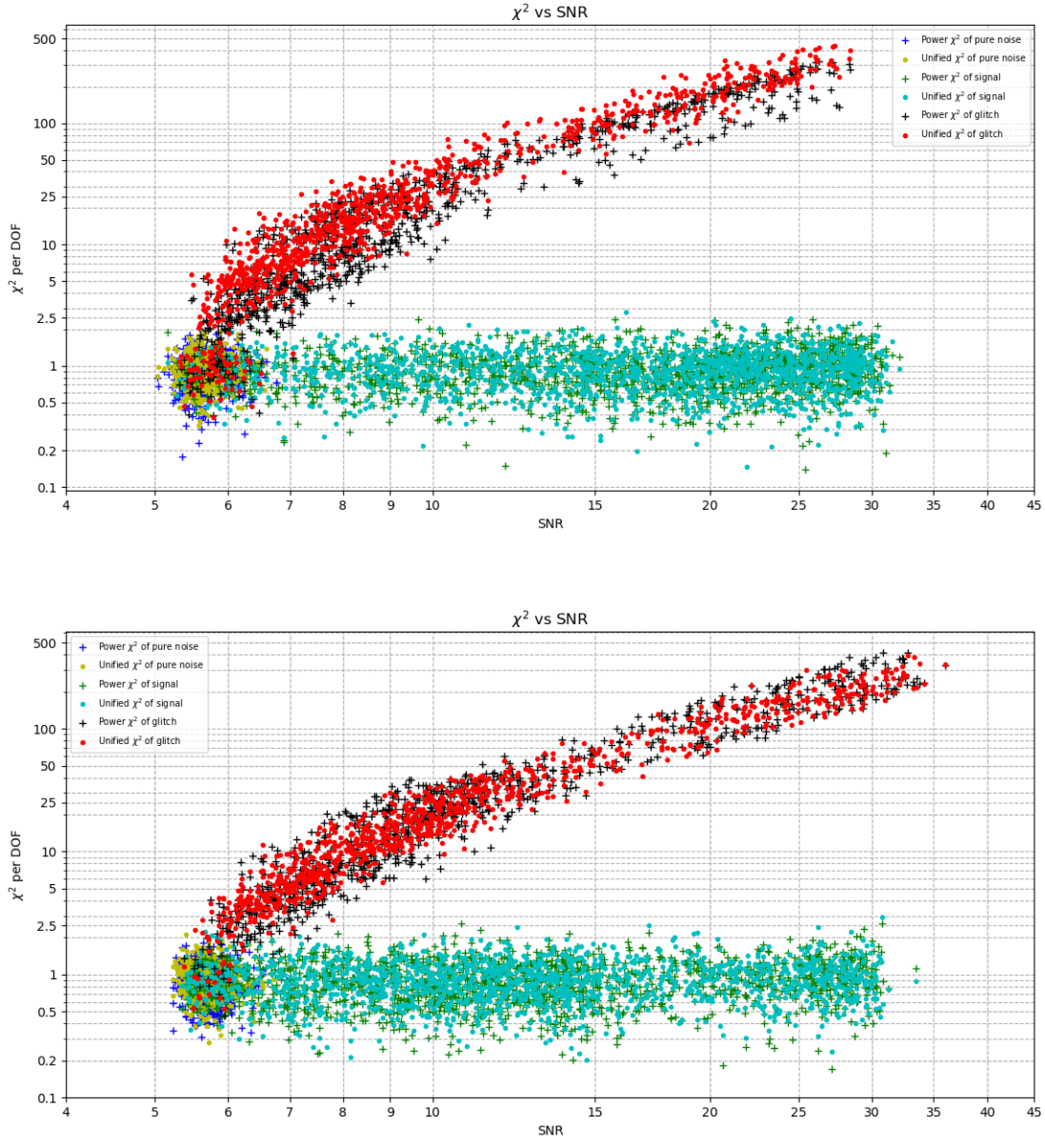


Figure 5.7: χ_r^2 v/s SNR plots for the region 'high Q and high f_0'. The upper and lower plots contain data from injections with a white and aLIGO Zero Detuned High Power PSD respectively

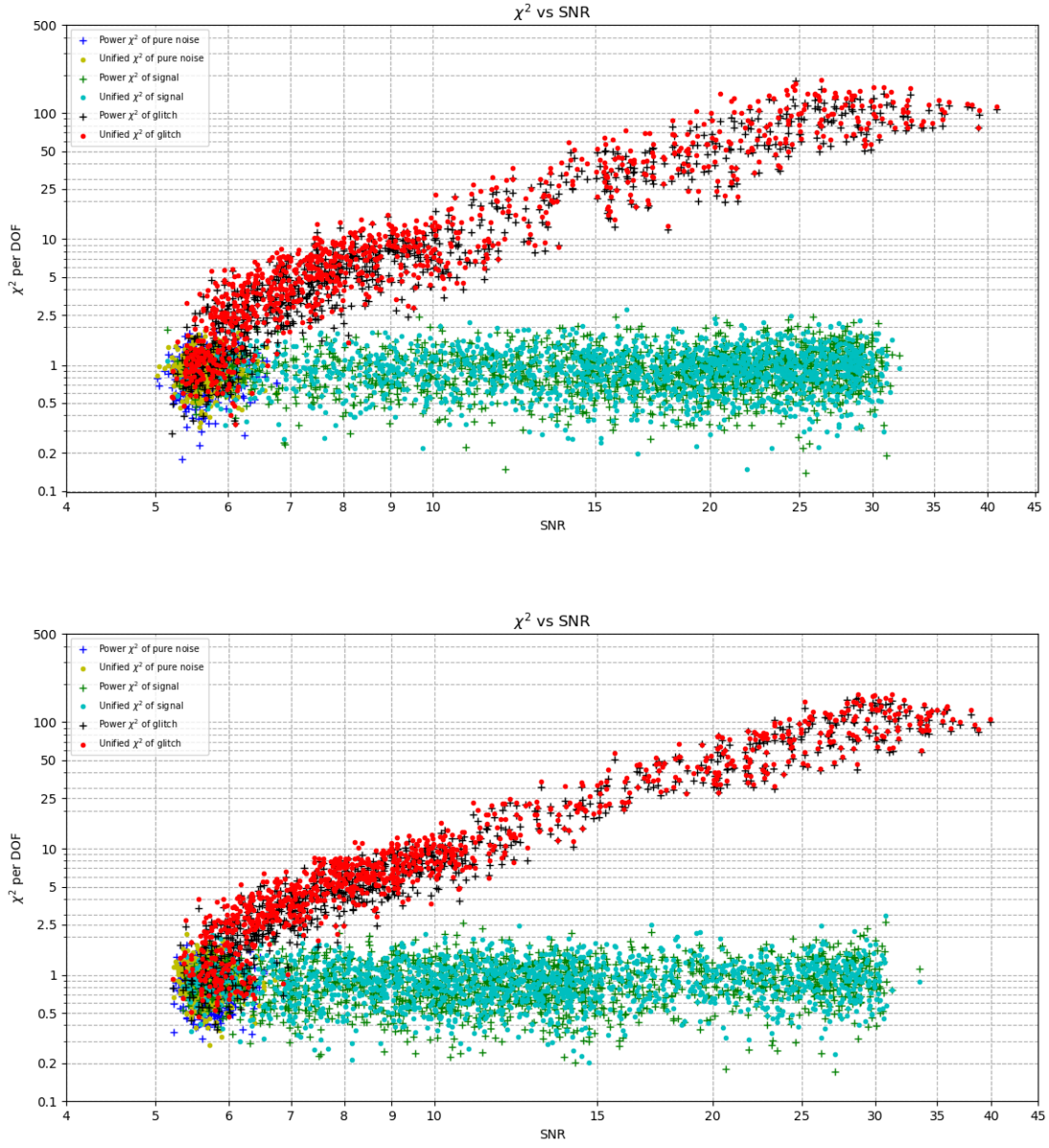


Figure 5.8: χ_r^2 v/s SNR plots for the region 'low Q and low f_0'. The upper and lower plots contain data from injections with a white and aLIGO Zero Detuned High Power PSD respectively

above, the points corresponding to the values of the two χ^2s are overlapping and therefore, it is not possible to say which statistic is better. Moreover, we do not get any quantitative information about their relative performance. We want to know the effectiveness of the statistic in terms of the detection probability of the trigger. A Receiver Operator Characteristic (ROC) curve contains precisely that information, and therefore is a more effective way to compare the two χ^2 statistics.

5.3.1 Definition of ROC Curves

Consider a data segment **x** from the detector. The problem at hand is to check whether this segment contains a signal or not. In order to do that, define a statistic z which is a function of **x** such that z can take values from z_{min} to z_{max}. Let z_0 ($z_{min} \leq z_0 \leq z_{max}$) be a threshold such that,

- $z(\mathbf{x}) \leq z_0 \implies$ A signal is present
- $z(\mathbf{x}) > z_0 \implies$ signal is absent.

Thus, z can classify data into two categories, signal present and signal absent. The effectiveness of the statistic can be checked by generating a large number of segments of simulated data, both containing and not containing signals. The statistic can be used to classify the segments into the two categories for a fixed value of threshold z_0 . In this process, there are four types of data segments as follows.

- **True positive** (TP): The simulated data contains a signal and it is classified by z as a signal.
- **False positive** (FP): The simulated data *does not* contain a signal but it is classified by z as a signal.
- **False negative** (FN): The simulated data contains a signal but it is classified by z as *not* a signal.
- **True negative** (TN): The simulated data *does not* contain a signal and it is classified by z as *not* a signal.

Let the number of data segments which are True positives be N_{TP}, which are False positives be N_{FP}, and so on. If N_p is the total number of data segments in which a signal is present and N_a is

the total number of data segments in which a signal is absent, then Detection Probability ($\mathscr{P}_D$) and False Alarm Probability ($\mathscr{P}_{FA}$) are defined as follows.

$$\mathscr{P}_D = \frac{N_{TP}}{N_p} \qquad\qquad \mathscr{P}_{FA} = \frac{N_{FP}}{N_a} \tag{5.1}$$

The Detection probability and False Alarm probability for different values of threshold z_0 can be calculated. Then, the Receiver Operator Characteristic is a plot of the Detection probability v/s the False Alarm probability. Such a plot is very useful for comparing effectiveness of two statistics. The two probabilities mentioned above are a way of quantifying the performance of the statistic z. The aim is to make a statistic that can classify maximum number of N_p as signals and the minimum number of N_a as signals. Thus, the statistic which has a high Detection probability for a small False Alarm probability qualifies as a good discriminator. In order to compare two statistics, we can compare the Detection probability of the two at the same value of False Alarm probability. Thus, the statistic with a ROC curve higher than the other is better.

5.3.2 ROC curves for unified χ^2

The performance of the discriminator depends heavily on the choice of the statistic. In the case of our problem, we have two statistics at our disposal, the SNR and the χ^2. As we have seen, SNR by itself is not a good statistic to discriminate between signals and glitches. That is why we have the χ^2 to supplement the SNR. We can use an initial SNR cutoff and later a χ^2 cutoff to discriminate between signals and glitches. But, we can achieve much better results by combining the SNR and the χ^2 and defining a new statistic. With the knowledge of the behaviour of the χ^2_r of glitches for increasing value of SNR from the χ^2_r v/s SNR plots, we use the following statistic.

Let $x = log_{10}(\rho)$ (ρ = SNR) and $y = log_{10}(\chi^2_r)$. Then the new statistic is defined as follows:

$$\zeta = x - \frac{(y+0.9)^2}{14} \tag{5.2}$$

If ζ_0 is a threshold on ζ, then $\zeta > \zeta_0 \implies$ signal is present and $\zeta < \zeta_0 \implies$ signal is absent. Curves of constant ζ on a χ^2_r v/s SNR plot look like parabolic functions in log-log scale. These curves are chosen such that they follow the linear trend that can been seen in χ^2_r of glitches at high SNRs.

For the 'low Q and low f_0' region of the parameter space, the linear trend on the χ^2 increasing with SNR has a different slope than the other two regions. Therefore, the statistic ζ as defined in Eq. (5.2) cannot work as a good statistic for discriminating between signals and glitches sampled from this region. But, we can make a different statistic using the same functional form, but changing the constants which aligns well with the linear trend of χ^2 values. This new statistic that will act as a good discriminator for glitches sampled from 'low Q and low f_0' region of the parameter space defined as follows. Let $x = log_{10}(\rho)$ (ρ = SNR) and $y = log_{10}(\chi^2)$.

$$\zeta_1 = x - \frac{(y+0.7)^2}{9} \tag{5.3}$$

If ζ_0 is a threshold on ζ_1, then $\zeta_1 > \zeta_0 \implies$ signal is present and $\zeta_1 < \zeta_0 \implies$ signal is absent. The ROC curves in different regions of the parameter space for the two different PSDs are in the Fig. (5.9), (5.10) and (5.11). The simulations used to plot these are the same ones that were used for the χ^2 v/s SNR plots.

In the 'high Q and low f_0' region, we see that the unified χ^2 is performing much better than the power χ^2 for both the types of noise. This observation is much clearer in the ROC curve that the χ^2_r v/s SNR plot. In the 'low Q and low f_0' region, the unified χ^2 is also performing marginally better in both the types of noise. The performance in the high Q and high f_0 region of the space is same for both the statistics. Altogether, the unified χ^2 is performing better than the power χ^2. This improvement is especially significant for triggers with low SNR.

We have seen in the χ^2_r v/s SNR plots that χ^2 for signals and glitches overlap in the region of SNR < 9. Due to this overlap, the False Alarm probability for triggers in this region is high. This is the region that we are probing at False Alarm probabilities of order 10^{-3}. The False alarm probabilities of triggers with SNRs higher than 9 are much smaller than 10^{-3}, since there is no overlap between χ^2 of signals and glitches. From the ROC curves we can clearly conclude that the unified χ^2 performs better than power χ^2 even at low SNRs. Since the unified χ^2 statistic has a higher detection probability, it will increase our confidence in detections at lower SNRs.

The region of 'low Q and low f_0' gives optimal results with a statistic different from the other regions of the parameter space. This is because the trend of the χ^2 in these regions is very different. We had tuned ζ to work as optimal discriminator for a certain region based on the trends that we observed in that region. Therefore, it is no surprise that it does not work optimally in other regions. We can construct a different statistic that works for all the regions, but that will give suboptimal

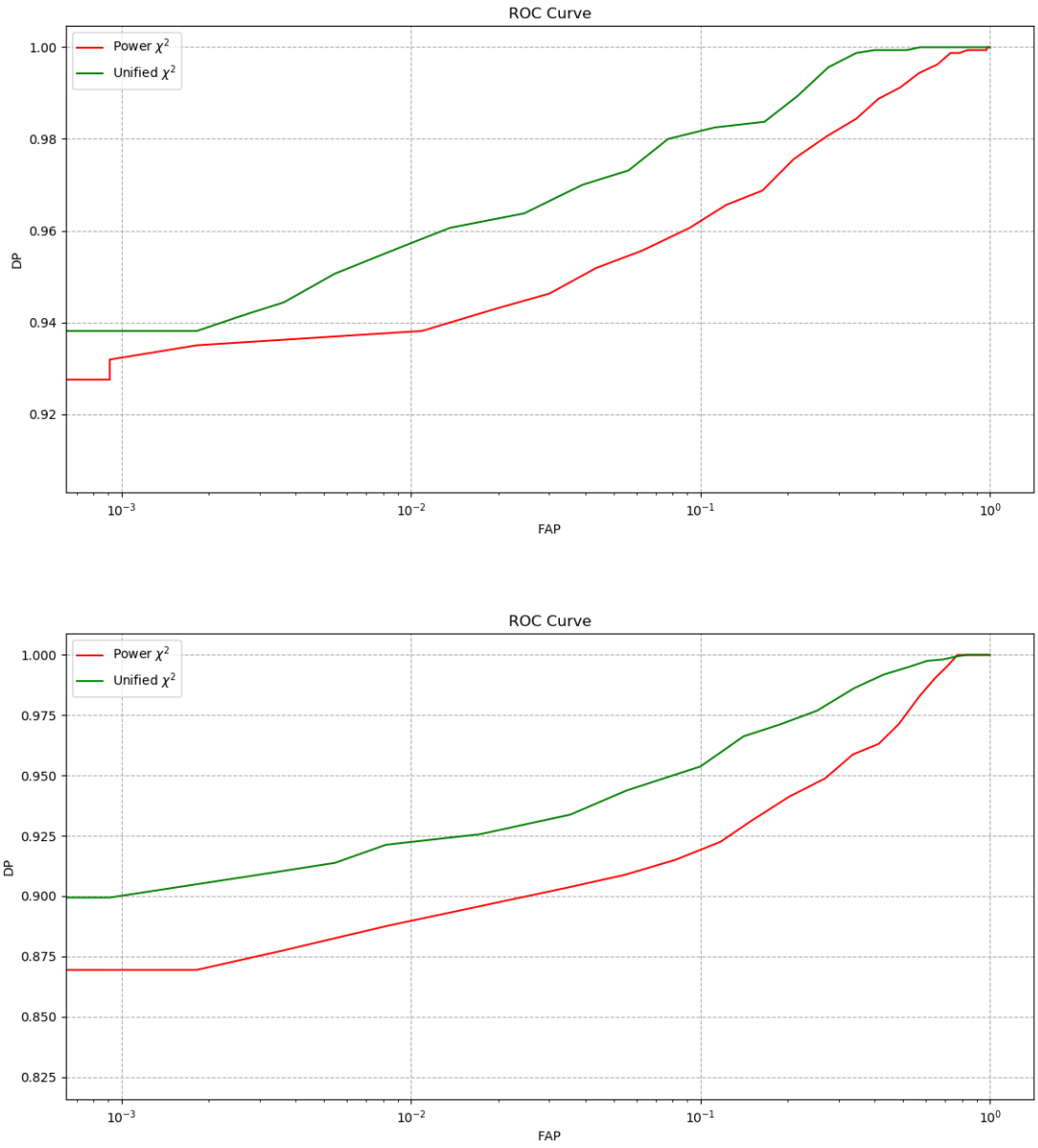


Figure 5.9: ROC curves for the region 'high Q and low f_0'. The upper and lower plots contain data from injections with a white and aLIGO Zero Detuned High Power PSD respectively

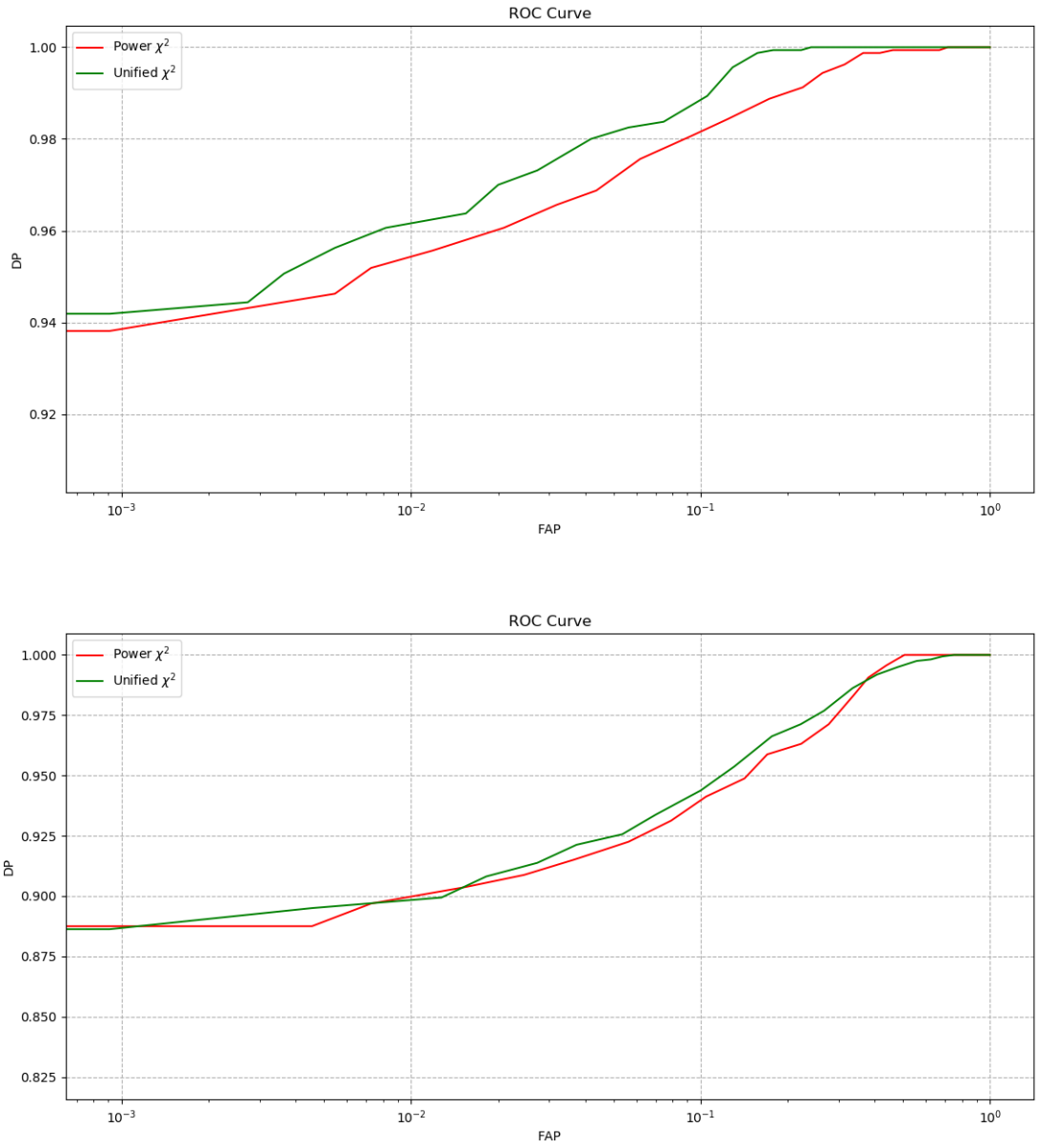


Figure 5.10: ROC curves for the region 'high Q and high f_0'. The upper and lower plots contain data from injections with a white and aLIGO Zero Detuned High Power PSD respectively

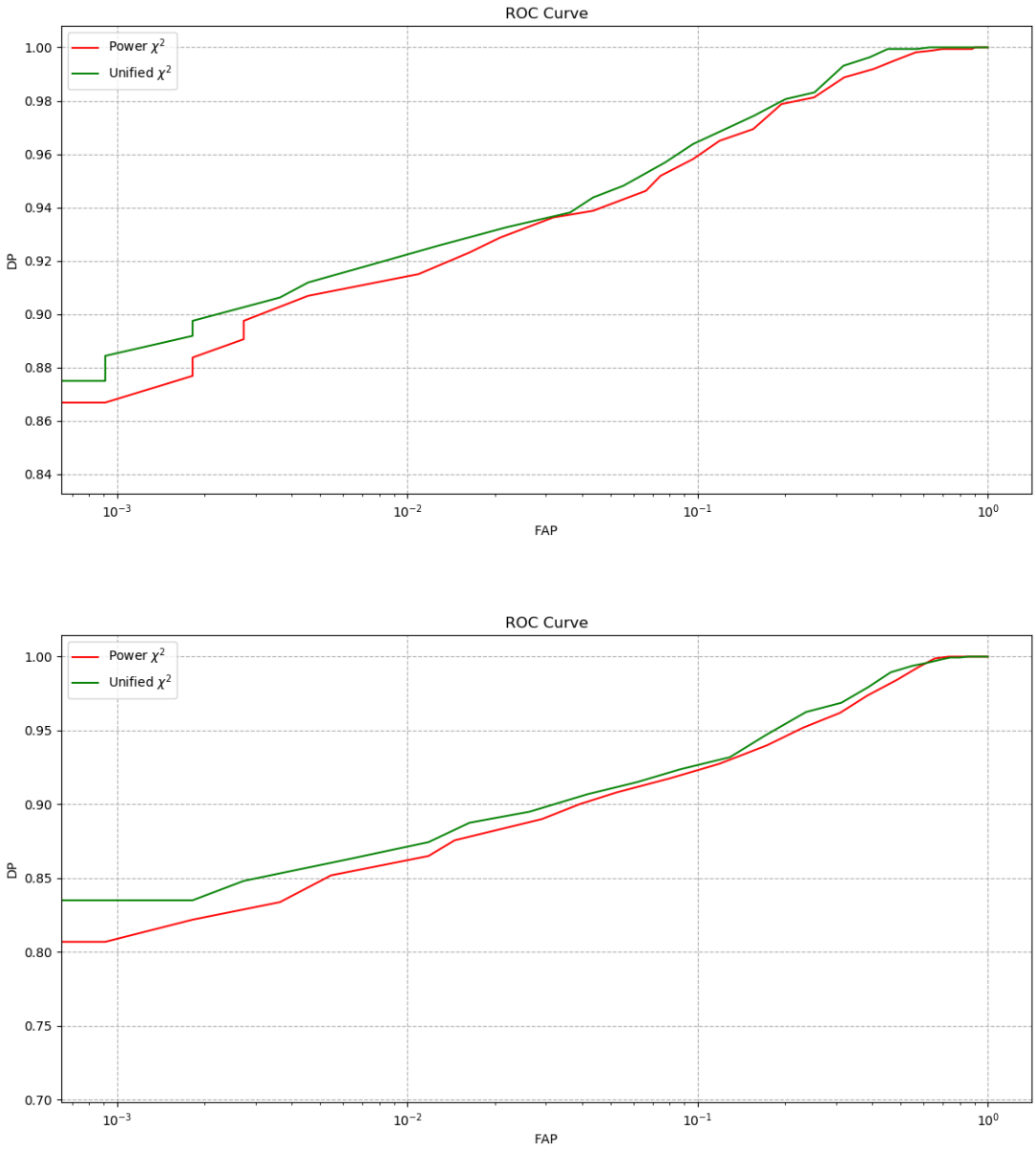


Figure 5.11: ROC curves for the region 'low Q and low f_0'. The upper and lower plots contain data from injections with a white and aLIGO Zero Detuned High Power PSD respectively

results in all the regions of the parameter space. Instead, it is a better idea to use both the statistics together. At least one of them will work as the optimal statistic for discrimination and based on which one does, we will also be able to gain information about the parameters of the glitch.

5.4 Conclusion

With the increasing sensitivity of the detectors, the number of glitches is also increasing. The transition from the first observation run to the third observation run has seen a drastic increase in the number and types of glitches. With more advanced detectors being built, it is important to develop better χ^2 tests to discriminate between signals and glitches. The unified χ^2 formalism provides a mathematical framework to discuss such χ^2 tests. Such a framework can provide insights into the performance of the χ^2 tests for different parameters of glitches an signals in a consistent mathematical way.

Using this framework, we have developed an optimal χ^2 test to discriminate between signals and transient glitches that can be modelled using sine-gaussian functions. In simulations with Gaussian noise, this optimal χ^2 is a more effective discriminator as compared to the power χ^2. Therefore, it is highly likely that it will also perform well in the real LIGO noise, which is not Gaussian. It works better in a large region of the parameter space of sine-gaussian glitches and the BBH templates. The improvement in the performance of the χ^2 at low SNRs is important for new detectors with improved sensitivities. The unified χ^2 can discriminate between signals and glitches more effectively at lower SNRs, further increasing our confidence in the low SNR detections.

The formalism of unified χ^2 can be applied to other types of glitches and optimal χ^2 tests can be constructed for those types as well. Those χ^2 can be combined with the above χ^2 by simply adding the two values. Such a combination can act as an effective χ^2 discriminator against all the different kinds of glitches.

Bibliography

[1] B. P. Abbott and et. al., "Observation of gravitational waves from a binary black hole merger," *Phys. Rev. Lett.*, vol. 116, p. 061102, Feb 2016.

[2] B. P. Abbott *et al.*, "GW150914: First results from the search for binary black hole coalescence with Advanced LIGO," *Phys. Rev.*, vol. D93, no. 12, p. 122003, 2016.

[3] G. M. H. and, "Advanced LIGO: the next generation of gravitational wave detectors," *Classical and Quantum Gravity*, vol. 27, p. 084006, apr 2010.

[4] C. W. Helstrom, *Statistical Theory of Signal Detection*. Pergamon, 1968.

[5] B. S. Sathyaprakash and S. V. Dhurandhar, "Choice of filters for the detection of gravitational waves from coalescing binaries," *Phys. Rev.*, vol. D44, pp. 3819–3834, 1991.

[6] S. V. Dhurandhar and B. S. Sathyaprakash, "Choice of filters for the detection of gravitational waves from coalescing binaries. 2. Detection in colored noise," *Phys. Rev.*, vol. D49, pp. 1707–1722, 1994.

[7] B. J. Owen, "Search templates for gravitational waves from inspiraling binaries: Choice of template spacing," *Phys. Rev.*, vol. D53, pp. 6749–6761, 1996.

[8] B. Allen, "χ^2 time-frequency discriminator for gravitational wave detection," *Phys. Rev.*, vol. D71, p. 062001, 2005.

[9] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, "FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries," *Phys. Rev.*, vol. D85, p. 122006, 2012.

[10] S. Babak, H. Grote, M. Hewitson, H. Lück, and K. A. Strain, "Signal based vetoes for the detection of gravitational waves from inspiralling compact binaries," *Phys. Rev. D*, vol. 72, p. 022002, Jul 2005.

[11] S. Dhurandhar, A. Gupta, B. Gadre, and S. Bose, "A unified approach to χ^2 discriminators for searches of gravitational waves from compact binary coalescences," *Phys. Rev.*, vol. D96, no. 10, p. 103018, 2017.

[12] D. George, H. Shen, and E. A. Huerta, "Classification and unsupervised clustering of ligo data with deep transfer learning," *Phys. Rev. D*, vol. 97, p. 101501, May 2018.

[13] D. Gabor, "Theory of communication. part 1: The analysis of information," *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, vol. 93, pp. 429–441, November 1946.

[14] S. Bose, S. Dhurandhar, A. Gupta, and A. Lundgren, "Towards mitigating the effect of sine-Gaussian noise transients on searches for gravitational waves from compact binary coalescences," *Phys. Rev.*, vol. D94, p. 122004, Dec. 2016.

[15] S. K. Chatterji, "*The search for gravitational wave bursts in data from the second LIGO science run*," *Ph.D. Thesis*, 2005.

[16] C. Eckart and G. Young *Psychometrika*, vol. 1, no. 26, pp. 211–218, 1936.

[17] "Advanced ligo anticipated sensitivity curves." `https://dcc.ligo.org/LIGO-T0900288/public`.