

# Use of the Bayes Factor to Improve the Detection of Binary Black Hole Systems

Avi Vajpeyi <sup>1</sup>, Rory J. Smith <sup>\*2</sup> and Jonah B. Kanner <sup>†2</sup>

<sup>1</sup>The College of Wooster, Wooster, OH 44691, USA

<sup>2</sup>LIGO Laboratory, California Institute of Technology, Pasadena,  
CA 91125, USA

September 22, 2016

## Abstract

On September 14th, 2015, aLIGO detected the first gravitational wave with a very large Signal-to-Noise Ratio (SNR). In contrast, there are several candidate events with low SNR values, which fall within the background distribution. This paper investigates an alternative detection statistic to SNR, known as the ‘Bayes Factor.’ The Bayes Factor is the ratio between the probability that the strain data contains a gravitational wave signal plus Gaussian noise, to the probability that strain data contains only Gaussian noise. In contrast the SNR is a maximum likelihood estimator, while the Bayes factor takes into consideration all possible binary configurations, including spin orientations and magnitudes. Hence the Bayes factor might prove to be more robust than SNR. This study focuses only on binary black hole systems.

---

<sup>\*</sup>smith\_r@ligo.caltech.edu  
<sup>†</sup>jkanner@caltech.edu

# 1 Introduction

The completion of the two Advanced Laser Interferometer Gravitational-Wave Observatories (aLIGO) has led to the discovery of a gravitational-wave signal [4]. This paper deals with a new detection statistic, involving Bayesian statistics, to rank the different candidate events (the strains in the data sets that could be due to gravitational waves). The candidate events we focus on are from binary black hole systems as they are believed to be common in the Universe [3].

Before we study this new detection statistic, we will discuss how data is currently being recorded and analysed.

## 1.1 Introduction to the physics of LIGO

Each of the aLIGO observatories uses a modified Michaelson Interferometer that measures the difference in the length of the orthogonal arms of the observatory to detect the presence of a gravitational wave [4]. On passing, a gravitational wave induces a difference in the length of the arms which is measured as a phase shift in the circulating laser light.

To determine if data recorded by LIGO store gravitational wave information, the data is processed with two search techniques. One search looks for generic transient waveforms (unmodelled or unexpected waveforms) in data [7]. The second is a match filtered search that compares the data with templates of waveforms generated by general relativity [7].

Both the search processes are made challenging due to the background noise present in the data. This background noise can result from defects in a mirror, the uncertainty in the number of photons travelling in a the laser beam (shot noise), seismic activity, or even thermal noise generated by the Brownian motion of electrons inside circuits [9]. To separate strains caused by background noise from those caused by gravitational waves, the data from one LIGO observatory is compared with another LIGO observatory's data [4]. To compare the data sets, one of the them can be time-shifted so that it matches the data in the other detector over the light-travel time between the detectors. That is, one data set can be time-shifted so that both of the data sets lie along the same interval of time.

After making the necessary time shift, events present in one data set but not the other become apparent. This process cuts a majority of the strains that may have been present due to noise [4]. We attempt to discount any remaining noise-strains by implementing a detection statistic to rank the strains according to the likelihood that the strain resulted from a gravitational wave.

The detection statistic currently being used to rank strains according to the

maximum likelihood of it being due to a gravitational wave is the signal-to-noise ratio (SNR) detection technique [7]. This method compares the power of the strain signal to the power of the remaining noise at a given point [9]. Although the power of the noise is difficult to calculate, we know that this ranking process can make some gravitational wave strains stand out, as seen from the Fig 1. However, we believe that this method lacks the sensitivity necessary to detect some gravitational waves that do not have SNR values as high as those of GW150914. Hence we would like to investigate an alternate detection statistic, specifically one that compares the relative probability that a data set contains a strain due to a gravitational wave to the probability that the data just contains noise. This ratio is known as the Bayes Factor.

Additionally, the templates that are used in the matched filtered searches account for only a small subset of the vast possibility of gravitational wave signals. This might result with some gravitational waves signals that do not have templates to be left undetected. In contrast, the LALInference program, which is what will be used to implement the new detection statistic, uses templates that can describe a far bigger set of signals. Hence, the Bayes factor might prove to be more sensitive than SNR, as we will have more information about the signal.

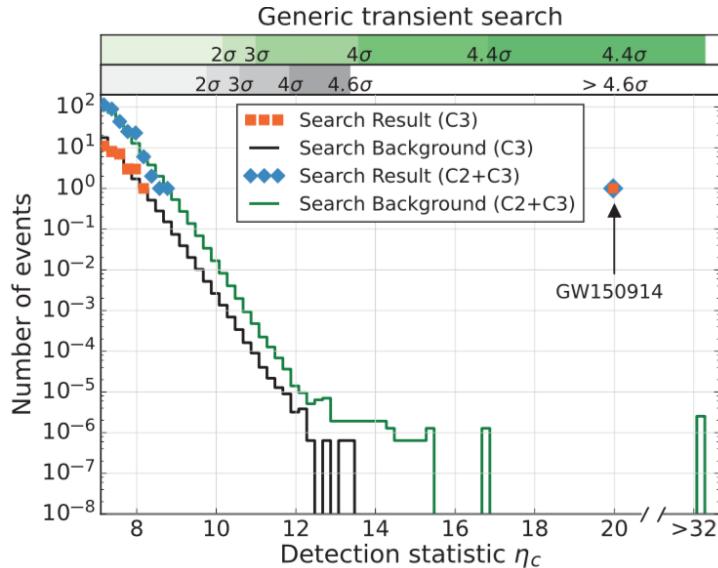


Figure 1: Search results of the generic transit search in which the first gravitational wave signal, marked GW150914, was detected. The wave strain of GW150914 stands out as the strongest strain in the entire search. SNR was used in this search to calculate the background noise and the events. Figure taken from [4]

## 1.2 Introduction to Bayesian Inference for Gravitational Wave Analysis

To compute the Bayes factor between two hypotheses, we need to first define the models we are comparing. The models are descriptions of the data  $d(t)$ , which contains either only noise, or noise along with a gravitational wave signal, parametrized by a certain vector  $\vec{\theta}$ . The data  $d(t)$  is recorded in the time domain, but can be written in the frequency domain at the frequency  $f$  of the detector as  $\tilde{d}(f) = \tilde{h}(f) + \tilde{n}(f)$ . The Fourier Transformed data  $\tilde{d}(f)$  is at times easier to work with.

The parameter vector  $\vec{\theta}$  contains information on several quantities describing the binary black hole system such as the masses of the black holes ( $m_1, m_2$ ), their spin vectors, and the distance of the black holes from Earth [10]. This vector also contains quantities such as the GPS time at the geocentre at which the gravitational wave passes the Earth ( $t_0$ ), and the phase of the signal at this time ( $\phi_0$ ) [10]. The observed signal is actually dependant on fifteen quantities [10]:

$$\vec{\theta} = \{m_1, m_2, t_0, \phi_0, D_L, \alpha, \delta, \psi, \iota, \vec{s}_1, \vec{s}_2\}.$$

The noise that is present in the data is assumed to be a Gaussian with a mean of zero, and a variance described by the noise spectral density of the data,  $S_n(f)$ , at a frequency  $f$  [10]. Under these standard assumptions, we can define the likelihood of noise to be given by the equation

$$P(n(f)) \propto \frac{1}{\sqrt{2\pi\sigma(f)^2}} e^{\frac{1}{2}\left(\frac{n(f)}{\sigma(f)}\right)^2}, \quad (1)$$

with  $\mu = 0$  and  $\sigma = S_n(f)$ .

Finally, we can formally define the models we will use as:

- $\mathcal{H}_{null}$ : the noise model, which corresponds to the hypothesis that  $d(t)$  contains only noise,  $\mathcal{H}_{null} : d(t) = n(t)$ .
- $\mathcal{H}_{GW}$ : the gravitational wave signal model, which corresponds to the hypothesis that the data contains noise, and a gravitational wave signal parametrised by  $\vec{\theta}$ . Hence the model is defined by  $\mathcal{H}_{GW} : d(t) = h(\vec{\theta}, t) + n(t)$ .

We can compare these two models by calculating the relative probabilities in the form of the posterior odds ratio  $O_{GW,null}$  between the two of them, as explained in Ref [10],

$$\begin{aligned} O_{GW, \text{null}} &= \frac{P(d|\mathcal{H}_{GW})}{P(d|\mathcal{H}_{\text{null}})} \frac{P(\mathcal{H}_{GW})}{P(\mathcal{H}_{\text{null}})} \\ &= B_{GW, \text{null}} \frac{P(\mathcal{H}_{GW})}{P(\mathcal{H}_{\text{null}})}, \end{aligned} \quad (2)$$

where  $B_{GW, \text{null}}$  is the ‘Bayes’ Factor,’ which is equal to :

$$B_{GW, \text{null}} = \frac{P(d|\mathcal{H}_{GW})}{P(d|\mathcal{H}_{\text{null}})}. \quad (3)$$

It should be noted that the relative probability, Eq 2, and the Bayes factor, Eq 3, contain no reference to the signal parameters  $\vec{\theta}$ . Hence, the Bayes factor  $B_{GW, \text{null}}$  can be calculated from our hypothesis, for any choice of model.

To be able to account for the set of parameters  $\vec{\theta}$  for a gravitational wave, the likelihood of the model  $\mathcal{H}_{GW}$  needs to be marginalised over all the parameters, weighted by their prior probability distribution, giving the marginal likelihood or evidence, given by  $P(d|\mathcal{H}_{GW})$  [10].

To calculate  $P(d|\mathcal{H}_{GW})$ , first we need to determine the *posterior probability density function* (PDF). For gravitational wave analysis, the PDF of the parameters  $\vec{\theta}$  is [9]:

$$p(\vec{\theta}|d, \mathcal{H}_{GW}) = \frac{p(\vec{\theta}|\mathcal{H}_{GW}) p(d|\vec{\theta}, \mathcal{H}_{GW})}{p(d|\mathcal{H}_{GW})}, \quad (4)$$

or in other terms,

$$\text{Posterior Probability Density Function} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}.$$

In other words, in Eq 4, we have each model  $\mathcal{H}_{GW}$ , to have a vector of parameters  $\vec{\theta}$ , with which we can calculate a ‘Prior’ distribution of  $P(\vec{\theta}|\mathcal{H}_{GW})$ . This prior states what specific values the model  $\mathcal{H}_{GW}$  might be expected to take from the data set  $d$ . We also have  $p(d|\vec{\theta}, \mathcal{H})$ , which is the ‘likelihood’ of the data, given that the model  $\mathcal{H}_{GW}$  is true. This likelihood can be calculated for a model  $\mathcal{H}_i$  with

$$\begin{aligned} p(d|h, \mathcal{H}_i) &\propto \exp \left[ -\frac{1}{2} (d - h|d - h) \right] \\ &\propto \exp \left[ -\frac{2}{T} \sum_{k>0} \frac{|\tilde{d}(f_k) - \tilde{h}(f_k)|^2}{S(f_k)} \right], \end{aligned} \quad (5)$$

as seen in Ref [10]. In this equation, the term  $(d - h|d - h)$  is referred to as a noise-weighted inner product, which is a mathematical tool used often in signal

analysis. The noise weighted inner product of two signals  $A$  and  $B$  can be calculated as shown in Ref [8],

$$(A|B) = 2 \int_0^\infty \frac{\tilde{A}^*(f)\tilde{B}(f) + \tilde{A}(f)\tilde{B}^*(f)}{S_n(f)} df.$$

Additionally, the term  $T$  is the duration of the data that contains the signal of the gravitational wave. It is equal to the inverse of the size of the discrete frequency being stepped over:  $df = 1/T$

To finally solve for  $p(d|\mathcal{H}_{GW})$ , we rearrange Eq 4 and integrate over  $\vec{\theta}$ , to get

$$P(d|\mathcal{H}_{GW}) = \int_{\Theta} P(\vec{\theta}|\mathcal{H}_{GW}) P(d|\vec{\theta}, \mathcal{H}_{GW}) d\vec{\theta} \quad (6)$$

$$Evidence = \int_{\Theta} (Prior) (Likelihood) d\vec{\theta} \quad (7)$$

since the integral of the PDF, by definition of a probability density, is

$$\int_{\Theta} p(\vec{\theta}|d, \mathcal{H}_{GW}) d\vec{\theta} = 1.$$

We can now use the value for  $p(d|\mathcal{H}_{GW})$  from Eq 6, and substitute it into our equation for the Bayes Factor, Eq 3. This Bayes' Factor may be able to better highlight the strains due to gravitational waves in our data.

We will demonstrate the use of the Bayes factor by generating a figure, like Fig 1. However, instead of using SNR to calculate the background noise and rank the candidate events, the Bayes factor will be used.

## 2 Approach

The main objective of this project will be to assess the sensitivity of the Bayes' Factor to rank gravitational waves that might have been missed using SNR as the detection-statistic. This analysis will be done with a procedure that we will write, with the help of the LALInference program.

We will begin by establishing the search background of the noise, with the help of the Bayes factor. This search background is made after time-shifting the data, as discussed in Section 1.1. We will then inject gravitational wave signal into data sets, and use the new detection statistic to study if it can detect the waves. If the Bayes Factor is unable to detect the gravitational wave signal, we will study the thresholds at which the the gravitational wave signals can finally be detected with the Bayes factor.

We will then study past data and see if the Bayes Factor detection statistic can help extract more gravitational waves signals from the noise in the various data sets. We will also study whether the new detection statistic will be able to detect gravitational waves from quieter events, that would appear as the background of SNR, as seen in Fig 1.

## 3 A Simple Bayesian Statistics Calculation

Bayesian Statistics can get complicated to program. To understand programming techniques to compute the Bayes factor between two hypothesis and to evaluate the posterior densities for unknown parameters in models, we began the summer by studying a simple example.

In the example, there was some data  $d(x)$  provided that had some Gaussian noise built in with a standard deviation of  $\sigma = 5$ . We were asked to fit the data with two models. The first model proposed that the data consists of a straight line passing through the origin and some noise, and the second model hypothesised that the data contains only noise. With this information, we can define the two models:

- $\mathcal{H}_m$ : the signal model, which corresponds to the hypothesis that  $d(x)$  contains a line passing through the origin ( $mx$ ) and some Gaussian noise  $n(x)$ ,

$$\mathcal{H}_m : d(x) = mx + n(x)$$

- $\mathcal{H}_n$ : the noise model, which corresponds to the hypothesis that the data contains only Gaussian noise, and is defined as

$$\mathcal{H}_n : d(x) = n(x)$$

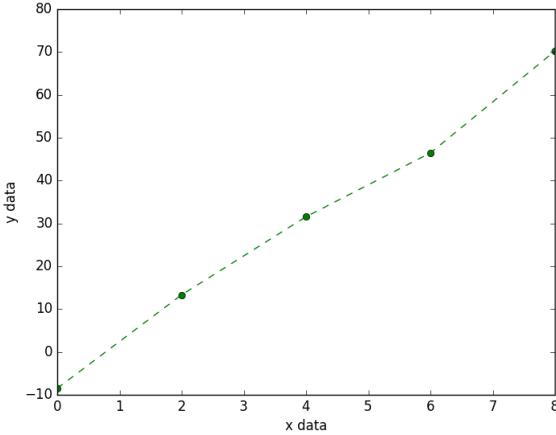


Figure 2: Plots of the raw data used in this example.

In both of the models, the noise can be defined by Gaussian as in the Eq 1. In this case, we set  $\sigma^2 = 5$  and  $\mu = 0$  to get:

$$n(x) = \frac{1}{5\sqrt{2\pi}} e^{\frac{1}{2}(\frac{x}{5})^2}. \quad (8)$$

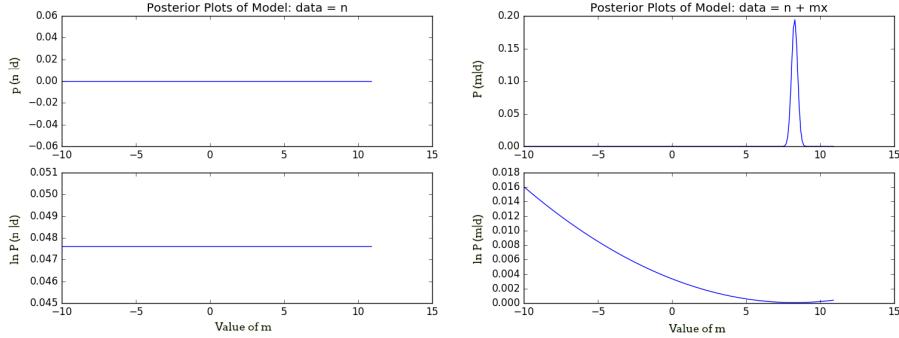


Figure 3: Plots of the Posteriors

Modifying the equation for the calculation of the posterior density function, Equation 4, we calculated the PDF for  $\mathcal{H}_m$  with

$$P(m|d, \mathcal{H}_m : d(x) = mx + n(x)) = \frac{P(m|\mathcal{H}_m) P(d|m, \mathcal{H}_m)}{P(d|\mathcal{H}_m)}, \quad (9)$$

and we calculated the posterior density for  $\mathcal{H}_n$  with

$$P(m|d, \mathcal{H}_n : d(x) = n(x)) = \frac{P(m|\mathcal{H}_n) P(d|m, \mathcal{H}_n)}{P(d|\mathcal{H}_n)}. \quad (10)$$

In both equations, we calculated the likelihood function using Eq 5. For example, the likelihood for the noise model can be written as

$$P(d|d(x) = n(x)) \propto \exp \left[ -\frac{1}{2} \sum_{i>0} \frac{|n(x_i)|^2}{\sigma} \right]$$

while the likelihood for the model  $\mathcal{H}_m$  is

$$\begin{aligned} P(d|d(x) = mx + n(x)) &= P(d|n(x) = d(x) - mx) \\ &\propto \exp \left[ -\frac{1}{2} \sum_{i>0} \frac{|d(x_i) - mx_i|^2}{\sigma} \right] \end{aligned}$$

Plots of the posterior densities for both the models can be seen in Figure 3. Looking at the plot for the posterior distributions for the  $\mathcal{H}_m$  model, we can see a peak at  $7.6 \pm 0.4$ , which is what the most probable value for the slope of the line would be. Looking at the plots for the noise model, we see that there is no regions of high probability, suggesting that the noise model does not fit the data well.

To see how well the model  $\mathcal{H}_m$  fits the data compared to model  $\mathcal{H}_n$ , we can compute the Bayes factors of the models. From Eq 3, we know that the Bayes factor is the ratio of the marginal likelihoods for the two models. Hence, in this case, the Bayes factor between the  $\mathcal{H}_m$  and  $\mathcal{H}_n$  models is

$$B_{m,n} = \frac{P(d|\mathcal{H}_m)}{P(d|\mathcal{H}_n)} = \exp [233170].$$

This value is very high, and hence, this demonstrates that the model  $\mathcal{H}_m$  is a much more likely model than  $\mathcal{H}_n$ . Hence, from this, we can conclude that the given data  $d(x)$  contains a line that passes through the origin and has a slope of  $7.6 \pm 0.4$  with some noise. In the next section, we will discuss how we can use Bayesian Inference to study gravitational wave strain data.

## 4 An Example of Bayesian Inference for Binary In-Spirals

As seen in the previous example, Bayesian inference can provide efficient and powerful approaches to model selection and parameter estimation. In this section, we describe the computational processes and theory needed to analyse some aLIGO strain data. We discuss how Bayesian inference can be used to calculate both the Bayes factor of the signal and the parameters that generated the signal. Note that this example uses gravitational wave strain data that was injected into noise data collected from the detectors during the observing run of 2015.

The starting point for data analysis begins with the strain data  $h(t)$  that is recorded by LIGO during the duration of a run. This data is passed through a band pass filter to allow only signals between 35-350 Hz, as this is the range in which LIGO is sensitive to frequencies. The data is then run through another filter to reject frequencies that are from known noise sources [6].

After the data is filtered, it is run through a series of generic gravitational wave templates [5]. If the data matches one of the generic templates, it is considered to be a candidate event, and is then looked at again more closely.

### 4.1 Model Selection

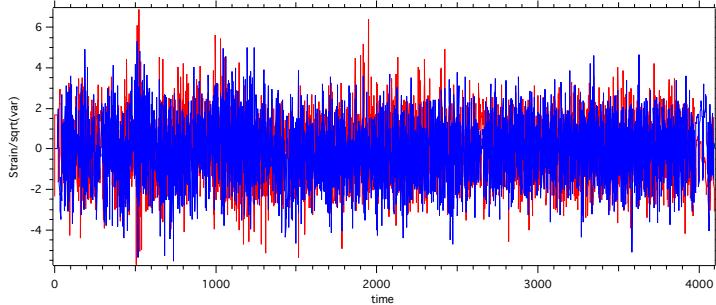


Figure 4: Whitened strain data's imaginary (red) and real (blue) parts of the frequency from one observatory. The data is from the first observing run in 2015, and is supposed to contain a candidate event.

In Figure 4 we have some data that is considered to contain a candidate event. The figure contains the whitened strain data (the strain data divided by the noise's standard deviation) plotted against the frequency. We obtain the real and the imaginary parts of the frequency by calculating a Fast Fourier Transform of the initial time data as discussed in the previous section.

Since the data in Figure 4 contains a candidate event, the data can be fit by either of the two models:

- $\mathcal{H}_{GW}$  model - a gravitational wave strain and noise
- $\mathcal{H}_{null}$  model - noise that triggered the program to consider it a candidate event

To select which model fits the data better, we can calculate the marginal likelihoods for both the models, with Eq 6. Since we have data from two observatories, and since in this case, we only look at two parameters, the equation for calculating the marginal likelihood can be re-written for the  $\mathcal{H}_{GW}$  model as

$$P(d_L, d_H | h \propto m_1, m_2) = \int_{m_1, m_2} P(m_1, m_2) P(m_1, m_2 | d_H) dm_1 dm_2 \\ \times \int_{m_1, m_2} P(m_1, m_2) P(m_1, m_2 | d_L) dm_1 dm_2 .$$

In this equation, the likelihood for the masses  $m_1$  and  $m_2$  given the data at Hanford and Livingston ( $d_H, d_L$ ) is calculated for each set of values for  $m_1, m_2$  ( $\forall m_2 < m_1$ ). After the likelihood is calculated for all sets of masses ( $\forall m_2 < m_1$ ), the evidence is calculated with Eq 6.

With the evidence for both models, we calculate the log of the Bayes factor, by modifying Eq 3 to

$$\log B_{GW, null} = \log \text{Evidence}_{GW} - \log \text{Evidence}_{null} \\ = 50,000,000.$$

This value suggests that there is a higher probability that the data contains a gravitational wave strain and some noise than the probability that the data contains just noise.

Note that the noise in this data is considered to be a Gaussian with a zero mean, as discussed in section 1.2. This is because if we plot strain data that contains only noise, and create a histogram of the data, we get a plot as seen in Figure 5. This is a Gaussian with a mean of  $\mu \approx 0$ . In most cases, the data that is recorded is mostly noise, and even data with gravitational wave strains appear to look like they only have noise. It is with the various analysis methods that we can find out is the data contains anything other than noise.

## 4.2 Parameter Estimation

Now that we know that the candidate event actually contains a signal, we can use the data for the posteriors we calculated when we were calculating the Marginal likelihood to calculate the posterior distribution for parameters. Here

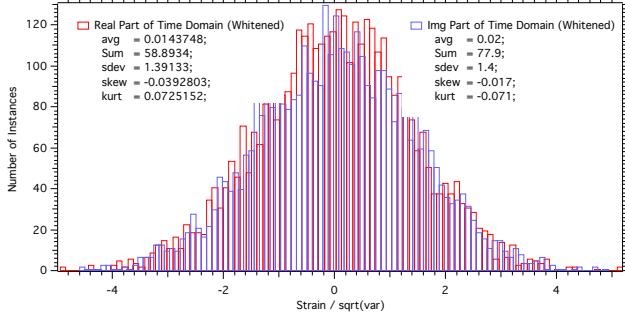


Figure 5: A histogram of strain data that does not have any strain due to a gravitational wave. This displays that noise in the data is Gaussian with a zero mean.

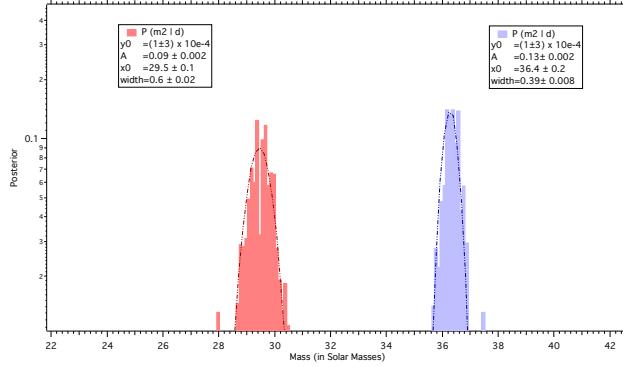


Figure 6: Posteriors for the masses of the black holes before the merger. This was created with data from an injected gravitational wave strain. The masses of the black holes are  $29.5 \pm 0.1$  and  $36.4 \pm 0.2$  solar masses in size.

we calculate the posterior distributions for the masses of the merging binary black holes. Using Eq 4 and 5, if we plot the likelihoods for the entire set of combinations of the masses, we can get a plot as seen in Figure 7. This indicates that the masses of the black holes are most likely to be in the range of 30 - 37 solar masses in size. Note that the plot does not include the posterior density for all the combination of masses. This is because after the diagonal, the combinations of the masses would be repeated.

To get a better estimation on what the mass for each black hole may be, we can integrate out the posterior for both masses, with the equation

$$P(m_1|d) = \int_{m_2} P(m_1, m_2|d) dm_2.$$

The plot of the posterior density of the masses can be studied in Figure 6.

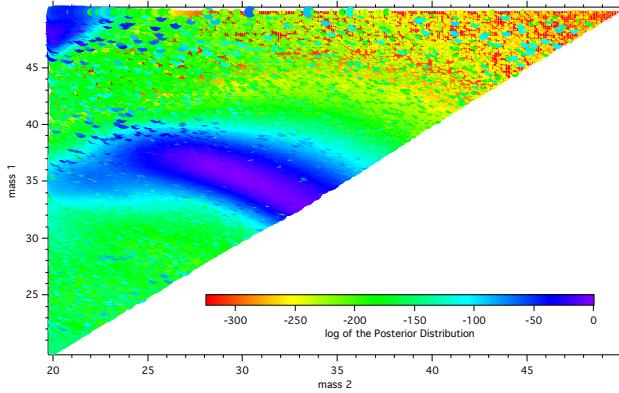


Figure 7: 2D plot of the log of the posterior distribution for the masses of the black holes before the merger. We can see a high posterior region near the 30–38 solar mass range. The data for this was generated from an injected gravitational wave strain.

From this, we can conclude that the masses for the black holes are approximately  $29.5 \pm 0.1$  and  $36.4 \pm 0.2$  solar masses in size. This method is similar to how parameter estimation is done for all the parameters for a gravitational wave signal. A discussion on how a gravitational wave signal is analysed for all the parameters is done in the next section.

## 5 Running LALinference to Analyse A Candidate Event

To analyse a candidate event, such as the one seen in Fig 8, there have been scripts written that use several computers to compare numerous waveform templates with the signal data. The templates are different for each parameter set, and once a certain number of templates are run for various parameters, numerous plots of posterior destnies for various parameters along with other information regarding the candidate event are displayed on a results page.

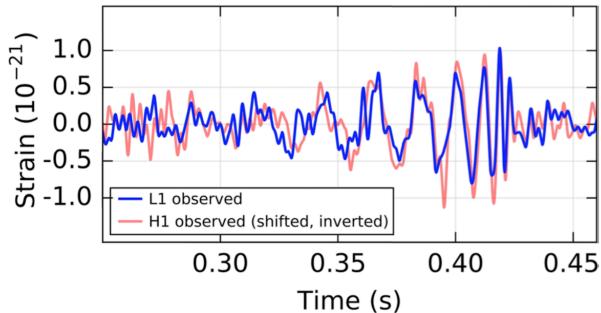


Figure 8: The filtered instrumental strain data in the Livingston detector (blue) and Hanford detector (red). The Hanford strain has been shifted back in time by 6.9 ms and inverted. This figure was taken from Ref [6].

Depending on the initial parameters passed to the function to analyse the candidate event, the results page will include the log of the Bayes factor and the Bayes factor of the gravitational wave strain model against the noise model. This value helps determine if the candidate event has information regarding a gravitational wave strain, or if it contains just noise.

This sections discusses an analysis of the *GW150914* gravitational wave, detected during the first observatory run of aLIGO. This analysis was conducted with the ‘`lalinference_nest`’ script, which uses a Nested Sampling algorithm to calculate the evidence.

The nested sampling algorithm randomly pick sets of parameters in  $\vec{\theta}$ , and first calculates likelihoods for each set of parameters. Once a certain amount of likelihoods are calculated, the algorithm calculates the evidence using the several sets of likelihoods. If the value of the evidence is below a certain threshold set before running the script, the nested sampling algorithm chooses another set of values of the parameters in  $\vec{\theta}$  and calculates the likelihood and the evidence again with the new set of parameters, along with the rest of the old sets of parameters. Note that after the initial sets of parameters are chosen, the new

set of parameters are used for the calculations only if the likelihood for the new parameter set is a value above a certain amount. This process of picking a set of the new parameter sets is repeated until the likelihoods are the maximum ones present in the parameter space.

On running the analysis, we received a Bayes Factor of  $5.77 \times 10^{36}$ , which suggests that there is some gravitational wave strain signal present in the data.

As the data contains a gravitational wave, we were able to calculate various parameters such as the masses, spins and orientations of the black holes. Figure 9 contains plots of the masses of the black holes, which informs us that the masses of the black holes are  $39.5 \pm 3.2$  and  $32.4 \pm 2.7$  solar masses. Several plots similar to the one in Figure 9 were generated for the other parameters of the binary black hole system. These help us state what values we believe that the system has.

We will run this analysis for several sets of instances where no gravitational wave strain is present to calculate the Bayes factor for these instances. As no gravitational waves are present in this data, we can create a background distribution of the value that we expect the Bayes factor to be if no gravitational wave data is present in the strain data. We will then run this analysis for the several candidate events that we observed to see if they have higher Bayes factors than the background distribution.

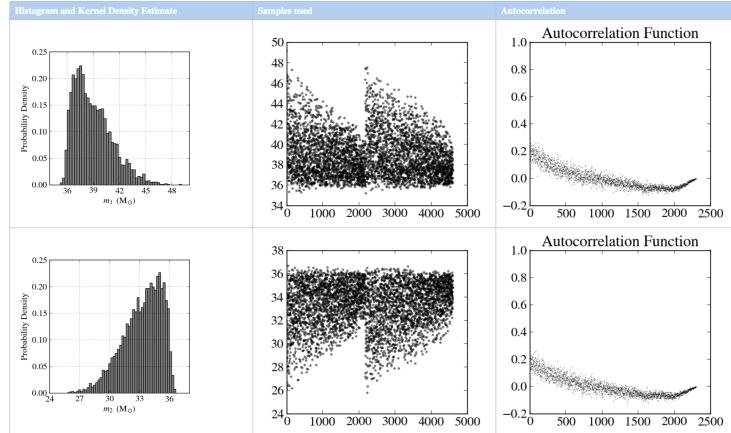


Figure 9: Posteriors for the black hole masses before the merger, created with the LALinference script. The masses are  $39.5 \pm 3.2$  and  $32.4 \pm 2.7$  solar masses in size. The data used for this was the same as the data from the *GW150914* detection.

## 6 Determining a New Detection Statistic

### 6.1 Preparing Sample Points

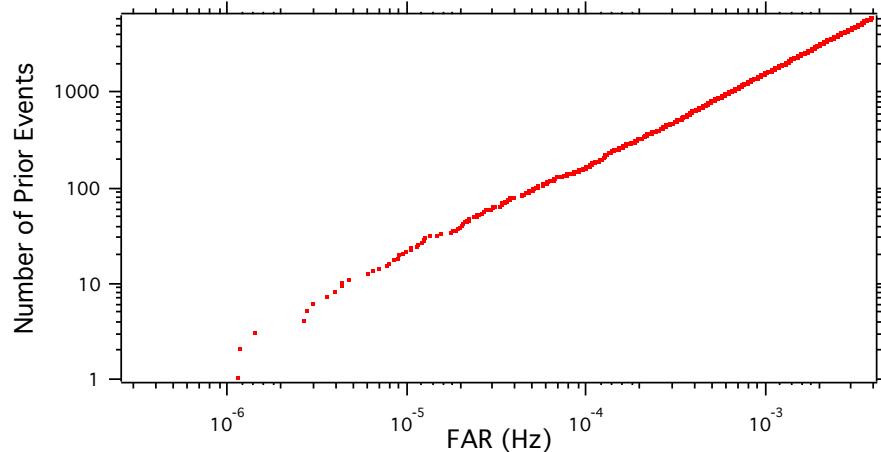


Figure 10: The False Alarm Rate (FAR) for background events from the GSTlal search plotted against the number of prior False Alarm Rates. The lowest FAR value in this graph is  $10^{-7}$ . We see that there is a larger spread of the FAR values between each data point for the first thousand data points.

In order to create the background distribution for the new detection statistic, we need to calculate the Bayes Factor for artificial noise events (also known as background events). Noise events are events in the strain data that can be due to noise sources, like seismic activity, and have a likelihood or SNR above a certain threshold. To create a list of artificial noise events (background events), a time shift is used to separate the Hanford and Livingston data streams. As no gravitational wave signal can be coincident in the data streams after the artificial time shift (time shift > light travel time), we can be certain that the time-shifted data contains only events due to noise. The background events we used in our study was generated from GSTlal and pyCBC searches. The events GSTlal were used as an initial study, after which we analysed the pyCBC events.

#### 6.1.1 Sampled points from GSTlal Searches

To create a list of background events, an artificial time off-set was applied between the Hanford and Livingston data streams.

For the GSTlal searches, we used a fixed time off-set of 25.13 s for each background event. The list of background events separated by this offset had been previously generated when a search for transient waveforms with the GSTlal

searches had been conducted.

The searches were conducted in the 17 highest mass bins. The values of chirp masses,  $\mathcal{M}$ , defined as

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}},$$

ranged from  $(6.6 \text{ to } 45.9) M_{\odot}$ . With a threshold of  $\text{SNR} \geq 4.0$ , 186,000 background events had been yielded. This data was retrieved from Ref [1].

For each background event generated from the time-shifts, the searches calculated a False Alarm Rate (FAR), associated to it. FAR for an event is the count of background events per year of the same likelihood. In other words, a low FAR value implies that there are fewer background events of such likelihoods.

A figure of the FAR for the entire set of background events from the GSTlal search can be seen in Fig 10. We selected the lowest thousand FAR values, and used these as the sample points for which we calculated the Bayes Factor for. The lower thousand FAR values were chosen as they differ in their FAR values from  $10^{-7}$  to  $10^{-3}$  per year, while in higher FAR regions, the change in FAR is very small. A plot of the FAR values chosen against the likelihoods for the sampled background event points can be seen in Fig 11.

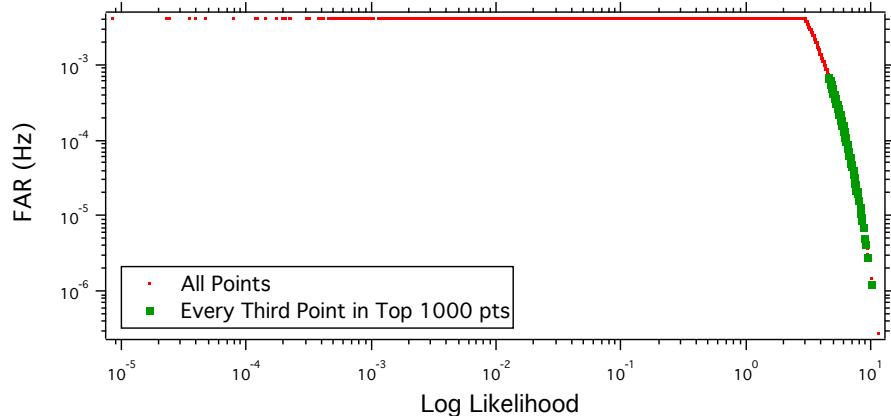


Figure 11: The False Alarm Rate plotted against the log Likelihood for the GSTlal background events. The green data points are those which we are using to generate the background distribution with the Bayes Factor. We see that for lower log Likelihood values, the FAR does not appear to be changing much. However, at higher values for the log Likelihood ( $\sim 3$ ), smaller changes cause a large variation in the value for the FAR. Additionally, we see that at higher log Likelihood values, the FAR is lower. This means that there is a smaller chance of having a background event of high log Likelihoods.

The Bayes Factors for these down sampled background GSTlal events were then calculated as a preliminary test. Discussion on the results for this can be found in Section 6.3. Following the analysis, we decided to look at background events that were calculated using varying time shifts so as to get lower FAR values. To do this, we used data from the pyCBC searches.

### 6.1.2 Sampled Events from pyCBC searches

Similar to how GSTlal searches created several thousand background events with one artificial time shift, the pyCBC searches had created lists of several millions of background events. This was done by taking the same data as used in the GSTlal search, and applying multiple different artificial time slides. As different time slides were used, the effective amount of background data that was present to be analysed was in the order of several million years.

Because of the massive volumes and abundance of the background events, the data from the pyCBC searches were heavily down sampled from several million to 300 background events. This data was taken from [2]. Each of these events were then analysed separately for the  $\mathcal{M}$  bins in the range of  $(6.6 \text{ to } 45.9) M_{\odot}$ . The analysis for these is discussed in Section 6.3.

Before we look at the results of the analysis, the next section discusses the different types of results the analysis produces.

## 6.2 The Different forms of the Bayes Factor

To calculate the Bayes Factor, the sampled background events were analysed with IMRPhenomPv2pseudoFourPN. The  $\mathcal{M}$  bins used were in the range of  $(6.6 \text{ to } 45.9) M_{\odot}$ . This produced several different useful Bayes Factor values:

- $\mathcal{B}_{SN}$  - the signal to noise coherent log Bayes Factor.
- $\mathcal{B}_{SN}^{I(D)}$  - the incoherent log Bayes Factors at a detector ‘D.’
- $\mathcal{B}_R$  - the coherent versus incoherent log Bayes Factor.

The signal to noise coherent log Bayes Factor is calculated as

$$\begin{aligned} \log \mathcal{B}_{SN} = & \log P(d_{HL} | \mathcal{H} : d_{HL} = h + n) \\ & - \log P(d_{HL} | \mathcal{H} : d_{HL} = n). \end{aligned} \quad (11)$$

Here,  $d_{HL}$  is the data from both the aLIGO observatories at Hanford and Livingston.

The incoherent log Bayes Factor  $\mathcal{B}_{SN}^{I(D)}$ , is what the Bayes Factor is calculated to be using only one detector  $D$ ’s data set.

Finally, the coherent versus incoherent log Bayes Factor, also known as the log Bayes Coherence ratio,  $\mathcal{B}_R$ , is calculated as

$$\mathcal{B}_R = \frac{\mathcal{B}_{SN}}{\mathcal{B}_{SN}^{I(H)} + \mathcal{B}_{SN}^{I(L)}}. \quad (12)$$

This is a ratio of the  $\mathcal{B}_{SN}$  and the sum of the  $\mathcal{B}_{SN}^{I(D)}$  of both the observatory's data sets.

We can now try to compare these different Bayes factors for the background events to the SNR, to understand how they can be used as a detection statistic.

### 6.3 Comparing Detection Statistics

#### 6.3.1 Bayes Factors as a Detection Statistic

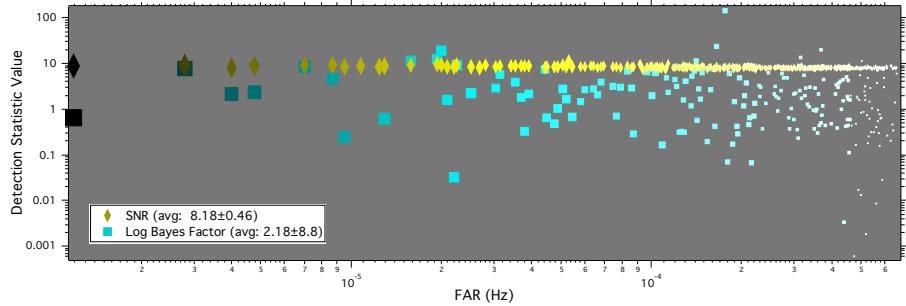


Figure 12: The SNR (yellow diamonds) and Log Bayes factor (cyan squares) plotted against FAR. The Data points are coloured and sized according to their Ln Likelihoods (dark end  $\sim 4.6$  to the white end  $\sim 10.3$ ).

After analysing the sampled points, we were able to compare the  $\mathcal{B}_{SN}$  (Coherent Bayes Factor) and the SNR for the background events. A plot of the SNR and the  $\mathcal{B}_{SN}$  against the FAR can be seen in Fig 12. From this plot, we can observe that for some of the sampled points, the  $\mathcal{B}_{SN}$  is very high. Again, in Fig 13, we see several large  $\mathcal{B}_{SN}$  for background events. When compared to the  $\mathcal{B}_{SN}$  for the gravitational waves events, some of the values for the background appear larger.

These high  $\mathcal{B}_{SN}$  for the background events could be due to the data not fitting the noise model well. This occurs if the noise in the data  $d_{HL}$  is not Gaussian, which is what is expected by the model.

Another reason for high  $\mathcal{B}_{SN}$ 's could be due to glitches in data. For example, the background event which occurred at the GPS time of 1127409058.94s has a log  $\mathcal{B}_{SN} = 142.82$ . On checking the incoherent log Bayes Factors obtained from

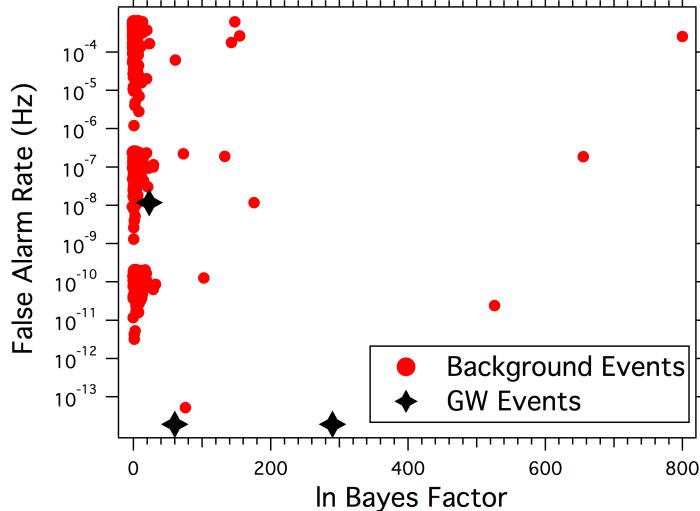


Figure 13: The Bayes Factor as a Detection Statistic, using only Coalescing Binary Back Hole Templates. We can see that there are some background events that have higher Bayes Factors than the gravitational wave events.

the Hanford and Livingston data sets ( $\log \mathcal{B}_{SN}^{I(H)}$  and  $\log \mathcal{B}_{SN}^{I(L)}$ ), we noted that they are

$$\begin{aligned}\log \mathcal{B}_{SN}^{I(H)} &= 152.58, \\ \log \mathcal{B}_{SN}^{I(L)} &= 0.91.\end{aligned}$$

Since in this case,  $\log \mathcal{B}_{SN}^{I(H)} \gg \log \mathcal{B}_{SN}^{I(L)}$ , we believe that a glitch was present in the Hanford data which in turn caused this background event's  $\log \mathcal{B}_{SN}^{I(H)}$  to be large. As  $\mathcal{B}_{SN}$  is calculated using data from both observatories, the glitch in the Hanford data causes the coherent  $\log \mathcal{B}_{SN}$  to be higher than it should have been if no glitch were present.

We would like  $\mathcal{B}_{SN}$  for background events to be high only if both  $\log \mathcal{B}_{SN}^{I(H)}$  and  $\log \mathcal{B}_{SN}^{I(L)}$  are high. Thus, we decided to investigate if the Bayes Coherence Ratio,  $\mathcal{B}_R$ , could prove to do this instead of the Coherent Bayes Factor,  $\mathcal{B}_{SN}$ .

### 6.3.2 Bayes Coherence Ratio as a Detection Statistic

As discussed in the previous section, we wanted to investigate the usefulness of  $\mathcal{B}_R$  (Bayes Coherence Ratio) as a detection statistic. Hence, we calculated the  $\mathcal{B}_R$  for the same set of sampled background events. Studying the  $\mathcal{B}_R$  values, we noticed that there were two outliers, which we realized were due to errors while calculating the  $\mathcal{B}_R$ . These values were removed from the data, and are

discussed in the Appendix 8.1.

A histogram of the  $\mathcal{B}_R$  for the pyCBC background events can be studied in Fig 14. From the plot, we see that  $\mathcal{B}_R$  values from the pyCBC data are low for background events. Most of the values fall between -1 and 1.

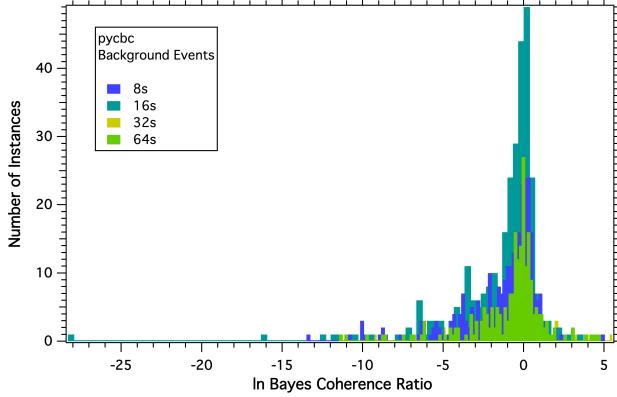


Figure 14: A histogram of the pyCBC data for the different mass bins. We can see that most of the background events yeilded a  $\mathcal{B}_R$  value 0.

We can compare these values of the of the  $\mathcal{B}_R$  and SNR of the background events in Fig 15. From this plot, we see that at higher FAR values, while the SNR fluctuates around its average value  $8.2 \pm 0.5$ , the  $\mathcal{B}_R$  appears to fluctuate more around its average,  $-0.23 \pm 1.2$ . It is important to note these values mean greatly different things. The  $\mathcal{B}_R$  informs us how well the model that the data contains noise and a gravitational wave signal compares against the model that the data contains only noise. As it compares models, to compute it, all possible configurations of templates and parameters must be considered. If the  $\mathcal{B}_R$  is low, this means that the background event has a low probability of being due to a gravitational wave signal, as we would expect.

Unlike the  $\mathcal{B}_R$ , the SNR value does not consider all the possible templates and possible parameter values. It instead demonstrates how likely it is for a specific template to match the given data. Hence we can conclude that the  $\mathcal{B}_R$  is more generic than the SNR of a background event as it is calculated using all possible templates and parameters, rather than just a singular template.

### 6.3.3 Studying SNR, $\mathcal{B}_{SN}$ and $\mathcal{B}_R$ Together

To get a better understanding of how the detection statistics differ, we can compare their values to the Likelihood of the background events in Fig 16.

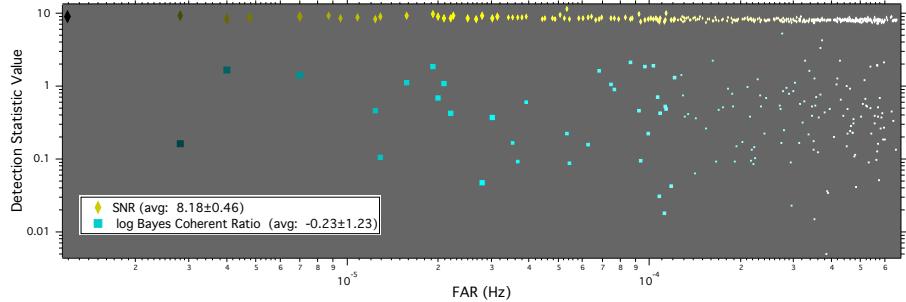


Figure 15: The SNR (yellow diamonds) and Log Bayes factor (cyan squares) plotted against FAR. The data points are coloured and sized according to their Ln Likelihoods (dark end  $\sim 4.6$  to the white end  $\sim 10.3$ ).

Studying the graphs of the detection statistics against the log likelihoods in Fig 16, we observe a correlation between the SNR and the log likelihood as the slope of its fitted line is  $0.25 \pm 0.02$ . In comparison, the correlation between the log  $\mathcal{B}_R$  and the log likelihood appears to be much weaker, given the slope of its fitted line is equal to  $0.05 \pm 0.04$ . If we extrapolate for what the data might be at higher likelihoods, we can hypothesize that at higher likelihoods, we get higher SNR values, but the same log  $\mathcal{B}_R$ s. This is because of the weaker correlation between log  $\mathcal{B}_R$  and the log likelihood of the background events. In other words, at higher likelihood background events, an SNR value might have a higher value than the log  $\mathcal{B}_R$  which is hypothesized to still be near zero in value. This makes sense since the log  $\mathcal{B}_R$  looks at the probability that the model fits the data, and since for background events, the model should not fit the data, the log  $\mathcal{B}_R$  should be low.

Since we expect that the  $\mathcal{B}_R$  will be low for background events at even lower FAR events, we believe that it may be a better detection statistic than SNR and  $\mathcal{B}_R$ .

#### 6.4 Studying Real Events with the Bayes Coherence Ratio

To compare the  $\mathcal{B}_R$  obtained with the background events with those due to real signal data, we calculated the  $\mathcal{B}_R$  for some candidate gravitational wave signal events. The candidate events we chose were the GW150914, GW151226 and LVT151012 events from the first aLIGO observing run.

Fig 17 contains a plot of the FAR of the background events and the events of the O1 data run, against the log  $\mathcal{B}_R$ , and the SNR. Studying the plots, it is evident that there is a clear separation between the  $\mathcal{B}_R$  for the gravitational wave events of the O1 data run, and the  $\mathcal{B}_R$  for the background. Most of the background events appear to cluster together with low  $\mathcal{B}_R$  values. The furthest

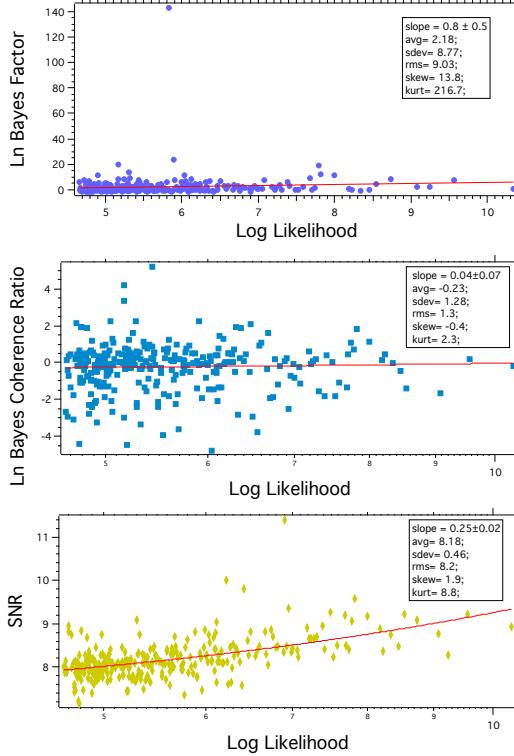


Figure 16: Plots of the various detection statistics against the Log Likelihood of the background events. There appears to be a correlation between the SNR and the log Likelihood, as seen by the slope of the fitted line equal to  $0.25 \pm 0.02$ . The correlation between the log Bayes Coherent Ratio and log Likelihood appears to be less, given that the slope of the fitted line for this is equal  $0.05 \pm 0.04$ . The plot of the Log Bayes Factor demonstrates that it may not be useful as a detection statistic.

background event from the pyCBC searches has  $\mathcal{B}_R = 146$ , and the furthest background event from the GSTlal searches has  $\mathcal{B}_R = 11,800$ . In contrast, the first gravitational wave event on the plot has a  $\mathcal{B}_R = 40,700$ , a value four times further away than the value of furthest background event.

Because of this large separation, we believe that the  $\mathcal{B}_R$  may be able to better discern gravitational wave signals from noise better than SNR and  $\mathcal{B}_{SN}$ .

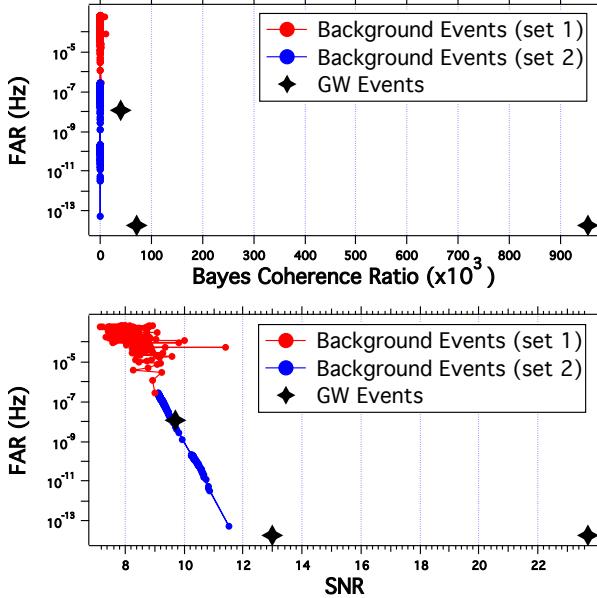


Figure 17: Plots of the FAR against the log  $\mathcal{B}_R$ , and the SNR. Background Events set 1 is from the GSTlal data, and Background Events set 2 is from the pyCBC data. The  $\mathcal{M}$  bins for the data are in the range of  $(6.6 \text{ to } 45.9) M_{\odot}$ . The GW events are the three events in the O1 aLIGO run.

## 7 Future Work

There is still much work left to be done before we can use the  $\mathcal{B}_R$  as a detection statistic. Since we excluded two outliers using the optimal SNR values of the outliers, we need to look at the coherent and incoherent optimal SNR values for all of the background events we have looked at. We also need to analyse the  $\mathcal{B}_R$  for lower FAR values. Additionally, the number of background events analysed is still small. To gain more confidence on the  $\mathcal{B}_R$ , we need to calculate the  $\mathcal{B}_R$  for many ore background events, and then compare the results with  $\mathcal{B}_R$  for real and injected events. Once these tasks are complete it will be interesting to repeat the study for binary neutron star systems.

## 8 Appendix

### 8.1 Removal of Outliers with the Optimal SNR

Here we will discuss two outliers we initially found, and how we were able to remove them. Fig 18 contains a plot of the FAR of the background events and the events of the O1 data run, against the log  $\mathcal{B}_R$ . Studying the plot, it is evident that the  $\mathcal{B}_R$  for two background events are much greater than the  $\mathcal{B}_R$  for

the gravitational wave events in the O1 data run. On independently studying both the background events with some of the additional information generated while calculating the  $\mathcal{B}_R$ , we found that the optimal SNR of the two outliers gave us useful information.

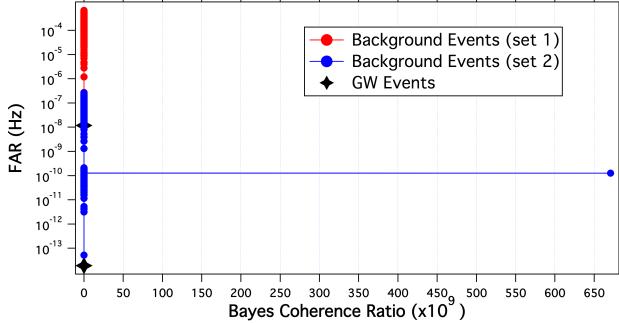


Figure 18: Plots of the FAR against the log  $\mathcal{B}_R$ . Background Events set 1 is from the GSTlal data, and Background Events set 2 is from the pyCBC data. The  $\mathcal{M}$  bins for the data are in the range of  $(6.6 \text{ to } 45.9) M_{\odot}$ . The GW events are the three events in the O1 aLIGO run, and are represented by the black diamonds. We can see that some background events have a higher  $\mathcal{B}_R$  than the GW wave.

When the background events are analysed, first the incoherent data sets (the separate data sets from Livingston and Hanford) are analysed for the specific GPS time of the background event. This produces an incoherent optimal SNR for the Hanford and Livingston data sets. Following this, when the coherent analysis is done (the analysis using both the Hanford and Livingston data sets together), a coherent optimal SNR is produced for both data sets.

Given that the incoherent analysis is done using only one detector's data set, the results should best match that specific detector's data set. In other terms, the results of the incoherent analysis for one observatory should be a better estimate for that observatory's data, than the results of the coherent analysis (using both data sets). This is because the coherent analysis uses both detectors data and hence, will try to estimate for both data sets together.

Because of this, we believe that the optimal SNR for an observatory should be highest when calculated with the incoherent analysis. Additionally, the coherent analysis should produce an optimal SNR that is lower or equal to that produced by the incoherent analysis.

If we look at the coherent and incoherent optimal SNRs for the two outliers in tables 1 and 2, we can see that in both these cases, one of the data set's coherent analysis has a higher coherent optimal SNR than its respective inco-

herent optimal SNR. With some quick checks we were able to study a few other background events and noticed that the coherent optimal SNR is always less than the incoherent optimal SNR. These two outliers are the anomaly.

We believe that the sampling algorithm which calculated the coherent and incoherent optimal SNRs for the background events might be making a mistake while calculating the incoherent optimal SNR. It might be looking at only one maxima value instead of considering other possible maxima values. In turn, the calculations for the  $\mathcal{B}_{SN}^{I(H)}$  and  $\mathcal{B}_{SN}^{I(L)}$  may be miscalculated which we think is causing the large  $\mathcal{B}_R$ .

Table 1: The optimal SNRs for the 8s mass bin's outlier, at the GPS time = 128443202.05. We can see that the Livingston data's coherent optimal SNR is greater than its incoherent optimal SNR.

	<b>Coherent</b>	<b>Incoherent</b>
Livingston	17.26	14.74
Hanford	2.68	4.26

Table 2: The optimal SNRs for the 64s mass bin's outlier, at the GPS time = 1127737863.12. We can see that the Hanford data's coherent optimal SNR is greater than its incoherent optimal SNR.

	<b>Coherent</b>	<b>Incoherent</b>
Livingston	8.05	8.74
Hanford	6.6	5.47

We think that using the  $\mathcal{B}_R$  along with information such as the optimal SNR, may help us discriminate gravitational wave events from noise events. Additionally, we have much more information along with the optimal SNR for each event, which may also be helpful in discriminating signals from noise. In the next section, we discuss the results after removing the outliers.

## References

- [1] GSTlal GW150914 Search Result Data. Available in UWM cluster at [/home/surabhi.sachdev/projects/rory-jonah/coins-time-slide/back-mod-230-247/new/H1.txt](http://home/surabhi.sachdev/projects/rory-jonah/coins-time-slide/back-mod-230-247/new/H1.txt).
- [2] PyCBC GW150914 Search Result C1 Data. Available online at <https://sugar-jobs.phy.syr.edu/~tjmassin/gw150914/gw150914-16day-c01-v1.3.4/>, verion 1.0.
- [3] Jea Abadie, BP Abbott, Richard Abbott, Matthew Abernathy, Timothee Accadia, Fausto Acernese, Carl Adams, Rana Adhikari, Parameswaran

Ajith, Bruce Allen, et al. Predictions for the rates of compact binary coalescences observable by ground-based gravitational-wave detectors. *Classical and Quantum Gravity*, 27(17):173001, 2010.

- [4] BP Abbott, R Abbott, TD Abbott, MR Abernathy, F Acernese, K Ackley, C Adams, T Adams, P Addesso, RX Adhikari, et al. Observation of gravitational waves from a binary black hole merger. *Physical Review Letters*, 116(6), 2016.
- [5] BP Abbott, R Abbott, TD Abbott, MR Abernathy, F Acernese, K Ackley, C Adams, T Adams, P Addesso, RX Adhikari, et al. Observing gravitational-wave transient gw150914 with minimal assumptions. *Physical Review D*, 93(12):122004, 2016.
- [6] LIGO Scientific Collaboration, Virgo Collaboration, et al. Properties of the binary black hole merger gw150914. *arXiv preprint arXiv:1602.03840*, 2016.
- [7] Jonah B Kanner, Tyson B Littenberg, Neil Cornish, Meg Millhouse, Enia Xhakaj, Francesco Salemi, Marco Drago, Gabriele Vedovato, and Sergey Klimenko. Leveraging waveform complexity for confident detection of gravitational waves. *Physical Review D*, 93(2):022002, 2016.
- [8] Lee Lindblom, Benjamin J Owen, and Duncan A Brown. Model waveform accuracy standards for gravitational wave data analysis. *Physical Review D*, 78(12):124020, 2008.
- [9] Rory James Edwin Smith. *Gravitational-wave astronomy with coalescing compact binaries: detection and parameter estimation with advanced detectors*. PhD thesis, University of Birmingham, 2013.
- [10] John Veitch and Alberto Vecchio. Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network. *Physical Review D*, 81(6):062003, 2010.