

School of Physics and Astronomy



Searching for Gravitational Waves

Iain Dorrington

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

16th May 2019

Summary of thesis

I love hitting my head against a wall, so I did that for a bit.

Declaration of authorship

- DECLARATION:

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed: (candidate) Date:

- STATEMENT 1:

This thesis is being submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (PhD).

Signed: (candidate) Date:

- STATEMENT 2:

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed: (candidate) Date:

- STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed: (candidate) Date:

Contents

1	Introduction	1
2	Gravitational Wave Astronomy in a Nutshell	2
2.1	General Relativity	2
2.2	Gravitational waves	2
2.2.1	Linearised Gravity	3
2.2.2	Gauge Transformation	3
2.2.3	Physical Effects of Gravitational Waves	4
2.3	Gravitational Waves Sources	7
2.4	Gravitational Wave Detectors	7
2.4.1	How to find something interesting in all that noise	7
3	Gamma-Ray Bursts	8
3.1	The History of Gamma-ray Burst Astronomy	8
3.1.1	Cold War Tension and an Unexpected Discovery	8
3.1.2	BATSE and the Galactic/Extra-galactic Controversy	9
3.1.3	The Long and Short of Gamma-ray Bursts	11
3.1.4	BeppoSAX and the First Afterglows	14
3.1.5	The Fireball Model	14
3.1.6	The Long GRB-Supernova Connection	17
3.1.7	The Short GRB-Compact Binary Connection	17
3.1.8	Gravitational Waves	20
3.2	GRB Progenitors	20
3.2.1	Compact Binary Coalescence	20
3.2.2	Core Collapse	24
3.3	GRB Detectors	24
3.3.1	Swift	25
3.3.2	Fermi	25
3.3.3	The InterPlanetary Network	25
3.4	GRB Gravitational Wave Search Strategies	26
3.4.1	Triggered and Untriggered Searches	26
3.4.2	Modeled and Burst Searches	26
3.5	GRB Astrophysics with Gravitational Waves	27
4	A Targeted Search for Gravitational Waves associated with Short GRBs	29
4.1	Introduction	29
4.2	PyCBC overview	30
4.2.1	Matched Filter Search	30

4.2.2	Consistency Checks	30
4.2.3	PyCBC Results	30
4.3	PyGRB	30
4.3.1	A Coherent Search for Coalescing Binary Systems	31
4.3.2	Binary Coalescence Waveform	31
4.3.3	Coherent SNR	32
4.3.4	Comparison to Coincident Search	34
4.3.5	Null SNR	36
4.3.6	The PyGRB Workflow	37
4.3.7	Reweighted SNR	38
4.3.8	Searching for Face on Signals	39
4.3.9	Searching over a Sky Patch	40
4.3.10	Event Significance	41
4.3.11	Calculating Search Sensitivity	43
4.4	O2 PyGRB Search	43
4.4.1	GRB sample	43
4.4.2	Results	44
4.5	Improvements to PyGRB	47
4.6	Future Plans	47
5	Machine Learning for GW Astronomy	48
5.1	Introduction	48
5.2	Xpipeline	48
5.2.1	Burst Search Background	49
5.2.2	Standard Likelihood	52
5.2.3	Null Energy	53
5.2.4	Incoherent Energy and Background Rejection	53
5.3	Machine Learning	54
5.3.1	Supervised Machine Learning	54
5.3.2	Boosted Decision Trees	57
5.3.3	Data Preprocessing	58
5.3.4	Optimisation and Validation	61

List of Figures

- 3.1 **BATSE GRB Fluence.** This plot shows the fluence (given by the colour of each point) and the sky position of each GRB detected by the BATSE mission. cite https://heasarc.gsfc.nasa.gov/docs/cgro/images/cgro/2704_grbs_fluence.jpg 10
- 3.2 **LogN-logP for BATSEVPO.** Here we plot the log of the number of GRBs against the log of the peak flux. The sample includes GRBs detected by BATSE and by PVO. The energy range for BATSE was 50-300keV, and the energy range for PVO was 100-2000keV. For uniformly distributed GRBs, we expect this plot to have a gradient of $-3/2$. The expected gradient is observed for high energy GRBs but not at lower energies. This suggests a limited distance to which GRBs can be observed. cite https://www.researchgate.net/figure/Distribution-log-N-log-P-for-a-combined-set-of-BATSEVPO-data-The-distributions-mafig5_242389649 cite <https://www.nature.com/articles/366040a0>. pdf 11
- 3.3 **T_{90} vs count.** Here we plot the T_{90} values against the the count of GRBs for the BATSE mission as of 1993. cite **identification of two classes of gamma ray bursts what is the difference between the dotted line and the solid line-read original paper.** 12
- 3.4 **T_{90} vs the spectral hardness ratio.** Here we plot the T_{90} values against the spectral hardness ratio for the BATSE GRBs. Those GRBs with the greatest ratio of energy in the X-ray to gamma ray band, generally those with a peak energy of less than 15 keV, are called *X-ray flashes* (XRF). Those with comparable energy in the gamma-ray and X-ray band are called *X-ray rich* (XRR) GRBs. All other GRBs are simply called GRBs. These different classes of GRBs are marked on the plot. Also shown is the 2 second dividing line between short and long GRBs. cite **What are Gamma-ray Bursts?** Joshua Bloom 13
- 3.5 **BATSE gamma-ray light curves.** cite <http://inspirehep.net/record/1358907/plots> 16

3.6	GRB170817A and GW170817. Here we see that the a coherent combination of the Hanford and Livingston strain data from GW170817 in the bottom panel. The top two panels shows the Fermi GRM curve in the 10-50keV and the 50-300keV range respectively. The INTEGRAL/SPI-ACS data is shown in the third plot. The background estimate for each detector is indicated by the red line. Note that the GRB was detected 1.7 seconds after the GW signal was detected. We can also see that Fermi detected a longer, softer signal in the 10-50 keV range, that lasted for a few seconds after the triggering pulse. cite Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A	21
3.7	Sky map for GW170817/GRB170817A. Here we see the 90% confidence sky localisation for the LIGO and Virgo collaborations in green, the GBM 90% localisation in purple, and the annulus formed by Fermi and INTEGRAL timing information in grey. cite Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A	22
3.8	NGC 4993. Image of NGC 4993 taken in 1992 (left) and August 18th 2017 (right). cite https://www.nature.com/articles/nature24291/figures/2	22
4.1	Null Stat Cut. Here we plot the coherent SNR against the null SNR. The blue crosses are background triggers. The red pluses are signal injections. The black line is the veto line, with all triggers in the shaded region above the line being discarded. The green line indicates the expected SNR for optimally oriented injections. The magenta and cyan lines show 1 and 2 sigma errors on the green line.	38
4.2	P-value for each GRB. This is the p-value distribution for the 41 GRBs other than GRB170817A. The GRBs with no trigger in the onsource window have upper and lower limits on the p-value. The upper limit is a p-value of 1. The lower limit is the fraction of offsource trials that also had no trigger. The distribution lays within the 2σ range for a beta distribution, shown by the upper and lower dotted lines.	45
4.3	Cumulative exclusion distance. This is the cumulative 90% exclusion distance for every GRB analysed by PyGRB except GRB170817A. The 90% exclusion distance is the distance at which 90% of injected simulated signals are recovered with a greater coherent SNR than the loudest trigger in the onsource.	46
5.1	Xpipeline Time-Frequency Map This figure shows a time-frequency map from Xpipeline for a $1.4 - 10M_{\odot}$ NSBH merger. The top figure shows the E_+ energy and the bottom figure shows the top 1% of pixels.	49
5.2	Xpipeline Cut This figure shows Xpipeline making a cut to eliminate many spurious signals.	50
5.3	A decision tree example To determine if a trigger is a signal or noise event the tree makes a series of cuts on the attributes $x[N]$. If the inequality in a node is true, then the next node is the branch to the left. Otherwise the next node is the one to the right.	58

5.4	Visualising the Classifier	In the bottom plot you can see the value for $\log(\text{Enull})$ and $\log(\text{Inull})$ for all the signal and background training data used to build the classifier. We chose one of these events at random (indicated by the star) and varied Enull and Inull to see how it changed the MVA score, indicated by the colour in the top plot. As you can see, increasing Inull and decreasing Enull leads to the event being more likely to be classed as a signal. This is akin to the xpipeline cut shown in fig.5.2.	59
-----	-----------------------------------	--	----

List of Tables

5.1	An example of MVA training data. Each event has a label and several attributes.	55
5.2	The attributes used by the machine learning classifier.	56

...as observations have accumulated, the subject has become, to my mind at least, more mysterious and more inapproachable.

William Parsons, Third Earl of Rosse

Chapter 1

Introduction

- Usual 'start of GW astronomy' stuff
- LIGO interferometer stuff. Virgo too. Soon Kagra and LIGO India
- GRBs are good candidate sources for GWs, especially after 170817A
- Triggered vs untriggered search
- There are several different search strategies for GRB GWs. They vary in the number of assumptions that they make about the source. This thesis covers two of these searches: The modelled CBC search, which assumes an inspiral, and the unmodelled Xpipeline search, which only a maximum duration and bandwidth of the GW.
- PyGRB and X are both triggered searches
- PyGRB assumptions are restrictive enough to allow a matched-filter search
- PyGRB can be made more sensitive by assuming circularly polarised GW emission, as the beam is perpendicular to the plane of the system
- Indicate the median exclusion distance for BNS systems or something like that to show how sensitive PyGRB is
- Xpipeline is a GWB search that makes minimal assumptions
- It only uses measures of coherence based on sky location
- O2 median distances
- Improved this search using machine learning
- Describe format of thesis

Chapter 2

Gravitational Wave Astronomy in a Nutshell

2.1 General Relativity

Christoffel Symbol

$$\Gamma_{\mu\nu}^{\lambda} = \frac{1}{2}g^{\lambda\rho}[\partial_{\nu}g_{\mu\rho} + \partial_{\mu}g_{\nu\rho} - \partial_{\rho}g_{\mu\nu}] \quad (2.1)$$

Riemann Curvature Tensor

$$R_{\lambda\alpha\beta}^{\mu} = \partial_{\alpha}\Gamma_{\lambda\beta}^{\mu} - \partial_{\beta}\Gamma_{\lambda\alpha}^{\mu} + \Gamma_{\nu\alpha}^{\mu}\Gamma_{\lambda\beta}^{\nu} - \Gamma_{\nu\beta}^{\mu}\Gamma_{\lambda\alpha}^{\nu} \quad (2.2)$$

Ricci Tensor

$$R_{\mu\nu} = g^{\alpha\beta}R_{\alpha\mu\beta\nu} = R_{\mu\beta\nu}^{\beta} \quad (2.3)$$

Ricci Scalar

$$R = g^{\alpha\beta}R_{\alpha\beta} = R_{\beta}^{\beta} \quad (2.4)$$

The Einstein Equations

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} \quad (2.5)$$

Alternative form of the Einstein equations

$$R_{\mu\nu} = \frac{8\pi G}{c^4} \left(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu} \right) \quad (2.6)$$

2.2 Gravitational waves

General Relativity shows that spacetime can curve and move. From this, it may seem obvious that waves can travel through spacetime. But this is not a trivial fact; Einstein himself, having introduced the concept of gravitational waves in 1916, later claimed they could not exist. In this section, we will show that gravitational

waves do exist. We will do this by considering a perturbation of the flat Minkowski metric. We will then calculate the Ricci Tensor and Ricci Scalar for the perturbed metric and, with a clever choice of Gauge transformation, see that these yield a wave solution to Einstein's Equations.

2.2.1 Linearised Gravity

Let $g_{\mu\nu}$ be the metric given by adding a small perturbation $h_{\mu\nu}$ to the Minkowski metric $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$. We write this metric as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1. \quad (2.7)$$

To first order in h , we calculate the Ricci tensor (2.3)

$$R_{\mu\nu} = \eta^{\alpha\beta} R_{\alpha\mu\beta\nu} = \frac{1}{2} (\partial_\mu \partial^\alpha h_{\alpha\nu} + \partial_\nu \partial_\alpha h_\mu^\alpha - \square h_{\mu\nu} - \partial_\mu \partial_\nu h) \quad (2.8)$$

and the Ricci scalar (2.4)

$$R = \eta^{\mu\nu} R_{\mu\nu} = \partial_\mu \partial_\alpha h^{\mu\alpha} - \square h, \quad (2.9)$$

where $\square = \partial_\mu \partial^\mu$ is the d'Alembertian operator and $h = h_\mu^\mu$ is the trace of h . From these, we find the resulting Einstein tensor is also linear in h

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}. \quad (2.10)$$

2.2.2 Gauge Transformation

With the following Gauge transformation, we can simplify the expression for the linear Einstein tensor

$$x^{\mu'} = x^\mu + \chi^\mu(x), \quad \chi \ll x. \quad (2.11)$$

As χ is small, we have $\partial_\mu \chi^\nu \ll 1$. This gives us

$$\frac{\partial x^\mu}{\partial x^{\alpha'}} = \delta_\alpha^\mu - \partial_\alpha \chi^\mu + \mathcal{O}(|\partial\chi|^2). \quad (2.12)$$

Applying these results to our metric (2.7) we find

$$g_{\alpha'\beta'} = \frac{\partial x^\mu}{\partial x^{\alpha'}} \frac{\partial x^\nu}{\partial x^{\beta'}} g_{\mu\nu} = g_{\alpha\beta} - \partial_\alpha \chi_\beta - \partial_\beta \chi_\alpha. \quad (2.13)$$

Subtracting the Minkowski metric from each side, we find

$$h_{\alpha'\beta'} = h_{\alpha\beta} - \partial_\alpha \chi_\beta - \partial_\beta \chi_\alpha. \quad (2.14)$$

We have some freedom in choosing our χ , so we impose the *harmonic gauge*

condition

$$\partial_\mu h_\nu^\mu = \frac{1}{2} \partial_\nu h, \quad (2.15)$$

where $h = h_\lambda^\lambda$. We can always choose a χ such that this is true. To see this, first note that $\partial'_\mu = \partial_\mu - (\partial_\mu \chi^\lambda) \partial_\lambda$. From this we find

$$(\partial'_\mu h_\nu^\mu - \frac{1}{2} \partial'_\nu h) \approx (\partial_\mu h_\nu^\mu - \frac{1}{2} \partial_\nu h) - \partial^2 \chi_\nu. \quad (2.16)$$

Thus, if we are given an h such that 2.15 is not true, we can choose a χ such that

$$\partial^2 \chi_\nu = (\partial_\mu h_\nu^\mu - \frac{1}{2} \partial_\nu h). \quad (2.17)$$

Using 2.15, we can simplify the Ricci tensor and scalar

$$R_{\mu\nu} = -\frac{1}{2} \square h_{\mu\nu} \quad (2.18)$$

$$R = -\frac{1}{2} \square h. \quad (2.19)$$

Using these in the linearised Einstein equations (2.10) gives us

$$\square h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} \partial^2 h = -\frac{16\pi G}{c^4} T_{\mu\nu}. \quad (2.20)$$

Alternatively, we can use 2.6 to write this as

$$\square h_{\mu\nu} = -16\pi G \left(T_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} T \right). \quad (2.21)$$

In a vacuum, the right hand side of this equation becomes zero, and we can recognise as the relativistic wave equation.

2.2.3 Physical Effects of Gravitational Waves

The plane wave solution for the vacuum wave equation is

$$h_{\mu\nu}(x) = \epsilon_{\mu\nu} e^{ik_\alpha x^\alpha} \quad (2.22)$$

where $\epsilon_{\mu\nu}$, the polarisation tensor for the gravitational wave, is symmetric and constant, and k^α is the 4-wave-vector given by $k^\alpha = (\omega, \vec{k})$. Substituting this into the vacuum wave equation, we find

$$k^2 = -\omega^2 + \vec{k}^2 = 0. \quad (2.23)$$

Hence, gravitational waves travel at the speed of light.

The polarisation tensor is not arbitrary: It must satisfy the wave equation and the harmonic gauge condition. Putting the wave solution 2.22 into the harmonic

gauge condition 2.15, we find that gravitational waves are transverse

$$k^\mu \epsilon_{\mu\nu} = 0 . \quad (2.24)$$

We can impose further gauge conditions as long as the harmonic gauge (and hence 2.24) is not violated. As any transformation with $\partial^2 \chi = 0$ will satisfy 2.17, and hence harmonic gauge condition, we express χ as

$$\chi_\nu = X_\nu e^{ikx} . \quad (2.25)$$

Using 2.25 and 2.22 in the transformation equation 2.14, we find the transformation equation for the polarisation tensor

$$\epsilon'_{\mu\nu} = \epsilon_{\mu\nu} - ik_\mu X_\nu - ik_\nu X_\mu . \quad (2.26)$$

Taking the trace of this, we find

$$\epsilon'^\mu{}_\mu = \epsilon^\mu{}_\mu - 2ik^\mu X_\mu . \quad (2.27)$$

Thus, we can impose the further gauge condition that the polarisation matrix be traceless by choosing coordinates such that $\epsilon^\mu{}_\mu = 2ik^\mu X_\mu$, and so fixing one element of X^μ . We can fix the other elements of X_μ by setting $\epsilon_{i0} = 0$ for $i = 1, 2, 3$. We do this using 2.26 to find

$$\epsilon'_{i0} = \epsilon_{i0} - ik_i X_0 - ik_0 X_i . \quad (2.28)$$

Now we see that by setting $\epsilon_{i0} = ik_i X_0 - ik_0 X_i$, we have $\epsilon_{i0} = 0$. As the polarisation tensor is symmetric, we have $\epsilon_i^0 = 0$ as well.

These conditions give us that $\epsilon_{00} = 0$ as well. Using the fact that the polarisation matrix can be assumed traceless $\epsilon^\mu{}_\mu$, the wave solution 2.22 reduces the harmonic gauge condition 2.15 to

$$\partial_\mu h^\mu{}_\nu = 0 . \quad (2.29)$$

As $\epsilon_{i0} = 0$, in the $\nu = 0$ case, only the $\mu = 0$ term is non-zero. Thus we are left with

$$\partial_\mu h_0^\mu = \epsilon_0^0 ik_0 e^{ik_\alpha x^\alpha} = 0 . \quad (2.30)$$

Which implies $\epsilon_0^0 = 0$. As the polarisation tensor is symmetric, we have $\epsilon_3^3 = \epsilon_3^i = 0$ as well.

We can make the polarisation tensor simpler by assuming the wave is traveling in the z-direction, that is $\vec{k} = (0, 0, \omega)$. The transverse condition 2.24 then forces all the z-components of the polarisation matrix to be zero, $\epsilon_i^3 = \epsilon_3^i = 0$. The wave

solution 2.22 now looks like

$$h_{\mu\nu}(x) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & a & b & 0 \\ 0 & b & -a & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} e^{i\omega(z-t)} .$$

2.3 Gravitational Waves Sources

- GW generation
- Sources
 - BBH - Multiple found already
 - BNS - One found and Hulse-Taylor
 - NSBH
 - GRB
 - SN
 - Unknown

2.4 Gravitational Wave Detectors

Can cite soaulson and the noises paper

- Michelson Interferometer (cite R.Weiss 1972 paper to look smart)
- Measure change in output power due to interference in the arms
- Noise sources and how they are mitigated
 - Fundamental noise
 - Environmental noise
 - Instrumental noise
 - Noise budget

2.4.1 How to find something interesting in all that noise

- Autocorrelation function
- PSD and ASD (tie back to the noise budget)
- Matched filtering (for 1 detector)
- Unmodelled search
- Network statistics (coincident and coherent)

Chapter 3

Gamma-Ray Bursts

Gamma-ray Bursts (GRBs) are exceptionally energetic flashes of gamma rays. They can last from just a few milliseconds up to several hours and have highly variable luminosity curves (see figure 3.5). They are detected at a rate of about one per day, are uniformly distributed over the sky, and are the most electromagnetically energetic objects in the universe. The short duration and huge energy emitted by GRBs suggests a violent origin, making them strong candidates for gravitational wave (GW) emission.

In this chapter we discuss GRB astrophysics. We will begin with some historical perspective, starting with the accidental first detection of a GRB in 1963 and continuing to the first GW detection associated with a GRB in 2017. We will see that GRBs can be classified as short-hard or long-soft, depending on their duration and spectral hardness. We will see the evidence that long GRBs are caused by core collapse supernova, and that short GRBs are caused by compact binary mergers involving at least one neutron star. We will discuss the physical processes that could be powering GRBs and what GW astronomy can teach us about these processes. We end this chapter with a discussion of current GRB detectors and search strategies for GW emission associated with GRBs.

3.1 The History of Gamma-ray Burst Astronomy

In this section we will discuss the key discoveries of GRB astronomy in their historical context. These discoveries are the motivation of GW searches associated with GRBs, and were used to develop the GW searches discussed in chapters 4 and 5.

3.1.1 Cold War Tension and an Unexpected Discovery

The partial nuclear test ban treaty, agreed between the USA and the Soviet Union in 1963, banned atmospheric, underwater, and outer space nuclear weapon tests. This created a technical challenge: how to enforce the ban? Seismic sensors could be used for on-Earth tests, but would not work for outer space tests. The solution was to look

for the flash of gamma-rays produced in the first milliseconds of a nuclear explosion. Thus the Vela and Kosmos gamma ray detecting satellites were produced by the USA and Soviet Union respectively. These satellites contained only rudimentary gamma ray detectors, and each individual satellite was not capable of localisation. Some localisation was possible using time delay and Earth blocking information.

These satellites started to detect brief bursts of gamma rays, which were first reported in 1973. These events did not look like those expected from a nuclear test, and did not seem to be coming from the Earth or any nearby astronomical bodies such as the moon. It appeared a new, high-energy astronomical phenomenon had been discovered.

This phenomena, called *Gamma-ray Bursts* (GRBs), would appear and fade away in as little as a milliseconds, and could be as brighter than the rest of the gamma ray sky combined. The brevity of these events placed constraints on the size of the source, as the crossing time for a region cannot be less than the light travel time. Thus a 1ms GRB must have a source smaller than 300km across. This limits the potential candidates down to compact objects, such as neutron stars and black holes, or to small regions of larger objects, such as the cores of massive stars. Another important feature of GRBs is that the bursts do not repeat. This suggests a source that is destroyed when the GRB is produced. More measurements were needed to narrow down the number of possible sources for GRBs.

3.1.2 BATSE and the Galactic/Extra-galactic Controversy

Most early GRB detectors could not localise well. It was known that GRBs were not coming from any planets in the solar system or from the galactic center, but otherwise the location of GRB sources was a mystery. In particular, it was not clear whether GRBs were coming from galactic or extra-galactic sources. Answering this question was an important step towards identifying the origin of GRBs, as an extra-galactic source would require far greater energy than a galactic source to produce a GRB of equal brightness.

Better sky localisation would help to answer this question. If GRBs mostly occur on the galactic plane, then they are probably of galactic origin. If they were clustered around nearby galaxies, then GRBs probably come from those galaxies. If, however, they are distributed isotropically on the sky, then it would be likely that GRBs form at cosmological distances.¹

Improved sky localisation was achieved in 1991, with the launch of the Burst and Transient Source Experiment (BATSE) on the Compton Gamma-ray Observatory. With BATSE, it became possible to determine the sky position of a GRB to within

¹It should be noted that there are other ways that GRBs could be isotropically distributed. If they are only detectable to a few hundred parsecs, they would be entirely within the disc of the galactic plane and so would also be isotropically distributed. Alternatively, as neutron stars receive a ‘kick’ during their formation, they could form a corona around the galaxy.

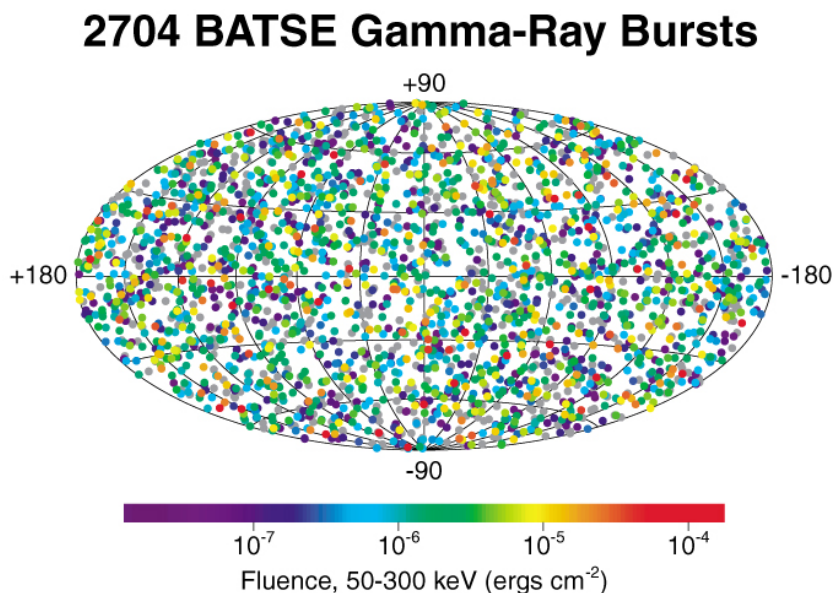


Figure 3.1: **BATSE GRB Fluence.** This plot shows the fluence (given by the colour of each point) and the sky position of each GRB detected by the BATSE mission. cite https://heasarc.gsfc.nasa.gov/docs/cgro/images/cgro/2704_grbs_fluence.jpg

half a degree. During its mission, BATSE detected approximately 2700 GRBs and determined the sky position of a large number of these. For the first time there was a statistically significant sample of well localised GRBs. In figure 3.1 we can see the sky location of every GRB detected by BATSE, together with the fluence² of that GRB. This plot shows that GRBs are isotropically distributed over the sky. This evidence strongly suggests GRBs are of cosmological origin.

The BATSE data also provided evidence that GRBs were uniformly distributed and that we were seeing a limited horizon, beyond which GRBs became much harder to detect. This evidence came in the form of a $\log N - \log P$ distribution, where N is the number of detected GRBs and P is the peak flux³. If GRBs are uniformly distributed in space then the number of GRBs out to a given distance increases as the cube of that distance. However, peak flux from the GRBs would decrease with the inverse square of the distance. Hence, plotting $\log N$ against $\log P$, we expect to find a gradient of approximately $-3/2$. Any short fall from this expected distribution suggests that we have reached a horizon for detectable GRBs. In figure 3.2 we see the $\log N - \log P$ distribution for a combined set of GRBs detected by BATSE and the Pioneer Venus Orbiter (PVO).⁴ The gradient for high energy GRBs

²Fluence is the time integral of the flux. Essentially a measure of the total detected energy.

³Some times the $\log N - \log S$ distribution is preferred, with S being the given flux. The essence of the plots is the same.

⁴ The PVO was less sensitive than BATSE, but it operated for 10 years and so observed a fairly large number of GRBs

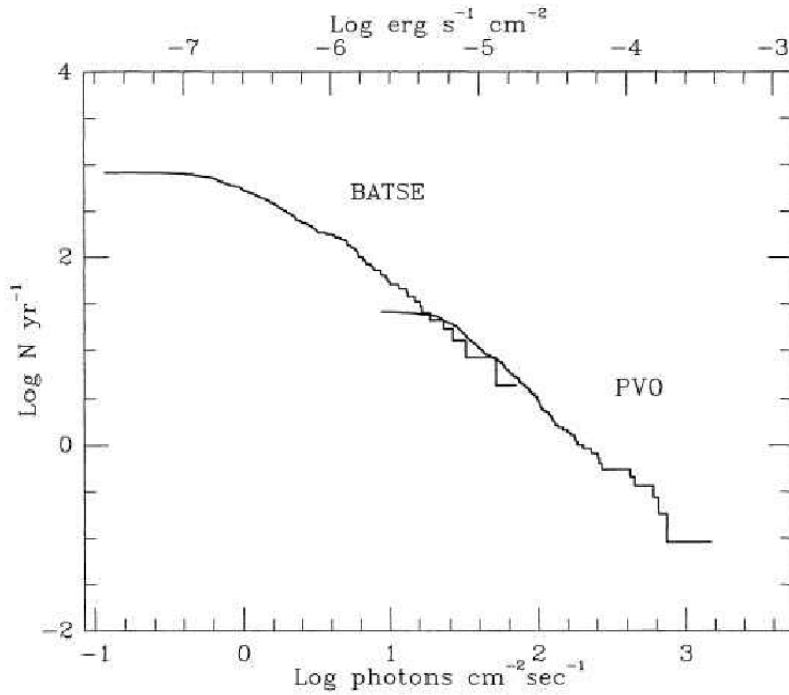


Figure 3.2: **LogN-logP for BATSE/PVO.** Here we plot the log of the number of GRBs against the log of the peak flux. The sample includes GRBs detected by BATSE and by PVO. The energy range for BATSE was 50-300keV, and the energy range for PVO was 100-2000keV. For uniformly distributed GRBs, we expect this plot to have a gradient of $-3/2$. The expected gradient is observed for high energy GRBs but not at lower energies. This suggests a limited distance to which GRBs can be observed. [cite https://www.researchgate.net/figure/Distribution-log-N-log-P-for-a-combined-set-of-BATSEPVO-data-The-distributions-match_fig5_242389649](https://www.researchgate.net/figure/Distribution-log-N-log-P-for-a-combined-set-of-BATSEPVO-data-The-distributions-match_fig5_242389649) [cite https://www.nature.com/articles/366040a0.pdf](https://www.nature.com/articles/366040a0.pdf)

is the expected $-3/2$ but there is a shortfall at low energies, suggesting there is a limited distance to which GRBs can be viewed. Unfortunately this data does not inform us of where that horizon is. It could be within our galaxy, or it could be the horizon of the universe.

3.1.3 The Long and Short of Gamma-ray Bursts

Another important piece of evidence into the origins of GRBs came from their duration. As every burst has different properties⁵ the duration of a GRB is not trivially defined. The measure most commonly used is the T_{90} , the time over which 90% of the total fluence is recorded.⁶

Figure 3.3 shows the number of bursts with a given T_{90} for the BATSE data. This plot shows that there are two populations of GRBs, the first with a T_{90} value

⁵For example, some bursts have multiple flares.

⁶Another common measure is the similarly defined T_{50} .

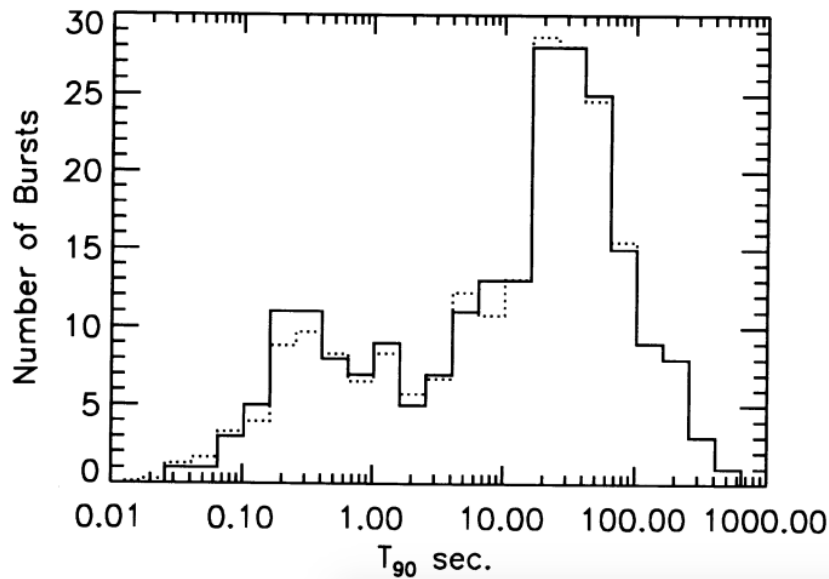


Figure 3.3: **T90 vs count.** Here we plot the T_{90} values against the the count of GRBs for the BATSE mission as of 1993. **cite identification of two classes of gamma ray bursts what is the difference between the dotted line and the solid line-read original paper.**

of about 0.5s and the second with a T_{90} of about 30s. It is also clear from this data that the longer population of GRBs are detected far more often. Plotting the spectral hardness of the BATSE GRBs against T_{90} , as has been done in figure 3.4, we see that the shorter GRBs also have a harder spectrum than the longer GRBs. This means short GRBs emit more high energy photons than long GRBs. Because of these two properties of the two populations, they are known as *short-hard* and *long-soft* GRBs. It is common to use the criteria that short GRBs are those that are less than 2s, and long GRBs are longer than 4s, with those inbetween being called *intermediate* GRBs.

It should be mentioned that while T_{90} is a very useful tool, it is instrument dependent. This is because more sensitive instruments will track GRBs for longer, and bursts have different durations in different energy bands. Also, the T_{90} is measured in the detector frame, and not the rest frame of the burst, which would make a GRB at a redshift of z appear a factor of $(1 + z)$ longer. For these reasons, using the T_{90} to classify short/long GRBs should only be considered approximate.

As a final note on GRB durations, it should be mentioned that there is some evidence that there may be intermediate and ultra-long populations of GRBs. It is an open question whether these GRBs represent new populations of GRBs or are part of the long and short GRB populations. As these are still contentious, we do not consider them further.

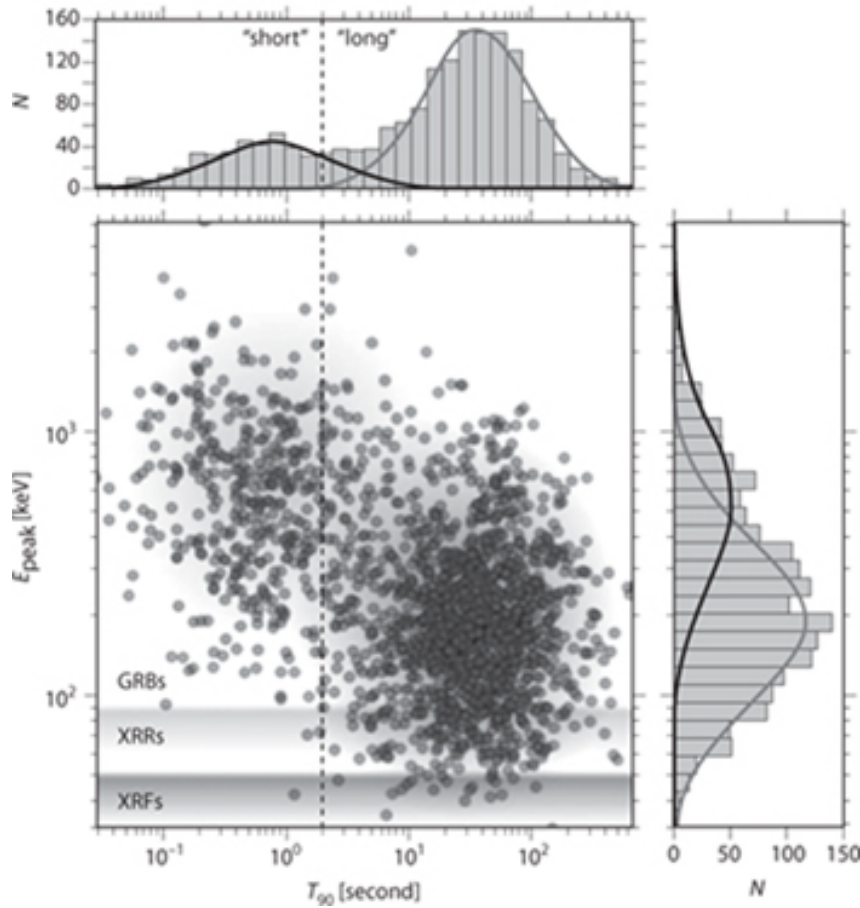


Figure 3.4: **T₉₀ vs the spectral hardness ratio.** Here we plot the T_{90} values against the spectral hardness ratio for the BATSE GRBs. Those GRBs with the greatest ratio of energy in the X-ray to gamma ray band, generally those with a peak energy of less than 15 keV, are called *X-ray flashes* (XRF). Those with comparable energy in the gamma-ray and X-ray band are called *X-ray rich* (XRR) GRBs. All other GRBs are simply called GRBs. These different classes of GRBs are marked on the plot. Also shown is the 2 second dividing line between short and long GRBs.

cite **What are Gamma-ray Bursts?** Joshua Bloom

3.1.4 BeppoSAX and the First Afterglows

Important clues had been found into the origin of GRBs, but still no one had found any trace of a GRB after the prompt emission (the initial flash of gamma-rays). The search area provided by GRB detectors at the time were too large for ground based telescopes to have a realistic chance of finding the source of the GRB, though attempts were made. This changed with the launch of the BeppoSAX satellite in 1996. BeppoSAX was able to localise to within a few arcminutes, much better than the half a degree BATSE was capable of.⁷ It also had a Narrow-Field x-ray Instrument (NFI) to search for X-ray counterparts to GRBs.

On February 28th 1997 a GRB was detected by BeppoSAX that was localised well enough for the satellite to aim the NFI at it. A new X-ray source was detected that faded away slowly over the next few days. The lower energy emission detectable in the days after the prompt emission is called the GRB *afterglow*, and this was what BeppoSAX had detected. The localisation was accurate enough to allow ground based telescopes to find the optical counterpart to this GRB as well. The optical images showed what some argued was a distant galaxy and others argued was a galactic nebula. On May 8th 1997, a GRB was detected by BeppoSAX and an optical counterpart quickly found. A spectrum was obtained for the counterpart which showed iron and magnesium absorption lines that had been significantly red-shifted. This showed that the GRB must have occurred at a distance greater than 5 Gpc and that the light passed through some gaseous cloud in a distant galaxy on its way to Earth. With this observation there could be no doubt that GRBs were originating at cosmological distances, as had been suspected.

3.1.5 The Fireball Model

The *fireball shock model* was developed in the 1990s. It attempted to explain some of the features of GRBs by making minimal assumptions about the *central engine*, the source of energy for the GRB. At this time the evidence was building that GRBs originated at cosmological distances. At such great distances the inferred energy emitted at the source is far higher, as much as a solar mass if emitted isotropically. Light-travel time arguments showed that GRB sources must also be small, at most hundreds of kilometers across (see section 3.1.1). Realising that all GRBs are small and highly energetic was the starting point of the fireball shock model.

The next step was to notice that photons emitted by GRBs are above the pair production threshold ($2 \times m_e c^2 \approx 1\text{MeV}$), and so should have created electron-positron pairs rather than gamma-rays. This problem is solved by assuming that the energy released drives a relativistic expansion from the source. In this case

⁷It is true that the InterPlanetary Network, a network of GRB detectors placed on various spacecraft throughout the solar system (see section 3.3), could localise better than BATSE. The problem was that it took too long find the sky position of the GRB using this method, and any trace of the GRB had disappeared before astronomers could find it.

the energy of the photons in their rest frame is inversely proportional to the bulk Lorentz factor Γ of the outflow, i.e. the energy in the rest frame of the photons is much lower. It also causes photons to bunch up just ahead of the relativistic matter that is emitting the photons, causing it to seem more energetic to an observer. Accounting for these factors, it can be shown that the Lorentz factor must be in the hundreds for a typical GRB. This is as compared to a Lorentz factor of approximately 1.001 for a fast supernovae, and corresponds to GRB velocities of at least 99.9% of the speed of light.

This means that a large amount of energy must be rapidly dumped into a small volume before the GRB is emitted. The energy density is so high that gamma-rays form electron-positron pairs which then annihilate to again form high energy gamma-rays. Most of the energy is in the form of photons, so this is called the *radiation dominated* phase. This mess of photons and some matter is the *fireball*. The fireball expands and the Lorentz factor grows with it. As it expands, the energy is absorbed by the matter (protons and electrons) in the form of kinetic energy. This is the *matter dominated* phase. Most of the particles are moving in the same direction now, with little random motion: The fireball has become cold.

A mechanism is needed to turn this kinetic energy into a GRB. The simplest way to do this is to have the matter collide with slower moving matter surrounding the system. The matter is too sparse for direct collisions to happen enough to form a GRB. Instead, it is thought that magnetic fields near the edge of the fireball can cause the matter to slow down and radiate its energy. When the fireball interacts with the surrounding matter, the rapid change in temperature, pressure, and density travel through the medium faster than the medium can react. Like the sonic boom of a supersonic jet, this causes *shocks* in the surrounding medium. It is these shocks that are thought to be the source of the GRB. Precisely how these shocks power a GRB is not known. *Fermi acceleration* possibly plays a role. This is where charged electrons enter the shock and are reflected back by magnetic fields, increasing their kinetic energy. After several iterations of this, the electron can travel even faster than the shock. The magnetic fields could then cause the electrons path to curve, causing them to emit energy as synchrotron radiation. Alternatively they might interact with photons, imparting their considerable energy onto the photon to create a high energy gamma-ray. This process is called *inverse Compton scattering*. There is still a lot to learn about the processes that power the GRB.

There are also two hypotheses as to what the slower moving matter could be. The first is simply material around the star, the *circumburst medium* (CBM). This is called an *external shock scenario*. The other theory is that multiple shells of material are emitted, and the shocks are created when faster moving shells catch up with slower moving shells. This is called the *internal shock scenario*. The internal shock method is generally favoured for the prompt emission. This is because GRB light curves are highly variable, with some showing multiple peaks (see figure 3.5).

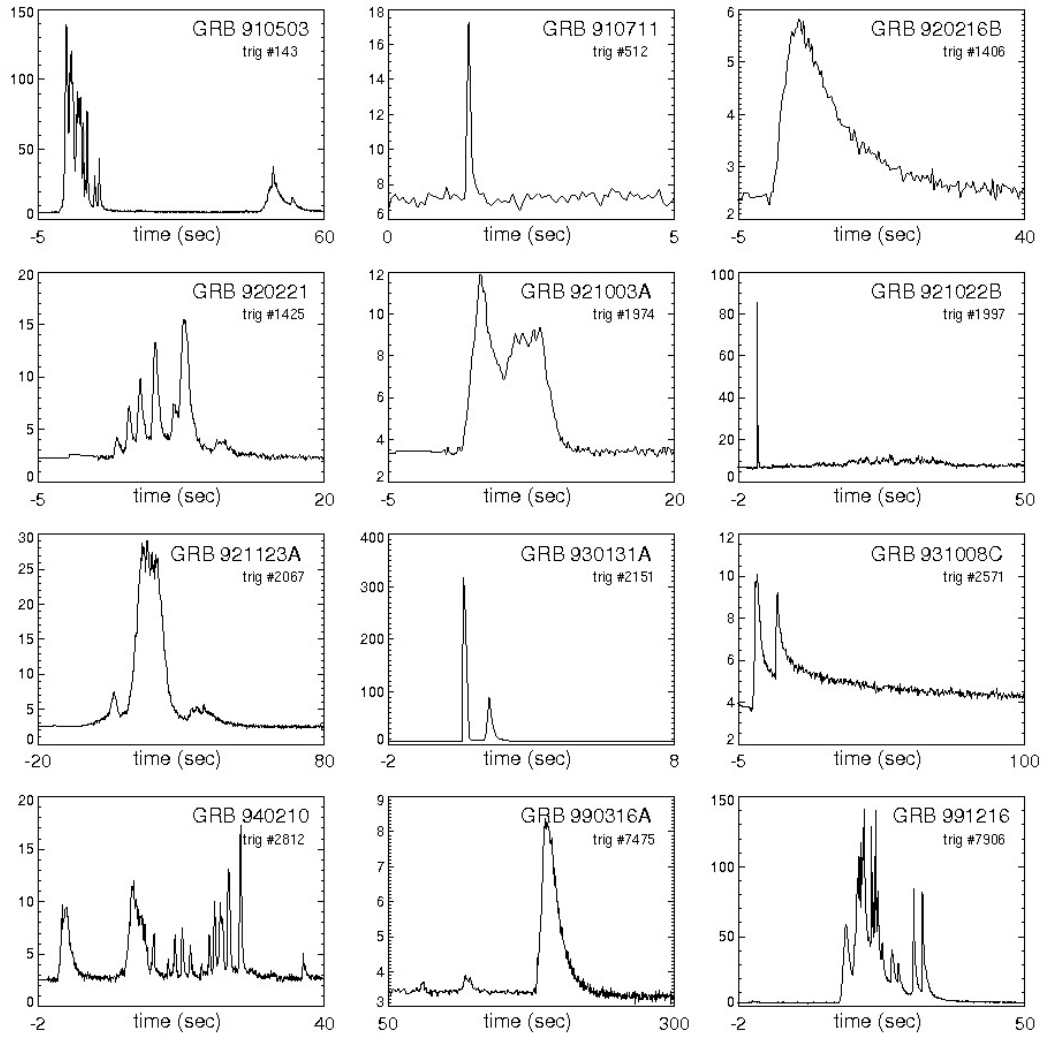


Figure 3.5: BATSE gamma-ray light curves. cite <http://inspirehep.net/record/1358907/plots>

This is easily explained by the multiple shocks of the internal shock scenario, but more difficult to explain for external shocks. External shocks are thought to be the cause of the GRB afterglow.

- Derive equation for 'bunching' of photons in front of relativistic matter
- Amount of energy and distance implies beaming (caused by relativistic effects)
- Inverse Compton scattering and other things to get energy

3.1.6 The Long GRB-Supernova Connection

As more afterglows were found, trends began to appear. Long GRBs tended to occur directly on host galaxies, not randomly in space.⁸ Spectroscopy of these galaxies often found the presence of emission lines excited by star formation. No long GRB has been found in a non-starforming galaxy. The host galaxies also tended to be relatively faint. Low luminosity (i.e. low stellar mass) suggests galaxies have a low metallicity [**cite someone**], and spectrographic studies confirmed this. That long GRBs originate from star forming galaxies with low metallicity raises the possibility that long GRBs are connected to the massive stars that can only exist in these kinds of galaxies. In particular, massive stars end their lives with violent supernova explosions, which could potentially power a GRB.

The strongest evidence of a supernova connection came with the detection of GRB 980425. This GRB was exceptionally close; at just $z = 0.0085$ it is still the closest GRB ever detected. Followup observations found a rising supernova, SN 1998bw. It was a very bright supernova, about ten times brighter than normal. It also showed no hydrogen or helium emission lines, making this a type Ic supernova. It seemed highly unlikely that the coincident GRB and supernova were unrelated. Questions remained though, because this GRB was exceptionally faint given its distance. Was it typical of other long GRBs?

After GRB 980425, searches were undertaken for other bright supernova afterglows associated with long GRBs. Not only were many found, but they were shown to have similar spectra to GRB 980425. In particular, GRB 030329 was another nearby GRB, at redshift $z = 0.17$, which had very detailed followup. It was 1000 times brighter than GRB 980425, but showed the same spectral features. With this discovery, the consensus grew that long GRBs were caused by type Ic supernova.

- Go into more detail about what a type 1c supernova is and its characteristics
- From this a consensus built that long GRBs were caused by broad lined (i.e. high velocity) type 1c SNe. **check understanding of broad lined**

3.1.7 The Short GRB-Compact Binary Connection

Much had been learned about long GRBs by studying their afterglows, but no afterglow had been found for short GRBs. The problem was that short GRBs faded rapidly, reducing the amount of time for astronomers to find the afterglow. For this reason, the Swift satellite (see section 3.3) was launched in 2004. Swift was able to autonomously followup a GRB with X-ray, UV, and optical measurements within minutes, just what was needed for short GRBs.

⁸We will see in section 3.1.7 that this is unlike short GRBs. which can originate quite far from any host galaxy.

The first detected short GRB afterglow was that of GRB 050509B. It was a faint X-ray afterglow that faded quickly, but it was localised well enough for ground based telescopes to determine the host galaxy, which had a redshift of $z = 0.225$. The afterglow and host galaxy were very different to those of long GRBs: the host galaxy was a massive elliptical galaxy which showed no evidence of star forming, there was also no evidence of a supernova, and the GRB took place far away from the galactic core. With more short GRB afterglow detections it became clear that, on average, the detected short GRBs occur much closer than long GRBs⁹, though this could partly be due to selection effects.¹⁰ Also short GRBs occur further from the center of their host galaxy than long GRBs, with an average offset of 4.5kpc and 1.5kpc respectively. Sometimes they are so far from any galaxy that determining the host is impossible; these are called *hostless* GRBs.

All of these observations are consistent with the theory that short GRBs are caused by the merger of a neutron star with either another neutron star or a black hole. The two objects would form in a binary system and slowly inspiral due to the loss of energy through gravitational wave emission, until they finally merge and emit the GRB. For neutron star - black hole (NS-BH) mergers, the black hole must be relatively low mass (less than $\sim 10M_{\odot}$)**[cite paper]**. This is because if the black hole has a high mass then the neutron star will be swallowed whole by the black hole, and there will be no material to produce a GRB. For a low mass black hole the neutron star will be tidally disrupted before merger, i.e. the black hole will pull apart the neutron star, providing matter that can then power a GRB. The merger time for a binary scales with a^4 , where a is the initial separation. This means a small change in the initial separation can lead to a very different merger time.¹¹ This explains their presence in elliptical galaxies that have long since stopped star forming.

It is also expected that these binary systems would receive a *kick* in their formation, which explains why so many short GRBs occur far from the center of their host galaxy. There appear to be two mechanisms through which the binary could receive a kick. The first is due to the supernovas that formed each component of the binary. As the supernova will cause a large amount of matter to become unbounded from the system at a particular point in the orbit, conservation of momentum forces the binary system to recoil. This cannot be the only mechanism that causes a kick to neutron stars though, as pulsar observations show that even lone neutron stars receive kicks from the supernova that forms them, though the mechanism that causes this kick is not known.

The binary inspiral theory also explains why short GRBs are short. The duration

⁹The average redshift of a short GRB is about $z = 0.5$, while less than 10% of long GRBs have a redshift less than one. **[cite <https://arxiv.org/abs/1702.03338>]**

¹⁰Short GRBs tend to be dimmer and so not observable at greater distances. Also, GRBs at high redshifts will appear longer in the detector frame.

¹¹In fact, some galactic binaries have merger times that exceed the age of the universe.

of a GRB depends on how long it takes the gamma-rays to break through any surrounding material and how long the central engine remains active. For a core collapse supernova there can be a lot of material around the central engine, and the amount of time the central engine is active could also be highly variable. This large amount of variability explains why long GRBs can last from seconds to hours. A binary inspiral is expected to clear out the surrounding system of any matter long before merger. The only source of matter around a binary merger is from the stars themselves. And as the GRB emitted at or immediately after the merger of the compact objects (see section 3.2), the amount of time the central engine is active is not very variable. Putting this together we conclude that binary mergers cannot produce long GRBs, as the central engine is not active for very long and there is not a lot of matter for the GRB to break through.

The observation of supernova remnants coincident with long GRBs demonstrated the long GRB - supernova connection. One might expect to find an analogous remnant for short GRBs. But what would the remnant of a neutron star merger look like? Would it even be observable? It is useful to first consider the circumstances and processes that cause supernovas, and then contrast them with those of a neutron star merger. Supernova ejecta emits light due to the radioactive decay of *s-process* elements. The s-process refers to when an atom captures a neutron, which makes the atom unstable, and then β -decays¹² to a heavier element. This is how the iron group elements are produced and it is also what makes supernovas radiate light.

The s-process occurs only if the unstable atom decays before it captures another neutron. This is unlikely in neutron rich, high temperature environments, such as a neutron star merger. Instead the atoms will capture multiple neutrons before they β -decay. This is called the *r-process* and it creates elements with a high atomic number, such as gold and iodine. The light emitted by the decay of the r-process elements after a neutron star merger is called a *kilonova*. The r-process elements are optically opaque, making kilonovas peak in the infrared (IR) and be about 10-100 times fainter than supernova.

It is difficult to detect IR using ground based telescopes, due to atmospheric distortion and a bright sky. Despite this difficulty, a kilonova was eventually observed in conjunction with GRB 130603B, a short GRB. The Hubble space telescope observed the event fade over 10 days until it was undetectable in optical light, but still clearly visible in IR. The IR light was brighter than would be expected by simply extrapolating from the afterglow. This additional IR component to the light was the kilonova. This was the strongest evidence yet that neutron star mergers were the progenitors of short GRBs.

¹² β -decay is when a neutron decays into a proton, electron, and an anti-neutrino.

3.1.8 Gravitational Waves

As discussed in chapter 2, compact binary coalescences (CBCs) are known to be strong emitters of GWs. This made short GRBs a promising target for GW astronomy, as a GW signal of a neutron star binary inspiral detected in coincidence with a short GRB would be the smoking-gun that neutron star mergers can produce short GRBs.

On the 17th of August 2017, the Fermi GRB detector (see section 3.3) observed a short GRB just two seconds after the LIGO and Virgo observatories detected a CBC GW signal (see figure 3.6). Comparing the sky localisation of the GW detector network and the Fermi GRB detector, a smaller sky patch was formed that must contain the GRB. Followup by ground based observatories quickly identified a bright object near the galaxy NGC 4993 [cite detection papers]. This object had not been there when the same galaxy had been previously observed (see figure 3.8), and this galaxy was at the right distance as determined from the GW signal. Further study revealed the new object to be a rapidly evolving kilonova, confirming the theory that short GRBs are caused by neutron star mergers.

- Should I go into more detail about what was learned from the BNS?
- Structured jet?
- blue to red spectrum?
- Heavy elements?

3.2 GRB Progenitors

In section 3.1 we discussed the evidence for short GRBs being produced by neutron star mergers and long GRBs being produced by core collapse supernova. In both of these cases, a large amount of energy is displaced rapidly from a small region. In section 3.1.5 we discussed the fireball model, which shows that these circumstances cause shocks and relativistic beaming which explain the high energy and variability of GRBs. But the fireball model only assumes that the *central engine*, the source of energy for the GRB, is small and highly energetic. There are many phenomena that satisfy this criteria. In this section we discuss some of these possibilities. We will see under which circumstances neutron star mergers and core collapse supernovae can power short and long GRBs respectively.

3.2.1 Compact Binary Coalescence

As mentioned previously, short-hard GRBs are thought to be powered by the CBC involving at least one neutron star, with a neutron star or black hole as the other

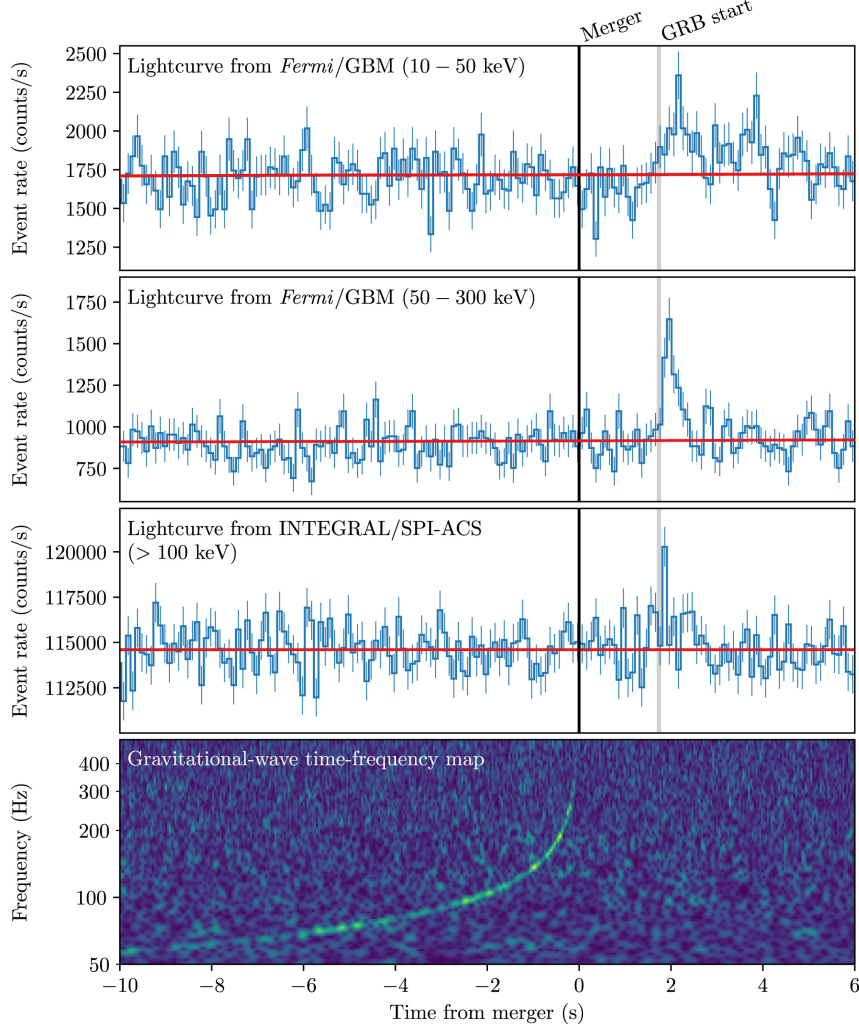


Figure 3.6: **GRB170817A and GW170817**. Here we see that the a coherent combination of the Hanford and Livingston strain data from GW170817 in the bottom panel. The top two panels shows the Fermi GRM curve in the 10-50keV and the 50-300keV range respectively. The INTEGRAL/SPI-ACS data is shown in the third plot. The background estimate for each detector is indicated by the red line. Note that the GRB was detected 1.7 seconds after the GW signal was detected. We can also see that Fermi detected a longer, softer signal in the 10-50 keV range, that lasted for a few seconds after the triggering pulse. **cite Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A**

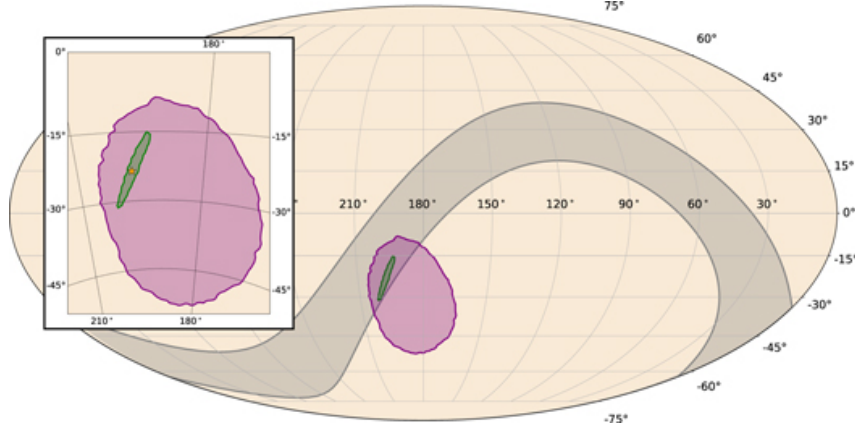


Figure 3.7: **Sky map for GW170817/GRB170817A.** Here we see the 90% confidence sky localisation for the LIGO and Virgo collaborations in green, the GBM 90% localisation in purple, and the annulus formed by Fermi and INTEGRAL timing information in grey. **cite Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A**

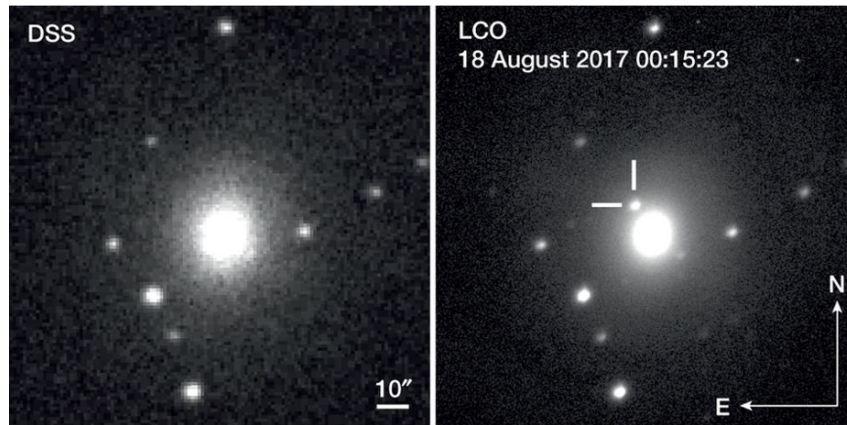


Figure 3.8: **NGC 4993.** Image of NGC 4993 taken in 1992 (left) and August 18th 2017 (right). **cite <https://www.nature.com/articles/nature24291/figures/2>**

component of the binary. With the detection of GW170817 in conjunction with GRB170817A, we now know this is the case for at least some short-hard GRBs. In this section we will discuss the immediate aftermath of a NS merger and the different GRB and GW signals that each could be produced.

NSBH Mergers The simplest case is that of an NSBH merger where the BH mass is significantly greater than that of the NS. If the BH to NS mass ratio is greater than 5:1 then the innermost stable circular orbit (ISCO)¹³ is greater than the tidal disruption radius¹⁴. This results in the NS being swallowed whole by the BH and leaving no accretion disk. The accretion disk is believed to be essential in powering a GRB after a NS merger, and so this case is not expected to produce a GRB. If the NSBH has a relatively low mass ratio, then the neutron star will be tidally disrupted before merging with the black hole. This will create a massive accretion disk and is expected to produce a GRB.

BNS Mergers For BNS mergers we must consider both the mass ratio and the total mass of the system. This is because it is possible for the merging neutron stars to either form a black hole or a *hypermassive neutron star* depending on the total mass of the binary system. The simplest of these is for a system with approximately equal mass ratio and a high total mass, such that a black hole can be produced immediately after merger. This requires the total mass to be above about $2.9M_{\odot}$ [cite], but the exact value depends on the EOS of neutron stars. In this case, no accretion disk will be produced and so no GRB is expected.

For an unequal mass ratio, the lighter neutron star is tidally disrupted by the larger neutron star. Matter from the lighter neutron star then accretes onto the more massive star, causing the more massive star to collapse into a black hole. This can potentially leave a massive accretion disk which could power a GRB.

If the total mass is less than about $3M_{\odot}$ [cite] then the merger is not expected to immediately form a black hole but to instead form a hypermassive neutron star. This is a neutron star that is supported by differential rotation and thermal pressure. The merger will produce an accretion disk which could power a GRB. It is also possible that the rapidly rotating hypermassive neutron star will be ellipsoidal in shape, making a powerful emitter of GWs. These GWs would be detectable with aLIGO up to about 20Mpc [cite]. Over time, the hypermassive neutron star will lose angular momentum due to GW emission and magnetic force, its rotation will become more uniform due to magnetic forces and viscosity, and it will radiate away its heat. These factors cause the hypermassive neutron star to eventually collapse to a black hole. This can happen on a timescale as short as milliseconds [cite someone].

¹³The ISCO is the smallest stable circular orbit for a test particle around a BH.

¹⁴The radius around a BH where the BH's tidal forces pull apart an in-falling star.

3.2.2 Core Collapse

Long GRBs are known to be caused by core collapse supernova. In this section we will discuss some of the possible central engines for long GRBs, focusing on those that could produce a GW signal strong enough to be detected with aLIGO.

Rotational Instabilities of Proto-Neutron Stars The cores of stars with initial masses in the range $10M_{\odot} \lesssim M \lesssim 25M_{\odot}$ are expected to collapse to rapidly rotating *protoneutron stars*; stars which are cooling and contracting to form a neutron star. If these protoneutron stars exhibit non-axisymmetric deformations, then rapid rotation of the star could drive significant GW emission, potentially detectable to 10Mpc with current detectors and even further if the protoneutron star accretes supernova material. There are several mechanisms that could cause non-axisymmetric deformations of the protoneutron star:

- Dynamical instabilities caused by rapid rotation and/or differential rotation.
- Secular (dissipation driven) instabilities caused by viscosity or GW emission.
- Magnetic distortion.
- **Add more detail (few sentences) to each of the mechanisms above. Include estimates of GW amplitude/frequency.**

Non-axisymmetric Instabilities of Accretion Disks For stars with initial mass greater than $\sim 30M_{\odot}$, the core collapse of the star will form a central black hole surrounded by an accretion disk. If the accretion disk has sufficiently high angular momentum and non-axisymmetric instabilities, then it can produce detectable GWs with a waveform similar to that of a low mass binary merger. For a stellar mass black hole with a clump of matter in its accretion disk of approximately $0.1M_{\odot}$, an accretion disk instability can potentially be detected out to 100Mpc with aLIGO [cite something]. An accretion disk instability can be caused by, for example, a high angular momentum such that the accretion disk is not gravitationally stable.

3.3 GRB Detectors

Most GRB detections have been come from the Swift and Fermi space telescopes, or a network of satellites called the Interplanetary Network. In this chapter we discuss the characteristics of these three missions, focusing on the aspects that affect a followup search for gravitational waves: sky coverage, source localisation, and sensitive energy range.

3.3.1 Swift

The Swift satellite is named for its ability to autonomously repoint itself towards a GRB within 90 seconds. It has three instruments: The first is the *Burst Alert Telescope* (BAT), the primary tool for GRB detection. It has a large field of view, approximately 2 steradians, and is highly sensitive in the 15-150keV energy range. The BAT can localise GRBs to within 1-4 arcminutes. The other two instruments are the *X-ray Telescope* (XRT) and the *UV/Optical telescope* (UVOT). These are used for followup observations of the GRB afterglow. They can also reduce the sky-error; XRT can localise to 3-5 arcseconds and UVOT to 0.5 arcseconds **cite** https://swift.gsfc.nasa.gov/about_swift/. The UVOT is also used to determine the redshift of the GRB. To date, Swift has detected over 1000 GRBs, with approximately 10% of these having $T_{90} < 2s$ **cite** https://swift.gsfc.nasa.gov/archive/grb_table/stats/

- How is redshift measurement done? Is it by finding the host galaxy?

3.3.2 Fermi

The Fermi Satellite [**cite** <https://www.nasa.gov/content/fermi/overview>] has two instruments: The first is the *Large Area Telescope* (LAT), and the second is the *Gamma-ray Burst Monitor* (GBM). The LAT is sensitive to higher energy photons (in the range 30MeV-300GeV), has sky coverage of 2 steradians, and can localise to within 1 arcminute of accuracy. The GBM is sensitive to lower energy photons (in the range 8keV-30MeV), is sensitive to a greater sky area than LAT (9.5 steradians), but cannot localise as well as LAT (typically 10s or 100s of square degrees).

3.3.3 The InterPlanetary Network

The interplanetary network (IPN) [**cite something**] is a network of GRB detectors on spacecraft that are in low Earth orbit, eccentric Earth orbit, traveling to other planets, or orbiting other planets already. None of the spacecraft in the IPN are dedicated GRB detectors; they are individually unable to determine the sky position of a GRB and are not as sensitive as dedicated GRB detectors. But the network as a whole acts as an all-time all-sky GRB monitor. And as the spacecraft are great distances apart, the network can triangulate sky position to within several arcminutes. The accuracy is greatly dependent on the number and position of the satellites that detect the GRB, with satellites at greater distances significantly improving the sky location error.

- Is there a list of IPN satellites for the O2 period?

3.4 GRB Gravitational Wave Search Strategies

In chapters 4 and 5, we discuss in detail two search pipelines for the CBC and burst search respectively. It is useful before discussing these searches to give an overview of the general strategy when searching for GWs associated with GRBs. This will give context to later discussions on the details of the searches, and also to the following section on the astrophysics that can be learned from detections of GWs from GRBs.

3.4.1 Triggered and Untriggered Searches

We can classify GW searches as being either *triggered* or *untriggered*. Untriggered searches will search all of the sky for all of the time when there is detector data. These all-sky all-time searches can be further divided into groups depending on the amount of time they take to run. Some searches, for example [cite **PyCBC live, gstlal, CWB papers**], are untriggered and low latency. They aim to find GW triggers and their sky position within minutes, allowing astronomers to followup the trigger. This is what happened for GW 170817, which was detected in low latency, in coincidence with a Fermi GRB, and followed up by ground and space based observatories. High latency untriggered searches, such as [cite **pycbc, gstlal papers**], work on much longer timescales, as long as months, but aim to achieve very high sensitivity with very low false alarm rates. These pipelines are more sensitive and have lower false alarm rates than the low latency searches.

Triggered searches use sky location and time information from other messengers, such as GRBs. These searches have the advantage of only needing to analyse a limited amount of data (as the trigger time is known) and being able to use information gleaned from the other messengers to restrict the search. For example, both long and short GRBs are suspected to emit their jets along the axis of angular momentum, which can be used to infer that the GRB is circularly polarised (see section 4.3.8).

Results from both triggered and untriggered searches can be used in *Multimessenger* searches [cite **subthreshold paper**]. These searches attempt to combine subthreshold triggers from multiple different messengers to make a confident detection. For example, a core collapse supernova could produce a long GRB, a GW signal, and neutrinos, but be too distant for any search using just one of these messengers to make a confident detection.

3.4.2 Modeled and Burst Searches

Searches for GWs can also be classified as either *modeled* or *burst*. Modeled searches have theoretical models of the waveforms they are searching for. For example, CBC signals can be modeled by numerical relativity and analytic methods [cite **something**]. These waveforms can then be used to build a more sensitive search, such as a matched filter search (see sections **list matched filter sections**). In

chapter 4 we will look at PyGRB, a targeted, modeled search for GWs associated with GRBs.

Burst searches use minimal assumptions about waveform morphology, instead relying on measures of coherence between detector data streams. This is in general less sensitive than if a waveform was known and a matched filter search could be carried out, but there are no such waveform models for supernovas/long GRBs. In chapter 5 we will discuss a burst search called Xpipeline.

3.5 GRB Astrophysics with Gravitational Waves

We have discussed what GRBs are, what causes GRBs, and seen that they are good emitters of GWs. We end this chapter with a discussion of how GW astronomy will add to our understanding of GRBs.

Short GRB Progenitors With the coincident detection of GRB 170817A and GW 170817, it is now known that at least some short GRBs are produced by the CBC of a binary neutron star system. As more detections are made it will become easier to determine whether this is the only source of short GRBs or if other mechanisms exist. Also, by comparing the electromagnetic counterpart to the knowledge gleaned from the GW signal about the binary, we can learn more about the central engine of the GRB. For example, a GRB detection in conjunction with a GW detection of a binary where we expect an accretion disk to form would indicate that some GRBs are accretion powered.

- Talk more about GW 170817. What was learned?

Long GRB Progenitors There is a lot of evidence that at least some long GRBs are powered by core collapse supernova. If this is the case, then it should be possible to detect GWs in coincidence with a nearby long GRB. The GW signal would provide clues as to the evolution of the core collapse. If there is no GW detection, then it is possible to constrain the dynamics of the core collapse.

- To what distance could we detect CCSNe?

Populations There are many unknowns about the populations of short GRBs and compact binary systems. As a network of GW detectors is essentially sensitive to the whole sky, it will be possible to better understand the population of compact binary systems within the well defined horizon of the detectors network. As GW signals also allow the direct inference of distance, it will even be possible to understand how the population of binary systems changes with distance, up to the horizon distance of the detector network. Due to the limited horizon of aLIGO and the low rate of neutron star mergers, this will have to wait for third generation GW detectors.

- Should I cite any population results LIGO already has?

GRB Distance and Luminosity For CBC GW signals where the inclination is approximately known, such as those associated with short GRBs, the distance to the source can be determined directly from the GW signal and without using a cosmological distance ladder. This allows independent verification of distances. Once the distance is known, the luminosity of the GRB can be reconstructed. This can also be used to as an alternative measure of Hubble’s constant [cite paper].

Jet Structure By comparing the rate of BNS and NSBH mergers detected through GWs that have a GRB counterpart to those that do not, it will be possible to determine the opening angle of short GRBs. Also, for strong GW signals, it should be possible to measure the polarisation of the GW. If the GW is face on, then it will be circularly polarised, but if it is edge on then it will be elliptically polarised. By measuring the polarisation of a large number of short GRBs and studying their prompt emission, it will be possible to fully describe the angular structure of short GRB jets.

- how many GRBs are needed to determine jet structure? citations needed
- Learn and write about what we learned about the jet structure of GRB 170718B

Neutron Star Equation of State The equation of state (EoS) of matter at neutron star densities is poorly constrained. There are many ways that studying the GW signal of neutron star binaries can be used to determine the EoS. For example, a GW signal for a system with a neutron star of mass $\sim 2M_{\odot}$ suggests a stiff EoS as a soft EoS could not support such a massive star.

Another way of constraining the EoS of neutron stars is to look at the orbital frequency of a neutron star binary when tidal disruption occurs, as this depends on the radius of the neutron star. Using this radius with the mass of the neutron star (as determined by the inspiral waveform), we can determine the density of the neutron star.

Tidal deformation can also affect the inspiral part of the waveform [cite same paper as bartos paper]. It has been shown [cite Read et al] that the advanced LIGO detectors can determine the radius of a neutron star to within $\sim 1\text{km}$ for a source at 100Mpc.

It has also been theorised that the core of neutron stars may be made of strange quark matter. If this is true, then it will have a significantly different GW signal for both CBC systems [cite same paper as bartos paper] and neutron star instabilities after a core collapse supernova. Thus, GW astronomy can determine if neutron stars contain quark matter cores.

Chapter 4

A Targeted Search for Gravitational Waves associated with Short GRBs

4.1 Introduction

Compact binary systems are strong emitters of gravitational waves (GWs), and many searches exist to search for signals from these systems [**cite pycbc gstlal and any others**]. These searches take theoretical waveforms [**cite someone**] and matched filter them against the strain data from GW detectors (**ref matched filter section**). It is known that the progenitor of at least some short GRBs are compact binaries, either binary neutron star (BNS) or low mass ratio neutron star - black hole (NSBH) systems. While the search pipelines mentioned above can detect the signals associated with short GRBs, they search for binary black hole (BBH) systems as well. This means they cannot use any of the EM information that GRB detectors collect. By combining the EM information from GRB detectors with the strain data from GW detectors, we can make a more sensitive search. This is what PyGRB was developed to do.

There are several ways EM information can improve a GW search. The first is that the time and sky position of the short GRB are known from GRB detectors (see section 3.3). This not only means that the analysis can run on less data but this information can also be used to make a more powerful detection statistic, which we will see in section 4.3.3. We can also assume that the binary system that produced the GRB was face on, as GRBs are emitted perpendicular to the orbital plane. We will see in section 4.3.8 that this can also be used to improve the detection statistic.

In this chapter we describe PyGRB [**cite all the papers**], a targeted matched filter search for short GRBs. PyGRB is integrated into the PyCBC data analysis software [**cite pycbc technical paper**], which is primarily an all-time, all-sky, matched filter search for compact binary coalescence (CBC) signals. We will begin

with an overview of the PyCBC pipeline, as this is the foundation that PyGRB is built upon. In section 4.3 we discuss the unique features of PyGRB that make it more sensitive to GRB signals than the all-sky, all-time search. Then we discuss the findings of PyGRB in the second advanced LIGO and advanced Virgo observing run. In section 4.5 we will discuss some recent improvements in speed and sensitivity to the PyGRB pipeline. We then end this chapter with a discussion on future improvements that can be made to PyGRB.

4.2 PyCBC overview

- Just the parts of the CBC search that are relevant for PyGRB
- Highlight some of the differences
- All sky all time
- Searches for a bigger population - i.e. not just BNS and low mass NSBH

4.2.1 Matched Filter Search

- Contents of this section depends on what goes into the background chapter. I think the mathematical formalism of matched filtering should go earlier in the thesis

4.2.2 Consistency Checks

Chisq tests

- Frequency bins
- Template bank
- Autocorrelation
- Need to learn about the effective degrees of freedom stuff.

4.2.3 PyCBC Results

- Detections made in O1 and O2 - Use catalog paper
- Sensitivity of search - Useful later for comparison between PyCBC and PyGRB

4.3 PyGRB

The PyCBC pipeline discussed above is an all-time, all-sky search. Some CBC systems¹ are expected to produce short GRBs (see chapter 3). *Targeted searches*, where

¹BNS and low mass ratio NSBH.

the time and sky position of the source are known, are not only computationally cheaper, but can make use of the additional information to make a more powerful detection statistic. In the case of a GRB search, as the GRB is emitted along the axis of rotation, we can also assume the binary is face on. This can be used to further reduce the background noise for a GW search.

In this section we will describe PyGRB; a targeted, matched filter search for GWs associated with short GRBs. The detection statistic of PyGRB is the *coherent SNR*, which uses the sky position and time of the GRB, as well as the antenna response function and PSD of each detector in the network. Using these, it is possible to calculate the amount of signal power expected in each detector. For example, consider the case of a GW directly over one of the detectors in a network of identical interferometers. Detectors are most sensitive to GW sources that are directly overhead, so the detector with the source directly overhead will have more power than any other detector in the network. If one of the other detectors in the network has more power, then it is less likely that this trigger is a true GW signal. Triggers with the expected ratio of signal power in each detector will have a higher coherent SNR than those that do not.

This qualitative description of coherence is made rigorous in the following sections up to 4.3.3. In section 4.3.5 we discuss the *null SNR*, another coherent statistic that measures the energy inconsistent with a GW. The rest of this section is about the PyGRB workflow and how to use coherent statistics to confidently detect GWs associated with GRBs.

4.3.1 A Coherent Search for Coalescing Binary Systems

- Show where the A's come from (don't know how to do this yet)

4.3.2 Binary Coalescence Waveform

PyGRB searches for CBC signals with circular orbits and aligned spin components. These waveforms depend on 11 parameters: The two component masses M_1 and M_2 , the component spins, the sky location (θ, ϕ) , the distance D , the coalescence time t_0 , the orientation ι , the polarisation angle ψ , and the coalescence phase ϕ_0 . We can reduce to nine parameters if we assume the sky position is known. Of these, the distance, binary inclination, polarisation, and coalescence phase affect the phase and amplitude of the waveform and not the signal morphology. That is, the waveform can be written as

$$h_+(t) = \mathcal{A}^1 h_0(t) + \mathcal{A}^3 h_{\pi/2}(t) \quad (4.1)$$

$$h_\times(t) = \mathcal{A}^2 h_0(t) + \mathcal{A}^4 h_{\pi/2}(t) \quad (4.2)$$

where $h_0(t)$ and $h_{\pi/2}(t)$ describe the two phases of the waveform, are usually assumed to be orthogonal, and depend only on the component masses. The amplitudes \mathcal{A}^i

are given by

$$\mathcal{A}^1 = \frac{D_0}{D} \left(\frac{(1 + \cos^2 \iota)}{2} \cos 2\phi_0 \cos 2\psi - \cos \iota \sin 2\phi_0 \sin 2\psi \right) \quad (4.3)$$

$$\mathcal{A}^2 = \frac{D_0}{D} \left(\frac{(1 + \cos^2 \iota)}{2} \cos 2\phi_0 \sin 2\psi + \cos \iota \sin 2\phi_0 \cos 2\psi \right) \quad (4.4)$$

$$\mathcal{A}^3 = -\frac{D_0}{D} \left(\frac{(1 + \cos^2 \iota)}{2} \sin 2\phi_0 \cos 2\psi + \cos \iota \cos 2\phi_0 \sin 2\psi \right) \quad (4.5)$$

$$\mathcal{A}^4 = \frac{D_0}{D} \left(-\frac{(1 + \cos^2 \iota)}{2} \sin 2\phi_0 \sin 2\psi + \cos \iota \cos 2\phi_0 \cos 2\psi \right) \quad (4.6)$$

where D_0 is used to scale the amplitude of the waveforms. For any \mathcal{A}^μ we can invert equations 4.3-4.6 to obtain the physical parameters up to a reflection symmetry of the system.

The response of GW detector X to a GW $h_{+,\times}$ is given by

$$h^X(t) = F_+(\theta^X, \phi^X, \chi^X)h_+(t^X) + F_\times(\theta^X, \phi^X, \chi^X)h_\times(t^X) \quad (4.7)$$

where $F_{+,\times}^X$ is the antenna response of detector X to the plus and cross polarisation of the GW, the angles $(\theta^X, \phi^X, \chi^X)$ map the detector frame to the radiation frame, and the time of arrival t^X at detector X depends on the sky location of the source and the time of arrival at the fiducial location, e.g. the Earth's center. Combining 4.7 with 4.1 and 4.2 we find

$$h^X(t) = \mathcal{A}^\mu h_\mu^X(t) \quad (4.8)$$

where the h_μ^X are given by

$$h_1^X = F_+^X h_0(t^X) \quad (4.9)$$

$$h_2^X = F_\times^X h_0(t^X) \quad (4.10)$$

$$h_3^X = F_+^X h_{\pi/2}(t^X) \quad (4.11)$$

$$h_4^X = F_\times^X h_{\pi/2}(t^X) . \quad (4.12)$$

4.3.3 Coherent SNR

We are about to derive the *coherent SNR*, a detection statistic for a multidetector matched filter search.² The idea is to find the likelihood that a trigger is a real GW in terms of \mathcal{A}^μ (assuming the detector noise is Gaussian). We then maximise the likelihood with respect to \mathcal{A}^μ , reducing the parameter space from nine dimensions to five, eliminating the noise associated with these extra dimensions and so improving the detection statistic. In section 4.3.4, we will compare the coherent SNR to the *coincident SNR*, the detection statistic currently used by the all sky search.

²Note that this is not the final detection statistic for PyGRB, which is a modified version of the coherent SNR. See sections 4.3.7 and 4.3.8

We begin by defining the data $s^X(t)$ from detector X to be the sum of the noise $n^X(t)$ and signal $h^X(t, \xi)$ components

$$s^X = n^X(t) + h^X(t, \xi) . \quad (4.13)$$

The noise power spectral density (PSD) S_h^X for detector X is defined by

$$\langle \tilde{n}^X(f) [\tilde{n}^X(f')]^* \rangle = \delta(f - f') S_h^X(f) \quad (4.14)$$

where the angle brackets denote the time average of the noise and tildes indicate that the function has been Fourier transformed. The matched filter between the GW waveform and detector data is given by the inner product

$$(s^X | h^X) = 4\text{Re} \int_0^\infty \frac{\tilde{s}^X(f) \cdot [\tilde{h}^X(f)]^*}{S_h^X(f)} e^{2\pi i f t} df . \quad (4.15)$$

The probability of obtaining detector data s given the presence of GW h is denoted $P(s|h)$. If the detector noise is Gaussian, then

$$P(s|h) = \frac{1}{2\pi} e^{-(s-h|s-h)/2} . \quad (4.16)$$

The likelihood ratio is the probability of obtaining detector data s when the GW signal h is present, divided by the probability of obtaining the data s in the absence of a GW, which we denote $P(s|0)$. This gives us

$$\Lambda(h) = \frac{P(s|h)}{P(s|0)} = \frac{e^{-(s-h|s-h)/2}}{e^{-(s|s)/2}} . \quad (4.17)$$

For convenience, we will use the log-likelihood

$$\log \Lambda = (s|h) - \frac{1}{2}(h|h) . \quad (4.18)$$

The log-likelihood defined above is a measure of the probability a trigger that has been found in a single detector is real. For a coherent search we need a likelihood measure that takes into consideration every detector in the network. First we must define the multidetector inner product, which we take to be the sum of individual detector inner products for the d detectors in the network

$$(\mathbf{a}|\mathbf{b}) = \sum_{X=1}^d (a^X | b^X) . \quad (4.19)$$

The multidetector log-likelihood then becomes

$$\log \Lambda = (\mathbf{s}|\mathbf{h}) - \frac{1}{2}(\mathbf{h}|\mathbf{h}) \quad (4.20)$$

where $\mathbf{h} = (\mathbf{F}_+ \mathbf{h}_0, \mathbf{F}_\times \mathbf{h}_0, \mathbf{F}_+ \mathbf{h}_{\pi/2}, \mathbf{F}_\times \mathbf{h}_{\pi/2})$. Using 4.8, we can rewrite the log-likelihood in terms of \mathcal{A}^μ

$$\ln \Lambda = \left[\mathcal{A}^\mu(\mathbf{s}|\mathbf{h}_\mu) - \frac{1}{2} \mathcal{A}^\mu \mathcal{M}_{\mu\nu} \mathcal{A}^\nu \right] \quad (4.21)$$

where

$$\mathcal{M}_{\mu\nu} = (\mathbf{h}_\mu | \mathbf{h}_\nu) . \quad (4.22)$$

We want to find the template that maximises the log-likelihood. The values for \mathcal{A}^μ for which the log-likelihood 4.21 is maximal are given by

$$\hat{\mathcal{A}}^\mu = \mathcal{M}^{\mu\nu}(\mathbf{s}|\mathbf{h}_\nu) \quad (4.23)$$

where $\mathcal{M}^{\mu\nu}$ is the inverse of $\mathcal{M}_{\mu\nu}$. The *coherent SNR* ρ_{coh} is then defined by

$$\rho_{\text{coh}}^2 = 2 \ln \Lambda|_{\text{max}} = (\mathbf{s}|\mathbf{h}_\mu) \mathcal{M}^{\mu\nu}(\mathbf{s}|\mathbf{h}_\nu) . \quad (4.24)$$

The \mathcal{A}^μ terms have been maximised over, leaving only five of our original nine waveform parameters to search over.

- Does the all sky search use all those parameters in its template bank?
- Quick discussion on template banks, EM bright etc.

4.3.4 Comparison to Coincident Search

In this section we will put the coherent SNR into a form such that it is easy to compare to the *coincident SNR*, which is the detection statistic currently used by all sky matched filter searches. The coincident SNR is defined to be the quadrature sum of the matched filter of the individual detectors

$$\rho_{\text{coinc}}^2 = \sum_{X,Y} \sum_{i=0,\pi/2} \left(s^X \left| \frac{h_i}{\sigma^X} \right| \right) [\delta^{XY}] \left(s^Y \left| \frac{h_i}{\sigma^Y} \right| \right) \quad (4.25)$$

where $\sigma^X = (h_i^X | h_i^X)$.

We can make 4.24 more easily comparable to 4.25 by making \mathcal{M} diagonal. To do this, first note that CBC signals spend a large number of cycles in the sensitive frequency range of the detectors. This causes the 0 and $\frac{\pi}{2}$ phases of the waveform to be approximately orthogonal. The slow frequency evolution means that the two phases have roughly equal amplitude. Hence we find

$$(h_0^X | h_{\pi/2}^X) \approx 0, \quad (4.26)$$

$$(h_{\pi/2}^X | h_{\pi/2}^X) \approx (h_0^X | h_0^X) = (\sigma^X)^2. \quad (4.27)$$

Using these, we see that \mathcal{M} simplifies to

$$\mathcal{M} = \begin{bmatrix} A & C & 0 & 0 \\ C & B & 0 & 0 \\ 0 & 0 & A & C \\ 0 & 0 & C & B \end{bmatrix}$$

with

$$A = \sum_X (\sigma^X F_+^X)^2 \quad (4.28)$$

$$B = \sum_X (\sigma^X F_\times^X)^2 \quad (4.29)$$

$$C = \sum_X (\sigma^X F_+^X)(\sigma^X F_\times^X) . \quad (4.30)$$

We can simplify further by rotating the detector frame to the *Dominant Polarization Frame*. In this frame, the antenna response functions to the plus and cross polarisation are orthogonal, and hence $C = 0$, making \mathcal{M} diagonal. As we included a polarisation angle between the equatorial and radiation frame (χ) and the radiation and source frame (ψ), we can rotate our network frame without placing further constraints on the system. In particular, we can rotate through an angle χ^{DP} such that $C = 0$. This gives us the following antenna response functions

$$F_+^{\text{DP},X} = F_+^X \cos 2\chi^{\text{DP}} + F_\times^X \sin 2\chi^{\text{DP}} \quad (4.31)$$

$$F_\times^{\text{DP},X} = -F_+^X \sin 2\chi^{\text{DP}} + F_\times^X \cos 2\chi^{\text{DP}} . \quad (4.32)$$

Using these antenna response functions in (4.30) and solving for χ^{DP} , we find

$$\chi^{\text{DP}} = \frac{1}{4} \arctan \left(\frac{2 \sum_X (\sigma^X F_+^X)(\sigma^X F_\times^X)}{\sum_X [(\sigma^X F_+^X)^2 - (\sigma^X F_\times^X)^2]} \right) . \quad (4.33)$$

This does not uniquely define the dominant polarisation frame, so we also require the network to be more sensitive to the plus polarisation

$$|F_+^{\text{DP},X}| \geq |F_\times^{\text{DP},X}| . \quad (4.34)$$

In what follows, we assume we are in the dominant polarisation frame and drop the DP superscript. Inverting \mathcal{M} and using (4.24), we see the coherent SNR in the dominant polarisation frame is

$$\rho_{\text{coh}}^2 = \frac{(\mathbf{s}|\mathbf{F}_+\mathbf{h}_0)^2 + (\mathbf{s}|\mathbf{F}_+\mathbf{h}_{\pi/2})^2}{(\mathbf{F}_+\mathbf{h}_0|\mathbf{F}_+\mathbf{h}_0)^2} + \frac{(\mathbf{s}|\mathbf{F}_\times\mathbf{h}_0)^2 + (\mathbf{s}|\mathbf{F}_\times\mathbf{h}_{\pi/2})^2}{(\mathbf{F}_\times\mathbf{h}_0|\mathbf{F}_\times\mathbf{h}_0)^2} . \quad (4.35)$$

We can rewrite this as

$$\rho_{\text{coh}}^2 = \sum_{X,Y} \sum_{i=0,\pi/2} \left(s^X \left| \frac{h_i}{\sigma^X} \right| \right) [f_+^X f_+^Y + f_\times^X f_\times^Y] \left(s^Y \left| \frac{h_i}{\sigma^Y} \right| \right) \quad (4.36)$$

where we define the orthonormal vectors

$$f_{+,\times}^X = \frac{\sigma^X F_{+,\times}^X}{\sqrt{\sum_Y (\sigma^Y F_{+,\times}^Y)^2}}. \quad (4.37)$$

We can now compare the coherent SNR as given in (4.36) to the coincident SNR given by (4.25). The coincident SNR is the quadrature sum of all the energy in each detector. In the space spanned by individual detector SNRs, where each trigger is represented by a point in this space, it is the distance from the origin to the trigger. It is the total energy of that trigger.

The coherent SNR is a projection of the total energy of a trigger onto the subspace spanned by f_+ and f_\times . This subspace is the space consistent with a gravitational wave signal from the given sky location and with the PSD of the detectors at the given time. Orthogonal to this space is the space inconsistent with a signal. Noise contributes energy to all components of the coincident SNR, and so projecting it onto the signal space will reduce its amplitude. The energy due to a GW signal lays entirely in the signal plane, so the projection does not change the distance to the point representing the trigger. In the case of a real GW detection, the coherent and coincident SNR are approximately the same as most of the energy of the trigger is due to the GW, though the coherent SNR will be slightly lower as some energy will be due to noise, and so is removed.

Another way to think of the difference between coherent and coincident SNR is in terms of the number of degrees of freedom. The coincident SNR has noise contributions from the phase and amplitude measurements in each detector, resulting in $2N$ degrees of freedom, where N is the number of detectors. From (4.35) we can see that the coherent SNR has four degrees of freedom: the 0 and $\pi/2$ phases of the plus and cross polarisation amplitudes of the gravitational wave. In the case where we have a non-degenerate³ two detector network, then both the coincident and coherent SNR have four degrees of freedom. In this case, we can show that the coherent and coincident SNRs are identical by calculating $f_+^X f_+^Y + f_\times^X f_\times^Y = \delta^{XY}$.

4.3.5 Null SNR

Gravitational waves have two polarisations. Therefore we can fully describe the signal with two non-aligned detectors. Adding additional detectors to our network allows us to construct additional data streams that have the GW signal removed. We can easily construct a such a data stream by subtracting the coherent SNR (the

³i.e. sensitive to both plus and cross polarisations.

energy consistent with a GW signal) from the coincident SNR (the total energy measured by the detectors). Any energy left from this is associated with noise, and we call this the *null SNR*

$$\rho_N^2 = \rho_{\text{coinc}}^2 - \rho_{\text{coh}}^2 = \sum_{X,Y} \sum_{i=0,\pi/2} \left(s^X \left| \frac{h_i}{\sigma^X} \right| \right) [N^{XY}] \left(s^Y \left| \frac{h_i}{\sigma^Y} \right| \right) \quad (4.38)$$

where

$$N^{XY} = \delta^{XY} - f_+^X f_+^Y - f_\times^X f_\times^Y. \quad (4.39)$$

Noise triggers are incoherent, and so for noise we expect $\rho_{\text{coinc}}^2 \gg \rho_{\text{coh}}^2$. Thus the null SNR is high in this case. For a true GW signal, most of the energy will be coherent, and so we expect $\rho_{\text{coinc}}^2 \simeq \rho_{\text{coh}}^2$. Thus, the null SNR is expected to be close to zero for a true GW signal.

The null SNR is similar to *null streams*, which are used by burst searches (see section 5.2). The null stream in a burst search is made to remove a GW signal of any morphology.

4.3.6 The PyGRB Workflow

We have seen several statistics that can indicate whether a trigger is a GW signal or noise. We apply cuts on these statistics, discarding any triggers that fail the cuts. This is especially important for the coherent search as the coherent statistics are slow to calculate and additional to all the statistics used in the coincident search. To speed up the coherent search, we discard as many triggers as possible at the earliest possible time in the workflow. In this section we will discuss the order of the workflow and the cuts made on each statistic.

The analysis begins with matched filtering the individual detector data. Using the known sky location of the GRB and the fact that gravitational waves travel at the speed of light, we timeshift the data from each detector so that a GW will be coincident in each detector. A list of triggers is formed for each detector by keeping only times when the individual detector SNR is greater than four, the rest of the data is discarded. We then discard any trigger that is not coincident in at least two detectors in the network.

The coincident SNR is then easy to calculate, and any trigger with a coincident SNR below six is discarded. The coherent SNR is then calculated for the remaining triggers. This reduces the SNR of noise, but not signals, and so we apply the same threshold of six to the coherent SNR.

Once we have the coincident and coherent SNR, the null SNR is easy to calculate. The null SNR of a trigger increases with the energy of the trigger for both noise and GW signals, so a very loud GW signal can still have a high null SNR. For this reason, we increase the null SNR threshold with the coherent SNR for loud triggers but keep the null SNR threshold fixed for quiet triggers, as can be seen in fig.4.1.

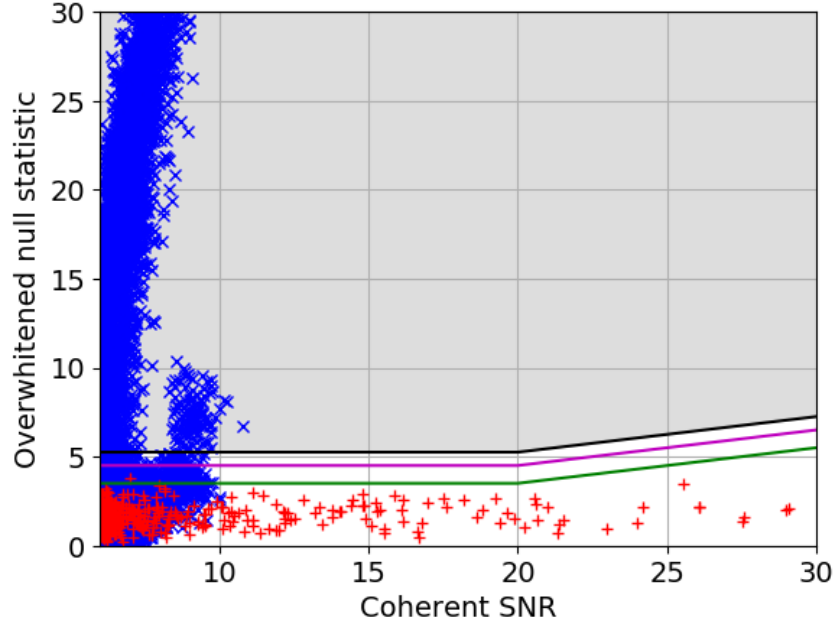


Figure 4.1: **Null Stat Cut.** Here we plot the coherent SNR against the null SNR. The blue crosses are background triggers. The red pluses are signal injections. The black line is the veto line, with all triggers in the shaded region above the line being discarded. The green line indicates the expected SNR for optimally oriented injections. The magenta and cyan lines show 1 and 2 sigma errors on the green line.

To be precise, we keep triggers meeting either of the following criteria:

$$\rho_N \leq 5.25 \text{ and } \rho_{\text{coh}} \leq 20 \quad (4.40)$$

or

$$\rho_N \leq \frac{\rho_{\text{coh}}}{5} + 5.25 \text{ and } \rho_{\text{coh}} > 20 . \quad (4.41)$$

The χ^2 tests (described in 4.2.2) are computationally expensive to calculate, so we calculate these only on those triggers that survive all the other tests. These tests are the same as used in the coincident search. In particular, they are not coherent but applied to the individual detector triggers.

- chi2 tests - Is this final? No coherent chi2 tests? Do we need to explain why we are no longer using coherent chi2 tests?
- Go through plot caption with Steve

4.3.7 Reweighted SNR

After all the cuts have been made, the remaining triggers are then reweighted twice. First they are reweighted according to their χ^2 values, according to (4.42), then they

are reweighted again according to their null SNR, according to (4.43).

$$\text{Chi2 reweighting goes here} \quad (4.42)$$

$$\text{null reweighting goes here} \quad (4.43)$$

Note that this reweighting only down-weights triggers with a high χ^2 or a high null SNR. The reweighting should not significantly alter the network SNR of a GW signal, but will reduce the SNR of noise. The reweighted SNR is detection statistic used by PyGRB.

4.3.8 Searching for Face on Signals

The opening angle of the jet from a short GRB is expected to be $< 30^\circ$, emitted perpendicular to the orbital plane. In a similar way to how the coherent SNR accounts for sky position and PSD of each detector to better reject noise, we can build a new detection statistic that takes into account that we expect a true GW signal to have an orbital inclination of $\iota \approx 0$ or $\iota \approx \pi$ with respect to the observer.

We can see from equations (4.3) to (4.6) that the GW amplitude terms depend linearly on $\cos \iota$ and $(1 + \cos^2 \iota)/2$. For $0^\circ \leq \iota < 30^\circ$, these are equal to within 1%. Thus, for $\iota \approx 0$, we can use the approximation $\cos \iota \approx (1 + \cos^2 \iota)/2$. For $\iota \approx \pi$, the two terms are approximately equal in magnitude, but with opposite signs.

By defining

$$\tilde{D} = \frac{D}{\cos \iota} \text{ and } \chi_{l,r} = \phi_0 \pm \psi \quad (4.44)$$

we can rewrite the GW amplitudes (4.3) to (4.6) for $\iota \approx 0$ as

$$\mathcal{A}^1 \approx \mathcal{A}^4 \approx -\frac{D_0}{\tilde{D}} \cos 2\chi_l \equiv \mathcal{B}_1 \quad (4.45)$$

$$\mathcal{A}^2 \approx -\mathcal{A}^3 \approx \frac{D_0}{\tilde{D}} \sin 2\chi_l \equiv \mathcal{B}_2. \quad (4.46)$$

We see that now there are only two amplitude factors, \mathcal{B}_1 and \mathcal{B}_2 . Therefore, the GW is circularly polarised. Similar results can be derived for the $\iota = \pi$ case.

Substituting these into (4.21) and working in the dominant polarisation frame⁴, we find

$$\ln \Lambda = \mathcal{B}_1(\mathbf{s}|\mathbf{F}_+\mathbf{h}_0 + \mathbf{F}_\times\mathbf{h}_{\pi/2}) + \mathcal{B}_2(\mathbf{s}|\mathbf{F}_\times\mathbf{h}_0 - \mathbf{F}_+\mathbf{h}_{\pi/2}) \quad (4.47)$$

$$- \frac{1}{2}[\mathcal{B}_1^2 + \mathcal{B}_2^2][(\mathbf{F}_+\mathbf{h}_0|\mathbf{F}_+\mathbf{h}_0) + (\mathbf{F}_\times\mathbf{h}_0|\mathbf{F}_\times\mathbf{h}_0)]. \quad (4.48)$$

⁴And recalling that $(h_{\pi/2}^X|h_{\pi/2}^X) \approx (h_0^X|h_0^X)$ for CBC signals.

Maximising the log likelihood with respect to \mathcal{B}_1 and \mathcal{B}_2 , we find

$$\rho_{\text{coh}}^2 = 2 \ln \lambda_{\text{max}} = \frac{\alpha^2 + \beta^2}{(\mathbf{F}_+ \mathbf{h}_0 | \mathbf{F}_+ \mathbf{h}_0) + (\mathbf{F}_\times \mathbf{h}_0 | \mathbf{F}_\times \mathbf{h}_0)} \quad (4.49)$$

where

$$\alpha = (\mathbf{s} | \mathbf{F}_+ \mathbf{h}_0 + \mathbf{F}_\times \mathbf{h}_{\pi/2}) \quad (4.50)$$

$$\beta = (\mathbf{s} | \mathbf{F}_\times \mathbf{h}_0 - \mathbf{F}_+ \mathbf{h}_{\pi/2}) . \quad (4.51)$$

In the ι 180° case, the SNR takes the same form, but with

$$\alpha = (\mathbf{s} | \mathbf{F}_+ \mathbf{h}_0 - \mathbf{F}_\times \mathbf{h}_{\pi/2}) \quad (4.52)$$

$$\beta = (\mathbf{s} | \mathbf{F}_\times \mathbf{h}_0 + \mathbf{F}_+ \mathbf{h}_{\pi/2}) . \quad (4.53)$$

We now have a detection statistic with only two degrees of freedom, as opposed to the four for the coherent SNR and 2N for the coincident SNR in the case of an N detector network. This gives us an additional null stream, and allows us to have a null stream even when analysing data from just two detectors. The circularly polarised coherent SNR allows for a 3% increase in range for a given FAP when compared to the generic coherent SNR. This corresponds to a 10% increase in the rate of GW signals detected.

4.3.9 Searching over a Sky Patch

The sky position of GRBs cannot be determined exactly. The uncertainty in sky position depends on which satellite detected the GRB. The BAT instrument on Swift (see section ??) can localise to within 1-4 arcminutes [**citation in Andrew's paper**]. This is a small enough uncertainty that the GW search can use a single sky point [**cite Andrew's paper?**]. Other GRB detectors cannot provide as precise localisation. The GBM on the Fermi satellite often provides a location with a roughly circular uncertainty with a radius of several degrees. The IPN uses triangulation to determine sky position, so depending on the number of satellites in the IPN that make the detection, sky localisation and vary from under a square degree to thousands of square degrees.

It is important to have approximately the correct sky position for a coherent search. If the sky position is incorrect, then the detector antenna patterns will be incorrect for the GRB, hurting the search sensitivity. More crucially, with the wrong sky position we will calculate the wrong time delay between detectors in the network and make it impossible to correctly find coincident triggers between detectors.

To search over a sky patch we fill the patch with concentric circles of points separated by δa , such that each ring has $2\pi n / \delta a$ points with $n = 0$ being the central point of the patch. These concentric circles extend to cover 3σ confidence region.

The value of δa depends on the timing uncertainty we are willing to accept. In practice, we use a timing uncertainty of $\delta t = 0.5\text{ms}$, which will limit the lost SNR to $< 5\%$ [cite andrew’s paper]. Searching over the sky grid is done after the computationally dominant step of matched filtering, so that searching over the sky grid does not significantly hurt analysis speed [Need some analyses to back this up].

It is interesting to note that in a two detector network, there is a ring of sky points that give the same time delay between the detectors. In this case, only the antenna response of the detectors change with sky position, but the antenna response functions drop out of the coherent SNR in the two detector case⁵. This would allow for us to analyse a greatly reduced set of sky points. Unfortunately, when we limit our search to looking for circularly polarised GWs, as discussed in section 4.3.8, then the antenna patterns again affect the detection statistic, and we must search over the sky grid.

- Include image of sky grid

4.3.10 Event Significance

Knowing the reweighted SNR of a trigger does not help us know if that trigger is a GW signal or not unless we know the false alarm probability (FAP) of that trigger. The FAP depends on many factors. For example, poor data quality at the time being analysed can lead to more high SNR triggers due to noise. Or the detectors may not be sensitive to the sky position of the GRB, which makes it less likely that a high SNR trigger is a GW signal. In this section we outline the process used to calculate the significance of an event.

As detector noise is not stationary, we calculate the FAP by using the data near the time of the GRB. A GW, if it is there, is assumed to be in the window starting five seconds before the GRB time and ending one second after the GRB time. This window would allow the detection of GWs from most GW models associated with short GRBs. We call this 6s window the *on-source* window. The loudest event, with the loudest template, in the on-source is taken as our trigger.

To evaluate the p-value of this trigger, we analyse approximately⁶ one hour of data around the on-source, called the *off-source* data. This data is assumed to not contain a GW signal. This data is divided up into as many 6s segments as possible, so that if we have T_{off} seconds of data then we have $N = T_{\text{off}}/T_{\text{on}}$ segments, where T_{on} is the length of the on-source. These segments are called the *off-trials*. Using

⁵This is because in the two detector case, both the coherent and coincident SNR have four degrees of freedom. Thus, any observed amplitude and phase is consistent with an astrophysical signal.

⁶The amount of data used before and after the on-source varies depending on the data available. For example, sometimes a GRB will happen less than an hour before (after) an interferometer loses (acquires) lock.

the same criteria as the for the on-source, we find the loudest trigger in each off-trial. As each off-trial is assumed to only contain noise, the FAP of the on-source trigger is the fraction of off-trials that have a louder trigger than the on-source, i.e. the probability that noise would produce a trigger with a reweighted SNR louder than the on-source.

This process alone is only capable of achieving a FAP of 10^{-3} [**Cite Andrew’s paper**] as there is not enough off-trials to claim a lower FAP. This is low enough to reject a signal hypothesis, but not low enough to claim a detection, which requires a FAP of less than 10^{-5} [**Cite Andrew again**]. One way to get around this would be to use a longer off-source, but as detector noise is not stationary, it is possible that this extra off-source data would have different properties to the time of the GRB. It would also significantly increase the computational time required to complete the analysis.

To find a lower FAP, the off-source data from each detector is time-shifted relative to the other detectors in the network and the data is then reanalysed. The time-shifts are longer than the light travel time between the detectors in the network and also longer than the duration of a typical glitch (less than one second), so that triggers in the time shifted data are unlikely to appear coherently.

There are two types of time shifting. The first is call *short slides* and the other *long slides*. The difference between them stems from the fact that PyGRB analyses data in segments of 128s [**Is this still true after the rewrite**]. Short slides refers to when only the data within a segment is time shifted. To be precise, in a HLV analysis, the Hanford data is not shifted, the Livingston data is shifted in increments of 6s, and the Virgo data is shifted in increments of 12 seconds. Short slides do not require the data to be matched filtered again, as the time shift happens after the filtering step, so only the network statistics must be recalculated. This means short slides are computationally cheap and can allow for a FAP as low as 10^{-4} . Long slides refers to when the segments themselves are shifted relative to one another. This requires the matched filtering to be redone, as the matched filter time series is not saved when moving to the next segment for analysis. This makes long slides much more computationally expensive than short slides. However, short slides can be done on top of the long slides, so that just 10 long slides are needed to show that a trigger has a $\text{FAP} < 10^{-5}$.

It should be noted that our treatment of time slides assumes that each combination of time shifted data is statistically independent of the others. This is not true, as each combination of time shifted data is still the same data being analysed.

- check segment length
- Are we still doing time slides the same way?

4.3.11 Calculating Search Sensitivity

It is useful to estimate the sensitivity of our search to GWs around the time of an analysed GRB. This allows us to know the range to which we are sensitive to GWs in the direction of the GRB, and so rule out the possibility that the GRB was within that distance and producing GWs.

To do this, we inject simulated signals into the off-source data and see if PyGRB can find them. The injected signals are CBC waveforms drawn from three populations: BNS mergers, NS-BH systems with spins aligned with the angular momentum, and NS-BH systems with spins generically aligned. The injections are drawn from an astrophysically motivated distribution of distances, component masses and spins, and binary inclination. The NS mass distribution is a normal distribution with mean $1.4M_{\odot}$ and standard deviation of $0.2M_{\odot}$ [refs in O2 GRB paper] restricted to the $1M_{\odot} - 3M_{\odot}$ range [refs in O2 GRB paper]. The NS spins magnitude is restricted to be less than the fastest observed pulsar spin, i.e. ≤ 0.4 [refs in O2 GRB paper]. The BH masses are drawn from a normal distribution of mean $10M_{\odot}$ with standard deviation of $6M_{\odot}$, restricted to the range $3M_{\odot} - 15M_{\odot}$. The BH spin magnitudes are ≤ 0.98 , motivated by X-ray observations [refs in O2 GRB paper]. The injections are limited to have binary inclinations of $0^{\circ} - 30^{\circ}$ or $150^{\circ} - 180^{\circ}$, as the opening angle for a CBC powered GRB is expected to be less than 30° .

We quantify the sensitivity of the search using the 50% and 90% exclusion distance. This is the distance at which 50% and 90% of the injected signals are recovered with a greater reweighted SNR than the most significant trigger in the on-source.

4.4 O2 PyGRB Search

From November 2016 to August 2017, the Advanced LIGO detectors were in their second observing run, O2, with the Advanced Virgo joining on August 1st 2017. During this time, PyGRB was used to search for GW signals associated with short and ambiguous GRBs. These results were reported in [cite O2 GRB paper], and we summarise the findings here. We begin by discussing the sample of GRBs chosen to analyse. Then we summarise the findings of the search.

4.4.1 GRB sample

The GRBs analysed were those detected Swift BAT, Fermi GBM, and the IPN (see section 3.3 for more details on these). There were in total 242 bursts detected by Swift and Fermi, and 52 detected by the IPN, with many GRBs appearing in both catalogues. Only short GRBs ($T_{90} < 2s$) and ambiguous GRBs ($2s < T_{90} < 4s$) were analysed using PyGRB, as long GRBs are not expected to have a CBC progenitor. At least 1664s of data is required in at least one detector for PyGRB to correctly estimate the background. Any GRB occurring at a time when this requirement

could not be met was not analysed. This left 42 short/ambiguous GRBs that could be analysed by PyGRB.

4.4.2 Results

The analysis found one GW signal, GW170817. It was associated with GRB 170817A and had been previously reported [cite BNS paper]. The p-value for this event is $\leq 9.38 \times 10^{-6}$ and the coherent SNR is 31.26 **do we want to explain data cleaning and gating?** [cite O2 GRB]. No other GWs were detected in conjunction with any other GRBs. Apart from GW170817, there were six GRBs with a p-value of less than 0.1. These candidate events had further data quality checks and were analysed using lalInference [cite Veitch like in GRB paper], a coherent Bayesian parameter estimation pipeline, to determine if the signal could be a subthreshold signal or if it was more likely to be background noise. No evidence was found to of a subthreshold signal.

In figure 4.2 we have plotted the p-value for the remaining 41 GRBs analysed after the removal GW170817 from the sample. For GRBs with no trigger in the onsource window, we provide upper and lower limits on the p-value. The upper limit is a p-value of 1. The lower limit is the fraction of offsource trials that also had no trigger. The distribution lays within the 2σ range for a beta distribution, shown by the upper and lower dotted lines. The population consistency with a no-signal hypothesis was calculated using the weighted binomial test outlined in [cite abadie et al]. This test considers the lowest 5% of p-values in the population, weighted by their prior probability of detection based on the time and sky position of the GRB. This analysis did not include GW170817 as it is a confirmed detection. The combined p-value for the search is 0.3. Thus we find no significant evidence for a population of subthreshold signals.

As GRB170817A is known to have originated in NGC 4993 [cite paper], the distance to this GRB is approximately 43Mpc. For the other GRBs in our sample we calculated the 90% exclusion distance. This is the distance at which 90% of injected simulated signals are recovered with a greater coherent SNR than the loudest trigger in the onsource. The 90% exclusion distances are plotted in figure 4.3. The median 90% exclusion distance for the BNS waveform is 80Mpc, for an NS-BH waveform with generic spin is 105Mpc, and for NS-BH with aligned spin is 144Mpc. These values are slightly lower than in O1, which were 90Mpc, 139Mpc, and 150Mpc respectively, though this seems to just be due to chance as we analyse a relatively small number of GRBs.

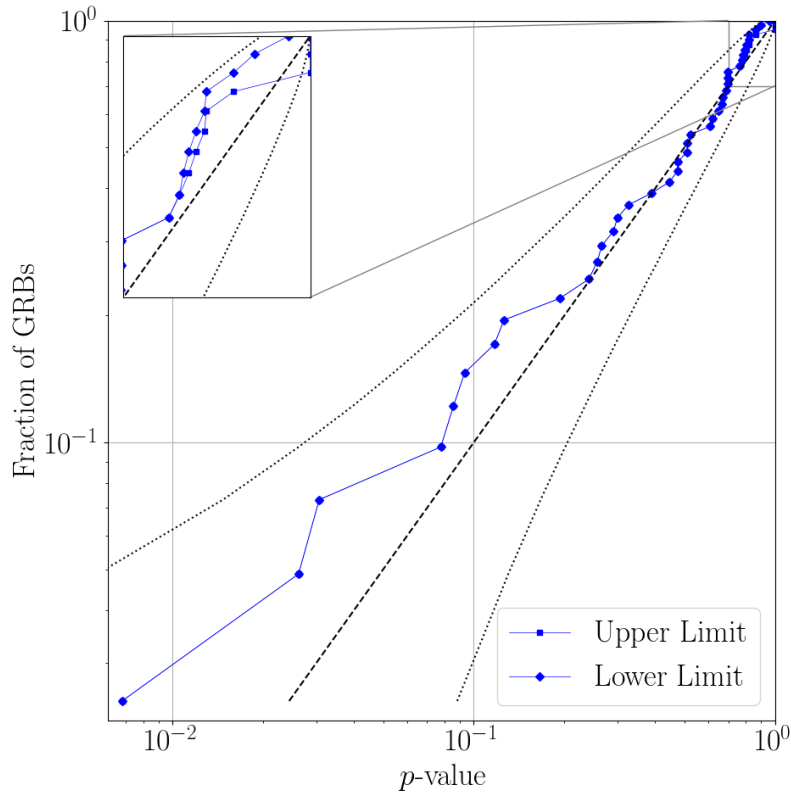


Figure 4.2: **P-value for each GRB.** This is the p-value distribution for the 41 GRBs other than GRB170817A. The GRBs with no trigger in the onsource window have upper and lower limits on the p-value. The upper limit is a p-value of 1. The lower limit is the fraction of offsource trials that also had no trigger. The distribution lays within the 2σ range for a beta distribution, shown by the upper and lower dotted lines.

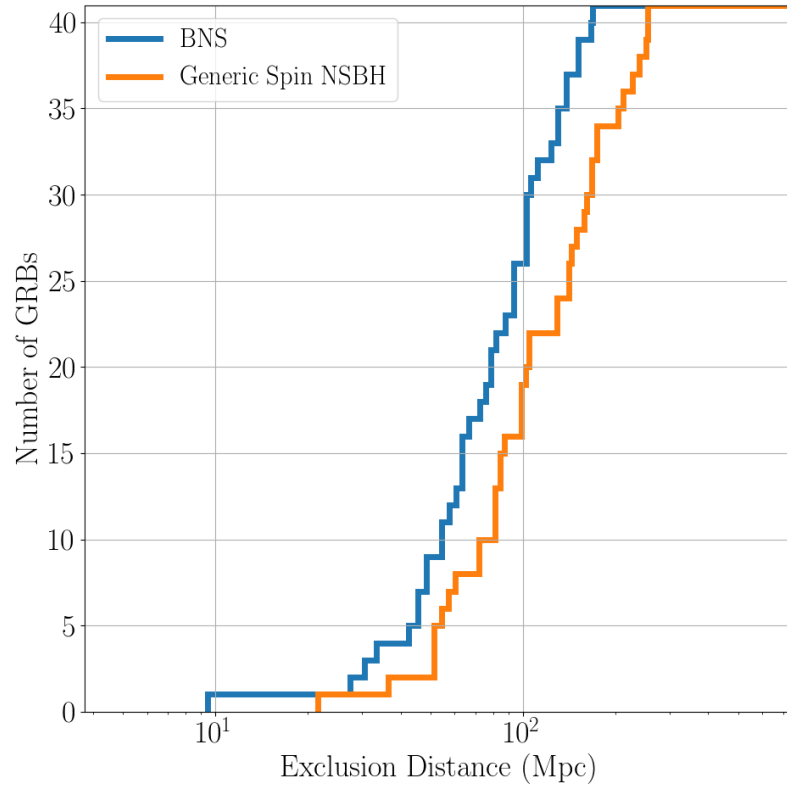


Figure 4.3: **Cumulative exclusion distance.** This is the cumulative 90% exclusion distance for every GRB analysed by PyGRB except GRB170817A. The 90% exclusion distance is the distance at which 90% of injected simulated signals are recovered with a greater coherent SNR than the loudest trigger in the onsource.

4.5 Improvements to PyGRB

- Made faster – matched filter engine and post processing
- allows new PyCBC stuff to be used
- Re-optimised the cuts and reweighted snr
- Ran on GRB with bigger sky grid than was previously possible or ran 170817 to be 5 sigma

4.6 Future Plans

- More detectors strengthens the case for coherent search
- Hierarchical search, using coincident search to find location
- Include astrophysical priors in the detection statistic

Chapter 5

Machine Learning for GW Astronomy

5.1 Introduction

The Laser Interferometer Gravitational Wave Observatory (LIGO) and Virgo collaborations operate ground-based Gravitational Wave (GW) detectors. They have detected signals of astrophysical origin, including the merger of a Binary Neutron Star (BNS) system and multiple Binary Black Hole (BBH) merger signals. As these sources have a known signal morphology, they can be found using a highly sensitive matched-filter search, as discussed in chapter ???. But some of the most interesting possible sources, such as core collapse supernova, do not have a known waveform morphology. For this reason, it is important to develop unmodelled searches as well.

In unmodelled searches we look for coherence between the data streams of multiple interferometers. We will consider the case of a Burst search where the sky position of the candidate source is known. This allows us to calculate the relative time of arrival in each detector and the relative signal power in each detector as well.

In this chapter, we will first discuss how an existing targeted Burst search, called *X-pipeline*, searches for Gravitational Wave Bursts (GWBs). We will then see how we can improve this pipeline by using *Multivariate Analysis* (MVA) to apply cuts and rank triggers.

5.2 Xpipeline

Xpipeline is a data analysis pipeline for the Burst search. It begins by using the known sky location of the GRB to time-shift the data from each detector so that the GW arrives simultaneously in each data stream. Xpipeline then makes coherent data-streams from the individual detector data. There are two types of coherent data-streams: *signal streams*, which increase the power of a GW signals, and *null*

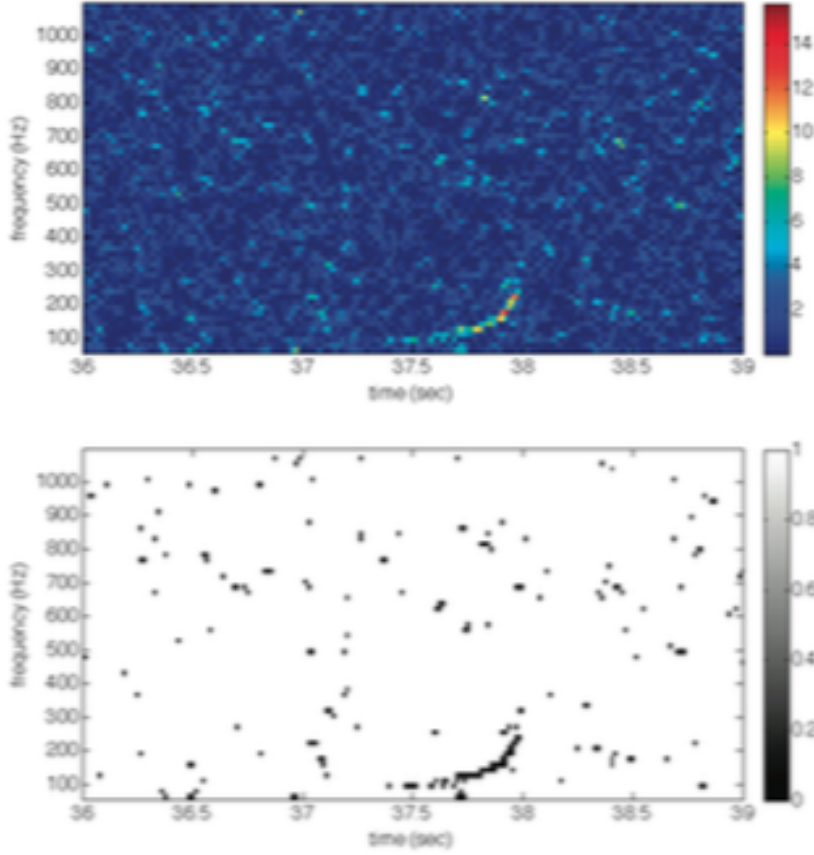


Figure 5.1: **Xpipeline Time-Frequency Map** This figure shows a time-frequency map from Xpipeline for a $1.4 - 10M_{\odot}$ NSBH merger. The top figure shows the E_+ energy and the bottom figure shows the top 1% of pixels.

streams, which reduce the power of true GW signals but not noise. *Triggers* are groups of neighbouring pixels that are above threshold in a time-frequency map of a signal stream (see fig.5.1). Xpipeline then removes triggers based on cuts of the various network data streams (see fig.5.2). The position of these cuts is set to give the best performance on a subset of the triggers that are used for tuning. For the MVA pipeline, the position of these cuts is instead set by a machine learning algorithm (as we will see in 5.3).

In this section, we will formulate two of the coherent data-streams generated by Xpipeline, as illustrative examples. The first is a signal-stream, called the *standard likelihood*. We will then use the standard likelihood to create a null-stream.

5.2.1 Burst Search Background

Suppose we have a network of D detectors. A gravitational wave, described by $h_+(t)$ and $h_{\times}(t)$, passes through the Earth from direction $\hat{\Omega}$. We describe the sensitivity of detector $\alpha \in \{1, \dots, D\}$ to the plus and cross polarizations using the *Antenna*

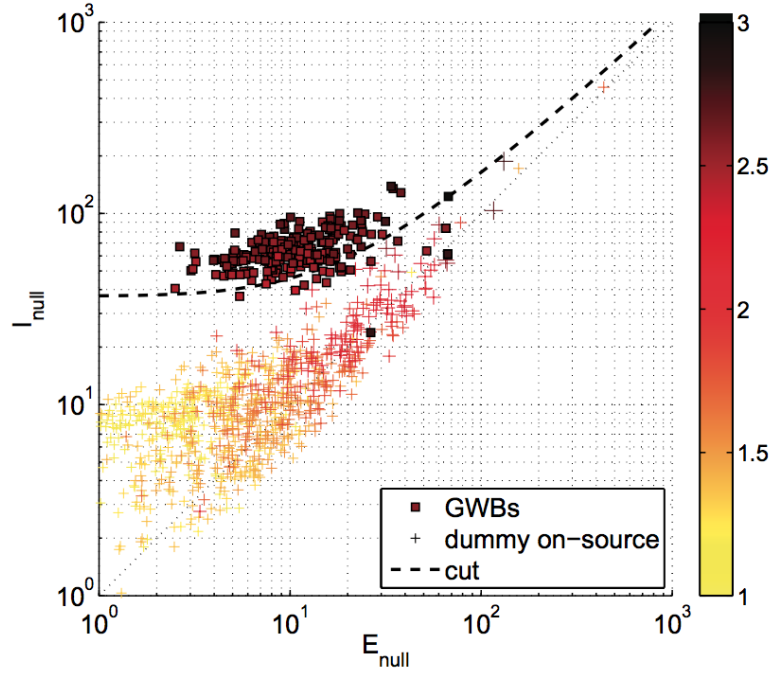


Figure 5.2: **Xpipeline Cut** This figure shows Xpipeline making a cut to eliminate many spurious signals.

Response Functions, denoted $F_{\alpha}^{+}(\hat{\Omega})$ and $F_{\alpha}^{\times}(\hat{\Omega})$. The position of detector α is denoted by \mathbf{r}_{α} and the noise in this detector is given by $n_{\alpha}(t)$. The detector output $d_{\alpha}(t)$ is then given by

$$d_{\alpha}(t + \Delta t_{\alpha}(\hat{\Omega})) = F_{\alpha}^{+}(\hat{\Omega})h_{+}(t) + F_{\alpha}^{\times}(\hat{\Omega})h_{\times}(t) + n_{\alpha}(t + \Delta t_{\alpha}(\hat{\Omega})) . \quad (5.1)$$

Here Δt_{α} is the time taken for the GW to reach the detector from some arbitrary reference point¹ \mathbf{r}_0

$$\Delta t_{\alpha}(\hat{\Omega}) = \frac{1}{c}(\mathbf{r}_0 - \mathbf{r}_{\alpha}) \cdot \hat{\Omega} . \quad (5.2)$$

From now on we will suppress explicit mention of the reference point \mathbf{r}_0 or the time delay Δt_{α} on the understanding that detector outputs need to be time-shifted by an appropriate amount.

In reality, detector outputs are not continuous but sampled discretely. The discrete Fourier transform $\tilde{x}[k]$ of the time series $x[j]$, and its inverse, are given by

$$\tilde{x}[k] = \sum_{j=0}^{N-1} x[j]e^{-i2\pi jk/N}, \quad x[i] = \frac{1}{N} \sum_{j=0}^{N-1} \tilde{x}[k]e^{i2\pi jk/N} . \quad (5.3)$$

For sampling rate is f_s and N data points in the time domain, we convert continuous

¹The center of the Earth is a fairly intuitive choice for a worldwide detector network.

to discrete notation by using

$$x(t) \rightarrow x[j] \quad (5.4)$$

$$\tilde{x}(f) \rightarrow f^{-1} \tilde{x}[k] \quad (5.5)$$

$$\int dt \rightarrow f_s^{-1} \sum_j \quad (5.6)$$

$$\int df \rightarrow f_s N^{-1} \sum_k \quad (5.7)$$

$$\delta(t - t') \rightarrow f_s \delta_{jj'} \quad (5.8)$$

$$\delta(f - f') \rightarrow N f_s^{-1} \delta_{kk'} . \quad (5.9)$$

For example, the one-sided noise spectral density for a detector with noise $n(t)$ can be written in continuous form as

$$\langle \tilde{n}^*(f) \tilde{n}(f') \rangle = \delta(f - f') \frac{1}{2} S_n(f) \quad (5.10)$$

where the angle brackets indicate an average over the noise. In the discrete notation listed above, this becomes

$$\langle \tilde{n}_\alpha^*[k] \tilde{n}_\beta[k'] \rangle = \frac{N}{2} \delta_{\alpha\beta} \delta_{kk'} S_\alpha[k] . \quad (5.11)$$

We will be working with the *noise-spectrum-weighted* quantities, defined by

$$\tilde{d}_{w\alpha}[k] = \frac{\tilde{d}_\alpha[k]}{\sqrt{\frac{N}{2} S_\alpha[k]}} \quad (5.12)$$

$$\tilde{n}_{w\alpha}[k] = \frac{\tilde{n}_\alpha[k]}{\sqrt{\frac{N}{2} S_\alpha[k]}} \quad (5.13)$$

$$F_{w\alpha}^{+, \times}(\hat{\Omega}, k) = \frac{F_\alpha^{+, \times}(\hat{\Omega})}{\sqrt{\frac{N}{2} S_\alpha[k]}} \quad (5.14)$$

The normalisation of the whitened data is

$$\langle \tilde{n}_\alpha^*[k] \tilde{n}_\beta[k'] \rangle = \delta_{\alpha\beta} \delta_{kk'} . \quad (5.15)$$

In vector notation, we can write 5.1 as

$$\tilde{\mathbf{d}} = \mathbf{F} \tilde{\mathbf{h}} + \tilde{\mathbf{n}} \quad (5.16)$$

where $\mathbf{F} = [\mathbf{F}^+ \quad \mathbf{F}^\times]$ and $\tilde{\mathbf{h}} = [\tilde{h}_+ \quad \tilde{h}_\times]^T$.

5.2.2 Standard Likelihood

Assume that the noise in our detectors is Gaussian. As a gravitational wave $\tilde{\mathbf{h}}$ passes through the detector from a known direction, the probability of attaining whitened detector output $\tilde{\mathbf{d}}$ in one time-frequency pixel is given by

$$P(\tilde{\mathbf{d}}|\tilde{\mathbf{h}}) = \frac{1}{(2\pi)^{D/2}} \exp \left[-\frac{1}{2} \left| \tilde{\mathbf{d}} - \mathbf{F}\tilde{\mathbf{h}} \right|^2 \right]. \quad (5.17)$$

For a set $\{\tilde{\mathbf{d}}\}$ of N_p time-frequency pixels, we have

$$P(\{\tilde{\mathbf{d}}\}|\{\tilde{\mathbf{h}}\}) = \frac{1}{(2\pi)^{N_p D/2}} \exp \left[-\frac{1}{2} \sum_k \left| \tilde{\mathbf{d}}[k] - \mathbf{F}[k]\tilde{\mathbf{h}}[k] \right|^2 \right]. \quad (5.18)$$

By comparing this value to the probability that the detector produces this output in the absence of any GW signal, we can calculate a likelihood of the signal being a real GW signal.

The *Likelihood Ratio* L is the log of the probability that the detector network will have output $\tilde{\mathbf{d}}$ in the presence of GW $\tilde{\mathbf{h}}$ divided by the probability of obtaining the same output in the absence of a gravitational wave ($\tilde{\mathbf{h}} = 0$)

$$L = \ln \frac{P(\{\tilde{\mathbf{d}}\}|\{\tilde{\mathbf{h}}\})}{P(\{\tilde{\mathbf{d}}\}|\{0\})} = \frac{1}{2} \sum_k \left[\left| \tilde{\mathbf{d}} \right|^2 - \left| \tilde{\mathbf{d}} - \mathbf{F}\tilde{\mathbf{h}} \right|^2 \right]. \quad (5.19)$$

For the above analysis to be applied, we would need to know the waveform $\tilde{\mathbf{h}}$ in advance. For GRBs and other unmodelled searches, this is not possible. One way to handle this problem is to fit the waveform in each time-frequency pixel to the data in such a way as to maximise the likelihood ratio. Hence we have

$$0 = \frac{\partial L}{\partial \tilde{\mathbf{h}}} \Big|_{\tilde{\mathbf{h}}=\tilde{\mathbf{h}}_{\max}}. \quad (5.20)$$

Solving this, we find

$$\tilde{\mathbf{h}}_{\max} = (\mathbf{F}^\dagger \mathbf{F})^{-1} \mathbf{F}^\dagger \tilde{\mathbf{d}} \quad (5.21)$$

where the superscript dagger † denotes the conjugate transpose.

Calculating the likelihood ratio for $\tilde{\mathbf{h}}_{\max}$ gives us the *Standard Likelihood* E_{SL}

$$E_{\text{SL}} = 2L(\tilde{\mathbf{h}}_{\max}) = \sum_k \tilde{\mathbf{d}}^\dagger \mathbf{P}^{\text{GW}} \tilde{\mathbf{d}} \quad (5.22)$$

where

$$\mathbf{P}^{\text{GW}} \equiv \mathbf{F}(\mathbf{F}^\dagger \mathbf{F})^{-1} \mathbf{F}^\dagger. \quad (5.23)$$

We can see from equation 5.1 that the contribution made to the data output by a passing GW from fixed sky location is restricted to the subspace spanned by \mathbf{F}_+

and \mathbf{F}_\times . Therefore the energy in this subspace is the energy that is consistent with a GW from a given sky location. We can show that \mathbf{P}^{GW} is a projection operator, projecting the data into this same subspace. The standard likelihood maximises the energy in this subspace, and so is the maximum energy contained in the whitened data that is consistent with a GW from the given sky location.² This is an example of the coherent signal-streams that Xpipeline uses.

5.2.3 Null Energy

We can use the standard likelihood to find a null-stream. First consider the *total energy*, given by

$$E_{\text{tot}} = \sum_k |\tilde{\mathbf{d}}|^2. \quad (5.24)$$

This is an incoherent statistic as it contains only auto-correlation terms, and no cross-correlation terms. If we subtract the standard likelihood from the total energy, we obtain the *null energy*

$$E_{\text{null}} \equiv E_{\text{tot}} - E_{\text{SL}} = \sum_k \tilde{\mathbf{d}}^\dagger \mathbf{P}^{\text{null}} \tilde{\mathbf{d}}. \quad (5.25)$$

This is the energy that is inconsistent with a GW from given sky location, and must therefore be associated with noise. This is the minimum amount of energy in our whitened data that is inconsistent with our GW. It is an example null-stream used by Xpipeline.

This shows one of the key advantages of coherent analysis. If we analysed our data incoherently, we would be working with just the total energy. But using coherent methods, we can project the whitened data into the subspace spanned by \mathbf{F}^+ and \mathbf{F}^\times , thus removing some fraction of the noise without removing any of the signal. The drawback is that if the sky position is not known in advance, then the analysis needs to be repeated for a set of directions that span the entire sky ($\gtrsim 10^3$ directions), each with different antenna response functions \mathbf{F}^+ and \mathbf{F}^\times . This will slow down analysis and increase the false alarm probability.

5.2.4 Incoherent Energy and Background Rejection

The diagonal elements of 5.25 are auto-correlation terms, and the other elements are cross-correlation terms. The auto-correlation part of the null energy are called the *incoherent energy*, and denoted by

$$I_{\text{null}} = \sum_k \sum_\alpha P_{\alpha\alpha}^{\text{null}} |\tilde{d}_\alpha|^2. \quad (5.26)$$

²In practice, some of this energy will be due to noise. We say that the energy due to noise is *inconsistent* with a GW signal. The rest of the energy must be due to the GW, so we say it is *consistent* with the GW.

Background triggers are typically not correlated between the different detectors of the network, so the cross-correlation terms are small relative to the auto-correlation terms. This means that for glitches, we have

$$E_{\text{null}} \approx I_{\text{null}} . \quad (5.27)$$

Compare this to the case of a GW signal. This will be correlated between the detectors. By construction, the energy of the correlated energy does not appear in the null-stream. Therefore, in the presence of a strong GW, the incoherent energy is much larger than the null energy

$$E_{\text{null}} \ll I_{\text{null}} . \quad (5.28)$$

Using 5.27 and 5.28, we see that the ratio of E_{null} and I_{null} is very different in the case of a glitch as opposed to a GW signal. We can use this to make the following cut to remove noise triggers from our sample

$$I_{\text{null}}/E_{\text{null}} > C \quad (5.29)$$

for some constant $C > 1$. This test does not work as well for small amplitude glitches, where the statistical fluctuations can lead to E_{null} being smaller than I_{null} . For this reason, Xpipeline varies C with the size of the trigger. The precise position of the cut is set to maximise performance on a set of injections.

5.3 Machine Learning

We have seen how Xpipeline makes cuts on coherent statistics to distinguish between noise and GW signals. You can see from 5.29 that the cut only uses two statistics, the null energy and the incoherent energy. Xpipeline generates many coherent statistics that could be used in conjunction to make a more powerful cut. In this section we discuss how to use Machine Learning to make achieve this. The software we use is the Toolkit for Multivariate Analysis package in the ROOT data analysis framework.

5.3.1 Supervised Machine Learning

The type of Machine learning we use is called *supervised machine learning*. Supervised machine learning algorithms are trained on data which has already been classified. They can then be shown a new, not-classified data point and determine the appropriate classification. Supervised machine learning requires data to be in a particular format (see table 5.1). It is a list of *events*, each with a *label* and corresponding *attributes*. The machine learning algorithm builds a *classifier* that can determine the label of an event when given the event's attributes. In the case of our GW search, the events are the triggers returned by Xpipeline (see section 5.2). The

Label	loghbayesiancirc	standard	circenergy	circinc
Background	12.3128	58.3523	44.7196	24.9015
Background	12.0349	67.5344	41.7045	22.4848
Signal	18.2145	59.8136	53.3320	22.0601
Signal	43.7113	123.9194	118.9774	43.9234
Signal	6422.1467	14426.9124	14167.2933	4991.7876

Table 5.1: An example of MVA training data. Each event has a label and several attributes.

labels are *signal* or *background*, and the attributes are the values of the signal and null data streams for those triggers as well as some statistics describing the trigger, e.g. peak frequency, trigger duration. A full list of the attributes used, together with a short description of the attribute can be seen in table 5.2. The signal triggers are generated by injecting signal waveforms into the data. Background triggers are triggers that do not coincide with an injected signal.

The signal and background trigger sets are each divided into two subsets: The *training set* and the *testing set*. The training set is used to build the classifier, while the testing set is used to measure the accuracy of the trained classifier. If the classifier performs much better on the training set than on the testing set, then the classifier is *overtrained*. This means that the classifier has learned exactly the properties of signals and noise in the training set, rather than learning the general properties of signals and noise. When an overtrained classifier is used on a different data set with different noise properties (such as the testing set), it will perform poorly.

Coherent Energy	Description
loghbayesiancirc	A likelihood ratio based on Bayesian methods, for the hypothesis of a circularly polarised GW vs Gaussian noise
E_{max}	The maximum energy in the whitened data that is consistent with a GW from a given sky location.
E_{null}	The minimum amount of energy in the whitened data that is inconsistent with a GW from a given sky location. Given by $E_{\text{tot}} - E_{\text{max}}$.
I_{null}	The sum of the autocorrelation terms of E_{null} .
E_{circ}	The maximum energy in the whitened data consistent with a circularly polarised GW from a given sky location.
I_{circ}	The sum of the autocorrelation terms of the E_{circ} .
E_{circnull}	The energy in the whitened data that is inconsistent with a circularly polarised GW from a given sky location. Given by $E_{\text{tot}} - E_{\text{circ}}$.
I_{circnull}	The sum of the autocorrelation terms of E_{circnull} .
E_{H1}	The energy in the H1 interferometer.
E_{L1}	The energy in the L1 interferometer.
E_{V1}	The energy in the V1 interferometer.
number of pixels	The number of pixels in the cluster.
Duration	Time duration of the trigger.
Bandwidth	The frequency range spanned by the trigger.
Power law	

Table 5.2: The attributes used by the machine learning classifier.

5.3.2 Boosted Decision Trees

A *Decision Tree* is a simple type of classifier. It is a flowchart of true/false statements about a trigger's attributes to determine the correct label. For example, consider the decision tree shown in fig.5.3. Here the attributes are labeled as the components of a vector x . We start at the top node and work downwards. If the node is true then we follow the branch to the left and if not then we follow the branch to the right. We then consider the statement at the end of whichever branch we follow. We continue this process until we reach a *leaf node*, which has no branches and contains a final classification for the trigger.

We can improve the performance of the classifier by using an *ensemble* (or *forest*) of trees. This means training multiple distinct trees, and each tree independently classifies the trigger. The final classification of each trigger is a normalised sum of the outputs of each tree, with +1 corresponding to signal, and -1 corresponding to background. This leads to different regions of the parameter space having different MVA scores, as can be seen in fig.5.4. The higher the score, the more likely an event is to be a signal.

To train a decision tree, we need methods to find the variable to cut on and the position of the cut for each decision node, and the label in each leaf node that best discerns between signal and background. Each of these values is set by brute force: Trying each possible combination of cuts and labels to get the best performance on the training events. Ensemble methods work best when each classifier in the ensemble is independent of the others, so training every tree on the same events is not going to give optimal results. For this reason each tree is trained on some subset of the training set. We could pick events at random to form these subsets, but a more powerful method is to use *Adaptive Boosting*.

When using adaptive boosting, each event in the training set is given a probability that it will be selected to train the next tree. Initially each event in the training set has the same probability of being selected. Then after each tree is trained the probability of each event being selected to train the next tree is updated such that misclassified events are more likely to be included in the training set for the next tree. The misclassified events have their probability of being selected for training the next tree updated by the *boost weight*, given by

$$\alpha = \frac{1 - \text{err}}{\text{err}} \quad (5.30)$$

where err is the misclassification rate. The weights are then renormalised.

Note that the boost weight is greater than one for³ $\text{err} < 1/2$, and that as the error gets smaller, the boost weight increases. The effect of this is that each new tree is more likely to be trained on the events that are most difficult to classify.

³The misclassification rate is always less than half when the labels only take two possible values as the algorithms we use are always better than chance alone.

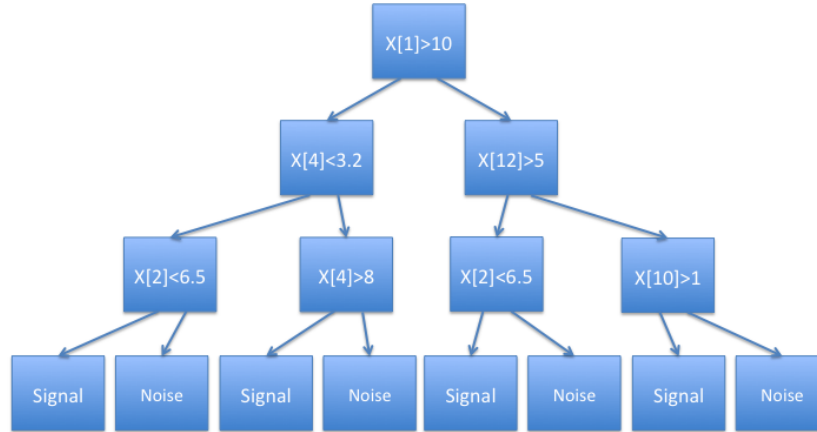


Figure 5.3: **A decision tree example** To determine if a trigger is a signal or noise event the tree makes a series of cuts on the attributes $x[N]$. If the inequality in a node is true, then the next node is the branch to the left. Otherwise the next node is the one to the right.

The ensemble output is also changed, so that it is weighted rather than being a simple sum. If the output of the i th tree is given by $h_i(\mathbf{x})$, with \mathbf{x} being the event attributes, then the ensemble output is given by

$$H(\mathbf{x}) = \frac{1}{N} \sum_{i=0}^N \ln(\alpha_i) h_i(\mathbf{x}) \quad (5.31)$$

where N is the number of tree in the ensemble. Small values indicate background events, while large values indicate a signal.

5.3.3 Data Preprocessing

Extra data preprocessing is required for the training set. Suppose a small amplitude signal injection is added close to a large amplitude glitch in the data. This trigger will be labeled as a signal, due to the injection, but the properties of the trigger will resemble a glitch, as the glitch has a much larger amplitude than the signal injection. It is essentially a background trigger labeled as a signal. This has two affects: It reduces our ability to detect real signals by making the properties of signals harder for the algorithm to learn, and it can lead to background triggers being misclassified as a Gravitational Wave signal.

For these reasons, it is important to make sure that signal injections do not overlap in time or frequency with background triggers. To prevent this from happening

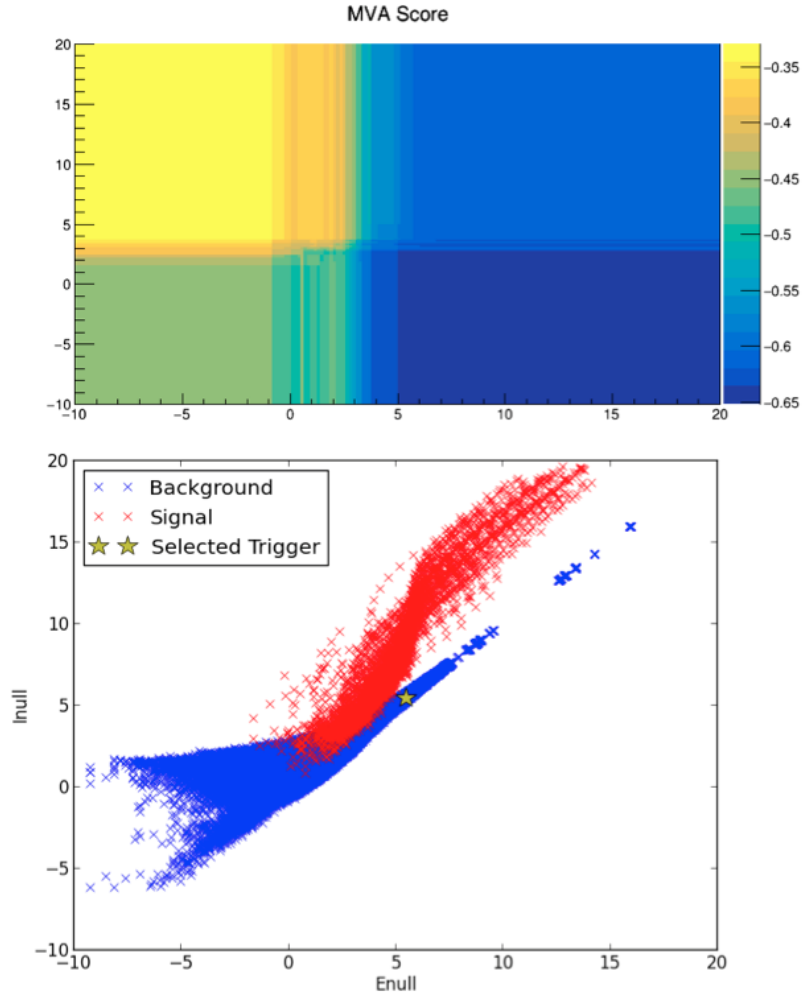


Figure 5.4: **Visualising the Classifier** In the bottom plot you can see the value for $\log(\text{Enull})$ and $\log(\text{Inull})$ for all the signal and background training data used to build the classifier. We chose one of these events at random (indicated by the star) and varied Enull and Inull to see how it changed the MVA score, indicated by the colour in the top plot. As you can see, increasing Inull and decreasing Enull leads to the event being more likely to be classed as a signal. This is akin to the xpipeline cut shown in fig.5.2.

we *clean* the data. This means finding all of the triggers in the data for the smallest signal injection scale. The injected waveforms are too small to be detected so all of these triggers must be background. We then increase the injection amplitude and look for triggers again. Any triggers that overlap in time or frequency with the noise triggers are then identified and removed from the signal set. **This description of cleaning may need to change if we change the cleaning codes.**

We must also not include injections in our signal set that are too small to be detected. If we do then we would again be including triggers labeled as signal that have the properties of a noise trigger, even if there is no glitch present. For this reason we apply a threshold on the amplitude of the signal set, so any injection below that amplitude is removed. The level of the threshold is set experimentally; If it is too low then we will increase the chance of a false positive and hurt our sensitivity to real signals, but if it is too high then we will limit the classifiers ability to detect low amplitude signals.

We also use *Generalised Clustering*. This is a change in the way Xpipeline defines a trigger. Without using generalised clustering, triggers are groups of neighbouring pixels in the time-frequency maps such as in fig.5.1. Generalised clustering allows these pixels to be separated by a user-specified number of pixels. This can boost the signal power of long triggers, as long triggers tend to contain breaks in their time-frequency maps that causes Xpipeline to list a single long trigger as multiple short triggers. The downside to using generalised clustering is that it can cause noise triggers to be grouped together as well, boosting the power of noise. This then causes our sensitivity to short triggers to be reduced slightly. Experimenting with generalised clustering in both the standard Xpipeline analysis and the MVA analysis suggests that the benefits outweigh the costs, with an improvement of between 7% and 50% for long inspiral and adi waveforms at the cost of a decrease of 3% to 9% for short sine-gaussian waveforms.

As XTMVA is to be used for the unmodelled search, it is also important to ensure that the search can find waveforms that are not in the training set. There are several tools that we use to achieve this. The first is to limit the amount of information the classifier is given about the waveform morphology. The classifier cannot know the precise morphology of any waveform because the only attributes that the classifier trains on is the time, peak-frequency, and various measurements of the coherent and null energy between the detectors in the network. This forces the classifier to use the coherence of the triggers to make a classification, rather than the waveform morphology.

There is still a possibility that the classifier will become too specialised to the waveforms in the training set, as the classifier is given peak frequency data and certain waveform morphologies may have particular characteristics than become apparent from the coherent and null streams. For this reason we also trained the

classifier on a variety of different waveforms. The training set included long and short waveforms, and a variety of different morphologies. Some of the signals are astrophysically motivated, such as compact binary coalescence signals, while some are artificial, such as the white noise burst.

The final tool we use to ensure the classifier is sensitive to generic waveforms is to test the classifier on waveforms that are not included in the training set. If the classifier can find waveforms not in the training set, then we can be reasonably confident that it is sensitive to generic waveforms. We also try removing certain waveforms from the training set to test the robustness of the classifier. This will lead to a drop in sensitivity for that waveform, but if the drop in sensitivity is small, then we can be confident that the classifier is robust.

5.3.4 Optimisation and Validation

With any machine learning algorithm, there are *hyperparameters* that must be tuned. The number of trees in the ensemble, the maximum depth of the trees, and many more hyperparameters must be chosen. We optimise the hyperparameters by repeatedly running an example analysis with different hyperparameters, trying to increase the number of injected signals we could recover from the testing set.

Setting our hyperparameters using the testing set can cause *data leakage*. Data leakage is when data from outside the training set is used to build a classifier. As we tune our classifier on the testing set, it is possible that we will implicitly tune our classifier to only work well on this training and testing set. This is very similar to overtraining discussed in section 5.3. To avoid this, once we have tuned the hyperparameters on a single GRB, we test the classifier on several other GRBs. If the performance drops significantly on these other GRB analyses, then we have had data leakage and we need to retune our classifier. This process of testing on previously unused data is called *validation*. If there is evidence of data leakage, then we must retune our classifier and validate it again, this time using a different (previously unused) GRB so as to avoid data leakage from the validation GRB.

Optimisation is a somewhat cyclical process, as once we have changed one hyperparameter, we must then go back and test that other parameters do not now need changing. But optimisation is not pure guess work, and in the rest of this section we will see the strategy we used to optimise our classifier. We first discuss optimising the training data to use. As mentioned in section 5.3.3, there are several choices to make regarding what data is used for training, such as setting the amplitude threshold. We then discuss optimising the BDT classifier itself.

Training Set

Many of the hyperparameters mentioned in section 5.3.3 had to be optimised. Consider the amplitude threshold applied to triggers before they are included in the

training set. If this is too low then noise triggers contaminate the signal training set, but if it is too high then we limit the sensitivity of our classifier to only higher amplitude waveforms.

We also optimised the waveforms that are included in the training set, whether to use cleaning or not, and whether to use generalised clustering. **Write more on this once we've done it.**

- Had to optimise the threshold as well
- Training waveforms
- cleaning or not
- Gen clustering or not

BDT Parameters

There are many hyperparameters that need to be set for a BDT analysis. In this section we discuss some of these hyperparameters and the method we used to optimise them for our analysis.

We began by setting values for $NTrees$, the number of trees in the ensemble, and the learning rate, discussed in section 5.3.2. These two parameters are set first to ensure the machine learning algorithm will converge in a reasonable amount of time. Setting the learning rate too low causes the classifier to take longer to converge. Setting the learning rate too high can cause the classifier to never converge. Similarly, using too many trees takes too long for the classifier to finish training, but too few and the training will terminate before the algorithm has converged. Setting these first ensures that we have a classifier that gives sensible results in a reasonable amount of time. While optimising, we set these values slightly high, to ensure that our classifier converges and does so quickly while we optimise our other parameters. Once the other parameters are set, we again optimise $NTrees$ and the learning rate, increasing the number of trees and decreasing the learning rate to ensure the algorithm reaches its optimum performance, even if it increases the time taken for training.

We now tune the tree-specific parameters. Unlike the number of trees and the learning rate which are primarily tuned to ensure the classifier will converge in a reasonable time, these parameters are set to ensure that the classifier is accurate but does not overtrain. Overtraining can happen when the trees are allowed to make cuts that are too fine, carving out regions of parameter-space around anomalous events in the training set rather than finding cuts that generalise beyond the training data. The way to prevent this is to limit how fine the cuts made by the decision trees are allowed to be, while allowing cuts that are fine enough to pick out the general features of signal and background events in our data. There are several hyperparameters we can set to do this, which must all be tuned.

The first of these is the maximum depth of the trees, which is the maximum number of cuts a tree can make before it reaches a leaf node. Each cut divides the parameter-space into ever smaller regions which it labels as background or signal. Setting the maximum number of cuts too low will therefore cause the classifier to be too coarse in dividing up the parameter-space, resulting in poor accuracy. Increasing the maximum depth allows the classifier to pick out smaller features in the parameter-space. If the maximum depth is too high then the classifier will overtrain; dividing the parameter-space into precisely the regions that work for the training set and losing generality. As we are using adaptive boosting, it is recommended **cite TMVA userguide** to use trees with fewer cuts. For this reason we tried values from 2-16, finding that for our problem a maximum depth of 8 was optimal.

A related hyperparameter is the minimum number of events that we allow in a leaf node. If we allow the training algorithm to have any number of events in a leaf node, then it will occasionally find cuts that results in a small number of events in one or more of the leaf nodes. This again allows for the carving out of anomalous events in the training set rather than finding cuts that can generalise beyond the training set. Conversely, setting the minimum number of events allowed in the leaf nodes to be too high does not allow the classifier to pick out the key features in the data. We used a grid search over the values 100-1600 for the minimum number of events per leaf node and found the optimal value to be 400.

The final hyperparameter we set is the number of cuts that the training algorithm scans over to find the best cut. When the classifier is training it searches for the best way to cut the parameter-space into a signal subspace and a background subspace. To do this we can try every possible cut on every parameter. This can be very slow and lead to overtraining. To speed up our analysis and reduce the chance of overtraining, we can choose the number of cuts to try on each parameter. For example, we may decide to use 10 cuts for each parameter. In this case the algorithm will tune the cut on a parameter which ranges from 0-100 by trying cuts at 0,10,20,... and selecting the best of these cuts. To tune the number of cuts we wanted to use, we tried values from 10-160 as well as allowing the classifier to try every possible cut. We found that 80 cuts was optimal for our purposes.

- This can definitely be phrased better

Result of Optimisation

- Show nice plots of improved sensitivity due to optimisation
- Show improvement when compared to Xpipeline

