DISTINGUISHING SIGNAL FROM NOISE: NEW TECHNIQUES FOR GRAVITATIONAL WAVE DATA ANALYSIS

by

Paul Thomas Baker

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Physics

MONTANA STATE UNIVERSITY
Bozeman, Montana

August 2013

APPROVAL

of a dissertation submitted by

Paul Thomas Baker

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to The Graduate School.

Dr. Neil J. Cornish

Approved for the Department of Physics

Dr. Richard J. Smith

Approved for The Graduate School

Dr. Ronald W. Larsen

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

Paul Thomas Baker

August 2013

PREFACE

The work presented in this dissertation was carried out within the LIGO Scientific Collaboration (LSC). The methods and results presented here are under review and are potentially subject to change. The opinions expressed here are those of the author and not necessarily those of the LSC. The author gratefully acknowledges the support of the National Science Foundation for the construction and operation of the LIGO Laboratory, which provided support for this work.

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

CONVENTIONS

Throughout this work we will use geometrical units, where $G = c = 1$, unless stated otherwise. The sign conventions related to the space-time metric will follow Carroll [15], using the signature $(-1, 1, 1, 1)$.

Tensors will be denoted by bold face: $\mathbf{R}$, and four-vectors will be denoted by italicized bold face: $\boldsymbol{x}$. Purely spacial three-vectors will be denoted by the vector symbol: $\vec{x}$.

The components of tensors and vectors will be labeled with greek indices running from 0-3 or roman indices running 1-3 for the spacial part only. The roman letters a, b will be reserved to run 1-2 for exclusive use in the transverse-traceless gauge. The indices x, y, z will be reserved to specify components in cartesian coordinates. Other common coordinates, like t, will similarly label specific components. We will follow the Einstein summation convention for all indices, where repeated raised and lowered indices imply summation:

$$\xi^{\mu}\,\xi_{\mu} = \sum_{\mu=0}^{3} \xi^{\mu}\,\xi_{\mu}.$$

Coordinate derivatives will be denoted by ∂ or a comma:

$$\frac{\partial \xi^{\mu}}{\partial x^{\nu}} = \partial_{\nu}\,\xi^{\mu} = \xi^{\mu}{}_{,\nu},$$
$$\frac{\partial \xi^{\mu}}{\partial t} = \partial_{t}\,\xi^{\mu} = \xi^{\mu}{}_{,t}.$$

Covariant derivatives with respect to the metric will be denoted by ∇ or a semicolon:

$$\frac{\partial \xi^{\mu}}{\partial x^{\nu}} + \xi^{\lambda}\,\Gamma^{\mu}{}_{\lambda\nu} = \nabla_{\nu}\,\xi^{\mu} = \xi^{\mu}{}_{;\nu}.$$

ABSTRACT

The principal problem of gravitational wave detection is distinguishing true gravitational wave signals from non-Gaussian noise artifacts. We describe two methods to deal with the problem of non-Gaussian noise in the Laser Interferometer Gravitational Observatory (LIGO).

Perturbed black holes (BH) are known to vibrate at determinable quasi-normal mode frequencies. These vibrational modes are strongly excited during the inspiral and merger of binary BH systems. We will develop a template based search for gravitational waves from black hole ringdowns: the final stage of binary merger. Past searches for gravitational waves developed *ad hoc* detection statistics in an attempt to separate the expected gravitational wave signals from noise. We show how using the output of a multi-variate statistical classifier trained to directly probe the high dimensional parameter space of gravitational waves can improve a search over more traditional means. We conclude by placing preliminary upper limits on the rate of ringdown producing binary BH mergers.

LIGO data contains frequent, non-Gaussian, instrument artifacts or glitches. Current LIGO searches for un-modeled gravitational wave bursts are primarily limited by the presence of glitches in analyzed data. We describe the BayesWave algorithm, wherein we model gravitational wave signals and detector glitches simultaneously in the wavelet domain. Using bayesian model selection techniques and a reversible jump Markov chain Monte Carlo, we are able determine whether data is consistent with the presence of gravitational waves, detector glitches, or both. We demonstrate BayesWave's utility as a data quality tool by fitting glitches non-Gaussian LIGO data. Finally, we discuss how BayesWave can be extended into a full-fledged search for gravitational wave bursts.

1. INTRODUCTION

Einstein's theory of gravitation, the General Theory of Relativity, was finalized in 1915 with the publication of the gravitational field equations. General relativity describes gravity as the curvature of space-time itself. It has withstood every test made of it, but one early prediction has eluded results despite years of work. Gravitational radiation, while indirectly detected via the measurement of energy loss of binary neutron star systems, has never been directly detected.

Gravitational radiation or gravitational waves are ripples in the fabric of space-time. Traveling at the speed of light and produced by a relatively generic time varying mass distribution, gravitational waves can provide a new vantage into astrophysical phenomenon. Where light, electro-magnetic radiation, is screened from view or distorted by matter in the cosmos, gravitational radiation travels nearly uninhibited. Primordial gravitational waves generated moments after the birth of the universe could exist to this day as a stochastic background much older than the Cosmic Microwave Background, the oldest light in the universe, originating 380,000 years after the big bang. Gravitational waves from tidally distorted neutron stars could provide insight into the equation of state of such objects. Gravitational waves generated by supermassive or intermediate mass black holes could constrain galactic formation models. Most excitingly, gravitational waves have the potential to detect new phenomenon completely unknown as of now. Gravitational waves will also allow for the

strongest tests of general relativity ever conducted. These tests will probe the extremely strong gravitational fields near the surfaces of black holes and neutron stars [62].

The detection of gravitational radiation is not a particularly new game. There are several experiments that over the years have attempted to measure gravitational radiation directly. The first such experiments used resonant bar detectors. These detectors look for induced resonant vibrations in a long cylindrical bar caused by passing gravitational waves. Such detectors are sensitive to only a narrow frequency band of gravitational radiation (near the bar's resonance frequency). Bar detectors never achieved the sensitivity to detect the truly minuscule effects of all but the rarest gravitational waves.

More recent endeavors focus on the precision timing of pulsars and laser interferometry. In the case of pulsar timing one looks for deviations in the expected time of arrival of regular radio bursts caused by intervening gravitational waves changing the distance between the pulsar and Earth. Interferometer experiments use the interference pattern of light to very precisely measure the differential length between two paths of travel. Again these deviations would be caused by gravitational waves changing the shape of space along these paths. Interferometer gravitational wave detectors are typically characterized by their arm length. The length of an interferometer sets a limit on the wavelength of gravitational waves that could be detected.

Figure 1.1. Aerial photos of the two LIGO observatories and the Virgo observatory. From the left: LIGO Hanford, LIGO Livingston, and Virgo. Photos courtesy the LIGO Laboratory.

The Laser Interferometry Gravitational Observatory (LIGO) is a terrestrial network of interferometers (IFO). In its initial configuration there where three IFOs: one 4 km IFO at the LIGO Livingston Observatory (LLO) in Louisiana, a second 4 km IFO at the LIGO Hanford Observatory (LHO) in Washington, and a second 2 km IFO at LHO. The LHO detectors share vacuum systems. The LIGO Scientific Collaboration also has a data sharing agreement with the 3 km Virgo detector located near Pisa, Italy.

LIGO is currently undergoing upgrades to its 'advanced' configuration. The two advanced detectors: one at each Hanford, WA and Livingston, LA, will be able to probe distances ten times larger than before. Using current rate estimates of astrophysical events, advanced LIGO should expect to detect at least several gravitational wave events per year [1]. In addition the Virgo detector is undergoing similar upgrades, and there are planned detectors in India (IndIGO) and Japan (Kagra).

Even with a network of gravitational wave detectors, direct detection has proven difficult. The difficulty of detecting gravitational waves is in some ways the focus of this work. We will discuss new methods for analyzing gravitational wave data

that improve the chances of detection. Each method addresses the central problem of gravitational wave data analysis for terrestrial interferometry: the non-stationary, non-gaussian detector noise. Most data analysis theory was developed assuming detector noise is stationary and gaussian. When this is not the case the theoretically optimal methods do not perform as well as expected.

We first address this problem by applying a sophisticated machine learning algorithm to a real gravitational wave search in LIGO data . After training the algorithm to distinguish gravitational wave events from random detector artifacts that trickle through our analysis pipeline, we can classify candidate events with a probability of being a true gravitational wave signal and compute a false alarm rate for each event. In our case of non-detection we then can place upper-limits on the astrophysical event in question.

Next we will attack the problem of non-gaussian noise directly. We will describe an algorithm to model non-gaussian noise and gravitational wave signals simultaneously. Using bayesian model selection, each event can be assigned a betting odds ratio that it is a gravitational wave. As a consequence, this method can be used to characterize and study the non-gaussian detector noise in and of itself.

This monograph will begin with a discussion of gravity, leading up to the indirect detection of gravitational waves in a binary neutron star system. In chapter 2 we will discuss sources of gravitational waves with a focus on the quasi-normal modes of vibration of black holes or black hole ringdowns. In chapter 3 we will discuss detecting

gravitational waves with the LIGO instrument, including some of its data analysis challenges. We will discuss a method for detecting gravitational waves from known sources in chapter 4, again focusing on black hole ringdowns. In chapter 5 we will look at how using a multi-variate statistical classifier can improve such searches.

We will take a brief interlude in chapter 6 to review Bayesian probability theory and the Markov chain Monte Carlo algorithm for drawing samples from a posterior distribution function. In chapter 7 we will introduce the BayesWave algorithm for detecting gravitational waves from unknown sources in non-stationary, non-gaussian noise. Next in chapter 8 we will demonstration of the efficacy of BayesWave under a variety of circumstances. Finally, we will conclude with a discussion of future work in chapter 9.

1.1. A Brief History of Gravity

People have known about gravity for as long as people have been. But throwing a spear and understanding the subtle forces at work are two very different labors. The Flemish mathematician and engineer Simon Stevin was one of the first to study gravity in detail. He proposed that all objects fall at the same rate regardless of their masses and demonstrated this by dropping lead balls out a church tower in 1586[53]. Galileo later championed these ideas and is often remembered as their originator.

Isaac Newton's theory of universal gravitation, which he described in his *Philosophiæ Naturalis Principia Mathematica* in 1687, combined the work of many who came

before him, including Galileo and Kepler, into one coherent theory. One aspect of his theory bothered many contemporaries, and even Newton himself: the idea of "action at a distance". According to Newton, the gravitational force acts instantaneously, influencing far flung bodies like the planets, without any direct physical interaction. Had Newtonian gravity not proven to be so wildly successful, it would have been dismissed by many natural philosophers of the time as no more sensical than angels pushing the planets.

Using his law of universal gravitation, Newton was able to derive Kepler's empirically defined laws of planetary motion. This effectively unified terrestrial and celestial mechanics under one set of simple laws. Newton also described the tides in detail, showing how the three body interaction of the Sun-Earth-Moon system results in the neap and spring tides. By the 19th century astronomers had confirmed Newton's laws to great precision. Some of the strongest support for Newtonian gravity came in 1846, when the French astronomer Urbain Le Verrier, using observations made by Alexis Bouvard, predicted the existence of Neptune. Le Verrier analyzed how Uranus' orbit differed from the predictions of Newtonian gravity and postulated that an as yet unseen planet was perturbing the motion of Uranus. Neptune was subsequently observed at the Berlin Fraunhofer refractor the same evening a letter from Le Verrier arrived, telling them where to point their telescope.

The first evidence against Newtonian gravity came in the 19th century. As astronomers mapped the solar system with greater and greater precision, one planet

defied Newton's mechanics. Mercury's elliptical orbit is observed to precess about 574 arc-seconds per century (0.16°); however, Newton's theory, even when accounting for perturbations from all of the planets in the solar system, only predicts 531 arc-seconds of precession [19]. These mere 43 arc-seconds per century (0.012°) could not be accounted for. Le Verrier, who performed the original measurements of Mercury's orbit, led the failed search to discover new mass in the solar system (like Neptune) to make up this small difference [36]. Finally, some proposed that Newton's theory must be incorrect, but no simple correction could fix Mercury's malfunction while leaving the rest of the solar system intact.

It is here where Albert Einstein entered the story. Einstein's new theory of gravity was not motivated by Mercury, but by the desire to extend his Special Theory of Relativity to new domains. Special Relativity (SR) was introduced in the landmark 1905 paper *On the Electrodynamics of Moving Bodies* [22]. Einstein built SR upon two foundational assumptions:

- the laws of physics are the same for all observers
- the speed of light *en vacuo* is the same for all observers

The first of these, often called the principle of relativity, is crucial to building a philosophically sound physical theory. The second was more controversial in its time, despite having been previously confirmed by Michelson and Morley[42]. While SR brings these two simple assumptions to unintuitive and spectacular conclusions, it

applies only for inertial observers: observers in uniform relative motion with respect to each other.

Einstein himself began published the first work generalizing SR to non-inertial observers in 1907 and again in 1911, considering the propagation of light through gravitational fields [23][24]. It was this work that lead to the General Theory of Relativity.

1.2. General Relativity

Einstein published *An Outline of a Generalized Theory of Relativity and of a Theory of Gravitation* with a former classmate and now colleague Marcel Grossmann in 1913, describing their first attempt at a fully covariant theory of gravity [25]. While they did not come to the correct field equations, it was this work that lead to the correct field equations being presented in 1915 and finally published in *The Foundation of the General Theory of Relativity* in 1916 [26]. In the latter paper Einstein finalizes his predictions of the slowing of clocks in gravitational fields and the deflection of light by massive bodies. The paper concludes with a computation of the relativistic correction to Mercury's orbit, arriving at exactly the 43 seconds of arc per century not accounted [26].

Einstein's theory of gravity would be a success based on its testable predictions alone, but, because it grew naturally as an extension of SR, it provided a certain

epistemological cohesiveness missing from Newtonian gravity. The philosophical triumph of General Relativity (GR) is two-fold. It eliminates the absolute time and space of Newtonian mechanics by generalizing SR to all reference frames. It also eliminates Newton's "action at a distance" by introducing a gravitational field that carries the influence of gravity across space and reacts to changes in finite time. Of course Einstein's gravitational field doesn't just carry gravity across space, like the electromagnetic field. It is space itself.

In GR there is no gravitational force. The gravitational field manifests as a space-time metric changing the way we measure distances or equivalently changing the shape of space-time itself. We know from classical mechanics that in the absence of forces particles move in straight lines, but this is only true if the particle lives in a flat universe. If space itself is curved, like the two dimensional surface of the Earth, then the particle is restricted to move along the curvature of the space on a geodesic. A geodesic is a generalized straight line. It is the path of shortest distance between two points. On a flat surface geodesics are indeed straight lines. On the surface of a sphere they are great circles.

The influence of the Einsteinian gravitational field is to alter the shape of space itself, causing particles to deviate from the straight line motion they would experience in flat space. Gravity is then a purely geometrical effect, and we may use the tools of Riemannian geometry to describe its influence.

The apparent deviation of a particle's motion from straight lines due to the gravitational field can be viewed as an acceleration, $\ddot{x}$:

$$\ddot{x}^{\lambda} = \Gamma^{\lambda}{}_{\mu\nu}\, \dot{x}^{\mu}\dot{x}^{\nu}. \tag{1.1}$$

where $\dot{\boldsymbol{x}}$ is the particle's 4-velocity and Γ is a Christoffel symbol given by derivatives of the gravitational field, $\mathbf{g}$, a symmetric space-time metric:

$$\Gamma^{\lambda}{}_{\mu\nu} = \tfrac{1}{2}\, g^{\lambda\rho}\left(g_{\nu\rho,\mu} + g_{\mu\rho,\nu} - g_{\mu\nu,\rho}\right). \tag{1.2}$$

The metric tells us how to measure distance in a particular space-time. We can express the space-time distance in terms of the coordinates of the metric as a line element:

$$ds^2 = g_{\mu\nu}\, \mathrm{d}x^{\mu}\, \mathrm{d}x^{\nu}. \tag{1.3}$$

It is often more convenient to express the coordinates of a space-time metric as it's line element.

From the Christoffel symbols for a given metric we can compute the components of its Riemann curvature tensor:

$$R^{\lambda}{}_{\rho\mu\nu} = \Gamma^{\lambda}{}_{\mu\rho,\nu} - \Gamma^{\lambda}{}_{\nu\rho,\mu} + \Gamma^{\lambda}{}_{\mu\sigma}\Gamma^{\sigma}{}_{\nu\rho} - \Gamma^{\lambda}{}_{\nu\sigma}\Gamma^{\sigma}{}_{\mu\rho}. \tag{1.4}$$

Everything we could want to know about the curvature of a space-time is given by its Riemann tensor. The Riemann tensor vanishes only in flat space. From the Riemann tensor we can compute the effects that the curvature of space has on relative particle motion. For example, we may consider two non-interacting particles moving along

neighboring geodesics. The relative acceleration between the two particles is zero in flat space, but in a curved space this acceleration or 'geodesic deviation' is not. We can define the relative velocity, $\boldsymbol{V}$, and acceleration, $\boldsymbol{A}$ of particles on nearby geodesics by using the vector field tangent to the geodesics, $\boldsymbol{T}$, and the separation vector field pointing from each geodesic to the next, $\boldsymbol{S}$:

$$V^\mu = T^\nu \nabla_\nu S^\mu, \tag{1.5}$$

$$A^\mu = T^\nu \nabla_\nu T^\mu = R^\mu{}_{\lambda\rho\nu}\, T^\lambda\, T^\rho\, S^\nu = \frac{D^2}{Dt^2} S^\mu. \tag{1.6}$$

The Riemann tensor is of great importance in GR. Not just because it contains information about how particles move in curved space, but because it is the relation between the components of the Riemann tensor and a mass (or energy) distribution that defines the gravitational field itself. That relation is given by Einstein's field equations:

$$G_{\mu\nu} = 16\pi\, T_{\mu\nu}, \tag{1.7}$$

where $G_{\mu\nu}$ and $T_{\mu\nu}$ are the components of the Einstein tensor and the matter stress-energy tensor that is the source the gravitational field, respectively. The Einstein tensor, $\mathbf{G}$, is built from contractions of the Riemann tensor:

$$R_{\mu\nu} = R^\lambda{}_{\mu\lambda\nu}, \tag{1.8}$$

$$R = R^\mu{}_\mu, \tag{1.9}$$

$$G_{\mu\nu} = R_{\mu\nu} + \tfrac{1}{2} R\, g_{\mu\nu}. \tag{1.10}$$

As the Riemann tensor is itself built from second derivatives of the space-time metric, Einstein's field equation is thus a non-linear second order differential equation for the gravitational field. Worse than that, each of the ten independent component pairs of **G** and **T** provides its own differential equation, leaving us with ten coupled, non-linear differential equations.

In Einstein's 1916 work he was forced to solve his own field equations perturbatively in order to arrive at the famed results. Despite the difficulty in solving the field equations, several exact solutions have been discovered since 1916. The first exact solution was found by the German physicist Karl Schwarzschild [49]. His solution for the gravitational field of a point mass *en vacuo* came only a few months after Einstein announced the correct field equations, and before Einstein himself published the in depth summary in 1916.

Since the work of Schwarzschild, many other solutions to the Einstein equation have been found. Some are exact solutions, but many more have found perturbatively. One of the most historically controversial solutions to Einstein's equation is the gravitational wave solution.

1.3. Gravitational Radiation

The existence of wave solutions to the GR field equations was a point of contention for several decades. Einstein first published the linearized version of his field equations in 1916, noting that such equations admit a wave-like solution. He later redacted the

wave solution as a consequence of linearization and not a real physical effect. Over the years the gravitational wave solution gained more acceptance. Finally, in the late 1960s and 1970s results showing that gravitational waves carry energy cemented the wave solutions as real, gauge invariant, measurable consequences of GR.

To linearize equation 1.7 one must assume a perturbative expansion around a known solution. It is simplest to work in a weak field expansion of the flat, Minkowski space-time. It is possible to expand about a general, non-Minkowski background. In this way one can perturb strong fields by demanding a 'high frequency' perturbation field, where the curvature scale of the perturbation is much smaller than the background space-time.

Starting from flat space, we may write the components of the gravitational field as:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1.11}$$

where $\boldsymbol{\eta}$ is the Minkowski metric and $\mathbf{h}$ is a small perturbation, satisfying $|\mathbf{h}| \ll 1$. Since $\boldsymbol{\eta}$ is flat, the components of the Riemann tensor will be functions of the perturbation field only:

$$R_{\lambda\rho\mu\nu} = \tfrac{1}{2}\left(h_{\lambda\nu,\rho\mu} + h_{\rho\mu,\lambda\nu} - h_{\rho\nu,\lambda\mu} - h_{\lambda\mu,\rho\nu}\right) + \mathcal{O}(h^2) \tag{1.12}$$

We can simplify things by defining the trace reversed perturbation field:

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}h \tag{1.13}$$

where $h = h^{\mu}{}_{\mu}$ is the trace of $\mathbf{h}$.

Next we will enforce the Lorenz gauge[1] condition, $\partial^\mu \bar{h}_{\mu\nu} = 0$, to further simplify matters. Finally, we may write the linearized Einstein tensor and thus the linearized Einstein equation (to first order):

$$G_{\mu\nu} = -\tfrac{1}{2}\Box\bar{h}_{\mu\nu} + \mathcal{O}(h^2) \tag{1.14}$$

$$\Box\bar{h}_{\mu\nu} = -16\pi\, T_{\mu\nu} \tag{1.15}$$

where $\Box = \partial^\mu\partial_\mu$ is the flat space D'Alembertian operator. The linearized Einstein equation is a driven wave equation. Outside of the source, when $T_{\mu\nu} = 0$, linearized gravity reduces to the simple, homogeneous wave equation with plane wave solutions:

$$\Box\bar{h}_{\mu\nu} = 0, \tag{1.16}$$

$$\bar{h}_{\mu\nu}(\boldsymbol{x}) = A_{\mu\nu}(\boldsymbol{k}) \exp\left(\imath\, \boldsymbol{k}\cdot\boldsymbol{x}\right). \tag{1.17}$$

Here $\boldsymbol{k}$ is the wave-vector and $A_{\mu\nu}$ are the constant components of the polarization tensor. Inserting this solution in equation 1.16, we find $k^\mu k_\mu = 0$: the wave vector is null. Gravitational waves travel at the speed of light. When we apply the Lorenz gauge condition to equation 1.17, we see that $k^\mu A_{\mu\nu} = 0$. The polarization is perpendicular to the direction of propagation, so gravitational waves are transverse.

But the Lorenz gauge does not fix the solution completely. Far from the source we may make an additional gauge choice, moving to the transverse-traceless gauge. In the transverse-traceless gauge we reduce the perturbation field to a mere two degrees

[1]First posed in the context of electricity and magnetism by Ludwig Lorenz, not to be confused with Hendrick A. Lorentz of special relativity fame.

of freedom by demanding[2]:

$$h^{t\mu} = 0, \quad h^{i}{}_{i} = 0, \quad \partial^{i} h_{ij} = 0. \tag{1.18}$$

If we choose the direction of propagation along the z-axis, we can express the solutions as:

$$h_{ab}^{\mathrm{TT}}(t,z) = \begin{pmatrix} h_{+} & h_{\times} \\ h_{\times} & -h_{+} \end{pmatrix}_{ab} \cos\left[\omega(t-z)\right], \tag{1.19}$$

where $a, b \in (x, y)$ are transverse spacial indices, $\omega = k_t$ is the wave frequency, and h_+ and $h_\times$ are the two independent components of the polarization tensor. We can now determine the effect of gravitational waves on test particles they pass.

Consider a collection of particles characterized by a velocity vector field, $\boldsymbol{U}$, and a separation vector, $\boldsymbol{S}$. Going to first order in $\mathbf{h}^{\mathrm{TT}}$ in the small velocity limit the geodesic deviation equation 1.6 reduces to:

$$\partial_t{}^2 S_a = R_{attb}\, S^b = \tfrac{1}{2} S^b\, \partial_t{}^2 h_{ab}^{\mathrm{TT}}, \tag{1.20}$$

keeping only the transverse part.

We can solve equation 1.20 for two cases given an initial separation vector, $\boldsymbol{S}_0$, where h_+ and $h_\times$ independently vanish. For the purely 'plus' polarized wave we find:

$$\begin{aligned} S^1 &= \left(1 + \tfrac{1}{2}\, h_+ \exp\left(\imath\, k_\mu\, x^\mu\right)\right) S_0^1, \\ S^2 &= \left(1 - \tfrac{1}{2}\, h_+ \exp\left(\imath\, k_\mu\, x^\mu\right)\right) S_0^2, \end{aligned} \tag{1.21}$$

[2]here we have dropped our overbars, because in a traceless gauge the trace reversal of equation 1.13 vanishes.

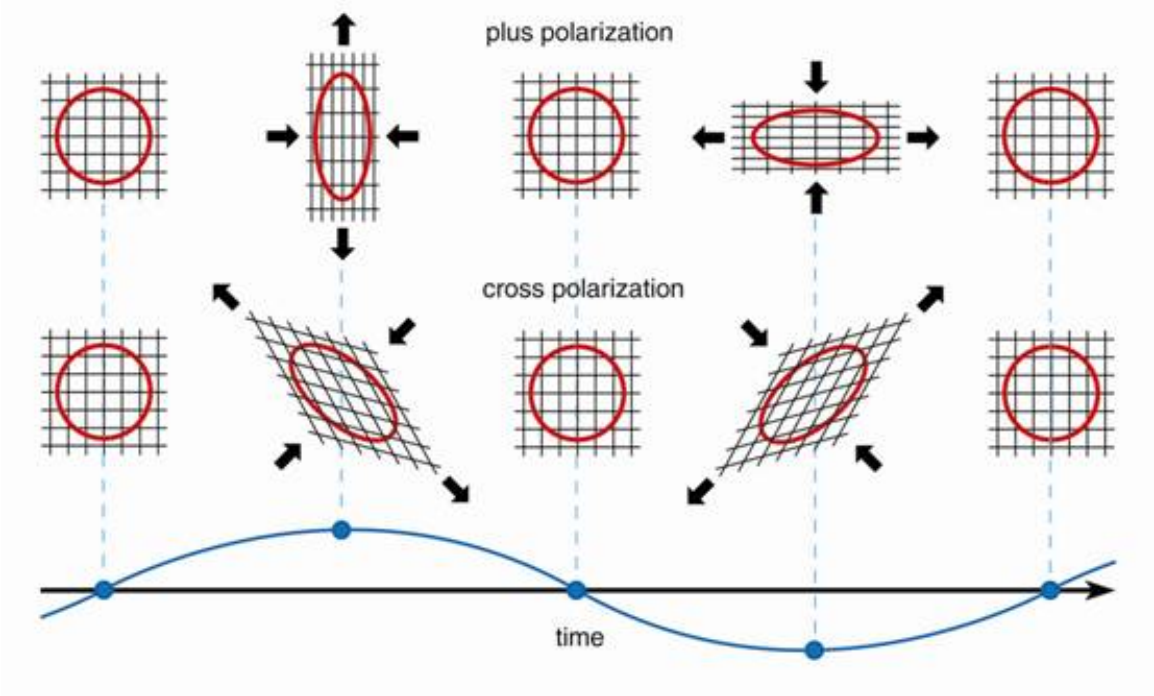

Figure 1.2. Action of 'plus' and 'cross' gravitational waves on a ring of test particles. The two polarizations have actions rotated 45° from each other.

and for the purely 'cross' polarized wave we find:

$$S^1 = S_0^1 + \tfrac{1}{2}\, S_0^2\, h_+ \exp\left(\imath\, k_\mu\, x^\mu\right),$$
$$S^2 = S_0^2 + \tfrac{1}{2}\, S_0^1\, h_+ \exp\left(\imath\, k_\mu\, x^\mu\right). \tag{1.22}$$

For the 'plus' case particles with initial separation in the x^1 will oscillate back and forth along that direction. Similarly for x^2. While the 'cross' case will induce shears across the direction of separation. This can be seen in figure 1.2.

If gravitational waves are to have real, measurable effects, such as changing the motion of test particles, then they must carry energy to accomplish this. The original controversy over gravitational radiation stems from this fact. In order to arrive at the

linearized wave solution we made two gauge choices. Are the effects of gravitational waves merely coordinate effects dependent upon our choice of gauge?

Isaacson answered this question definitively in 1968 [32][33]. Perturbing an arbitrary space-time metric (instead of flat space) in the high frequency limit of the radiation, he showed that gravitational waves produce gauge invariant effects on the curvature tensor. He also derived the energy loss of a system due to gravitational radiation.

While it is unclear how to locally define the energy of a gravitational field and thus a gravitational wave. One can define an effective stress-energy tensor for the gravitational waves by averaging over a small region of space compared to the background curvature scale. In the transverse-traceless gauge the components of the effective stress-energy are:

$$\tau_{\mu\nu} = \frac{1}{32\pi} \left\langle \left(\partial_\mu \, h^{\mathrm{TT}}_{\lambda\rho}\right) \, \left(\partial_\nu \, h^{\lambda\rho}_{\mathrm{TT}}\right) \right\rangle . \tag{1.23}$$

From this stress-energy we can compute the energy radiated, E, through a surface, Σ:

$$E = \int_\Sigma \tau_{tt} \, \mathrm{d}^3 x. \tag{1.24}$$

If we take Σ to be the two-sphere at spacial infinity, Σ^2_∞, we can compute the total rate of energy loss or power, P, radiated by a gravitational wave producing system:

$$P = \int_{\Sigma^2_\infty} \tau_{t\mu} \, n^\mu \, r^2 \, \mathrm{d}\Omega, \tag{1.25}$$

where $\boldsymbol{n}$ is the unit space-like vector normal to Σ^2_∞. In spherical coordinates (t, r, θ, ϕ) $\boldsymbol{n}$'s components are $n^\nu = (0, 1, 0, 0)$.

This loss of energy will have measurable effects on the radiating system. In the 1970s Weisberg, Taylor, and Fowler began studying a binary neutron star system containing the pulsar, PSR B1913+16. Because of the pulsars frequent and regular emission they were able to determine the orbital parameters of the system very accurately. With the orbital parameters they could compute the gravitational radiation produced by the system and in turn the rate of change of the orbital parameters due to energy loss from gravitational radiation. They tracked the system for several years and published the first indirect measurement of gravitational radiation in 1981 [60].

Weisberg and Taylor continue to track the system to this day. With each new measurement they have confirmed the predictions of GR to greater and greater accuracy. Figure 1.3 shows more recent results from 2004 comparing their measurements with theoretical predictions.

To this day, no direct detection of gravitational waves has been made and not for lack of trying. The indirect detection along with several independent tests of GR give us great confidence that gravitational waves are being produced by astrophysical systems in the manner described by GR.

In the work that follows we will discuss how gravitational waves are generated and what sorts of systems produce them. With this knowledge we can begin to discuss how we hope to directly detect gravitational waves, and the challenges that make the direct detection of gravitational waves so difficult.

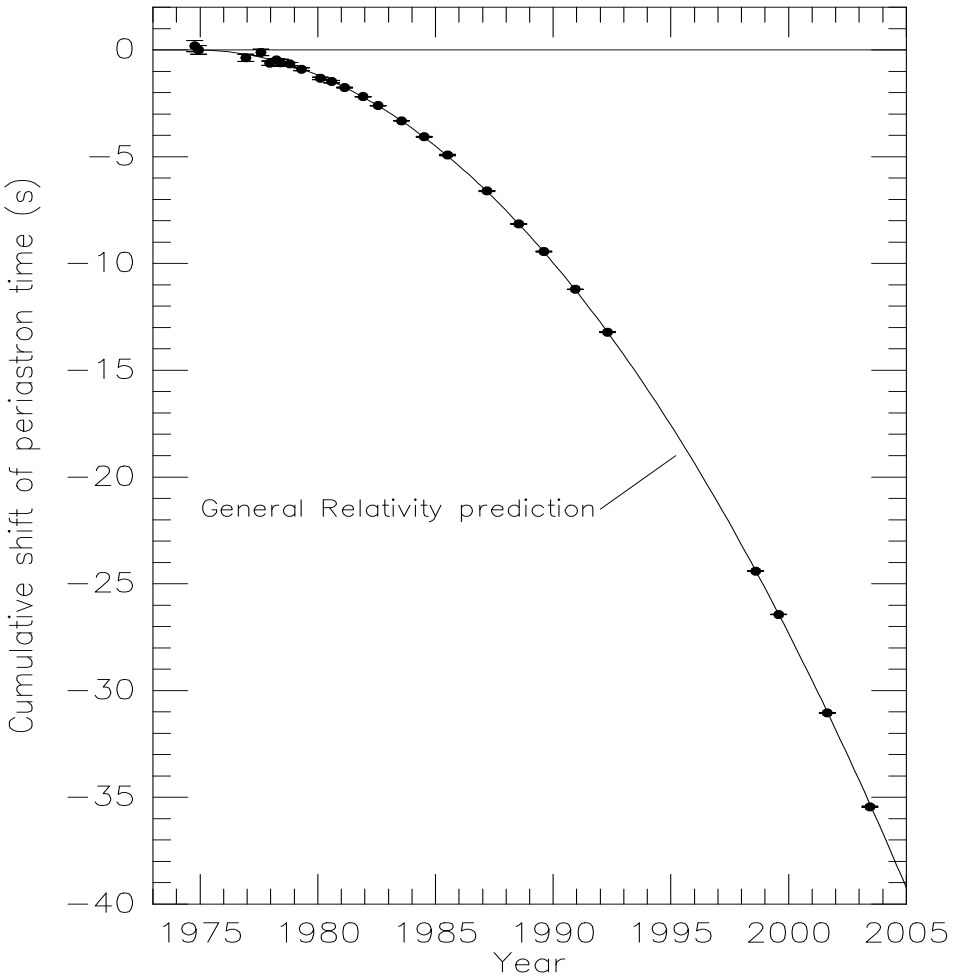

Figure 1.3. Orbital decay of PSR B1913+16. The data points indicate the observed change in the epoch of periastron with time while the parabola illustrates the theoretically expected change in epoch for a system emitting gravitational radiation, according to general relativity [59].

2. SOURCES OF GRAVITATIONAL WAVES

The name of the game is detecting gravitational waves. As Sherlock Holmes would be happy to tell us, we can improve our ability to detect anything by knowing what to look for. Knowing the specifics of the gravitational wave events we hope to find can drive instrument design and data analysis techniques. Here we will examine first how gravitational waves are produced in general and then consider two specific examples of gravitational wave production: compact binary orbits and black hole ringdown.

If we wish to determine which sorts of astrophysical events will produce gravitational waves we must consider solutions to the inhomogeneous, linearized Einstein equation from section 1.3:

$$\Box \bar{h}_{\mu\nu} = -16\pi T_{\mu\nu}. \tag{1.15}$$

Equation 1.15 is most easily solved using a Green's function:

$$\bar{h}_{\mu\nu}(t, \vec{x}) = 4 \int \frac{T_{\mu\nu}\left(t - |\vec{x} - \vec{x}'|, \vec{x}'\right)}{|\vec{x} - \vec{x}'|} \, \mathrm{d}^3 x'. \tag{2.1}$$

In this case to determine the linearized gravitational field at a space-time location $(t, \vec{x})$ we must evaluate the position of the source, $T_{\mu\nu}$, at a time in the past to account for the finite travel time of the gravitational effect. This same procedure is used in the case of electro-magnetic fields.

To simplify the integration, we will evaluate equation 2.1 in the 'far field' zone. We assume that the distance to the source, $|\vec{x} - \vec{x}'|$, is much larger than the wavelength of the radiation, λ, which is in turn much larger than the characteristic size of the

source, R:

$$r \gg \lambda \gg R \tag{2.2}$$

The far field assumption implies that the source is moving slowly. The velocity of the source is related to its size and the frequency of the radiation: $v \approx R\,\omega \approx R/\lambda$. Under these assumptions we may approximate the distance to the source as simply the radial component of the field's spacial location vector: $|\vec{x} - \vec{x}'| \approx r$. We will define the retarded time, $t_r = t - r$, and rewrite equation 2.1:

$$\bar{h}_{\mu\nu}(t, \vec{x}) = \frac{4}{r} \int T_{\mu\nu}\left(t_r, \vec{x}'\right)\, \mathrm{d}^3x'. \tag{2.3}$$

We can use conservation of the stress-energy tensor in our flat background space, $T^{\mu\nu}{}_{,\nu} = 0$, to show

$$\int T^{ij}\, \mathrm{d}^3x = \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2} \int T^{tt} x^i x^j\, \mathrm{d}^3x. \tag{2.4}$$

For slowly moving sources, which was implicitly part of our far field assumption, T^{tt} is simply the mass density of the source. So we define the mass quadrupole moment of our source:

$$I_{ij}(t) = \int x_i\, x_j\, T^{tt}(t, \vec{x})\, \mathrm{d}^3x. \tag{2.5}$$

Finally, the spacial components of the far field gravitational wave amplitude is given by:

$$\bar{h}_{ij}(t, r) = \frac{2}{r}\frac{\mathrm{d}^2}{\mathrm{d}t^2} I_{ij}(t_r). \tag{2.6}$$

To leading order gravitational waves are produced by accelerating mass quadrupoles. Any system with such a mass quadrupole will radiate gravitational waves. We will

consider two such systems binary orbits and the quasi-normal vibrations of black holes.

2.1. Gravitational Waves from Binary Systems

Consider two bodies in a circular bound orbit. We will assume the bodies are widely enough separated that their trajectories are described by Newtonian gravity. For simplicity we will assume that the objects are of equal mass, $m_1 = m_2 = m$, and we will define the system total mass, $M = m_1 + m_2 = 2m$. Because of their equal masses, the bodies are equidistant from the system's center of mass. We will call this distance R. Finally, we will align the orbital angular momentum with the z-axis, so the motion is constrained to the xy-plane.

The trajectory of such a Keplerian orbit for one of the bodies is given in cartesian coordinates by:

$$x = R\cos(\Omega\, t), \qquad y = R\sin(\Omega\, t), \tag{2.7}$$

where $\Omega = (M/R^3)^{1/2}$ is the orbital angular frequency. The other object has a similar trajectory:

$$x = -R\cos(\Omega\, t), \qquad y = -R\sin(\Omega\, t), \tag{2.8}$$

for point masses the mass density is given by:

$$T^{tt} = \frac{M}{2}\,\delta(z)\times$$
$$[\delta(x - R\cos\Omega\, t)\delta(y - R\sin\Omega\, t) + \delta(x + R\cos\Omega\, t)\delta(y + R\sin\Omega\, t)] \tag{2.9}$$

From the mass density we compute the components of the mass quadrupole by integration as in equation 2.5:

$$I_{xx} = \tfrac{1}{2} MR^2 \left(1 + \cos 2\Omega\, t\right) \tag{2.10}$$

$$I_{yy} = \tfrac{1}{2} MR^2 \left(1 - \cos 2\Omega\, t\right) \tag{2.11}$$

$$I_{xy} = I_{yx} = \tfrac{1}{2} MR^2 \, \sin 2\Omega\, t \tag{2.12}$$

$$I_{zi} = 0 \tag{2.13}$$

Finally we can determine the gravitational wave perturbation for an observer on the z-axis by differentiation as in equation 2.6, remembering to evaluate the source at the retarded time:

$$\bar{h}_{ij}(t, r) = \frac{4M^2}{rR} \begin{pmatrix} -\cos 2\Omega\, t_r & -\sin 2\Omega\, t_r & 0 \\ -\sin 2\Omega\, t_r & \cos 2\Omega\, t_r & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.14}$$

Note that the gravitational wave frequency is twice the orbital frequency. Because of the symmetry in the system, the configuration, and thus the quadrupole moment, repeats after one half orbital period. This would be true even for unequal mass systems, because, while the configuration is not identical, the quadrupole is invariant under rotations by π.

The gravitational waveform of equation 2.14 only applies along the z-axis, but we can extend the result to apply to waves propagating in an arbitrary direction, $\vec{k}$, by performing a rotation. We define the inclination angle, ι, as the angle between $\vec{k}$, the propagation direction from the system center of mass to the observer, and $\vec{z}$, the orbital angular momentum direction. We can allow $\vec{k}$ to define a new coordinate

system $(\hat{p}, \hat{q}, \hat{k})$ such that $\hat{p} \cdot \hat{k} = 0$ and $\hat{p} \times \hat{k} = \hat{q}$. In this coordinate system the gravitational waveform can be expressed in the transverse-traceless gauge as its plus and cross components:

$$h_+(t, r) = \frac{2M^2}{rR}(1 + \cos^2 \iota) \cos 2\Omega\, t_r, \tag{2.15}$$

$$h_\times(t, r) = \frac{4M^2}{rR} \cos(\iota) \sin 2\Omega\, t_r. \tag{2.16}$$

In the case of a real binary system the energy loss from the radiation would cause the orbital separation to shrink and consequently the orbital frequency to grow. As the system evolves, one must alter the particle trajectories, and therefore the quadrupole moment, to recompute the gravitational waveforms. In practice the time scale of energy loss is much longer than the orbital period. In this case we could hold the orbital separation and frequency fixed for several orbits and then determine the total energy radiated in that time. We adjust the orbital parameters accordingly and compute the next several orbits. By iterating this process one can build a psuedo-adiabadic approximation to a longer duration gravitational waveform.

As more time passes, the frequency would eventually become so large that the wavelength of the radiation will be of a similar scale as the orbital separation. At this point the far-field assumptions break down, and we cannot use our simple methods to compute waveforms.

We can now estimate the amplitude at Earth of a gravitational wave from a reasonable astrophysical source. We will choose an optimally oriented binary neutron star system, so $\iota = 0$, $m \approx 1.5 M_\odot$ and $M \approx 3 M_\odot$. The Milky Way is about 35 kpc

in diameter, so we will pick a galactic system at a distance of 10 kpc. At a time late in the evolution of such a system the total orbital separation will be about 200 km, so $R = 100$ km. From Kepler's law the orbital frequency would be about 100 Hz at this time, giving a radiation frequency of about 200 Hz.

$$h_{\oplus} \approx \frac{4M^2}{rR} \approx 3 \times 10^{-18} \left(\frac{M}{3M_{\odot}}\right)^2 \left(\frac{10\,\mathrm{kpc}}{r}\right) \left(\frac{100\,\mathrm{km}}{R}\right) \tag{2.17}$$

This source is, in the grand scheme of things, very close to Earth. The Virgo super cluster, of which the Milky Way is a member, is about 33 Mpc in diameter. A reasonable extra galactic neutron star binary could be at a distance of 10 Mpc. Radiation from such a source would have an amplitude of order 10^{-21} at Earth. This amplitude is the gravitational wave strain felt by an observer on Earth. Needless to say, we will need very precise detectors if we hope of detecting such a gravitational wave signal directly.

2.2. Gravitational Waves from Black Hole Ringdowns

In 1916 Karl Schwarzschild published the unique stationary, spherically symmetric vacuum solution to the Einstein equation 1.7 [49]. In Schwarzschild coordinates (t, r, θ, ϕ) where the radial coordinate geometrically defines the Gaussian curvature and area of a sphere, the Schwarzschild solution can be expressed as a space-time metric, which defines the line element:

$$ds^2 = -\left(1 - \frac{2M}{r}\right) \mathrm{d}t^2 + \left(1 - \frac{2M}{r}\right)^{-1} \mathrm{d}r^2 + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2\right) \tag{2.18}$$

As mentioned, r defines the radius of curvature at a point in the Schwarzschild space; however, it does not map directly to the usual Euclidean radial distance. Because of this, all four coordinates are only asymptotically related to the flat-space spherical coordinates, as $r \to \infty$.

The Schwarzschild metric defines the gravitational field outside of a spherically symmetric body of mass, M. The correspondence between the metric parameter M and the mass of the body results from taking a weak field limit, where $r \gg M$, and comparing to the Newtonian gravitational solution.

There are two particularly interesting places in a Schwarzschild space. At both $r = 0$ and $r = 2M$ the metric components become infinite. To consider what is happening at these locations we turn to the curvature tensor defined in equation 1.4. We may contract the Riemann tensor for Schwarzschild space with itself, forming the Kretchmann scalar:

$$R^{\mu\nu\lambda\rho} R_{\mu\nu\lambda\rho} = \frac{48M^2}{r^2}. \tag{2.19}$$

We see that this curvature invariant blows up as $r \to 0$. This point is a true curvature singularity, where the tools of Riemannian geometry breakdown. The other problem location, actually a surface where $r = 2M = r_s$, is not as bad. The distance r_s is known as the Schwarzschild radius, and it turns out that all contractions of the Riemann tensor are well defined here. In this case a simple change of coordinates can

show that the apparent singularity at r_s is merely an effect of the coordinate system. For instance, in Eddington-Finkelstein coordinates the metric is well behaved at the Schwarzschild radius [58].

For most real scenarios the Schwarzschild radius is inside of the gravitating body in question. In these cases the Schwarzschild metric, being a vacuum solution, does not apply. Consider the sun. Its mass is about $M_\odot = 2 \times 10^{30}$ kg, giving a Schwarzschild radius of about 1.5 km. The sun's radius is much larger than that, $R_\odot = 7 \times 10^5$ km. For real stars there is nothing strange going on at the Schwarzschild radius, because the solution does not apply. But if the entire mass of an object is inside the Schwarzschild radius, then $r = r_s$ is now a valid point in the space[1].

The Schwarzschild radius defines the 'event horizon' of the space-time. For a particle inside the event horizon ($r < r_s$) to exit ($r > r_s$) it would need an infinite amount of energy. One way to describe this is to say that the escape velocity for a particle at the horizon is the speed of light. Because of this, light that originates inside of the horizon cannot escape. It is this fact that gives us the term 'black hole', popularized by John Wheeler in the late 1960s. Black holes are defined by the presence of their event horizon.

Schwarzschild space-time provides the simplest black hole, but other black hole solutions exist. In general, a black hole may be spinning and/or charged, with a slightly modified metric defining the space in each case. Surprisingly, the so-called

[1]The central singularity might not be in the vacuum and therefore it is still not a valid point for the solution.

'no hair theorem' states that if we consider situations where the only fields present are the gravitational and electro-magnetic, all black holes are uniquely characterized by their mass, angular momentum, and charge. Astrophysical black holes, like most macroscopic objects, are unlikely to be significantly charged, so we will restrict ourselves to massive and spinning black holes.

The Kerr solution describes spinning black holes. In Boyer-Lindquist coordinates (t, r, θ, ϕ), which are asymptotically ellipsoidal (in the same way Schwarzcshild were spherical):

$$ds^2 = -\frac{\Delta}{\Sigma}\left(\mathrm{d}t - \bar{a}\sin^2\,\mathrm{d}\phi\right)^2 + \frac{\sin^2\theta}{\Sigma}\left((r^2+\bar{a}^2)\mathrm{d}\phi - \bar{a}\,\mathrm{d}t\right) + \frac{\Sigma}{\Delta}\mathrm{d}r^2 + \Sigma\,\mathrm{d}\theta^2 \quad (2.20)$$

$$\Delta = r^2 - 2Mr + \bar{a}^2 \quad (2.21)$$

$$\Sigma = r^2 + \bar{a}^2\cos^2\theta \quad (2.22)$$

here M is again the black hole mass and $\bar{a} = J/M$ is the angular momentum per unit mass. The Kerr black hole has an event horizon at $r_- = M - \sqrt{M^2 - \bar{a}^2}$ and a secondary horizon at $r_+ = M + \sqrt{M^2 - \bar{a}^2}$. The secondary horizon defines the 'ergosphere', inside of which the space-time co-rotates with the black hole. We note that this solution breaks down for $\bar{a} > M$. Because of this, it is often more useful to work with the dimensionless spin parameter, $a = \bar{a}/M = J/M^2 \in [0, 1]$.

So what happens to the black hole space-time as an object falls into the black hole? When an in-falling object is outside of the event horizon the Kerr solution is no longer valid. The surrounding space is not a vacuum. Because of the no-hair theorem, once the object passes through the horizon, the outside space-time must

return to the Kerr solution (albeit with potentially modified mass and spin). We can find approximate solutions for the space-time during this process by considering small perturbations to the black hole space-time.

The first examination of perturbations to black hole space-times was conducted by Regge and Wheeler in 1957 [47]. This analysis both predated the Kerr solution and the coining of the term 'black hole'. For the Schwarzschild space-time gravitational perturbations can be handled by a separation of variables, which leads to a spherical harmonic decomposition.

Teukolsky went on to show, that by working with curvature invariants instead of the metric directly, the perturbations lead to a similar separation of variables in these quantities [57]. The Fourier transform of a spin-2 perturbation field[2], ψ_s, on the curvature invariants is expanded as spin weighted spheroidal harmonics:

$$\psi_s(t, r, \theta, \phi) = \frac{1}{2\pi} \int \exp(-\imath\omega t) \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} {}_sS_{\ell m}(a\omega, \theta, \phi)\, R_{\ell m}(r)\, \mathrm{d}\omega \tag{2.23}$$

where ${}_sS_{\ell m}$ are the components of the spin weighted spheroidal harmonics, $R_{\ell m}$ are the components of the radial part of the field, ω is the angular oscillation frequency of the perturbation, and a is the black hole's dimensionless spin parameter. The radial equation for $R_{\ell m}$ is often referred to as the Teukolsky equation. For vacuum perturbations we have:

$$\Delta \partial_r{}^2 R_{\ell m} + 2(s+1)(r-M)\partial_r R_{\ell m} + V\, R_{\ell m} = 0, \tag{2.24}$$

[2]The field is spin-2 for gravitational perturbations, but in general the perturbation field can have any spin. We will consider spins of both $s = \pm 2$.

where $\Delta = (r - r_-)(r - r_+)$,

$$V = 2\imath s\omega r - a^2\omega^2 - {}_sA_{\ell m} + \frac{1}{\Delta}\Big[(r^2+a^2)\omega^2 - 4Mam\omega r + a^2m^2 + 2\imath s\left(am(r-M) - M\omega(r^2-a^2)\right)\Big],$$

M is the black hole mass, and $A_{\ell m}$ are separation constants determined by the angular equation.

Solving the Teukolsky equation can be thought of as an eigen-problem with complex eigenvalues, ω. Only certain values of ω satisfy equation 2.24, leading to a spectrum of discrete vibrational modes, $M\omega_{\ell m}$, characterized by the mass of the black hole. Because the eigenvalues are complex, the perturbations will be exponentially damped, resulting in quasi-normal modes (QNM) of vibration. In the case where $s = 2$ this damping is the result of energy being lost into the black hole horizon. Where $s = -2$, the damping is caused by energy radiated away to $r = \infty$. For these cases the perturbation fields, ψ_s, are actually Weyl scalars: $\psi_2 = \Psi_0$ and $\psi_{-2} = \Psi_4$. Each QNM can be broken into it's real part, related to the vibrational frequency, f, and it's imaginary part related to the damping time, τ, or quality factor, Q of the perturbation:

$$f_{\ell m} = 2\pi\,\mathrm{Re}(\omega_{\ell m}) \tag{2.25}$$

$$\tau_{\ell m}^{-1} = \mathrm{Im}(\omega_{\ell m}) = \pi\frac{f_{\ell m}}{Q_{\ell m}} \tag{2.26}$$

Solving the Teukolsky eignen-problem is no easy task. Leaver published the first compendium of QNM frequencies in 1985 [37], some of which are compiled in table

ℓ, m	a	$M\omega_{\ell m}$
2,+2	-0.98	$+0.2927 + 0.0881\imath$
	-0.60	$+0.3168 + 0.0889\imath$
	0	$\pm 0.3737 + 0.0890\imath$
	+0.60	$-0.4940 + 0.0838\imath$
	+0.98	$-0.8254 + 0.0386\imath$
2,+1	-0.98	$+0.3439 + 0.0837\imath$
	-0.60	$+0.3490 + 0.0876\imath$
	0	$\pm 0.3737 + 0.0890\imath$
	+0.60	$-0.4360 + 0.0846\imath$
	+0.98	$-0.5642 + 0.0516\imath$
2,0	$\mp$0.98	$\pm 0.4222 + 0.0735\imath$
	$\mp$0.60	$\pm 0.3881 + 0.0860\imath$
	0	$\pm 0.3737 + 0.0890\imath$

Table 2.1. A list of a few black hole quasi-normal mode frequencies for various black hole spins. Adapted from [37].

2.1. Now it is more common to solve the problem numerically, resulting in long lists of QNM frequencies for every imaginable black hole space-time and perturbation type. To supplement all of this tabulation it is possible to determine analytic fits to numerical results as a function of black hole mass and spin:

$$f_{\ell m} = \frac{1}{2\pi M}\left(f_1 + f_2(1-a)^{f_3}\right), \tag{2.27}$$

$$Q_{\ell m} = q_1 + q_2(1-a)^{q_3}. \tag{2.28}$$

The constants f_i and q_i are tabulated for individual QNMs. For example the fundamental mode of $\ell = m = 2$ has:

$$f_1 = 1.5251, \qquad f_2 = -1.1568, \qquad f_3 = 0.1292,$$

$$q_1 = 0.7000, \qquad q_2 = 1.4187, \qquad q_3 = -0.4990,$$

providing a fit within 5% of numerical results, as computed by Berti *et al.* [13]. Figure 2.1 shows the frequency fitting function for the $\ell = m = 2$ QNM at four dimensionless spin values.

As we mentioned, a perturbed black hole will emit gravitational radiation to shed energy and return to a state described by only its mass and spin. The Weyl scalars, Ψ_0 and Ψ_4, themselves describe the in-going and out-going radiation directly. The Teukolsky problem translates the gravitational perturbations directly to the gravitational radiation. Because the QNM vibrations are expressed as spheroidal harmonics, the resulting radiation is a spheroidal wave. We can easily discern the form of the

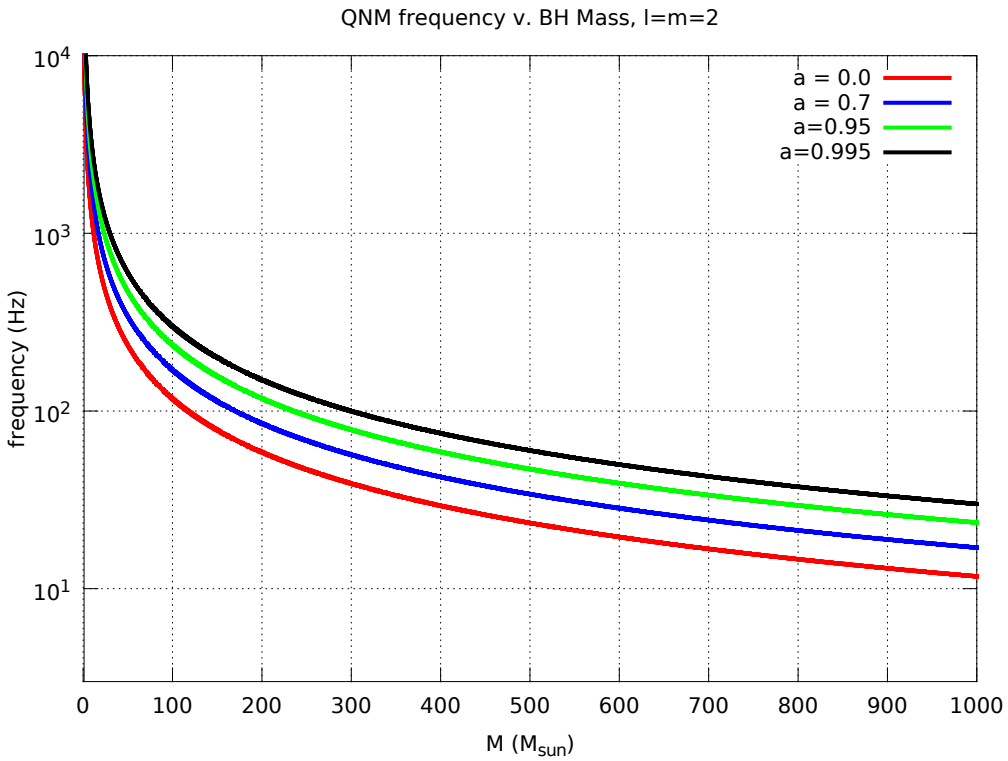

Figure 2.1. $\ell = m = 2$ quasi-normal mode frequency as a function of black hole mass.

out-going gravitational waves from Ψ_4.

$$h_+(t,r) = \frac{M}{r}\mathrm{Re}\left[A^+_{\ell m}\exp\imath(2\pi f_{\ell m}t_r + \phi^+_{\ell m})\,\exp(-\pi f t_r/Q_{\ell m})\,{}_{-2}S_{\ell m}(\iota,\beta)\right], \quad (2.29)$$

$$h_\times(t,r) = \frac{M}{r}\mathrm{Im}\left[A^\times_{\ell m}\exp\imath(2\pi f_{\ell m}t_r + \phi^\times_{\ell m})\,\exp(-\pi f t_r/Q_{\ell m})\,{}_{-2}S_{\ell m}(\iota,\beta)\right], \quad (2.30)$$

where $A^{+/\times}_{\ell m}$ and $\phi^{+/\times}_{\ell m}$ are the real amplitude and phase of the wave. The As and ϕs are determined by the source of the QNM driving perturbation. ${}_{-2}S_{\ell m}$ are the components of the spin-weight -2 spheroidal harmonics, and the angles (ι,β) define the line of sight from the black hole coordinate system to an observer. As gravitational waves are fundamentally coupled to the mass quadrupole, the $\ell = 2$ modes will be the first to produce radiation.

We can also define radiation efficiency, ϵ, as the fraction of a black hole's total mass that is radiated away as gravitational waves. The radiation efficiency will greatly affect the gravitational wave amplitude seen by an observer at Earth. The radiation efficiency per QNM was approximated in [13] as:

$$\epsilon_{\ell m} \approx \frac{Q_{\ell m} M f_{\ell m}}{16}\left[(A^+_{\ell m})^2 + (A^\times_{\ell m})^2\right]. \quad (2.31)$$

The radiation efficiency depends strongly on the nature of the perturbation. In order to produce a sufficient amplitude of gravitational waves for detection the $\ell = 2$ modes need to be maximally excited. Binary mergers that result in black holes are prime examples of a method that produces high amplitude $\ell = 2$ perturbations. Assuming quasi-circular orbits and initially non-spinning black holes, Berti *et al.* approximated the radiation efficiency for two dominate modes resulting from binary

black hole mergers. For a binary system of initially non-spinning black holes the fractional energy radiated is given in terms of the system mass ratio, q [11]:

$$\epsilon_{22} \approx 0.271 \frac{q^2}{(1+q)^4}, \qquad \epsilon_{33} \approx 0.104 \frac{q^2(q-1)^2}{(1+q)^6}. \tag{2.32}$$

For an equal mass merger, $q = 1$ we see that the $\ell = m = 3$ mode is not excited, and about 1.7% of the final black hole's total mass is radiated in the $\ell = m = 2$ mode. On the other hand when $q = 10$, only about 0.2% of the total mass is radiated in the $\ell = m = 2$ mode and about 0.05% is radiated in the $\ell = m = 3$ mode. As the mass ratio increases, a larger fraction of the total radiated energy is in higher order modes.

We can again estimate the amplitude of a gravitational wave seen on Earth. In this case we will assume a ringdown from a black hole that was created by an equal mass merger. We will choose an optimally oriented source with a final mass of $100\mathrm{M}_\odot$. If we assume the two initial black holes are not spinning, then the final angular momentum will come from the orbital angular momentum alone. In the equal mass case the final black hole will have dimensionless spin $a \approx .7$. We will choose an extra galactic source with $r = 10$ Mpc.

Using equations 2.27 and 2.28, we find $f_{22} \approx 170$ Hz and $Q_{22} \approx 3.3$. We will assume a purely plus polarized wave and determine the amplitude, A^+_{22}, by combining equations 2.31 and 2.32: $A^+_{22} \approx .98$. The peak gravitational wave amplitude seen at Earth is then:

$$h_\oplus \approx \frac{M}{r} A^+_{22} \approx 5 \times 10^{-19} \left(\frac{M}{100 M_\odot}\right) \left(\frac{10\,\mathrm{Mpc}}{r}\right) \left(\frac{A^+_{22}}{.98}\right) \tag{2.33}$$

This extra-galactic source has an estimated peak amplitude two orders of magnitude larger than the similarly distant binary system discussed in section 2.1. We should note that the black hole ringdown will be of much shorter duration than a binary inspiral, which will make detection harder. Further, given our current understanding of black hole populations, a black hole-black hole system with these sorts of parameters should be exceedingly rare. We will discuss this more in chapter 4.

Now that we have discussed the sorts of gravitational waves that might reach an Earth based detector, we can begin to examine how we could practically detect gravitational waves.

3. DETECTING GRAVITATIONAL WAVES

We will turn now to how gravitational waves interact with instruments on Earth. In section 1.3 we discussed how 'plus' and 'cross' polarized gravitational waves cause geodesic deviation in the direction perpendicular to their propagation. To measure this deviation, we will simply measure the distance between two objects. To begin we return to equation 1.19, the gravitational wave solution to the vacuum Einstein equations in the transverse-traceless gauge:

$$h_{ab}^{\mathrm{TT}}(t,z) = \begin{pmatrix} h_+ & h_\times \\ h_\times & -h_+ \end{pmatrix}_{ab} \cos\left[\omega(t-z)\right]. \tag{1.19}$$

To measure the space-time interval, or proper distance, between two points in space while a gravitational wave is passing, we will use the gravitational wave line element in the transverse-traceless gauge. Equation 1.19 gives only the components of the perturbation metric. We will need the components of the full space-time metric $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}^{\mathrm{TT}}$, where $\boldsymbol{\eta}$ is the flat Minkowski background space. The line element is thus:

$$\begin{aligned} ds^2 = &- \mathrm{d}t^2 + \big(1 + h_+ \cos\left[\omega(t-z)\right]\big)\mathrm{d}x^2 + \big(1 - h_+ \cos\left[\omega(t-z)\right]\big)\,\mathrm{d}y^2 \\ &+ 2h_\times \cos\left[\omega(t-z)\right]\mathrm{d}x\mathrm{d}y + \mathrm{d}z^2. \end{aligned} \tag{3.1}$$

If we have two objects located at coordinates $(t, x_1, 0, 0)$ and $(t, x_2, 0, 0)$, respectively, then the proper distance between them is:

$$ds = (x_2 - x_1)\sqrt{1 + h_+ \cos\omega t}. \tag{3.2}$$

We will define the coordinate separation of the two objects as $L = x_2 - x_1$. Using the fact that the amplitude of the gravitational waves is small, we may approximate equation 3.2 as:

$$ds = L(1 + \tfrac{1}{2}h_+ \cos \omega t). \tag{3.3}$$

In section 2.1 we estimated the amplitude of a galactic, neutron star binary system at Earth to be order 10^{-18} (equation 2.6). For two objects separated by a few km this results in a proper distance change of 10^{-15} m. For terrestrial gravitational wave detectors to have any hope of detection, they must measure distances to incredible precision. To do so, we turn to laser interferometry.

3.1. Interferometers as Gravitational Wave Detectors

In a Michelson interferometer laser light is sent through a beam-splitter and then along two separate paths. The light reflects off of distant mirrors and returns to the beam-splitter, where it is recombined and focused on a photodetector. When the light is recombined the two beams will interfere, depending on the relative length of the two paths. From the interference pattern we can infer the differential path length, ΔL, to fractions of the wavelength of the light. A diagram of a Michelson interferometer is shown in figure 3.1.

The existent interferometer gravitational wave detectors are kilometer scale Michelson interferometers with their two light travel paths, or arms, at right angles to each

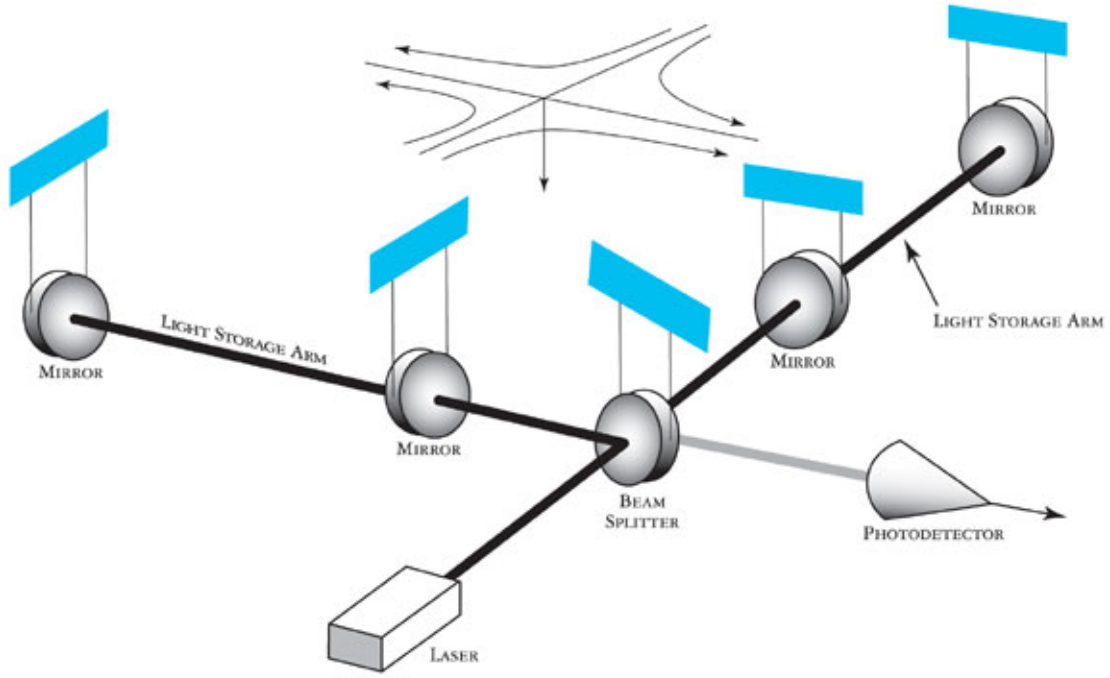

Figure 3.1. Simplified diagram of a Michelson interferometer gravitational wave detector. Laser light after passing through a beam-splitter is sent down each arm where it bounces back and forth multiple times in the light storage arm. It is eventually recombined and measured by a photodetector. A gravitational wave with polarization aligned to the arms of the detector is incident from above (courtesy LIGO Laboratory).

other. If a gravitational wave is incident from above, then the two arms will stretch and shrink, depending on the polarization of the wave.

Let us define the origin of a coordinate system at the beam-splitter, and align the x and y axes along the interferometer arms. We choose to place the two mirrors at the end of the arms, each at a distance L from the origin. If a 'plus' polarized gravitational wave is incident along the z-axis, the proper distance between two points in the space-time is given by equation 3.1, which reduces to:

$$ds^2 = -\mathrm{d}t^2 + (1+h)\,\mathrm{d}x^2 + (1-h)\,\mathrm{d}y^2, \tag{3.4}$$

where $h = h_+ \cos[\omega(t-z)]$. A photon will travel in one of our interferometer's arms along a null geodesic, $ds^2 = 0$. Its path will be modified depending upon which arm it is in:

$$\begin{aligned} \mathrm{d}x &= \mathrm{d}t\sqrt{1+h} \approx \left(1+\tfrac{1}{2}h\right)\mathrm{d}t = L_x, \\ \mathrm{d}y &= \mathrm{d}t\sqrt{1-h} \approx \left(1-\tfrac{1}{2}h\right)\mathrm{d}t = L_y. \end{aligned} \tag{3.5}$$

Here the approximation assumes $h \ll 1$. The differential path is then:

$$L_x - L_y = h\,\mathrm{d}t = hL, \tag{3.6}$$

where L is the flat-space path length: the coordinate length of the arms. The gravitational wave produces a strain, or fractional length change: $h = \Delta L/L$.

Gravitational waves that are incident from an arbitrary direction will induce a smaller strain than the optimal case described above. Geometrically, we can gain some intuition as to the angular sensitivity of an interferometer to gravitational wave strain. A purely 'plus' polarized gravitational wave incident from above with the polarization aligned to the detector arms produces the largest strain. In this case one arm shrinks while the other grows. If we rotate the polarization by 45°, then both detector arms will stretch and shrink together. Even though the proper distance between the mirrors and beam-splitter is changing, both arms stretch symmetrically, so the strain is zero.

If a gravitational wave is incident along one of the arms, then that arm will not be affected, but the other will still experience a strain dependent upon the polarization

of the waves. If the polarization is aligned with the second arm, the affect will be maximal (although only one fourth as strong as the first case). These geometrical concerns are summarized by the detector response functions, computed by considering the strain measured from an arbitrary gravitational wave source:

$$F_+ = -\tfrac{1}{2}\left(1+\cos^2\theta\right)\cos 2\phi\,\cos 2\psi - \cos\theta\,\sin 2\phi\,\sin 2\psi,$$
$$F_\times = \tfrac{1}{2}\left(1+\cos^2\theta\right)\cos 2\phi\,\sin 2\psi - \cos\theta\,\sin 2\phi\,\cos\psi, \tag{3.7}$$

where ψ is the polarization angle relative to the detector's reference frame and (θ, ϕ) defines the propagation direction of the gravitational wave in the detector frame (in spherical polar coordinates). The detector response functions are shown as a surfaces in figure 3.2.

From the detector response we can compute the strain produced in a detector from an arbitrary gravitational wave source:

$$h(t) = h_+(t)\,F_+(\theta,\phi,\psi) + h_\times(t)\,F_\times(\theta,\phi,\psi) \tag{3.8}$$

It will also prove useful to define the effective distance to a gravitational wave source, D_{eff}:

$$D_{\text{eff}} = \frac{D}{\sqrt{{F_+}^2(1+\cos^2\iota)^2 + {F_\times}^2\cos^2\iota}}, \tag{3.9}$$

where D is the physical distance to the source and ι is the angle between the z-axis of the radiating system and the line of sight to the detector. This inclination angle was introduced in chapter 2. The effective distance is a single number summarizing the ease or difficulty of detection. A source that is misaligned with our detector will

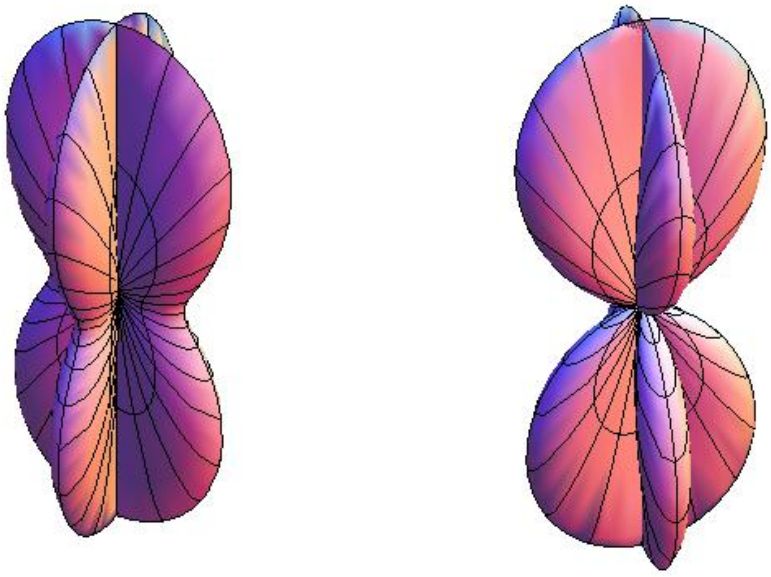

Figure 3.2. Detector response of a Michelson interferometer to gravitational waves with fixed polarization as a function of gravitational wave propagation direction. Left panel is the response to 'plus' polarized waves and the right panel is 'cross'.

appear farther away and have a larger effective distance. The effective distance can be inferred directly from the amplitude of a gravitational wave observed in a detector even when no accurate astrophysical point of origin is known.

The total strain measured in a gravitational wave detector, $s(t)$, is the sum of many independent noise sources, $n(t)$, and any gravitational wave strain $h(t)$. The detection problem is to determine whether or not an $h(t)$ is present. If we have $h(t) \gg n(t)$, this is not a problem. However, when $h(t) \lesssim n(t)$, we may run into difficulties. In some sense $n(t)$ represents an approximate threshold for detecting

gravitational waves. By comparing $n(t)$ for different detectors, we can compare their relative performance.

For now we will assume that our detector noise originates from a stationary random process. In this case the expectation value of the noise is a time average:

$$\langle n \rangle = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} n(t)\,\mathrm{d}t \tag{3.10}$$

Without loss of generality we can assume that $\langle n \rangle = 0$. In this case the more useful measure is $\langle n^2 \rangle$. We will assume a finite time measurement, saying that $n(t) = 0$ for $t < T_-$ and $T_+ < t$. In this case taking the limit for the integration is trivial. Then we can move to the Fourier domain:

$$\begin{aligned}
\left\langle n^2 \right\rangle &= \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} n^2(t)\,\mathrm{d}t, \\
&= \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} \left|\tilde{n}^2(f)\right|\,\mathrm{d}f, \\
&= \lim_{T\to\infty} \frac{2}{T} \int_{0}^{\infty} \left|\tilde{n}^2(f)\right|\,\mathrm{d}f, \\
&= \int_{0}^{\infty} S_n(f)\,\mathrm{d}f,
\end{aligned} \tag{3.11}$$

where the tilde denotes the Fourier transform of a function and we define $S_n(f)$ as the noise power spectral density:

$$S_n(f) = \lim_{T\to\infty} \frac{2}{T} \left| \int_{-T/2}^{T/2} n(t) \exp(-2\pi\imath f t)\,\mathrm{d}t \right|^2. \tag{3.12}$$

From the noise power spectral density we can define a noise-weighted inner product:

$$\langle a \,|\, b \rangle = 2 \int_{-\infty}^{\infty} \frac{\tilde{a}(f)\,\tilde{b}^\star(f)}{S_n(f)}\,\mathrm{d}f, \tag{3.13}$$

where the asterisk denotes complex conjugation.

What sort of strain sensitivity can we achieve with an interferometer gravitational wave detector? As discussed in section 2.1, to detect an extra-galactic neutron star binary we will need to measures strains of order 10^{-21}. We will begin by conservatively estimating that our interferometer can measure differential length to an accuracy of the wavelength of its light. For an infrared laser of wavelength $\lambda = 1\mu$m and a kilometer scale interferometer we can measure strains of order $h \sim 10^{-9}$. We will need to do better.

An obvious improvement to make is to increase the interferometer arm length. Due to the constraints of terrestrial construction, it is impractical to build an interferometer larger than the kilometer scale, but we can increase the effective arm length by sending our laser light back and forth along each arm multiple times before recombining. This is the function of the light storage arms of figure 3.1. Additional mirrors at the beam-splitter end of the interferometer send the light back towards the end mirror. This traps the laser light in a Fabry-Perot cavity, increasing the total path traveled. The effective arm length can increase by a factor of 10^3 [34], which brings us to a strain sensitivity of $h \sim 10^{-12}$ – not even close to what we need.

Our estimate of the differential arm length measure was conservative. The limiting factor in measuring the small levels of interference is photon shot noise. If we collect N photons at a photodetector, our ability to detect small fluctuations in light beyond normal Poisson fluctuations affects our differential arm length measure as:

$$\Delta L \approx \frac{\lambda}{\sqrt{N}}. \tag{3.14}$$

To resolve one wavelength of a gravitational wave, we need to collect all of our photons in one period of the wave, $T = 1/f_{\mathrm{GW}}$. The number of photons collected depends both on the gravitational wave frequency, f_{GW}, and the laser power P:

$$N = \frac{P}{E}T = \frac{2\pi\, P\lambda}{\hbar\, f_{\mathrm{GW}}}, \tag{3.15}$$

where E is the photon energy and $\hbar$ is Planck's constant. For a 1 Watt laser with wavelength 1μm and a 200 Hz gravitational wave, we can collect $N \sim 10^{18}$ photons. From equation 3.14 our differential arm length sensitivity is $\Delta L \approx 10^{-9}\lambda$. Our strain sensitivity is now $h \sim 10^{-21}$, which is exactly what we need.

Further improvements can be made to increase the sensitivity beyond this level. A power recycling mirror can be used to increase the photon count in the Fabry-Perot cavity, raising the effective laser power. While photon shot noise is the dominant noise source for interferometers at high frequencies greater than about 200 Hz, at low frequency seismic noise grows rapidly. Better suspension systems on the mirrors can provide boosts in low frequency sensitivity. In the intermediate regime thermal noise, causing expansion and contraction of the mirror surfaces and suspension wires, takes over. KAGRA, a Japanese interferometer gravitational wave detector currently under construction, will cryogenically cool its systems to combat this problem [5]. It is a combination of all of these noise sources that defines the detector noise floor. The levels of these three major interferometer noise sources are shown in figure 3.3 for a first generation, 4 km interferometer gravitational wave detector.

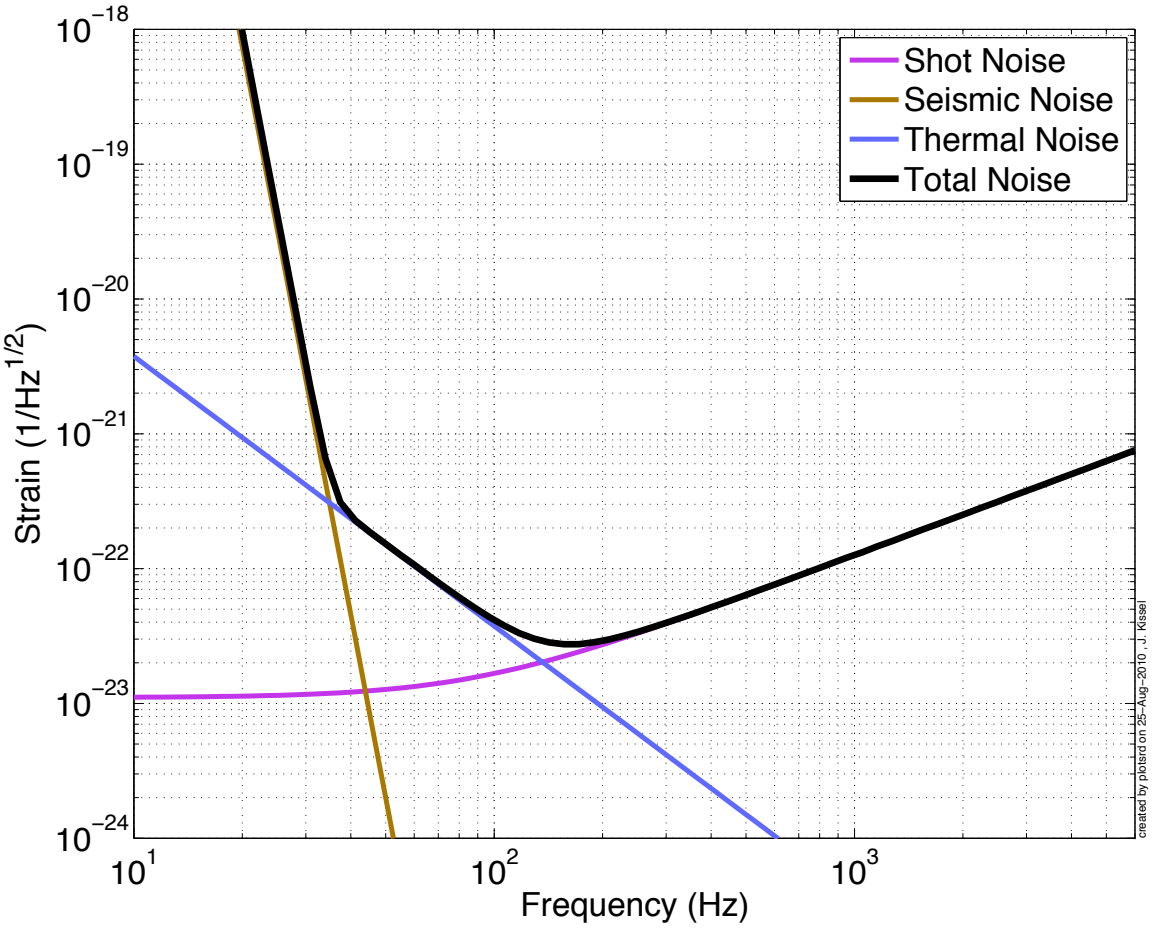

Figure 3.3. Strain sensitivity limitations from the three major noise sources in initial LIGO: seismic, thermal, and photon shot, as a function of frequency (taken from [34]).

Operating a power-recycled, Fabry-Perot, Michelson interferometer is a tricky business. The control systems for the power recycler must hold the two mirrors at a separation that is an integer multiple of the wavelength of laser light, so, when new photons are added, they constructively interfere [34]. When the two mirrors are properly spaced and the power recycler is functioning, we say the interferometer has achieved 'lock'. In order to hold the detectors in a locked state we must apply forces

to the mirrors, holding them at the correct separation. This active feedback system provides non-linear coupling between many of the control subsystems.

If each noise source were an independent random processes, then the resulting detector noise would be Gaussian with a colored spectrum. Unfortunately, the non-linear detector couplings break this simple noise scenario. There are two effects to contend with. First is non-stationarity. Depending on the equilibrium state of the various control systems, the level of the noise floor will fluctuate on differing time scales. In the best cases the noise floor fluctuates slowly, providing near stationarity for tens of minutes at a time. In these cases it is still possible to accurately estimate the noise power spectral density, which is necessary for most data analysis techniques.

The worst non-Gaussian noise appears as sudden, short term, high amplitude fluctuations. These so called 'glitches' can cause false positive gravitational wave events in a data analysis pipeline. Entire data analysis groups are dedicated to detector characterization, where the appearance of certain common glitches is correlated with environmental factors in hopes of tuning the control systems to eliminate them in the future.

3.2. The Laser Interferometer Gravitational Observatory

After the preceding discussion it should be no surprise that the Laser Interferometer Gravitational Observatory (LIGO) is a network of power-recycled, Fabry-Perot, Michelson interferometers dedicated to the direct detection of gravitational waves.

LIGO collected its first gravitational wave data in 2002. This initial era, or iLIGO, continued until late 2007 and included five distinct periods of data collection or 'science runs'. Three interferometers operated during iLIGO: one 4 km detector in Livingston, Louisiana (L1); and two detectors in Hanford, Washington – a 4km and a 2 km (H1, H2).

During S5, the final iLIGO science run, the 3 km Virgo detector came online for its first science run (VSR1). S5 collected data for approximately two years achieving the initial LIGO design sensitivity. In 2008 the LIGO and Virgo detectors went offline for hardware upgrades. The enhanced detector era spanning 2009 to 2010 followed these upgrades. The 2 km detector at Hanford was decommissioned during this time. Due to stability problems in the enhanced LIGO (eLIGO) detectors, they never significantly exceeded iLIGO in sensitivity. A summary of the initial and enhanced ground based detector eras is shown in table 3.1.

The LIGO and Virgo detectors are currently undergoing a series of additional upgrades to their advanced designs. In addition the LIGO Laboratory will soon begin building a third detector in India (IndIGO), and the cryogenically cooled KAGRA detector mentioned above is currently under construction in Japan. One key design element of the advanced detectors is a greatly improved suspension system to mitigate seismic noise. Better seismic isolation will provide enhanced low frequency sensitivity. This is especially important for black hole sources, as high mass systems radiate at low frequency as seen in chapter 2.

The advanced LIGO (aLIGO) detectors have begun preliminary engineering runs and are poised to collect their first science data in 2015. aLIGO is expected to make the first direct gravitational wave detections upon reaching design sensitivity. The full advanced network of detectors, expected around 2020, should usher in the era of gravitational wave astronomy, where routine gravitational wave detections can compliment current electro-magnetic observations.

The LIGO detectors, with a peak sensitivity in the 100 to 1000 Hz range, were designed with binary neutron star (NS) systems in mind. As discussed in section 2.1, NS-NS binary systems will supply many gravitational wave cycles in this frequency range as the two stars inspiral toward each other and collide. By coherently adding the effects of multiple gravitational wave cycles it is possible to detect gravitational waves with instantaneous amplitudes below the LIGO noise floor. We will discuss such data analysis techniques in chapter 4.

Phase	Run Name	Detectors	Start Date	End Date	Duration
iLIGO	S1	H1,H2,L1	23 Aug 2002	8 Sept 2002	17 d
	S2	H1,H2,L1	14 Feb 2003	14 Apr 2003	60 d
	S3	H1,H2,L1	31 Oct 2003	9 Jan 2004	71 d
	S4	H1,H2,L1	22 Feb 2005	24 Mar 2005	31 d
	S5	H1,H2,L1	4 Nov 2005	Oct 2007	2 yr
Virgo	VSR1	V1	May 2007	Dec 2007	6 mo
eLIGO	S6	H1,L1	7 July 2009	20 Oct 2010	1 yr
Virgo	VSR2	V1	7 July 2009	11 Jan 2010	6 mo
	VSR3	V1	11 Aug 2010	20 Oct 2010	2 mo

Table 3.1. LIGO and Virgo Science runs of the initial and enhanced eras.

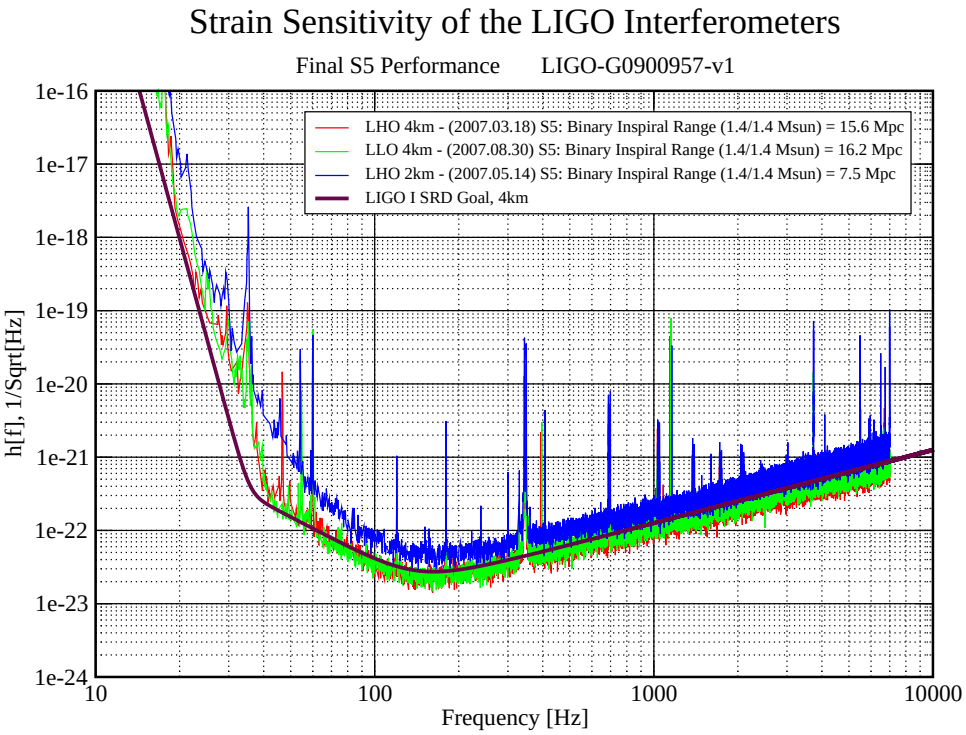

Figure 3.4. Strain sensitivity of the LIGO detectors during the fifth science run (S5). The vertical spikes correspond to resonance frequencies in the various LIGO subsystems (courtesy LIGO Laboratory).

A common performance summary statistic is the 'Binary Inspiral Range'. This is the distance to which an optimally oriented NS-NS binary would produce a signal-to-noise ratio of 8 in the detector. Similar range statistics can be made for any source. Figure 3.4 shows the stain sensitivity achieved by iLIGO in S5. This sensitivity corresponds to a binary inspiral range of about 10 Mpc, covering the local galactic group ($\sim$ 3 Mpc) and beyond.

The 'Black Hole Ringdown Range' is shown in figure 3.5. Similar to the binary inspiral range, this is the distance to which an optimally oriented black hole will

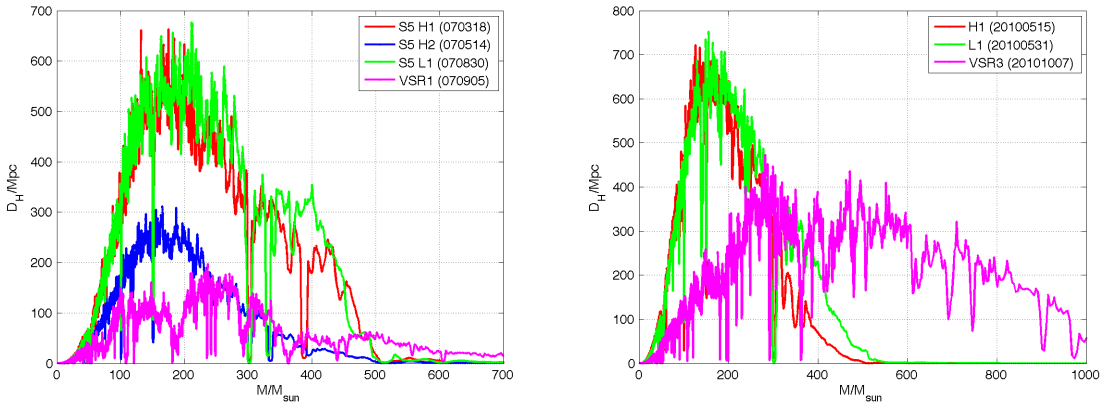

Figure 3.5. Black Hole Ringdown Range achieved during the initial and enhanced detector eras as a function of black hole mass. The left panel shows LIGO S5 and Virgo VSR1, while the right shows LIGO S6 and Virgo VSR3. Note how enhanced Virgo's superior seismic isolation systems provide sensitivity to much higher masses.

produce a signal-to-noise ratio of 8 ringdown signal in a detector. For this distance we assume the black hole has a dimensionless spin of $a = 0.9$ and a radiation efficiency of $\epsilon = 1\%$.

We should note that the initial and enhanced detectors are not particularly sensitive to stellar mass black holes with masses of a few to tens of solar masses. They could, however, detect Intermediate Mass Black Holes (IMBH) with masses of a few hundred solar masses beyond even the Milky Way's parent Virgo Super Cluster (diameter ~ 33 Mpc). In this way LIGO could provide the first concrete evidence as to the existence of IMBHs.

Here we have discussed the interaction of gravitational waves with terrestrial interferometers. With such detectors we can achieve sensitivity to the minute gravitational

wave strains from astrophysical sources. The engineering feat of the existent and future gravitational wave detectors can not be discounted. A large collaboration of scientists and engineers worked for years to build the initial LIGO detectors. While iLIGO did not directly detect gravitational waves, the advanced detectors should achieve the necessary sensitivity to make this discovery.

4. A SEARCH FOR GRAVITATIONAL WAVES FROM BLACK HOLE RINGDOWNS

Identifying black holes (BH) via electromagnetic observation is fundamentally hampered by their blackness. As no light can escape the event horizon, we must rely on inferring the black hole's presence by other means. The existence of Sagitarius A*, the supermassive black hole (SMBH) at the Milky Way's center is inferred by observing the orbits of several nearby stars. Much like Neptune's discovery in 1846, the motion of bodies in the galactic center could only be accounted for with the inclusion of a large unseen mass.

We don't always have the luxury of nearby, resolvable, massive stars. In these cases we must infer BHs by their effects on in-falling matter. A compact object with a companion star will pull material from its companion, forming an accretion disk. As this material approaches the surface of the object, it is heated to very high temperatures and can emit X-ray flares. These so called X-ray binaries are divided by the mass of the accreting object. The low mass X-ray binaries (LMXB) are possibly neutron stars (NS), but the high mass X-ray binaries (HMXB) with masses $\gtrsim 4$ $M_\odot$ are almost certainly BHs. There are about two dozen stellar mass BH (SBH) candidates with masses ranging from 3-20 $M_\odot$. SBHs are believed to be the end state of very massive stars after core collapse.

Black holes can also form columnated jets, ejecting originally inward falling material along the BH spin axis. Such jets are characteristic of SMBHs, found at the

center of most galaxies. There are several dozen SMBH candidates including Sag A*. They have masses ranging from $10^6 - 10^9$ $M_\odot$. The formation of SMBH is matter of debate. The most massive SMBHs likely gained mass by a combination of large scale accretion in galactic centers and repeated mergers with other SMBHs during galactic mergers.

It is no secret that there is a large mass gap in the black hole candidate population. The intermediate mass black holes (IMBH) with masses in the range $10^2 - 10^4$ $M_\odot$ are predicted to occupy this space. While IMBHs have eluded convincing observation, there are some preliminary candidates, suggesting a population may be on the brink of discovery [27].

An astute reader should note that we have referred to all observed BHs as 'candidates'. All BH observations rely on a series of assumptions, any of which could prove false. The only way to ensure we have observed a BH as described by general relativity is by its event horizon. A compact five solar mass object is not a BH unless all of its mass is within a radius of about 7.5 km. Gravitational waves from BH ringdowns probe the event horizon of BHs in a way that electromagnetic observations cannot. By observing BH ringdowns we could confirm the existence of BHs beyond all possible doubt. Gravitational wave observations could also confirm or place strong limits on the existence of IMBHs.

As we discussed in section 2.2, the vibrations of perturbed BHs follow a characteristic spectrum. While in general any of these quasi-normal modes (QNM) can be

excited, we are most concerned with the modes that strongly couple to gravitational wave production. We will also focus on astrophysical events that radiate the most energy, leading to higher amplitude signals.

Any mass or energy will produce QNMs as it falls into a BH, but which astrophysical scenarios will produce the most gravitational radiation? One obvious way to get matter into a BH is through accretion. Unfortunately, any reasonable accretion model will have multiple inflows leading to destructive interference of the QNMs [12]. Numerical simulations of BH-NS mergers confirm this. When the NS is tidally disrupted, the NS material is accreted. The resulting gravitational waveforms have no ringdown, showing an abrupt cutoff at the tidal disruption frequency [51].

Stellar mass BHs (with masses of a few to a few tens of solar masses) likely form via massive stellar core collapse. If the collapse occurs in an asymmetric manner, then QNMs will be excited. Recent simulations show that the mass quadrupole does not change rapidly during collapse, resulting in radiation efficiencies of $\epsilon \sim 10^{-6}$ [6]. Gravitational waves from core collapse are not likely detection candidates.

This brings us to the mergers of compact objects resulting in a BH. We already mentioned why NSs, and matter in general, can dampen QNMs. This leaves BH-BH mergers as the most promising sources of ringdown gravitational waves. In section 2.2 we saw that the $\ell = m = 2$ mode carries most of the radiated energy of a BH ringdown. Because of this fact, we will focus a search for gravitational waves from BH ringdowns on this particular QNM. In general it would be best to search for all QNMs:

frequency (Hz)	spin	Mass ($M_\odot$)	Quality
2000	0	5.86	2.12
	0.7	8.51	3.29
	0.95	11.8	7.03
50	0	234	2.12
	0.7	341	3.29
	0.95	471	7.03

Table 4.1. Black hole masses to produce given quasi-normal mode frequencies for various dimensionless spin parameters.

as the mass ratio of merging bodies grows, a greater fraction of energy is radiated in other modes. However, for high mass ratio mergers less energy is radiated period. This makes the dominant mode harder to detect, but the sub-dominant modes are harder still. A single mode search has the two-fold advantage of capturing most of the power from the most promising sources and computational simplicity [16].

The Laser Interferometer Gravitational Observatory (LIGO) is most sensitive to gravitational waves on the order of 100 Hz. In its initial and enhanced configurations LIGO had good sensitivity in the range 50 Hz to 2000 Hz. Which black holes will produce gravitational waves in this range? We can invert equations 2.27 and 2.28 to determine what BH masses and spins will produce radiation detectable by LIGO. The limiting cases are shown in table 4.1

In order to search for BH ringdowns we will use a template waveform, comparing this known waveform with the data. For computational ease we will use the simplest possible wave form:

$$h(t;\, t_0, f, Q) = A \exp\left(-\pi \frac{f}{Q}(t - t_0)\right) \cos\left(2\pi f (t - t_0)\right), \tag{4.1}$$

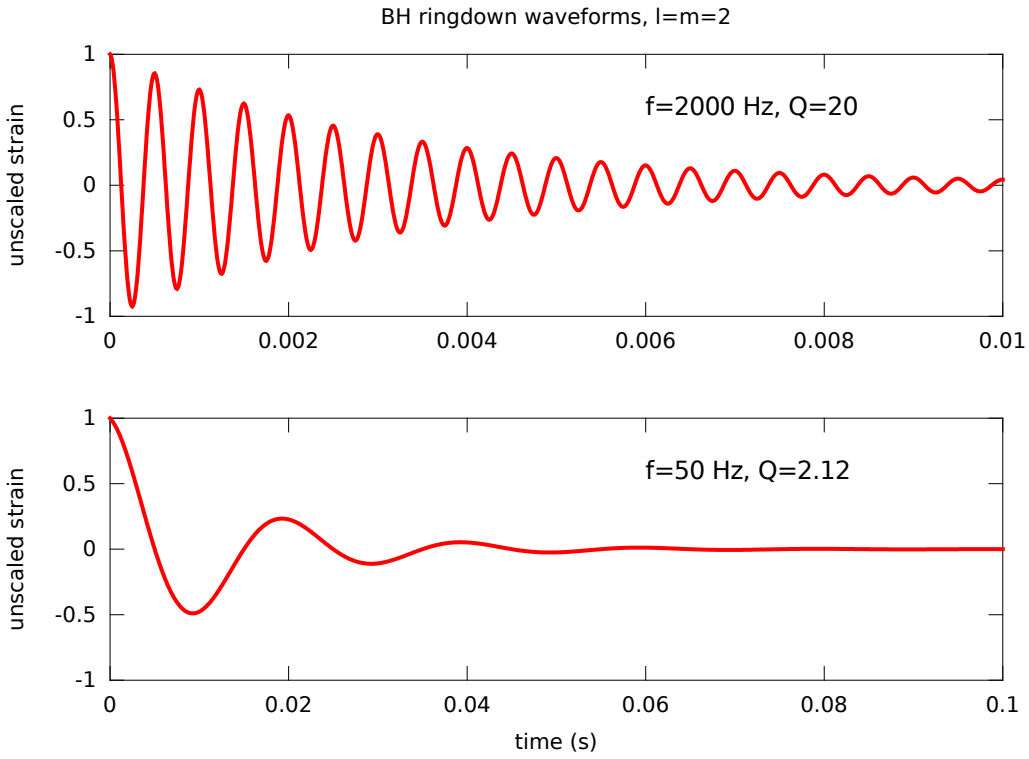

Figure 4.1. Characteristic black hole ringdown waveforms for the $\ell = m = 2$ mode. Top panel is high frequency and high quality, while the bottom is low frequency and quality. The amplitudes are unscaled, and would in practice carry a factor of $A = \frac{M}{r} A_{22}$.

where A is the strain amplitude containing polarization, geometrical, and distance factors for the source and t_0 is the time of arrival in the detector. In addition we have dropped the ℓ, m indices[1]. Creighton and Nakano *et al.* showed that even this simplest of waveforms is sufficient for detection [21] [30]. We can safely ignore the initial phase of the waveform, because of a partial degeneracy with amplitude and time of arrival.

[1]We will follow this practice from here on, assuming to always be working with the $\ell = m = 2$ mode unless otherwise stated.

4.1. Single Interferometer Analysis

Armed with a known gravitational waveform, we can now attempt to extract such signals from our detector output. A gravitational wave detector experiences a strain, s, from environmental noise sources, n, and, possibly, gravitational waves, h:

$$s(t) = n(t) + h(t). \tag{4.2}$$

We can determine if there is a signal present by transforming to a new basis via an integral transform.

$$Z(t_0) = \int \mathcal{K}(t - t_0)s(t)\,\mathrm{d}t \tag{4.3}$$

Our new basis is not orthogonal, so we cannot reconstruct our original strain by adding up all of our basis functions like a Fourier transform. However we can try to find one basis function, $\mathcal{K}$, that by itself contains the most information about the original strain. So we want to choose basis functions that maximize the transformed strain, $Z(t_0)$. To this end we break Z into the two parts of the sum, defining N and H:

$$\begin{aligned} Z(t_0) &= \int \mathcal{K}(t - t_0)n(t)\,\mathrm{d}t + \int \mathcal{K}(t - t_0)h(t)\,\mathrm{d}t \\ &= N(t_0) + H(t_0) \end{aligned} \tag{4.4}$$

We may take an ensemble average over many instances of $N(t_0)$, assuming the noise results from a stationary, random process that is Gaussian distributed with zero mean

(even though we know this not to be true):

$$\begin{aligned}\left\langle N^2 \right\rangle &= \lim_{T\to\infty} \frac{1}{T} \int_0^\infty \left| \tilde{N}^2 \right| \, \mathrm{d}f \\ &= \int_0^\infty \left| \tilde{\mathcal{K}}(f, t_0) \right|^2 S_n(f) \, \mathrm{d}f, \end{aligned} \tag{4.5}$$

recalling the definition of the noise spectral density in equation 3.12. Using the convolution theorem, we may write the transformed signal as:

$$H(t_0) = 2 \int_0^\infty \tilde{\mathcal{K}}^\star(f, t_0) \tilde{h}(f) \, \mathrm{d}f. \tag{4.6}$$

We now define the signal-to-noise ratio, ξ. We can determine if a signal is present in in our detector strain, by determining whether ξ is large.

$$\begin{aligned}\xi(t_0) &= \frac{H^2(t_0)}{\left\langle N^2 \right\rangle} \\ &= \frac{4 \left| \int_0^\infty \tilde{\mathcal{K}}^\star(f, t_0) \, \tilde{h}(f) \, \mathrm{d}f \right|^2}{\int_0^\infty \left| \tilde{\mathcal{K}}(f, t_0) \right|^2 S_n(f) \, \mathrm{d}f} \\ &= \frac{4 \left| \int_0^\infty \left[\tilde{\mathcal{K}}^\star(f, t_0) {S_n}^{1/2}(f) \right] \left[\tilde{h}(f) {S_n}^{-1/2}(f) \right] \, \mathrm{d}f \right|^2}{\int_0^\infty \left| \tilde{\mathcal{K}}(f, t_0) \right|^2 S_n(f) \, \mathrm{d}f} \qquad (4.7) \\ &\leq \frac{4 \left[\int_0^\infty \left| \tilde{\mathcal{K}}^\star(f, t_0) {S_n}^{1/2}(f) \right|^2 \, \mathrm{d}f \right] \left[\int_0^\infty \left| \tilde{h}(f) {S_n}^{-1/2}(f) \right|^2 \, \mathrm{d}f \right]}{\int_0^\infty \left| \tilde{\mathcal{K}}(f, t_0) \right|^2 S_n(f) \, \mathrm{d}f} \qquad (4.8) \end{aligned}$$

If the numerator is to be maximized, then the Cauchy-Schwartz inequality connecting equations 4.7 and 4.8 should be a true equality. This is only the case if the two integrands on the right hand side are equivalent up to a constant factor:

$$\tilde{\mathcal{K}}^2(f, t_0) \, S_n(f) = A^2 \frac{\tilde{h}^2(f)}{S_n(f)} \tag{4.9}$$

We now determine the optimal basis to represent $s(t)$:

$$\tilde{\mathcal{K}}(f, t_0) = A\frac{\tilde{h}(f)}{S_n(f)}. \tag{4.10}$$

Our basis functions are the gravitational wave strain itself (scaled by the noise spectral density). If there are no gravitational waves present, then to optimal representation must be $\mathcal{K} = 0$. This is not surprising, as we assumed the noise has zero mean. If a basis function, $\mathcal{K}$, maximizes $Z(t_0)$ then there is likely to be a gravitational wave strain of the form $\mathcal{K}$ in our data.

We will apply the new expression for $\mathcal{K}$ back into equation 4.3. After using the convolution theorem we have:

$$Z(t_0) = \int_{-\infty}^{\infty} \frac{\tilde{s}(f)\,\tilde{h}^{\star}(f, t_0)}{S_n(f)}\,\mathrm{d}f = \langle s \,|\, h(t_0)\rangle\,. \tag{4.11}$$

We see that this is the noise-weighted inner product defined in equation 3.13. We may also compute the variance of our basis function:

$$\sigma^2 = \int_{-\infty}^{\infty} \frac{\tilde{h}(f)\,\tilde{h}^{\star}(f)}{S_n(f)}\,\mathrm{d}f = \langle h \,|\, h\rangle\,. \tag{4.12}$$

Using this, we can scale the transformed detector strain in the new basis, which is equivalent to normalizing our basis function. The new basis representation is a signal-to-noise ratio (SNR) time series:

$$\rho(t_0) = \frac{Z(t_0)}{\sigma} = \frac{\langle s \,|\, h(t_0)\rangle}{\sqrt{\langle h \,|\, h\rangle}}. \tag{4.13}$$

$\rho(t_0)$ tells us how much the detector strain at a given time is consistent with there being a gravitational wave of the form h arriving at time t_0. Our basis functions

act as a gravitational wave template. As expected, by knowing *a priori* what sort of gravitational waves we wish to find we improve our ability to detect them. This method is sometimes referred to as 'matched filtering' and has been used in signal processing since the early 20th century.

We can use the SNR time series to determine whether candidate gravitational waves are present in the detector strain by establishing a simple threshold $\bar{\rho}$.

$$\text{If} \quad \rho \begin{cases} < \bar{\rho}, & \text{no candidate GW present.} \\ \geq \bar{\rho}, & \text{candidate GW present.} \end{cases} \tag{4.14}$$

In the presence of Gaussian noise choosing a threshold would be trivial. A candidate event with $\rho = 5$ would be a 5σ occurrence with a false alarm probability of 5×10^{-7}. But we do not have Gaussian noise. We will come back to the issue of setting a threshold in section 4.5.

Because the QNM eigenfrequencies $\omega_{\ell m}$ are a function of the BH mass and dimensionless spin alone, by measuring f and Q of a gravitational wave in our detector accurately we can measure the fundamental parameters of the vibrating BH. This will be a key point in terms of the astrophysics we can learn from gravitational waves. To determine the parameters of the source, we need to try multiple templates each with different frequencies and qualities. Since the template with the exact same parameters as the source maximizes the SNR, the template that provides the largest SNR must have parameters closest to the actual signal. Due to computational constraints we cannot possibly try every template against the data. We need a clever scheme

to limit the number of templates we use and still give accurate source parameter recovery.

As a first method to cut the number of necessary templates, we can restrict our search to the frequencies and qualities that LIGO is most sensitive to. We will only search for signals with frequencies between 50 and 2000 Hz. In addition we will only search for signals with qualities between 2.1187 and 20.0. This range of ringdown qualities covers BH dimensionless spins from 0 to about 0.995.

To determine how many templates we actually need to cover this range of ringdown parameters, consider the match, $M_{ij} = \langle h_i \,|\, h_j \rangle$, between any two normalized templates i and j. We will lay out a grid of points in the template parameter space such that no two neighboring points differ by more than 3%. That is to say any neighboring h_i and h_j will have $M_{ij} \geq 0.97$. This most easily accomplished with a template space metric. We will discuss how to compute such metric in the next section in the context of the multi-detector coincident analysis.

For now it will suffice to say that we can layout a bank of templates that satisfies these requirements. It turns out the ringdown quality factor is not well determined in our signals. This is because missing or adding one cycle, particularly in the low amplitude tail, does not affect the match very much. This is especially true of high quality signals where there are already plenty of cycles carrying the power. In contrast with this, small changes in ringdown frequency will have a large effect on template match. This mismatch depends on the fractional frequency difference between one

template and the next, so our templates should be densest at low frequencies. We can see a small region of the template bank in figure 4.2. In actuality it takes only about 600 templates to cover the range of parameters with the parameter recovery accuracy we require. This template bank is nothing more than a simple list of parameters, f and Q, used to generate gravitational wave templates for our search.

To find a gravitational wave signal we will try each of the templates in our bank, computing the SNR as a function of time. If the SNR at a given time is greater than

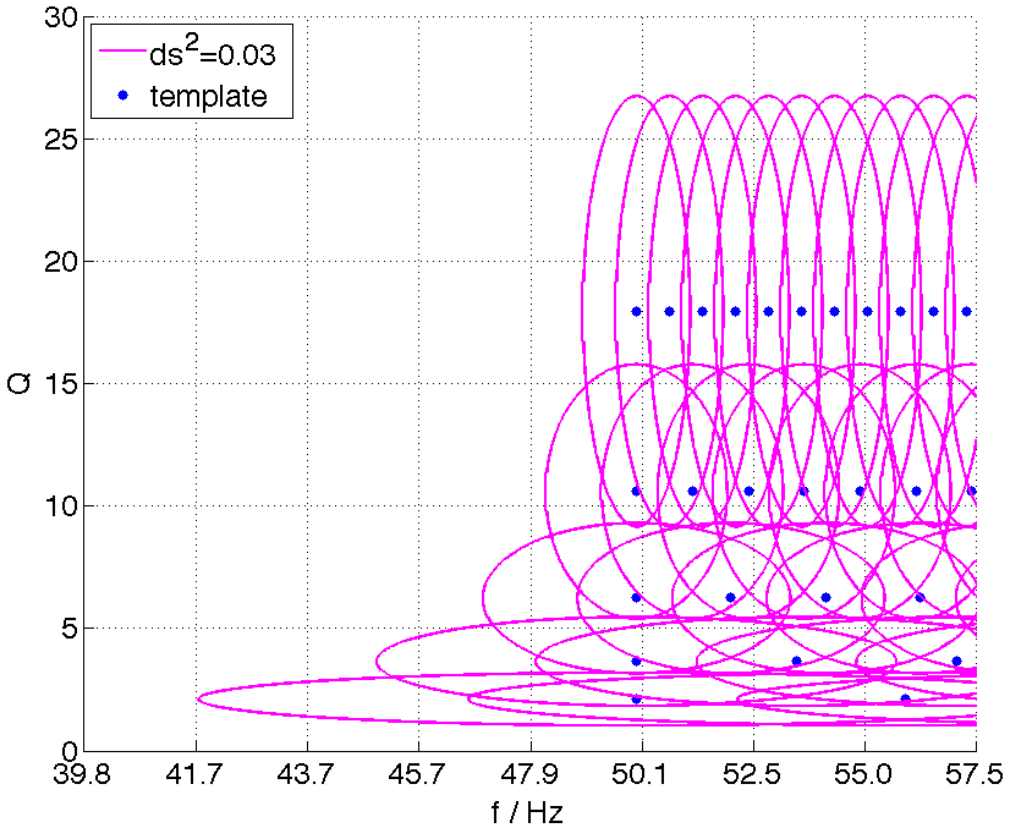

Figure 4.2. Template bank layout in the LIGO S5 and S6 search for black hole ringdowns. Templates are spaced uniformly in log frequency and quality. Contours show a parameter space metric distance of 0.03, or 97% match.

some predetermined threshold, then we have a candidate gravitational wave event or analysis 'trigger'. We will store the template parameters, the time, and the SNR of the trigger for later use. We can also compute some additional parameters that will prove useful. One of these is a reference gravitational wave amplitude $\mathcal{A}$:

$$\mathcal{A} = \frac{M}{r} A_{22} \tag{4.15}$$

This detector amplitude is computed using equation 2.31 to determine A_{22} assuming a radiation efficiency of 0.01. As a mere reference amplitude, we assume a distance to the source of 1 Mpc. The choice of radiation efficiency adds a bias to these reference amplitude estimates, but it is order unity for the systems to which we are most sensitive.

From the reference gravitational wave amplitude we can compute the detector sensitivity, ς, to candidate signal at 1 Mpc:

$$\varsigma^2 = \sigma^2 \mathcal{A}^2, \tag{4.16}$$

where σ is the template variance defined in equation 4.12. Now we can compute the effective distance, D_{eff}, to the source. We defined the effective distance geometrically in equation 3.9. But now we can determine it based on the SNR, $\rho(t_0)$, seen in a detector:

$$D_{\text{eff}} = \frac{\varsigma}{\rho(t_0)} \times 1\,\text{Mpc}. \tag{4.17}$$

The effective distance will act as an intuitive measure of the detectability of a particular source. It will be especially useful when testing the efficacy of the search using simulated gravitational wave signals.

4.2. Coincident Analysis

After conducting a single interferometer search using each of our detectors, we have a list of candidate event triggers containing several parameters. Due to the non-Gaussian detector noise, this list will contain many triggers that result from noise not gravitational waves. The first line of defense against these false alarms is to combine the information from all of our detectors. Because gravitational waves are of astrophysical origin, we expect gravitational wave signals to appear in all detectors. Noise on the other hand arrises from sources local to each detector. False alarms therefore should not pass a multi-detector coincidence test.

The simplest coincidence test compares the start time of triggers in multiple detectors, passing events that are within some time window and failing events that are not. We can do better than this, since we have access to several parameters in addition to just the start time of the triggers. We can compare several of these parameters and measure the compatibility of triggers from different detectors.

As we mentioned in the previous section, we can measure the compatibility of two triggers with the match M_{ij}. Equivalently, we will define the distance between

two triggers in parameter space, or mismatch as:

$$ds^2 = 1 - \langle h_i \,|\, h_j \rangle \,. \tag{4.18}$$

for two different triggers h_i and h_j. Our triggers are actually the recovered gravitational waveform templates parametrized by a vector of parameters $\boldsymbol{\theta}$. If we assume the difference between the parameters of our two triggers is small, we can take the continuum limit and express the full parameter space distance in terms of a metric:

$$ds^2 = g_{\alpha\beta}\, \mathrm{d}\theta^\alpha \mathrm{d}\theta^\beta, \tag{4.19}$$

where α and β run over all parameters. The components of the parameter space metric are given by [20]:

$$g_{\alpha\beta} = \frac{\langle \partial_\alpha h \,|\, \partial_\beta h \rangle}{\langle h \,|\, h \rangle} - \frac{\langle h \,|\, \partial_\alpha h \rangle \, \langle h \,|\, \partial_\beta h \rangle}{\langle h \,|\, h \rangle}^2 \,. \tag{4.20}$$

The ringdown templates used in our search are parametrized by four parameters (f, Q, t_0, A): the ringdown frequency and quality, the time of arrival in the detector, and the amplitude at the detector. The relative amplitude observed in two detectors is a function of the point of origin of the gravitational wave, because of the two detectors differing angular response. As our single interferometer triggers contain no sky location information, we cannot make any statements about the expected relative amplitude in two detectors. Because of this, we will work with normalized templates in a three dimensional parameter space $\boldsymbol{\theta} = (f, Q, t_0)$.

The simplicity of the ringdown waveform lets us approximate equation 4.20 analytically. Consider the noise spectral density, $S_n(f)$, in the denominator of our

noise-weighted inner product. For small $\mathrm{d}f$ we can assume that $S_n(f) \approx S_n(f + \mathrm{d}f)$. In this case we can factor the noise-weight out of the inner product and simplify the calculation:

$$\langle a \,|\, b \rangle \approx \frac{2}{S_n(f)} \int_0^\infty a(t)\, b(t)\, \mathrm{d}t \tag{4.21}$$

The necessary differentiation and integration is easily done analytically, resulting in a line element for the ringdown parameter space:

$$\begin{aligned} ds^2 =& \mathrm{d}f^2 \frac{1 + 6Q^2 + 16Q^4}{4f^2(1 + 2Q^2)} + \mathrm{d}Q^2 \frac{1 + 28Q^4 + 128Q^6 + 64Q^8}{4Q^2(1 + 6Q^2 + 8Q^4)^2} \\ &+ {\mathrm{d}t_0}^2 \pi^2 f^2 \frac{1 + 4Q^2}{Q^2} - \mathrm{d}f\mathrm{d}Q \frac{1 + 2Q^2 + 8Q^4}{2fQ(1 + 6Q^2 + 8Q^4)} \\ &- \mathrm{d}t_0 \mathrm{d}f\, 2\pi Q \frac{1 + 4Q2}{1 + 2Q2} + \mathrm{d}t_0 \mathrm{d}Q\, 2\pi f \frac{1 - 2Q^2}{(1 + 2Q^2)^2} \end{aligned} \tag{4.22}$$

There is one further complication. Because the detectors are spatially separated, the gravitational wave will arrive at the various detectors at different times. This can be accounted if we know the point of origin of the gravitational wave. Without that information we can only check whether the two recovered times are consistent with any possible source sky location. The maximum signal travel time between the two LIGO detector sites is about 14 ms. On top of this there could be measurement error in the time of arrival due to noise. To avoid artificially increasing the metric distance between triggers due to signal travel time, we will assume[2] that the signal travel corrected $\mathrm{d}t_{\mathrm{eff}}$ between triggers is the one that minimizes ds^2:

$$\mathrm{d}t_{\mathrm{eff}} = -\frac{g_{tf}\mathrm{d}f + g_{tQ}\mathrm{d}Q}{g_{tt}}, \tag{4.23}$$

[2]Unless of course $\mathrm{d}t_{\mathrm{eff}}$ is outside of the range allowable by signal travel time. In this case we take $\mathrm{d}t_{\mathrm{eff}}$ to be the maximum allowed by signal travel.

where g_{tf}, *etc.* are components of the ringdown template metric that can be read from equation 4.22.

To use this coincidence test in practice we pick a trigger from one of our detectors. Next we make a list of all triggers in a second detector that have compatible start times considering the maximum signal travel and some measurement error. Then we compute the metric distance, ds^2, between the two triggers. If ds^2 is less than some threshold $\overline{ds}^2$, we say the two form a new doubly coincident trigger or 'double'. If there are three detectors in operation, then we can make a list of all triggers in the third detector that are signal travel compatible with both individual triggers of the double. We compute ds^2 twice, once for the third detector trigger with each original trigger. If both of these distances are below threshold, we add this event to the existent double creating a triply coincident trigger or 'triple'.

We can follow this scheme for all of the triggers produced by the single interferometer analysis. In the end we will have a new list of coincident triggers. Only triggers that are coincident in two or more interferometers will be considered gravitational wave candidates.

We will label our coincident triggers by the detectors where the single triggers occurred. For instance we may have an 'H1L1' double trigger consisting of one single trigger in the H1 detector and one trigger in the L1 detector. We might also have a 'H1L1V1' triple trigger in a similar way. A triple trigger may only ever occur when three interferometers are operating. A double trigger on the other hand could occur

when either two or three detectors operating. We will distinguish these two cases by stating that we have a 'double in double time' or a 'double in triple time'.

Just like for the single interferometer triggers we can produce some new parameters for our new coincident triggers. We can form the SNR weighted mean for the ringdown parameters f and Q. These are the recovered parameters for the coincident trigger:

$$f_{\mathrm{coinc}} = \frac{\sum_{i=1}^{N} f_i \rho_i}{\sum_{i=1}^{N} \rho_i}, \qquad\qquad Q_{\mathrm{coinc}} = \frac{\sum_{i=1}^{N} Q_i \rho_i}{\sum_{i=1}^{N} \rho_i}, \tag{4.24}$$

where N is the multiplicity of the coincident trigger. Another example is the network SNR, ρ_{net}, which is given by the quadrature sum of the individual detector SNRs:

$$\rho_{\mathrm{net}}{}^2 = \sum_{i=1}^{N} \rho_i{}^2. \tag{4.25}$$

Another useful SNR based quantity is the 'chopped-L' statistic, $\mathcal{L}$, defined for triples as:

$$\mathcal{L} = \min \begin{pmatrix} \sqrt{\rho_i{}^2 + \rho_j{}^2 + \rho_k{}^2}, \\ \rho_i + \rho_j + A, \\ \rho_i + \rho_k + A, \\ \rho_j + \rho_k + A \end{pmatrix}^2, \tag{4.26}$$

were the indices i, j, k refer to the three different detectors and A is a tunable parameter. 'Chopped-L' can also be defined for doubles as:

$$\mathcal{L} = \min \begin{pmatrix} \rho_i + \rho_j, \\ B\rho_i + C, \\ B\rho_j + C \end{pmatrix}, \tag{4.27}$$

where B and C are tunable parameters. The 'chopped-L' statistic proves useful, because it down weights triggers where the SNR in one detector is much larger or smaller than in the others. In addition, for triples chopped-L appears as an SNR-squared,

providing extra relative weight to triples over doubles. The tunable parameters for chopped-L were chosen for the S5 and S6 ringdown search to be: $A = 0.75$, $B = 2.0$, and $C = 2.2$.

4.3. Background Estimation

In order to compute false alarm probabilities for gravitational wave candidate events, we need to know how often random detector noise will cause coincident triggers. If there are no gravitational waves in our detector strain, then we could just run a coincident analysis and count how often coincidences happen by accident. We cannot easily simulate detector noise, so how can we generate a data stream that accurately reproduces the non-Gaussian noise of a LIGO detector, while at the same is guaranteed to contain no gravitational waves?

The individual detector triggers are not enough to establish detection alone. Provided gravitational waves are rare, we can treat all of the individual detector triggers as being the result of non-Gaussian noise. This is not as crazy at it sounds. What is the difference between $\rho = 8$ trigger resulting from a gravitational wave and dozens of $\rho = 8$ triggers resulting from noise? The only distinguishing fact is that gravitational waves occur in multiple detectors near simultaneously. We note that if gravitational waves were the primary source of triggers, we could not do this.

Starting from our individual detector triggers, we want to estimate how many triggers will accidentally agree with triggers in another detector. We cannot perform

a simple coincidence analysis, because the gravitational waves would pass the test and corrupt our background estimation. In order to ensure that we get no gravitational wave coincidences, we may time shift the triggers from one detector relative to the others. After this 'time-slide' takes place we may form coincidences in the usual way. If this time-slide is large enough, it will be impossible for any coincident triggers to have resulted from gravitational waves. We have effectively generated a whole new observation period with only detector noise. We can repeat this process, sliding by a different amount each time, to generate as many gravitational wave free observations as we want. To estimate the false alarm rate of coincidence triggers, we need only to count the coincidences resulting from a time-slide analysis.

The time-slides actually occur on rings with the end of each detector observation periodically connected back to the beginning. Figure 4.3 attempts to demonstrate this process.

There is one problem to using time-slides to estimate the background. If two detectors have correlated noise, then the time-sliding process will break the correlations and underestimate the true background. Unfortunately, this is the case with the colocated Hanford detectors H1 and H2. We will choose to not time-slide these two detectors relative to each other, effectively holding them fixed while time-sliding only the L1 detector for the S5 analysis. We have chosen to shift L1 by 5 s for each time-slide, because the longest ringdown we can look for has a duration less than 1 s. In this manner L1 will be shifted by 10 s total for the second time-slide.

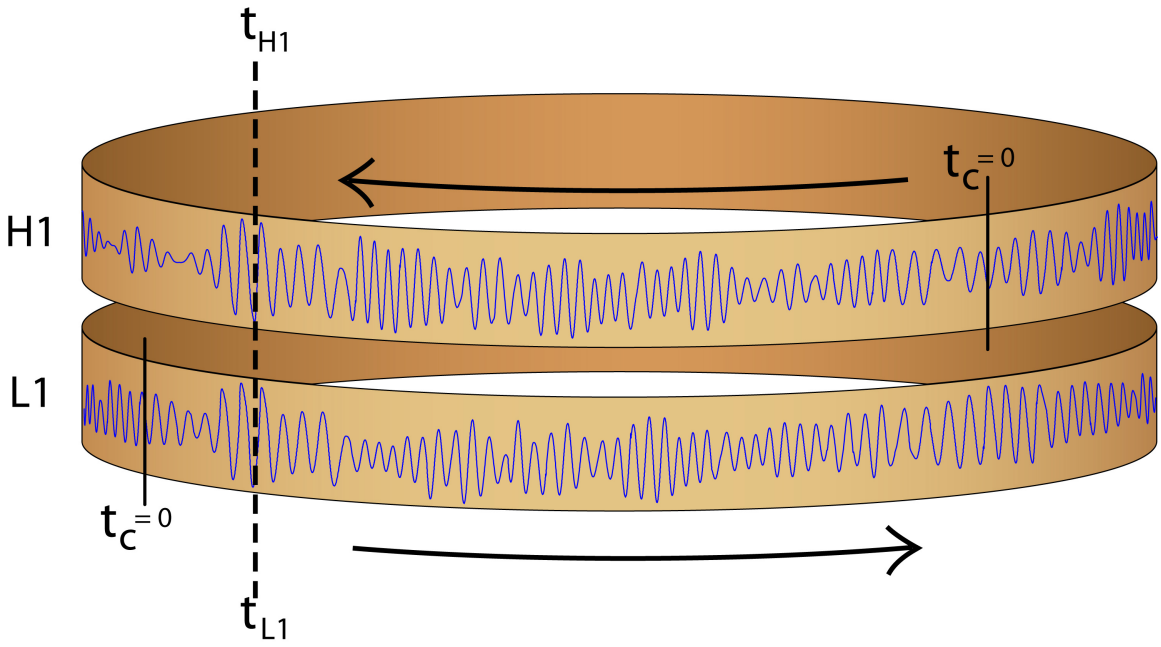

Figure 4.3. Cartoon demonstrating time-slides. Data from two or more detectors are time shifted relative to each other. In this way any coincident triggers cannot be the result of gravitational waves (courtesy of R. Tucker).

By not time-sliding H1 and H2 we have no way of knowing the false alarm rate for H1H2 doubles. We will not be able to make any detection claims on such triggers, so we will simply discard them. In addition we can save time by not analyzing any data where only these two detectors are operating. We make this choice and discard data in order to have the most accurate background estimation for triple triggers, which provide the best hope for detection.

When we have three separated interferometers, as is the case in LIGO's sixth and Virgo's second and third science runs. We must slide all of the detectors relative to each other. In this case we choose arbitrarily to hold H1 fixed. We slide L1 5 s relative to H1 and V1 10 s relative to H1 per time slide.

4.4. Detecting Simulated Signals

Before we can make any statements about the detection or non-detection of gravitational waves, we should verify our ability to detect the sorts of gravitational waves we are looking for. We do this by adding simulated gravitational wave signals directly to the strain data and performing an analysis on this modified data. When this analysis is combined with the time-slide analysis of section 4.3, we can determine the detection confidence for gravitational waves from known sources. We refer to this as an 'injection' analysis, because simulated signals are 'injected' into the data.

We will inject a large number of signals with a variety of source parameters to map out our ability to detect gravitational waves more precisely. Previous searches only considered 'naked ringdown' signals. The naked ringdowns are exactly the templates we use in recovery, with no precursor signal attached. We expect astrophysical ringdown signals resulting from black hole coallesence to have inspiral and merger gravitational wave signals attached to the beginning of the ringdown. Even in the case of stellar core collapse there is likely some 'ring-up' to observe. These precursors can act as a distraction to our templates and cause problems with parameter recovery. While naked ringdown injections can directly test our template bank for coverage, they do not provide a good sense of our ability to detect astrophysical sources.

We need gravitational waveforms for more astrophysically motivated sources. Recently a host of realistic gravitational waveforms covering the full inspiral, merger, and ringdown (IMR) of compact binary coalesence have become available. One of

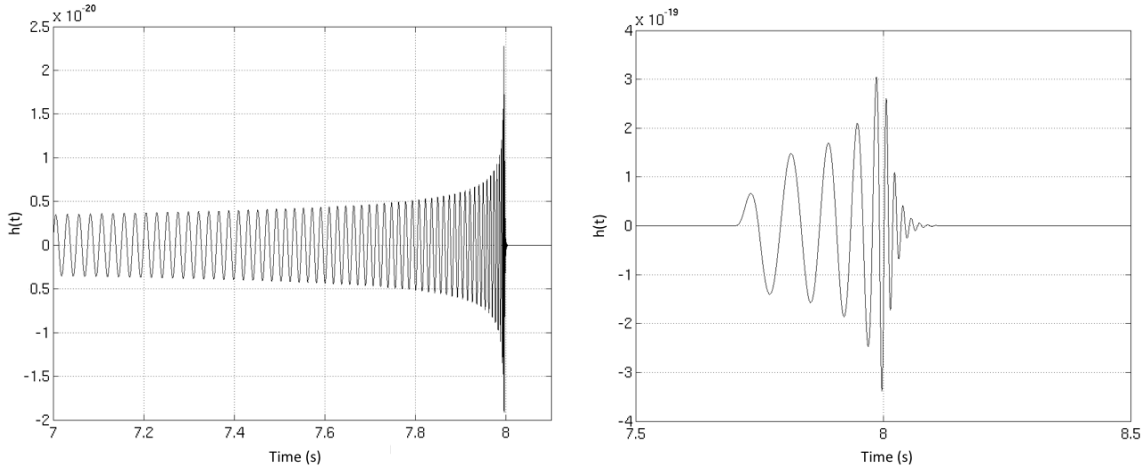

Figure 4.4. Two example EOBNRv2 waveforms showing equal mass binary black hole coalescences. On the left each component has a mass of 10 $M_{\odot}$, werea the right shows component masses of 150 $M_{\odot}$.

these was developed using the effective one-body (EOB) approach and calibrated against numerical relativity (NR) results [14]. More recent developments have lead to higher order terms being introduced to these 'EOBNR' waveforms resulting in the EOBNRv2 waveforms. The EOBNR family of waveforms have been used for injections in other LIGO searches and will likely be used as templates in future searches. We will use the second generation waveforms extensively to make injections in our analysis. We show two example EOBNRv2 waveforms in figure 4.4.

Another full IMR family of waveforms was constructed by a Fourier domain piecewise fit to NR simulations. These parameterized, phenomenological waveforms are known as IMRPhenom [3]. We will choose to use the second generation IMRPhenomB in our analysis. The IMRPhenomB waveforms have the advantage of including spins for the two initial black holes.

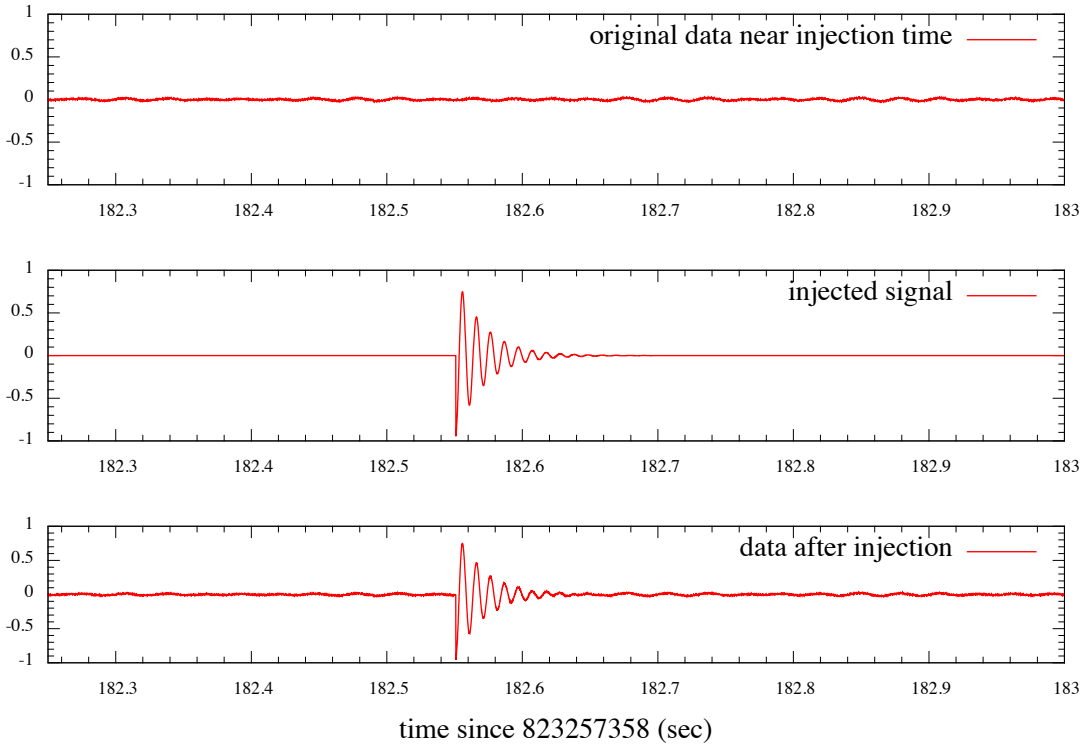

Figure 4.5. Process of injecting a simulated signal into data. In this case a very loud 'naked ringdown' signal. y-axis units are unscaled strain.

The actual process of making injections into data is very simple. We first pick a waveform. Then we generate a list of parameters, randomly, for a given number of injections. When running the single detector analysis we simply add a time domain waveform to the stain data as it is read in. This process is demonstrated for a naked ringdown injection in figure 4.5.

The hard part of an injection analysis is selecting the ranges and distributions for the various source parameters. As naked ringdown injections are used primarily for pipeline verification, we will choose the same parameter ranges as our template bank. For simplicity we will choose a radiation efficiency of 1% for all ringdown injections.

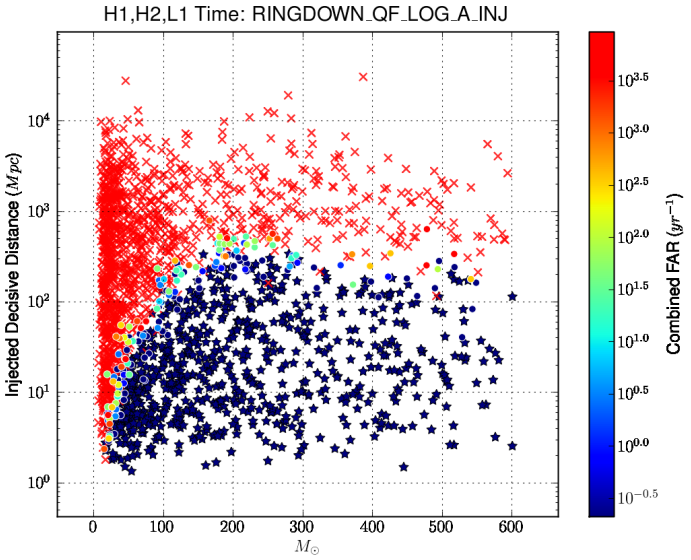

Figure 4.6. A missed / found plot for a 'naked ringdown' injection run in 2 months of S5 data. Injections detected with zero false alarm rate are blue stars, missed injections are red crosses. False alarm rate (FAR) will be defined in chapter 5.

We will cover a wide range of distances from about 1 Mpc to 10 Gpc. The result of one of many ringdown injection runs is shown in figure 4.6.

For IMR injections we will focus on a more astrophysically motivated distribution of parameters. We must also keep in mind the range of validity of the IMR waveforms. The waveforms are only trustworthy to system total masses of about 450 $M_\odot$ and mass ratios of about $q = 4$. We chose to hold to these limits for a large scale injection campaign. For some of our smaller pipeline verification runs we choose to push the mass ratio limit, making some injections with mass ratios up to $q = 10$. We show the results of one EOBNRv2 and one IMRPhenomB injection run in figure 4.7.

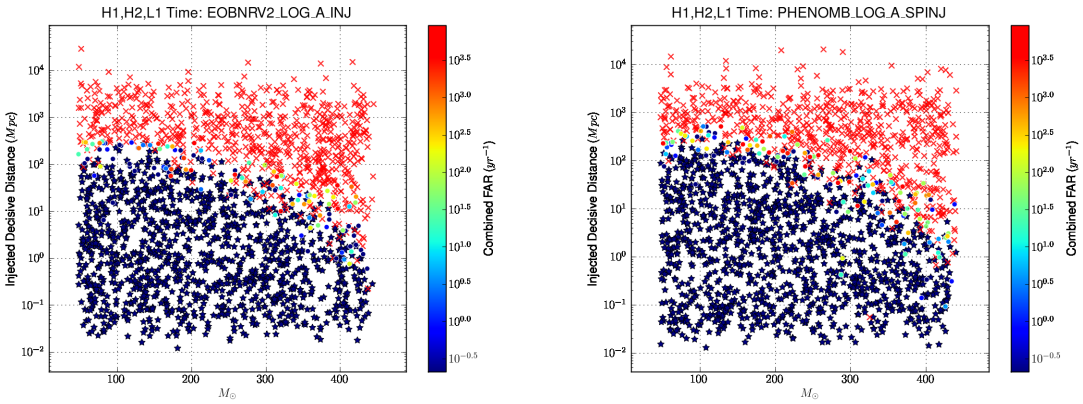

Figure 4.7. Missed / found plots for two IMR injection runs in 2 months of S5 data. EOBNRv2 injections are on the left and IMRPhenomB are on the left. Injections detected with zero false alarm rate are blue stars, missed injections are red crosses. False alarm rate (FAR) will be defined in chapter 5.

In the case where our analysis does not detect any gravitational waves we will use our signal injections to place rate upper limits on BH-BH mergers. Provided our injections accurately represent astrophysical sources, we can estimate the volume of space our analysis probed by our ability to find signal injections. Our limits would be based on excluding any BH-BH mergers from happening in that volume over the course of the observation time.

4.5. The Data Analysis Pipeline

LIGO's fifth science run (S5) collected data for approximately two years. For the last six months of S5, the Virgo detector also collected data. However, due to complications with the analysis pipeline, we chose to only use data from the three LIGO detectors. In sum there was approximately 1 year total of 'triple time' data

collected, when all three interferometers were operating. In addition there was another half year of 'double time' with two interferometers operating.

LIGO's sixth science run (S6) operated the enhanced detector network consisting of only two LIGO interferometers for about 1 year. Virgo's second and third science runs (VSR2 and VSR3) were conducted during this time providing about eight months total with three interferometers potentially operating. In the end there were about 37 days of triple time and about 138 days of double time, giving about 6 months of total analysis time. For ease we will refer to S6/VSR2,3 science runs as simple 'S6'. S6 was divided into 4 main data collection periods denoted A through D. The detectors were taken offline for maintenance between each data collection.

The data analysis pipeline was largely the same for the S5 and S6 searches. The methods we describe herein are applicable to both analyses unless state explicitly. We must note that the results from S5/S6 search for gravitational waves from BH ringdowns are currently under LIGO Scientific Collaboration review. All of the results show are preliminary.

The data analysis software for the LIGO search for BH ringdowns is built using the LIGO Algorithms Library Suite (LALSuite), a collection a C libraries maintained by the LSC Data Analysis Software Working Group. These contain low level functions for data analysis all well as search specific tools [38]. High level user interfaces are handled in python and large scale data and meta-data structures produced by the analysis are held and manipulated as SQL databases. The data analysis software

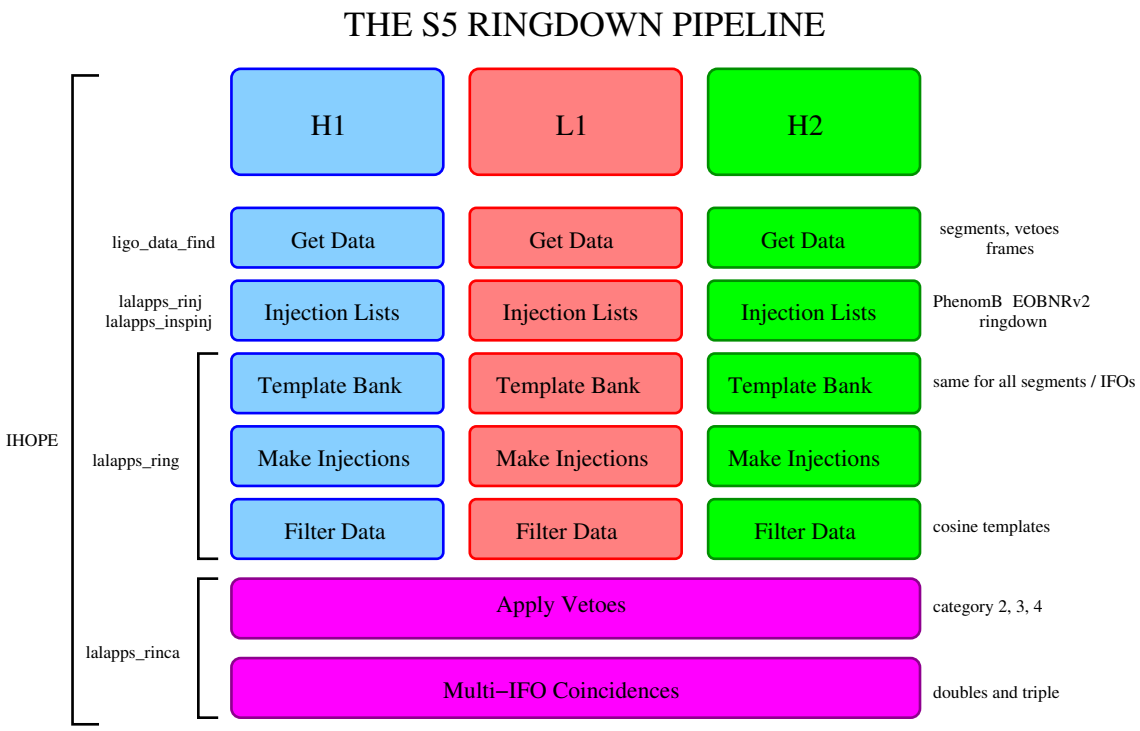

Figure 4.8. Outline of S5 data analysis pipeline. The S6 analysis used the same structure, replacing the H2 interferometer with V1. Some of the steps shown are specific to an analysis with software injections. Individual software applications that handle various tasks are shown on the left..

used in the S5 and S6 analysis was built on top of existing software from the S2 and S4 ringdown searches developed primarily by J. Creighton, D. Brown, and L. Goggin. The software for the S5/S6 search was developed by S. Caudill, C. Capano, K. Hodge, D. Talukder and the author.

We described the primary steps of a search for BH ringdowns. Now we must put the pieces together in a practical way to get from 2 years of detector strain data to candidate gravitational wave events. In the following we will discuss the ringdown data analysis pipeline, omitting many details of implementation. The data analysis

pipeline for an analysis with software injections is diagramed in figure 4.8. The same basic steps apply to all analyses.

First the data is divided into reasonably lengthed analysis chunks. For S5 the two years of collection time was divided into 12 approximately 2 month long analysis periods. During S6 analysis divisions were naturally made between the four data collection periods. Further divisions were made when necessary. For example S6B was divided into two chunks, split when the new Hydraulic External Pre-Isolation (HEPI) system came online at the Livingstone observatory. S6D had two divisions: one marking the beginning of VSR3, while the other was choosen to split VSR3 roughly in half.

After selecting an analysis chunk to consider, there is still a good deal of data conditioning that must be done. LIGO collects data at 16384 Hz, as our analysis is only interested in signals up to 2000 Hz we will down sample the data by a factor of two. This gives a Nyquist frequency of 4096 Hz. The data is then low pass filter at the Nyquist frequency to prevent aliasing, and high pass filtered at 35 Hz, removing data where the detectors have little sensitivity. The data is next cut up into 2176 s blocks. These blocks are the fundamental data unit passed into the single interferometer analysis.

The single detector analysis requires the noise power spectral density $S_n(f)$ to perform inner products. We estimate this directly from the data. Starting with one 2176 s block, we further divide it into sixteen 256 s overlapping segments. To do

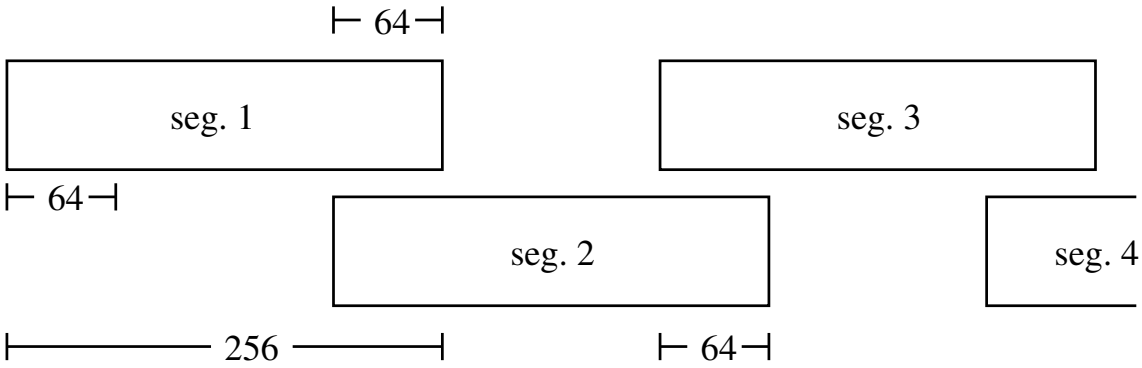

Figure 4.9. A cartoon demonstrating the data segmentation process. Each 256 s segment has a 64 s overlap with the segments before and after it..

this the second segment begins 64 s before the end of the first. In this way only the central 128 s of each segment is unique (except for the first and last). This redundancy prevents edge effects from corrupting the search. This method is shown in figure 4.9. Each of the sixteen segments is given a labeling number.

We can now Fourier transform each data segment and estimate its power spectral density by discretizing equation 3.12. We determine the median spectrum twice: once using only the even segments and again using only the odd. We finally calculate the mean of the even and odd median spectra, taking this to be the power spectral density for the whole 2176 second data block. This average power spectral density is used to over-whiten each individual segment, $\tilde{s}_i(f)$:

$$\tilde{s}_i^{\mathrm{w}}(f) = \frac{\tilde{s}_i(f)}{S_n(f)} \tag{4.28}$$

Using the over-whitened data segments, $\tilde{s}_i^{\mathrm{w}}(f)$, saves computational cost in our inner products.

Finally, we can generate a gravitational wave template h_j and compute its inner product with each of the data segments. We select parameters f and Q from the template bank described in section 4 and generate the corresponding gravitational waveform using equation 4.1. We Fourier transform this template and normalize it:

$$\tilde{h}_j^{\mathrm{n}} = \frac{\tilde{h}_j}{\sqrt{\langle h_j \,|\, h_j \rangle}}. \tag{4.29}$$

Now we can compute the necessary inner product:

$$\rho(t_0) = \frac{\langle h_j \,|\, s_i \rangle}{\sqrt{h_j h_j}} = 2 \int_0^\infty \tilde{h}_j^{\mathrm{n}\star}(f, t_0)\, \tilde{s}_i^{\mathrm{w}}(f)\, \mathrm{d}f, \tag{4.30}$$

and look for places were $\rho(t_0)$ exceeds a threshold, $\bar{\rho}$. For the S5 and S6 search we set $\bar{\rho}$ on a per detector basis. Following other LIGO searches we choose to set $\bar{\rho} = 5.5$ for the two 4 km detectors, H1 and L1. Breaking with previous searches we decided to reconsider the 2 km Hanford detector threshold. Because H2 is half as sensitive as its 4 km companion, all triggers we find in H1 should appear in H2 with half the SNR. We decided to set the threshold for H2 lower than the two 4 km interferometers. Setting $\bar{\rho}_{\mathrm{H2}}$ too low, will cause memory overflows as the number of triggers that must be recorded becomes very large. With this in mind we set $\bar{\rho}_{\mathrm{H2}} = 4.0$.

Our triggers are determined by finding all of the instances where $\rho(t_0) > \bar{\rho}$. To cut memory requirements we only end up keeping one trigger per template per second of data. This one second limit was chosen to be about as long as our longest template. As we have ~ 600 templates, we can still have many triggers in a short period of time. This can be seen in figure 4.10, where we show the single interferometer output for a

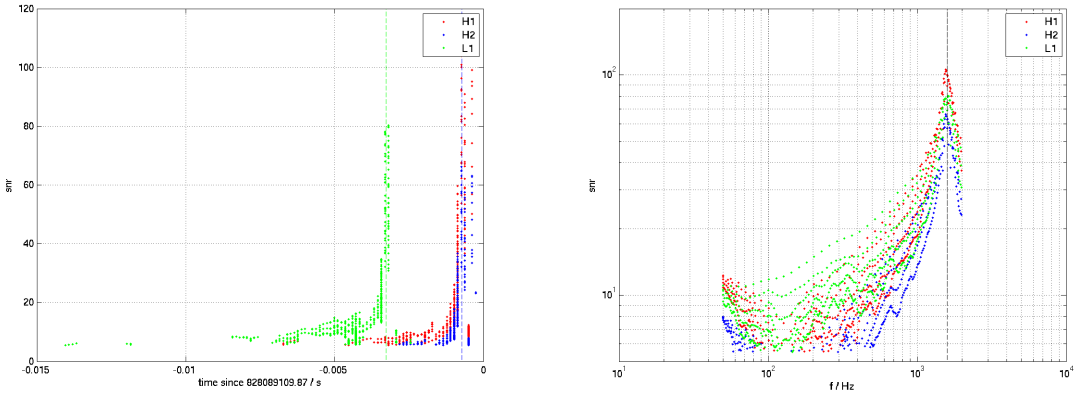

Figure 4.10. Signal-to-noise ratio time series and frequency series output by the single interferometer analysis. SNR v. time is on the left. The time of arrival for a simulated signal is shown as dashed lines, one for each detector site. SNR v. frequency is on the right, with the frequency of the simulated signal shown as a dashed line. Note there is only one trigger per template. This is most easily seen in the frequency series.

few milliseconds of data near a software injection. We also see how the SNR varies across the template bank as a function of frequency.

The process of computing $\rho(t_0)$ and identifying triggers is be repeated for our template h_j for each of the 16 data segments. Afterwards we choose a new template h_{j+1} and repeat the process. In this way we build a list of single interferometer triggers for each 2176 s data block in a large scale analysis chunk (and then for all of the chunks). Do to the independence of each block analysis, we can massively parallelize the single interferometer analysis. We follow the exact same procedure to construct single interferometer triggers for data that has been injected with simulated gravitational wave signals.

These single interferometer triggers are fed into a coincidence analysis described in section 4.2. It is here that we apply data quality vetoes. Many of our single

interferometer triggers will be the result of noise in the detector and not gravitational wave signals. To combat this, we can identify periods of time when the detector is known to be behaving badly and throw away triggers that fall in those time periods. There are many ways of identifying data quality problems. LIGO collects data from thousands of monitors in the environment and control systems. One of the most effective ways to generate data quality vetoes is to correlate periods where the detector noise is strongly non-Gaussian to particular behaviors in some of these auxiliary channels. Then in the future when a particular auxiliary monitor exhibits similar behavior, we can be confident that the detector will not be producing high quality strain data [18][34].

The LSC Detector Characterization group compiles lists of times when the detector is misbehaving and provides recommendations to the gravitational wave data analysis groups as to which vetoes to apply. The veto lists are divided into several 'categories', depending upon how confident we are that the detector is not functioning properly. We apply vetoes at the coincident stage, because the data is severely fragmented by vetoes. This fragmentation makes the Fourier transforms required by the single interferometer analysis problematic.

Vetoes are our first line of defense against non-Gaussian detector noise. Unfortunately, by throwing away data we reduce our observation time, thereby reducing our ability to find gravitational waves. For this reason we will use the less strict 'category 2' vetoes for the purpose of identifying candidate detections, potentially allowing loud

false alarms through the pipeline in the process. If no detections are made, we will use the more tight lipped 'category 3' vetoes to produce rate upper limits[3].

We need to run the multi-detector coincidence stage of our pipeline multiple times: once for each veto level. The coincidence stage is also were we construct time-slide triggers for background estimation, as discussed in section 4.3. We conduct 100 different time-slide analyses to accurately estimate our background. All of these circumstances boil down to selecting some subset of triggers from the various interferometers and computing the template metric distance ds^2 by equation 4.22. For the S5/S6 search we chose to set a metric distance threshold of $\overline{ds}^2 = 0.4$. This is a very wide separation to be true, but the template metric distance does such a good job of eliminating false alarms that we were not producing enough triggers in the time-slide analysis to accurately estimate the background. This is demonstrated in the comparison with the S4 analysis in section 4.6. At the beginning of the coincidence analysis we maximize over all templates in each interferometer, keeping only one trigger in any half second window. We choose the keep the trigger with the highest SNR. In figure 4.10 we would keep three triggers total, one for each detector. These three triggers would easily pass the coincidence test as they are all very close to the frequency and time of arrival for the same injected signal.

With a list of coincident triggers we can do a bit more trigger rejection before arriving at our candidate events. First as discussed in 4.3, we cannot accurately

[3]There was a veto category restructuring between S5 and S6. For the S6 analysis the equivalent detection versus upper limits veto scheme is achieved using 'category 3' and 'category 4' vetoes, respectively. Yes, it is confusing.

estimate the H1H2 coincident background with time-slides, because the colocated detectors have correlated noise. We reject any H1H2 doubles when all detectors were operating. Without an accurate background we cannot determine whether such events are gravitational waves or not. We have already ignored any time when only the two Hanford detectors were operating from our analysis.

Next we will apply an H1H2 consistency test. Coincident H1H2 triggers can only exist as part of a triply coincident event, so this test will apply to triples only. We expect the SNR of a trigger in H2 to be half of that in H1, and we will use this to eliminate triggers from obvious detector glitches. If all three interferometers are operating, then any trigger in H2 should have a companion trigger in H1. If there is no companion trigger, then the H2 trigger must be the result of a glitch. We reject H2L1 doubles found in triple time.

Similarly if we have a trigger in H1, then there should be a companion trigger in H2. It is possible that the H2 trigger could be absent, because it is below the SNR threshold, $\bar{\rho} = 4.0$. We will reject any H1L1 double in triple time, if the H1 trigger has SNR greater than 16. In this case we would expect an H2 trigger of $\rho_{\mathrm{H2}} \approx 8$ well above threshold. Finally we reject any triply coincident trigger where $\rho_{\mathrm{H2}} > \rho_{\mathrm{H1}}$. In the S6 analysis H2 was not operating, so the H1H2 consistency test cannot be applied.

After all of this we have a list of candidate gravitational wave events, expressed as multi-detector coincident triggers. We also have a list of coincident triggers that

were produced by the time-slide analysis and a list of triggers from the injection analysis many of which correspond to found signal injections. These triggers are next fed to a series of post processing steps. The post processing analysis identifies which signal injections were found by the pipeline. Next it ranks all of the triggers: found injections, time-slide triggers, and gravitational wave candidates, with a detection. Finally, it estimates the background using the time-slide triggers, and computes a false alarm probability for all of the triggers.

4.6. Comparison to the S4 pipeline

Let us stop here for a moment to consider how the pipeline described in the previous section differs from that used in the S4 search for BH ringdowns. Most of the main changes to the analysis pipeline are under the hood, so to speak. The least visible change is the great deal of automation that was implemented, much to the relief of the analysts. This automation is incredibly important for the large scale data analysis of the S5/S6 ringdown search. The trial and error processes of tuning the various thresholds was made much easier.

Another largely invisible improvement is the ability to analyze Virgo data. This involves minor changes to data conditioning and handling. We also updated some of the ringdown QNM specific fitting functions in the code to represent the most recent theoretical and numerical results. A slightly more visible change is our choice to make full IMR injections. As discussed in section 4.4, this provides an astrophysically

motivated source population to test the analysis pipeline and compute rate upper limits.

The most significant change from the S4 pipeline is an improvement to the coincidence test. In S4 the coincidence test was performed using a two dimensional template metric on f and Q only. Time coincidence was established by accounting for signal travel time and a constant dt measurement error. This choice ignores the correlation between signal frequency and timing accuracy. Consider a pure sinusoid signal. We can understand in this simple case how poor detector timing effects the signal-template match. If a sinusoid template is aligned with this signal properly, we will find perfect agreement. When the template is misaligned by one half wavelength, there will be zero agreement. Any amount of misalignment between these two cases will provide a continuous, monotonically decreasing match. We can then characterize the template match in terms of the misalignment as a fraction of signal wavelength. This tells us we should demand greater timing accuracy for higher frequency signals. This is a feature of the S5/S6 template metric where $g_{tt} \sim f^2$.

With the S5/S6 coincidence test we are able to reject a great number of background events that passed the S4 test. In figure 4.11 we see how for the colocated H1 and H2 detectors we are able to place very strong timing constraints on high frequency coincidences. Doing so reduces the number false positive H1H2 coincidences by a factor of 6. We do not analyze H1H2 coincidences alone, but S5 triple coincident

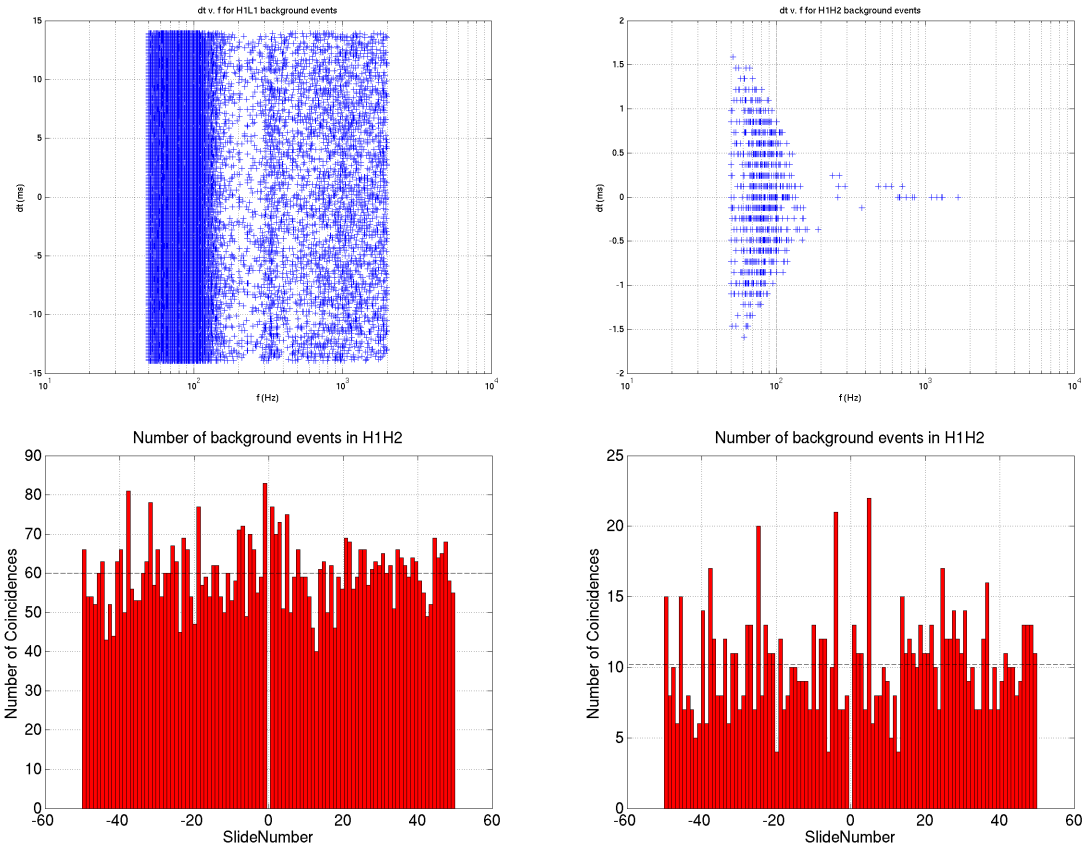

Figure 4.11. S4 to S5 background comparison for H1H2 doubles: left column is S4, right is S5. The top row compares the distribution of dt as a function of df for background events. The bottom row compares the number of coincident triggers per timeslide. Because H1H2 doubles are not used in the search this improvement directly maps to a reduction in triply coincident background events..

triggers contain an H1H2 double. This feature of the new test translates directly to the triple coincident background.

For other combinations of interferometers we still have signal travel uncertainty. The effect is not as pronounced as the colocated case, but we still find a 20% reduction in background events for H1L1 doubles shown in figure 4.12. These background

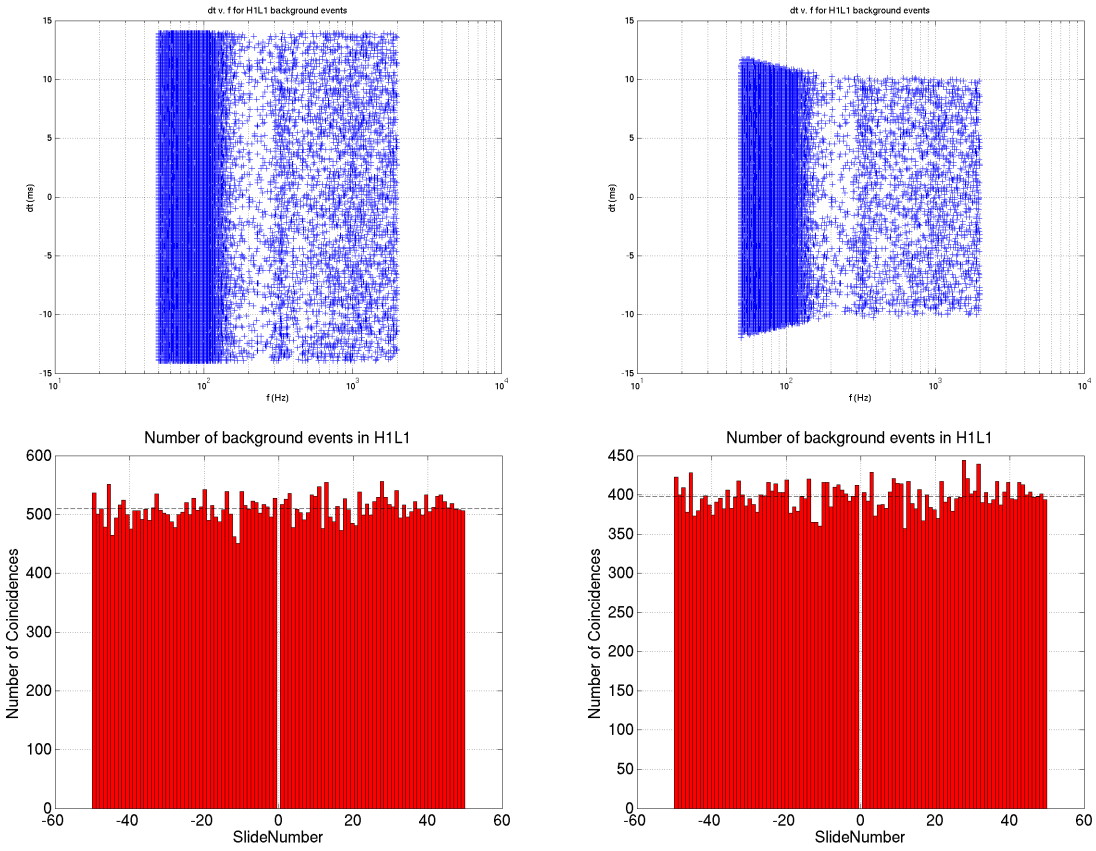

Figure 4.12. S4 to S5 background comparison for H1L1 doubles: left column is S4, right is S5. The top row compares the distribution of dt as a function of df for background events. The bottom row compares the number of coincident triggers per timeslide. The improvement is not as great as the H1H2 case due to the signal travel time between detectors..

reductions contribute directly to increasing the confidence in any coincident events that pass the multi-detector coincidence test.

Finally, an entirely new post processing pipeline was used. This new pipeline incudes new 'injection finding' methods, which improves our ability to accurately match events from injection runs with their parent simulated signals. This results in more realistic missed / found lists and, consequently, more accurate rate upper

limits. The real centerpiece of the new post processing pipeline is a multi-variate statistical classifier, which we use to sort gravitational wave signals from noise caused false alarms. It is here we turn our attention.

5. USING A MULTI-VARIATE STATISTICAL CLASSIFIER IN A GRAVITATIONAL WAVE SEARCH

Multi-Variate Statistical Classifiers (MVSC[1]) have been used for a number of applications in and outside of physics. However, they have not yet been used extensively in gravitational wave astronomy. There are several MVSC algorithms, but they all attempt to accomplish the same goal: distinguish between classes of objects. In simple cases, where the objects are parameterized in only a few dimensions, a human can look at one-dimensional histograms or two-dimensional scatterplots and devise cuts that distinguish between the classes. However when the objects are parameterized in a high dimensional space the problem is not so simple. For instance in a ten-dimensional space the distinguishing cuts would be nine-dimensional surfaces. Dealing in these sorts of abstract spaces directly is a problem for any human. One solution is to project the high dimensional space to a lower, more workable dimension. Unfortunately, using only one or two-dimensional projections of the parameters can hide higher order correlations, again limiting the capabilities of the human analyst.

MVSC algorithms are able to explore these high dimensional spaces directly, finding surfaces that separate objects into distinct classes. In gravitational wave astronomy the classes of interest are 'signal' (real gravitational waves) and 'background' (non-gaussian detector artifacts or glitches). The analysis pipeline described in chapter 4 outputs a list of coincident 'triggers' characterized by many parameters (such as signal-to-noise ratio (SNR) and frequency). It is then MVSC's job to determine which

[1]typically pronounced "music"

of these triggers are true gravitational wave signals and which are merely background events.

The LIGO S5/S6 search for gravitational waves from black hole ringdowns uses the StatPatternRecognition (SPR) software package developed by Narsky for high energy physics data analysis [44][7]. This analysis uses the Random Forest of Bootstrap AGGregatED (bagged) Decision Trees algorithm. The bagged decision tree algorithm has proven effective in a wide class of problems [43]. It is also considered more stable in the presence of outliers and in very high dimensional parameter spaces than other MVSC algorithms such as neural networks and boosted decision trees [8].

5.1. Random Forest of Bagged Decision Trees

A random forest of bagged decision trees uses a collection of many decision trees to assign a probability that a given event belongs to one of two sets. To understand this process we must first understand how a single decision tree can classify events.

A decision tree is generated with a collection of training data: events that we know *a priori* belong to either the set of signals or background. It is important to remember that decision trees cannot be used to classify the same events on which they were trained. The performance of this algorithm is highly dependent on the quality of the training set. That is to say: how representative of the data we wish to classify with the decision tree the training data are.

To construct a decision tree we make a series of one-dimensional cuts on a random subset of the available parameters. These cuts are chosen to best separate our two sets: signal and background. Deciding which cut best separates signal from background is a non-trivial matter. We choose to measure the goodness of a cut with a binary information entropy function:

$$H(p) = -p \log p - (1-p) \log (1-p) \,, \tag{5.1}$$

where p is the fraction of events in a node that are signals, and $1-p = q$ is the fraction of events in the node that are background (as is customary in information theory, the logarithms are base 2). We notice that if a node contains as many signal events as background, then $H(\frac{1}{2}) = 1$, and a perfectly sorted node has $H(1) = H(0) = 0$. By minimizing the binary information entropy of our nodes, we can find the optimal sort.

To illustrate this process suppose we have an initial node N, containing s_N signal events and b_N background events. We randomly choose a parameter and a theshold value. All events with this parameter greater than the theshold go to a new node, A; the rest go to a second new node B. Now we have two nodes containing s_A, s_B and b_A, b_B signal and background events respectively. We measure the 'goodness' of this cut by the sum of the information entropy of the resulting nodes:

$$H_{\text{cut}} = H(p_A) + H(p_B) \tag{5.2}$$

$$p_i = \frac{s_i}{s_i + b_i} \tag{5.3}$$

here the ps are the fraction of signals in each new node. We can compute H for many cuts on a given parameter and in this way find the 'optimal' cut on that parameter, best separating signal from background. In practice at each node of the tree, we choose a random subset of all parameters, find the optimal cut on each, and then pick the best cut from any of the chosen parameters.

When a node meets one of two conditions, it becomes a 'leaf' and it is no longer divided: either the node is perfectly sorted, containing only signal or background, or there are fewer events in the node than a preset limit, the 'leaf size'. When there are no more nodes to divide (*i.e.* all have become leaves) we have created a decision tree.

To use the decision tree to rank an event, we begin by placing the event in the initial node or 'trunk'. We then follow all of the cuts, passing the event along the various tree 'branches' until it arrives at a leaf. The probability that this event is a signal is given by p_ℓ, the fraction of signal events in the leaf. The MVSC rank, $\mathcal{M}$, is given by the ratio of probabilities that the event is a signal or background:

$$\mathcal{M} = \frac{p_\ell}{1 - p_\ell} \tag{5.4}$$

This rank is a Bayes-factor (see section 6.1), which tells us how much one model is favored over another. In this case the two models are *this event is a signal* and *this event is a background*.

Because this tree is generated by a random process, each time we make a new tree, even when using the same training data, we can get a different result. To combat this effect and arrive at a final solution, we generate a random forest of many trees.

Each tree is generated in the same way. The individual trees will be different, because a different subset of parameters is chosen at each node to optimize over. To rank an event from a forest, we find its final leaf in all trees. The probability that this event is a signal is given by the fraction of signal events in all leaves:

$$p_{\text{forest}} = \frac{\sum s_i}{\sum s_i + b_i} = \frac{1}{n} \sum_{\text{leaves}} s_i, \tag{5.5}$$

where s_i and b_i are the number of signal and background events in the ith leaf and n is the total number of events (signal and background) in all leaves. The MVSC rank is defined in the same way:

$$\mathcal{M}_{\text{forest}} = \frac{p_{\text{forest}}}{1 - p_{\text{forest}}}. \tag{5.6}$$

For a random forest of bagged decision trees as described above there are three forest parameters that can be chosen. Choosing these parameters effectively will alter the sorting performance of the forest. First is 'leaf size'. The leaf size is the minimum number of events allowed in a node before it becomes a leaf. It determines when the training of a tree is stopped. A good choice of leaf size is affected by how similar the training data are to the actual data we hope to classify and how many events are in the training data. By choosing too small a leaf size the forest becomes 'over-trained': the forest can separate the training data well, but the sort is specific to those data. An over-trained forest will assign all events a very high or very low MVSC rank. It will only correctly classify events that are nearly identical to events in the training data. Over-trained forests have a large systematic bias.

Alternatively, if we choose a leaf size that is too large, our forest will be 'under-trained'. An under-trained forest did not have enough chances to make good cuts on the training data. It will assign almost all events an intermediate MVSC rank, leaving us unsure if any event is signal or background.

The next forest parameter is the number of 'sampled parameters'. Recall that at each node we choose a random subset of the parameters and find the optimal cut among this subset. The number of sampled parameters is how many parameters are in the considered subset at each node. If this number is too large, each node will have essentially the same parameters available to make cuts. The same few parameters will be used over and over again, and the forest will not be fully exploring the space of possible cuts. Furthermore, because individual trees will be making cuts based on the same parameters, all of the trees in the forest will be very similar. This is another example of over-training.

If we were to select a number that is too low, then each node would have very few options to make cuts. If some parameters were bad at separating signal from background, we would be forced to use them, making inefficient cuts. The tree would run up against the leaf size limit before the training events were well sorted. This is an example of under-training. The classification in this case would be highly dependent on the presence (or lack thereof) of 'junk parameters'.

Finally, we may adjust the number of trees in our forest. Using more trees provides a more stable MVSC rank by increasing the total number of training events used in a

given rank calculation. Eventually, we can max out the utility of adding more trees. There is only so much information in the training data, and adding a large number of additional trees will reproduce results found in earlier trees. Adding more trees will also increase the computational cost of training linearly.

5.2. A Toy Problem

To build intuition as to how MVSC performs, we will first consider a toy problem in two dimensions. Consider a set of events characterized by two parameters t_1 and t_2. Within this set there are actually two types of events: 'signals' where t_1 and t_2 are correlated and 'background' where they are not. To generate the uncorrelated 'background' we will draw t_1 and t_2 independently from a uniform distribution $[0, 100]$. The 'signals' will have t_1 drawn from the same uniform distribution, but t_2 will be given by a small Gaussian deviation (variance 1) from t_1. For these 'signals' t_1 and t_2 are highly correlated and densely distributed along the $t_1 = t_2$ line.

In this manner we will generate 10 000 of each background and signal events. In order to classify these events we must train forests of decision trees with some data for which we already know their classifications. We will select 2000 of each background and signal as an 'evaluation' set that we hope to classify and use the other 16 000 events as a 'training' set to construct the decision trees. By repeating this process for five evaluation sets we may rank all 20 000 simulated events.

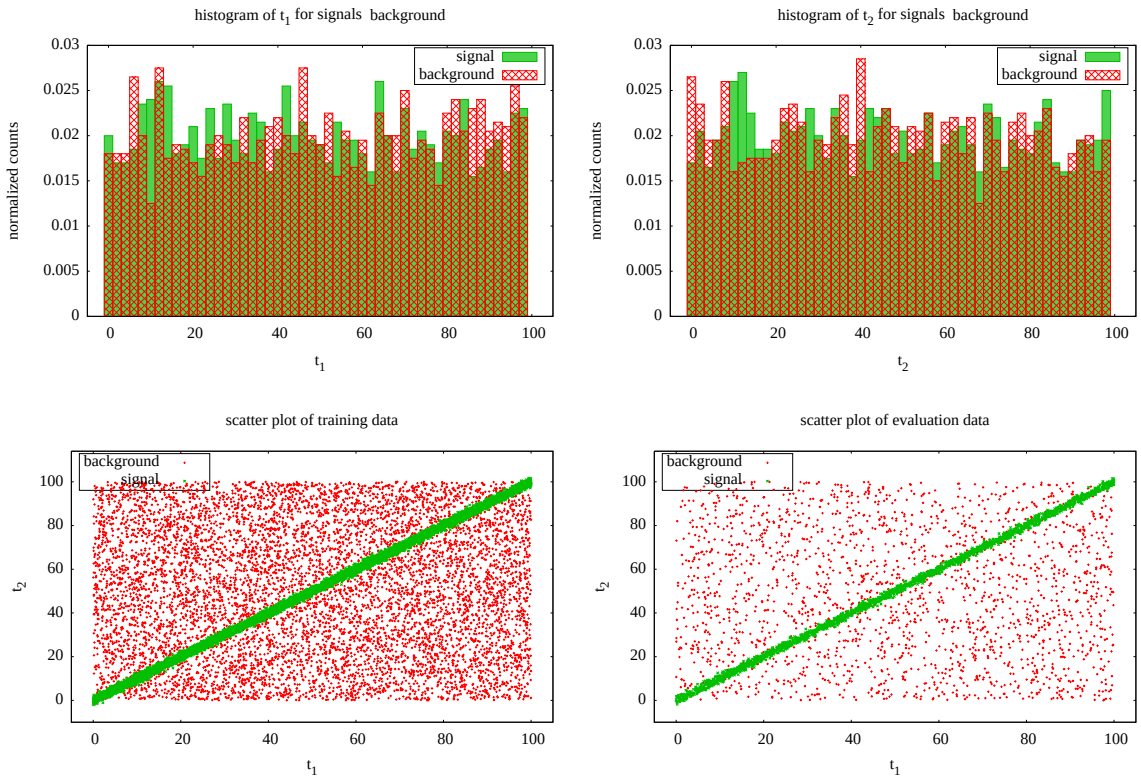

Figure 5.1. Simulated 2D data for a MVSC toy problem. The top panels show 1D histograms of the two parameters, t_1 and t_2, for both 'signal' and 'background'. The two sets appear indistinguishable in these projections. The bottom panels are 2D scatter plots of a 'training' and 'evaluation' data sets, where the correlation in the signals becomes obvious.

A naive analysis might attempt to use one of the given parameters, t_1 or t_2, to make a cut that separates signal from background. As can be seen in figure 5.1, this would be a poor choice. A better choice would be to make cuts in the full 2D parameter space. Knowing that for most ($\sim 95\%$) signals t_2 is within 2σ of t_1, we may make cuts along the $t_2 = t_1 \pm 2$ lines, calling everything that lies between them 'detected signals'. Even with these cuts we cannot achieve perfect separation, since some background events fall along the $t_1 = t_2$ line. We may geometrically estimate

the fraction of background events that are misclassified as detected signals or false alarm fraction (FAF).

$$\text{FAF} = \frac{\text{area between cuts}}{\text{total area}} \approx .05 \tag{5.7}$$

By knowing about the signal correlation and making cuts in the full parameter space we should be able to achieve 95% detection efficiency with only 5% false alarm fraction. When we train a random forest of decision trees and classify one evaluation set, this is exactly the performance we see in figure 5.2: without any *a priori* knowledge of the parameter correlation. MVSC is able to take in only the individual parameters and, by making a series of 1D cuts on these parameters, construct the surface embedded in the full parameter space that best separates background from signal.

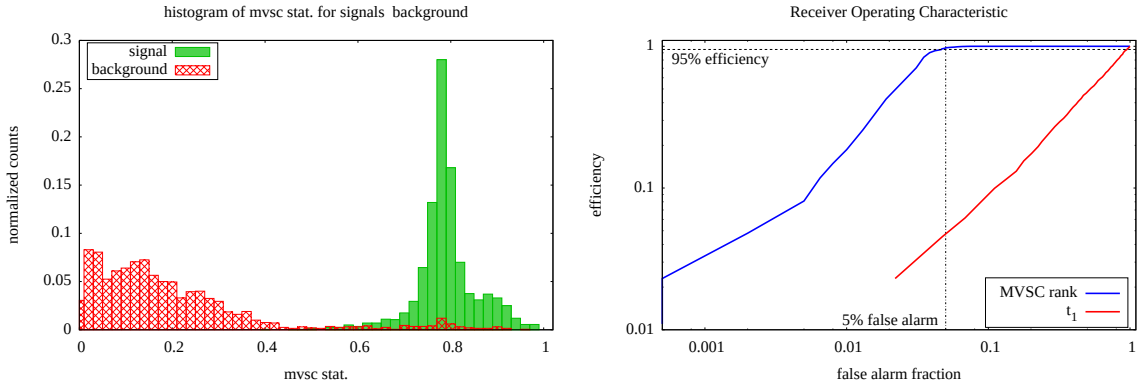

Figure 5.2. Histogram and reciever operating characteristic (ROC) for MVSC toy problem. The classifier is able to separate the two sets nicely. The right panel is an ROC curve comparing MVSC rank to t_1 as a detection statistic. False alarm fraction is the fraction of background events that are misclassified, and efficiency is the fraction of signal events that are correctly classified.

In this simple case it is trivial to discover that our two parameters are correlated by examining the data in two dimensions. However, in real data analysis problems the parameter space is not so tractable. For the S5 LIGO black hole ringdown search we identified 20 parameters of interest. For the S6 analysis we choose to include an additional four more. No human analyst could ever examine the full parameter space. Just like how the one-dimensional projections of our toy data proved useless, there is little chance that one, two, or even three dimensional projections of such data would provide insight.

5.3. LIGO Implementation

In LIGO data analysis we hope to distinguish between coincident gravitational wave events (signal) and random noise events in the data (background). The candidate gravitational wave triggers are selected by the coincidence analysis pipeline discussed in chapter 4. For simplicity our MVSC analysis operates on doubly coincident triggers (doubles) only, assigning each a rank, $\mathcal{M}$. After doubles have been ranked we can then form ranks for triply coincident triggers (triples) by combining the ranks of their constituent doubles. We can think of the MVSC rank for each of a triple trigger's member doubles as the result of an independent experiment on the single triple trigger. The MVSC ranks can then be combined by simple multiplication: the rank for the triple is the product of the ranks for the doubles:

$$\mathcal{M}_T = \mathcal{M}_{D1} * \mathcal{M}_{D2} * \mathcal{M}_{D3} \tag{5.8}$$

Remembering that the MVSC rank is a Bayes-factor, this is akin to treating the outcome of one experiment as the prior belief applied to the next experiment (see section 6.1).

There are many parameters that are calculated during coincident pipeline for each trigger. Some characterize single interferometer events such as SNR in a given interferometer. Others are only computed for doubly coincident events (including those that make up triples) such as the template metric distance between the single detector triggers. These are unique to each double. Still others are computed for coincident triggers that are unique to the coincident event (rather than an individual double). These parameters are the same for each double that makes up a triple, such as the chopped-L statistic defined in equation 4.26.

Because MVSC combines information from all parameters, it will always[2] produce rankings that are better than any individual parameter. Further, giving MVSC parameters that are not useful in separating signal from background has little to no effect for certain choices of the number of sampled parameters.

The training set for the LIGO ringdown search consists of triggers output by the coincident analysis pipeline. The training 'signals' are the found signal injections discussed in section 4.4. The training 'background' are triggers output by the time-slide analysis of section 4.3. The actual events we wish to rank are the candidate

[2]If the decision trees are not trained properly, MVSC will not beat a good single parameter (like chopped-L). Seeing this behavior, tells us that certain forest parameters, like number of trees per forest or minimum leaf size, need to tuned. In a worst case this can mean the training data are not sufficient to properly train MVSC.

gravitational wave triggers identified after the coincidence stage of our pipeline. All of these triggers are characterized by the same parameters. The full list of all parameters used in the S5/S6 ringdown analysis is found at the end of this chapter in table 5.1. These parameters were chosen to represent a wide variety of quantities that are known to have individual utility in separating signal from background.

For diagnostic purposes we wish to assign MVSC ranks to the injection and background triggers used to train our forest of decision trees. We do this following the same method as the toy problem. First we randomly break the training data into 10 subsets. Next we choose one of these subset to rank, and train a forest of trees using the nine remaining subsets as training data. This forest is used to assign ranks to the subset of interest. The process is repeated until all 10 subsets have been assigned MVSC ranks.

In practice we use the MVSC rank as a frequentist detection statistic. We can determine the false alarm rate (FAR) from the time-slide triggers, determining how often we expect to find a trigger with a particular rank as the result of detector noise and not a real gravitational wave event.

One issue that arose in using MVSC rank as a detection statistic had to do with injection finding. The training set uses coincident triggers associated with injections, so we must devise some criterion to determine whether a given injection has been found. This is accomplished using time coincidence. We know when an injection was

made into each interferometer data stream. If a trigger occurred within some window of the injection time, we say that the injection was found by that particular trigger.

This scheme can cause some complications. Consider a low amplitude injection made into a data stream. A loud trigger could occur at the injection time that is the result of a glitch or even a gravitational wave. This trigger would be incorrectly identified as a found injection. The exact same trigger could end up as a candidate gravitational wave event, because it was associated with an event in the real strain data not the injected signal. In this case MVSC will be trained with an event (the found injection) that is in the evaluation set. MVSC would train forests assuming such an event is a signal, and the evaluation event would be much more highly ranked than expected. This situation is the cause of the so-called 'MVSC bump' problem seen in a previous LIGO implementation of MVSC. MVSC was told in training that detector glitches are 'signals'. Those glitches were then classified as signals, causing a large excess of high significance events (the bump) at the end of the analysis.

To combat this problem, we use different criteria to determine whether an injection is found for the purposes of FAR calculation or MVSC training. All injections found with the original method are considered found and used with all background events to calculate FAR. This way injections 'found' with high FAR can be identified by the pipeline as not really found. For the injections that enter the MVSC training set we use a stricter criterion. For each candidate gravitational wave trigger, we choose a time window for each constituent single detector trigger. Any injections

made into these windows will be ignored for the purposes of training MVSC. In this way we can ensure that no injection triggers find their way into both the training and evaluation sets.

Finally, we can verify the efficacy of using MVSC rank as a detection statistic. By using MVSC to rank known signal injections and time-slide background events, we can compute a simple false alarm fraction and an injection finding efficiency. Then we can compare MVSC's performance to the performance of another statistic that is known to be effective.

The 'chopped L' statistic was defined for doubly coincident events in the LIGO S4 search for black hole ringdowns, where it was used as the detection statistic for doubles [2]. We extended this statistic to triples for the S5/S6 analysis in equation 4.26. Chopped L has shown to be an effective detection statistic, outperforming the S4 triple detection statistic in detection statistic comparison study [56]. We can see how MVSC rank compares to chopped L for a short two detector analysis in figure 5.3. For all values of false alarm fraction MVSC rank correctly identifies more injections than chopped L. This is to be expected as chopped L is one of the parameters given to MVSC.

5.4. MVSC as a LIGO Detection Statistic

After ranking all of the coincident pipeline triggers we can begin to compute the detection confidence from the MVSC ranks. We do this by computing a per year

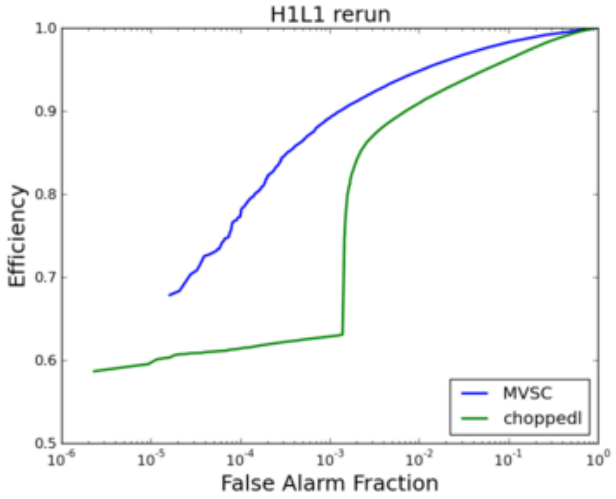

Figure 5.3. Receiver Operating Characteristic for MVSC and the chopped-L statistic. Efficiency is the fraction of correctly identified signal injections, and false alarm fraction is the fraction of incorrectly identified background events..

FAR for each event. For a detection statistic, $\mathcal{D}$ and the detection threshold $\bar{\mathcal{D}}$, we define the FAR as a function of detection threshold:

$$\mathrm{FAR}(\bar{\mathcal{D}}) = \frac{B(\mathcal{D} > \bar{\mathcal{D}})}{T_{\mathrm{obs}}} \tag{5.9}$$

where B is the number of background events with rank, $\mathcal{D}$, greater than the threshold and T_{obs} is the observation time in years. For each trigger it's false alarm rate is computed by FAR($\mathcal{M}$), where $\mathcal{M}$ is the MVSC rank of that trigger. We are asking: if we set the detection threshold such that this event is a detection, how many false alarm detections would we make?

The background can then be described as a cumulative number of 'gravitational wave' detections in a given time-slide analysis as a function of FAR. We find that the inverse false alarm rate (IFAR) is useful for interpretation. Large IFAR corresponds

to more confidence that a trigger is a detection. In figures 5.4 and 5.5 we show the cumulative event counts as a function of IFAR for two months of S5 data and 39 days of S6 data. The gray lines each represent one time-slide analysis. The blue triangles are the candidate gravitational wave events output by the coincident pipeline. There are root-N error contours, that estimate the 1 and 2 σ expected background for

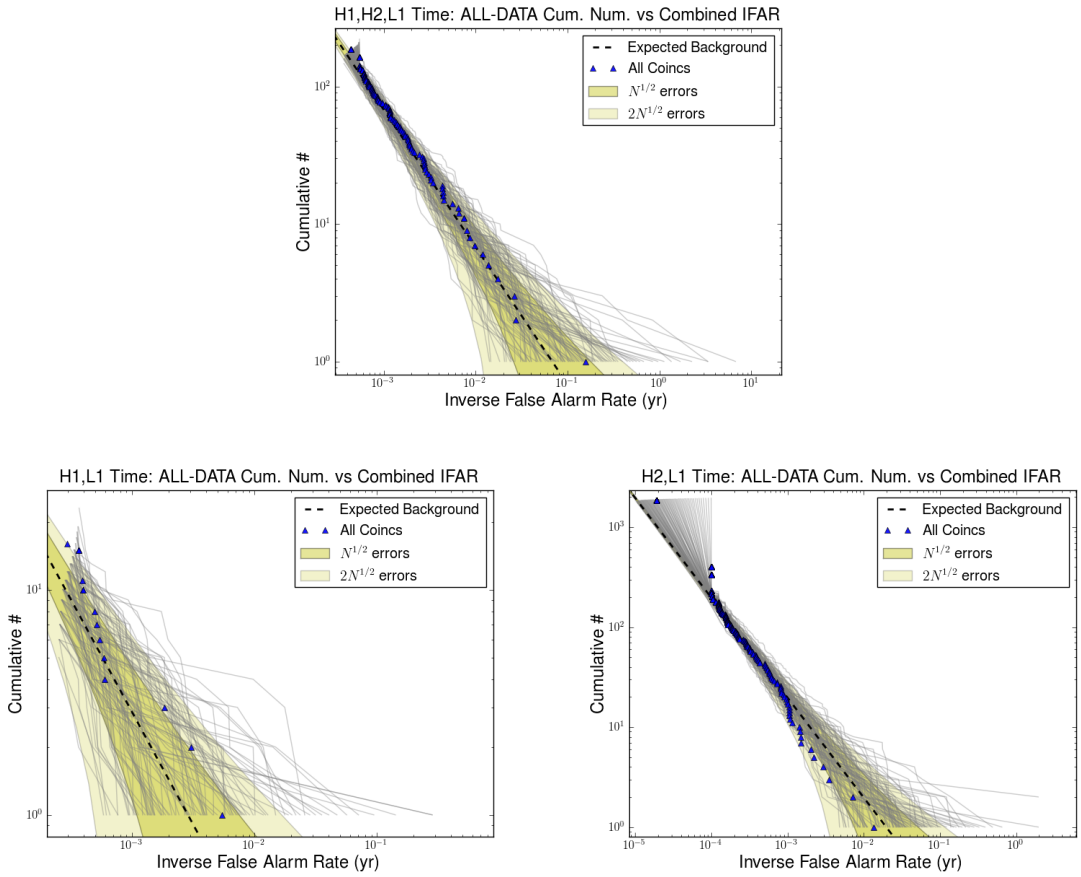

Figure 5.4. Inverse false alarm rate for 2 month analysis of LIGO S5 data. Trigger counts are consistent with background. There is non-Gaussian background excess for high IFAR. There was not much H1L1 time resulting in poor background estimation for this period. H2L1 shows a 'fanning' feature. This effect results from the discreteness of the MVSC rank.

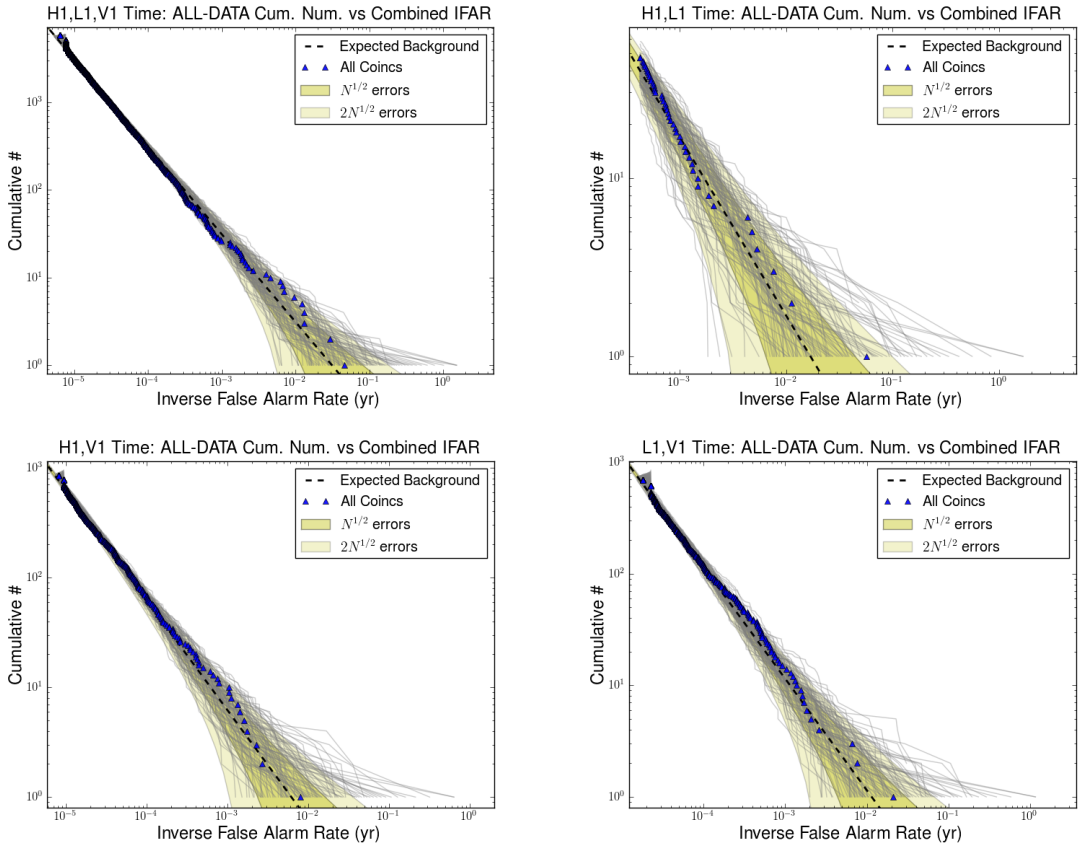

Figure 5.5. Inverse false alarm rate for 39 day analysis of S6, VSR3 data. Trigger counts are consistent with background. There is non-Gaussian background excess for high IFAR. The 'fanning' feature seen in S5 is absent, thanks to better tuning of the MVSC parameters and an increased training set size.

Gaussian noise. We see that in all runs the actual background is in excess of the Gaussian estimates. In each case the event rates are consistent with the background.

Because there were no plausible gravitational wave signals in our analysis, we can use our observations to set rate upper limits on compact binary coalescence events that result in black holes. We do this by determining the volume of space observed in which we were capable of making a detection and the time we observed this space.

We then say that the astrophysical rate of events must be less than 1 in that volume of space per that amount of time. Determining the observation time is quite simple, but determining the volume of space observed can be a bit tricky.

In section 4.4 we discussed how simulated signal software 'injections' are made in the data. We then determine which injected signals we recovered by comparing the detected time of arrival in the various detectors with the injected time of arrival. Each of these injections was assigned a MVSC rank themselves, and so we can compute the FAR for each recovered injection. An injection with a MVSC rank greater than the most significant background event will be assigned a FAR of zero. Other injections that are quiet in one or more detectors will be assigned non-zero FAR.

In figure 5.6 we see a large population of the software injections made into two months of S5 data, while all three interferometers were operating. These injections were made by all three waveforms discussed in section 4.4: the full inspiral-merger-ringdown (IMR) Phenomenological-B and Effective One Body-Numerical Relativity version 2 (EOBNRv2) in addition to the less physical naked ringdown. They also cover a very wide range of astrophysical systems. The final black hole masses range from 50 to 450 $M_{\odot}$ for the IMR and 5 to 900 for naked ringdown. For upper limits we want to determine our ability to detect an astrophysical population, so we will restrict ourselves to the EOBNRv2 waveform injections. Further we will restrict the mass ratios considered to $q \in [1, 4]$.

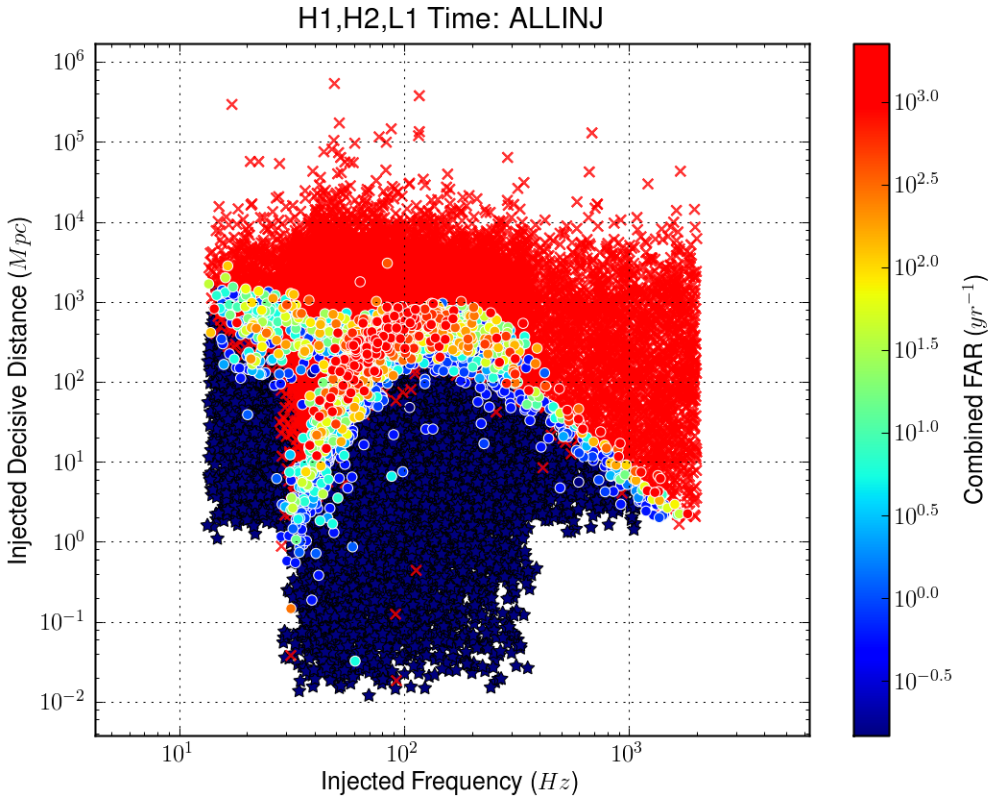

Figure 5.6. Missed / found plot for 2 months simulated signal injection analysis of S5 data by injected ringdown frequency and decisive distance. Note that the two populations of injections are discernible: the full IMR waveform injections are 'on top' and include to very small injected distances. The naked ringdown injections remain detectable to very low frequency due to an aphysical radiation efficiency artificially increasing their amplitude.

Next we need to determine the detection threshold for a found injection. Some distant injections will be assigned a high FAR. If a candidate event from our search pipeline has a FAR of say 300 per year, we could not call that a gravitational wave detection. In this case we will use the 'loudest event statistic'. If an injection is found to be more significant than the most significant candidate event from our search we will say it was found.

To get more meaningful upper limits we will divide the injections into several sub groups by final black hole mass and initial mass ratio. This binning scheme will allow us to quote upper limits for different astrophysical systems specifically. We chose to bin our injections in 50 $M_\odot$ ranges characterized by their midpoint. We can then count the fraction of injections found by the analysis pipeline in each mass bin for various source distances. This fraction of found injections is known as the injection efficiency, ϵ. It is from this efficiency that we estimate the effective volume of space probed by our search.

$$V_{\text{eff}} = 4\pi \int \epsilon(r) r^2 \mathrm{d}r \tag{5.10}$$

Preliminary injection finding efficiency for four of our eight total mass bins are shown in figure 5.7. These injections used EOBNRv2 waveforms for a fixed mass ratio of $q = 1$.

In LIGO's fourth science run the black hole ringdown search observed an volume of $V_{\text{eff}} = 2 \times 10^6$ Mpc^3 [2]. The S5 analysis probed approximately $V_{\text{eff}} = 4 \times 10^9$ Mpc^3, a three order of magnitude improvement [17]. While the S5 and S6 results are still preliminary, we can estimate the false alarm rate by knowing the observation time. For S5 LIGO detectors had about 70% live time of the total two year analysis after category 3 vetoes a were applied. Live time refers to any time when two or more detectors were operating (excluding H1H2 time).

Assuming that binary BH mergers are a Poisson process, we can note that the probability of observing no events is 10% for a Poisson distribution with mean rate of

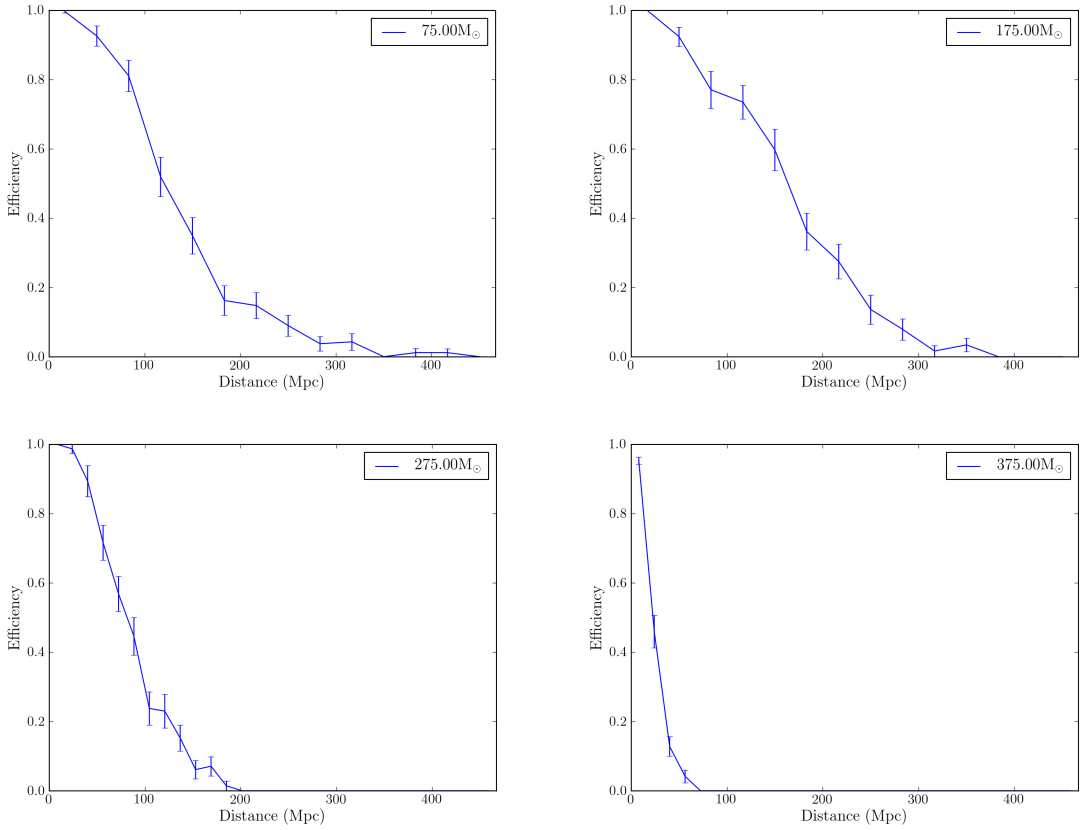

Figure 5.7. Injection finding efficiency curves for four total mass bins. There are an additional four mass bins centered at 125, 225, 325, and 425 $M_\odot$. Distance here is physical distance to the simulated source, not effective distance..

$\log 10$ [3]. So we are 90% sure that the mean rate is less than $\log 10$ per our observed space-time volume. We place a preliminary 90% confidence rate upper limit as:

$$R_{90\%} \lesssim \frac{\log 10}{V_{\text{eff}}\, T_{\text{live}}} \approx 4 \times 10^{-10}\,\text{Mpc}^{-3}\,\text{yr}^{-1} \tag{5.11}$$

The current rate of binary intermediate mass black hole (IMBH) mergers is computed to be 2.1×10^{-11} Mpc^{-3} yr^{-1} from astrophysical arguments [1]. With advanced

[3]The probability of observing k events in an observation period for a Poisson process with mean rate λ is $P(\lambda, k) = \frac{\lambda^k \exp(-\lambda)}{k!}$.

LIGO's predicted factor of 1000 increase in detectable volume, the advanced detector era promises to either detect gravitational waves from IMBH mergers or place the strongest constraints on the existence of IMBH sources to date.

5.5. Addendum: the MVSC Parameters

Here we list the many parameters used by MVSC in the S5 and S6 search for gravitational waves from black hole ringdowns. The full 24 parameters used in S6 are shown in table 5.1. Only the first twenty of were used in the S5 analysis. In addition we define any parameters that are not defined elsewhere in the text.

After the coincidence stage of the analysis there are potential many coincident events that occur at very similar times. We discard many of these in order to reduce the number candidate events and keep only the most likely to be real gravitational waves. This is done by keeping only one coincidence per 10 second window. We choose to keep the event with the largest 'chopped-L' statistic. The number of clustered coincidences is a simple count of how many coincident events there were before this discard is made. This is useful because gravitational wave ringdown events have a narrow frequency band whereas detector glitches tend to be broad-band. We expect detector glitches to produce many coincident triggers covering a broad range of frequencies in a small time window. In contrast gravitational wave signals should have only a few coincident triggers concentrated near the gravitational wave frequency. This parameter was only used in the S6 analysis

In S6 we also chose to allow MVSC to take data quality flags. Each single interferometer (IFO) trigger carries a boolean value as to whether a certain data quality flag was in effect at that IFO. In general this can be any data quality flag, but we choose to use the output of the hierarchical veto or 'hveto' pipeline. The hveto pipeline looks for correlations between periods of excess power in the detectors with hundreds of environmental and control channels monitored at each detector site [18]. It is hierarchical in the sense it finds the most significant correlations first and then looks for secondary correlations, *etc.* until no more are found. The hveto is particularly good at identifying families of glitches by their multiplicity of auxiliary channel correlations. It can then flag the entire family of glitches across long analysis periods. As a boolean, this parameter contains no information about what the hveto pipeline actually found, only that it identified a particular time as likely to contain glitches.

The S4 detection statistic is not defined elsewhere in the text, so we present it here without further comment:

$$\rho_{\mathrm{stat}} = \min \begin{cases} \sqrt{{\rho_a}^2 + {\rho_b}^2}, \\ 2\rho_a + 2.2, \\ 2\rho_b + 2.2 \end{cases} \tag{5.12}$$

quantity	definition	description
dt	$\lvert t_a - t_b \rvert$	difference of time of arrival between IFOs
df	$\lvert f_a - f_b \rvert$	difference of frequency recovered between IFOs
dQ	$\lvert Q_a - Q_b \rvert$	difference of quality factor recovered between IFOs
ds^2	$g_{ij} d\lambda_i d\lambda_j$	metric distance between recovered templates in two IFOs, $\lambda \in (t, f, Q)$
g_{tt}	eq. 4.22	$t-t$ template metric coefficient
g_{ff}	eq. 4.22	$f-f$ template metric coefficient
g_{QQ}	eq. 4.22	$Q-Q$ template metric coefficient
g_{tf}	eq. 4.22	$t-f$ template metric coefficient
g_{tQ}	eq. 4.22	$t-Q$ template metric coefficient
g_{fQ}	eq. 4.22	$f-Q$ template metric coefficient
ρ_a	$(s_a \lvert h_a)$	SNR in IFO$_a$
ρ_b	$(s_b \lvert h_b)$	SNR in IFO$_b$
φ	$\max\left(\frac{\rho_a}{\rho_b}, \frac{\rho_b}{\rho_a}\right)$	SNR ratio
ρ^2	$\sum_{\text{IFOs}} {\rho_i}^2$	network SNR
$\mathcal{L}$	eq. 4.26	'chopped-L' statistic
ρ_{stat}	eq. 5.12	S4 detection statistic
D_a	ς_a / ρ_a	effective distance to IFO$_a$, ς is the sensitivity of the detector
D_b	ς_b / ρ_b	effective distance to IFO$_b$
κ	$\max\left(\frac{D_a}{D_b}, \frac{D_b}{D_a}\right)$	effective distance ratio
dD	$\lvert D_a - D_b \rvert$	difference of effective distance to each IFO
${N_a}^\dagger$	see above	number of clustered coincidences in IFO$_a$
${N_b}^\dagger$	see above	number of clustered coincidences in IFO$_b$
${H_a}^\dagger$	see above	boolean value, whether a data quality flag is in effect for IFO$_a$
${H_b}^\dagger$	see above	boolean value, whether a data quality flag is in effect for IFO$_b$

Table 5.1. The parameters used by MVSC. The subscripts a and b refer to the values of quantities in the individual IFOs. All of these parameters are invariant under interchange of a and b. The quantities marked by a † were only used in the S6 analysis.

6. BAYESIAN MODEL SELECTION

If we were to know the exact workings of a physical process, then we could, using some initial data, predict the outcome of any experiment. The goal of the scientific endeavor is the converse. We wish to reverse engineer the rules governing physical processes from the outcomes we observe. Science then could be described as a method to determine an 'iterative forward solution' to nature. We repeatedly propose hypotheses and determine if they agree with our observations. Of course once a hypotheses has proven to be consistent with observations, it does not mean the work is complete.

As time passes and new data become available, we can further refine our hypotheses. For example, better experiments continue to reinforce our hypothesis that the photon is massless. Based on several experiments between 2004 and 2012, we are able to place stronger bounds on the rest mass of a photon, reducing from less than 6×10^{-17} eV/c^2 to 1×10^{-18} eV/c^2 [10].

The reverse can also happen. As we saw in section 1.1, Newtonian gravity was consistent with all observations until Mercury's perihelion precession was well measured. This one chink in the Newtonian armor was enough to cast doubt on the whole theory.

The question of science then becomes: how can we quantify the level of agreement of a hypothesis with observations? Without a tool to measure this agreement how

can we distinguish competing hypotheses that each have some level of agreement with observations?

To this end we turn to probability theory. The Reverend Thomas Bayes developed a probability theory that applied to binomial hypotheses [9]. His formulation of probability theory was concerned with how to decide between two choices given some observation. Simon-Pierre Laplace extended Bayes' work beyond simple *yes/no* decisions and became the first to apply Bayesian probability to problems of celestial mechanics [54].

Let us begin to develop Bayesian probability theory by defining $p\left(A \,|\, B\right)$ as the probability that a statement, A, is true given that another statement, B, is known. Similarly, we may write $p\left(A \,|\, B, C\right)$: the probability that A is true given both B and C are known. Finally, we may write $p\left(A, B \,|\, C\right)$: the joint probability that both A and B are true given C is know. In both of these combined statements the comma represents the logical *and*. Such probabilities have a product rule:

$$p\left(A, B \,|\, C\right) = \, p\left(A \,|\, C\right) \, p\left(B \,|\, A, C\right) = p\left(B \,|\, C\right) \, p\left(A \,|\, B, C\right). \tag{6.1}$$

Only when A and B are independent statements does this reduce to the simple product rule:

$$p\left(A_{\text{ind}}, B_{\text{ind}} \,|\, C\right) = \, p\left(A_{\text{ind}} \,|\, C\right) \, p\left(B_{\text{ind}} \,|\, C\right). \tag{6.2}$$

As scientists, we are interested in determining the probability that a hypothesis, $\mathcal{H}$, is true given both some observed data, d, and some information we already know, $\mathcal{I}$. Formally, this is $p\left(\mathcal{H} \,|\, d, \mathcal{I}\right)$ and is called the *a posteriori* probability. It tells us

the probability that a hypothesis is true after we have observed some new data, d. Using the product rule of equation 6.1, we may write:

$$p\left(\mathcal{H}\,|\,d,\mathcal{I}\right)=\frac{p\left(\mathcal{H}\,|\,\mathcal{I}\right)\,p\left(d\,|\,\mathcal{H},\mathcal{I}\right)}{p\left(d\,|\,\mathcal{I}\right)}. \tag{6.3}$$

This statement is now known as Bayes theorem. It provides a purely algorithmic prescription for how we update our belief in a hypothesis after making a new observation. On the right hand side $p\left(\mathcal{H}\,|\,\mathcal{I}\right)$ expresses the probability that $\mathcal{H}$ is true given our previous knowledge $\mathcal{I}$ and is known as the *a priori* probability. $p\left(d\,|\,\mathcal{H},\mathcal{I}\right)$ is the probability that some data d would be observed if our hypothesis $\mathcal{H}$ is true and is called the Bayesian likelihood. Both of these quantities can be easily defined and computed. The Bayesian evidence, $p\left(d\,|\,\mathcal{I}\right)$, is another story. We may notice that this term is independent of the hypothesis. When comparing competing hypotheses, we may treat the evidence as a normalization constant to be determined by direct integration.

If we have more than one observation, we can use the product rule to find the joint *a posteriori* probability for both:

$$\begin{aligned}p\left(\mathcal{H}\,|\,d_1,d_2,\mathcal{I}\right)&=\frac{p\left(\mathcal{H}\,|\,\mathcal{I}\right)\,p\left(d_1,d_2\,|\,\mathcal{H},\mathcal{I}\right)}{p\left(d_1,d_2\,|\,\mathcal{I}\right)}\\&=\frac{p\left(\mathcal{H}\,|\,\mathcal{I}\right)\,p\left(d_1\,|\,\mathcal{H},\mathcal{I}\right)}{p\left(d_1\,|\,\mathcal{I}\right)}\,\frac{p\left(d_2\,|\,\mathcal{H},d_1,\mathcal{I}\right)}{p\left(d_2\,|\,d_1,\mathcal{I}\right)}.\end{aligned} \tag{6.4}$$

We recognize the first term of equation 6.4 as simply $p\left(\mathcal{H}\,|\,d_1,\mathcal{I}\right)$. The Bayesian method provides a natural way for us to combine observations. After observing d_1 we can compute the *a posteriori* probability for our hypothesis. This new probability

becomes part of the information we already know. That is to say we use the old *a posteriori* probability as the new *a priori* knowledge for the next observation. We are left with:

$$\begin{aligned} p\left(\mathcal{H} \mid d_1, d_2, \mathcal{I}\right) &= \frac{p\left(\mathcal{H} \mid d_1, \mathcal{I}\right)\, p\left(d_2 \mid \mathcal{H}, d_1, \mathcal{I}\right)}{p\left(d_2 \mid d_1, \mathcal{I}\right)}, \\ &= \frac{p\left(\mathcal{H} \mid \mathcal{I}'\right)\, p\left(d_2 \mid \mathcal{H}, \mathcal{I}'\right)}{p\left(d_2 \mid \mathcal{I}'\right)} = p\left(\mathcal{H} \mid d_2, \mathcal{I}'\right), \end{aligned} \tag{6.5}$$

where $\mathcal{I}' = (d_1 \,\text{and}\, \mathcal{I})$. This prescription allows us to iteratively update our belief in a hypothesis over the course of multiple independent observations.

Some have criticized Bayesian probability for its dependance on our *a priori* beliefs. This is not a failing of the Bayesian method but a feature. All analyses carry with them the preconceived beliefs of the analyst. In a Bayesian analysis these beliefs are transparently laid bare for all to see. While it is true that the specific choice of *a priori* information effects the *a posteriori* probability, the Bernstein–von Mises theorem states that given a large number of observations, the final *a posteriori* probability is independent of the choice of *a priori* information. So even if several observers begin with differing *a prior* beliefs, the data will convince all to agree in the end.

For convenience we will anglicize the *a posteriori* and *a priori* probabilities as the posterior and prior respectively.

6.1. Bayesian Inference

Once we have a posterior probability in hand we can begin to make inferences. Lets say we have two hypotheses, $\mathcal{H}_1$ and $\mathcal{H}_2$, and we want to know which to favor. We can compute the posterior probability, $\mathcal{P}$, for each, and from these we can compute the Bayes factor:

$$\mathcal{B}_{12} = \frac{\mathcal{P}_1}{\mathcal{P}_2}. \tag{6.6}$$

The Bayes factor tells us exactly how much the data, d, favors $\mathcal{H}_1$ over $\mathcal{H}_2$. When $\mathcal{B}_{12} \gg 1$, $\mathcal{H}_1$ is favored; when $\mathcal{B}_{12} \ll 1$, $\mathcal{H}_2$ is favored; and when $\mathcal{B}_{12} \approx 1$, neither hypothesis is favored over the other. We then use the Bayes factor to update our belief in the hypotheses. In the case that we have a prior belief that one hypothesis should be favored over the other the new observation will update the true odds that one hypothesis is favored over the other:

$$\mathcal{O}_{12} = \frac{p_1\,\mathcal{P}_1}{p_2\,\mathcal{P}_2} = p_{12}\,B_{12}, \tag{6.7}$$

here p_i is the prior belief in hypothesis $\mathcal{H}_i$ and p_{ij} is the prior ratio. $\mathcal{O}_{12}$ is the odds ratio between hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$. The odds ratio can be thought of as a betting odds. If $\mathcal{O}_{12} = 100$, then we favor hypothesis $\mathcal{H}_1$ at 100:1. This method allows us to compare any two hypotheses, but what if we have more than two competing hypotheses?

If there are only a small number of hypotheses, we can compute the odds ratio for each combination of two. Using a variant of the above method, we can also compute

the joint odds for some subset against a second subset: one 'null' hypothesis versus all others for example. This method must be extended if we hope to cover many competing hypotheses.

Consider a hypothesis that is parameterized by θ. This case is equivalent to having an infinite number of competing hypotheses, each characterized by a different value of θ. Instead of having a posterior probability $\mathcal{P}_i$ for each of two hypotheses $\mathcal{H}_i$, we have a posterior probability for each value of θ. Our posterior probability can be expressed as a function of θ:

$$\mathcal{P}(\theta) = p\left(\mathcal{H}(\theta) \,|\, d, \mathcal{I}\right) \tag{6.8}$$

In this case the posterior probability is a probability density function. We assume that one of the hypotheses is correct and θ allows us to cover all possible hypotheses. It follows that:

$$\int_{\theta} \mathcal{P}(\theta)\, \mathrm{d}\theta = 1. \tag{6.9}$$

The 'best' hypothesis is the one that maximizes the posterior probability. We will refer to this choice of θ as the *maxima a posteriori* (MAP).

Using equation 6.9 we may place limits on subsets of our hypotheses given our observed data and prior knowledge. For example, we could compute the probability that θ is smaller than some threshold, θ_0:

$$p\left(\theta < \theta_0 \,|\, d, \mathcal{I}\right) = \int_{\theta<\theta_0} \mathcal{P}(\theta)\, \mathrm{d}\theta. \tag{6.10}$$

We could also use equation 6.9 to place confidence bounds on our hypotheses. For instance we could determine the value θ_0 such that we are 90% sure the true value

of θ is less than θ_0 by inverting:

$$\int_{\theta<\theta_0} \mathcal{P}(\theta)\,\mathrm{d}\theta = 0.9. \tag{6.11}$$

If we determined the MAP hypothesis, $\theta_\star$, we can determine an equidistant 90% confidence interval, $\theta_\star \pm \Delta\theta$, by inverting:

$$\int_{\theta_\star-\Delta\theta}^{\theta_\star+\Delta\theta} \mathcal{P}(\theta)\,\mathrm{d}\theta = 0.9. \tag{6.12}$$

Similarly, we may determine an equiweight 90% confidence interval by finding $\Delta\theta_+$ and $\Delta\theta_-$:

$$\int_{\theta_\star-\Delta\theta_-}^{\theta_\star} \mathcal{P}(\theta)\,\mathrm{d}\theta = \int_{\theta_\star}^{\theta_\star+\Delta\theta_+} \mathcal{P}(\theta)\,\mathrm{d}\theta = 0.45 \tag{6.13}$$

We can make these same inferences for a set of hypotheses parameterized by any number of parameters. Consider a hypothesis, $\mathcal{H}(\boldsymbol{\theta})$, parameterized by a N-dimensional vector $\boldsymbol{\theta}$. Using the fact that $\mathcal{P}(\boldsymbol{\theta})$ is normalized:

$$\int_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta})\,\mathrm{d}^{\mathrm{N}}\boldsymbol{\theta} = 1, \tag{6.14}$$

we may make inferences on $\mathcal{H}(\boldsymbol{\theta})$ in any one dimension by marginalization, or in multiple dimensions jointly.

The posterior probability gives us a powerful tool for scientific inference, but it all hinges on our posterior being a true, normalized probability density. Typically, the prior belief, $p\,(\mathcal{H}\,|\,\mathcal{I})$, and the Bayesian likelihood, $p\,(d\,|\,\mathcal{H},\mathcal{I})$, are known functions. As we mentioned above, in this case we can compute the Bayesian evidence (the all so important normalization), $p\,(d\,|\,\mathcal{I})$, by direct integration. Unfortunately, when

$\mathcal{H}$ is parameterized in a high dimensional space this integration can be impractical. Without the evidence all we are able to compute is the relative posterior weight between two points in our parameter space, not the true posterior probability.

All is not lost. There are several methods for drawing samples from an unnormalized probability distribution. With enough samples we may construct a numerical approximation of the true, normalized probability density to make inferences.

6.2. Markov chain Monte Carlo

A Markov chain is a random process of ordered, memoryless transitions between states. Markov chains have many interesting mathematical properties, but we will focus on their ability to draw samples from a probability distribution[46]. Formally we may generate a Markov chain with a transition matrix, $\mathbf{p}$. Starting in a state, x_i, the probability of moving to a new state, x_j, is given by p_{ij}. If the transition matrix, $\mathbf{p}$, is stationary, then the states, $\{x\}$, are samples from a stationary distribution. So by judiciously choosing $\mathbf{p}$, we can ensure that we generate samples from the posterior probability distribution of interest.

The detailed balance condition helps us further. If the probability of transitioning from state x_i to state x_j is the same as the reverse:

$$x_i\, p_{ij} = x_j\, p_{ji}, \tag{6.15}$$

then our samples, $\{x\}$, are from a stationary distribution, even if $\mathbf{p}$ is not stationary. Enforcing detailed balance will allow us to choose transition matrices that are functions of time or even change randomly.

Of course in practice we may not be able to figure out which $\mathbf{p}$ (or even $\mathbf{p}(t)$) will give us our posterior probability. Luckily for us, the Metropolis-Hastings algorithm tells us how to use an arbitrary transition probability, along with a determinable acceptance rate, to mimic $\mathbf{p}$ and achieve samples from any stationary distribution [31].

Let q_{ij} be a proposal distribution to transition to a new state and k_j be the acceptance rate for that transition. So $q_{ij}\, k_j$ is the actual transition probability. We wish to draw samples from a posterior probability distribution, so we will label them accordingly: $\mathcal{P}_i$. To maintain detailed balance we must have:

$$\mathcal{P}_i\, q_{ij}\, k_j = \mathcal{P}_j\, q_{ji}\, k_i \tag{6.16}$$

We may freely choose $k_i = 1$ and always accept the reverse transition. Now we can determine what acceptance rate for the proposed transition, k_j, ensures detailed balance:

$$k_j = \frac{\mathcal{P}_j\, q_{ji}}{\mathcal{P}_i\, q_{ij}} = H \tag{6.17}$$

This acceptance rate is known as the Hastings ratio. Because k_j is a probability, we note that $k_j = \min(1, H)$. From the Hastings ratio, we see that the Markov chain prefers states where $\mathcal{P}_j > \mathcal{P}_i$ and $q_{ij} < q_{ji}$. If the posterior is large at a given state, then we need many samples to accurately describe the posterior. Therefore, we will

accept more transitions to collect those samples. If the probability of choosing a particular state is low, we may not get another chance to go there. In this case we should take samples now rather than later. In this way by using an arbitrary proposal distribution, we may draw samples from our posterior.

There is one more complication to consider: accessibility. We can only sample our posterior, if it is possible to propose a transition to any state allowed by the posterior. The posterior is defined on the domain of the prior, so the volume of allowable posterior states is called the 'prior volume'. If there are states in the prior volume that can never be proposed, they will never be sampled. In this case the Markov chain samples are drawn from a sub-set of the posterior, not the true posterior. We don't need to have access to every possible state in the prior volume from any particular state, we just need to be able to make it from any one to any other in a countable number of transitions.

The Metropolis-Hastings algorithm does more than draw posterior samples. The posterior probability appears only as a ratio in our acceptance rate. This allows us to draw samples from a non-normalized posterior. We no longer need to compute the Bayesian evidence by costly direct integration to make inferences. We can generate samples from posterior probability as a Markov chain and use the samples to approximate the true posterior for inference.

Over the course of generating samples as a Markov chain, we computed the unnormalized posterior density for many locations in parameter space. The best fit, or

MAP, parameters are those at which the posterior density is maximized. Since the normalization is the same for all points in parameter space, we can determine this easily by searching through the posterior samples.

The posterior samples can be used directly to infer the posterior probability density. For instance the probability that a parameter θ is in some range $\Delta\theta$ is given by the fraction of samples that fell in that range. In this way we may use our posterior samples to determine parameter confidence or exclusion limits by approximating the integrations in equations 6.11 through 6.14.

6.3. Reverse-Jump Markov chain Monte Carlo

In Markov chain Monte Carlo simulations we draw samples from a posterior probability density for a parameterized hypothesis. In this case we treat the underlying problem of model selection as a problem of parameter estimation. Each point in parameter space represents one of a continuum of models. We seek to find the best fit model, which defines the best fit parameters. From this simple situation it is possible to construct more complicated sets of hypotheses. For example, consider two hypotheses, $\mathcal{H}_1(\boldsymbol{\theta}_1)$ and $\mathcal{H}_2(\boldsymbol{\theta}_2)$, where each is parameterized by a different vector of parameters, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. For each hypothesis we can draw samples from its respective posterior, finding the best fit $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. But how can we determine whether $\mathcal{H}_1(\boldsymbol{\theta}_1)$ or $\mathcal{H}_2(\boldsymbol{\theta}_2)$ is preferred?

In this case we have a hierarchical model selection problem, where sub-models are nested inside of larger models. The basic Metropolis-Hastings algorithm can be extended to accommodate these multiple parametrized hypotheses. The Reversible Jump Markov chain Monte Carlo (RJMcMC) draws samples from among all hypotheses generating a combined posterior that contains information about which hypothesis and which parameters in that hypothesis are preferred [29].

To accomplish this sampling we need to establish an acceptance rate for transitions across the multiple hypotheses. We modify the Hastings ratio as follows:

$$H = \frac{\mathcal{P}_j \, q_{ji}}{\mathcal{P}_i \, q_{ij}} \det\left(J_{ij}\right) \tag{6.18}$$

where $\mathcal{P}_i = p\left(\mathcal{H}_i(\boldsymbol{\theta}_i) \,|\, d, \mathcal{I}\right)$ is the posterior density in the i^{th} model, q_{ij} is the proposal distribution for selecting parameters $\boldsymbol{\theta}_j$ in the j^{th} model from the original location $\boldsymbol{\theta}_i$ in the i^{th}, and

$$J_{ij} = \frac{\partial(\boldsymbol{\theta}_i, d\boldsymbol{\theta}_j)}{\partial(\boldsymbol{\theta}_j, d\boldsymbol{\theta}_i)} \tag{6.19}$$

is the Jacobian defining the coordinate transformation from the $\boldsymbol{\theta}_i$ to the $\boldsymbol{\theta}_j$ basis. In the case that $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ are of different dimension, new parameters, $d\boldsymbol{\theta}_i$, must be drawn to fully specify $\boldsymbol{\theta}_j$ (which then contribute to the proposal density). If the i^{th} and j^{th} models are independent, then $\boldsymbol{\theta}_j$ has no functional dependence on $\boldsymbol{\theta}_i$ and J_{ij} is unity.

Similarly, if the i^{th} model is nested in the j^{th}, with $\boldsymbol{\theta}_i$ occupying a subspace of $\boldsymbol{\theta}_j$, then we can carry the old $\boldsymbol{\theta}_i$ into the new $\boldsymbol{\theta}_j$ and independently generate any

new parameters. The Jacobian is unity for this case as well. In what follows we will restrict ourselves to nested models, adding or subtracting one degree of freedom at a time.

Once we have our posterior samples we can determine which hypothesis, $\mathcal{H}_i$, is favored. The posterior probability for each hypothesis is given by:

$$\mathcal{P}_i = \lim_{N_{\rm tot}\to\infty} \frac{N_i}{N_{\rm tot}}, \tag{6.20}$$

where $N_{\rm tot}$ is the number of posterior samples and N_i is the number of samples that came from the $i^{\rm th}$ hypothesis. While we can never have an infinite number of samples, we can approximate $\mathcal{P}_i$ by using a large number of samples. The most favored hypothesis is the one with the most samples. The Bayes factor between any two hypotheses defined in equation 6.6 is just:

$$\mathcal{B}_{ij} \approx \frac{N_i}{N_j} \tag{6.21}$$

We can also compute the Bayes factor comparing any one hypothesis against all others:

$$\mathcal{B}_i = \frac{\mathcal{P}_i}{1-\mathcal{P}_i} \approx \frac{N_i}{N_{\rm tot}-N_i}. \tag{6.22}$$

Littenberg describes several more methods to estimate the Bayes factor for model selection problems [40].

Once we have determined the hypothesis that is most favored by the data, we can determine the MAP parameters for that hypothesis. This defines the true *maxima a posteriori*. Note how this is not simply the largest unnormalized posterior value

found in the chain. If the competing hypotheses are of differing dimension, then the prior volume will be different for each. In this case the posterior normalization will be different for each hypothesis. Any inference we make on the parameters, $\boldsymbol{\theta}$, must be exclusively at one level of our hierarchical model selection problem.

7. THE BAYESWAVE ALGORITHM

In chapters 4 and 5 we focused on detecting gravitational waves from known sources. With foreknowledge of the gravitational waveform, we were not only able to detect signals, but also extract their parameters by maximizing the inner product $\langle h \,|\, s \rangle$. When we don't have *a priori* information as to the particular waveform in question we will need a new method.

The BayesWave algorithm is a method for detecting un-modeled gravitational waves. We will look for generic excesses of power in a gravitational wave detector. If the noise in a detector is stationary and Gaussian, then any short term deviation from the norm could be considered a candidate gravitational wave. Of course in terrestrial interferometer detectors like LIGO, the noise is neither stationary nor Gaussian. We must demand that a true gravitational wave signal appears in multiple detectors, just like in the known signal case.

There are several existing algorithms that attempt to detect un-modeled gravitational wave bursts. Some of the most promising are the coherent Wave Burst (cWB) algorithm [35] and the X-pipeline [55]. Both of these seek to reconstruct gravitational wave bursts coherently. In coherent methods one uses the sum of all detector responses to build a true network signal-to-noise ratio (SNR). The coherent sum requires information about the point of origin, or 'sky location', of a gravitational wave signal in order to account for the angular detector response and the time of arrival in the various detectors. A coherent detection then provides a best fit sky location for the

gravitational wave source in a way that the coincidence method discussed in section 4.2 does not. Both cWB and X-pipeline create a list of sky locations and construct a coherent data stream, summing the individual detector data, for each. Finally, either algorithm must analyze each coherent data stream individually, looking for excess power.

In any case we will need to be able to characterize the transient bursts in a meaningful way. In the X-pipeline this is accomplished by performing many short duration Fourier transforms of the coherent data. These are then compared to a longer term estimation of the power spectral density. Any deviation from the expected levels becomes a candidate gravitational wave event and is in turn characterized by its time, frequency, and sky location (that is in which coherent data set it was found). cWB operates in the wavelet domain. It estimates a wavelet domain 'power spectrum' and looks for individual wavelet coefficients that deviate from the expected spectrum. Again, the candidate gravitational waves can be characterized by a time, frequency, and sky location.

There are several methods to reject any detector noise or 'glitches' that accidentally passes either of these coherent analyses. One of the most powerful is the use of 'null streams'. With at least three detectors one can construct a linear combination of time shifted, single detector data such that there is zero response to gravitational waves in the new data stream [61]. For exactly three detectors labeled 1,2,3 and a

signal with sky location (θ, ϕ):

$$N(\theta, \phi, t) = A_{23}\, h_1(t) + A_{31}\, h_2(t + \tau_{12}) + A_{12}\, h_3(t + \tau_{13}), \tag{7.1}$$

where

$$A_{ij} = F_i^{+}(t_i, \theta, \phi)\, F_j^{\times}(t_j, \theta, \phi) - F_j^{+}(t_j, \theta, \phi)\, F_i^{\times}(t_i, \theta, \phi), \tag{7.2}$$

$F_i^{+/\times}$ is the angular detector response function for the i^{th} detector, τ_{ij} is the signal travel time between detectors i and j for the sky location (θ, ϕ). Since the null stream has no response to gravitational waves, any excess power in it must be the result of a detector glitch. By using this fact if we find a candidate gravitational wave in the coherent stream at the same time that there is an excess in the null stream, then we can reject the candidate event as the result of a glitch.

BayesWave shares some similarities with the two methods we have described, but there are a few key difference. BayesWave works in the wavelet domain like cWB, looking for wavelet coefficients that are larger than the expected average. However, cWB treats the gravitational wave detection problem as a linear algebra problem, inverting the detector response using regularization techniques. BayesWave uses a Markov chain Monte Carlo (McMC) to sample a posterior distribution function, repeatedly proposing gravitational wave solutions to find the best one. This means that BayesWave can treat the signal sky location as a model parameter and explore all possible sky locations within the McMC, rather than constructing coherent data streams in advance.

The most striking difference between BayesWave and the existing gravitational wave burst detection methods is their treatments of detector noise. BayesWave models the detector noise and the gravitational wave signal simultaneously. In essence BayesWave has a three layer model for the detector strain: it models gravitational waves as a strain in a fiducial coordinate system at the geocenter, $h(t)$; it models non-Gaussian noise glitches in each independent detector, $g_i(t)$; it models the remaining noise in the detector as a piece-wise fit to colored, Gaussian noise, $n_i(t)$. This can be summarized as:

$$s_i(t) = \mathcal{A}_i(t, \theta, \phi)\, h(t) + g_i(t) + n_i(t), \tag{7.3}$$

where $\mathcal{A}_i$ is a projection operator mapping the gravitational wave strain from the geocenter to the i^{th} detector. Because all of these models operate simultaneously, BayesWave can actively determine whether or not there are gravitational waves and/or glitches present, using the reversible jump McMC (RJMcMC) method. After fitting all of the non-Gaussian components of the detector strain (h and g), the remaining noise, n, will be Gaussian by construction. We can safely use data analysis techniques that have assumed Gaussian detector noise. This method of active noise modeling was introduced by Littenberg and Cornish [39].

We will now discuss some of the details of the BayesWave algorithm. First we will remind ourselves of wavelet transforms and their utility. Then we will discuss how we apply the Bayesian model selection techniques discussed in chapter 6 to real data analysis problems.

7.1. Wavelet Decomposition

A time series by definition gives ideal time resolution at the expense of any frequency information. Similarly, a Fourier series sacrifices all time resolution to gain frequency resolution. This trade off is summarized by the Fourier uncertainty principle: *A nonzero function and its Fourier transform cannot both be sharply localized*[28]. While this is true of all functions, we can hope to represent data in a basis that captures at least some of both the time and the frequency content.

A wavelet decomposition is one solution to this problem. A wavelet transform, like a Fourier transform, is nothing more than an integral transform:

$$\tilde{h}(\xi) = \int_{-\infty}^{\infty} h(t)\mathcal{K}(t,\xi)dt. \tag{7.4}$$

In the case of a Fourier transform the kernel function, $\mathcal{K} \propto \exp(-\imath\,\xi t)$, is purely oscillatory and persists for all time. This acts to put a time series into a basis purely defined by frequency, where we have lost all timing information. A wavelet transform uses kernel functions that have both a well defined frequency and a finite duration. A sine-Gaussian:

$$\mathcal{K}(t,\xi) \propto \sin(\xi t)\,\exp\left(-(t-t_0)^2/2\sigma^2\right) \tag{7.5}$$

fits this description. Unfortunately, such a choice contains two new free parameters, t_0 and σ. Further, a basis defined by sine-Gaussians is not orthogonal or complete. Technically the sine-Gaussian transform is a wavelet transform, just not the most

useful one. We wish to use a orthogonal, complete wavelet basis so that any time domain signal can be accurately and uniquely constructed with our wavelets.

The basic method to construct a kernel for the integral transform is to use the dilation equation [45]. Before we get started we will define a dimensionless time coordinate, $u = t/T$, where T is the analysis period. We can also define the fundamental frequency of the analysis as $f_0 = 1/T$. We will begin the construction of our kernel function or 'mother wavelet' by defining a scaling function:

$$\phi(u) = \sum_{k=0}^{n-1} c_k \, \phi(2u - k), \tag{7.6}$$

in terms of dilated versions of itself[1] and n real generating coefficients, c_k. It is usually not possible to solve the dilation equation directly, so instead we must construct $\phi(u)$ iteratively. We can define

$$\phi_j(u) = \sum_{k=0}^{n-1} c_k \, \phi_{j-1}(2u - k) \tag{7.7}$$

and pick an arbitrary, non-zero function, ϕ_0, as the starting point (in practice it is common to begin with a step function). After many iterations of constructing ϕ_j from ϕ_{j-1}, ϕ_j will reach a steady state. This is the solution to the dilation equation, and thus our time domain scaling function.

From the scaling function we can construct the mother wavelet:

$$W(u) = \sum_{k=0}^{n-1} (-1)^n \, c_k \, \phi(2u + k - N + 1), \tag{7.8}$$

[1]Technically the scaling function is built from compressed versions of itself, at higher frequency.

All other wavelets are simply dilations and translations of the mother wavelet:

$$W_{ij}(u) = W(2^i u - j). \tag{7.9}$$

Note how the i index tells us about the wavelet frequency relative to the fundamental f_0, and the j index tells us about the time translation as a fraction of the analysis time.

The mother wavelet is completely determined by the choice of generating coefficients, c_k, and depending upon that choice we can bestow our wavelets with a number of usefully properties. For instance the wavelets are orthogonal if:

$$\sum_k c_k c_{k+2m} = 0, \qquad \text{for all} \quad m \neq 0. \tag{7.10}$$

There are other conditions on the coefficients to achieve wavelet transformations that are area preserving (in the time frequency plane), or meet various accuracy requirements. These three conditions: orthogonality, area, and accuracy, produce a set of n equations on any n generating coefficients. For every choice of coefficient number n there is a unique set of coefficients that produces a 'good' wavelet basis. By using more generating coefficients the accuracy condition gets stricter and stricter, effectively making the wavelets more like perfect top-hats in the Fourier domain.

We can finally define the discrete wavelet transform of our time domain data, $s(u)$. Assuming periodic boundary conditions it is given by:

$$a_{2^i+j} = 2^j \int_0^1 s(u)\, W_{ij}(u) \mathrm{d}u, \tag{7.11}$$

$$a_0 = \int_0^1 s(u)\, \phi(u) \mathrm{d}u, \tag{7.12}$$

and the data is reconstructed by:

$$s(u) = a_0\phi(u) + \sum_i \sum_{j=0}^{i-1} a_{2^i+j} W_{ij}(u). \tag{7.13}$$

The a_ks are the wavelet amplitudes, the wavelet domain representation of the data, s. For convenience we have chosen to define a single wavelet index $k = 2^i + j$. The wavelet amplitudes are a one dimensional series in k of equal length to the original time series. In final wavelet domain representation we often refer to individual wavelet amplitudes as 'pixel' amplitudes or simply 'pixels'.

Actually computing the DWT by direct integration is costly. Luckily, there is an algorithm that can perform DWTs at a similar cost as a fast Fourier transform. The actual construction of a time domain mother wavelet functions from the generating coefficients can be posed as a linear algebra problem, if the coefficients meet the orthogonality condition, equation 7.10. Because of this, the whole wavelet transform can be mapped to another linear algebra problem using only the generating coefficients. This fact makes it possible to compute a DWT without ever constructing the actual time domain mother wavelet. This implementation of the DWT is known as Mallat's pyramid [41]. Mallat's pyramid effectively performs a series of high and low pass filters, extracting the wavelet amplitudes for the various 'frequency layers', i.

The description above applies to a dyadic wavelet decomposition. In a dyadic decomposition each frequency layer defined by the i index spans twice the frequency range of the previous one, $i-1$. Each layer also contains twice as many wavelets spread out in time as the previous, as can be inferred from the summation of equation 7.13.

This dyadic tiling scheme is shown in in figure 7.1. The dyadic decomposition has the advantage that all of the wavelets have the same time domain form (modulo dilation and translation).

By modifying the Mallat's pyramid algorithm it is possible to create other decomposition schemes. This is accomplished by applying the high and low pass filters in a different order. One common choice is the binary decomposition. In a binary wavelet decomposition one can achieve regular tilings of the time-frequency plane, where each wavelet has the same rectangular shape. In the time domain the binary wavelets do not have the same functional form. Although at each layer, i, they are the same modulo translations. A binary decomposition is also shown in figure 7.1

A particular binary decomposition is further defined by it's 'decomposition depth', which characterizes the actual modification of Mallat's pyramid. The decomposition depth determines the frequency coverage or 'resolution' of each wavelet layer. For different decomposition depths the wavelet pixels of the basis have different shapes in the time-frequency plane.

An earlier version of BayesWave used a dyadic wavelet decomposition, but we have found the regular tiling of the binary decomposition to be advantageous. In the work that follows we will use a binary wavelet decomposition. The decomposition depth will be chosen for the particular data in question. In the future it may even become one of BayesWave's model parameters. BayesWave uses the same wavelet family as cWB, the Meyer wavelets. Meyer wavelets use 1024 generating coefficients, permitting

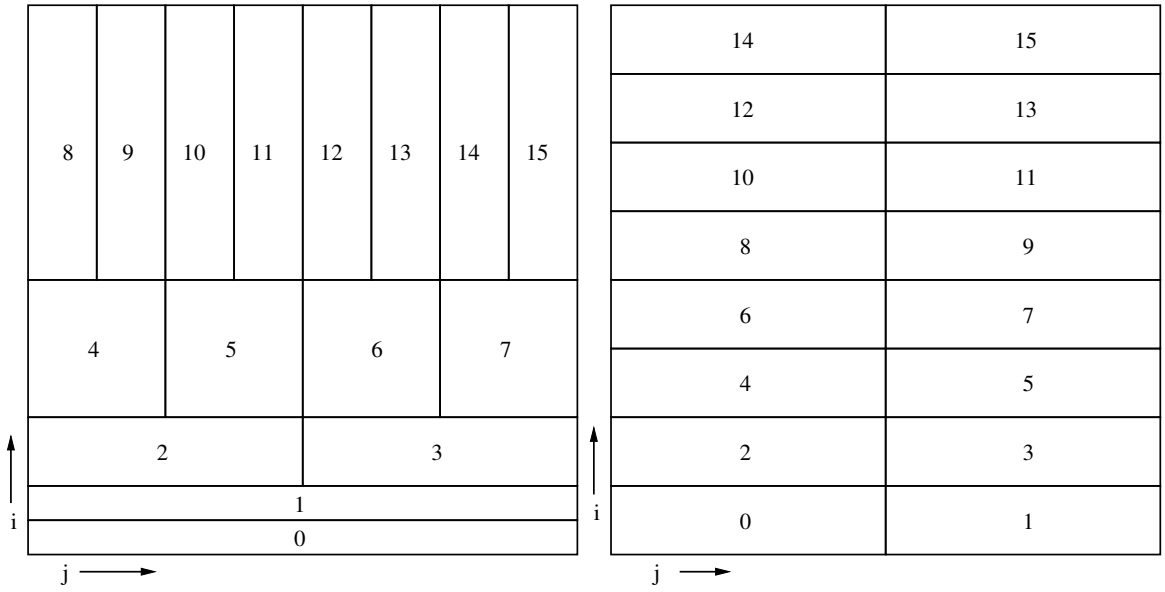

Figure 7.1. A dyadic and binary tiling of a time frequency plane. The left panel is dyadic, the right binary. In each case the all wavelets cover equal area. The numbering of the wavelet pixels is the k index. For the dyadic tiling $k = 2^i + j$, and for the binary $k = \ell * j + i$, where ℓ is the number of wavelet pixels in each layer. In this particular binary decomposition the tiling is the same as the dyadic $j = 1$ layer. In general any dyadic layer could be chosen, giving different tilings.

much less spectral leakage than the historically important Haar family (two generating coefficients) or the commonly used Symlet family (60 generating coefficients) [35].

BayesWave operates entirely in the wavelet domain. There is only ever one wavelet transform computed, when the input data is transformed into a binary Meyer wavelet decomposition. After BayesWave finishes, we can optionally compute time domain reconstructions of the best fit gravitational wave signal or glitches. By limiting the number of wavelet transforms to the bare minimum, we can decrease the computational cost of our analysis.

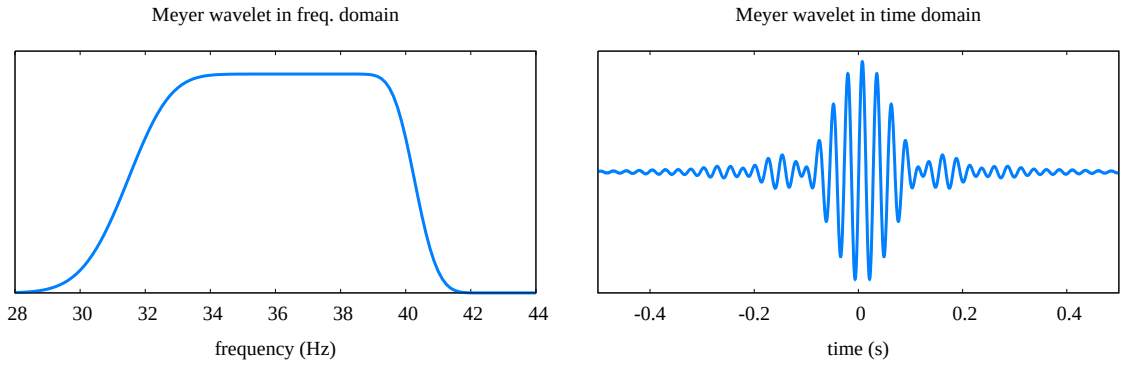

Figure 7.2. A Meyer wavelet in the frequency and time domain. This particular wavelet decomposition has 8 Hz resolution. This wavelet nominally covers the range 32-40 Hz.

Working in the wavelet domain, we hope to capture most of a signal or a glitch with a small number of wavelet pixels. The choice of wavelet decomposition level will be driven by this fact. If the types of events we wish to detect are short duration and broad band, then we will want to use wavelets that mimic these properties. The actual gravitational wave and glitch model parameters used by BayesWave are the number of wavelet 'pixels' in a fit and the amplitudes and k or equivalently i, j indices of the used pixels.

7.2. McMC Implemented

To implement a RJMcMC for BayesWave, we will need to be able to compute the likelihood for our model as a function of the model parameters and our prior belief for each combination of model parameters. In addition we will need proposal distributions to randomly choose the next state in the Markov chain. At every step

in the Markov chain we choose an new state and compute the Hastings ratio from equation 6.18:

$$H = \frac{\mathcal{P}_j \, q_{ji}}{\mathcal{P}_i \, q_{ij}} \det\left(J_{ij}\right). \tag{6.18}$$

We accept this move with probability $\min(1, H)$. If the move is not accepted, then the Markov chain stays in its current state, and we try again. Computing the Hastings ratio is of central importance to any McMC. To compute the unnormalized posterior we will need the likelihood and prior. It will prove convenient to work in the log of these quantities.

The likelihood, $p\left(d \,|\, \mathcal{H}, \mathcal{I}\right)$, is the probability that the data would be observed if our hypothesis is true. Our hypothesis is that the wavelet domain strain from each detector $s_i(k)$ is described by our model:

$$s_i(k) = \mathcal{A}_i \, h(k) + g_i(k) + n_i(k) \tag{7.14}$$

If h and g are good fits to the actual gravitational wave signal and glitches, then the noise, n, will be Gaussian distributed. Our likelihood will use this fact. Consider the probability that one point of data, d, was drawn from a Gaussian distribution with zero mean and variance σ:

$$p\left(d \,|\, \text{Gaussian}\right) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-d^2}{2\sigma^2}\right). \tag{7.15}$$

The probability that a set of n data points $\{d\}$ were drawn from the same Gaussian distribution is:

$$p\left(\{d\} \,|\, \text{Gaussian}\right) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left(\sum_{i=1}^{n} \frac{-{d_i}^2}{2\sigma^2}\right). \tag{7.16}$$

In principle we could draw these data from several Gaussian distributions, each with a different variance. In this case each d_i would have an associated σ_i. Our likelihood is then the probability that the noise was drawn from the piece-wise, colored Gaussian distribution described by the noise model. The noise of our model is defined by the residual:

$$r_i(k) = s_i(k) - \mathcal{A}_i\, h(k) - g_i(k), \tag{7.17}$$

what is left over after subtracting the signal and glitches from the data. The log likelihood for each detector is:

$$\log\left(p\left(s_i(k)\,|\,\mathcal{H},\mathcal{I}\right)\right) = \log L = \sum_k \log\left(\frac{1}{\sqrt{2\pi}\,\sigma_k}\right) + \sum_k \frac{-r_i(k)^2}{2{\sigma_k}^2}, \tag{7.18}$$

and the total log likelihood is the sum of the likelihoods for each detector. The second sum in equation 7.18 can be though of as a wavelet domain noise weighted inner product:

$$\sum_k \frac{-r_i(k)^2}{2{\sigma_k}^2} = \langle r_i \,|\, r_i \rangle\,, \tag{7.19}$$

In practice only a small subset of the total pixels will ever change at a time, so we can save computational cost by computing only the change in the likelihood, $\Delta \log L$, instead of the full likelihood. The noise fit is at the same location then we can compute, $\Delta \log L$, considering a change in the glitch model in one detector, $\Delta g_i(k)$, for some subset of pixels, k, that changed:

$$\begin{aligned}\Delta \log L &= \langle \Delta g_i \,|\, r_i \rangle - \tfrac{1}{2}\langle \Delta g_i \,|\, \Delta g_i \rangle\,,\\ &= \sum_{k\in\text{changed}} \frac{1}{2\sigma_k}\big(2\,\Delta g_i(k) r_i(k) - \Delta g_i(k)\Delta g_i(k)\big),\end{aligned} \tag{7.20}$$

where $r_i = s_i - \mathcal{A}_i h - g_i$ is the original residual in the i^{th} detector. Equivalently, we can compute the change in log likelihood for changes in the signal model by replacing Δg_i with $\Delta \mathcal{A}_i h$.

If the noise fit does change, then each σ_k could potentially change. In this case we will need to compute the full log likelihood at much greater computational cost. To take advantage of the available savings we will compartmentalize our proposals. Only attempting to change one fit: noise, glitch or signal, at each step of the McMC.

The other half of our posterior is the prior belief in a parameter space location. We need to ask ourselves what sorts of properties do we expect gravitational wave bursts and glitches to have. We will use the same model for both, with the only distinguishing characteristic between gravitational waves and glitches being the gravitational waves' network coherence. Because we treat both as the same, in the discussion that follows the term 'signal' will stand in for 'gravitational wave signal and/or glitch'. We can think of both gravitational waves and glitches as two different types of signal, each worthy of detection and distinct from the Gaussian detector noise.

Our signal model is fundamentally defined by the number of non-zero wavelet coefficients, the number of pixels in the model. Each model dimension acts as an independent hypothesis. A bit of combinatorics tells us that the number of possible states in each dimension is not the same. That is to say, the prior volume is a function of the model dimension. For data of length N, there are $\binom{N}{d}$ states for model dimension d. The default 'ignorance' prior would state: *each individual state is equally*

likely. This places an effective prior on the model dimension: *model dimensions with more states are preferred*. Not knowing anything about the presence or lack-thereof of signals, we would like to have a prior that states: *each model dimension is equally likely*. We do this by applying a prior on the model dimension that scales with the volume of that dimension:

$$P_{\mathrm{dim}}(d) = \binom{N}{d}^{-1}. \tag{7.21}$$

As it is the prior ratio that appears in the Hastings ratio, we will choose to compute these directly for the case of moving up or down one dimension, rather than computing large factorials at each step:

$$\frac{P_{\mathrm{dim}}(d+1)}{P_{\mathrm{dim}}(d)} = \frac{d+1}{N-d}, \qquad \frac{P_{\mathrm{dim}}(d-1)}{P_{\mathrm{dim}}(d)} = \frac{N-d+1}{d}. \tag{7.22}$$

Another important prior is on the amplitude of a signal. The individual detectors are more or less sensitive to incident gravitational waves depending upon their point of origin. This dependence includes some locations where the detector is completely 'blind', having no response at all. Even for a network there can be sky locations where all detectors have very low response. For these sky locations, it would be consistent with our model to have incident gravitational waves of very large amplitude producing little to no effect in the detectors.

Because high amplitude gravitational waves are very rare, it is unlikely that such occurrences would happen. Further if the network is not sensitive to these events, it could not detect them even if they did occur. To prevent our analysis from 'detecting' loud gravitational waves in our detector blind spots, we will place a prior on the

amplitude of incident signals. This prior also makes sense for glitches. There are very loud glitches in LIGO, but they are not particularly common. Further when noise events get to high enough amplitude, they knock the detectors out of 'lock' (see section 3.1). Any glitches in a continuous analysis segment must have lower amplitude than this.

We do not want to place a hard threshold on the amplitude of signals. We simply want to discourage high amplitude signals that have no impact on the likelihood. If there really is a high amplitude signal somewhere, the data should be able to overcome the prior by showing a large improvement in the likelihood. Fitting a real high amplitude signal will remove a large amount of power from the residual, making it much more Gaussian.

To this end we may apply a 'Gaussian amplitude prior'. We assume the amplitudes of the signals should be roughly Gaussian distributed. While this is not strictly true, the effect is that we have more prior belief in low amplitude signals than high amplitude signals, consistent with our assumption that high amplitude signals are rare. cWB deals with this same problem of detector blind spots by using regularization conditions on its matrix inversions. These regularizations are precisely equivalent to a Gaussian amplitude prior [50].

It is convenient to set a prior on the SNR of a single pixel where our variable, $x(k) = |a(k)/\sigma_k|$, is the absolute value of the ratio of the wavelet amplitude to the noise level (as estimated by the noise model). We expect our signals to be louder

than the noise, so we can set the width of the Gaussian distribution on the SNR to be larger than one. The log prior on the wavelet SNR, $x(k)$, is then:

$$\log P_{\rm amp}(x) = \sum_k \log\left(\frac{1}{\sqrt{2\pi}\,\kappa}\right) + \sum_k \frac{-x(k)^2}{2\kappa^2}, \tag{7.23}$$

where κ is the variance of the prior, determining how many times louder than the noise we will allow a signal to be. For a choice of $\kappa = 5$, a single pixel with SNR of 10 would be a somewhat common 2σ event.

We could choose to express this prior in terms of a correlation matrix, $\mathbf{C}$. If there are N pixels in our signal model, then in this form the prior is:

$$\log P_{\rm amp}(x) = -\tfrac{N}{2}\log(2\pi) - \tfrac{1}{2}\log(\det \mathbf{C}) - \tfrac{1}{2}\vec{x}\cdot\mathbf{C}^{-1}\cdot\vec{x}, \tag{7.24}$$

where the inverse of $\mathbf{C}$ is dotted into the wavelet SNR vector for all N pixels of the signal model. For the prior of equation 7.23 the correlation matrix, $\mathbf{C}$, is simply a scaled identity matrix $\mathbf{C} = \kappa^2\mathbf{I}$. We will come back to this form of the amplitude prior, as we further constrain our signal model.

We do not expect to find dozens of gravitational waves, or for that matter dozens of glitches, in a single analysis time. So we want to place a prior on the number of signals we find. We identify a particular signal as a 'cluster' of pixels. A cluster is defined as one or more pixels that share a boundary point on the wavelet grid. These could be left/right, up/down, or diagonal neighbors giving 8-point connectivity. We identify clusters using a standard connected-component labeling algorithm [48]. Figure 7.3 shows three example clusters, including a lone pixel.

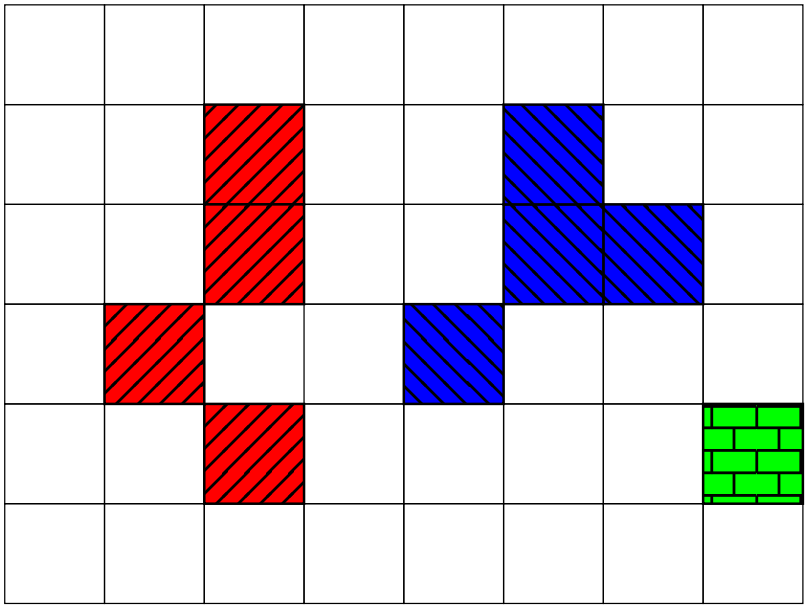

Figure 7.3. A wavelet signal model with 3 clusters. Cluster membership includes direct up/down and left/right neighbors as well as diagonal neighbors. There are several places where adding a new pixel could cause two clusters to merge, reducing the total number of clusters. There are also ways to remove an existing pixel and split a cluster in two, raising the total.

To limit the number of clusters, K, we provide the following penalty:

$$P_{\text{clust}}(K) \propto \alpha^{-K}, \tag{7.25}$$

where α is a tunable parameter. By choosing α we can strengthen or weaken the penalty for adding a new cluster. Consider a proposal to add a new cluster (by adding a lone pixel). The gain in log likelihood will be given by the SNR this new pixel fits, $\Delta \log L = (\text{SNR})^2/2$. We can choose an α, such that the prior penalty exactly cancels this gain in likelihood:

$$\Delta \log L = (\text{SNR})^2/2 = \log \alpha = -\Delta \log P \tag{7.26}$$

In this way we can set a *de facto* limit on the SNR fit by a new pixel. We will choose $\alpha = e^2 \approx 8$, so each new pixel should fit an SNR of at least 2 or it is penalized. Of course we can add pixels to existing clusters with no penalty regardless of the fit.

A cluster is in effect a new fundamental unit of our model. Since a cluster represents a single signal, the pixels that make up each cluster should be related in some way. This brings us back to the amplitude prior of equation 7.24. Instead of setting prior on the individual pixel amplitudes, we want to set a prior on the pixel amplitudes on a cluster by cluster basis.

First let us consider the parameter κ, which determines how much louder than the noise we expect a signal to be. Since we expect the amplitudes of all members of a cluster to be related in some way, we will let κ be a model parameter for each cluster, m. This means there will be several $\kappa_{(m)}$s, with each cluster actively deciding how much louder than the noise it needs to be by adjusting its $\kappa_{(m)}$.

Beyond this first level of intra-cluster relations, we can choose the correlation matrix, $\mathbf{C}$, to be non-diagonal. We will build a correlation matrix that enforces a degree of correlation between neighboring pixels. We will say every pixel should be correlated to its direct up/down and left/right neighbors to some degree, and it should be correlated to its diagonal neighbors to a lesser degree. These relations are neatly summarized by the Gaussian smoothing kernel:

$$\mathcal{K}_{\mathcal{G}} = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1/4 \end{pmatrix} \tag{7.27}$$

We will construct a correlation matrix using this kernel as a model:

$$C_{ij} = \kappa^2 \times \begin{cases} 1, & \text{if } i = j \\ 1/2, & \text{if } i \text{ up/down } j \\ 1/2, & \text{if } i \text{ left/right } j \\ 1/4, & \text{if } i \text{ diagonal } j \end{cases} \tag{7.28}$$

If we label the pixels in each cluster of figure 7.3 starting from the top left the three correlation matrices will be:

$$C_{ij}^{(\text{red})} = {\kappa_{(\text{red})}}^2 \begin{pmatrix} 1 & 1/2 & 0 & 0 \\ 1/2 & 1 & 1/4 & 0 \\ 0 & 1/4 & 1 & 1/4 \\ 0 & 0 & 1/4 & 1 \end{pmatrix}, \tag{7.29}$$

$$C_{ij}^{(\text{blue})} = {\kappa_{(\text{blue})}}^2 \begin{pmatrix} 1 & 1/2 & 1/4 & 0 \\ 1/2 & 1 & 1/2 & 1/4 \\ 1/4 & 1/2 & 1 & 0 \\ 0 & 1/4 & 0 & 1 \end{pmatrix}, \quad C_{ij}^{(\text{green})} = {\kappa_{(\text{green})}}^2. \tag{7.30}$$

The amplitude prior now must be computed differently for each cluster, m, with a different correlation matrix dotted into the vector of pixel SNRs in that cluster only, $\boldsymbol{x}^{(m)}$:

$$\log P_{\text{amp}}(x) = -\tfrac{N}{2}\log(2\pi) - \sum_{m \in \text{clust}} \left(\tfrac{1}{2}\log(\det \mathbf{C}^{(m)}) + \tfrac{1}{2}\boldsymbol{x}^{(m)} \cdot {\mathbf{C}^{(m)}}^{-1} \cdot \boldsymbol{x}^{(m)} \right), \tag{7.31}$$

Because $\kappa_{(m)}$ is now a model parameter, we need to place a prior on it. We don't expect our signals to be quieter than the noise, so we will place a lower bound of $\kappa_{(m)} \geq 1$. This parameter is enforcing our amplitude prior that signals should not be arbitrarily loud, so $\kappa_{(m)}$ should not be free to grow arbitrarily. We place a half Gaussian prior on the $\kappa_{(m)}$s with mean of 1 and variance ς:

$$\log P_\kappa = \sum_{m=1}^{K} \begin{cases} -\infty, & \kappa_{(m)} < 1 \\ -\log(2\pi\,\varsigma^2) - \frac{\left(\kappa_{(m)}-1\right)^2}{2\varsigma^2}, & \kappa_{(m)} \geq 1 \end{cases}, \tag{7.32}$$

where K is the number of clusters. Choosing $\varsigma = 5$ allows plenty of room for the $\kappa_{(m)}$s and therefore the pixel SNRs, $x(k)$, to adjust.

We can compare the impact of this correlated signal prior relative to the signal prior with a diagonal correlation matrix. First we shall consider the Akaike Information Criterion (AIC). The AIC measures the relative quality of a statistical model by considering both the model fit and the model complexity [4]. The model fit is given by the log likelihood and the complexity by the number of degrees of freedom. Two models are of equal quality if their AIC is the same:

$$\mathrm{AIC} = 2D - 2\log L, \tag{7.33}$$

where D is the model dimension. Because $\log L$ is proportional to χ^2, the AIC is closely related to the frequentist χ^2 *per degree of freedom* statements. For BayesWave the model dimension is effectively the number of wavelet pixels in a fit. Each wavelet pixel actually adds a bit more than $D = 1$, because the choice of which pixel k adds a partial degree of freedom. We will ignore this fact for a simplified discussion.

According to the AIC, if a model of higher dimension is to be equivalent to one of lower dimension, the higher dimension model must provide an improvement to the likelihood. For each new pixel we add, we should gain $\Delta \log L \geq 1$ to get an equal or better quality model. This can be thought of as an 'Occam penalty' where the simplest model that provides an adequate fit is preferred. This feature is a natural effect of the RJMcMC algorithm. The Hastings ratio depends on both the likelihood and the prior, so we can overcome this penalty by improving either the model fit or our prior belief in the model.

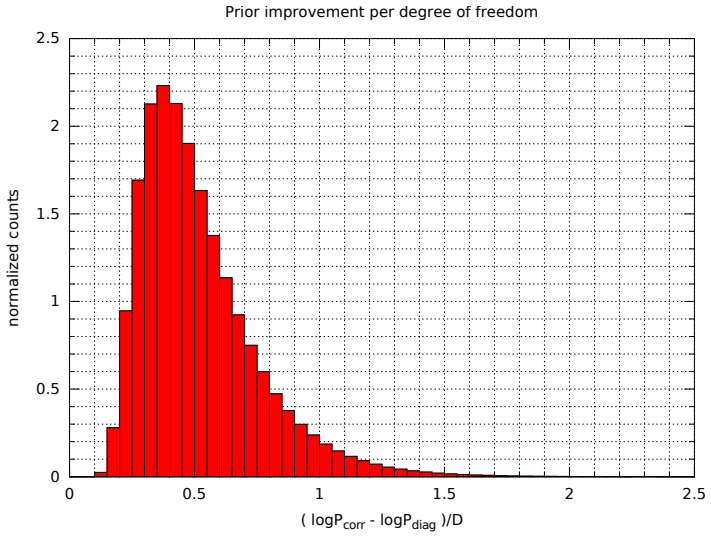

Figure 7.4. Distribution of log prior increase using a correlated signal prior. The prior gain is peaked around 0.4 per pixel used in the model. This provides about half of the 'Occam penalty' defined by the AIC.

In figure 7.4 we compare the difference in log prior per degree of freedom for the correlated and diagonal priors for a 5×10^6 step RJMcMC on a simulated signal. The correlated signal prior provides a small improvement. In many instances the prior provides about half of the 'Occam penalty', meaning the improvement in fit, $\Delta \log L$, only needs to be about 0.5 to favor a move that increases the model dimension. In less than 5% of the samples is the prior improvement greater than 1. In these instances a new pixel can be added with no improvement to the fit.

In addition to this choice of correlation structure we implemented several others. We found that matrices built with other correlation structures could become singular

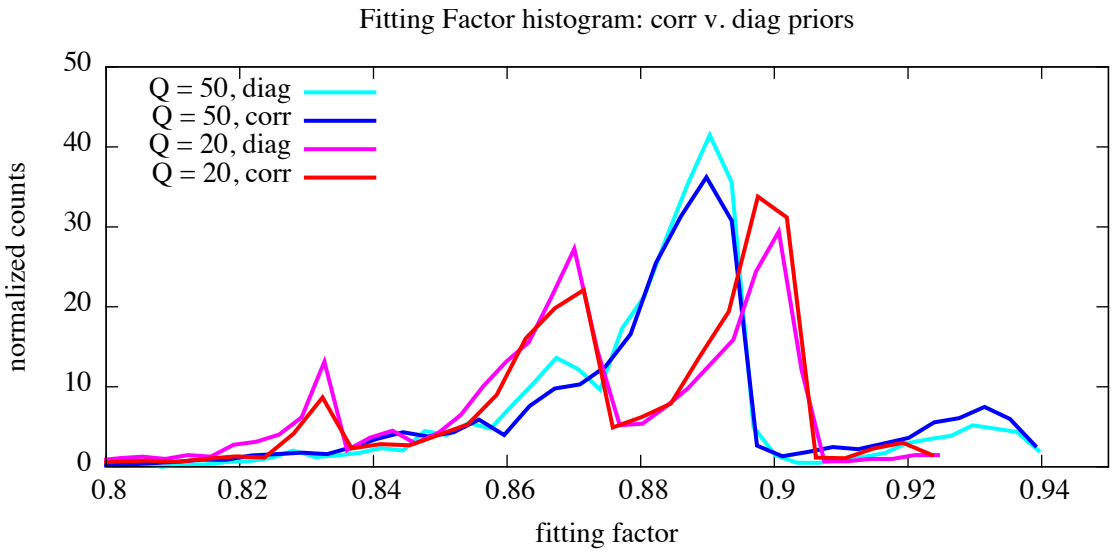

Figure 7.5. Comparison of correlated and diagonal amplitude priors for two different simulated signals. The correlated prior gives a small increase over diagonal prior. The multimodal structure is a result of the addition or omission of individual pixels in the fit.

or non-positive in certain instances where the Gaussian smoothing motivated correlation did not. We will leave it to future work to determine if there is a criterion for a well behaved correlated prior.

We also tried several *ad hoc* signal priors, in an attempt to favor clusters of pixels over lone pixels. In all cases the improvement in tangible parameters, like fitting factor (see chapter 8), was of order a few percent. Because all choices were functionally equivalent, we choose to use the correlation prior, as it is well motivated theoretical and simply posed: *clusters of pixels with correlated amplitudes are preferred.*

In figure 7.5 we compare the correlated and diagonal amplitude priors by examining the posterior fitting factor distribution for two RJMcMC runs: one for each case.

The correlated amplitude prior adds a small amount of posterior weight to the high fitting factor modes, relative to the diagonal. It is not a particularly large gain.

It remains an open question as to what, if any, signal prior can provide significant improvement. It is also worth considering altering the choice of likelihood. The Gaussian residual likelihood only considers independent residual amplitudes. True Gaussian noise should produced uncorrelated amplitudes. Instead of having a prior belief that the signal amplitudes are correlated, we could choose a prior that enforces a locally uncorrelated residual.

We have not yet given the true details of our noise model. The noise model, as mentioned, is a piece-wise fit to colored gaussian noise. Each wavelet layer covers some range of frequencies. This provides a natural boundary for the pieces in our fit. At each frequency layer, i, there is one noise level, σ_i. This noise level is free to grow or shrink to best fit the Gaussian distribution that all ℓ residual amplitudes in that layer are drawn from. The starting location for the noise model is based a wavelet domain power spectrum of the data. We determine the median amplitude of the original data for each wavelet layer, and take that to be the variance of the Gaussian noise, $\bar{\sigma}_i$.

We must place a prior on the noise levels in each frequency layer. This is based on the original $\bar{\sigma}_i$ in each layer, and the number of samples in the data. We assume the noise levels are Gaussian distributed around the original estimate. The more wavelet pixels in our model, the better the original estimate should be, so the variance, ς_{n},

is given by the root-N error of the estimate. We choose to work with a scaled noise level $u_i = \sigma_i/\bar{\sigma}_i$, so the variance is the same for all layers:

$$\log P_{\text{noise}} = -\tfrac{N_\ell}{2} \log(2\pi\, {\varsigma_{\text{n}}}^2) - \sum_{i=1}^{\ell} \frac{(u_i - 1)^2}{2{\varsigma_{\text{n}}}^2}, \tag{7.34}$$

$$\varsigma_{\text{n}} = \ell^{-1/2}, \tag{7.35}$$

where N_ℓ is the number of frequency layers and ℓ is the number of wavelet pixels in each layer.

Now that we have discussed the likelihood and priors for our Hastings ratio all that remains are the proposal distributions, q_{ij}, which tell us how to transition from one state, i, to another, j. We will exploit the detailed balance of our McMC in order to use a whole suite of proposals instead of just one. There will be transitions that change the noise fit. There will be transitions that change the signal model dimension by adding or subtracting pixels. There will be transitions that adjust the signal model parameters of the same dimension, namely the pixel amplitudes and the cluster parameters, κ.

It is important to remember that the choice of proposals can have no effect on the end result of an McMC, as long as all proposals are 'memoryless' and provide 'accessibility'. For a proposal to be memoryless, it must depend on only the current state of the Markov chain. A proposal that is the same regardless of the current state is trivially memoryless. To achieve 'accessibility' it must be possible for the Markov chain to reach any state in the prior volume in a countable number of moves. Most of our priors included arbitrarily large values, although at very low probability. Since all

moves are of finite distance any one of these infinities can be reached in a countable number of moves. This is acceptable, as a finite number of countable infinities is still countable [52].

The most basic proposals, are those to adjust the pixel amplitudes, the noise levels, and the cluster parameters, κ. There are two basic methods to accomplish all of these: proposing a new state as a draw from the prior distribution, or proposing a new state as a small step from the old location. Whenever a new pixel is added or a new cluster is added, we propose its initial amplitude or κ, as a draw from the prior. In this case the proposal distribution, q_ij, is the prior:

$$\log q_{ij} = \log P_j. \tag{7.36}$$

When we want to explore a region of the available space more finely, it makes sense for the new proposed state to depend on the current state. In these cases we choose to move a parameter by Gaussian draw. We define a Gaussian draw as:

$$\mathcal{G}(\bar{x}, \sigma), \tag{7.37}$$

a number drawn from a Gaussian distribution of mean $\bar{x}$ and variance σ. For instance, if we want to propose an adjustment to a noise level, u, we can try a new state that is about a one standard deviation move from the current state. The noise variance is given by ς_{n} so the proposed new noise level, w is:

$$w = u + \mathcal{G}(u, \varsigma_{\mathrm{n}}), \tag{7.38}$$

We could also propose to change several noise levels simultaneously. In this case the variance of the Gaussian draw is scaled by the number of changing parameters, so the net result is a change of about one standard deviation:

$$w_i = u_i + \mathcal{G}\left(u_i, \frac{\varsigma_n}{N}\right). \tag{7.39}$$

Because the variance is the same for all points in the parameter space the proposal distribution to attempt a transition to the new state is the same as the reverse, proposing to transition from the new to the old. The ratio of proposal distributions which appears in the Hastings ratio is unity.

In practice we will randomly choose to propose a new state for one of these three parameters: amplitude, noise level, or cluster parameter. Then we will select a random subset from the total number and propose new locations for those as in equation 7.39. In this way we propose to make small changes to some of these parameters at each step of our McMC. After the proposal we can compute the likelihood and priors, at the original and new state necessary for the Hastings ratio, H. Finally, one last random draw determines if the proposed move is accepted with probability $\min(1, H)$.

We can also propose changes of states that add or subtract a pixel from a particular signal model. These proposal change the model dimension, but they will hold all of the existent pixels at their previous values. The Jacobian in the RJMcMC Hastings ratio, equation 6.18 is unity for all of these proposals.

The simplest way to add a new pixel is by a uniform draw of all unused pixels. The proposal distribution is then given by the number of unused pixels, N_{off}:

$$q_{ij} = \frac{1}{N_{\text{off}}}. \tag{7.40}$$

The new pixel will have an amplitude given by a prior draw. These proposals are useful for exploring the full parameter space, but will not be accepted very often. In a large analysis period with many potential pixels, but only a few glitches or gravitational waves it is unlikely that any given proposed pixel will happen to land on a signal. Even if the new pixel is part of a signal, it must also match the signal's amplitude.

The reverse of this uniform pixel addition, is a uniform pixel subtraction. We propose to remove a pixel via a uniform draw of the current pixels in the model. The proposal distribution is simply given by the number of used pixels, N_{on}:

$$q_{ij} = \frac{1}{N_{\text{on}}}. \tag{7.41}$$

This uniform method of subtraction is again inefficient. It will try to remove pixels that are good fits to a signal as often as those which are not.

In each of these cases there are a few complications when considering clusters of pixels. When new pixels are added they can create a new cluster or join an existing cluster. In some cases a new pixel can span a gap, connecting two (or more) previous disparate clusters into one new cluster. When a pixel is removed from the model it could cause a schism in a single cluster, creating two (or more) new clusters in the

process. In each of these cases we must handle the cluster parameter κ properly to ensure that any one cluster is described by only one κ. Further we must cover these cases in a way that preserves detailed balance in our McMC.

There are more efficient ways to add and subtract pixels. Since we expect signals to form clusters of pixels we can propose new pixels to the edges of existing clusters more often than elsewhere. Similarly we could propose to remove isolated pixels more often than those in larger clusters. Both of these situations can be achieved by re-weigthing the proposal distributions. For instance we can choose a distribution where an unoccupied pixel is more likely to be proposed to add, if some of it's neighboring pixels are already in the model. In the reverse of this we are less likely to propose to remove a pixel with neighbors in the model.

We could also use the input data itself to construct a proposal distribution. For each point in the data $s(k)$ we can compute $s^2(k)/\sigma_k$, using the current noise model. We can treat this as a proposal distribution for adding new pixels. In this way we would be 25 times more likely to add a pixels with SNR of 5 than a pixel at the noise level with SNR of 1. In practice it is good to place a lower and upper bounds on this distribution. For instance if a pixel had an amplitude of exactly zero, this proposal would not provide accessibility. To propose pixels to remove, we can use the point-wise inverse of this distribution, making us unlikely to try to remove a very loud pixel.

It is worth reiterating that the choice of proposals has no effect on the outcome of our analysis. The McMC draws samples the posterior distribution function no matter how we propose transitions to new states. By selecting proposals that successfully transition more often, we can achieve more samples in less time. The choice of proposal distributions affects the convergence time of an McMC, not the end result. The convergence time is a measure of how many samples we need to accurately represent the posterior. Usually we say an McMC has converged after it reaches a 'steady state'. That is to say when subsets of samples from different parts of the McMC are statistically equivalent. For example the samples from two subsets should have the same mean likelihood.

8. APPLYING BAYESWAVE TO GRAVITATIONAL WAVE DATA

After discussing the BayesWave algorithm in the last chapter, we are finally ready to see how it functions in practice. There are two quantities of merit that we will use to quantify the performance of BayesWave . The first is the Bayes factor. We will compute the Bayes factor for detection of a signal, $\mathcal{B}$, by comparing the number of McMC samples from the signal model with no pixels, N_0, to the number of samples in all other models $N_{\rm tot} - N_0$:

$$\mathcal{B} = \frac{N_{\rm tot} - N_0}{N_0}$$

If there are many samples from models with any pixels the Bayes factor will be large, and we can say that BayesWave has detected a signal. The lager the Bayes factor the more confident we are in detection. We will say that $\mathcal{B} \geq 10^3$ is a reasonable level for detection. In some cases we will find that BayesWave takes no samples from the model with no pixels, $N_0 = 0$. In this case we set a lower limit on the Bayes factor of $\mathcal{B} \geq N_{\rm tot} - N_0$.

The next quantity of merit is the fitting factor. The fitting factor tells us how well a model fits some known waveform. In the cases where we inject simulated signals into data, the fitting factor measures what fraction of the injected signal-to-noise ratio (SNR) BayesWave recovered in a model. The fitting factor is defined:

$$\mathrm{FF} = \frac{\langle h_{\rm inj} \,|\, h_{\rm rec} \rangle}{\sqrt{\langle h_{\rm inj} \,|\, h_{\rm inj} \rangle \, \langle h_{\rm rec} \,|\, h_{\rm rec} \rangle}}, \tag{8.1}$$

where $h_{\rm inj}$ is the injected waveform and $h_{\rm rec}$ is the recovered waveform. Because of the normalization, the fitting factor decreases when there is extraneous power in the recovered signal. We can compute the fitting factor at each step of the McMC using the current model as the recovered waveform. We can also compute the fitting factor for the *maxima a posteriori* (MAP) or best fit parameters.

8.1. BayesWave as a Data Quality Tool

As a first testing ground, we will examine the detectability of signal software injections into a single detector. In this case it is impossible for BayesWave to distinguish between coherent signal and independent glitches. We will restrict ourselves by only fitting for non-Gaussian glitches and colored Gaussian noise. All of these same factors apply when fitting for signals as well.

We will begin with an exploration into the detectability of a single loud pixel. BayesWave is designed to find clusters of loud pixels, but in order to construct a cluster we must begin with a single pixel. By studying the detection of single pixels, we can discover the SNR per pixel threshold required by BayesWave to find any feature in the data. When we inject a single loud pixel, we can choose the true SNR of that injection. BayesWave in order to determine the SNR of a pixel, must first establish the noise level for that particular wavelet frequency layer. If the noise estimate is wrong, then BayesWave may not be able to distinguish our injection from the Gaussian noise background.

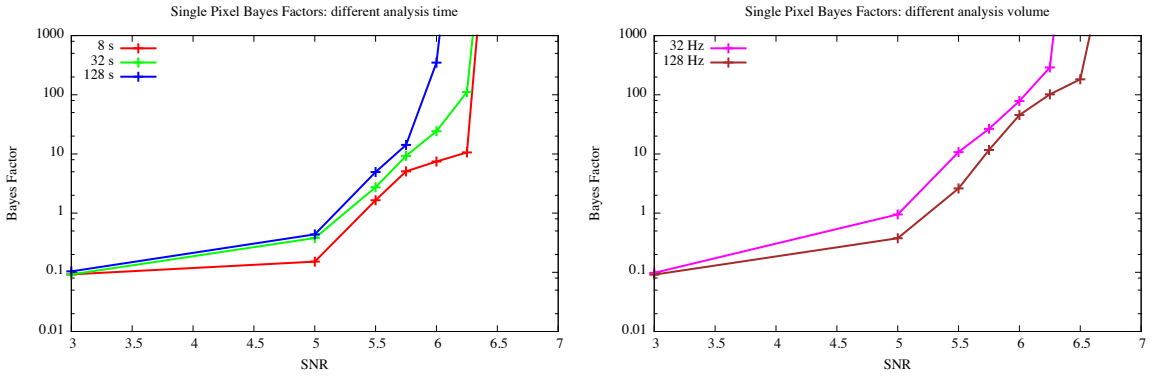

Figure 8.1. Detectablity of a single loud pixel by BayesWave. In the left panel we examine how extending the analysis time, while holding the total time-frequency volume fixed, affects detection. By extending the analysis for longer times, we are able to get a better fit to the colored, Gaussian noise, translating to a lower SNR threshold for detection. In the right panel we examine the effect of time-frequency volume on detection, holding the analysis time fixed and changing the analyzed frequency range.

Consider a wavelet decomposition with 8 Hz frequency resolution. The wavelet pixels are then spaced by $\frac{1}{16}$ s. If we choose to analyze 8 s of data, there are only 128 wavelet coefficients in each frequency layer. To estimate the noise level, BayesWave is fitting a Gaussian to only 128 data points. If one pixel is slightly above the expected noise level, it is easier to adjust the Gaussian noise level than to add a glitch model. There is not enough information in this short duration of data to establish accurate noise estimation. This effect can be seen in figure 8.1. By increasing the analysis time, there are more 'time samples' per frequency layer. More data points leads to a more accurate fit to the underlying noise level and, therefore, a lower threshold for detection. We demonstrate this effect in figure 8.1.

We cannot just allow BayesWave to run on an arbitrary long segment of data. If the noise is non-stationary over a long segment of data (as is the case with LIGO),

our noise estimate will be skewed by the fluctuating levels and be inaccurate. On the practical side of things, the BayesWave run time is a function of the number of pixels in the wavelet grid. By adding more analysis time, it takes BayesWave longer and longer to arrive at an answer. In the future we will outsource the noise fitting from BayesWave to a separate utility. This utility will fit the noise over a longer period of time using more detailed methods than BayesWave. The data can then be whitened with this fit, and passed to BayesWave for glitch fitting in several smaller time intervals.

The analyzed time-frequency volume also affects detectability. For a large volume we are more likely to see a loud pixel somewhere in the volume by chance, lowering the significance of any loud pixels actually seen. This effect is demonstrated in figure 8.1. When computing a Bayes factor for detection purposes, we must account for this effect. We effectively have a uniform prior on the time and frequency of a signal's appearance in the data. The Bayes factor then carries with it a factor of $\frac{1}{V_\mathrm{a}}$, where V_a is the analysis volume. To compare two Bayes factors from different analyses we must rescale them by the total analyzed volume. This prior volume scaling is equivalent to the frequentist 'trials factor' and the 'look-elsewhere' effect. For instance in the right panel of figure 8.1 the 32 Hz analysis window is a factor of 4 smaller than 128 Hz analysis. When the Bayes factors for the 32 Hz analyses are scaled by $\frac{1}{4}$, they agree with those computed from the 128 Hz analyses.

As a more realistic exercise in detectability, we will inject sine-Gaussian glitches. In the time domain our sine-Gaussian (actually a cosine-Gaussain) glitch is given by:

$$A(t) = A_0 \exp\left(-\bar{t}^2/2Q^2\right) \cos\left(\bar{t} + \phi_0\right) \tag{8.2}$$

$$\bar{t} = 2\pi\, f_0\, \left(t - t_0\right)$$

where A_0 is an amplitude factor, f_0 is the central frequency, t_0 is the peak time, ϕ_0 is the starting phase and Q is the quality factor. We will tune the amplitude of an injection to achieve a desired signal to noise ratio.

While a sine-Gaussian is relatively compact in both time and frequency, it will still take several wavelet pixels to capture the entire waveform. If the injection's power is divided among too many pixels, then the SNR per pixel will be low. In this way a 'loud' signal will not be detectable. BayesWave will not be able to find a starting point to latch onto and begin building a cluster. It is possible that by moving to a different wavelet basis where the individual pixels have a different shape, we can better fit the particular sine-Gaussian with fewer pixels. Therefore, we can characterize the SNR per pixel by a combination of the sine-Gaussian quality factor and the time-frequency resolution of the wavelet basis.

We will begin exploring this at a fixed wavelet resolution and by varying the sine-Gaussian quality factor. For this case we will use wavelet pixels that are 8 Hz 'tall' and separated by $\frac{1}{16}$ s. These wavelets have most of their power concentrated in about $\frac{1}{4}$ s, but their low amplitude tails reach out for more than 1 s. For the low quality sine gaussian injections are very compact in time, so by the Fourier uncertainty principle

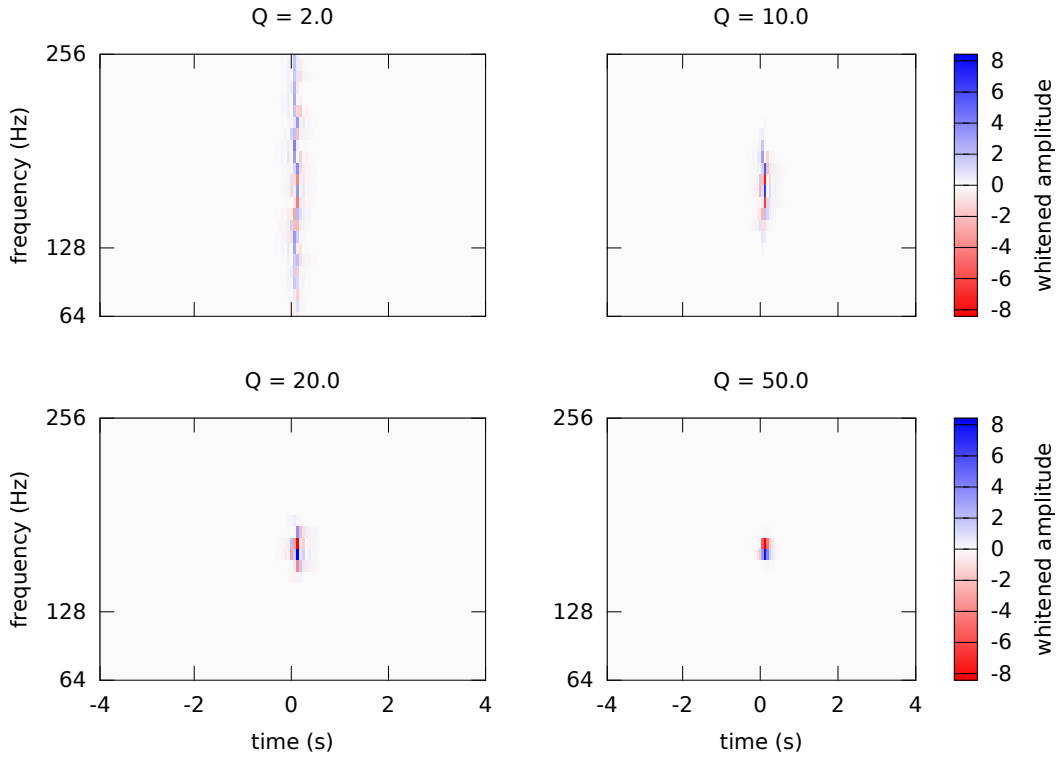

Figure 8.2. Time frequency plot of four sine-Gaussians with varying quality factors. All are in the same wavelet basis with 8 Hz frequency resolution. The low Q sine-Gaussians spread across many wavelet pixels, whereas those with high Q are compact in this wavelet basis.

they span a broader frequency band than their high quality brethren. We can see this in the spectrograms of figure 8.2. This broad band spread means the SNR per pixel wavelet pixel will be low. This intuition is confirmed in figure 8.3, where we see the distributions of SNR per pixel for our sine-Gaussians of differing quality factor. Also in figure 8.3 is the cumulative fractional power for our four sine-Gaussians. We see that we can capture 80% of the power for high quality signals using only a few pixels.

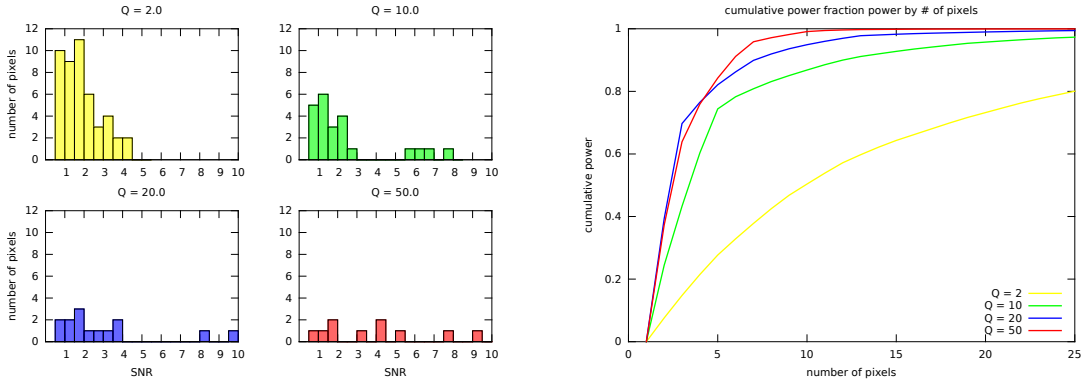

Figure 8.3. SNR per pixel and cumulative fractional power for varying sine-Gaussian quality factor. The left plot is based on a SNR 15 injection.

Finally, we can see the impact sine-Gaussian quality factor has on detection. The more compact signals are easily detectable as seen by their Bayes factors in figureQbayes. As detection is established, BayesWave is able to recover 80% of the compact signals as seen in figure 8.4. For more broad signals BayesWave leaves a good deal of power behind, even with a confident detection. This study of sine-Gaussian quality factor should be seen as a lesson not for these waveforms alone, but for BayesWave's ability to detect any signal. The compactness of a signal in the analysis basis plays a key role in detection of any signal.

In the case of the template based searches that we discussed in section 4.1, we tried to choose basis functions that exactly matched the expected signal. Our wavelets act just like templates if a signal can be captured using only one pixel. The fewer wavelets we need to represent a signal, the more our un-modeled burst search acts like a template based search.

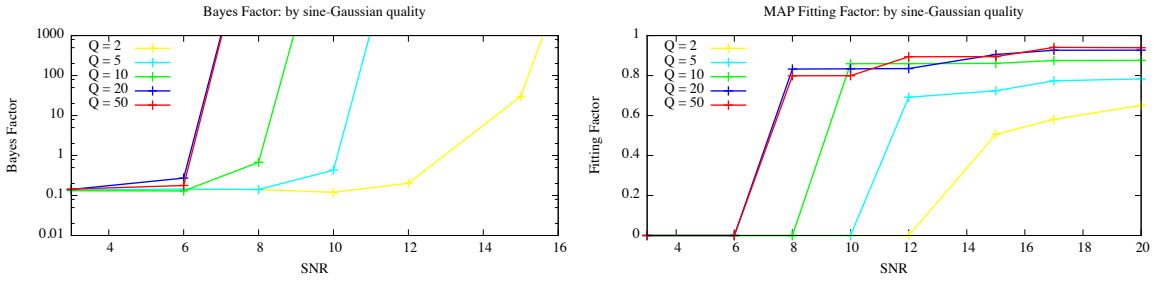

Figure 8.4. Bayes factors and MAP fitting factors as a function of injected sine-Gaussian SNR. We see that when the injected signal is more compact in the wavelet basis it is easier to detect. We will can capture more power with the best fit parameters for the case of compact signals.

To delve deeper into how signal compactness affects detection, we will choose a fixed sine-Gausian of $Q = 10$ and examine how the resolution of the wavelet basis affects the SNR per pixel. Depending upon the time-frequency content of a signal it will be more compactly represented in a particular wavelet basis, leading to easier detection. We see spectrograms of the same sine-Gaussian glitch in four different wavelet bases in figure 8.5.

As expected, this broad band signal is best represented in a basis with broad band wavelets. Figure 8.6 shows SNR per pixel histograms and cumulative fractional power for each wavelet basis representation of the signal. We compare the fitting factor for BayesWave running on the same sine-Gaussian injection at different wavelet resolutions in figure 8.7. The results match the intuition established by considering the SNR per pixel.

The choice of wavelet resolution gives us one additional knob to turn for the BayesWave algorithm. Because this choice is dependent on the particular signal, we

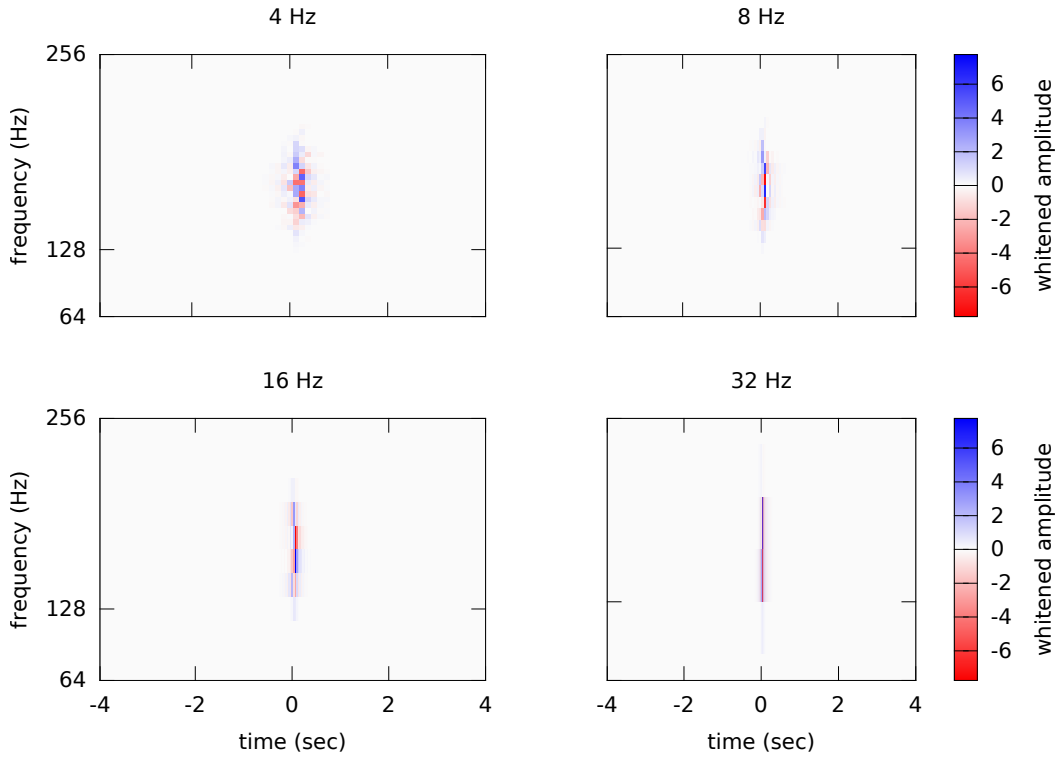

Figure 8.5. Time frequency plot of a sine-Gaussian at four different wavelet resolutions. This sine-Gaussian has $Q = 10$ and $f = 160$ Hz. Note how there is a good deal of leakage for the wavelet resolution with the smallest df.

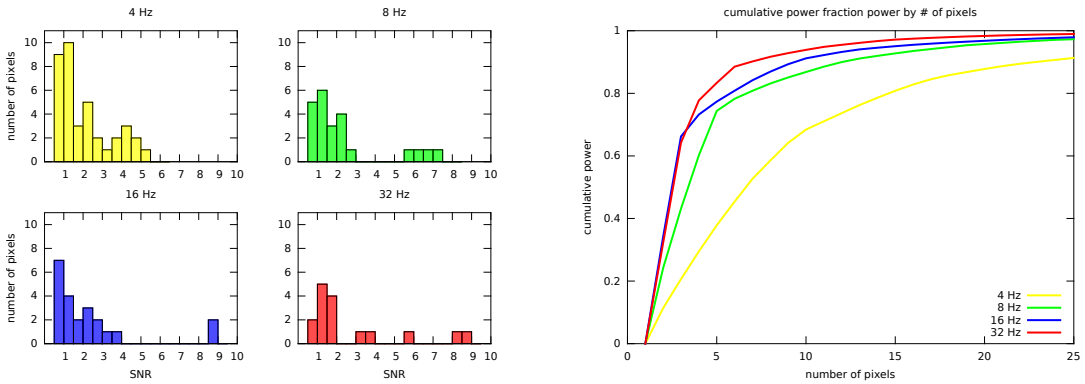

Figure 8.6. SNR per pixel and cumulative fractional power four four different wavelet resolutions. Note how there are two pixels for the 16 Hz (blue) case with SNR$>$ 8.5. The next loudest pixels have SNR$<$ 4. In the 8 Hz (green) case there are four pixels with SNR$>$ 5.5.

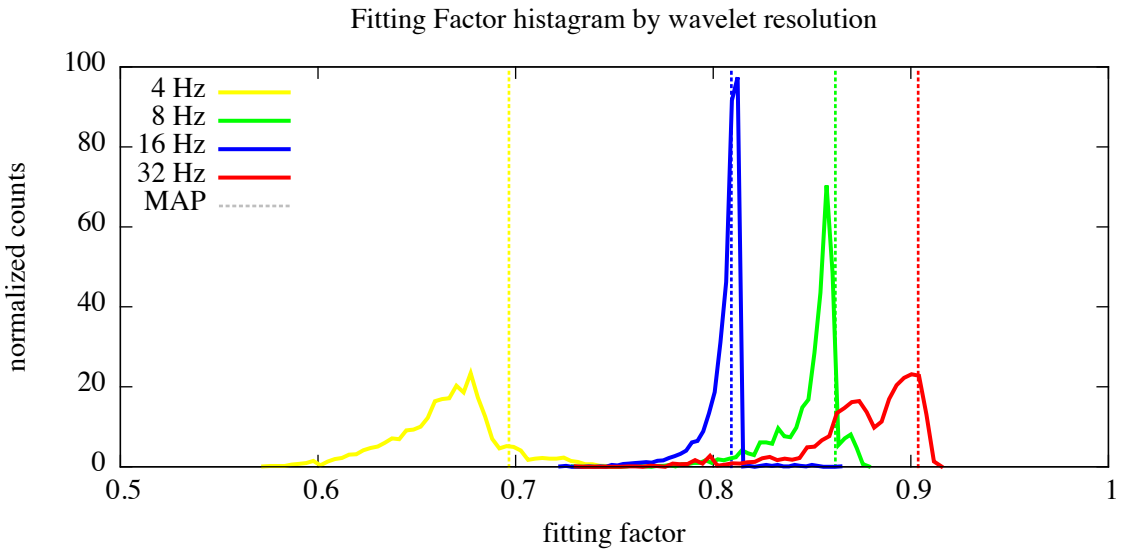

Figure 8.7. Histogram of recovered fitting factor for a sine-Gaussian injection at four wavelet resolutions. This sine-Gaussian has SNR=15 and is easily detectable at all resolutions. The MAP fitting factors for each resolution are shown as dashed vertical lines. The 16 Hz (blue) case used only the two loudest pixels in its MAP fit, whereas the 8 Hz (green) case used 4. This allowed for the 8 Hz case to achieve a better fit. The bimodal behavior in the 32 Hz (red) case results from the inclusion or exclusion of an additional pixel in the model.

cannot always choose the best basis *a priori*. As of now, we can run BayesWave multiple times at different wavelet resolutions and select which basis performed best after the fact. If the posterior distributions recovered from each run are properly normalized then we can even use Bayesian model selection to compare the results from the different runs. This adds an additional level to our hierarchical model selection problem.

After learning about the performance of BayesWave in a controlled environment we are ready to apply it to real LIGO data. We will still only consider a single detector at a time, using BayesWave as a glitch search. In this context we can think of

BayesWave as a data cleaner. By removing glitches it can clean LIGO data of its non-Gaussianity. Such data can then be analyzed by other methods, such as a template based search. As a glitch detecting method, BayesWave can also feed information to data quality analyses. By understanding individual glitches, their underlying causes in the physical detector can be found and corrected. BayesWave has only recently been applied to LIGO data for the first time. These results have not been reviewed by the LIGO Scientific Collaboration and should be considered very preliminary.

In the coherent Wave Burst analysis of LIGO's sixth science run (S6) an new family of common glitches was discovered at the LIGO Livingstone Observatory. The Detector Characterization group gave them the name 'flip glitch' based upon their visual appearance in high resolution time-frequency plots (see figure 8.8). While the flip glitch has shown to correlate with activity in a particular auxiliary channel, no one has yet to determine its true cause.

In figure 8.9, we see how BayesWave performed analyzing a segment of data containing a flip glitch. Flip glitches are of short duration, on the order of 10 ms, and they are broad in frequency. This analysis used 8 Hz wavelets, but it is likely that better results could be had at a different wavelet resolution. In the top left panel we see the original wavelet domain data. The flip glitch appears as an excess of power at $t = 0$. On the top right we see the best fit parameters recovered by BayesWave. In the middle left is the residual: the original data with the best fit glitch subtracted. In the residual the flip glitch is properly removed, and the data looks like Gaussian

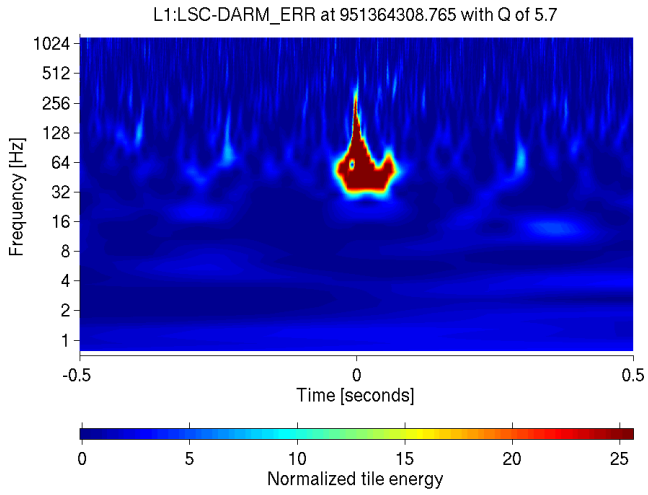

Figure 8.8. A 'flip glitch' in the LIGO L1 detector, as seen by the LIGO Data Quality group's Ω-pipeline.

noise. In the middle right panel we quantify the Gaussianity of the data before and after the glitch removal. This histogram shows the distribution SNR per wavelet pixel in the data. The original data has fat, non-Gaussian tails, resulting from the high amplitude glitch. The residual on the other hand is in much better agreement with Gaussian distributed noise, $N[0, 1]$. The bottom left panel shows the likelihood computed by BayesWave for 4 million steps of the McMC. Finally, the bottom right panel compare the time domain reconstructed glitch with the original data.

For another example we turn to a 'glitch storm' from LIGO's fifth science run (S5). Instead of a single, loud, isolated glitch, a glitch storm is an extended period marked by heightened noise. In this particular time there are periodic bursts of broad band noise. The first effect we can see of this glitch storm is its effect on LIGO's line removal routines. Before data is analyzed a good deal of preconditioning occurs. One

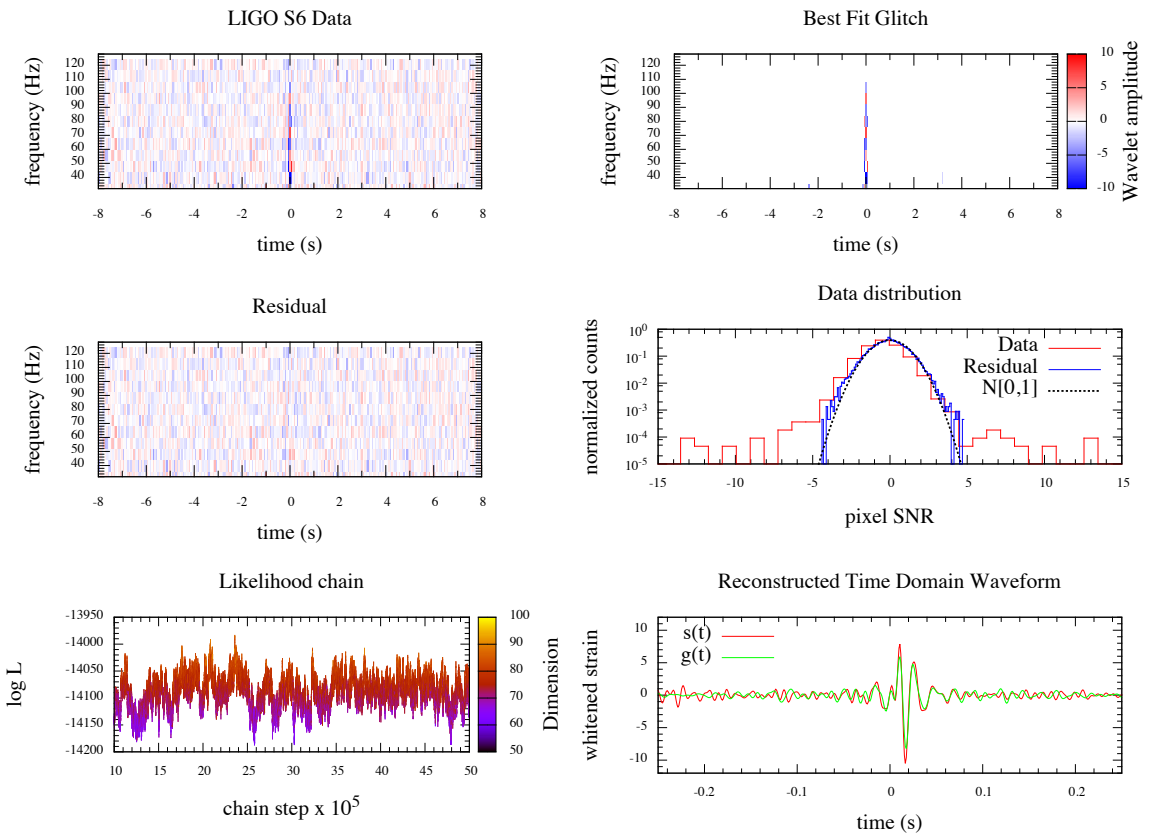

Figure 8.9. Summary of a BayesWave run on a LIGO S6 'flip glitch'. BayesWave fits the glitch accurately, removing the large non-Gaussian tails of the input data.

step is to remove the large resonance lines from the detector power spectrum. The largest resonance line in the spectrum is at 60 Hz, the frequency of input AC electrical current. In the top left panel of figure 8.10 we can plainly see that the line removal routines blanked out a wide range of frequencies around 60 Hz. There is a second line removal artifact around 110-120 Hz.

This glitch storm is a perfect candidate for iterative noise fitting. An outside routine can perform line removal and spectral estimation and pass whitened data to BayesWave. After it fits for glitches, BayesWave can pass the cleaned data back to

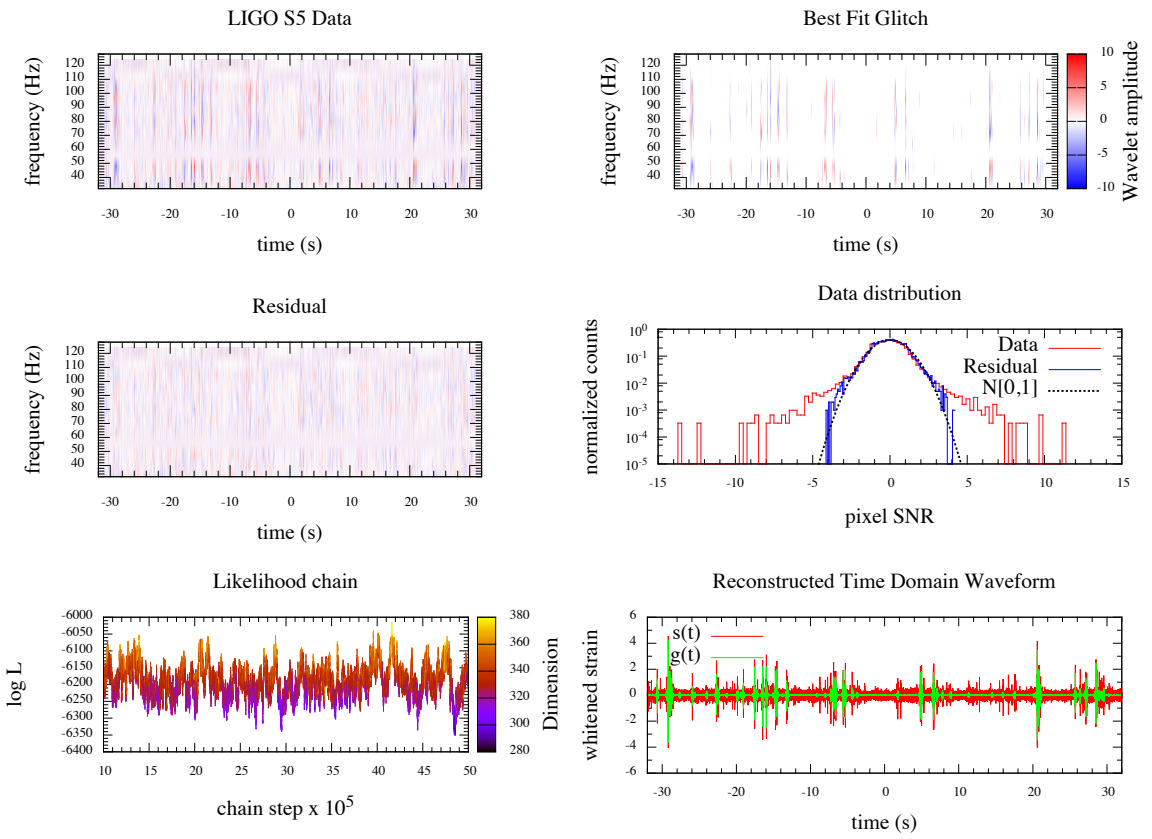

Figure 8.10. Summary of a BayesWave run on a LIGO S 'glitch storm', BayesWave fits many of the glitches, removing the large non-Gaussian tails of the input data. There is still some residual power that can be seen by eye. Note how this glitch storm was sufficient to cause problems with LIGO's preconditioning 'line removal' routines.

the noise fitting routines to achieve a better fit. In this manner a better fit from each method can be achieved after several iterations.

Even in these adverse conditions BayesWave performed admirably. In the middle right panel of figure 8.10 we see the residual is much more like non-Gaussian noise than the original data. However, in the middle left panel there is still residual power following the morphology of the glitch storm. The human eye can detect these small

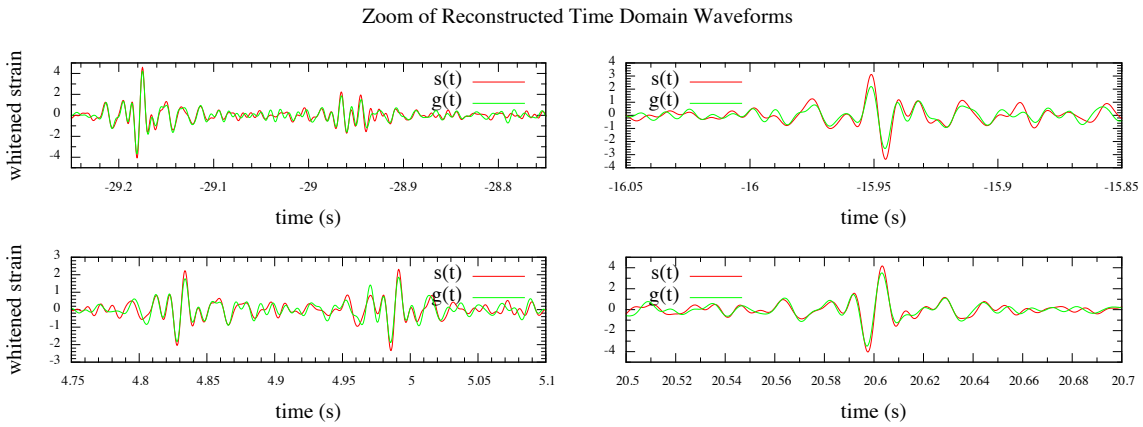

Figure 8.11. Detail of 'glitch storm' time domain reconstruction.

correlations much better than any algorithm, especially when told what to look for by the panel above.

In figure 8.11 we see four detail plots of the reconstructed time domain glitches. BayesWave accurately reconstructs the multiple glitches in this time segment.

8.2. BayesWave as a Gravitational Wave Search

We saw that BayesWave is able to accurately detect and reconstruct glitches in a single gravitational wave detector, but what about actual gravitational waves? To detect a gravitational wave it must appear in multiple detectors, consistent with a single point of origin in the sky. BayesWave models this gravitational wave in the geocenter, then depending upon the actual sky location and polarization of the gravitational wave, the various detectors will respond differently. This action was summarized in chapter 7 by a projection operator for each detector, $\mathcal{A}_i$.

The simplest case for gravitational wave detection is the case colocated, aligned detectors. In this case $\mathcal{A}_i$ is the same for all detectors, and the wavelet domain representation of the gravitational wave will be identical in all detectors. To examine BayesWave's performance in this simplest of scenarios, we will consider two gravitational wave detectors. They will be aligned and located at the geocenter. In this case the geocenter model is the gravitational waveform in each detector, $\mathcal{A}_i = 1$.

Consider a sine-Gaussian gravitational wave burst with $Q = 14$ and $f = 120$ Hz. We will inject this burst with a network SNR of about 14, requiring individual detector SNRs of about 10 each. We will also inject one loud glitch per detector. In the first detector we will inject a very broad spectrum $Q = 1$, $f = 100$ Hz sine-Gaussian about 3.5 s before the signal This glitch has a SNR of about 15, making it borderline detectable given our single detector studies. In the second detector we will inject a $Q = 5$ sine-Gaussian at about 175 Hz less than 1 s after the signal. This glitch was injected with a SNR of about 25. It should be easily detected by BayesWave.

In figure 8.12 we see the simulated data and residuals, along with the injected and recovered gravitational wave signal. BayesWave recovered the gravitational wave signal and glitches all with Bayes factors of $\mathcal{B} \geq 5 \times 10^6$. The signal was recovered very well with a fitting factor of 0.89. In the residuals both glitches are still somewhat visible. The glitch in the first detector was recovered with a fitting factor of about 0.6. A good deal of power was left behind in the low amplitude high frequency tail. The glitch in the second detector was recovered with a fitting factor of about 0.85. In

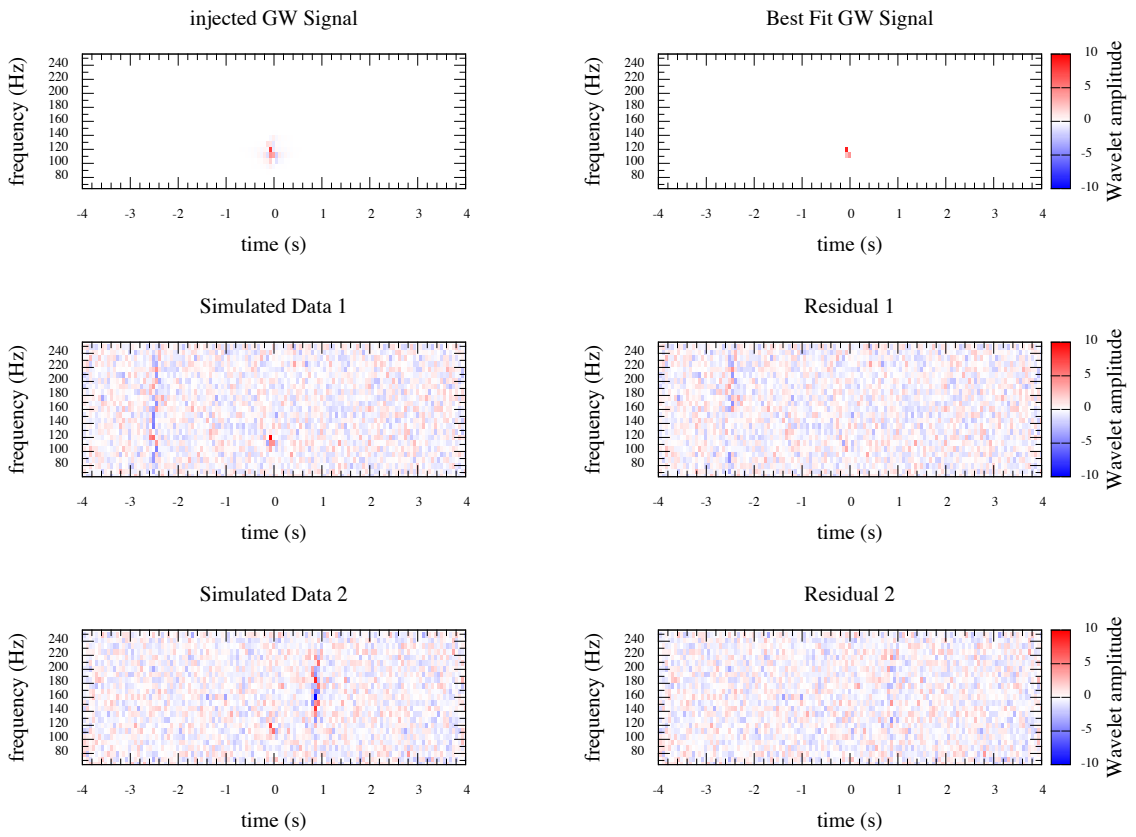

Figure 8.12. Summary of a BayesWave run on simulated two detector gravitational wave data with detector glitches. The top row shows the injected and recovered wavelet domain gravitational wave signal. The second and third rows show the simulated data, and residuals after subtracting the best fit gravitational wave signal and glitches.

figure 8.13 we see time domain reconstructions for all compared with their respective injected waveforms. In the data distribution we see our simulated data has much sparser non-Gaussian tails than the real LIGO cases considered in section 8.1.

The colocated detector case is the simplest testing ground for a gravitational wave search. In order to implement a true multi-detector search we will need to know

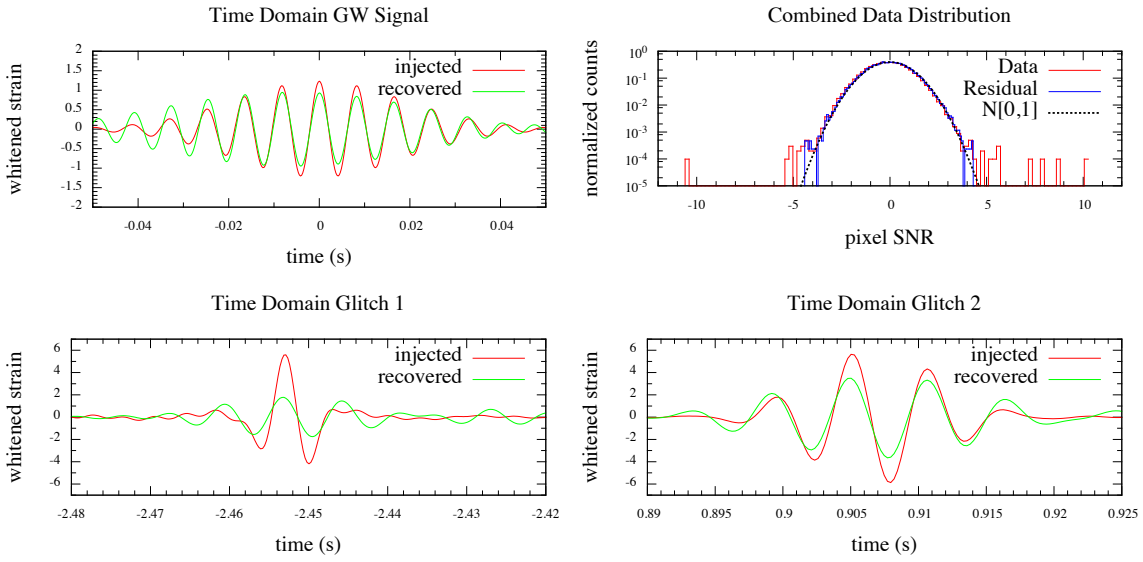

Figure 8.13. Time domain reconstructions of gravitational wave signal and two glitches compared to the injected waveforms. The top right panel shows the data SNR distribution combined for both detectors.

the projection operators $\mathcal{A}_i$ to project our geocenter modeled signal onto the various detectors of our network.

Applying the angular detector response for an incident gravitational wave is easy. At any given time a detector's reference frame is defined by its location on the Earth, the time of day, and the time of year. For the existent and planned future ground based detectors the location on the Earth is known. The time of observation is essentially arbitrary. For the high frequency, short duration signals for which we are looking, we can safely assume the detector is fixed in space. In this way we can use the same angular response for a detector for the whole of a brief, order minutes observation. The angular response acts simply to scale the amplitude of incident

gravitational waves in each detector. This same scaling can be applied to all wavelet pixels in each detector for a given sky location.

The other effect of the projection operator is the alter the time of arrival of a signal for the various detectors. Because the detectors in our network are spacially separated, a gravitational wave will arrive at each detector at different times depending upon its sky location. To account for this we need to shift the gravitational wave signal in time relative to the geocenter model in each detector. The maximum distance the signal can travel between any two detectors is the earth's diameter, leading to a difference in arrival time of about 42 ms. Our gravitational model is in the geocenter, so at worst we will only need to be able to project a pixel forward and backward in time by about 22 ms.

Time shifting waveforms is trivial in the time domain. Unfortunately, in the wavelet domain unless the shift is by exactly the pixel time spacing, there are problems. Because wavelets form a complete basis, we can construct any function out of a linear combination of individual wavelet pixels. This of course includes a time shifted wavelet pixel. We must determine how to represent an arbitrarily time shifted pixel in terms of the original wavelet basis.

To do this we generate a time domain wavelet waveform and shift it in time. Then we wavelet transform the shifted waveform to determine its wavelet representation. In general the time shifted pixel will require many wavelet pixels for a full representation. If we are willing to accept a 1% loss of power, we can greatly reduce the number

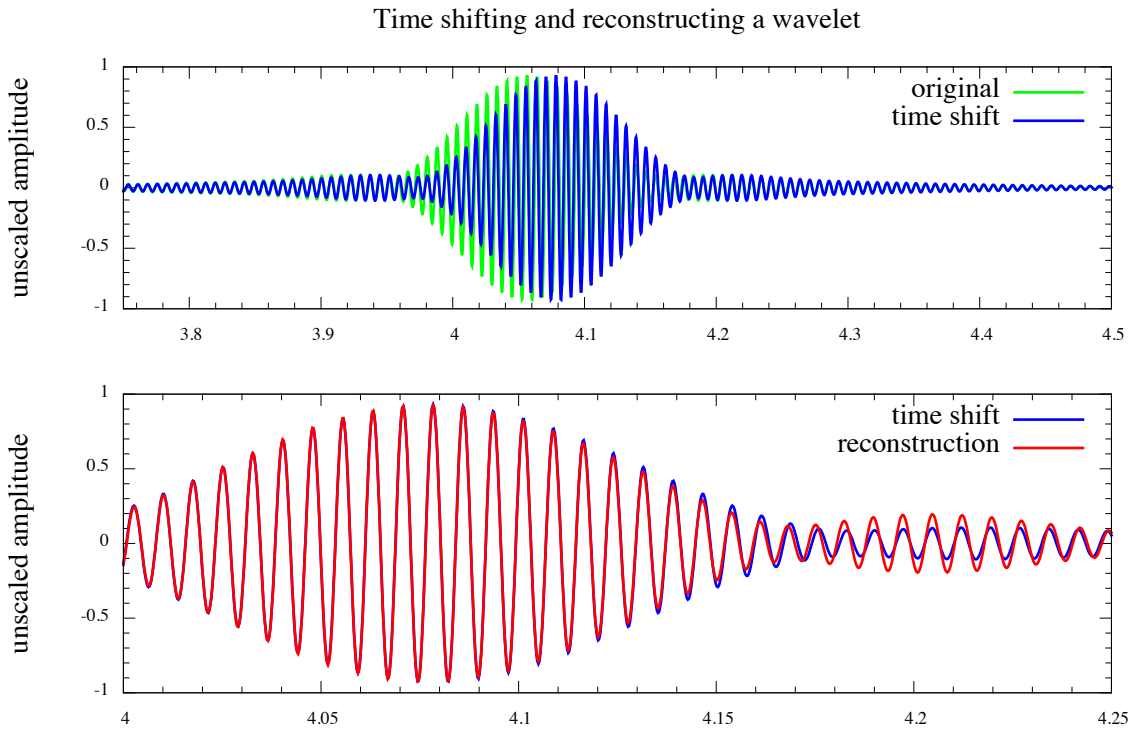

Figure 8.14. Time shifting and reconstructing a wavelet pixel. This pixel is from an 8 Hz decomposition spanning the range 120 to 128 Hz. It is shifted forward in time by 22 ms and then reconstructed in the wavelet domain demanding less than 1% error. The resulting reconstruction fits very well in the high amplitude core with error developing on the tails. Here we achieve a 0.994 fitting factor using only 7 pixels..

of pixels in the wavelet domain reconstruction. Figure 8.14, shows a wavelet pixel shifted by 22 ms and reconstructed as sum of only seven wavelet pixels. For any of the individual time shifts we need for terrestrial detectors, the shifted waveform can be reproduced with about 20 pixels. This method of constructing time shifted wavelets directly in the wavelet domain is functionally equivalent to cWB's method of 'time delay filters'.

For each wavelet layer we can construct a look-up table telling how pixels from that layer shift in time. Then when we want to time shift an entire signal model,

we can simply shift each pixel individually using this table. In this way we can shift arbitrary signals in time directly in the wavelet domain. We never have to perform any wavelet transforms during the BayesWave run time, saving a great deal of computational cost.

This wavelet domain time shifting is in the process of being implemented and tested. In order to analyze data from the LIGO detectors this work must be completed. One problem with analyzing data that contains both gravitational waves and glitches is convergence time. If a particular pixel from a gravitational wave is fit by our glitch model it will be hard to remove it from that fit and then add it to the gravitational wave fit.

To ease this process we must implement proposals that attempt to transform pixels in the glitch model into pixels in the gravitational wave model and *vice versa*. Proposing to change a gravitational wave to individual detector glitches is not a problem. This simply involves removing a pixel from the gravitational wave model and replacing it with the proper projections in each detector for the glitch model.

Transforming glitches into signals is another matter. To do this we must identify which pixels in each glitch model could combine to be the result of a gravitational wave. In general several pixels are need to account for the 99% time shift accuracy, but we can reduce this problem to one that considers only the largest amplitude pixel in each model. For the time shifts we are considering, the loudest pixel in the reconstruction is always the original pixel or one of its left/right neighbors. For

spacially separated detectors we can propose moves that transform the proper set of glitches for the current sky location of the signal model. If there are no pixels in the signal model, we can simply propose a sky location first and then choose a set of consistent glitches.

When we have a network of 3 spacially separated detectors, it is possible to measure the polarization of gravitational waves. In this case we may model the plus, h_+, and cross, $h_\times$, gravitational wave polarizations independently or we may choose a specific polarization where $h_\times$ is function of h_+. In this case we only need to model one polarization. The second is computed from the first, then both can be projected onto the individual detectors in the usual way.

For elliptical polarized gravitational waves we can define right and left handed polarizations, h_r and h_l, that are related in the Fourier domain by an amplitude rescale and a constant phase shift. In this case if we wish to compute one from the other it is convenient to do so in the Fourier domain. Similarly to the time-shifts we discussed earlier, we can represent a phase shifted wavelet as a sum of other wavelets. We may even combine this process, producing time-shifted wavelets with and without the relative phase shift. In this way we can model polarized gravitational waves as a single polarization in the geocenter and project both polarizations onto the network.

9. CONCLUSIONS

The first direct detection of gravitational waves will be momentous event for General Relativity. The first detections will likely arise from analysis techniques that can analyze data rapidly after collection. The data analysis techniques we have discussed herein will likely not be applied in 'low-latency' analyses, but they still have an important role in gravitational wave detection. Simply detecting gravitational waves is not sufficient, we wish to use gravitational waves as a tool for astronomy. To that end we must accurately measure the parameters of the radiating systems. It is in the characterization of gravitational waves that non-gaussian noise can be most troubling. Detector glitches may not entirely spoil a detection. They may weaken the confidence of a detection or reduce the accuracy of parameter estimation. A more computationally expensive 'follow-up' analysis will be critical in establishing confidence in the first detections.

Multi-Variate Statistical Classifiers (MVSC) are a very promising tool for gravitational wave data analysis. The Random Forest of Bagged Decision Trees algorithm out performed traditional statistics in the LIGO search for black hole ringdowns. Over the course of the ringdown analysis we learned much about the application of MVSC algorithms to gravitational wave problems that differed from the conventional wisdom developed for high energy physics analysis.

Several LIGO data analysis groups are exploring MVSC algorithms to enhance their efforts. Within the Compact Binary Coalescence (CBC) group other searches

will soon be carried out using MVSC. The search for high mass binary black hole inspirals is poised to be the second search to use the output of a MVSC algorithm as a detection statistic. Detector characterization analysts seek to use MVSC algorithms to classify detector glitches into families in software instead of by hand by human analysts. Such efforts can provide a windfall of information to data quality studies.

The direct noise modeling BayesWave can be applied to template based searches as well. There is an ongoing effort to integrate BayesWave noise modeling techniques into the parameter estimation follow-up pipelines of the LIGO CBC group. In these efforts the gravitational wave burst model is replaced with a template model either an inspiral only post-Newtonian expansion or a full inspiral-merger-ringdown waveform. These efforts should provide superior parameter estimation in noisy LIGO data, potentially probing data that was vetoed for data quality concerns.

In the near future we will concentrate on adapting the BayesWave algorithm to a full network analysis capable of coherent gravitational wave detection. Such an analysis may provide better sky localization than existing techniques. To this end a priority will be comparing the performance of BayesWave directly against coherent WaveBurst (cWB). Due to cWB's highly developed infrastructure, BayesWave will never replace it as a low latency gravitational wave search, but BayesWave can provide a Bayesian follow-up pipeline to provide full posterior distributions for reconstructed gravitational waves, instead of mere maximum likelihood estimates.

In the coming years we will see the first detection of gravitational waves or be forced to amend our understanding of General Relativity. The future of gravitational wave astronomy will depend not just on understanding astrophysical systems but also on understanding the gravitational wave detectors. By planning to analyze advanced detector data containing non-gaussian noise, we position ourselves to detect gravitational waves and produce the best possible astrophysical results.

REFERENCES CITED

[1] Abadie, J., et al. (LIGO Scientific Collaboration). Predictions for the rates of compact binary coalescences observable by ground-based gravitational-wave detectors. *Classical and Quantum Gravity 27* (2010), 173001.

[2] Abbott, B. et al. (LIGO Scientific Collaboration). Search for gravitational wave ringdowns from perturbed black holes in ligo s4 data. *Phys. Rev. D 80* (2009), 062001.

[3] Ajith, P. et al. A phenomenological template family for black-hole coalescence waveforms. *Classical and Quantum Gravity 24* (2007), S689.

[4] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19* (1974), 716–723.

[5] Aso, Y., Michimura, Y., Somiya, K., Ando, M., Miyakawa, O., Sekiguchi, T., Tatsumi, D., and Yamamoto, H. Interferometer design of the KAGRA gravitational wave detector. *arXiv:1306.6747 [gr-qc]* (2013).

[6] Baiotti, L., Hawke, I., and Rezzolla, L. On the gravitational radiation from the collapse of neutron stars to rotating black holes. *Classical and Quantum Gravity 24* (2007), S187.

[7] Baker, P. T. Using a multi-variate statistical classifier in the ligo black hole ringdown search. *LIGO Technical Document*, T1200215-v2 (2013).

[8] Bauer, E., and Kohavi, R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning 36* (1999), 105–142.

[9] Bayes, T. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. 53* (1763), 370–418.

[10] Beringer, J. et al. (Particle Data Group). Review of particle physics. *Phys. Rev. D 86* (2012), 010001.

[11] Berti, E., Cardoso, V., Gonzalez, J. A., U., S., Hannam, M., Husa, S., and Brügmann, B. Inspiral, merger, and ringdown of unequal mass black hole binaries: A multipolar analysis. *Physical Review D 76* (2007), 064034.

[12] Berti, E., Cardoso, V., and Starinets, A. O. Quasinormal modes of black holes and black branes. *Classical and Quantum Gravity 26*, 16 (2009).

[13] Berti, E., Cardoso, V., and Will, C. M. On gravitational-wave spectroscopy of massive black holes with the space interferometer LISA. *Physical Review D 73* (2006), 064030.

[14] Buonanno, A. et al. Approaching faithful templates for nonspinning binary black holes using the effective-one-body approach. *Physical Review D 76* (2007), 104049.

[15] Carroll, S. M. *An Introduction to General Relativity Spacetime and Geometry.* Addison Wesley, 2004.

[16] Caudill, S., Baker, P. T., Talukder, D., Benacquista, M., Cornish, N., Bose, S., and Hodge, K. Searches for gravitational wave ringdowns. *LIGO Technical Document*, T1200214-v2 (2012).

[17] Caudill, S. E. *Searches for Gravitational Waves from Perturbed Black Holes in Data from LIGO Detectors.* PhD thesis, Lousiana State University, July 2012.

[18] Christiansen, N. Ligo s6 detector characterization studies. *Classical and Quantum Gravity 27* (2010), 194010.

[19] Clemence, G. M. The relativity effect in planetary motions. *Rev. Mod. Phys. 19* (1947), 361–364.

[20] Cornish, N. J., and Porter, E. K. Detecting galactic binaries with lisa. *Classical and Quantum Gravity 22* (2005), S927–S934.

[21] Creighton, J. D. E. Search techniques for gravitational waves from black-hole ringdowns. *Physical Review D 60* (1999), 022001.

[22] Einstein, A. Zur elektrodynamik bewegter kröper. *Annalen der Physik 17* (1905).

[23] Einstein, A. Chapter from. *Jahrbuch für Radioakt. und Electronik 4* (1907).

[24] Einstein, A. Über den einfluss der schwekraft auf die ausbreitung des lichtes. *Annalen der Physik 35* (1911).

[25] Einstein, A. Entwurf einer verallgemeinerten relativitätstheorie und einer theorie der gravitation. *Zeitschrift für Mathematik und Physik 62* (1913), 225–261.

[26] Einstein, A. Die gundlage der allgemeinen relativitätstheorie. *Annalen der Physik 49* (1916).

[27] Farrell, S., Webb, N., Barret, D., Godet, O., and Rodrigues, J. An intermediate-mass black hole of over 500 solar masses in the galaxy eso 243-49. *Nature 460* (2009), 73.

[28] Folland, G. B., and Sitaram, A. The uncertainty principle: A mathematical survey. *Journal of Fourier Analysis and Applications 3* (1997), 207–238.

[29] Green, P. J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82* (1995), 711–732.

[30] H., N., Takahashi, H., Tagoshi, H., and Sasaki, M. An effective search method for gravitational ringing of black holes. *Physical Review D 68* (2003), 102003.

[31] Hastings, W. K. Monte Carlo sampling methods using Markov chain and their applications. *Biometrika 57* (1970), 97–109.

[32] Isaacson, R. A. Gravitational radiation in the limit of high frequency. i: The linear approximation and geometrical optics. *Phys. Rev. 166* (1968), 1263–1271.

[33] Isaacson, R. A. Gravitational radiation in the limit of high frequency ii: Nonlinear terms and the effective stress tensor. *Phys. Rev. 166* (1968), 1272–1279.

[34] Kissel, J. S. *Calibrating and Improving the Sensitivity of the LIGO Detectors*. PhD thesis, Lousiana State University, August 2010.

[35] Klimenko, S., Yakushin, I., Mercer, A., and Mitselmakher, G. Coherent method for detection of gravitational wave bursts. *Classical and Quantum Gravity 25* (2008), 114029.

[36] Le Verrier, U. Lettre de m. le verrier à m. faye sur la théorie de mercure et sur le mouvement du périhélie de cette planète. *Comptes rendus hebdomadaires des séances de l'Académie des sciences 49* (1859), 379–383.

[37] Leaver, E. An analytic representation for the quasi normal modes of kerr black holes. *Proceedings of the Royal Society of London 402* (1985), 285–298.

[38] LIGO Data Analysis Software Working Group. LSC Algorithm Library Suite. *https://www.lsc-group.phys.uwm.edu/daswg/projects/lalsuite.html*.

[39] Littenberg, T., and Cornish, N. Separating gravitational wave signals from instrument artifacts. *Physical Review D 82* (2010), 103007.

[40] Littenberg, T. B. *A Comprehensive Bayesian Approach to Gravitational Wave Astronomy*. PhD thesis, Montana State University, April 2009.

[41] Mallat, S. A theory for multi-resolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11* (1989), 674–693.

[42] Michelson, A. A., and Morley, E. W. On the relative motion of the earth and the luminiferous ether. *American Journal of Science 34* (1887), 333–345.

[43] Narsky, I. Optimization of signal significance by bagging decision trees. *arXiv:physics/0507157v1 physics.data-an* (2005).

[44] Narsky, I. StatPatternRecognition: A C++ Package for Statistical Analysis of High Energy Physics Data. *arXiv:physics/0507143v1 physics.data-an* (2005).

[45] Newland, D. E. *An Introduction to Random Vibrations, Spectral and Wavelet Analysis*. Longman Scientific & Technical, 1993.

[46] Norris, J. *Markov Chains*. Cambridge University Press, 1998.

[47] Regge, T., and Wheeler, J. A. Stability of a schwarzschild singularity. *The Physical Review 108* (1957), 1063.

[48] Samet, H., and Tamminen, M. Efficient component labeling of images of arbitrary dimension represented by linear bintrees. *IEEE Transactions on Pattern Analysis and Machine Intelligence 10* (1988), 579 – 586.

[49] Schwarzschild, K. Über das gravitationsfeld eines massenpunktes nach der einsteinschen theorie. *Sitzungsberichte der Königlich-Preussischen Akademie der Wissenschaften 3* (1916), 189–196.

[50] Searle, A. C., Sutton, P. J., and Tinto, M. Bayesian detection of unmodeled bursts of gravitational waves. *Classical and Quantum Gravity 26* (2009), 155017.

[51] Shibata, M., Kyutoku, K., Yamamoto, T., and Taniguchi, K. Numerical simulations of compact object binaries. *Physical Review D 79* (2009), 044030.

[52] Stein, E. M., and Shakarchi, R. *Princeton Lectures in Analysis III: Real Analysis, Measure Theory, Integration, and Hilbert Spaces*. Princeton University Press, 2005.

[53] Stevin, S. *The Principal Works of Simon Stevin*, vol. 1. C. V. Swets & Zeitlinger, 1955.

[54] Stigler, S. M. *The History of Statistics*. Harvard University press, 1986.

[55] Sutton, P. J. et al. X-pipeline: An analysis package for autonomous gravitational-wave burst searches. *New Journal of Physics 12* (2010), 053034.

[56] Talukder, D., Bose, S., Caudill, S., and Baker, P. T. Improved coincident and coherent detection statistics for searches for gravitational wave ringdown signals. *LIGO Technical Document*, T1300343-v2 (2013).

[57] Teukolsky, S. Perturbations of a rotating black hole: I. fundamental equations for gravitational, electromagnetic and neutrino-field pertrubations. *The Astrophysical Journal 185* (1973), 635–647.

[58] Wald, R. M. *General Relativity*. University of Chicago Press, 1984.

[59] Weisberg, J. M., and Taylor, J. H. Relativistic binary pulsar b1913+16: Thirty years of observations and analysis. In *Binary Radio Pulsars, Proc. Aspen Conference, ASP Conf. Series* (2004), F. Rasio and I. Stairs, Eds., Binary Radio Pulsars, ASP.

[60] Weisberg, J. M., Taylor, J. H., and Fowler, L. A. Gravitational waves from an orbiting pulsar. *Scientific American 245* (1981), 74–82.

[61] Wen, L., and Schutz, B. F. Coherent network detection of gravitational waves: the redundancy veto. *Classical and Quantum Gravity 22* (2005), S1321S1335.

[62] Yunes, N., and Siemens, X. Gravitational wave tests of general relativity with ground-based detectors and pulsar timing arrays. *Living Reviews in Relativity* (in preparation).