A COMPREHENSIVE BAYESIAN APPROACH TO GRAVITATIONAL WAVE ASTRONOMY

by

Tyson Bailey Littenberg

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

of

Doctor of Philosophy

in

Physics

MONTANA STATE UNIVERSITY
Bozeman, Montana

April, 2009

APPROVAL

of a dissertation submitted by

Tyson Bailey Littenberg

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 Division of Graduate Education.

Dr. Neil J. Cornish

Approved for the Department of Physics

Dr. Richard Smith

Approved for the Division of Graduate Education

Dr. Carl A. Fox

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. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Tyson Bailey Littenberg

April, 2009

DEDICATION

To Quinn for motivating me to start, Milo for motivating me to finish, and Sabrina for occupying Quinn and Milo so I could write.

ACKNOWLEDGEMENTS

This dissertation could not have been completed at this time or in this state without the help of a large supporting cast. Particularly: Neil Cornish for dreaming all this stuff up and for being an unparalleled resource and mentor during the time it took me to do it. My committee for their helpful insight, especially Charles Kankelborg for his valuable advice regarding the statistical side of this dissertation. David McKenzie and the Solar Physics group for precious CPU hours as I neared the completion of this work. And finally, Margaret Jarrett for allowing graduate students to focus on graduate school while she is quietly taking care of everything else. On the personal side, I especially want to thank Joey Key for being my office mate, co-manager, co-miserator, and co-celebrator since the day I arrived in Bozeman, and along with Camas, Cayle, Steve, and Debora, for being my family's family away from home. Thanks go out to my Mom for keeping me company Sunday mornings over coffee, my Dad for instilling just the right amount of "nerd" in me, and Eli for dragging me out of my apartment when I moved to Bozeman, and up and down mountains when our paths would cross each winter. Thank you to Steve and Carolyn for showing me what academia and science were all about – perhaps your intention was to warn me off but either way I'm happy here. And finally, thank you Sabrina, for putting your dissertation on hold while I finished mine. I look forward to returning the favor.

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

ABSTRACT

The challenge of determining whether data from a gravitational wave detector contains signals which are cosmic in origin is the central problem in gravitational wave astronomy. The "detection problem" is particularly challenging for low amplitude signals embedded in "glitchy" instrument noise. It is imperative that we can robustly distinguish between the data being consistent with instrument noise alone, or noise and a weak gravitational wave signal. In response to this challenge we have set out to develop a robust, general purpose approach that can locate and characterize gravitational wave signals, and provided odds that the signal is of cosmic origin. Our approach employs the Markov Chain Monte Carlo family of algorithms to construct a fully Bayesian solution to the challenge – the Parallel Tempered Markov Chain Monte Carlo (PTMCMC) detection algorithm. The PTMCMC detection algorithm establishes which regions of parameter space contain the highest posterior weight, efficiently explores the posterior distribution function of the model parameters, and calculates the marginalized likelihood, or evidence, for the models under consideration. We illustrate our approach using simulated LISA and LIGO-Virgo data.

1

INTRODUCTION

The first direct detection of gravitational waves (GWs) is now within reach, ushering in the beginning of a new field of astronomy. The *Laser Interferometer Gravitational wave Observatory* (LIGO) [2, 3] completed the acquisition of one year of triple-coincidence data at design sensitivities (the S5 science run) at the end of 2007 and analysis efforts are in full swing [4, 5]. Currently, upgrades to the LIGO and Virgo [6] detectors are nearing completion in preparation for the S6/VSR2 data collection period which is scheduled to begin in mid-2009. This "enhanced" iteration of the original LIGO and Virgo design will pave the way for the advanced generation of observatories and routine gravitational wave detection. Meanwhile, plans for the space based *Laser Interferometer Space Antenna* (LISA) [7] continue to mature. All key LISA technologies are now in place and will be flown in 2010 on the *LISA Pathfinder* [8] mission. While hardware developments have brought the first detection within reach, exploiting the full potential of these instruments requires similar improvements in the techniques used to analyze the data.

Until recently, the primary tool of gravitational wave data analysis has been that of frequentists statistics. Starting with references [9, 10] the community has begun to break away from this approach in favor of Bayesian methods. Inference of the astrophysical processes responsible for a gravitational wave signal is most appropriately performed in the Bayesian framework and this approach, when possible, should be the one taken. Frequentist methods are appropriate, for example, in high energy particle physics as the experimenter has detailed control of the experiment parameters (e.g., the conditions under which particles collide) allowing the event to be repeated many times, building a large enough sample that statistics are meaningful. Gravita-

tional wave astronomy involves one observation of an event which cannot be repeated. Therefore, to establish statistics the frequentist must analyze simulated ensembles of data in an attempt to understand *what would happen* if multiple occurrences of the same event were to be recorded. This is in contrast to Bayesian analysis which is applied only to the data acquired.

In addition to GW astronomy, Bayesian probability theory is slowly making its way into the mainstream of astrophysics and astronomy [11, 12, 13, 14]. In the Bayesian framework, the data are used to update our belief in the validity of a hypothesis according to the likelihood that the hypothesis in question could have generated the data under the performed experiment, weighted by our prior belief in that hypothesis. This is often referred to as "Bayesian learning" to emphasize how this quantitatively mirrors how we as individuals assimilate new information into our world view. It takes significant evidence to sway our opinion if we are operating under some prior belief. Although codifying such "biases" into our analysis may initially seem to contradict the scientist's goal of objectivity, we will later see that the inclusion of this prior information forces us to confront and address these biases.

Bayesian inference can be split into two sub-topics: Parameter estimation and model selection. Parameter estimation is, to date, the most thoroughly employed branch of Bayesian methods in gravitational wave data analysis. As the name implies, this involves finding the "best fit" parameters under a specific model of the gravitational wave data. An example of which would be determining the area on the celestial sphere where an emitting source is most likely to be located.

The less commonly employed model selection techniques assign probabilities to different hypotheses used to describe the data (e.g., the data contains a GW signal either from a core-collapse supernova or the merger of a binary neutron star system). For GW astronomy applications, model selection problems are solved in consort with

parameter estimation, as it would be a mistake to ask about the relative validity of two different models (a supernova or a binary merger) without first finding the highest probability configuration (the best-fit parameters) for each model. We will focus our attention on the model selection aspect of gravitational wave astronomy, while still giving due diligence to the underlying parameter estimation problems.

The most important example of model selection in gravitational wave astronomy is that of the "detection problem." Data from a gravitational wave detector is the sum of two independent contributions: Random instrument noise and the incident gravitational waves. The detection problem involves accurately extracting the GW signal from this noisy data. Obviously, for exceptionally bright GW signals this challenge is easily met. Instead, our interest lies in the regime of marginally detectable signals. This is of great importance for the current state of ground-based detectors, as large signal-to-noise ratio (SNR) events are unlikely due to the limited sensitivity of these instruments. The difficulty of making a detection is intensified as the detectors on Earth are subject to frequent transient noise events resulting in a high probability that "candidate" events are instrumental, instead of cosmic, in origin. It is therefore imperative for maximizing the success of this class of detector that we can robustly distinguish between the data being consistent with "glitchy" instrument noise alone, or noise and a weak gravitational wave signal. Currently the analysis of data from ground based gravitational wave detectors is mostly conducted using classical statistics, and methods such as the Neyman-Pearson criteria are used for model selection.

Future space based detectors, such as LISA, will also pose unique model selection problems, as it is expected that this class of detector will produce rich data streams containing signals with a high degree of overlap. For example, an important class of LISA sources are the millions of stellar-mass binary systems within our own galaxy,

tens of thousands of which will be detectable. Those that cannot be resolved blend together to form "confusion noise" which will dominate the instrument noise over a certain frequency range.

In addition to the number of sources which can be resolved being unknown, the number of parameters required to model the waveforms is also in question. For example, a significant subset of the resolvable galactic binaries will exhibit orbital frequency evolution, while a smaller number will have measurable eccentricity. Over-parameterizing signal model leads to over-fitting of the data and thus spurious results. It is important to avoid this over-fitting in order to maintain the desired precision measurements of the emitters.

The galactic binary model selection problems are critical to the LISA mission success. The subtraction of the galactic foreground is an important data-processing step and it is vital that a "clean" residual remains after its removal, allowing us to dig deeper into the data for fainter, and arguably more interesting, cosmological sources.

In response to these challenges we have set out to develop a generic, comprehensive, and robust solution to the detection problem which will pave the way for future studies in gravitational wave astronomy. In this thesis we will iteratively construct such an approach. We tackle the problem in four phases:

1. Explore different Bayesian model selection techniques on a simple one-parameter test case.

2. Expand these techniques to a large-dimension detection problem.

3. Build on the previous developments by improving the efficiency of the technique, allowing us to approach more complicated signals.

4. Develop robustness against the realities of real instrument noise, specifically non-stationarities.

This requires the employment of many established techniques in Bayesian gravitational wave data analysis, specifically the heavy use of the Markov Chain Monte Carlo family of algorithms. In addition, we introduce to the field novel methods, particularly in our calculation of the model evidence and noise modeling for both space- and ground-based detectors.

What follows is an outline for the remainder of this dissertation. The references contained in this outline point to work that was original to this thesis.

We will begin in chapter 2 with a brief discussion of gravitational waves, the machines used to detect them, and the response of the detectors to incident gravitational radiation. This chapter closes with a technique for rapidly modeling LISA's response to a signal which is slowly increasing in frequency and amplitude (i.e., "chirping") found in [15].

We follow our discussion of gravitational wave astronomy with an introduction to Bayesian probability theory in chapter 3, detailing the parameter estimation and model selection applications of Bayes' theorem. We then outline the computational machinery necessary to the implementation of these techniques, specifically the Markov Chain Monte Carlo algorithm and the specifics involved therein when applying these techniques to gravitational wave astronomy.

In chapter 4, originally appearing in [15], we test a number of Bayesian model selection techniques as applied to the simple case of distinguishing under what conditions the frequency evolution of a single galactic binary in the LISA band should be included as a parameter. The methods explored are the Laplace approximation to

the evidence, Schwarz-Bayes Information Criteria, Savage-Dickie density ratios, and of notable importance, the Reverse Jump Markov Chain Monte Carlo approach.

Chapter 5 defines the gravitational wave detection problem and makes the argument that it is most suitably solved using Bayesian methods. We then outline our basic strategy, dubbed the Parallel Tempered Markov Chain Monte Carlo (PTMCMC) detection algorithm [16], for resolving the detection problem.

Chapters 6 and 7 describe in detail the specific implementations of the PTMCMC detection algorithm for LISA [16] and LIGO [17], respectively. The details unique to each application involve the types of signals being analyzed and, most importantly, the specifics of the detection algorithms implementation, including how the noise modeling is approached.

In chapter 8 we will summarize the results of this work followed by a discussion in chapter 9 of the future outlook for the continued evolution of this complete and general analysis technique.

7

GRAVITATIONAL WAVES

Introduction

Gravitational waves are a yet-to-be observed, but necessary, consequence of one of science's most rigorously tested theories – General Relativity. Instead of the familiar picture of a wave – some disturbance propagating through a medium – gravitational waves are perturbations of space-time itself. The gravitational waves that are currently bathing our region of the Universe carry with them invaluable information that is unattainable from electromagnetic observations or high-energy cosmic ray detections. We will briefly outline the value of gravitational wave astronomy before moving into the details of detection. For more thorough background there exists a variety of review articles including Thorne (1987) [18] and Will (1993) [19] .

Throughout history, revolutions in astronomy, and ultimately humanity's understanding of the Universe, have resulted from new techniques for collecting photons. Galileo's telescope magnified the night sky expanding the size of the cosmos and dethroning the Earth as the center of the Universe. Edwin Hubble's observations showed that our own galaxy is just one of countless others and that the Universe is in fact expanding, seeding the field of cosmology. These accomplishments, while shifting paradigms, were made solely utilizing the visible light portion of the electromagnetic spectrum.

Access to other frequencies brought further, previously unimaginable, discoveries. To name a few: Microwave telescopes provided powerful evidence in support of the Big Bang model for the origin of the Universe, gamma ray observatories revealed the most powerful explosions in the Universe, infrared telescopes have allowed us to peer deep into the heart of the galaxy showing the birth of stars and ultimately the

formation of solar systems like our own, and X-ray telescopes verified the existence of such exotic objects as black holes. Without folding in the notable contributions from particle astrophysics, nearly everything we know about the Universe comes to us via light. The inclusion of a new messenger into our arsenal, gravitational waves, will incalculably expand our bank of astrophysical knowledge.

According to General Relativity, gravitational waves should be a ubiquitous presence in the Universe. Unfortunately, the effect of these waves borders on the infinitesimal. Space-time is very "stiff" and the amplitude of the disturbances is so incredibly small that they have evaded direct detection. Despite this observational hardship, extremely compelling and indirect evidence for the existence of GWs comes from the long term monitoring of the binary system in the galaxy containing pulsar PSR 1913+16 – the so called "Hulse-Taylor Pulsar" [20]. The orbital decay of this system, monitored over decades, precisely matches that which is predicted by General Relativity via the loss of energy and angular momentum due to the emission of gravitational waves [21]. Spurred by this evidence, and the incredible potential for making new discoveries, the design and construction of gravitational wave detectors has grown in sophistication such that we are now on the threshold of the first direct detection and the dawn of gravitational wave astronomy.

Like its electromagnetic counterpart, the gravitational wave spectrum spans several decades in frequency with each regime hosting unique information. Gravitational wave detectors are typically broad-band, all-sky instruments allowing us to achieve significant coverage of the entire spectrum with a comparatively small number of observatories. Figure 1 was compiled by Sterl Phinney for NASA's Beyond Einstein Program and depicts some possible sources of gravitational radiation as well as the instruments sensitive to those frequency regimes.

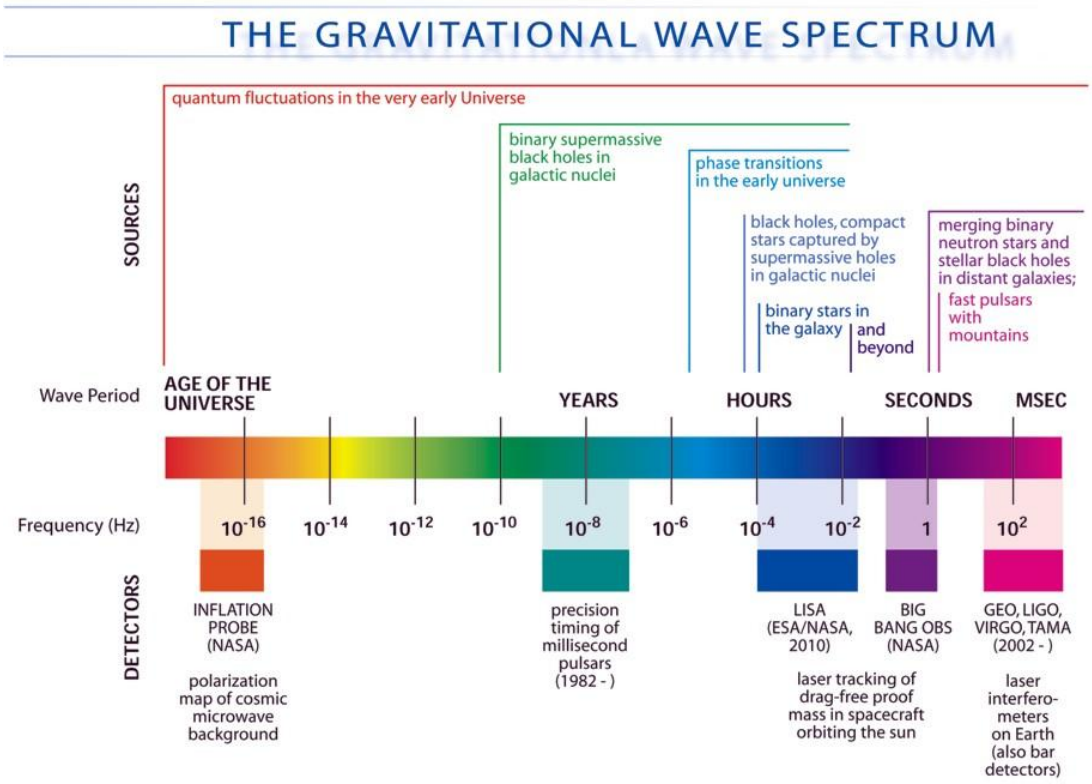


Figure 1: The gravitational wave spectrum with sources for different decades in frequency as well as current and future experiments which will probe certain frequency regimes.

A list of potential discoveries to be made by gravitational wave astronomy reads as a "who's who" of astrophysical and cosmological mysteries. What we hope to learn from GW astronomy ranges from the grandiose (details of cosmological inflation and dark energy) to the detailed (such as strong field tests of General Relativity). An abbreviated list of topics in astrophysics and cosmology targeted by gravitational wave detectors includes

- the formation and evolution of galaxies and black holes,
- the mechanism behind core-collapse supernovae,

- the validity with which General Relativity predicts the curvature of space-time in the presence of strong gravitational fields,
- the physics of the inflationary period of the Universe,
- the nature of dark energy, and
- the equation of state of neutron stars.

A more thorough discussion of the potential discoveries that can be made via the detection of gravitational waves can be found in Cutler and Thorne (2002) [22].

To make these contributions to physics and astronomy, dedicated gravitational wave "observatories" have been constructed around the world and will one day be flown in space. The word "observatories" is in quotes because GW detectors do not *observe* the Universe in a manner similar to telescopes. Gravitational wave detectors do not point or focus, and are more aptly thought of as antennae, responding to passing GWs from every direction at once. If telescopes are the "eyes" of astronomy then GW detectors are analogous to the "ears." Much like our own auditory senses, we are still able to determine where a particular gravitational wave signal is coming from, albeit without the precision of a telescope. However, what we lose in precision we gain in all-sky coverage.

Gravitational wave astronomy is a very young field and while the engineering technologies have recently achieved the sensitivities to realistically anticipate a positive detection, there are still many analysis obstacles which have not yet been surmounted. Unlike conventional astronomy, where the data can be used to construct an image (something our brains are excellent at interpreting), the output of a gravitational wave detector is a time series of length measurements. Contained in a gravitational waveform which we hope to extract from the data is precise information about the

emitter and decoding this information allows us to accurately infer characteristics about the emitting source unavailable to electromagnetic observations. An example of such quantities which we hope to measure are the individual masses of a binary black hole system to better than one percent accuracy. Doing so offers a unique set of signal processing challenges which must be addressed in order to maximize what can be learned from these instruments.

While many of the analysis techniques for making GW detections are already in place, still lacking is a robust method for distinguishing between a *weak* gravitational wave signal and instrument noise. The current state of the art is to use Monte Carlo simulations of data to establish a detection threshold based on a permissible false alarm rate. We will show that this approach is sub-optimal as the "detectability" of a gravitational wave signal embedded in instrument noise significantly depends on the particular noise realization with which the GW is competing. Our proposed solution is to abandon the thresholding approach in favor of a case-by-case evaluation of potential GW events. A candidate detection is assessed by calculating the odds that the data contains a signal that is cosmic in origin versus containing strictly instrument noise. While this is a common approach to model selection (or hypothesis testing) it has only recently gained favor in the GW community.

Prior to detailing this unique approach to GW data analysis we will briefly introduce a subset of the current and future GW detectors, and outline how these instruments respond to impinging gravitational waves. This will lay the foundation for using the subsequent data to answer some of the questions accessible to GW astronomy.

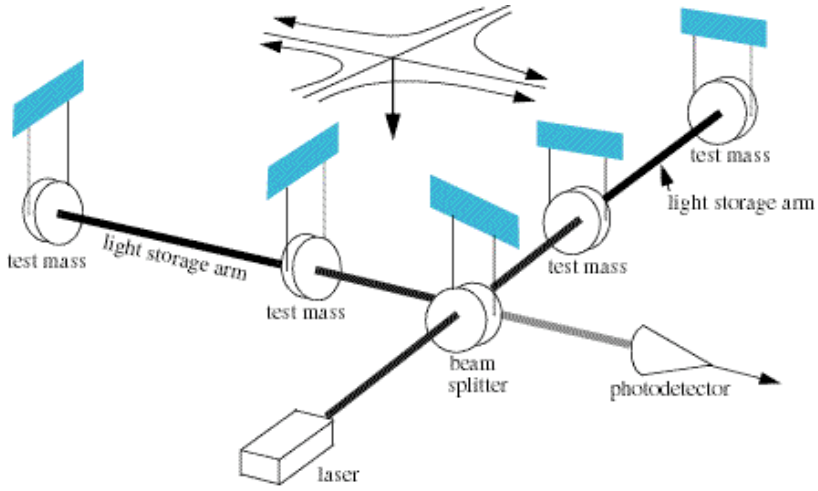
Gravitational Wave Detectors

Making gravitational wave astronomy a reality requires instruments of fantastic sensitivity. There are three primary approaches to GW observation: Interferometers, pulsar timing, and resonant mass bar detectors. While bar detectors marked the first attempts at making GW measurements, it is the interferometers that have achieved the highest sensitivities and thus have the highest probability (in the immediate future) of making the first confirmed detections. An intriguing alternative to constructing detectors is found in the precise monitoring of millisecond pulsar timing residuals which can indicate the shrinking and stretching of the space-time between us and the pulsar due to passing gravitational waves. While pulsar timing could potentially yield the first GW discovery we will focus our attention on the kilometer-scale ground-based interferometers, particularly the LIGO detectors, and the proposed space-based interferometer, LISA.

The Laser Interferometer Gravitational wave Observatory (LIGO)

LIGO is the state of the art for operational gravitational wave detectors. The instrument is a large "L" shaped Fabry-Perot-Michelson interferometer designed to detect passing gravitational waves by measuring the distance between the mirrors of the resonant cavity (see Figure 2a). In order to make the precision measurements necessary for gravitational wave detection, the entire interferometer must operate under vacuum. There are two sites in the United States which are host to LIGO interferometers. The site in Hanford, Washington (LHO) is actually two interferometers, referred to as H1 and H2, with arm lengths of four kilometers and two kilometers, respectively. Each share the same vacuum system and should respond identically to passing gravitational waves. The second site in Livingston, Louisiana (LLO) is home

to the third interferometer (L1) identical in design to H1. Multiple interferometers of similar capabilities are vital because any candidate detection would necessarily be coincident (up to the light travel time between the detectors) at both locations. This allows the analysis efforts to veto a large fraction of transient noise events that are an inevitable outcome for instruments of such unprecedented sensitivity located on Earth.



(a) The Fabry-Perot-Michelson interferometer concept.

(b) LHO in Hanford, Washington

(c) LLO in Livingston, Louisiana

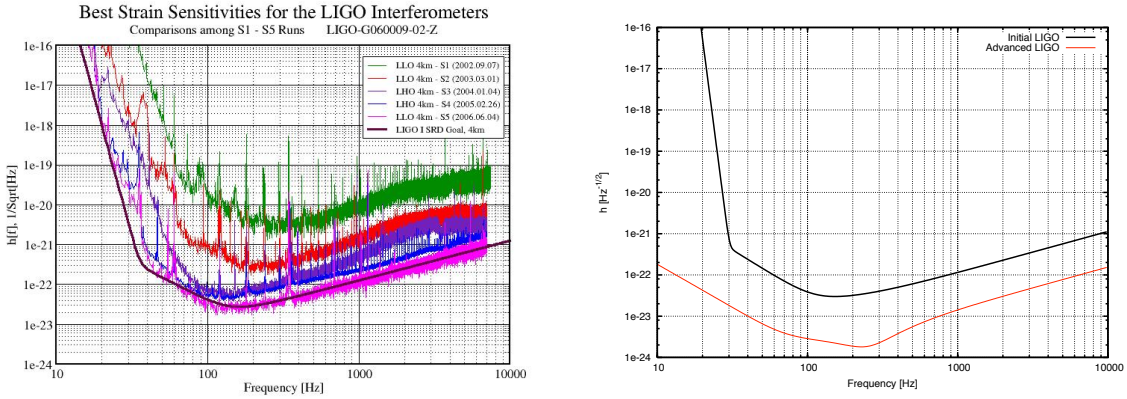
Figure 2: The LIGO design (a) with aerial photographs of the LHO (b) and LLO (c) cites.

The LIGO Scientific Collaboration (LSC) [23] is a joint effort between LIGO and two sub-kilometer interferometers, GEO 600 [24] and TAMA300 [25]. In 2007 the LSC entered into a data sharing agreement with the French-Italian three kilometer Virgo [6] interferometer creating a world-wide network of cooperating gravitational wave detectors. Adding additional detectors to the network strengthens the coincidence tests while improving source localization, as each individual detector has rather poor sky resolution.

To date, LIGO has successfully completed five science collecting runs (S1 through S5) with commissioning phases between each data acquisition period to increase the sensitivity of the instruments (see Fig. 3a). S5, which produced one year of triple coincidence data between H1, H2, and L1 was completed in the autumn of 2007 and has the distinction of being the first LIGO data collected at design sensitivities. At the time of this thesis, the analysis of the S5 data is still ongoing while attention is now turning towards the sixth science run (S6) due to begin sometime during 2009. S6 will follow the Enhanced LIGO commissioning that will have increased the sensitivity to gravitational waves by roughly a factor of two to three, thereby increasing the volume of the "audible" Universe by a factor between 8 and 30. Enhanced LIGO will run in consort with a similarly improved Virgo detector in its second science run (VSR2).

The sensitivity of these instruments is remarkable. A "strong" and therefore "easily" detectable gravitational wave will produce a differential length measurement, or strain, in the detector of $\Delta L/L \sim 10^{-21}$ requiring sensitivity to changes in length of about one thousandth the width of a proton. Yet these design specifications have been met and are now being improved upon. After the conclusion of S6/VSR2, LIGO and Virgo will both undergo major upgrades which will produce the advanced generation of gravitational wave detectors with sufficient sensitivity and dynamic range to make frequent detections, truly ushering in the era of gravitational wave astronomy.

The improvements that will be made between Initial LIGO and Advanced LIGO include increased bandwidth, and provide an overall factor of at least ten improvement in sensitivity across all frequencies. Figure 3b depicts the theoretical noise levels for Initial and Advanced LIGO. This improved sensitivity effectively increases the volume of the audible Universe such that a few hours of Advanced LIGO operation will have the same detection probability as observations made over a one year integration of Initial LIGO data.



(a) (b)

Figure 3: The progress of the sensitivity to gravitational waves by the LIGO detectors. **a)** The measured strain sensitivity curves for Initial LIGO spanning S1 to S5 [green to magenta] compared to the design sensitivity [black]. **b)** Theoretical noise power spectral density for Initial and Advanced LIGO.

LIGO and other ground based GW detectors are limited at low frequencies by seismic noise and thus focus their attention on the high frequency band of the gravitational wave spectrum, from ~ 10 Hz to ~ 10 kHz. Some events which are targeted by LIGO include gravitational wave emission by nearby core-collapse supernova, binary neutron star and stellar mass black hole binary mergers, the subsequent ring-down from black hole formation, near-by pulsars, and cusps or kinks propagating along

cosmic strings. A great deal of effort and innovation has also gone towards the identification and characterization of GW "bursts" – un-modeled, short duration signals – the detection of which is key to unlocking the discovery potential of gravitational wave astronomy.

The Laser Interferometer Space Antenna (LISA)

It is desirable to overcome the low-frequency noise wall attributed to seismic activity on Earth in order to access a frequency regime rich in gravitational wave signals. To do so we must leave the noisy planet and place a gravitational wave detector in space. LISA, a joint endeavor between the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA) [7], is a proposed space-based interferometer composed of three satellites configured as an equilateral triangle with legs of length 5×10^9 meters. This translates to sensitivities of $\sim 10^{-11}$ m in relative displacement of the spacecraft – less than the width of an atom but a much simpler task than is faced by LIGO. This, however, does not mean that LISA is less sensitive to gravitational waves. Since gravitational waves produce a strain the longer baseline of LISA translates to a markedly more substantial effect on the distance between the test-masses.

The LISA constellation will trail the Earth in its orbit by 20^o. Each satellite's orbital inclination will be such that the entire triangle will "tumble" in its path around the Sun causing the instrument's antenna pattern to sweep across the entire celestial sphere, making LISA an all-sky detector over the course of one year. The individual satellites will be approximately two meters across and will house a pair of test masses, each four centimeters on a side, as well as the appropriate optical system to precisely monitor the distance between itself, the two test masses, and the other two satellites.

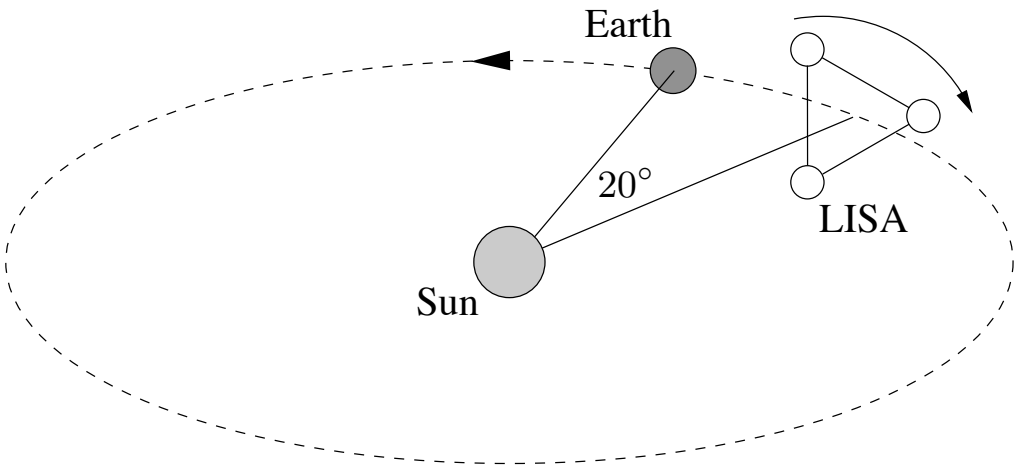


Figure 4: The configuration of the LISA orbit. LISA will trail the Earth by 20°, 1 AU from the Sun. Each satellite will have a different orbital inclination allowing the constellation's configuration to rotate throughout the year making LISA an all-sky instrument.

The relative distance measures between each satellite will be combined to form three coupled Michelson-type interferometer channels.

LISA, with interferometer arms about 10^6 times larger than those of LIGO, is a low frequency detector with sensitivity between 0.1 and 10 mHz. This region of the gravitational wave spectrum, particularly between 10^{-1} and 10 mHz, is dominated by compact galactic binaries [28, 31] which will warrant considerable attention in this thesis. The most compelling scientific results from LISA will emerge from cosmological sources including the merger of massive black holes (MBHs) which are theorized (and supported by mounting evidence) to exist at the cores of galaxies, the in-spiral and capture of stellar mass black holes by MBHs (i.e. extreme mass ratio inspirals or "EMRIs"), and perhaps the cosmological background of gravitational waves which would have been produced during the inflationary period of the Universe, thus

supplying us with information closer in time to the Big Bang than is possible from electromagnetic measurements.

The technologies required to successfully fly an interferometer with the desired sensitivity are orders of magnitude beyond what has previously been flown and have posed a formidable engineering challenge. Despite these obstacles, a demonstration of these key technologies, the LISA pathfinder mission [8], is due to launch in 2010. Furthermore, the *analysis* of a time series so densely populated by gravitational waves has presented a new set of challenges for gravitational wave astronomy and preparation for this onslaught of data is a key component to the mission's flight readiness. This need has spawned a set of iteratively more sophisticated tests, the Mock LISA Data Challenges [26], each a collection of simulated data sets containing blind injections of signals designed to address some of the more compelling challenges and rewards of the LISA mission.

Gravitational Wave Data Analysis

Coupled with the excitement of collecting information about the Universe from a new messenger is the technological challenge of decoding the message. The existence of programs such as the MLDC highlights the need to develop and demonstrate analysis methods for taking the output of a gravitational wave instrument (a noisy time series) and inferring astrophysical information. This vital cog in the machinery for successful gravitational wave observatories has spawned the active field in physics research of gravitational wave data analysis. What follows is a very basic introduction to the interaction between passing gravitational waves and the machines commissioned to detect them. Forward modeling of the detector response to incident

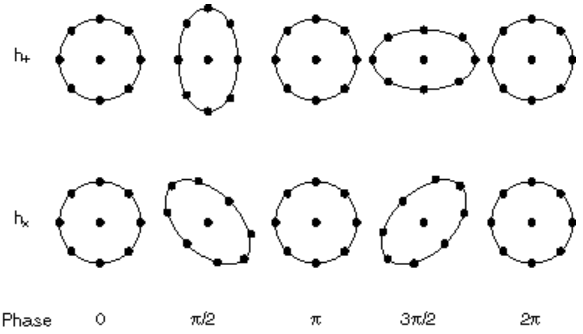


Figure 5: The effect of passing gravitational waves on a ring of test masses. The top and bottom rows are for the plus and cross polarizations, respectively. One can qualitatively see how monitoring the relative distance between three test masses will reveal the effect of the passing gravitational waves.

radiation forms the foundation of the matched filtering analysis techniques on which gravitational wave astronomy so heavily depends.

The coordinate frame of an arbitrary gravitational wave is defined by the propagation direction, $\hat{k}$, and the metric tensor of the gravitational wave can be written at some location $\vec{x}$ as the linear sum of two independent polarization states

$$\mathbf{h}(\xi) = h_+(\xi)\varepsilon^+ + h_\times(\xi)\varepsilon^\times \tag{1}$$

where the wave variable $\xi = t - \hat{k} \cdot \vec{x}$ gives the surfaces of constant phase. We take as the origin some reference location at the detector (e.g., the vertex of a LIGO-like interferometer). If the wavelength is much larger than the detector (often a good approximation), $\hat{k} \cdot \vec{x}$ can be neglected, making $\xi \sim t$. The effect that each polarization of a passing gravitational wave would have on a ring of acceleration-free masses is illustrated in Fig. 5. From this cartoon one can see how monitoring the relative distances between any two points along the ring relative to a common vertex can reveal the GW's presence.

The polarization tensors can be expanded in terms of the basis tensors $\mathbf{e}^+$ and $\mathbf{e}^\times$ as

$$\begin{aligned} \varepsilon^+ &= \cos(2\psi)\mathbf{e}^+ - \sin(2\psi)\mathbf{e}^\times \\ \varepsilon^\times &= \sin(2\psi)\mathbf{e}^+ + \cos(2\psi)\mathbf{e}^\times \,, \end{aligned} \tag{2}$$

where ψ is the polarization angle and

$$\begin{aligned} \mathbf{e}^+ &= \hat{u}\otimes\hat{u} - \hat{v}\otimes\hat{v} \\ \mathbf{e}^\times &= \hat{u}\otimes\hat{v} + \hat{v}\otimes\hat{u}\,. \end{aligned} \tag{3}$$

The vectors $(\hat{u}, \hat{v}, \hat{k})$ form an orthonormal triad which may be expressed as a function of the source location on the celestial sphere (θ, ϕ) as

$$\begin{aligned} \hat{u} &= \cos\theta\cos\phi\,\hat{x} + \cos\theta\sin\phi\,\hat{y} - \sin\theta\,\hat{z} \\ \hat{v} &= \sin\phi\,\hat{x} - \cos\phi\,\hat{y} \\ \hat{k} &= -\sin\theta\cos\phi\,\hat{x} - \sin\theta\sin\phi\,\hat{y} - \cos\theta\,\hat{z}\,. \end{aligned} \tag{4}$$

The basis vector $\hat{k}$ is anti-parallel to the line of sight from the detector to the source, and $\hat{u}$ and $\hat{v}$ define the plane in which the space-time is being distorted.

To reveal how this perturbation will influence an interferometer we must contract the gravitational wave (Eq. 1) with the *detector tensor*

$$\mathbf{d}(f, t, \hat{k}) = \mathbf{D}(t)\mathcal{T}(f, t, \hat{k}) \tag{5}$$

where $\mathbf{D}$ is the interferometer arm tensor. The transfer function $\mathcal{T}$ relates the output of the detector to its input, akin to the point-spread-function of a telescope, and is specific to each instrument design. For a simple 90° interferometer with arms in the $\hat{x}$ and $\hat{y}$ directions, $\mathbf{D} = \hat{x}\otimes\hat{x} - \hat{y}\otimes\hat{y}$ [27]. We can now write the detector output to an incident gravitational wave as

$$h(t) = \mathbf{d}(f, t, \hat{k}) : \mathbf{h}(t) \tag{6}$$

where the double contraction $\mathbf{a} : \mathbf{b} \equiv a^{ij}b_{ij}$ using the Einstein summation convention (repeated indices – one superscripted and one subscripted – are summed). This is commonly rewritten by defining the antenna, or beam pattern, functions $F_{+,\times} \equiv \mathbf{d} : \varepsilon^{+,\times}$ which can take values between zero and one. Basically, F_A^2 is the fraction of the incident gravitational wave power, with polarization A, which is transmitted to the detector output. F_+ and $F_\times$ are functions of the sky location of the source and, depending on convention, time and the polarization angle ψ. Eq. 6 can then be further simplified into its more recognizable form

$$h(t) = F_+ h_+(t) + F_\times h_\times(t). \tag{7}$$

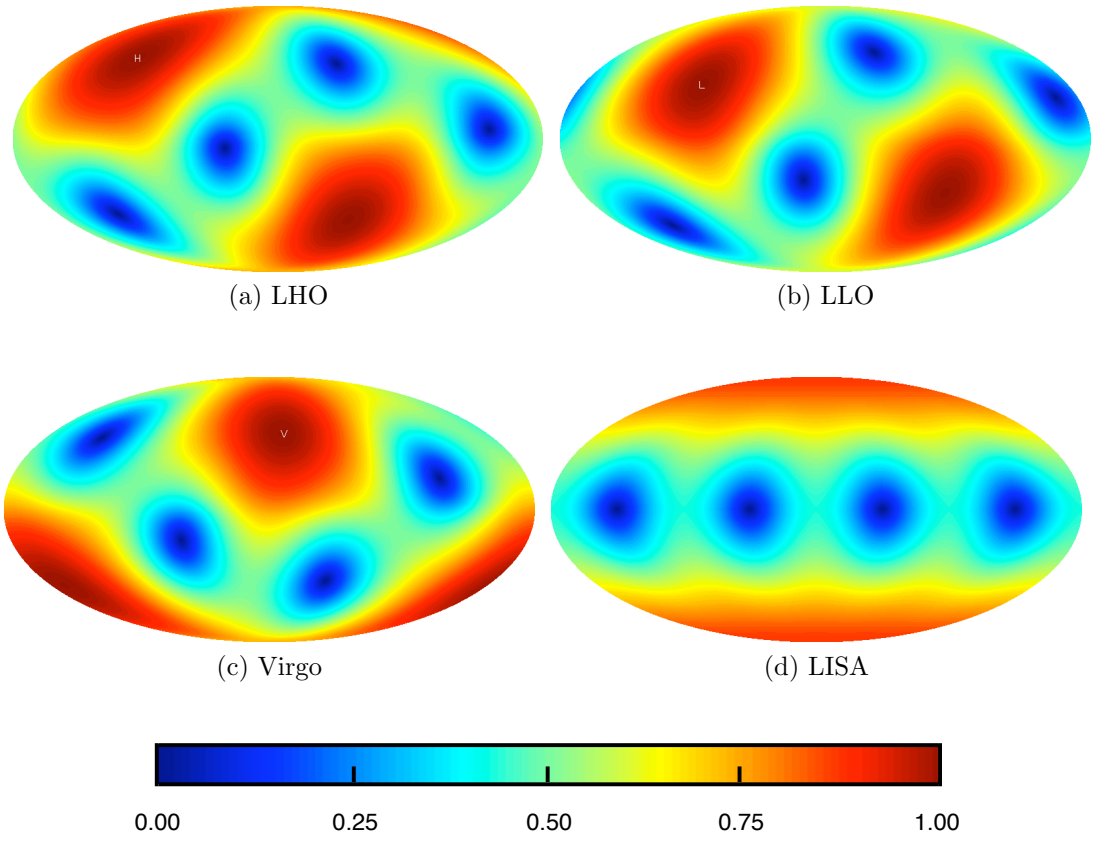
The antenna patterns for LHO, LLO, Virgo, and LISA are depicted in Fig. 6 As the detectors go through their orbital motion the antenna patterns sweep across the entire sky.

Ideally, the instrument noise that will necessarily accompany the gravitational wave signal is the result of uncorrelated stochastic processes which result in stationary, normally distributed noise written in the Fourier domain as $\tilde{n}(f)$. While this may seem an unrealistic assumption, the noise in the LIGO detectors (after conditioning) very nearly satisfies these conditions, apart from occasional, large-amplitude, transient events which will later warrant attention. Assuming that the above conditions are met, the noise can be uniquely determined by its mean which, without loss of generality, is taken to be $\langle \tilde{n}(f) \rangle = 0$ and its autocorrelation $R(\tau) \equiv \langle n(t+\tau)n(t) \rangle$ which we use to define the one-sided noise power spectral density

$$\frac{1}{2} S_n(f) \equiv \int_{-\infty}^{\infty} d\tau \; R(\tau) e^{2\pi i f \tau}. \tag{8}$$

By inverse Fourier transforming Eq. 8 we see that

$$R(\tau) \equiv \langle n(t+\tau)n(t) \rangle = \frac{1}{2} \int_{-\infty}^{\infty} df \; S_n(f) e^{2\pi i f \tau} \tag{9}$$



(a) LHO (b) LLO

(c) Virgo (d) LISA

Figure 6: Antenna patterns for different interferometers averaged over all GW polarization angles. The sky maps for LHO, LLO, and Virgo are as seen from Earth in equatorial coordinates with the location of each ground-based interferometer marked on its respective antenna pattern. The equator for the LISA sky map corresponds to the ecliptic plane.

and in particular, for $\tau = 0$,

$$\langle n^2(t) \rangle = \int_0^\infty df \; S_n(f). \tag{10}$$

This is more commonly encountered in the Fourier domain with samples $\tilde{n}(f)$ drawn over a finite time interval T from a normal distribution with zero mean and variance

$$\langle |\tilde{n}(f)|^2 \rangle = \frac{1}{2} S_n(f) T. \tag{11}$$

We have generically outlined the steps necessary to model the output $s(t) = h(t) + n(t)$ of a gravitational wave detector to some arbitrary, low-frequency gravitational wave signal $h(t)$ and ideal stationary gaussian noise $n(t)$. What follows is a practical implementation that we have developed which takes advantage of a fast-slow decomposition to produce an algorithm which can rapidly simulate waveforms from compact galactic binaries emitting in the LISA band. This is of great utility as it allows for detailed treatment of the galactic foreground without requiring extensive computational resources.

Modeling the LISA Response

The galaxy is predicted to contain millions of binary systems composed of compact stellar objects such as white dwarf and neutron stars. Here we outline a technique to rapidly simulate the response of the LISA detector to such a signal. The simulation takes advantage of the predominantly monochromatic nature, and subsequently slow orbital evolution, of such signals to produce template waveforms at very low computational cost. This then allows for accurate simulation of the entire galactic foreground in a reasonable amount of time (a few hours on a single GHz processor) opening the door to detailed studies of how source confusion noise will affect LISA data analysis. For more information on the galactic foreground problem see [28, 29, 30, 31, 32] and chapters 4 and 6.

Determining the response of a gravitational wave detector boils down to reconciling three different reference frames that are convenient for solving the three different phases of the problem: The source frame where the gravitational waves are extracted, the propagation frame of the gravitational waves, and the detector frame. For LISA

we choose as the origin of the detector frame the Solar System barycenter (SSB) with a z-axis pointed along the angular momentum vector of Earth's orbit.

To leading order in the eccentricity, e, the time dependent Cartesian coordinates of the $i^{\rm th}$ LISA spacecraft relative to the SSB are given by [33]

$$\begin{aligned} x_i(t) &= R\cos(\alpha) + \frac{1}{2}eR\Big(\cos(2\alpha-\beta_i) - 3\cos(\beta)\Big) \\ y_i(t) &= R\sin(\alpha) + \frac{1}{2}eR\Big(\sin(2\alpha-\beta_i) - 3\sin(\beta)\Big) \\ z_i(t) &= -\sqrt{3}eR\cos(\alpha-\beta_i)\,. \end{aligned} \tag{12}$$

$R = 1$ AU is the radial distance from the SSB to the guiding center of the LISA constellation, $\alpha = 2\pi f_m t + \kappa$ is the orbital phase of the guiding center with $f_m \sim 1\times 10^{-8}$ Hz being the modulation frequency of the orbit, and $\beta_i = 2\pi(i-1)/3+\lambda$ with $i = 1,2,3$ is the relative phase of the spacecraft within the constellation. The parameters κ and λ give the initial ecliptic longitude and orientation of the constellation. The distance between the spacecraft is $L = 2\sqrt{3}eR$. Setting $e = 0.00985$ results in the desired spacecraft separation of $L = 5\times 10^9$ m.

For mildly chirping binary sources, the gravitational wave tensor is simply

$$\mathbf{h}(\xi) = \Re\Big[\left(A_+\varepsilon^+ + e^{i3\pi/2}A_\times\varepsilon^\times\right)e^{i\Psi(\xi)}\Big] \tag{13}$$

where

$$\begin{aligned} A_+ &= \frac{2\mathcal{M}(\pi f)^{2/3}}{D_L}\left(1+\cos^2\iota\right) \\ A_\times &= -\frac{4\mathcal{M}(\pi f)^{2/3}}{D_L}\cos\iota\,. \end{aligned} \tag{14}$$

Here $\mathcal{M}$ is the chirp mass defined in terms of the binary's individual masses m_1 and m_2 as

$$\mathcal{M} = \frac{(m_1m_2)^{\frac{3}{5}}}{(m_1+m_2)^{\frac{1}{5}}}\,, \tag{15}$$

D_L is the luminosity distance and ι is the inclination of the binary's angular momentum vector to the line of sight. Higher post-Newtonian corrections, eccentricity of the orbit, and spin effects will introduce additional harmonics; however, low-frequency binaries (such as those which will be in the LISA band) are expected to have orbits which have predominantly been circularized, while spin effects for stellar mass compact binaries are negligible. For chirping sources we wish to evolve the orbit adiabatically, requiring that the frequency evolution $\dot{f}$ occurs on a timescale long compared to the light travel time in the interferometer: $f/\dot{f} \ll L$. Under such conditions the gravitational wave phase can be approximated as

$$\Psi(\xi) = 2\pi f_0 \xi + \pi \dot{f}_0 \xi^2 + \varphi_0 \,, \tag{16}$$

where f_0, $\dot{f}_0$, and φ_0 are the initial frequency, df/dt and phase, respectively. The instantaneous frequency is given by $2\pi f = \partial_t \Psi$:

$$f = (f_0 + \dot{f}_0 \xi)(1 - \hat{k} \cdot \mathbf{v}) \,. \tag{17}$$

In general, the expression for the path length variation between two test masses i and j caused by a passing gravitational wave involves an integral in ξ from ξ_i to ξ_j. Writing $\xi = \xi_i + \delta\xi$ we have

$$\Psi(\xi) \simeq 2\pi (f_0 + \dot{f}_0 \xi_i)\delta\xi + \text{const} \,. \tag{18}$$

Thus, we can treat the wave as having fixed frequency $f_0 + \dot{f}_0 \xi_i$ for the purposes of the integration, and then increment the frequency forward in time in the final expression [34]. The path length variation is then given by [33, 34]

$$\delta\ell_{ij}(\xi) = L \Re\Big[\mathbf{d}(f, t, \hat{k}) : \mathbf{h}(\xi)\Big] \,, \tag{19}$$

where $\mathbf{a} : \mathbf{b} = a^{ij} b_{ij}$. The one-arm detector tensor is given by

$$\mathbf{d}(f, t, \hat{k}) = \frac{1}{2}\Big(\hat{r}_{ij}(t) \otimes \hat{r}_{ij}(t)\Big) \mathcal{T}(f, t, \hat{k}) \,, \tag{20}$$

with $\hat{r}_{ij}(t)$ being the unit vector from spacecraft i to j. The LISA transfer function is

$$\begin{aligned}\mathcal{T}(f,t,\hat{k}) &= \mathrm{sinc}\left(\frac{f}{2f_*}\Big(1-\hat{k}\cdot\hat{r}_{ij}(t)\Big)\right)\\ &\times \exp\left(i\frac{f}{2f_*}\Big(1-\hat{k}\cdot\hat{r}_{ij}(t)\Big)\right), \end{aligned} \tag{21}$$

where $f_* = 1/(2\pi L)$ is the LISA transfer frequency and $f = f_0 + \dot{f}_0\xi$. The expression can be attacked in pieces. It is useful to define the quantities

$$d^+_{ij}(t) \equiv (\hat{r}_{ij}(t)\otimes\hat{r}_{ij}(t)) : \mathbf{e}^+ \tag{22}$$

$$d^\times_{ij}(t) \equiv (\hat{r}_{ij}(t)\otimes\hat{r}_{ij}(t)) : \mathbf{e}^\times . \tag{23}$$

and $y_{ij}(t) = \delta\ell_{ij}(t)/(2L)$. Then

$$y_{ij}(t) = \Re\left[y^{\mathrm{slow}}_{ij}(t)e^{2\pi i f_0 t}\right], \tag{24}$$

where

$$\begin{aligned} y^{\mathrm{slow}}_{ij}(t) &= \frac{\mathcal{T}(f,t,\hat{k})}{4}\Big(\big(d^+_{ij}(t)(A_+(t)\cos(2\psi)\\ &+ e^{3\pi i/2}A_\times(t)\sin(2\psi))\\ &+ d^\times_{ij}(t)(e^{3\pi i/2}A_\times(t)\cos(2\psi)\\ &- A_+(t)\sin(2\psi))\big)\, e^{(\pi i\dot{f}_0\xi^2 + i\varphi_0 - 2\pi i f_0\hat{k}\cdot\mathbf{x})}\Big) \end{aligned} \tag{25}$$

In the Fourier domain the response can be written as

$$y_{ij}(t) = \Re\left[\left(\sum_n a_n e^{2\pi i n t/T_{\mathrm{obs}}}\right)e^{2\pi i f_0 t}\right], \tag{26}$$

where the coefficients a_n can be found by a numerical FFT of the slow terms $y^{\mathrm{slow}}_{ij}(t)$. The number of time samples needed in the FFT will depend on f_0, $\dot{f}_0$ and T_{obs}, but is less than $2^9 = 512$ for any galactic sources we are likely to encounter when $T_{\mathrm{obs}} \leq 2$ years (the assumed mission lifetime).

To ensure sufficient sampling we estimate the bandwidth of a source as

$$B = 2\left(4 + 2\pi f_0 R\sin(\theta)\right) f_m + \dot{f}_0 T_{\rm obs} \tag{27}$$

and require the number of samples to exceed $2BT_{\rm obs}$. The Fourier transform of the fast term can be done analytically:

$$e^{2\pi i f_0 t} = \sum_m b_m e^{2\pi i m t/T_{\rm obs}} \tag{28}$$

where

$$b_m = T_{\rm obs}\,{\rm sinc}(x_m) e^{i x_m} \tag{29}$$

and

$$x_m = f_0 T_{\rm obs} - m\,. \tag{30}$$

The cardinal sine function in Eq. 29 ensures that the Fourier components b_m away from resonance, $x_m \approx 0$, are quite small. It is only necessary to keep $\sim 100 \rightarrow 1000$ terms either side of $p = [f_0 T_{\rm obs}]$, depending on how bright the source is, and how far $f_0 T_{\rm obs}$ is from an integer. We now have

$$y_{ij}(t) = \Re\left[\left(\sum_j c_j e^{2\pi i j t/T_{\rm obs}}\right)\right] \tag{31}$$

where

$$c_j = \sum_n a_n b_{j-n}\,. \tag{32}$$

The final step is to ensure that our Fourier transform yields a real $y_{ij}(t)$. This is done by setting the final answer for the Fourier coefficients equal to $d_j = (c_j + c^*_{-j})/2$. But since x_m never hits resonance for positive j (we are not interested in the negative frequency components $j < 0$), we can neglect the second term and simply write $d_j = c_j/2$.

Basically what we are doing is heterodyning the signal to the base frequency f_0, then Fourier transforming the slowly evolving heterodyned signal numerically. We

then convolve these Fourier coefficients with the analytically derived Fourier coefficients of the carrier wave.

With the differential arm lengths in hand, we can construct the Michelson type time delay interferometry (TDI) variables [35]

$$\begin{aligned} X(t) &= y_{12}(t-3L) - y_{13}(t-3L) + y_{21}(t-2L) \\ &- y_{31}(t-2L) + y_{13}(t-L) - y_{12}(t-L) \\ &+ y_{31}(t) - y_{21}(t), \end{aligned} \tag{33}$$

$$\begin{aligned} Y(t) &= y_{23}(t-3L) - y_{21}(t-3L) + y_{32}(t-2L) \\ &- y_{12}(t-2L) + y_{21}(t-L) - y_{23}(t-L) \\ &+ y_{12}(t) - y_{32}(t), \end{aligned} \tag{34}$$

$$\begin{aligned} Z(t) &= y_{31}(t-3L) - y_{32}(t-3L) + y_{13}(t-2L) \\ &- y_{23}(t-2L) + y_{32}(t-L) - y_{31}(t-L) \\ &+ y_{23}(t) - y_{13}(t). \end{aligned} \tag{35}$$

Note that in the Fourier domain

$$\begin{aligned} X(f) &= \tilde{y}_{12}(f)e^{-3if/f*} - \tilde{y}_{13}(f)e^{-3if/f*} + \tilde{y}_{21}(f)e^{-2if/f*} \\ &- \tilde{y}_{31}(f)e^{-2if/f*} + \tilde{y}_{13}(f)e^{-if/f*} - \tilde{y}_{12}(f)e^{-if/f*} \\ &+ \tilde{y}_{31}(f) - \tilde{y}_{21}(f)\,. \end{aligned} \tag{36}$$

This saves us from having to interpolate the differential arm lengths in the time domain as it is unlikely that a time shift of nL will land precisely on another time sample. Instead, we can combine phase shifted versions of our original Fourier transforms.

To complete the forward modeling we will assume stationary Gaussian instrument noise with no contribution from a confusion background. We also prefer to work with the noise orthogonal A, E, T data channels, which can be constructed from linear

combinations of the Michelson type X, Y, Z signals [35]:

$$\begin{aligned} A &= \frac{1}{3}\left(2X - Y - Z\right), \\ E &= \frac{1}{\sqrt{3}}\left(Z - Y\right), \\ T &= \frac{1}{3}\left(X + Y + Z\right). \end{aligned} \tag{37}$$

For galactic binaries the T channel can be neglected as it is insensitive to GWs at the frequencies we are considering. The noise spectral density is identical in the A and E channel and has the form

$$\begin{aligned} S_n(f) &= \frac{4}{3}\left(1 - \cos(2f/f_*)\right)\Big((2 + \cos(f/f_*))S_s \\ &+ 2(3 + 2\cos(f/f_*) + \cos(2f/f_*))S_a\Big)\ \mathrm{Hz}^{-1}. \end{aligned} \tag{38}$$

The acceleration noise S_a and shot noise S_s for LISA are simulated at the levels

$$\begin{aligned} S_a &= \frac{9 \times 10^{-30}}{(2\pi f)^4 L^2}\ \mathrm{Hz}^{-1}. \\ S_s &= \frac{10^{-22}}{L^2}\ \mathrm{Hz}^{-1} \end{aligned} \tag{39}$$

Armed with this prescription for generating mildly chirping galactic binaries we are able to efficiently simulate the 3×10^7 binaries in the galaxy allowing for rapid, detailed studies of the galactic foreground and how it will effect LISA's performance. We will now turn our focus to the task of producing an end-to-end Bayesian detection algorithm to analyze the data produced by LIGO or LISA, ultimately allowing us to answer some of the pressing astrophysical questions posed in chapter 2.

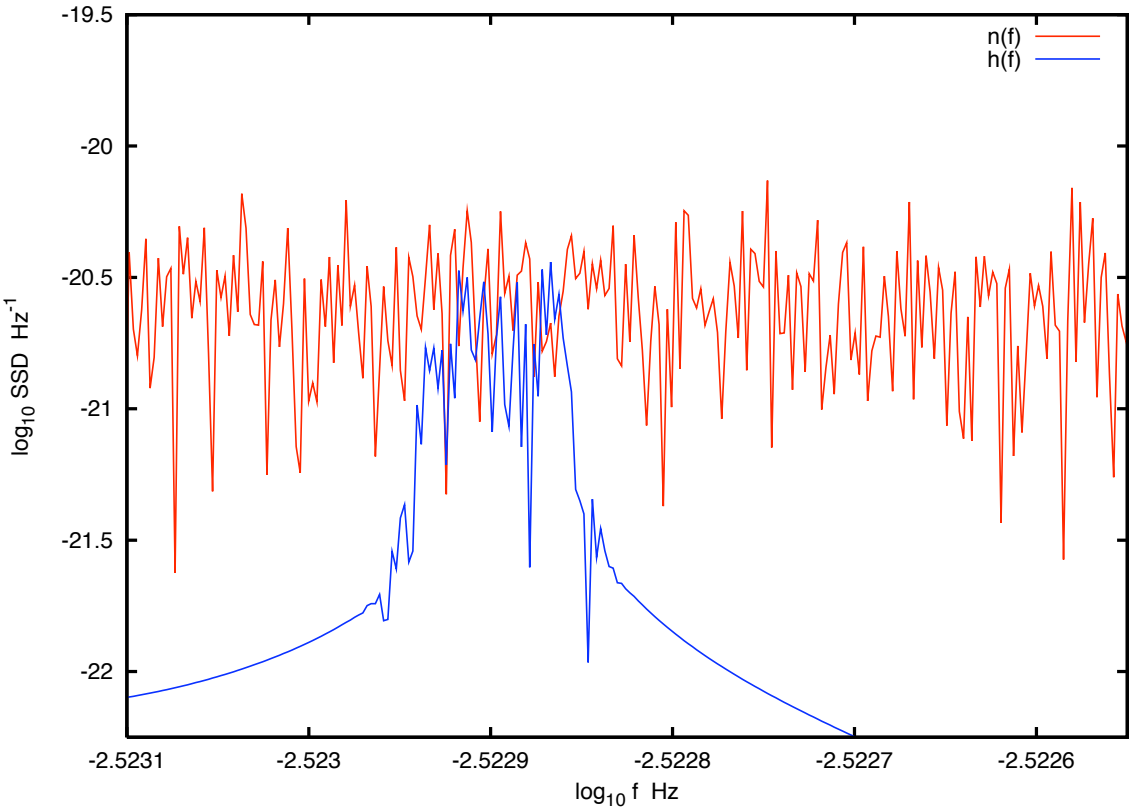


Figure 7: The A-channel LISA response with noise $|\tilde{n}(f)|$ [red] and galactic binary waveform $|\tilde{h}(f)|$ [blue]. This waveform was injected with signal-to-noise ratio of 8.

BAYESIAN PROBABILITY THEORY

If given a set of initial conditions, it is often possible to conclude via deduction what outcomes may be realized. Determining the chances of succeeding at a hand of poker is part of making good decisions about what, if anything, to wager. The goal for many branches of science is quite the opposite: to take an observed result and work backwards via inference towards a possible cause. Regrettably, the amount of information about the phenomenon in question is, to some level, incomplete. These gaps in our description of natural processes, however small, prevent us from ever absolutely assigning absolute truth to our hypothesis about the underlying mechanisms which drive the observed event. We need a reliable method for assessing how well a proposed hypothesis is supported by our observations. We can then learn about some process in question by iteratively collecting data and using this new information to determine if our hypothesis is supported by reality, updating our knowledge about the tested phenomenon.

Qualitatively assessing the completeness of a hypothesis is a subtle task, and the quest for such machinery led Reverend Thomas Bayes to develop a method of using the logic of deduction to devise a scheme for inference specifically for binomial distributions [36, 37], posthumously published in 1763. Bayes' method was later (and perhaps independently) generalized by Laplace to solve problems in celestial mechanics, medicine, and law [36]. Referred to as Bayesian Probability Theory, this formalism serves as the foundation for scientific logical inference, allowing us to algorithmically compute the relative probabilities between two or more hypotheses and thereby asses their relative validity without ambiguity.

This is in stark contrast to the more common statistical approach to inference, where each hypothesis is treated individually. Statistical inference tells us if the data succeeds or fails to reject a hypothesis. This differs from making an assessment about a hypothesis' *validity*. Statements of truth can only be made when all but a single hypothesis have been rejected. While this discrepancy may seem rather semantic the implementation of the opposing methods is vastly different and is subject to disagreement. See Gregory (2006) [37] for a very clear comparative discussion between the two approaches.

The Bayesian approach to inference defines the probability $p(A|B)$ as the plausibility of a proposition A, given the truth of the information represented by B. Valid manipulations of these probabilities include the product rule:

$$\begin{aligned} p(A,B|C) &= p(A|C)p(B|A,C) \\ &= p(B|C)p(A|B,C) \end{aligned} \tag{40}$$

where a comma indicates both propositions are true, so A, B means A *and* B are valid. Propositions to the right of the vertical bar are assumed to be true making $p(A,B|C)$ the joint probability of A and B being true given, or conditional on, the truth of C. Bayes' theorem is then written using the product rule as

$$p(A|B,C) = \frac{p(A|C)p(B|A,C)}{p(B|C)}. \tag{41}$$

Bayes' theorem can be extended to scientific inference problems by writing Eq. 41 as

$$p(\mathcal{H}|d,\mathcal{I}) = \frac{p(\mathcal{H}|\mathcal{I})p(d|\mathcal{H},\mathcal{I})}{p(d|\mathcal{I})} \tag{42}$$

which mechanically tells us how the information $\mathcal{I}$ available to us about the Universe is updated upon the acquisition of new data d. This is done so by quantifiably testing a hypothesis $\mathcal{H}$. To do so we begin with some *a priori* (before the experiment) belief

in $\mathcal{H}$ which is established in advance of the newly acquired data being analyzed, written as a conditional probability (density) $p(\mathcal{H}|\mathcal{I})$ (i.e., the probability that $\mathcal{H}$ is true given the available information $\mathcal{I}$). Henceforth, we will refer to $p(\mathcal{H}|\mathcal{I})$ simply as "the prior distribution" or, more abruptly, "the prior".

The degree with which the hypothesis accurately predicts the data establishes some "goodness of fit," or likelihood, for $\mathcal{H}$ written in the language of conditional probabilities as $p(d|\mathcal{H},\mathcal{I})$. The "evidence" $p(d|\mathcal{I})$ ensures that the $\sum_i p(\mathcal{H}_i|d,\mathcal{I}) = 1$. The *a posteriori* distribution, anglicized to "the posterior probability" can either support or reject the hypothesis, or be inconclusive. Upon the acquisition of new (and hopefully more informative) data the posterior distribution then becomes the prior and the hypothesis can be further refined.

Those who are less familiar with this approach tend to be wary of the inclusion of one's prior belief in the analysis of the data. This effectively encodes ones biases about the data under analysis and at first glance appears to be in direct opposition to the objectivity that is so necessary for successful science. The explicit inclusion of priors is in fact a feature rather than a flaw. Someone who performs an experiment will have some preconceived notion of the outcome, and these prior biases will affect the experimenter's interpretation. It is normal to expect that different views of the same data can yield valid yet differing hypotheses to describe some phenomena. What is necessary, and guaranteed by the Bayesian approach, is that the data produce identical results under the same prior beliefs.

An excellent example in astrophysics is the decades-long puzzle over the origin of Gamma Ray Bursts (GRBs). For 40 years after their discovery, the progenitors for these extremely powerful explosions in the Universe were a complete mystery. In the presence of poorly informative data (individual GRB observations were rare), a wide range of mechanisms for the events were proposed. Researchers had a differing set of

prior beliefs about the cause of the phenomenon, and the information content of the available data was insufficient to significantly favor any one hypothesis over the others. It was not until the operation of dedicated cosmological gamma-ray observatories, and the data rich in GRB events they supplied, that most of the competing models were rejected and we arrived at the widely accepted view of two classes of GRB: The merger of binary neutron stars and the collapse of extremely massive (~ 100 $M_\odot$) stars.

One (over)simplified specific example of Bayesian learning in action would be the debate over whether GRBs were galactic or cosmological in origin. In the early years after the discovery of GRBs the bias was that they would have to be near-by because of the incredible gamma ray flux. If one were to analyze a new catalogue of GRB events with this hypothesis one must adopt a prior that the events should be distributed similarly to the mass of the Galaxy. The observed isotropic distribution of GRBs would penalize the posterior probability (calculated from Eq. 42) because of the low prior weight given to events outside of the Galactic plane. In contrast, those who proposed a cosmological origin would presumably adopt a uniform prior on the celestial distribution of events and that analysis would incur no such penalty. The cosmological model would thus significantly outperform that of the galactic-origin model. For a more detailed treatise on the discovery and historical attempts at explaining GRBs, see Schilling (2002) [38].

For gravitational wave detections our state of "ignorance" about source parameters is generally high: We do not know *a priori* the physical parameters of an emitting source until we have made a detection. In this situation we typically divide model parameters into two classes: Location parameters and scale parameters. Location parameters depend on the definition of an origin and can take on positive and negative values. The choice of prior distribution should be invariant to shifts in the location

of the origin, satisfied by a uniform distribution for parameter x:

$$p(x|\mathcal{I}) = \frac{1}{\Delta x} \tag{43}$$

over the allowed parameter range $\Delta x = x_{\max} - x_{\min}$.

Scale parameters, such as the distance to a GW emitting source, are always positive and an ignorance prior should be invariant over the the units used to measure the scale, which can potentially span several decades in parameter space. For example, merging stellar-mass black holes are detectable by LIGO out to a few hundred Mpc. Now suppose we were to draw from a uniform prior ranging from zero to 1,000 Mpc. Ninety percent of the samples would be within the final decade of the distribution (distances from 100 to 1,000 Mpc). However, our state of ignorance about merging black holes does not reflect that these objects are much more likely to occur at great distances. It is therefore more common to either use the logarithm of the scale parameter in the modeling, or to adopt a Jeffery's prior [37] for scale parameter x:

$$p(x|\mathcal{I}) = \frac{1}{x \ln \frac{x_{\max}}{x_{\min}}}. \tag{44}$$

In some scenarios our degree of ignorance is less severe and the information on hand can be used to establish a prior which more closely represents the situation. A relevant example would be the use of conventional astronomical observations as triggers for directed gravitational wave searches, as is done in the LIGO analysis for supernovae and gamma ray bursts. The astronomical observation can help determine the location of the event and this information can be folded into the prior used for the location parameters in the gravitational wave search. One such scenario that was of particular interest occurred during the LIGO S5 run, details of which can be found in Ref. [39].

In the more familiar frequentist approach, priors manifest themselves as constraints which enter into the calculation of the statistic. Obviously, placing illogical

constraints on the analysis will maliciously bias the results. The advantage of the Bayesian approach is that one must explicitly declare the prior assumptions under which the analysis is being performed. This is in opposition to a statistical approach to inference where such constraints and biases can exist but are veiled by the way the analysis is performed [40].

In summary, the results of a Bayesian analysis are affected by one's choice of priors and for a completely blind analysis it is expected that one would use non-informative, or ignorance priors. However it is often the case that there is *good reason* to include an informative prior, such as a supposed nearby astrophysical event existing in the Galaxy. It is therefore important to use priors which accurately reflect our current understanding of the phenomenon to be tested by our experiment. The explicit declaration of the priors used alerts the reader to these underlying assumptions as opposed to a constrained frequentist analysis which can implicitly impose illogical restrictions.

We now wish to make use of this powerful mathematical tool. To do so it makes sense to classify inference problems into two generic groups: Parameter estimation and model selection. Although Bayes' theorem is the foundation for solutions to both of these problems, the interpretation of the components takes on a slightly different form and we will treat them separately.

Parameter Estimation

For parameter estimation applications, the hypothesis states that the acquired data are the result of some process (modeled by $\mathcal{M}$) which is parameterized by a vector of physical quantities $\vec{\theta}$ which are used in $\mathcal{M}$. To evaluate the PDF we use

Bayes' theorem

$$p(\vec{\theta}|d,\mathcal{M}) = \frac{p(d|\vec{\theta},\mathcal{M})p(\vec{\theta},\mathcal{M})}{p(d,\mathcal{M})} \tag{45}$$

where the normalization constant is the marginalized likelihood, or evidence, calculated by

$$p(d,\mathcal{M}) = \int d\vec{\theta}\; p(\vec{\theta},\mathcal{M})p(d|\vec{\theta},\mathcal{M}). \tag{46}$$

The reason for this naming convention will become evident in our discussion of model selection. In practice it quickly becomes too costly to perform a direct evaluation of the posterior distribution function because of the cost in evaluating Eq. 46 and so we are typically concerned only with the relative posterior weight between different locations in parameter space.

If the model is characterized over some continuous parameter space, the posterior distribution function $p(\vec{\theta}|d,\mathcal{M})$ (further abbreviated as the PDF) is the probability density function for the model parameters. Furthermore, $p(\vec{\theta},\mathcal{M})$ is the prior probability density for $\vec{\theta}$ and $p(\vec{\theta},\mathcal{M})d\vec{\theta}$ is the probability that the parameters lie in the interval $[\vec{\theta},\vec{\theta}+d\vec{\theta}]$. The global maximum of the PDF, or maximum a posteriori (MAP) point, occurs where the model parameters are the "best fit" values for the hypothesis in question, weighted by the prior distribution for those parameter values. The distribution about this location in parameter space reveals how well the data constrains the hypothesis.

The inherent width of $p(\vec{\theta}|,\mathcal{M})$ does not imply some intrinsic uncertainty in the parameter values taken on by the system being studied (as is the case in the frequentist interpretation), but instead our incomplete knowledge of the system. In the (often unachievable) limit of complete knowledge, $p(\vec{\theta}|d,\mathcal{M})$ would collapse to a delta function about the physical parameter values of the object being studied. In practice

this distribution has some width which represents our uncertainty (due to our own ignorance) about the system.

Notice that we have dropped the information $\mathcal{I}$ from the conditional probabilities above. This is for clarity, as the arguments of the probabilities will shortly become crowded, but it is implicitly included in the discussion. $\mathcal{I}$ does not enter into the calculus of the PDF, but it is of paramount importance when *interpreting* the result. For example, suppose a candidate event in LIGO data is being analyzed using waveforms predicted by simulations of supernova explosions and we assume in the analysis that the noise in the detectors is gaussian. The parameters would be those describing the supernova progenitor and the results of the analysis are obviously dependent on the assumption that the event *is the result of a supernova explosion*, encoded by $\mathcal{M}$, and that the noise is truly gaussian in nature, encoded by $\mathcal{I}$. If this assumption is false, the result is obviously moot, but the analysis can only answer what is being asked of it. In this case it will tell us about which supernova progenitor parameters are the most likely to produce such a signal in gaussian noise, but is incapable of informing us about the accuracy of our assumption regarding the origin of the data.

In the event that a model contains several parameters but we would like to focus our attention on one quantity in particular, the axioms of logical inference (namely the sum rule) allow us to integrate the posterior distribution function over all "nuisance parameters" leaving behind the distribution about the parameter of interest, historically referred to as the marginalized PDF. For a model parameterized by two quantities, (x, y), the marginalized PDF for parameter x is

$$\begin{aligned} p(x|d) &\propto \int p(x, y|d)dy \\ &\propto \int p(d|x, y)p(x, y)dy. \end{aligned} \tag{47}$$

For high dimension problems it is difficult to sufficiently sample the full PDF and marginalization is one common approach to studying combinations of parameters which are of particular interest.

If we were only ever interested in detecting gravitational waves this rigor would be completely unnecessary, however the goal of data analysis is to divulge as much detail about the PDF as is possible (and practical). Although the values for the model parameters that result in the maximum posterior weight receive most of the attention when one wants to answer astrophysical questions with gravitational waves, it is the *distribution* of the parameters that mirrors our state of knowledge about the system and thereby what can be learned. In the familiar statistical approach, a collection of "best fit" parameters is of little value without appropriately assigned error bars. Bayesian parameter estimation allows us to avoid the murky business of confidence intervals and instead supplies us with a mechanism for producing the actual distribution for each parameter.

How to go about implementing this set of tools is another question. One approach that may come to mind would be to evaluate Eq. 45 along discrete points distributed in some way throughout the parameter space. Similar grid-based methods are the backbone of most search algorithms used to analyze the data from the current generation of gravitational wave detectors. Unfortunately, in many GW applications the number of parameters needed to describe the signal can be very large (ranging from a few to hundreds of thousands), and the regions of significant posterior weight generally occupy a minute fraction of the total prior volume. For instance, GW detectors are broad-band instruments with extremely precise frequency resolution. Resolving the peaks in the posterior distribution function subsequently requires very fine sampling of the parameter space and if this is extended over the entire prior volume the total number of samples can become prohibitively large, making stochastic search methods

a necessity. We will exclusively focus on such techniques for parameter estimation problems in this work.

Implicit in this discussion is the assumption that the model which is described by parameters $\vec{\theta}$ is in fact the true mechanism which has produced the data. This, of course, is not always a reasonable assumption in practical applications, especially in such ill-conditioned analysis problems posed by gravitational wave astronomy. Perhaps a more natural interpretation of Bayes' theorem is that of model selection, where we wish to evaluate which model is most appropriately represented by the data.

Model Selection

Consider a scenario where one is attempting to fit data with two competing models of differing dimension. In general, a higher dimensional model will produce a better fit to a given set of data. This can be taken to the limit where there are as many model parameters as there are data points allowing one to perfectly match the data. The process of deciding which model is most appropriate can therefore not solely depend on which model has a better fit. Fortunately, we can use probability and logic to infer if either model should be favored. In the context of gravitational wave detection, these disparate hypotheses could involve the inclusion of additional physical parameters used to model the source or the inclusion of additional sources in the data.

One such example which will be confronted by the LISA data analysis effort is known as the "White Dwarf Transform" [10]. One of the most sought after anticipated LISA sources is the merger of binary systems containing two massive black holes. Furthermore, one of the most reliably anticipated forms of gravitational waves results from the orbit of stellar mass compact objects (such as white dwarf stars) within the galaxy. The model confusion lies in the fact that a binary black hole in-spiral

waveform can be written as a linear combination of white dwarf binaries. Therefore, if a black hole inspiral waveform is encoded within the data it could be mistaken for a collection of white dwarf binaries, and a model selection analysis should be performed. This turns out to be a somewhat trivial model selection problem but not all gravitational wave model selection problems will be as transparent.

We will again begin with Bayes' theorem with model $\mathcal{M}$ as our hypothesis such that the posterior is calculated by

$$p(\mathcal{M}|d) = \frac{p(d|\mathcal{M})p(\mathcal{M})}{p(d)}. \tag{48}$$

In model selection problems the hypotheses are discrete. The data are described by model $\mathcal{M}_i$ or $\mathcal{M}_j$, so $p(\mathcal{M}|d)$ is to be interpreted as the probability that $\mathcal{M}$ is the appropriate description of the data. While the normalization constant was unimportant in parameter estimation problems, it is necessary to arrive at the appropriate number for the posterior. This requires prior knowledge of *all possible models* testable by our experiment, which is impractical. Instead we must ask our data more carefully constructed questions, such as "how do two specific models compare?" Since the normalization in Eq. 48 is the same for all models, ratios of the posterior odds eliminate the need to know all hypotheses *a priori* and yields a useful quantitative measure of confidence, the odds ratio, for one model over another:

$$\mathcal{O}_{10} \equiv \frac{p(\mathcal{M}_1|d)}{p(\mathcal{M}_0|d)} = \frac{p(\mathcal{M}_1)p(d|\mathcal{M}_1)}{p(\mathcal{M}_0)p(d|\mathcal{M}_0)} \tag{49}$$

$\mathcal{O}_{10}$ is most simply interpreted as the "betting odds" of $\mathcal{M}_1$ over $\mathcal{M}_0$. If $\mathcal{O}_{10} = x$, we should read this as $\mathcal{M}_1$ being preferred over $\mathcal{M}_0$ with $x : 1$ odds. An odds ratio of ~ 1 means both models are on equal footing as far as the analysis is concerned and both must be kept under consideration. The question of when the odds ratio becomes sufficiently informative to favor one model over the other is more subtle. We

will adopt the interpretation of A.E. Raftery (1996) [1] in which the number of bits of information contained in favoring a particular model is used to guide our confidence in that model's selection. Typically, an odds ratio ≥ 3 is when one begins to assign preference to $\mathcal{M}_1$. See Table 1

$\mathcal{O}_{10}$	Evidence for model 1
$< 1/3$	Negative (supports model 0)
1/3 to 3	Not worth more than a bare mention
3 to 12	Positive
12 to 150	Strong
> 150	Very Strong

Table 1: "Confidence" levels for adopting model 1 over model 0 as a function of increasing odds ratio based on the relative number of bits of information gained from the two models. Typically odds ratios greater than three (less than 1/3) signify distinguishable models. This table was adapted from Raftery (1996) [1].

The odds ratio is typically decomposed into two components: The prior odds $\mathcal{P}_{10} \equiv p(\mathcal{M}_1)/p(\mathcal{M}_0)$ reflects any preference between models before the experiment is conducted while the Bayes factor

$$B_{10} \equiv \frac{p(d|\mathcal{M}_1)}{p(d|\mathcal{M}_0)} \tag{50}$$

is the likelihood ratio of one model over the other in describing the available data. If we are to adopt a stance which states that one of the two models is preferred the Bayes factor must overwhelm the prior odds signaling the data are sufficiently informative to distinguish between the competing hypotheses. For the current generation LIGO detectors $\mathcal{P}_{10}$ would likely down-weight the chances of a resolvable GW signal based on some combination of reasonable event rates, expected "glitch" rates, and detector behavior around the time of the candidate event. This is analogous to setting a higher frequentist threshold which must be exceeded for a candidate detection to warrant further scrutiny. For the LISA extension of this same problem (due to the large

number of anticipated detections) the prior odds would likely be far less restrictive, leaving the analysis more open to the possibility of a signal being present.

If we were to revisit equation 45 and explicitly include $\mathcal{M}$ as a condition on the probabilities, we would see that the likelihood in the model selection context is identical to the previously ignored normalization constant in the parameter estimation formalism. The impracticality of successfully integrating Eq. 45 which was previously dismissed must now be confronted. The potential power of these methods coupled with the difficulty encountered in their implementation has produced an active field of research [1, 41, 42, 37, 36, 43, 44] from which we gratefully harvest.

Advocates of Bayesian inference cite a natural parsimony in the model selection approach application akin to Occam's Razor. The claim is that if two models give similar quality "fits" to the data, the lower dimensional model will be preferred. This line of reasoning is only valid when accompanied by a strict qualifying statement: *Excess model parameters must be constrained by the data.*

For example, suppose we are interested in two models, $\mathcal{M}_0$ and $\mathcal{M}_1$, and we wish to calculate the Bayes factor between them when used to describe data d. Furthermore, suppose $\mathcal{M}_1$ has a single parameter, θ, and $\mathcal{M}_0$ is un-parameterized, with $\theta = \theta_0$. For this example we will adopt a uniform prior for θ over the range $\Delta\theta$ and assume that the likelihood for model $\mathcal{M}_1$ is a gaussian with variance σ^2 and mean $\theta_{\rm MAP}$.

The marginalized evidence for $\mathcal{M}_0$ is just the likelihood $p(d|\theta_0, \mathcal{M}_0)$. For $\mathcal{M}_1$ we must integrate over the model parameter

$$\begin{aligned} p(d|\mathcal{M}_1) &= \int_{\Delta\theta} p(\theta|\mathcal{M}_1) p(d|\theta, \mathcal{M}_1) d\theta \\ &= \frac{1}{\Delta\theta} \int_{\Delta\theta} p(d|\hat{\theta}, \mathcal{M}_1) d\theta \\ &= p(d|\hat{\theta}, \mathcal{M}_1) \frac{\delta\theta}{\Delta\theta} \end{aligned} \tag{51}$$

where $\delta\theta$ is the characteristic width of the posterior which, for our gaussian example, is equal to $\sqrt{2\pi}\sigma$. The Bayes factor for these two models is then

$$B_{10} = \frac{p(d|\hat{\theta}, \mathcal{M}_1)}{p(d|\theta_0, \mathcal{M}_0)} \frac{\delta\theta}{\Delta\theta} \tag{52}$$

and we can see that for the Bayes factor to favor $\mathcal{M}_1$ the improvement in the likelihood courtesy of the additional degree of freedom over $\mathcal{M}_0$ must overcome the "Occam factor" $\delta\theta/\Delta\theta$.

The marginalized posterior distribution function of an unconstrained parameter will match that quantity's prior distribution (for our previous example $\delta\theta \sim \Delta\theta$), meaning the data was incapable of updating our understanding of that parameter and can not be used to answer questions about whether or not that quantity "belongs" in the model (See Ref. [42] for an excellent discussion of this and other misconceptions regarding Bayesian inference). That model then incurs no penalty for having unconstrained variables included in its parameterization. If the remaining parameters do as well describing the data as some other, lower dimensional model, both are equally viable. Reverend Bayes brandishes his razor only if a particular model *requires* more parameters to achieve a similar fit. This is a somewhat subtle and often overlooked point that, when addressed, enhances the conceptual understanding of Bayesian model selection.

We now turn our attention to the practical implementation of these powerful tools of inference. Although specific implementations will be discussed in detail for the different analysis problems solved in this work, our approach contains a common thread: The use of Markov Chain Monte Carlo methods to sample from the target posterior distribution.

The Markov Chain Monte Carlo Algorithms

Markov Chain Monte Carlo (MCMC) [45, 46, 47] techniques are becoming ubiquitous in large dimensional minimization problems and are growing in familiarity among the gravitational wave community thanks in part to their great effectiveness in solving parameter estimation problems [48, 49]. Variants of these techniques have also proven quite effective as search algorithms [10, 32, 50, 51, 52]. Here a brief description of the MCMC technique will be followed by more in-depth "case by case" discussions of the particular implementations that we have used.

The MCMC approach encompasses a powerful set of techniques for producing samples from the posterior distribution. Directly sampling from the desired PDF becomes prohibitively costly as the model dimension increases because of the need to evaluate the marginalized likelihood (see Eq. 46). Instead, these algorithms focus their attention on regions of high posterior weight, and neatly avoid the problem of computing the model evidence.

The utility of the MCMC approach stems from its ability to stochastically explore the parameter space in such a way that the resulting samples represent the target posterior distribution function. In other words, the "chain" of samples is like a time series where the duration of "time" spent in a particular region of parameter space is proportional to the posterior weight contained in that region.

We are assured that subsequent samples (x_i and x_{i+1}) are independently drawn from a target distribution function π if they satisfy the *detailed balance* condition:

$$\pi(x_i)p(x_{i+1}|x_i) = \pi(x_{i+1})p(x_i|x_{i+1}) \tag{53}$$

where $p(x_{i+1}|x_i)$ is the probability of transitioning from x_i to x_{i+1}. In principle, if we knew this transition probability we could easily generate samples from the

target distribution. Unfortunately this is not the case for inference problems, so we must replace $p(x_{i+1}|x_i)$ with some arbitrary *proposal distribution* $q(x_{i+1}|x_i)$ – the probability of proposing a move to x_{i+1} given that we are at x_i – which will most likely violate the detailed balance condition. Without loss of generality we will assume that the left-hand side of Eq. 53 is excessively favored by our choice of proposal distribution

$$\pi(x_i)q(x_{i+1}|x_i) > \pi(x_{i+1})q(x_i|x_{i+1}). \tag{54}$$

To recover the detailed balance condition we must include an *acceptance probability* $\kappa(x_{i+1}|x_i)$ – the probability of *accepting* the proposed move – on each side of the inequality

$$\pi(x_i)q(x_{i+1}|x_i)\kappa(x_{i+1}|x_i) = \pi(x_{i+1})q(x_i|x_{i+1})\kappa(x_i|x_{i+1}). \tag{55}$$

For our choice of proposal distribution, transitions from x_{i+1} to x_i were less probable than the reverse move, suggesting that we should take $\kappa(x_i|x_{i+1}) = 1$. We can then rearrange Eq. 55 to solve for the acceptance probability of the transition from x_i to x_{i+1}:

$$\kappa(x_{i+1}|x_i) = \frac{\pi(x_{i+1})q(x_i|x_{i+1})}{\pi(x_i)q(x_{i+1}|x_i)}. \tag{56}$$

The term on the right hand side of Eq. 57 is called the Hastings Ratio [46], H, which can take on values between 0 and ∞. To ensure that $\kappa(x_{i+1}|x_i)$ is a probability we re-write Eq. 57 as

$$\kappa(x_{i+1}|x_i) = \min\left[1, H\right]. \tag{57}$$

In the scenario where we are interested in solving a parameter estimation problem we wish to infer the posterior distribution function of the model parameters $\vec{\theta}$. The PDF is resolved by first adopting some position in parameter space $\vec{\theta}_i = \vec{\theta}_x$ as the first "link" in the chain. A subsequent move to a new position $\vec{\theta}_{i+1} = \vec{\theta}_y$ is proposed using a normalized distribution $q(\vec{\theta}_y|\vec{\theta}_x)$ and the proposed location $\vec{\theta}_y$ is adopted with

probability $\kappa = [1, H]$ using the Hastings ratio

$$H_{\vec{\theta}_x \to \vec{\theta}_y} = \frac{p(d|\vec{\theta}_y)p(\vec{\theta}_y)q(\vec{\theta}_x|\vec{\theta}_y)}{p(d|\vec{\theta}_x)p(\vec{\theta}_x)q(\vec{\theta}_y|\vec{\theta}_x)}. \tag{58}$$

The MCMC always accepts transitions to locations with higher posterior weight ($\kappa = 1$) and has some finite probability of accepting moves to places which are "worse." In practice, we compare κ to a random number γ drawn from $U[0, 1]$. If $\gamma > \kappa$, the move is rejected and $\theta_{i+1} = \theta_x$. If $\gamma < \kappa$ then the new position is accepted, and $\theta_{i+1} = \theta_y$.

Scrutiny of Eq. 58 reveals how Bayes' theorem is at work. In the Hastings Ratio we see the un-normalized posterior weight for positions $\vec{\theta}_y$ and $\vec{\theta}_x$ which appear in the numerator and denominator respectively as the product of the likelihood and the prior. Because $\vec{\theta}_x$ and $\vec{\theta}_y$ are from the same model, the normalization of the posterior is identical for both numerator and denominator, allowing us to avoid the costly integration of the evidence while still recovering samples from the posterior. The ratio of the proposal probabilities is used to penalize regions of parameter space which will be frequently suggested and to promote rarely visited regions in such a way that the chains are not biased by the choice of proposal distribution.

A common practice is to choose a proposal distribution that is symmetric, thus eliminating the need to include it in the calculation of H. The choice of $q(\vec{\theta}_y|\vec{\theta}_x)$, by construction, *can not* alter the recovered posterior distribution function. The proposal distribution does, however, dramatically affect the acceptance rate of trial locations in parameter space and, therefore, the number of iterations required to satisfactorily sample the joint PDF.

Early iterations, referred to as the "burn-in" samples of the chain, are likely to be excessively weighted in the final distribution as it is improbable that the first few steps of the chain accurately represent the target distribution – in parameter estimation problems we most likely initialize the chains at positions with very low

posterior weight. These initial samples are discarded such that all remaining samples are from a stationary distribution. The number of post burn-in iterations spent in a particular region of parameter space, normalized by the total number of post burn-in steps in the chain, is representative of the probability that the source parameters take values in that interval.

For large dimension search spaces the number of samples is typically far fewer than one would need to sample the full PDF. Marginalized posterior distribution functions are of more value and are easily inferred by binning the values taken by parameters over the entire post burn-in chain of samples.

MCMC techniques can be generalized so that the transitions between steps in the Markov Chain include "tans-dimensional" moves between different models, resulting in what are termed Reverse Jump Markov Chain Monte Carlo (RJMCMC) algorithms. These methods can be constructed to simultaneously sample the within-model posterior (as described above) *and* the model posterior which is necessary for model selection. Details of the RJMCMC techniques can be found in chapters 4 and 5.

While the overall prescription for implementing an MCMC approach is remarkably simple, attention to detail is necessary to obtain satisfactory efficiency. While MCMCs require specific tuning for each problem that they are used to tackle, there are sufficient similarities between any gravitational wave analysis application to warrant a general GW-specific implementation discussion.

Practical MCMC Implementation for GW Data Analysis

We will now describe a basic implementation of the MCMC algorithm to analyze gravitational wave data. Here we will define our evaluation of the likelihood

and discuss the role of the proposal distribution, $q(\vec{y}|\vec{x})$. We will also outline the most commonly used proposal distribution in our work. This involves constructing a multivariate gaussian approximation to the posterior using the Fisher information matrix and using the eigenvalues and eigenvectors of the distribution to propose new parameters.

Evaluating the Likelihood

To calculate the likelihood we first consider binned data containing no contribution from gravitational waves. Furthermore we will assume that we have some information $\mathcal{I}$ about the noise in the detector, namely that the noise in each bin is independently drawn from a stationary, normal distribution with zero mean and variance σ^2. Considering a single bin of data from a single detector, the probability that the instrument would measure noise n_i is simply

$$p(n_i|\mathcal{I}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{n_i^2}{2\sigma^2}} \tag{59}$$

The probability that the instrument would produce N independent noise samples can be written as $p(n_1, n_2, ..., n_N|\mathcal{I})$ where $p(A, B) = p(A)p(B)$ is the probability that A and B are true. Therefore, the probability of acquiring a particular noise realization $\vec{n} \to (n_1, n_2, ..., n_N)$ is

$$p(\vec{n}) = \prod_i p(n_i) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} e^{-\sum_i \frac{n_i^2}{2\sigma^2}}. \tag{60}$$

As in chapter 2 we will assume that the signal s consists of two contributions. The gravitational wave convolved with the transfer function and antenna pattern of the instrument produces some response h, while the instrument itself produces noise n. The analysis is typically performed in the Fourier domain because of the relative simplicity of $\tilde{n}(f)$ as opposed to $n(t)$.

It is not always the case that the noise spectrum is independent of frequency (i.e., white), however the noise power spectral density (PSD) of a detector, symbolized as $S_n(f)$, will, to some extent, be understood. We then can rework the above derivation for a frequency-dependent noise variance. This, however, complicates the normalization term in Eq. 60 as σ now depends on frequency. Fortunately, what is important in MCMC methods is the *likelihood ratio*. If we assume that the theoretical (or measured) noise spectral density is a true representation of the detecter noise then the normalization is identical for any noise realization and will not contribute to the ratio. We can then focus our attention on the variable part of the likelihood.

At this point it is convenient to define the noise-weighted inner product

$$(a|b) \equiv 2 \int_0^\infty \frac{\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df \tag{61}$$

which will be of great utility to us throughout this work. For discrete Fourier domain data from observation time T the inner product takes the form

$$(a|b) = \frac{2}{T} \sum_k \sum_f \frac{\tilde{a_k}^*(f)\tilde{b}_k(f) + \tilde{a}_k(f)\tilde{b}_k^*(f)}{S_n^k(f)} \tag{62}$$

where the index k indicates different interferometer channels. These could be the A, E, and T channels of LISA or the different interferometers of the ground-based network.

The factor of two results from the integration (or summation) being performed over only positive frequencies, as $S_n(f)$ is a one-sided PSD only defined over $f > 0$. This is valid as long as the time series $a(t)$ and $b(t)$ are real which is clearly satisfied data from gravitational wave detectors. We can then write the probability of a particular noise realization more concisely as

$$p(\tilde{n}(f)) \propto e^{-\frac{1}{2}(n|n)} \tag{63}$$

Extending this discussion to data that potentially contains a gravitational wave signal, the likelihood that $\tilde{s}(f)$ could have been generated by some template waveform $\tilde{h}(f)$ parameterized by $\vec{\theta}$ is the probability of the instrument producing a noise realization $\tilde{n}(f)$ *identical to the residual of the data with the template.*

$$p(s|\vec{\theta}) \equiv Ce^{-\frac{1}{2}(s-h|s-h)} \tag{64}$$

where the C is the normalization constant which is independent of signal parameters.

Having defined the inner product we will also take this opportunity to define the signal-to-noise ratio (SNR) which we use as a metric to compare the relative "detectability" of the signal in question. The optimal matched filtering SNR is

$$\text{SNR}_{\text{optimal}} = \sqrt{(h|h)} \tag{65}$$

while a more "data dependent" definition which is more commonly seen in the gravitational wave literature is

$$\text{SNR} = \frac{(s|h)}{\sqrt{(h|h)}}. \tag{66}$$

When quoting SNRs we will use Eq. 65 (the optimal matched filtering version) unless otherwise stated.

Proposal Distributions

The scheme for proposing trial positions in the parameter space has no bearing on the ultimate results of any MCMC analysis but it greatly affects the amount of time required to arrive at that solution. Convergence time is at a minimum if the proposal distribution matches the posterior distribution. While we can not expect to know the "answer" (the PDF) until after we have done the analysis, physical considerations can help in the design of proposal distributions that approximate the target distribution. It is therefore imperative to the efficiency of an MCMC-based approach that proposal

distributions are constructed with the details of the problem in mind as a well-tuned MCMC will drastically outperform a more general approach.

Although we will heed our own advice and tune each implementation to the specifics of the problem for which it is designed to solve, there are some common proposal distributions used in each of the cases that we will study in detail. Most of these are trivial, such as uniform draws over the prior volume, however one in particular warrants further detailed description.

Signal parameters can often exhibit significant correlations which, if not accounted for, can conspire to maliciously affect the convergence time of an MCMC as the rate at which new positions are accepted can become prohibitively low. Figure 8 is a two-dimensional cartoon to illustrate how independent parameter jumps can easily result in proposing moves to locations of very low posterior weight, and therefore very low probability of being accepted. When jumps are made which incorporate the correlations between parameters the chain can much more rapidly explore the high density regions of the posterior. Accurately determining these correlations requires the calculation of the full covariance matrix – second order derivatives for an $N \times N$ dimensional matrix. For moderate to large dimensional problems this can be computationally prohibitive. A common approach to ameliorate this dilemma is to use the Fisher Information Matrix ($\mathbf{\Gamma}$) as an approximation to the inverse correlation matrix. The components of $\mathbf{\Gamma}$

$$\Gamma_{ij} = (h_{,i}|h_{,j}) \tag{67}$$

require first order partial derivatives, denoted by $h_{,i} \equiv \partial h/\partial\theta_i$ [53]. The matrix is then numerically inverted to find its eigenvectors and eigenvalues. We arrive at newly proposed parameters by "jumping" in the eigen-directions of the Fisher by drawing from $N[0,1]$ and scaling it by the associated eigenvalue. These jumps are decomposed

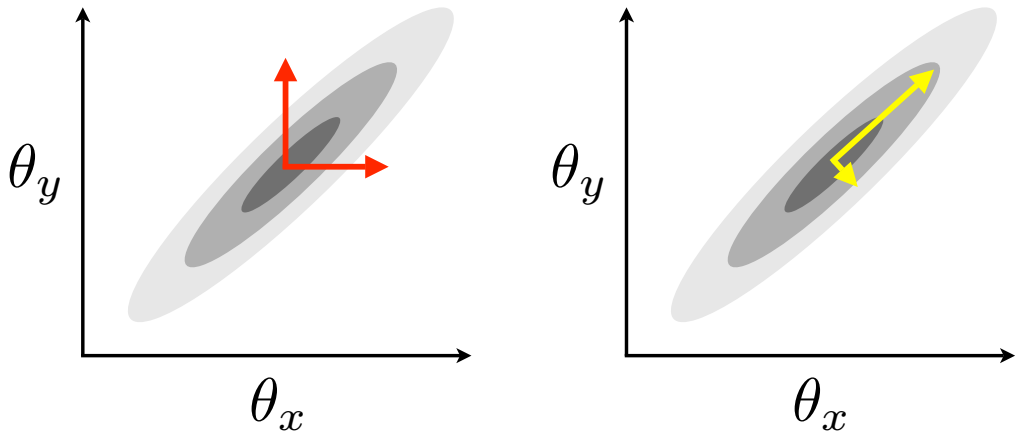


Figure 8: This cartoon depiction of the target distribution is to illustrate the need for incorporating parameter correlations when proposing new locations in $\vec{\theta}$ space. The concentric ellipses depict 1 σ contours of the PDF, with darker grey signifying a region of higher posterior weight. The panel on the left shows typical independent parameter jumps (red) from the MAP value for θ_x and θ_y scaled by the marginalized variance for each parameter. The correlation prevents these proposals from landing in a region of substantial weight and therefore have very poor chance of being accepted. This is in contrast to the panel on the right where correlated jumps (yellow) are proposed thus allowing the chain to efficiently explore the posterior.

into the parameter directions and additionally rescaled by $1/\sqrt{D_\theta}$, where D_θ is the model dimension. This ensures that typical jumps are about one standard deviation in size. Although this type of proposal distribution very efficiently samples the posterior near the peak of the distribution it is a poor choice for sampling the entire posterior, as large jumps are very unlikely to occur.

Since the Fisher Information Matrix is a local approximation to the inverse covariance matrix, it cannot reasonably predict the parameter correlations far from the current location. For the large jumps we use a simpler proposal that takes the estimated variance for each parameter (as determined from $\mathbf{\Gamma}$) and use this to scale additional draws from $N[0,1]$ to jump in parameter directions. This proposal distribution cocktail is then rounded out by occasionally proposing uniform draws on the prior, ensuring that trial samples cover the entire prior volume during the exploration of the target distribution. This collection of proposal distributions has the added benefit of being symmetric in $\vec{\theta}_x$ and $\vec{\theta}_y$, allowing us to neglect their contribution to the Hastings ratio.

We have now laid the foundation for using the Markov Chain Monte Carlo family of algorithms to infer what can be learned by data obtained by gravitational wave detectors. From here we will demonstrate three examples where Bayesian techniques, spearheaded by the MCMC machinery, are used to simultaneously solve parameter estimation and model selection problems that are pertinent to the success of the current and near-future GW instruments.

TESTS OF BAYESIAN MODEL SELECTION TECHNIQUES FOR GRAVITATIONAL WAVE ASTRONOMY

Introduction

Here we explore various methods for computing Bayes factors in the context of determining which galactic binaries have measurable frequency evolution. The methods explored include a Reverse Jump Markov Chain Monte Carlo (RJMCMC) algorithm, Savage-Dickie density ratios, the Schwarz-Bayes Information Criterion (BIC), and the Laplace approximation to the model evidence. We find good agreement between all of the approaches.

The LISA observatory [7] is designed to explore the low frequency portion of the gravitational wave spectrum between $\sim 0.1 \rightarrow 100$ mHz. This frequency region will be heavily populated by signals from galactic binary systems composed of stellar mass compact objects (e.g., white dwarfs and neutron stars), of which millions are theorized to exist. Tens of thousands of these GW sources will be resolvable by LISA and the remaining sources will contribute to a confusion-limited background [31]. This is expected to be the dominant source of low frequency noise for LISA. Detection and subsequent regression of the galactic foreground is of vital importance in order to then pursue dimmer sources that would otherwise be buried by the foreground. To do so efficiently requires a global solution to the "galactic foreground problem" and stochastic methods have proven to be the most successful [32].

Although the problem of determining the number of resolvable galactic binaries is arguably the most pressing, the need to determine the number of measurable source parameters is an important endeavor that serves as a testbed for techniques which can ultimately be used for the source-by-source model selection problem. One could argue

the decision regarding the inclusion of a single model parameter is better described as "approximation selection" since we are choosing between different parameterizations of the full physical model that describes the signals from binary systems of point masses in general relativity. However, many similar modeling problems in astrophysics and cosmology [13], as well as in other fields such as geophysics [54], are considered to be examples of model selection, and we will adopt that viewpoint here.

For example, the GW signal from a binary inspiral, as detected by LISA, may involve as many as 17 parameters. However, for massive black hole binaries of comparable mass we expect the eccentricity to be negligible, reducing the model dimension to 15, while for extreme mass ratio systems we expect the spin of the smaller body to have little impact on the waveforms, reducing the model dimension to 14. In many cases the inspiral signals may be described by even fewer parameters. For low mass galactic binaries spin effects will be negligible (removing six parameters), and various astrophysical processes will have circularized the orbits of the majority of systems (removing two parameters). Of the remaining nine parameters, two describe the frequency evolution – *e.g.* the first and second time derivatives of the GW frequency, which may or may not be detectable [1].

Here we investigate the application of Bayesian model selection to LISA data analysis in the context of determining the conditions under which the first time derivative of the GW frequency, $\dot{f}$, can be inferred from the data. We wish to distinguish between two models with signals parameterized by eight quantities

$$\mathcal{M}_0 : \vec{\theta}_0 \rightarrow (A, f, \theta, \phi, \psi, \iota, \varphi_0, \dot{f} \equiv 0)$$

[1] While this count is only strictly correct for point-like masses, frequency evolution due to tides and mass transfer can also be described by the same two parameters for the majority of sources in the LISA band.

$$\mathcal{M}_1 : \vec{\theta}_1 \rightarrow (A, f, \theta, \phi, \psi, \iota, \varphi_0, \dot{f}) \tag{68}$$

where A is the amplitude, f is the initial gravitational wave frequency, θ and ϕ are the ecliptic co-latitude and longitude, ψ is the polarization angle, ι is the orbital inclination of the binary and φ_0 is the initial phase of the GW. The parameters f, $\dot{f}$ and φ_0 are evaluated at some fiducial time (*e.g.* at the time the first data are taken). For this study only a single source is injected into the simulated data streams.

Model Selection

Bayes Factor Estimates

Apart from carefully concocted toy problems, direct calculation of the evidence, and therefore B_{10}, is impractical. To determine B_{10}, the integration required to compute $p(s, \mathcal{M}_i)$ cannot generally be solved analytically and for high dimension models Monte-Carlo integration proves to be inefficient. To employ this powerful analytic tool various estimates for the Bayes Factor have been developed with different levels of accuracy and computational cost [1, 13]. We have chosen to focus on four such methods: Reverse Jump Markov Chain Monte Carlo and Savage-Dickie density ratios, which directly estimate the Bayes factor, and the Schwarz-Bayes Information Criterion (BIC) and Laplace approximations of the model evidence.

The Reverse Jump Markov Chain Monte Carlo: This breed of MCMC was originally developed by P.J. Green in 1995 [41] and is unique in that it has the ability to transition between competing models, effectively making the model one of the search parameters. This technique, given adequate mixing and convergence, has the advantage of directly sampling from $p(\mathcal{M}|d)$ as opposed to approximating or calculating the model evidence. Like its fixed dimension counterpart, the RJMCMC

is guaranteed to converge to the appropriate PDF. The chain will tend to spend more iterations using the model most appropriately describing the data, making the decision of which model to favor a trivial one. The model likelihood is determined by the number of iterations the chain spends in each model (normalized by the total number of iterations). With a small number of models present (as in our example) a more useful representation of the information garnered by the RJ algorithm is to calculate the Bayes factor for two competing models

$$B_{10} = \frac{\#\text{ of iterations in } \mathcal{M}_1}{\#\text{ of iterations in } \mathcal{M}_0}. \tag{69}$$

Allowing for the exploration of different models (which may differ in dimension) requires a separate Hastings step which proposes to move the chain from one model to another. Parameters for trial model $\mathcal{M}_i$ are drawn from $q(\vec{\theta}_i, \mathcal{M}_i)$. If the models are nested (perhaps proposing additional parameters to include in the existing set) all of the like parameters are held fixed while the new parameters are drawn from the proposal distribution.

Once the new model's parameters are in hand the trans-model Hastings ratio (again, satisfying detailed balance) is calculated by

$$H_{\mathcal{M}_i \to \mathcal{M}_j} = \frac{p(s|\vec{\theta}_j, \mathcal{M}_j) p(\vec{\theta}_j, \mathcal{M}_j) q(\vec{u}_i, \mathcal{M}_i)}{p(s|\vec{\theta}_i, \mathcal{M}_i) p(\vec{\theta}_i, \mathcal{M}_i) q(\vec{u}_j, \mathcal{M}_j)} |J_{ij}| \tag{70}$$

where the determinant of the Jacobian

$$|J_{ij}| \equiv \left| \frac{\partial(\vec{\theta}_j, \vec{u}_i)}{\partial(\vec{\theta}_i, \vec{u}_j)} \right| \tag{71}$$

accounts for any change in dimension between models $\mathcal{M}_i$ and $\mathcal{M}_j$ which are parameterized by $\vec{\theta}_i$ and $\vec{\theta}_j$ respectively. The vector $\vec{u}_j$ is the set of random numbers used to generate the parameters for model j and is of size r_j. If model $\mathcal{M}_j$ is of dimension D_j then the Jacobian is a $(D_i + r_i) \times (D_j + r_j)$ matrix [54].

For example, assume that $\mathcal{M}_0$ has one free parameter, θ, while $\mathcal{M}_1$ is parameterized by a two quantities, θ, ϕ. When proposing a transition from $\mathcal{M}_0 \to \mathcal{M}_1$ we will adopt the same value of θ and require a single random number u_ϕ which is used to generate ϕ via $\phi = u_\phi \Delta\phi$. In this case the Jacobian in Eq. 71 becomes

$$|J_{10}| = \left| \begin{array}{cc} \frac{\partial\theta}{\partial\theta} & \frac{\partial\theta}{\partial u_\phi} \\ \\ \frac{\partial\phi}{\partial\theta} & \frac{\partial\phi}{\partial u_\phi} \end{array} \right| = \Delta\phi \tag{72}$$

If we instead choose to draw the model parameters directly from some distribution, $|J_{ij}|$ is unity and can be neglected.

Selecting an efficient proposal distribution for model transitions is typically the major obstacle in the implementation of an efficient RJ routine. If the proposal distributions yield the model parameters directly, instead of a set of random numbers which are then used to determine the new model parameters, the Jacobian is unity and can be neglected. RJMCMC algorithms are notoriously difficult to implement because dimension changing moves are typically accepted with very low frequency causing the chains to converge slowly. This problem is exacerbated when competing models are of high dimensionality. It is therefore vital to the success of the algorithm to construct trans-model proposals which allow for decent mixing between the two models. Without adequate between-model transitions the inferred Bayes factors are suspect.

For the trans-dimensional implementation applicable to the LISA data analysis problem studied here, the choice of model parameters becomes one of the search parameters. The algorithm proposes parameter 'birth' or 'death' moves (proposing to include or discard the 'extra' parameter(s)) while holding all other parameters fixed.

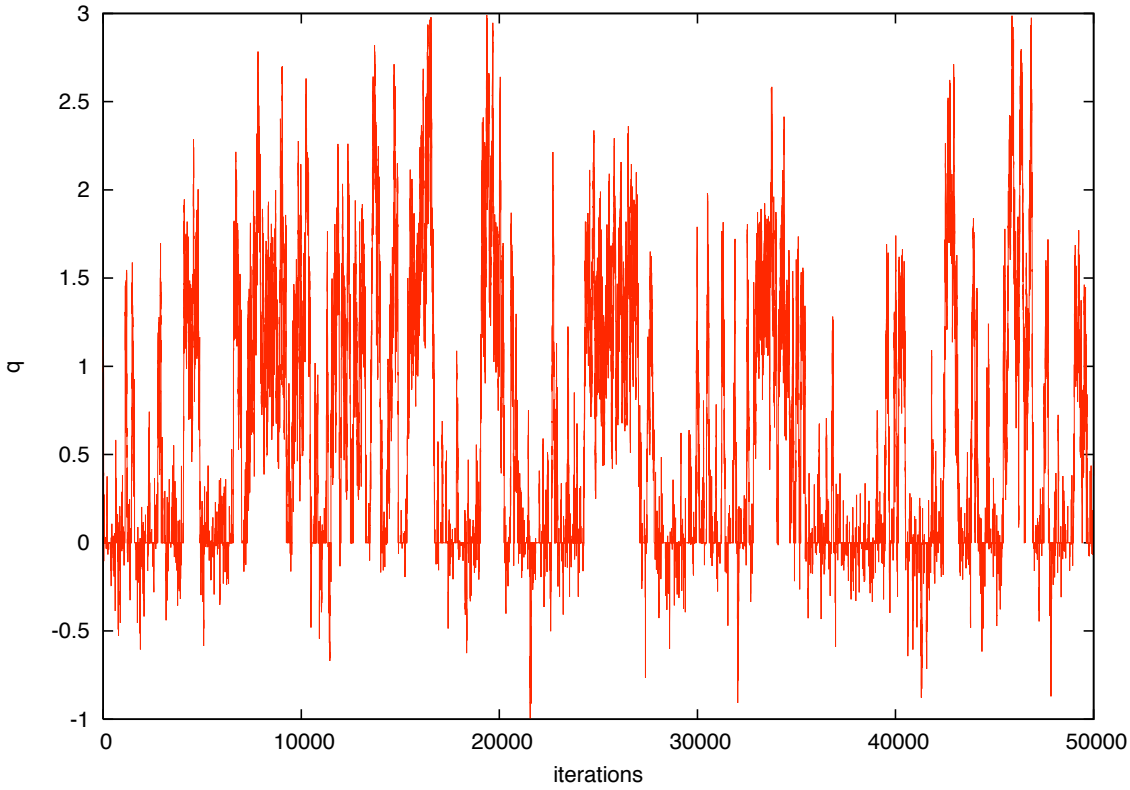


Figure 9: 50,000 iteration segment of an RJMCMC chain moving between models with and without frequency evolution. This particular example is for a source with $q = 1$ and SNR $= 10$, yielding $B_{10} \sim 1$.

Figure 9 shows an example chain for an RJMCMC search of a simulated LISA data stream containing the signal from a galactic binary with $q \equiv \dot{f} T_{\rm obs}^2 = 1$ and an observation time of $T_{\rm obs} = 2$ years. The unitless parameter q is the number of frequency bins the carrier frequency of the GW signal moves over the course of the observation time. The RJMCMC chain moved freely between the 7-dimensional model with $q = 0$ and the 8-dimensional model which included the frequency evolution as a free parameter.

The Laplace Approximation: A common approach to model selection is to approximate the model evidence directly. Working under the assumption that the PDF

is Gaussian, the integral in equation 46 can be estimated by use of the Laplace approximation. This is accomplished by comparing the volume of the models parameter space V to that of the parameter uncertainty ellipsoid ΔV

$$p(s|\mathcal{M}_i) \simeq p(\vec{\theta}_{i,\mathrm{MAP}}|s,\mathcal{M}_i)\Big(\frac{\Delta V_i}{V_i}\Big). \tag{73}$$

The uncertainty ellipsoid can be determined by calculating the determinant of the Hessian $\mathcal{H}$ of partial derivatives of the posterior with respect to the model parameter evaluated at the MAP value for the parameters.

$$p(s|\mathcal{M}_i) \simeq p(\vec{\theta}_{i,\mathrm{MAP}}|s,\mathcal{M}_i)\frac{(2\pi)^{D/2}}{\sqrt{\det\mathcal{H}}} \tag{74}$$

The Fisher Information Matrix (FIM), $\mathbf{\Gamma}$, can be used as a quadratic approximation to $\mathcal{H}$ yielding

$$p(s|\mathcal{M}_i) \simeq p(\vec{\theta}_{i,\mathrm{MAP}}|s,\mathcal{M}_i)\frac{(2\pi)^{D/2}}{\sqrt{\det\mathbf{\Gamma}}} \tag{75}$$

We will refer to this estimate of the evidence as the Laplace-Fisher (LF) approximation. The LF approximation breaks down if the priors have large gradients in the vicinity of the MAP parameter estimates. The FIM estimates can also be poor if some of the source parameters are highly correlated, or if the quadratic approximation fails. In addition, the FIM approximation gets progressively worse as the SNR of the source decreases.

A more accurate (though more costly) method for estimating the evidence is the Laplace-Metropolis (LM) approximation which employs the PDF as mapped out by the MCMC exploration of the likelihood surface to estimate $\mathcal{H}$ [55]. This can be accomplished by fitting a minimum volume ellipsoid (MVE) to the D-dimensional posterior distribution function. The principle axes of the MVE lie in eigen-directions of the distribution which generally do not lie along the parameter directions. Here we employ the `MVE.jar` package which utilizes a genetic algorithm to determine the

MVE of the distribution and returns the covariance matrix of the PDF [56]. The determinant of the covariance matrix can then be used to determine the evidence via

$$p(s|\mathcal{M}_i) \simeq p(\vec{\theta}_{i,\mathrm{MAP}}|s,\mathcal{M}_i)(2\pi)^{D/2}\sqrt{\det \mathbf{C}}. \tag{76}$$

In the MCMC literature the LM approximation is generally considered to be second only to the RJMCMC method for estimating Bayes Factors between high dimension models.

Savage Dickie Density Ratio: Both RJMCMC and LM require exploration of the posterior for each model under consideration. The Savage-Dickie (SD) approximation estimates the Bayes Factor directly while only requiring exploration of the highest dimensional space [13, 57]. This approximation requires that two conditions are met: $\mathcal{M}_j$ must be nested within $\mathcal{M}_i$ (adding and subtracting parameters clearly satisfies this condition) and the priors for any given model must be separable, i.e.,

$$p(\vec{\theta}_i,\mathcal{M}_i) = \prod_k^D p(\theta_i^k,\mathcal{M}_i) \tag{77}$$

which is satisfied in our example. The Bayes factor B_{ij} is then calculated by comparing the weight of the marginalized posterior to the weight of the prior distribution for the 'extra' parameter at the default, lower-dimensional, value for the parameter in question.

$$B_{ij} \simeq \frac{p(\theta_i^0|s,\mathcal{M}_i)}{p(\theta_i^0,\mathcal{M}_i)} \tag{78}$$

It is interesting to note that if the above conditions are precisely satisfied it can then be shown that this is an exact calculation of B_{ij} (assuming sufficient sampling of the PDF), as opposed to an approximation.

Schwarz-Bayes Information Criterion: All of the approximations discussed so far depend on the supplied priors $p(\vec{\theta}_i, \mathcal{M}_i)$. The Schwarz-Bayes Information Criterion (BIC) method is an approximation to the model evidence which makes its own assumptions about the priors – namely that they take the form of a multivariate Gaussian with covariance matrix derived from the Hessian $\mathcal{H}$ [55, 58]. The BIC estimate for the evidence is then

$$\ln p(s|\mathcal{M}_i) \simeq \ln p(\vec{\theta}_{i,\mathrm{MAP}}|s, \mathcal{M}_i) - \frac{D_i}{2} \ln N_{\mathrm{eff}} \tag{79}$$

where D_i is the dimension of model $\mathcal{M}_i$ and N_{eff} is the *effective* number of samples in the data. For our tests we defined N_{eff} to be the number of data points required to return a (power) signal-to-noise ratio of $\mathrm{SNR}^2 - D$, where SNR is the signal-to-noise one gets by summing over the entire LISA band. This choice was motivated by the fact that the variance in SNR^2 is equal to D^2, so N_{eff} accounts for the data points that carry significant weight in the model fit. The BIC estimate has the advantage of being very easy to calculate, but is generally considered less reliable than the other techniques we are using.

Case Study

To compare the various approximations to the Bayes Factor we simulated a 'typical' galactic binary. The injected parameters for our test source can be found in table 2. For detached binaries $\dot{f} \propto f^{11/3}$ so higher frequency sources are more likely to have a measurable $\dot{f}$. On the other hand, galaxy synthesis models show the number of binaries per frequency bin falls off as $\sim f^{-11/3}$, so high frequency systems are fairly rare. As a compromise, we selected a system with a GW frequency of 5 mHz. To describe the frequency evolution we introduced the dimensionless parameter

$$q \equiv \dot{f} T_{\mathrm{obs}}^2, \tag{80}$$

which measures the change in the Barycenter GW frequency in units of $1/T_{\rm obs}$ frequency bins. For $q \gg 1$ it is reasonable to believe that a search algorithm will have no difficulty detecting the frequency shift. Likewise, for $q \ll 1$ it is unlikely that the frequency evolution can be detected (at least for sources with modest SNR). Therefore we have selected $q \sim 1$ to test the model selection techniques. Achieving $q = 1$ for typical galactic binaries at 5 mHz requires observation times of approximately two years. A range of SNRs were explored by varying the distance to the source while keeping all other signal parameters, and the noise realization with which the signal is competing, fixed.

We can rapidly calculate accurate waveform templates using the fast-slow decomposition described in chapter 2. Our waveform algorithm has been used in the second round of Mock LISA Data Challenges [59, 60] to simulate the response to a galaxy containing some 26 million sources. The simulation takes just a few hours to run on a single 2 GHz processor.

f/mHz	$\cos\theta$	ϕ/degree	ψ/degree	$\cos\iota$	φ_0/degree	q	$T_{\rm obs}$/year
5.0	1.0	266.0	51.3	0.2	205.0	1.0	2.0

Table 2: Galactic binary signal parameters

We simulated a 1,024 frequency-bin snippet of LISA data around 5 mHz that included the injected signal and stationary, gaussian instrument noise. The Markov chains were initialized at the injected source parameters as the focus of this study is the character of the detection, and not the search (a highly efficient search algorithm is described in Ref. [32]). We initially used uniform priors for all of the parameters, with the angular parameters taking their standard ranges. We took the frequency to lie somewhere within the data snippet, and $\ln A$ to be uniform across the range $\frac{1}{2}\ln(S_n/(2T))$ and $\frac{1}{2}\ln(1000S_n/(2T)$, which roughly corresponds to SNRs between 1

and 1000. We took the frequency evolution parameter q to be uniformly distributed between -3 and 3 and adopted $q = 0$ as the default value when operating under the 7-dimensional (no frequency evolution) model. In reality, astrophysical considerations yield a very strong prior for q that will significantly impact model selection. We decided to use a simple uniform prior to compare the various approximations to the Bayes Factor, before moving on to consider the effects of the astrophysical prior.

Comparison Of Techniques

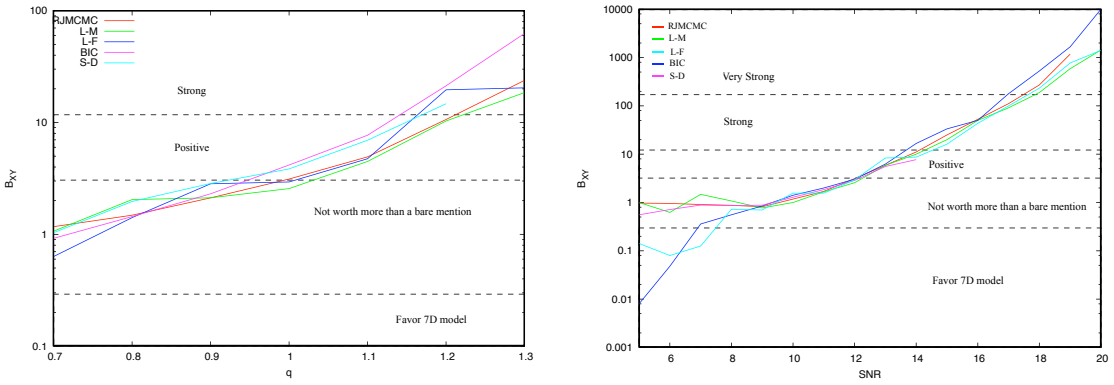
We compared the Bayes Factor estimates obtained using the various methods in two ways. First, we fixed the frequency derivative of the source at $q = 1$ and varied the SNR between 5 and 20 in unit increments. Second, we fixed the signal to noise ratio at SNR = 12 and varied the frequency derivative of the source.

The results of the first test are shown in Figure 10a. We see that all five methods agree very well for SNR > 7. As expected, the Laplace-Metropolis and Savage-Dickie methods provide the best approximation to the "Gold Standard" RJMCMC calculation, showing good agreement all the way down to SNR = 5. Most importantly, all five methods agree on when the 8-dimensional source model is favored over the 7-dimensional model, placing the transition point at SNR $\simeq$ 12.2. To get a rough estimate for the numerical error in the various Bayes Factor estimates we repeated the SNR= 15 case 10 times using different random number seeds to initialize the chains. We found that the numerical error was enough to account for any quantitative differences between the estimates returned by the various approaches.

It is interesting to compare the Bayesian model selection results to the frequentist "3-σ" rule for positive detection:

$$|\bar{q}| > 3\sigma_q, \tag{81}$$

where $\bar{q}$ is the MAP estimate for the frequency change and σ_q is the standard deviation in q as determined by the FIM. For the source under consideration we found satisfaction of the "3-σ" rule required SNR $\simeq 13$ for a detection, in good agreement with the Bayesian analysis. This lends support to Seto's [61] earlier FIM based study of the detectability of the frequency evolution of galactic binaries, but we should caution that the literature is replete with examples where the "3-σ" rule yields results in disagreement with Bayesian model selection and common sense [62]. The results of the second test are displayed in Figure 10b. In this case all five methods produced results that agree to within numerical error.



(a) (b)

Figure 10: Bayes factor plots for each of the methods described in the text as a function of a) q with SNR $= 12$ and b) SNR with $q = 1$.

While the results shown here are for a particular choice of source parameters, we found similar results for other sets of source parameters. In general all five methods for estimating the Bayes Factor gave consistent results for signals with SNR > 7. One exception to this general trend were sources with inclinations close to zero, as then the PDFs tend to be highly non-gaussian. The Laplace-Metropolis and Laplace-Fisher approximations suffered the most in those cases.

Astrophysical Priors

Astrophysical considerations lead to very strong priors for the frequency evolution of galactic binaries. The detached systems, which are expected to account for the majority of LISA sources, will evolve under gravitational radiation reaction in accord with the leading order quadrupole formula:

$$\dot{f} = \frac{3(8\pi)^{8/3}}{40} f^{11/3} \mathcal{M}^{5/3}\,, \tag{82}$$

where $\mathcal{M}$ is the chirp mass. Contact binaries undergoing stable mass transfer from the lighter to the heavier component are driven to longer orbital periods by angular momentum conservation. The competition between the effects of mass transfer and gravitational wave emission lead to a formula for $\dot{f}$ with the same frequency and mass scaling as equation 82, but with the opposite sign and a slightly lower magnitude [28].

Population synthesis models, calibrated against observational data, yield predictions for the distribution of chirp masses $\mathcal{M}$ as a function of orbital frequency. These distributions can be converted into priors on q. In constructing such priors one should also fold in observational selection effects, which will favor systems with larger chirp mass (the GW amplitude scales as $\mathcal{M}^{5/6}$). To get some sense of how such priors will affect the model selection we took the chirp mass distribution for detached systems at $f \sim 5$ mHz from the population synthesis model described in Ref. [28] and used equation 82 to construct the prior on q shown in Figures 11 and 12 (observation selection effects were ignored). The prior has been modified slightly to give a small but non-vanishing weight to sources with $q = 0$. The astrophysically motivated prior has a very sharp peak at $q = 0.64$, and we adopt this as the default value for the frequency derivative parameter when in the 7-dimensional model.

To explore the impact on model selection when such a strong prior has been adopted we simulated a source with $q = 1$ and varied the SNR. The RJMCMC algo-

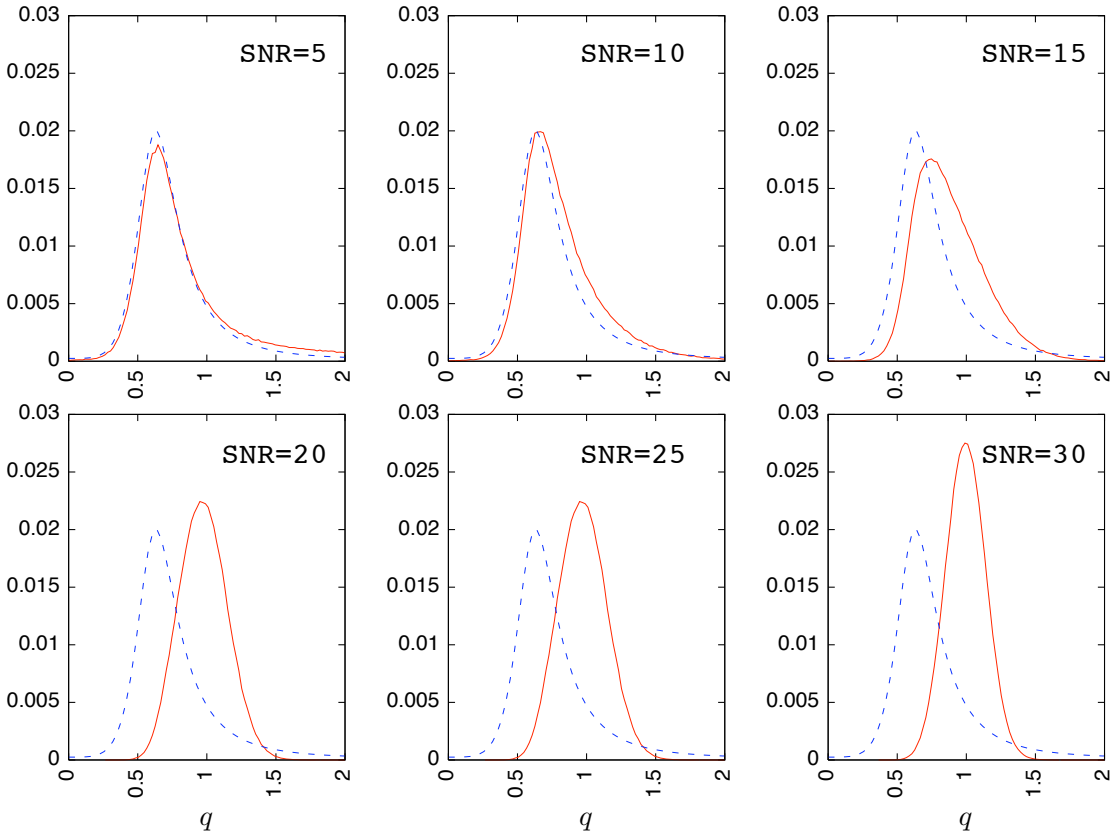


Figure 11: Comparison between astrophysically motivated prior distribution of q for $f = 5$ mHz and $T_{\rm obs} = 2$ years (dashed, blue) to marginalized PDF (solid, red) for sources injected with $q = 1$ and SNRs varying from 5 to 30.

rithm was applied using chains of length 10^7 in conjunction with a fixed 8-dimensional MCMC (also allowed to run for 10^7 iterations) in order to compare the RJMCMC results with the Savage-Dickie density ratio.

The results of this first exploration are shown in Fig. 11. We found that for SNR < 15 the marginalized PDF very closely resembled the prior distribution. This demonstrates that the information content of the data is insufficient to change our prior belief about the value of the frequency derivative. As the SNR increased, however, the PDF began to move away from the prior until we reached SNR $= 30$ when

SNR	B_{10} (SD)	B_{10} (RJMCMC)
5	0.926	1.015
10	0.977	0.996
15	0.749	0.742
20	0.427	0.427
25	0.176	0.177
30	0.060	0.056

Table 3: Savage-Dickie density ratio estimates of B_{10} for sources with $q = 1$ and SNRs varying from 5 to 30. Comparisons with RJMCMC explorations of the same data set show excellent agreement between the two methods.

the astrophysical prior had a negligible effect on the shape of the posterior, signaling high confidence in the MAP value of q and its inclusion as a model parameter. This qualitative assessment of model preference is strongly supported by the Bayes factor estimation made by the RJMCMC algorithm as can be seen in Table 3. It should also be noted that there is outstanding agreement between the RJMCMC and S-D estimates for Bayes factor B_{10}. Both methods indicate that for the chosen value of $q = 1$, the signal-to-noise needs to exceed SNR ~ 25 for the 8-dimensional model to be favored. This is in contrast to the case discussed earlier where a uniform prior was adopted for the frequency derivative, and the model selection methods began showing a preference for the 8-dimensional model around SNR = 12.

q	B_{10} (SD)	B_{10} (RJMCMC)
0.35	1.412	1.414
0.48	1.381	1.388
0.83	1.059	1.052
1.15	0.432	0.428

Table 4: Savage-Dickie and RJMCMC density ratio estimates of B_{10} for sources with SNR=15 and q at FWHM and FWQM of astrophysical prior

Figure 12 shows the impact of the astrophysically motivated prior when the SNR was held at 15 and four different injected values for q were adopted, corresponding

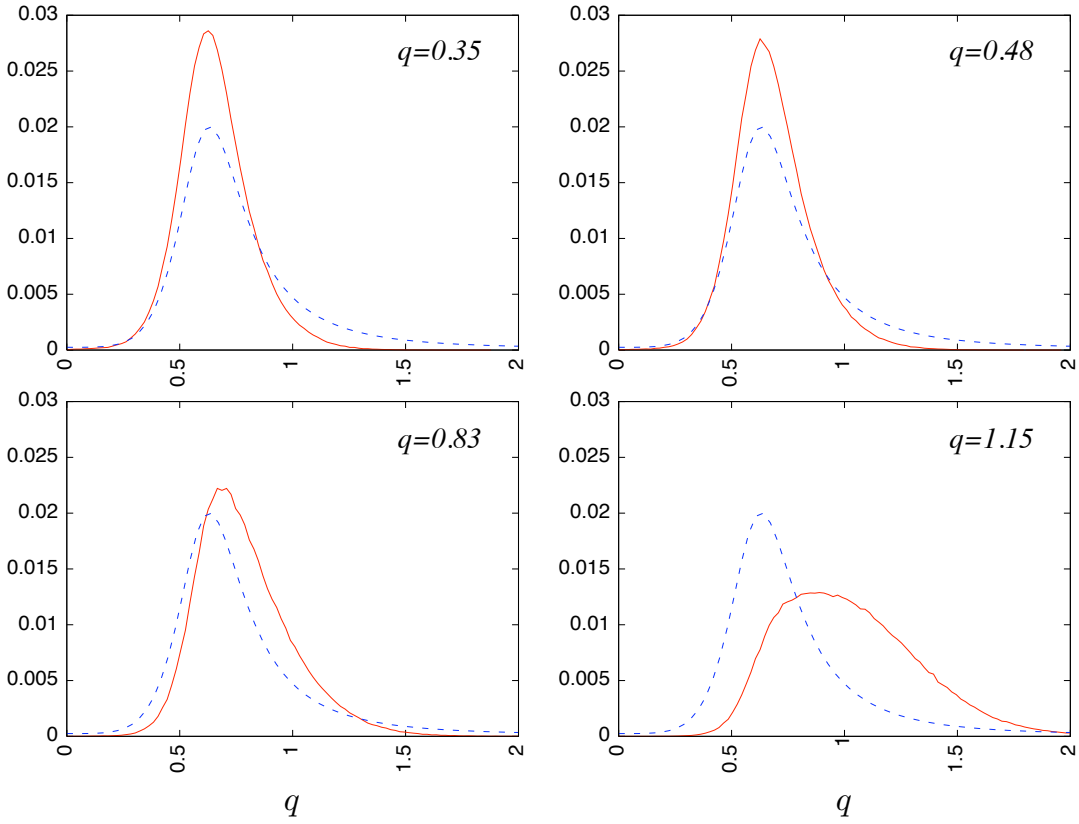


Figure 12: Marginalized PDF (solid, red) for fixed SNR=15 injected sources with q corresponding to FWHM and FWQM of the astrophysical prior (dashed, blue)

to the full width at half maximum (FWHM) and full width at quarter maximum (FWQM) of the prior distribution. The Bayes factors listed in Table IV indicate that for modestly loud sources with SNR=15 the model selection techniques do not favor updating our estimate of the frequency derivative until the frequency derivative exceeds $q \sim 1.2$.

Discussion

We have found that the several common methods for estimating Bayes Factors give good agreement when applied to the model selection problem of deciding when

the data from the LISA observatory can be used to detect the orbital evolution of a galactic binary. The methods studied require varying degrees of effort to implement and calculate, and although found to be accurate in this test case, it is clear that some of these methods would be inappropriate approximations for more physically relevant examples.

If a RJMCMC algorithm is used as the sole model selection technique, the resistance of the algorithm to change dimension, especially when making multi-dimensional jumps, can result in invalid model selection unless the chains are run for a very large numbers of steps. In the examples we studied the trans-dimensional jumps only had to span one dimension, and our basic RJMCMC algorithm performed well. However, a more sophisticated implementation, using *e.g.* rejection sampling or coupled chains, will be required to select the number of sources, as this requires jumps that span seven or more dimensions.

The Laplace-Metropolis method for approximating the model evidence is more robust than the commonly used Fisher Information Matrix approximation of the Hessian of the PDF. However, implementing an LM evidence estimation is drastically more costly because of the need to fit the posterior to a minimum volume ellipsoid.

The Savage-Dickie approximation is more economical than the RJMCMC or LM methods, but is limited by the requirement that the competing models *must* be nested.

The Bayes Information Criterion approximation to the evidence is by far the cheapest to implement, and is able to produce reliable results when the SNR is high. It has therefore shown the most promise as an 'on the fly' model determination scheme. More thorough (and therefore more costly) methods such as RJMCMC and LM could then be used to refine the conclusions initially made by the BIC.

Our investigation using a strong astrophysical prior indicated that the gravitational wave signals will need to have high signal-to-noise (SNR > 25), or moderate

signal-to-noise (SNR > 15) and frequency derivatives far from the peak of the astrophysical distribution, in order to update our prior belief in the value of the frequency derivative. In other words, the frequency derivative will only be needed as a search parameter for a small number of bright, high frequency sources.

THE GRAVITATIONAL WAVE DETECTION PROBLEM

Statement of the Problem

The ability to determine whether data contains just instrument noise, or noise plus a resolvable gravitational wave signal, is the central problem in gravitational wave astronomy. The solution to the "detection problem" can be divided into three stages:

1. *Search*: Determine the region(s) in parameter space containing the maximum posterior weight to establish a candidate gravitational wave signal.

2. *Characterization*: Establish confidence intervals for the parameters that describe the signal model and the noise model.

3. *Evaluation*: The detection hypothesis is compared to alternatives through model selection.

The data from the current generation of gravitational wave detectors is, for the most part, analyzed using frequentist techniques, whereby some "detection statistic" is introduced that is designed to respond differently to gravitational wave signals and instrument noise. Standard choices of detection statistics include signal-to-noise or likelihood ratios, and measures of correlation across the network. Candidate events are identified using these statistics, and the significance of a given event is established from Monte Carlo studies of signal injections and time slides of the data. The hypotheses that are being compared, and the assumptions that are being made, are not clearly spelled out in this approach, and in some cases may even be unphysical [40].

The Bayesian approach to model selection forces us to define explicit models for the hypotheses under consideration. Once the models are defined the analysis is

entirely mechanical. The end result is an odds ratio for one hypothesis over another. We will see the Bayesian approach does not yield a fixed signal-to-noise threshold at which a class of signal becomes detectable. Rather, the signal-to-noise ratio at which a particular instance of the signal, in a particular noise realization, becomes detectable can vary substantially.

The competing models are written as

$$\begin{aligned} \mathcal{M}_0(\vec{\theta}_0) &: \tilde{s}(f) = \tilde{n}(f) \\ \mathcal{M}_1(\vec{\theta}_1) &: \tilde{s}(f) = \tilde{n}(f) + \tilde{h}(f) \end{aligned} \tag{83}$$

with

$$\begin{aligned} \vec{\theta}_0 &= \sum_j \vec{\eta}_j \\ \vec{\theta}_1 &= \vec{\lambda} + \sum_j \vec{\eta}_j \end{aligned} \tag{84}$$

where $\vec{\lambda}$ denotes GW signal parameters and $\vec{\eta}_j$ are the noise parameters for the j^{th} interferometer channel. A note of caution here: The analysis is not answering the question "is there a gravitational wave signal present in the data, or is it just instrument noise?", but rather, "is the data most consistent with our model of the gravitational wave signal and instrument noise, or our model of the instrument noise alone?" In particular, if our model of the instrument noise is poor - such as not allowing for occasional glitches - then the odds may favor the detection hypothesis even when no signal is present in the data. The goal is to construct models for the signals and instrument noise that are sufficiently realistic so that the hypotheses that are being tested closely approximate the physical situations we hope to compare.

Here we will describe a fully Bayesian solution to the detection problem. All three stages of the analysis are handled by a single technique - the Parallel Tempered Markov Chain Monte Carlo (PTMCMC) detection algorithm. The PTMCMC

algorithm establishes which regions of parameter space contain the highest posterior weight, efficiently explores the posterior distribution function (PDF) of the model parameters, and calculates the marginalized likelihood, or evidence, for the model. This procedure is repeated for the competing hypotheses under consideration (e.g. instrument noise, instrument noise plus a black hole inspiral signal, instrument noise plus N galactic binary signals) and the evidence ratios, or Bayes factors, are then used to identify the hypothesis that is most consistent with our prior belief and the current data. The Bayes factors computed from the PTMCMC algorithm are checked against those computed using a Reverse jump Markov Chain Monte Carlo (RJMCMC) algorithm, and are found to be in very good agreement.

Several solutions to components of the GW detection problem have recently been successfully demonstrated using methods very similar to our own. Parallel tempering has been shown to effectively satisfy the search and characterization phase of ground-based data containing a candidate detection of a spinning binary inspiral [49]. A model selection study of when an inspiral waveform injected into simulated LIGO data has sufficient signal-to-noise to be favored over a model containing only instrument noise has been performed using the Nested Sampling algorithm [63, 64]. Our work here is unique in that it employs a single analysis algorithm to satisfy *all three phases* of the detection problem, doing so without any constraints supplied by some candidate detection.

The PTMCMC Detection Algorithm

Parallel Tempering

Although our proposal distributions permit exploration of the entire prior volume, the bolder jumps are rarely accepted as the chance of making large changes to a

large number of parameters and still ending up at a location with decent likelihood is vanishingly small. Furthermore, it is our desire that this Markovian chain can also be used to *find* the injected signal even if it is started randomly within the prior volume. A straight forward implementation of the MCMC algorithm will never (practically) find the MAP for parameter spaces of this complexity without some assistance. Usually this assistance comes in a way that violates the detailed balance condition between subsequent iterations and thus nullifies the Markovian nature of the chain (resulting in biased samples which do not mirror the PDF). This is of no negative consequence as long as these non-Markovian steps in the chain are discarded as burn-in samples, but it does put tremendous pressure on the burn-in to put the chain very near to the global maximum.

To encourage efficient global sampling of the target distribution, which in turn expedites the convergence time of the chain, while preserving detailed balance we have adopted the powerful technique of *parallel tempering* [65], where multiple chains explore the data simultaneously, each sampling from an iteratively higher "temperature" target distribution

$$p(\vec{\theta}|s,\beta) \propto p(\vec{\theta})p(s|\vec{\theta})^{\beta} \tag{85}$$

where β is analogous to an inverse temperature. This effectively "softens" the likelihood surface allowing the chain easier access to positions in parameter space that would otherwise be difficult to reach.

We explore the parameter space with N_c chains, each at higher temperature, where new locations are adopted for each chain using the Hastings ratio in Eq. 58 with the likelihood from Eq. 85. The high temperature (low β) chains effectively sample the prior distributions as the likelihood ratio, raised to the power of β, is suppressed for

small β. Cold chains are more apt to explore regions of high posterior weight and are prone to "sticking" on local features of the likelihood surface.

The great benefit of parallel tempering derives from the ability for information from "hot" chains to propagate towards colder chains (and vice versa) by randomly proposing parameter exchanges between temperature levels. An exchange of parameters between the i^{th} and j^{th} chain satisfies detailed balance if the conditional probability is evaluated with Hastings ratio:

$$H_{i\leftrightarrow j} = \frac{p(s|\vec{\theta}_i, \mathcal{M}_i, \beta_j)p(s|\vec{\theta}_j, \mathcal{M}_j, \beta_i)}{p(s|\vec{\theta}_i, \mathcal{M}_i, \beta_i)p(s|\vec{\theta}_j, \mathcal{M}_j, \beta_j)} \tag{86}$$

Only points in the $\beta = 1$ chain sample the true posterior and are therefore permitted to contribute to the Markov chain from which the PDF is inferred. However, by exchanging parameters with the hot chains the $\beta = 1$ chain rapidly explores the full target distribution revealing a wealth of information about the posterior. This is in contrast to chains without parallel tempering which are prone to "sticking" at or near the maximum of the posterior, or worse, some secondary maxima. Although non-parallel-tempered chains efficiently sample the region around some maxima they can completely miss (for a finite run time) the more complex structure of the target distribution.

Although parallel tempering increases the computational cost of each iteration, for well constructed heating schemes, N_c chains running for I iterations generically converges faster than one chain allowed to run for $N_c \times I$ iterations. Figure 13 clearly demonstrates the advantage of augmenting a typical MCMC with parallel tempering. The figure shows two dimensional marginalized posterior distribution functions as sampled by an MCMC as described above. The data contained simulated LISA noise and the injected source was a single galactic binary with a SNR = 7. For demonstrative purposes the chains were initiated at the injected signal parameters

and underwent 1.5×10^6 iterations, the first 5×10^5 being discarded as burn in samples. The PDFs are for the $\ln A - fT_{\rm obs}$ and $\cos\theta - \phi$ planes. The MCMC without parallel tempering clearly samples the region around the injected parameters but is confined to that location in parameter space and would need to run beyond a time that is practical in order to sample the full distribution. The parallel tempered MCMC samples the full PDF very efficiently, and for this example discovers a global maximum that significantly differs from the injected parameters. This demonstrates how the MAP parameters can be pushed away from the injected parameters by the noise.

This particular example also demonstrates the harm in being solely interested in the "best fit" parameters, as opposed to the full distribution. Had a search been performed over this data (as opposed to starting the chain at the injected parameters) through some MCMC driven technique (such as using simulated annealing during the burn-in phase) without parallel tempering it is likely that *only* the global maximum will have been sampled by the post burn-in MCMC. The recovered signal parameters would be measurably different from the physical parameters of the source which is producing the gravitational waves. This may be of little consequence for a single galactic binary, but for more exotic signals where source parameters will be used to learn about important astrophysics, or where optical counterparts will be of great interest, it is clear that the full PDF is vital for accurate understanding of the underlying process. Alternatively, a well constructed PTMCMC algorithm will simultaneously locate maxima in the search space (satisfying the search phase of the detection problem) while accurately and efficiently mapping out the posterior distribution function (characterizing the gravitational wave source).

Although only the $\beta = 1$ chain samples the target distribution in the PTMCMC approach, the higher temperature chains serve as more than just a convergence aid.

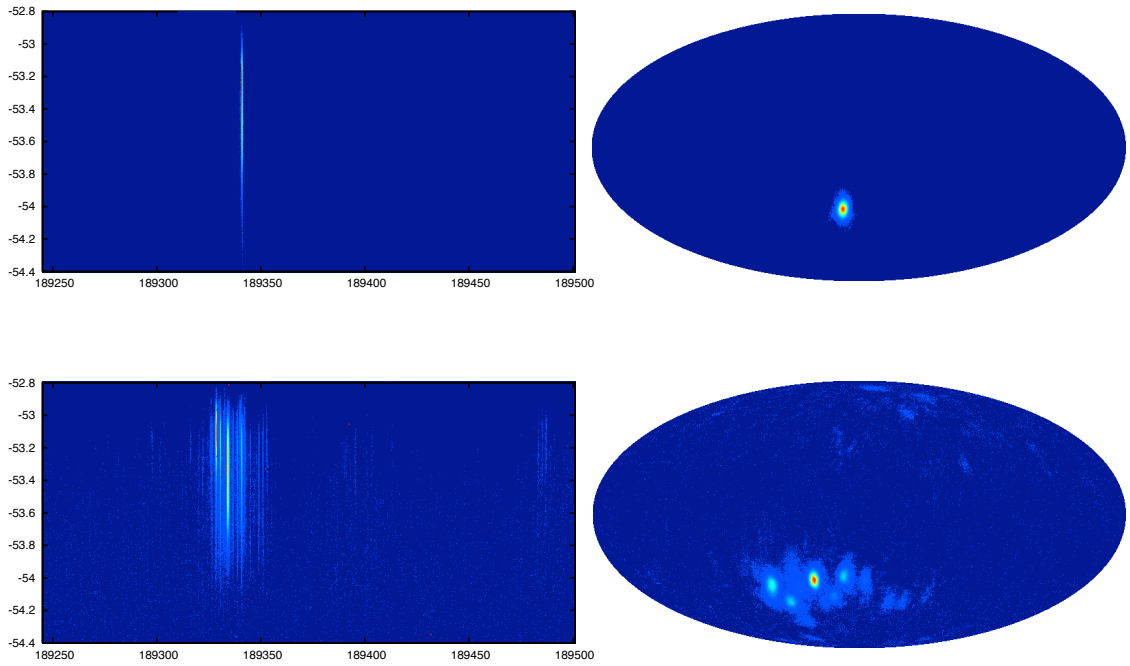


Figure 13: Two dimensional marginalized posterior distribution functions of a single galactic binary with $SNR = 7$ as sampled by a MCMC. On the left is the $\ln A - fT_{\rm obs}$ plane, the right shows a Molweide projection of the sky location (parameters θ and ϕ). Red (white) locations have the highest log probability density while blue (black) has the least. The top panel is for an MCMC without parallel tempering where the chain was started at the injected parameters and allowed 500,000 steps of burn-in. The bottom panel is the same scenario only now with twenty parallel chains spaced geometrically in heat with a maximum temperature of 100 ($\beta_{\rm max} = 0.01$). Notice how much more efficiently the PTMCMC samples the entire posterior, resolving the more detailed structure of the PDF, while the straight-forward MCMC never leaves the region around injected parameter values, missing the global maximum entirely.

These additional chains can also be utilized to calculate the model evidence. For normalizable priors we can evaluate $p(d|\mathcal{M})$ from Eq. 46 by using the expectation value of the likelihood for each temperature level. First we consider the evidence for each temperature's posterior distribution function as part of some partition function $Z(\beta)$ where

$$\begin{aligned} Z(\beta) &\equiv \int d\vec{\theta}\; p(d|\vec{\theta},\mathcal{M},\beta)p(\vec{\theta},\mathcal{M}) \\ &= \int d\vec{\theta}\; p(d|\theta,\vec{\mathcal{M}})^{\beta}p(\vec{\theta},\mathcal{M}). \end{aligned} \tag{87}$$

Because the prior is independent of β we can write the partition function as

$$\frac{d}{d\beta}\ln Z(\beta) = \langle \ln p(d|\vec{\theta},\mathcal{M})\rangle_{\beta} \tag{88}$$

where $\langle \ln p(d|\vec{\theta},\mathcal{M})\rangle_{\beta}$ is the expectation value of the likelihood for the chain with temperature $1/\beta$. Then the evidence can be found by integrating over β via

$$\ln p(d|\mathcal{M}) = \int_0^1 d\beta\; \langle \ln p(d|\vec{\theta},\mathcal{M})\rangle_{\beta}. \tag{89}$$

Alternatively, after the data has been analyzed under both models the Bayes factor is calculated by

$$\begin{aligned} \ln B_{10} &= \int_{-\infty}^{0} \beta(\langle \ln p(d|\vec{\theta}_1,\mathcal{M}_1)\rangle_{\beta} \\ &- \langle \ln p(d|\vec{\theta}_0,\mathcal{M}_0)\rangle_{\beta})\; d\ln\beta. \end{aligned} \tag{90}$$

The expectation value of the likelihood is evaluated over the post-burn-in chain iterations [37, 11]. This technique, known as thermodynamic integration (TI), is a direct calculation (instead of an approximation) to the evidence [43]. It is not necessary to thermodynamic integration that the different temperature chains be allowed to exchange parameters, however having them to do so ensures healthy mixing and convergence.

The evidence ratio between competing models $\mathcal{M}_1$ and $\mathcal{M}_0$ then yields the Bayes factor for the models in question. We now have a single algorithm that, if implemented effectively, solves all three facets of the detection problem free of any approximations which may render the result suspect.

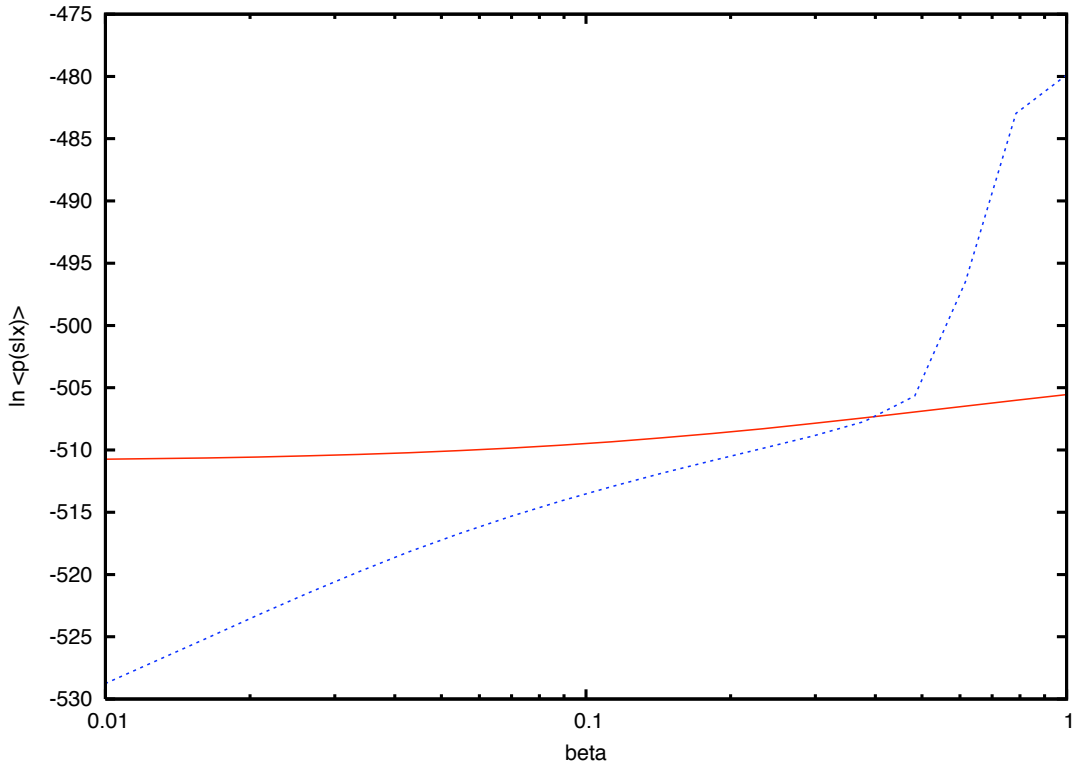


Figure 14: Average likelihood for chains of different heats $\beta = 1/T$. The red (solid) line is for the noise only model $\mathcal{M}_0$ while the blue (dashed) line is for the signal plus noise model $\mathcal{M}_1$. This particular example was for data containing a SNR = 8 source.

Reverse Jump Markov Chain Monte Carlo

Although TI is robust and converges predictably, we feel that for "high stakes" model selection problems (such as the first GW detection) several complementary methods should be used to support the detection model over any alternative hypotheses in order to establish additional confidence in the detection. One approach

for such a case is to use alternative techniques for calculating the Bayes factor. For this additional checking stage we have chosen the "gold standard" of Bayes factor calculators: The reverse jump Markov chain Monte Carlo (RJMCMC) algorithm.

Trans-dimensional Proposal Distribution

It took us many attempts to get the RJMCMC routine to work. Our initial efforts used the Fisher approximation to the posterior in the neighborhood of the MAP parameters found from the PTMCMC search/characterization. This technique showed initial promise by producing qualitatively reasonable Bayes factors (monotonically increasing as a function of SNR) and stability (producing similar results through different random seeds.) However, further testing revealed disagreement with the thermodynamic integration and erratic behavior including large increases in Bayes factor for very small increases in SNR, and occasionally questionable results (such as "detectable" sources with SNR well below five). This was confounding until we noticed that the RJMCMC chains were producing PDFs that differed significantly from the PTMCMC results. This signaled poor exploration of the prior and a lack of convergence towards the true distribution. It wasn't until the Fisher approximation became sufficiently accurate (the SNR became sufficiently large) that this proposal distribution mimicked the target distribution well enough to allow for adequate mixing. It was difficult to immediately diagnose, based on the behavior of a RJ chain alone, what the problem was, and without having the thermodynamic integration results for comparison, it would have been difficult to know if alternative approaches fixed the problem.

We settled on a method for constructing trans-model proposals first suggested by Green [66] to start with fixed dimension MCMC trial explorations of each model in question. The resultant joint PDF from this learning period can be used as the

proposal distribution used to move into that model. The implementation of this scheme which would yield the most efficient between-model exploration would be to sample from the fixed-dimension joint PDF directly. Unfortunately, thoroughly sampling the PDF for high dimension models rapidly becomes prohibitively costly. Instead, we approximate the full PDF using the chain from the PTMCMC analysis used in the initial detection.

Sufficiently sampling the full signal posterior would require $10^{10} - 10^{12}$ samples, whereas our typical Markov chains are only of length 10^6 samples. Clearly the full PDF will be woefully under-sampled, but thanks to the excellent mixing from the fixed dimension exploration (courtesy of the parallel tempering) we are confident that the high probability density regions of the PDF *have* been adequately sampled and qualitatively reflect the true joint distribution. To construct the signal parameter proposal distribution $q(\vec{\lambda}, \mathcal{M}_1)$ we first divide the prior volume for each parameter λ^i by the average standard deviation $\bar{\sigma}^i_\Gamma$ of that parameter. This average is taken over the parameter variances calculated each time the Fisher Information Matrix is updated during the PTMCMC analysis.

The prior volume can now be thought of as (for the galactic binary parameterization) an 8-dimensional hypercube with cells of width $\bar{\sigma}^i_\Gamma$. The volume of each cell is $dV = \prod_i \bar{\sigma}^i_\Gamma$ and the total number of cells in the hypercube is $N_{\rm cell} = \prod_i \Delta p(\lambda^i)/\bar{\sigma}^i_\Gamma$ where $\Delta p(\lambda^i)$ is the prior range for parameter i. We then sort through the fixed dimensional signal-model Markov chain from the PTMCMC to populate the cells in the hypercube. Typically between 10^3 to 10^5 cells are actually occupied for fixed-dimension chains of length $N_{\rm MCMC} = 10^6$. If n points in the chain land in the cell with reference parameters $\vec{\lambda}$ then the probability density of jumping into that cell is

$$q_s(\vec{\lambda}, \mathcal{M}_1) = \frac{n(\vec{\lambda})}{dV N_{\rm MCMC}} \tag{91}$$

This is not yet a valid proposal distribution, however, because drawing from $q_s(\vec{\lambda}, \mathcal{M}_1)$ will not allow the signal model transitions to access the entire prior volume. To rectify this we add to the existing distribution a uniform distribution over the entire cube, by giving each cell a single occupant. The uniform piece of the proposal has probability density

$$q_u(\mathcal{M}_1) = \frac{1}{dV N_{\text{cell}}} \tag{92}$$

The full proposal distribution is then the sum of the sampled and uniform cell populations, with relative weighting w

$$q(\vec{\lambda}, \mathcal{M}_1) = wq_s + (1-w)q_u. \tag{93}$$

To ensure proper normalization of the proposal distribution $w \in [0, 1]$.

To draw from this hybrid distribution we first compare a number from $U[0, 1]$ to w. If this exceeds w we uniformly draw the signal parameters from the prior. We then must determine to which cell of the hypercube the new set of parameters belongs so we may evaluate $q(\vec{\lambda}, \mathcal{M}_1)$. When drawing from the PTMCMC-populated portion of the distribution we randomly select one "link" from the fixed dimension chain to select an occupied cell of the hypercube. This guarantees that moves to a particular occupied cell are proposed at the same frequency with which the cell was sampled during the search. We then draw the signal parameters uniformly from within the selected cell. For transitions into $\mathcal{M}_0$ we must again determine where in the cube the current signal parameters reside to accurately calculate the trans-dimensional Hastings Ratio.

RJ Parallel Tempering

Between trans-dimensional proposals the chains update within their current dimension identically to the description in chapter 3. We only allow the $\beta = 1$ chain to

undergo trans-dimensional moves. When this chain is in a given model it has $N_c \sim 20$ chains of the same model, each of increasing temperature, with which to exchange parameters. When trans-dimensional moves are accepted by the "cold" chain these parallel tempered chains stop updating and store their current locations in parameter space, while the analogous chains in the other model resume from their most recent positions. This way the chain sampling from the target distribution is always in contact with a "heat reservoir" of parallel chains, but we can avoid wasting computations for chains which have no way of communicating with the $\beta = 1$ parameters. The schedule for these different types of updates (inter-dimensional MCMC, parallel tempering parameter exchanges, and trans-dimensional proposals) is part of the overall proposal distribution and therefore does not affect the outcome. This schedule does, however, need to be tuned in order to see convergence within a reasonable amount of time.

CASE I: COMPACT GALACTIC BINARIES – LISA

To test our analysis scheme we want to use a toy problem that allows for fast computation of likelihoods, while still being fairly realistic. To that end we generate a 256 frequency-bin segment of simulated LISA data containing instrument noise and the signal from a single chirping galactic binary over an observation time of $T_{\rm obs} = 2$ years.

We first generate a noise realization for the frequency range of interest by drawing from the theoretical LISA noise spectral density (see chapter 5). We then inject a test source somewhere in the data away from the edges of the segment with some tunable amplitude so that we may establish at what signal to noise ratio the source becomes detectable. To be considered effective our algorithm must find the MAP parameters for two distinct models

$$\begin{aligned} \mathcal{M}_0 : \tilde{s}(f) &= \tilde{n}(f) \\ \mathcal{M}_1 : \tilde{s}(f) &= \tilde{n}(f) + \tilde{h}(f), \end{aligned} \tag{94}$$

thoroughly resolve each models PDF, and calculate the evidence for each model (up to some constant factor common between $\mathcal{M}_1$ and $\mathcal{M}_0$). We will consider the prior odds between the two models to be unity, thus allowing us to interpret the evidence ratio (Bayes factor) as the odds ratio.

Modeling the Source: Galactic Compact Binaries

The choice of using galactic binaries as the test source was motivated by the fact that waveforms for these objects can be rapidly modeled using the fast-slow decomposition described in chapter 3. Despite the simplicity of generating these

waveforms, they posses sufficient complexity that we do not anticipate significant additional complications related to the modeling/fitting of other, more exotic sources (the same techniques described here, with a few enhancements, are proving sucessful at handling spinning black holes and extreme mass ratio inspirals).

Also, because of the great number of galactic sources, and the ensuing overlap between individual sources, a one-by-one detection/regression is inaccurate [10]. Therefore a global fit to all of the galactic sources is required. Because of the uncertainty in the number of resolvable sources one can not fix the model dimension *a priori* which presents a crucial model selection problem. Over-fitting the data will result in an inaccurate regression which would then remove power from other sources in the data-stream, negatively impacting their detection and characterization. The Reverse Jump Markov Chain Monte Carlo approach to Bayesian model selection has been used to determine the number of resolvable sources in the context of a toy problem [67, 68] which shares some of the features of the LISA foreground removal problem. Meanwhile the Laplace approximation to Bayesian model selection has been employed to estimate the number of resolvable sources as part of a MCMC based algorithm to extract signals from simulated LISA data streams [10, 32].

Detached, mildly chirping compact binaries are well modeled, and can be parameterized using eight quantities which we will consider as components of a parameter vector $\vec{\lambda}$:

$$\vec{\lambda} \rightarrow (\ln A, fT_{\rm obs}, \dot{f}T_{\rm obs}^2, \cos\theta, \phi, \psi, \cos\iota, \varphi_0) \tag{95}$$

The physical parameters are the amplitude A, GW frequency f (twice the binary's orbital frequency), frequency derivative $\dot{f} = df/dt$, co-latitude and azimuthal sky angles θ and ϕ, GW polarization angle ψ, orbital inclination ι, and phase φ_0 of the

gravitational waves. The quantities f, $\dot{f}$ and φ_0 are defined at some fiducial time such as when the first data are taken.

Modeling the Instrument Noise

The contribution to the data from instrument noise in each Fourier bin, $\tilde{n}(f)$, has spectral density $S_n(f)$ with mean and variance

$$\begin{aligned} \langle \tilde{n}(f)_i \rangle &= 0 \\ \langle \tilde{n}(f)_i \tilde{n}(f)_j \rangle &= \frac{T_{\rm obs}}{2} \delta_{ij} S_n(f) \end{aligned} \tag{96}$$

respectively. For operational detectors this is approximately true but excursions from the expected level are to be anticipated. In preparation for this reality during the LISA mission we developed our algorithm to simultaneously fit for the noise level in the data. We parameterize the noise by assuming that departures from $S_n(f)$ can be characterized by a rescaling of the noise spectral density. This rescaling is unique to each channel (A and E) while $S_n(f)$ is identical for both. A more sophisticated approach would be to fit for the twenty-four components of noise in the LISA constellation (the laser phase noise between, and the position noise of, each test mass). This approach has been developed by Adams and Cornish [69], but has not been implemented here.

The simulated segment of LISA data is divided into four equal length sub-regions in frequency space of size Δf. Because the noise spectral density hardly varies over the entire 256-frequency-bin data segment it can be treated as constant. The central frequency of the data segment is used to evaluate S_n while the noise level in each sub-region i, and interferometer channel α, is rescaled by η^i_α. The modeled noise level

in each frequency window is

$$S_n^{\alpha}(f) \rightarrow \eta_i^{\alpha} S_n. \tag{97}$$

For the case where there are two non-negligable interferometer channels (A and E) a total of eight noise parameters are required. An example of these parameters for a particular noise realization can be seen in Fig. 15.

A consequence of this noise fitting procedure is that the noise spectral density is no longer fixed, and care must be taken to account for this when computing the likelihood. In particular, the normalization of the likelihood depends on the noise parameters. Recall from chapter 3 that we had defined the likelihood as

$$p(s|\vec{\theta}) \equiv Ce^{-\frac{1}{2}(s-h|s-h)}. \tag{98}$$

We account for the varying noise level by elevating the non-constant components of C into the argument of the exponential of equation 98. The likelihoods are now written as

$$\begin{aligned} p(s|\vec{\eta}, \mathcal{M}_0) &\equiv C' \exp\left[-\frac{1}{2}\left((s|s) + N\sum_{\alpha,j} \ln \eta_j^{\alpha}\right)\right] \\ p(s|\vec{\lambda}, \vec{\eta}, \mathcal{M}_1) &\equiv C' \exp\left[-\frac{1}{2}\left((s-h|s-h) + N\sum_{\alpha,j} \ln \eta_j^{\alpha}\right)\right] \end{aligned} \tag{99}$$

where $N = T_{\rm obs}\Delta f$ is the number of Fourier bins over range Δf and the sum over j ranges over the number of Δf segments in the data. Meanwhile our definition of the noise weighted inner product (eq. 62) must also reflect the parameterization of the noise by replacing $S_n^{\alpha}(f)$ with $\eta_i^{\alpha} S_n$.

We have now fully parameterized two competing models which can be used to explain the data in question. Our attention will subsequently turn to applying Bayesian tools to establish the most appropriate configuration of each model given the simulated data, and to learn which model most appropriately describes the data in question.

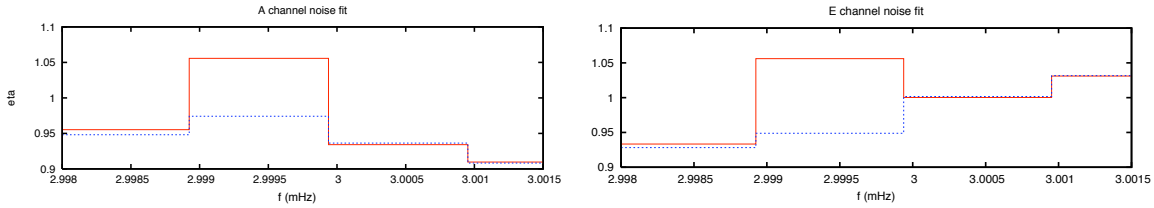


Figure 15: MAP noise parameters for $\mathcal{M}_0$ (red, solid) and $\mathcal{M}_1$ (blue, dashed). The data consists of two interferometer channels, each containing 256 Fourier bins which are divided into four sub-regions. Each sub-division is fitted with a unique noise parameter. The signal is injected somewhere in the second frequency interval of the data. Model $\mathcal{M}_0$ elevates the noise parameter in the second window to account for the excess power caused by the gravitational wave signal. Model $\mathcal{M}_1$ successfully fits to that gravitational wave leaving the noise parameter closer to unity. Notice how the noise parameters for remaining portions of the data are nearly identical between the two models.

Results

Three test sources, chosen to provide representaitve examples over the sky and frequency range of compact binaries, were constructed and injected into several different noise realizations per signal.

Source	f (mHz)	$\cos\theta$	ϕ	$\cos\iota$	ψ	φ_0	$\dot{f}T^2_{\rm obs}$
1	1	0.713	0.452	0.534	2.334	0.624	3.241
2	3	0.598	2.972	0.827	1.240	5.938	0.139
3	5	0.326	4.644	0.133	0.314	3.878	0.643

Table 5: Injected galactic binary parameters.

For each case the analysis was repeated for a range of amplitudes, corresponding to signals with a range of signal-to-noise ratios defined by

$$\mathrm{SNR} \equiv \frac{(s|h(\vec{\lambda}))}{\sqrt{(h(\vec{\lambda})|h(\vec{\lambda}))}} \tag{100}$$

We are then able to learn how the Bayes factor increases as a function of SNR to infer when that particular source, in that particular noise realization, becomes "de-

tectable" ($B_{10} \geq 3$). The data was analyzed using several runs of the fixed dimension PTMCMC (one for each model) with different seeds used to initiate the chains, thus testing the stability of the evidence calculation and establishing the error bars seen in later figures. We used 20 chains in the parallel tempering scheme with a maximum temperature of 100. The temperature spacing of the chains was geometric in β.

Search Phase

The chains were initialized randomly within the parameter space and allowed 5×10^5 iterations of "burn-in" before the samples were considered to be accurately representing the target distribution. As can be seen from an example chain in Figure 16 the $\beta = 1$ chain reached a stationary solution in the vicinity of the maximum log-likelihood within the burn-in phase (the red [solid] portion of the plot) thus signaling the arrival of the search to the region(s) of maximum posterior weight. This demonstrates the success of the search phase in a reasonable number of iterations without non-Markovian convergence aids.

Characterization

The 10^6 iterations post burn-in were used to characterize the candidate detection. The parallel tempering allowed the chains to thoroughly explore the entire posterior distribution function within this alloted time. We were surprised to see a great deal of structure in the signal posterior even in data where no signals were injected. Figure 17 shows the marginalized posterior distributions for both the sky location and the $fT - \ln A$ plane when no souce was present in the data. Notice the concentration of posterior weight at one localized sky location and frequency bin, despite the data being devoid of any injected signal. This strengthens our opinion that rigorous model selection steps are necessary, as even the analysis of noise-only data recovered a

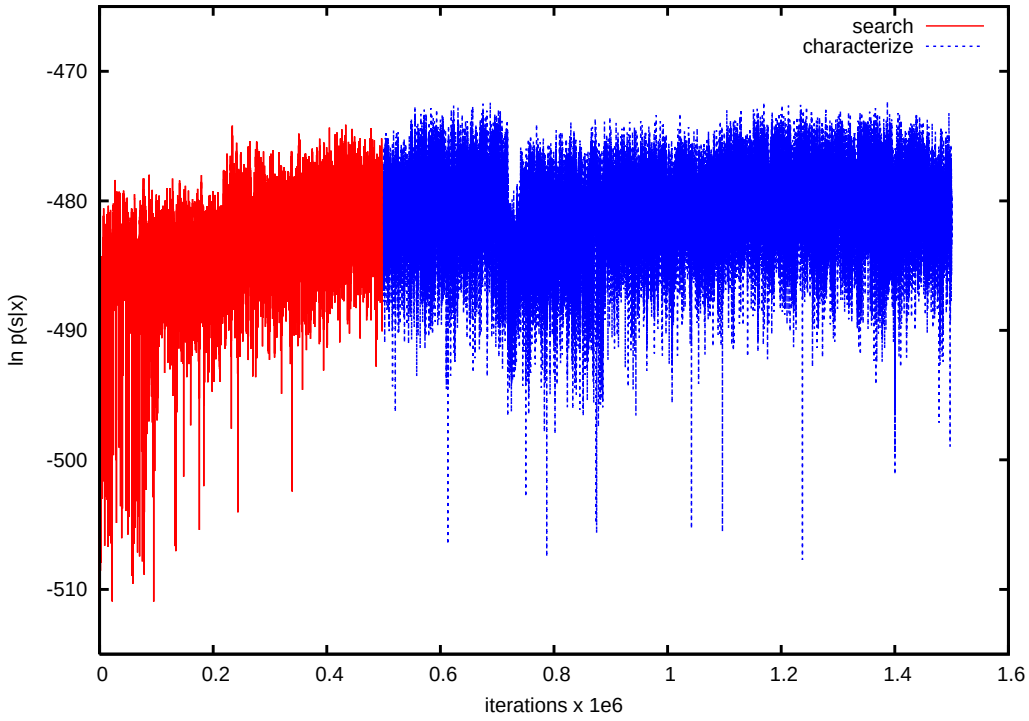


Figure 16: $\ln p(s|\vec{\theta})$ during search (red, solid) and characterization (blue, dashed) phase of the analysis for a SNR $= 8$ source.

"believable" signal (SNR ~ 5) which was well localized in parameter space but easily rejected by our Bayes factor calculations. PDFs of higher signal-to-noise sources continue to demonstrate complicated secondary features in addition to the regions of parameter space near to the injected parameters. Figure 13 (bottom panel) in chapter 5 is an example of this multi-modality.

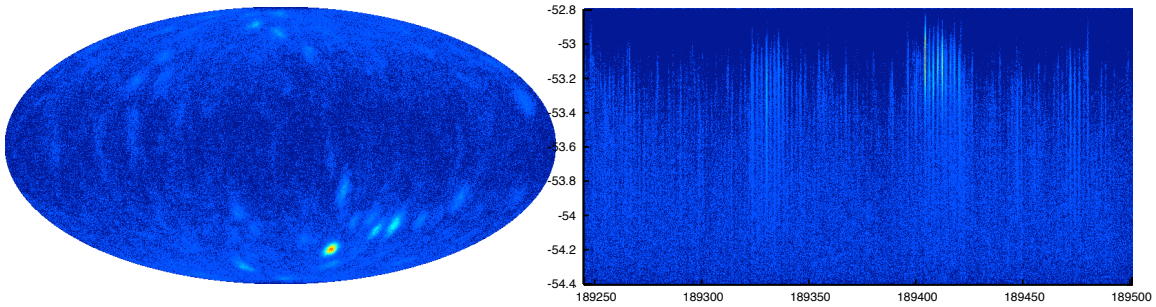


Figure 17: Marginalized PDFs of the $\theta - \phi$ (left) and the $\ln A - fT_{\rm obs}$ (right) plane for data containing only stationary-gaussian noise. Although this data contained no gravitational wave signal the PDFs show organized locations in parameter space that *look* like GW signals. The evaluation step of the analysis easily discarded these potential detections returning a Bayes factor of ~ 0.7. The $\theta - \phi$ PDF is shown in a Molweide projection on the sky.

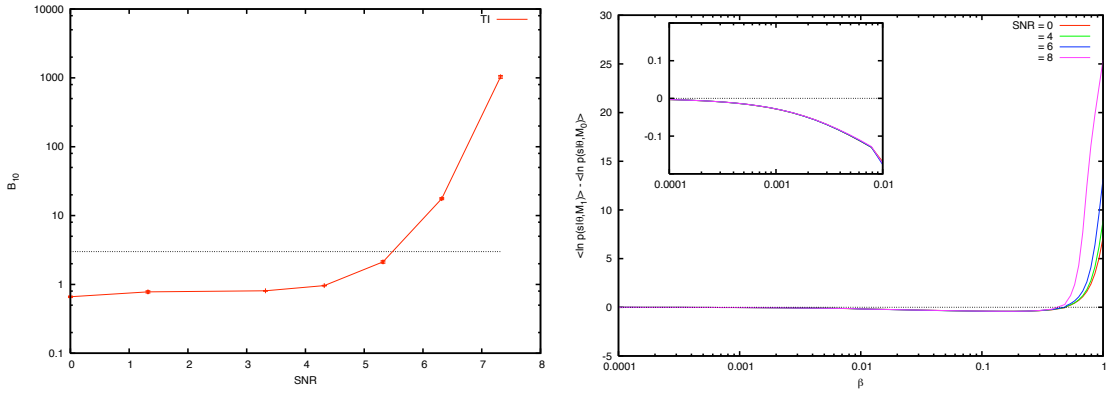
Evaluation

The model evidence for each SNR and each hypothesis ($\mathcal{M}_0$ and $\mathcal{M}_1$) is calculated using the 20 chains from the fixed dimension exploration with the integral over β being evaluated from 10^{-2} to 1. Figure 18a shows, for a single source and noise realization, the Bayes factor as a function of increasing SNR. The source becomes "detectable" ($B_{10} \geq 3$) in the vicinity of the fiducial signal-to-noise ratio of five

threshold. Figure 18b shows the integrand from eq. 90 for injected signals of SNR = $[0, 8]$ [1].

From this figure we can learn much about thermodynamic integration, including how a maximum temperature of 100 is clearly insufficient to capture all of the weight in the integrand. Running chains out to a more appropriate temperature ($T \sim 1 \times 10^4$) with sufficiently dense spacing is inconvenient. This problem can be circumvented, however, because at sufficiently high temperature the expectation value of the likelihood becomes independent of the injected signal. This manifests itself in the Figure 18b as the different curves become identical at lower β. This is easily understood if one considers that the $1/T$ term in the likelihood decreases the effective SNR of the injected source by $\sim 1/\sqrt{T}$. Because our analysis did not extend beyond sources with SNR = 10, a maximum temperature of 100 ($\beta = 0.01$) sufficiently hides any contribution to the data by the gravitational wave. Therefore we can perform the evidence integration up to $\beta = 0.01$ once for each noise realization and only need to do the case by case study in the low temperature regime. To take advantage of this the analysis was repeated once for each noise realization with 20 chains spaced geometrically with $[T_{\min}, T_{\max}] = [100, 10000]$. The evidence ratio from this high temperature analysis is a common multiple which needs to be applied to the Bayes factors found for each SNR of that noise realization. The inset in Figure 18b shows that the integrand has very nearly gone to zero at the maximum temperature of our analysis, signaling convergence of the integral.

[1] The signals were injected with amplitudes set using the definition SNR $\equiv \sqrt{(h|h)}$ which differs from Eq. 100. It is expected that the SNRs when calculated via Eq. 100 and $\sqrt{(h|h)}$ will differ (for data comprised of a finite number of frequency bins). The results of the Bayes factor vs. SNR studies use Eq. 100 to ensure that the interpretation of the SNR is the same between noise realizations.



(a) Bayes factor

(b) Evidence ratio integrand

Figure 18: Thermodynamic integration results for B_{10} on data with signals injected at increasingly higher signal-to-noise ratio. a) Error bars on the Bayes factor calculation were established by starting the chains with different random number generator seeds. The horizontal line is the Bayes factor where one would consider the result a positive detection. b) The horizontal line marks the regimes where the integrand supports $\mathcal{M}_1$ (above) or $\mathcal{M}_0$ (below). The inset shows that the integrand has sufficiently vanished at the maximum temperature analyzed.

Cross Validation

To verify the Bayes factor calculation from the thermodynamic integration the post burn-in samples from the $T = 1$ chain are sorted into the signal-model hypercube that will be used for the trans-dimensional proposal distribution in the RJMCMC. We experimented with a variety of different weightings for the sampled and uniform portions of the proposal, and found little variation in the resultant Bayes factors. Because the result is supposed to be independent of the proposal distribution this serves as a cheap test to ensure stability and convergence. For production-level runs the relative weighting between the two contributions to the proposal distribution was 0.5. The first 10^5 RJ iterations were burn in samples, after which the samples were assumed to be drawn from the model posterior. Figure 19 shows the achieved

agreement between the TI and RJMCMC calculations of the Bayes factor as a function of increasing SNR.

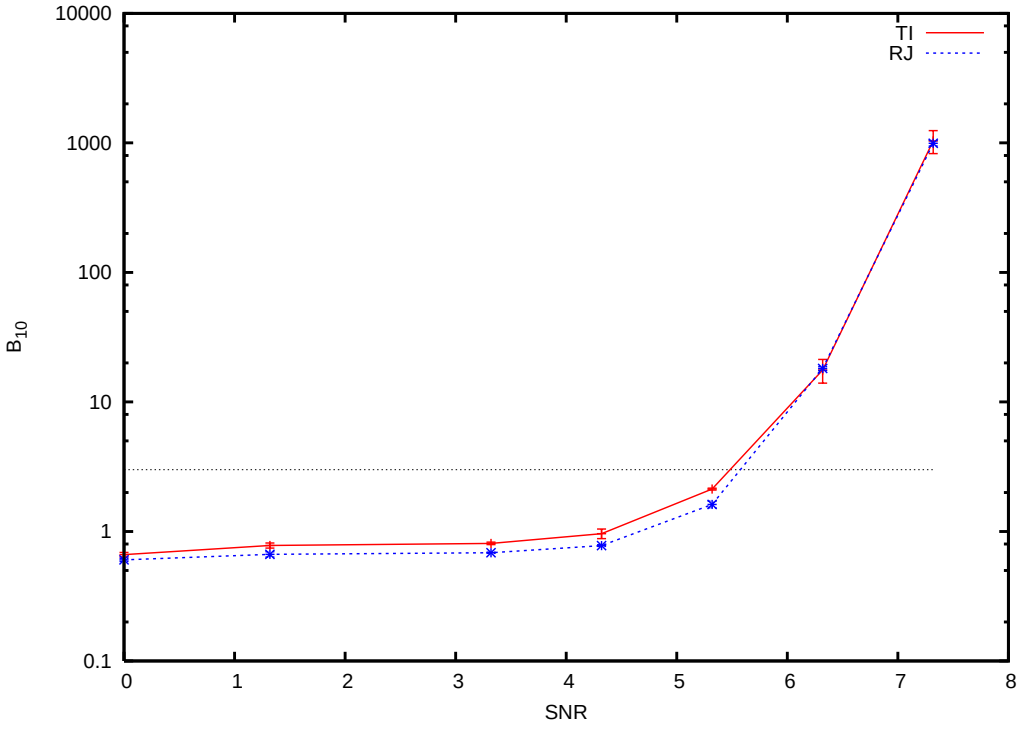


Figure 19: Thermodynamic integration and RJMCMC results for B_{10} on data with signals injected at increasingly higher signal-to-noise ratio. The horizontal line is the Bayes factor where one would consider the result a positive detection.

Dependence on Priors

As discussed previously the analysis was repeated adopting a lower bound on the amplitude prior such that the minimum allowed template SNR was ~ 5. Although the SNR at which the signal became detectable was unchanged, the restrictive prior volume did allow the analysis to definitively prefer the noise model over the signal

model when the injected signal was well below this threshold. This is in opposition to the examples which allowed an arbitrarily small amplitude, which caused the two models to be indistinguishable at very low SNR despite the additional degrees of freedom of the signal model. Figure 20 shows the Bayes factor as a function of SNR for the two cases with the red (solid) line depicting the results with the uninformative prior while the blue (dashed) line resulted from the constrained amplitude prior case. What should be taken from this demonstration is how the results of Bayesian methods depend upon what is being asked of the data. The two questions asked here, "Does the data contain *any* gravitational waves?" and "Does the data contain gravitational waves *above a certain amplitude*?" have very different interpretations and thus result in very different responses from the analysis. This is a feature (rather than a flaw) of Bayesian methods that underlines the role and importance of ones *a priori* beliefs about the experiment.

Discussion

We have implemeted and tested an end-to-end Bayesian analysis algorithm which proved successful at solving the detection problem for a single galactic bininary signal in simulated LISA data. This involves locating the regions in parameter space of high posterior weight (search phase), thoroughly sampling the distribution about these locations (characterization phase), and quantifying the confidence of the detection (evaluation phase). All three phases of the detection problem have been successfully completed using a single routine which resolves the PDF and calculates the evidence for each of the models under consideration. The construction of such a routine is relatively simple, and this simplicity results in a reliable "one stop" solution to the detection problem.

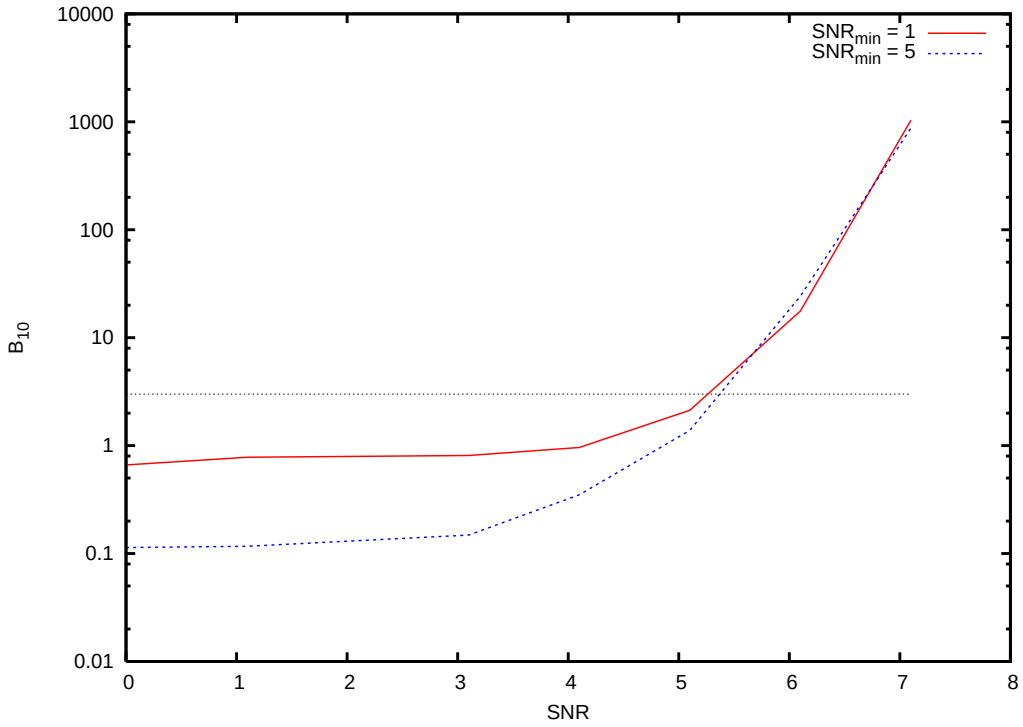


Figure 20: Thermodynamic integration results for B_{10} on data with signals injected at increasingly higher signal-to-noise ratio. The horizontal line is the Bayes factor where one would consider the result a positive detection. The red (solid) line is for the analysis done with uniform priors in the signal parameters (apart form $\dot{f}$). The blue (dashed) line is the same data analyzed with a restrictive prior on amplitude such that the minimum Amplitude corresponds to a SNR of ~ 5.

The search and characterization phases were simultaneously accomplished through a parallel tempering MCMC analysis which located the regions of interest within a few hundred thousand iterations. The efficient exploration of the full posterior was clearly demonstrated as the results revealed a surprising amount of additional structure in the signal PDFs. This could conspire to "fool" a less sophisticated analysis by preferentially sampling from some secondary maxima, thus failing to properly characterize the source. This could lead to a poor estimation of astrophysical parameters, or worse, a missed or false detection.

Our algorithm was *not* fooled thanks to the model selection analysis provided by the thermodynamic integration across the PTMCMC chains. The performance of the model selection phase of the problem was verified with a RJMCMC analysis. It is our belief that thermodynamic integration has proven to be the most reliable technique, and is exceptionally appealing when we consider extending this algorithm to a more general GW detection problem. The implementation of the thermodynamic integration calculation is independent of the model dimension and should remain as robust for more complicated scenarios.

Although we also had success with the RJMCMC algorithm, we caution that the RJMCMC mixing and convergence time was very sensitive to the construction of the algorithm, and is prone to "fooling" the user by appearing to converge to a stationary distribution that is *not* the target distribution. The diffuculty, a well documented one in the employment of the RJMCMC, is constructing an efficient trans-dimensional proposal distribution. Although the proposal is not supposed to affect the end result of the analysis, we found that sub-optimally tuned proposals led to very poor inter-dimensional mixing which in turn leads to very slow convergence: Well beyond the 10^6 iterations used to generate the displayed results. Once we adopted the PTMCMC-PDF proposal it was discovered through repeated testing that the chains were stable

and rapid in their convergence and in good agreement with thermodynamic integration. Figure 19 demonstrates the RJMCMC's robustness against different starting seeds under this efficient trans-dimensional proposal scheme.

For further verification in the model selection phase of the problem we (unsuccessfully) experimented with constructing a Nested Sampling [44, 36] evidence calculator. Nested Sampling, originally conceived by Skilling, is receiving attention in the astrophysics and cosmology community (e.g., [12, 70]) and more recently in gravitational wave detection problems similar to our own [63, 64]. Nested Sampling is an alternative to the MCMC-driven methods discussed previously. The algorithm stochastically explores the posterior by first drawing from the prior distribution with some fixed number of samples. The "worst-fit" sample is updated to higher likelihood by iteratively reducing the prior volume available for exploration via a minimum likelihood constraint. As this constraint approaches the maximum likelihood the samples are tightly clustered in the regions(s) of high posterior weight. The evaluation of the evidence reduces eq. 46 in chapter 3 to a one-dimensional integral over the fractional prior volume contained within the iteratively more restrictive likelihood constraint. A more detailed description of the step-by-step workings of the algorithm can be found in Ref. [44].

Despite a great deal of effort we were unable to tweak the algorithm to give us reliable results. It is our belief that the high dimensionality and multi-modality of the problem makes it exceptionally difficult to successfully create a healthy Nested Sampling evidence calculator. Although Skilling demonstrates Nested Sampling's robustness against local extremum we found in earlier implementations (when the entire prior volume was initially sampled) that the global likelihood maximum would often be missed without excessive numbers of initial samples and iterations.

The main obstacle was evaluating the fraction of the available prior volume. Although making a single such calculation is not prohibitive, the fact that this needs to be performed repeatedly over many iterations to achieve some reasonable convergence made the full implementation unfeasible, while all of our efforts at estimating the volume proved to be unstable.

One modification over Skilling's original design that we had attempted in our own implementation of Nested Sampling is to return to the PTMCMC chains and begin our exploration with a single sample placed in each occupied cell of the signal-model hypercube (see chapter 5). This approach hinges on the valid assumption that unoccupied cells do not contribute substantially to the posterior and therefore can be neglected from the evidence integration. This reduces the amount of time the algorithm spends integrating regions in parameter space which fail to measurably contribute to the overall evidence, while also ensuring that the global maximum will be sampled.

Despite this "short-cut" the results were still unsatisfactory. Undoubtedly a Nested Sampling routine *could* work if properly constructed, but in our experience the size of the search space coupled with the multi-modality of the target distribution make this a very difficult prospect and it is hard to imagine that Nested Sampling could produce results as rapidly as TI and RJMCMC.

When comparing the two successfully demonstrated model selection techniques (TI and RJMCMC) we have found the TI to be the most straight-forward in its implementation. The calculation is certainly time consuming, as many chains operating up to very high temperature are necessary to see convergence of the integration over β. The trade off for this cost is a large degree of confidence in the result as convergence tests are simple and decisive. Careful construction and ample testing is necessary for one to have such assurance in the output of the RJMCMC at this

level of dimensionality, although once achieved, the RJMCMC is significantly less computationally expensive to operate. It is therefore our conclusion that the thermodynamic integration should be the Bayes factor calculator of choice, while a carefully designed and tested RJMCMC algorithm serves as an excellent *additional* means of calculating the odds ratio between the two models. We would be uncomfortable with a RJMCMC-only approach to high-dimensional model evaluation.

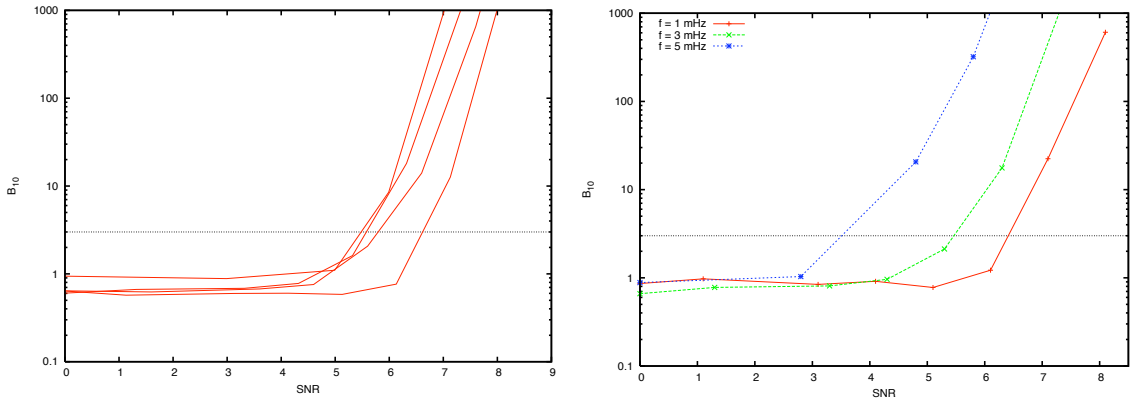
By comparing the SNR at which a signal becomes detectable between different noise realizations the weakness of using fixed SNR detection thresholds becomes apparent. Figure 21a shows the Bayes factor as a function of injected SNR for four identical sources injected into three different noise realizations. The point at which a signal becomes detectable varies appreciably between noise realizations even for benign, stationary and Gaussian noise. Even under ideal circumstances the detectability of a source is dependent on the particular noise realization with which the source is competing, as opposed to a hypothetical ensemble of possible noise realizations.

In this study we have ignored many real world complications. For example, a 256-bin segment of LISA data will likely contain several galactic binaries with varying degrees of overlap. To properly address the LISA-specific detection problem we would need to allow for an arbitrary number of sources and our algorithm would be responsible for sampling the model posterior $p(\mathcal{M}_i, s)$ to determine the most likely number of sources, instead of the binary 0- or 1-source case. An example of this type of study made with sinusoidal waveforms in simulated LISA data can be found in Ref. [30]. The RJMCMC approach nicely generalizes to the "N or $N+1$ sources" detection problem. One problem that has been faced while doing global galaxy fitting is the occasional tendency for the algorithm to fit a single galactic binary with two highly-overlapping waveforms. There are specific trans-dimensional moves, split/merge proposals [30], which are designed to resolve problems with overlapping signals that can be applied

the RJMCMC algorithms. The model evidence, with its built in Occam factor, will favor solutions where a single waveform template is used to model each signal.

Our single-source toy problem is more akin to the ground-based application of model selection where events are expected to be rare. However, our toy problem is unrealistic because our treatment of the noise does not allow for non-stationarity or departures from Gaussianity. Noteworthy work has been done to understand how to properly model realistic noise in a Bayesian framework [71] and application of such techniques in the ground-based problem is necessary.

In future work we plan to extend our model to include multiple, overlapping signals - in the context of LISA observations - and non-stationary and non-Gaussian instrument noise - in the context of LIGO observations.



(a) (b)

Figure 21: B_{10} for data with signals injected at increasingly higher signal-to-noise ratio via Thermodynamic Integration. The horizontal line is the Bayes factor where one would consider the result a positive detection. a) Different curves represent different noise realizations but identical signal parameters. This demonstrates that the detection "threshold" is sensitive to the particulars of the noise in the detector, even when the noise characteristics are identical (stationary, gaussian, with known spectral density). b) Different curves represent different signal parameters but identical noise realizations. The red, green, and blue curves are for injected source parameters 1, 2, and 3 respectively (see Table 6). This demonstrates that the detection "threshold" is sensitive to the details of the waveforms.

CASE II: INSPIRALING MASSIVE BLACK HOLE BINARIES – LIGO

For ground based GW detectors the detection problem is of utmost importance because strong GW signals are rare, while large-amplitude transient noise events in the detectors are common. One must then be able to distinguish between these noise glitches and GWs. The majority of glitches can be excluded by looking at auxiliary instrument monitoring channels (data quality tests and vetoes in LIGO parlance), and by applying coincidence tests among the world-wide network of detectors. Remaining events must then be analyzed in some model-selection sense.

To date, this analysis is almost exclusively done using frequentist methods, where a "false positive" level is established by analyzing time-slides of the data and detection efficiencies are established using signal injections. We feel that this problem is best addressed in a Bayesian framework but the Bayesian approach has been little used due to the difficulty of accurately modeling the instrument noise. Our aim is to generalize the PTMCMC detection algorithm so that it can perform robustly in the presence of non-gaussian, non-stationary noise.

The first step, however, is to enhance the performance of the existing algorithm in the presence of "ideal" noise by increasing the efficiency of the search phase. While the search method for galactic binaries eventually recovers the global MAP values for the candidate signals, our "brute force" method for finding them (a large number of chains, for many iterations) becomes exceedingly costly as the complexity of the signal, and the difficulty in finding it, increases.

To this end, the bulk of the following chapter will be dedicated to improving the algorithm we applied to the galactic binary search in LISA data. In this case we will be analyzing the performance of the global network of ground based detectors, simu-

lating data from LHO, LLO, and Virgo. We will develop an improved version of the PTMCMC detection algorithm to establish when the evidence for a model claiming the data contains a signal from an inspiraling black hole becomes distinguishable from a model describing the data as exclusively containing instrument noise. While our focus in this test case will be on the search phase of the analysis, the continued development of the algorithm will lay the foundation for a more sophisticated treatment of the noise modeling.

During this chapter we will describe the changes that have been made to the detection algorithm that have increased the efficiency and made it less "taylored" for galactic binary searches. We first outline the signal and noise models for the ground-based network of interferometers containing the signal from an inspiraling binary black hole system. We devise a unique noise model which uses a discrete wavelet transform of the data, and assigns a noise level to blocks of wavelet pixels, thus retaining spectral and temporal resolution in order to account for noise glitches in the instruments. The gravitational waveform that we use to test our algorithm is that of a pair of black holes which are non-spinning and have circular orbits. We then discuss the changes that have been made to the PTMCMC detection algorithm, specifically in the search-phase where we have adopted a scheme which maximizes the likelihood over several parameters. We finally demonstrate the success of the algorithm in efficiently finding the injected signals, characterizing the PDFs, and computing the Bayes factor for increasing signal-to-noise ratio. Similar to the galactic binary study, we again see significant variability in the level where the signals become "detectable".

To test the PTMCMC detection algorithm with enhanced search capabilities we generate a 16 second segment of simulated LIGO-Virgo data containing instrument noise and the signal from a single black hole binary inspiral. We first generate a noise realization for the frequency range of interest by drawing from the theoretical

LIGO/Virgo noise spectral density (see chapter 6). We then inject a test source into the data with tunable distance (inversely proportional to amplitude) so that we may establish at what SNR the source becomes detectable. To be considered effective we maintain the same standards as in chapter 6: Specifically, our algorithm must find the MAP parameters for two distinct models

$$\begin{aligned} \mathcal{M}_0 : s_{ij} &= n_{ij} \\ \mathcal{M}_1 : s_{ij} &= n_{ij} + h_{ij}, \end{aligned} \tag{101}$$

where the signals are now written in the wavelet domain for reasons that will be discussed in the following section. The PTMCMC detection algorithm must then thoroughly resolve each model's PDF, and calculate the evidence for each model (up to some constant factor common between $\mathcal{M}_1$ and $\mathcal{M}_0$). We will continue to consider the prior odds between the two models to be unity, thus allowing us to interpret the evidence ratio (Bayes factor) as the odds ratio. When using a gaussian noise model this is, in fact, a poor choice of model prior for initial LIGO data as the predicted event rates for GW signals are very low, while instrument glitches are common. A real-world implementation of this approach would undoubtedly require a more stringent prior ratio for the Bayes factor to overcome before claiming a detection or a much more realistic noise model.

Modeling the Instrument Noise

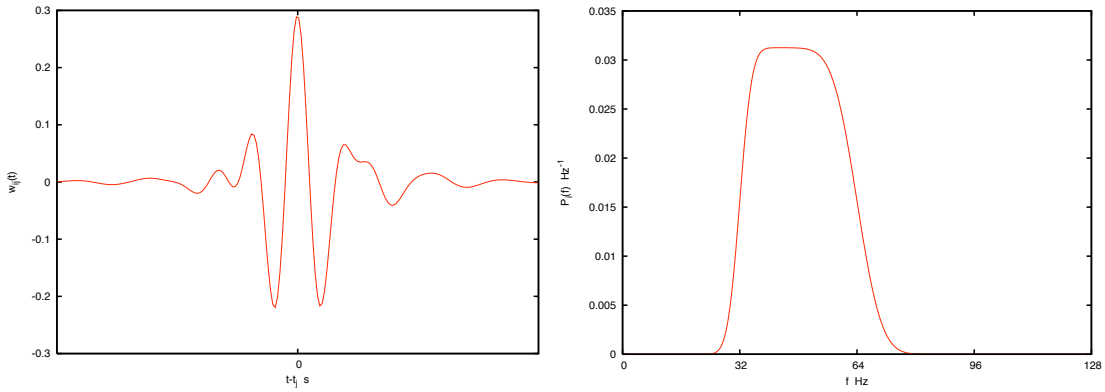
Adequately modeling the LIGO/Virgo noise is currently the main obstacle to seeing Bayesian methods take hold in the ground-based analysis community. Existing techniques assume stationary and gaussian noise and modify the model selection thresholds to be robust against glitches [63, 64]. While this can be done, it comes at

the cost of decreased sensitivity to actual GW signals [64]. Our desire is to ultimately improve upon this approach by adopting a noise model that is more faithful to the realities of LIGO data. For most LIGO sources the issue of slowly-varying, non-stationary noise is one that can be neglected as the noise is typically stable over the duration of time the source is in band. The obvious exceptions to this are the long-lived signals from isolated pulsars which we do not consider here.

The problematic features in the data are the so-called glitches – large amplitude transient events – which, on occasion, can produce coincident triggers between detectors. The ability for our noise modeling to account for these events is a vital component to maximizing the effectiveness of the model selection approach. To do so we must abandon the familiar fourier domain approach to signal processing as we need both temporal and spectral resolution. Preserving frequency-domain content is still important, as the noise is colored, and the signals have distinct spectral features. To this end, time-frequency methods have been used in a variety of GW search algorithms both for LIGO/Virgo data as well as simulated LISA data. The foundation for our time-frequency approach is to decompose the data into a wavelet basis to perform the analysis. The most prevalent use of wavelets in GW data analysis is in the Coherent Wave Burst algorithm [72] which serves as the flagship search pipeline for the LIGO Burst Group analysis.

Individual wavelet basis functions ψ_{ij} are compact in both frequency and time, and can be used to form an orthogonal basis [73]. Each element of the basis is a scaled, time-shifted version of the "mother wavelet" ψ of which a large variety exist. We have chosen the Meyer wavlet to form our basis functions which is depicted both in the time- and fourier-domain in Fig. 22. The Fourier power of a single wavelet function is perhaps most illuminating. Each wavelet basis function acts as a high-

and low-pass filter, selecting for a specific range of frequencies, during a specific time interval of the signal.



(a) (b)

Figure 22: The Meyer wavelet basis function for frequency layer $i = 9$ and signal duration of 16 seconds. a) $\psi_{ij}(t)$ for arbitrary time index j. b) Notice how, for $i = 9$ and $T = 16$ s, $|\tilde{\psi}_{ij}(f)|^2$ will filter the data for frequencies $f \sim [32, 64]$ Hz.

Similar to the more familiar spectral analysis, we use a discrete wavelet transform (DWT) of a time series $s(t)$ to determine the wavelet coefficients w_{ij}. Each w_{ij} represents a pixel in the time-frequency plane with volume $\Delta t \Delta f = 1$. We can create "scalograms" of the data, similar to fourier spectrograms, to graphically portray which pixels in the time frequency plane contain the most power. Unique to the DWT is the way the pixels "tile" the time-frequency plane: Low frequency wavelets have long durations with high frequency resolution. As we move to higher frequency the bandwidth of the wavelet increases while the time resolution becomes more compact (maintaining $\Delta f \Delta t = 1$). A cartoon representation of this time-frequency tiling can be found in Fig. 23. Each frequency layer, denoted by the index i contains 2^i divisions in time and, for signal duration of T, is $2^i/T$ wide in frequency.

By decomposing the data into this basis we can use as our noise-model parameters the expected noise level in blocks of wavelet pixels. This allows the model to elevate

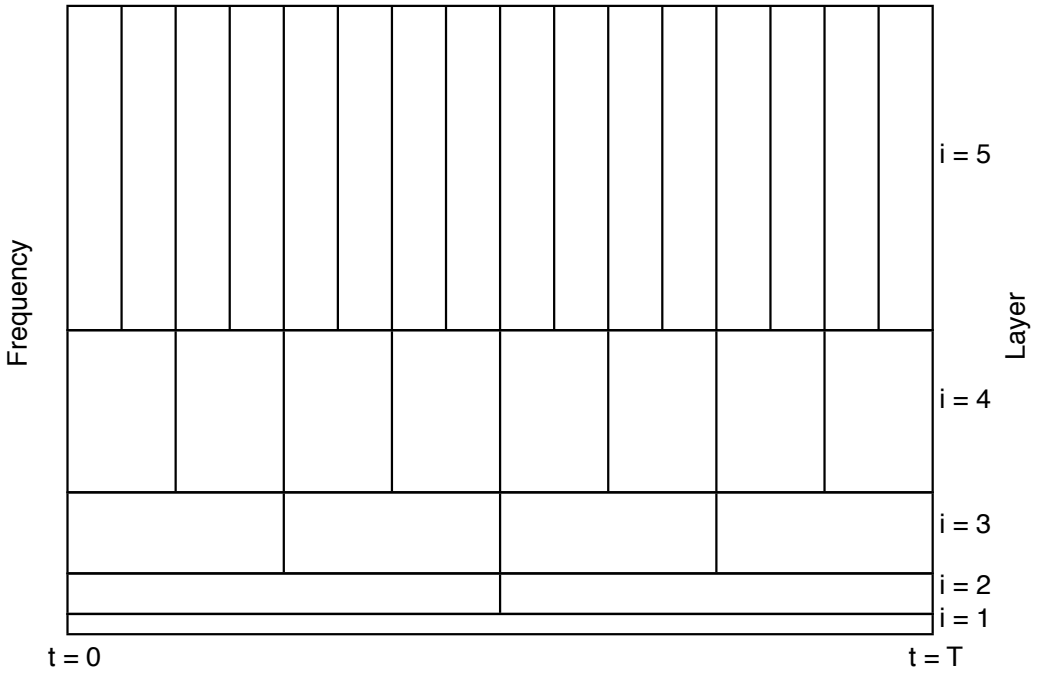


Figure 23: A cartoon depicting the tiling of the time-frequency plane by a discrete wavelet transform (DWT). Each pixel has time-frequency volume $\Delta t \Delta f = 1$.

the noise during a small duration of the signal to account for transient excess power, which may be the result of some instrumental glitch, without enforcing similar excess power on the remainder of the data. The choice of tiling is obviously a tunable feature in the algorithm. We have elected to group wavelet pixels across a layer such that the noise level contains 128 such samples. This is a sufficiently large number of noise samples per block to approximate the expected distribution of the noise as a gaussian.

While this approach to noise modeling gives us the flexibility to adjust to glitches, ultimately we wish to construct a model which accurately *anticipates* such events. One such approach would be to have a variety of glitch models, one for each of the

imaginable noise transient, and to analyze each candidate detection with a signal model as well as each glitch model. This would be extremely inefficient and too narrowly focused on the characteristics of the modeled glitches to account for the unanticipated noise excursion.

Instead we aim to adopt the approach of [74] where the likelihood is treated as a weighted sum of two normal distributions. The majority of the weight would be alloted to the distribution with variance S_n so that the vast majority of noise samples would be drawn from the standard picture of the instrument noise. However, some non-zero weight would come from a significantly wider distribution. This effectively broadens the tails of the noise distribution such that typical ($< 3\sigma$) events reflect the theoretical stationary, gaussian noise probability distribution while allowing for more frequent large sigma events.

This is, to a good approximation, how the actual instrument noise behaves, as after the data is conditioned it is nearly stationary and gaussian, but with occasional large glitches. Implementing such a change will merely require a new likelihood function and we soon anticipate this feature of the algorithm to be implemented, tested, and ready for S5 data. Using a superposition of two normal distributions is a standard tool in Bayesian modeling of non-gaussian data [36], and has seen some attention in modeling the daily behavior of the stock market but, despite having been suggested in [74], has not yet been *implemented* in gravitational wave signal processing.

Modeling the Source: Black Hole Binaries

We take as our test signal the inspiral of a binary system composed of two stellar-mass black holes. The full gravitational waveform from the inspiral of binary black holes is composed of an inspiral, merger, and ringdown phase. Analytic solutions exist

for the inspiral and ringdown waveforms courtesy of the Post-Newtonian expansion and Black Hole Perturbation Theory, respectively. Recent advances in Numerical Relativity have led to the production of merger waveforms allowing, for specific initial conditions, the construction of complete inspiral-merger-ringdown (IMR) templates. For this work we will only simulate the inspiral phase of the waveform, both for our injections as well as our analysis, while we eagerly anticipating further developments which will hopefully yield generic IMR waveforms.

While neglecting the merger- and ringdown-phase of the waveforms, we apply additional simplifications by neglecting spin effects, as well as any orbital eccentricity. What remains is a model of the waveforms which is characterized by nine quantities which are written as components of the signal-model parameter vector:

$$\vec{\lambda} \rightarrow (m, \mathcal{M}, t_c, \log D_L, \sin\delta, \alpha, \cos\theta_L, \phi_L, \varphi_c) \tag{102}$$

where m and $\mathcal{M}$ are the total and chirp mass of the binary system, t_c is the time until the binary coalesces, D_L is the luminosity distance to the system, δ and α are the declination and right ascension as defined on the celestial sphere, θ_L and ϕ_L are the polar coordinates of the angular momentum vector, and φ_c is the GW phase at coalescence. Although it is reasonable to expect that the orbits will have circularized [75] it is possible that the spin of the individual black holes will be important.

Including the spin effects introduces additional modulations in the waveform and the performance of LIGO [49] as well as LISA using spinning templates has been extensively studied [76]. While we recognize the value of including these effects, the additional parameters and the complexity of generating the waveforms unnecessarily increases the computational cost of our test-of-principle. We can not foresee any reason why the approach to the detection problem taken here will not easily adapt

to the spinning case and thus willingly opt for the more time-efficient modeling. We will further simplify the waveform by neglecting all but the strongest harmonic of the wave, that which corresponds to the quadrupole moment of the source. Studies of the performance of gravitational wave detectors when higher harmonics are included in the modeling can be found in [77].

We will describe the modeling of the signal similar in manner and notation to [48]. The gravitational wave in the post-Newtonian (PN) approximation are given as

$$\begin{aligned} h_+ &= \frac{2m\eta}{D_L}(1+\cos^2\iota)x\cos(\Phi), \\ h_\times &= -\frac{4m\eta}{D_L}\cos\iota x\sin(\Phi), \end{aligned} \tag{103}$$

where $\eta = m_1 m_2/m^2$ is the reduced mass of the binary and $x = (m\omega)^{2/3}$ is the post-Newtonian velocity parameter. The orbital frequency of the binary, out to 2 PN order is [78]

$$\begin{aligned} \omega(t) &= (8m)^{-1}\left[\Theta^{-3/8} + \left(\frac{743}{2688}+\frac{11}{32}\eta\right)\Theta^{-5/8} - \frac{3\pi}{10}\Theta^{-3/4}\right. \\ &+ \left.\left(\frac{1855099}{14450688}+\frac{56975}{258048}\eta+\frac{371}{2048}\eta^2\right)\Theta^{-7/8}\right] \end{aligned} \tag{104}$$

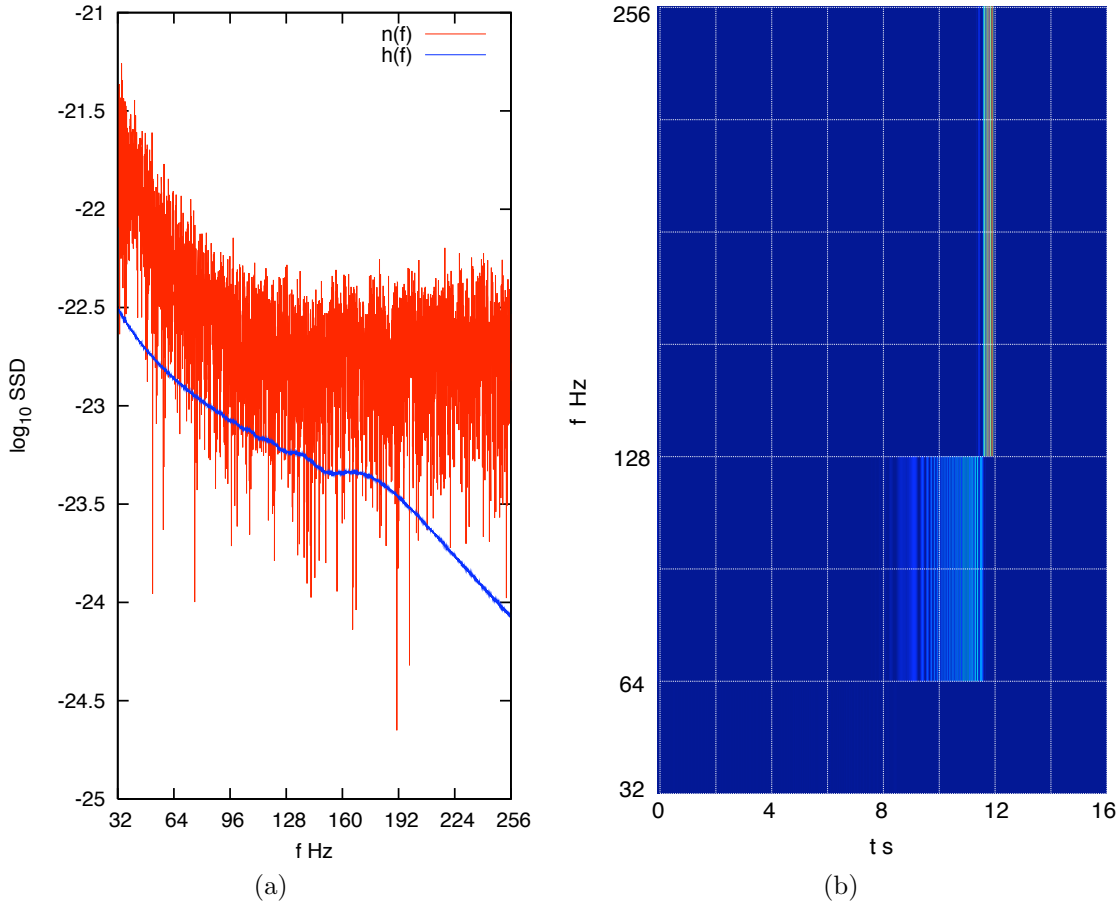
The gravitational wave phase Φ in Eq. 103 is defined as twice the orbital phase of the binary which, again to 2 PN order is defined as

$$\begin{aligned} \Phi(t) &= \varphi_c - \frac{2}{\eta}\left[\Theta^{5/8} + \left(\frac{3715}{8064}+\frac{55}{96}\eta\right)\Theta^{3/8} - \frac{3\pi}{4}\Theta^{1/4}\right. \\ &+ \left.\left(\frac{9275495}{14450688}+\frac{284875}{258048}\eta+\frac{1855}{2048}\eta^2\right)\Theta^{1/8}\right]. \end{aligned} \tag{105}$$

where $\Theta(t,t_c) = \frac{\eta}{5m}(t_c - t)$. Finally, we must gradually stop the simulation of the waveform prior to the PN approximation breaking down around $x_{\max} \sim 1/7$ so as to not corrupt the spectral content of the waveform by having an abrupt "edge" in the time-domain waveform. We therefore taper the waveforms with the function

$$\mathcal{F}(x, x_{\max}) = \frac{1}{2}\left[1 - \tanh(150(x - x_{\max})\right]. \tag{106}$$

A fourier- and wavelet-domain representation of a simulated black hole waveform can be found in Fig. 24. This signal was injected with SNR = 40 and $t_c = 12$ s. The peak wavelet power (as seen in the right panel) occurs in the high-frequency wavelet layer at the coalescence time.



(a) (b)

Figure 24: Black hole signal with SNR of 40 as seen by LHO. a) An example of frequency domain noise (red) and the injected waveform (blue). b) Wavelet domain scalogram of injected signal whitened by $S_n(f)$. For this source $t_c = 12$ seconds.

Details of the PTMCMC algorithm implementation

Enhanced Search Techniques

The earlier generation of the PTMCMC solution to the detection problem, described in detail in chapters 5 and 6, involved a brute-force approach to the search phase of the analysis. Specifically, using a large number of parallel chains (20) and a long burn-in phase (5×10^5 iterations). While this approach proved reliable it is admittedly inefficient and becomes impractical for the more challenging black hole binary inspiral signals. The prominent frequency evolution of chirping black hole binaries renders it impossible to take advantage of the fast-slow decomposition that we have relied on to rapidly generate white dwarf waveforms. This drastically slows down the waveform generation which, in turn, magnifies the inefficiency of using long burn-in periods during the search.

Finally, the key parameters to achieving an adequate match to the signal are the time to coalescence (t_c), the chirp mass ($\mathcal{M}$), the total mass (m), and the phase at coalescence (φ_c), each of which exhibits strong correlations with the other three key parameters *and* occupies a miniscule region of its respective allowed range. We estimate the volume of parameter space occupied by the signal by computing the volume contained in a $3\sigma_\Gamma$ hyper-cube extending in parameter directions about the injected signal parameters, where σ_Γ is determined by the Fisher information matrix. The volume of this hyper-cube is compared to the entire volume of parameter space (or prior volume). We find for SNRs ranging from 5 to 40 that the fractional volume is of order $\sim 10^-14$ to $\sim 10^{-18}$, illuminating what a small target in the search space these signals present.

An MCMC search of the parameter space would struggle to randomly pick values for these four parameters with enough accuracy to converge onto the MAP parameters. The difficulty is exacerbated by the complicated structure of secondary features

as depicted in Fig. 25. Analysis of these signals (when at marginal SNR) poses a formidable challenge and the search phase warrants more assistance than is applied the PTMCMC algorithm as described in chapter 5.

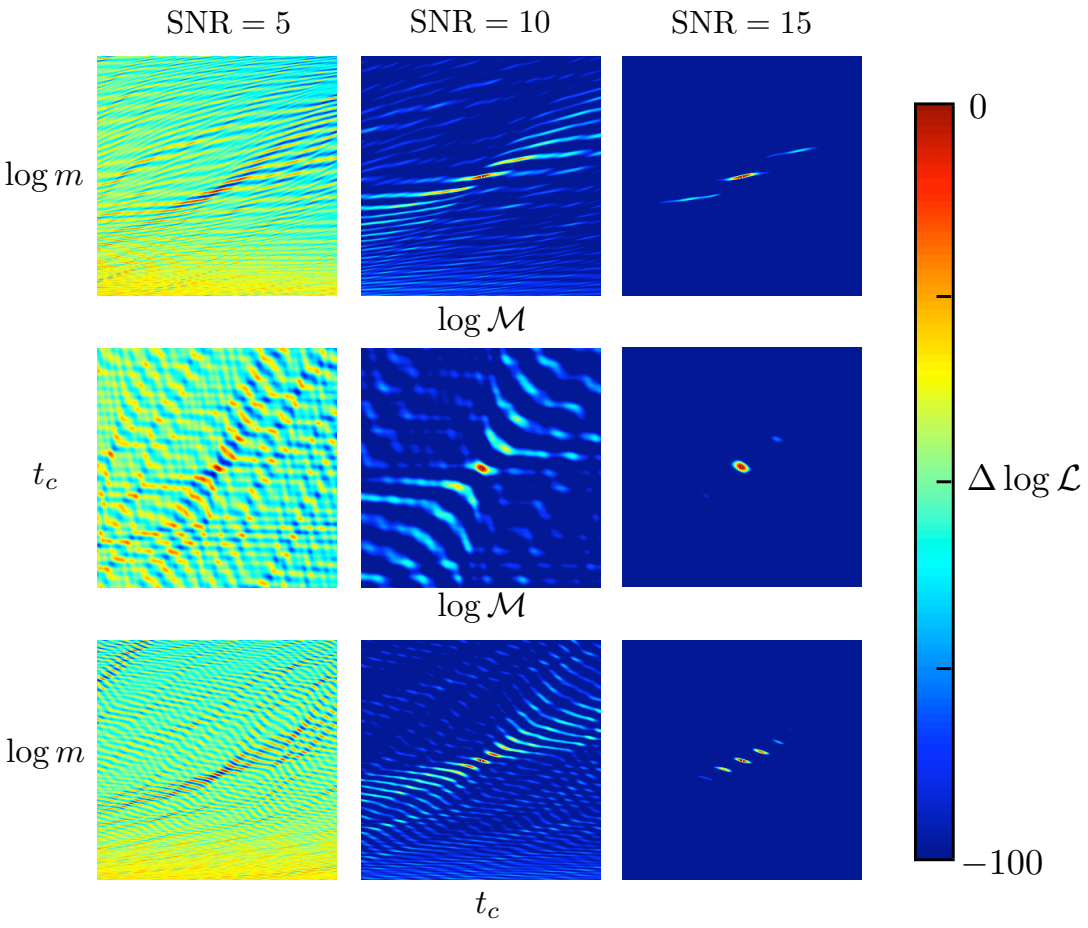


Figure 25: Slices of the log-likelihood surface for a binary black hole waveform with different SNR injected into LIGO data showing the degree of correlation for a selection of the signal parameters. $\Delta \log \mathcal{L}$ is the difference in log likelihood relative to that of the MAP signal parameters. The ranges are $\pm 10\sigma$ from the injected signal parameters as estimated using the Fisher information matrix.

Several methods exist for optimizing a search using Metropolis-Hastings algorithms as the driver for the exploration, many of which have been extensively studied in similar scenarios to our current example [10, 79, 80, 51]. The most effective are

those which break the Markovian nature of the chain (e.g. simulated annealing, F-statistic extremization), thus biasing the chain samples towards high $\ln p(s|\vec{\theta})$ and rendering the chain useless for model selection calculations or parameter estimation (beyond MAP finding). There do exist search enhancements which preserve the independence of the chain samples, such as using delayed rejection, but we have found the implementation of such methods unnecessarily complicated. Our approach is to use non-Markovian search techniques to locate the regions of high posterior weight and to use the biased PDF from this "illegal" search as a proposal distribution used to sample from the target distribution.

We have adopted an extremely simple, general, scheme which maximizes the likelihood over t_c, D_L, and φ_c. To do so we begin by generating a template $Ah(\vec{\lambda}')$ where $\vec{\lambda}'$ are *not* the model parameters over which the maximization is performed, and A is the amplitude of the gravitational waveform. We then define the phase-blind correlation $R(\tau)$ by:

$$\begin{aligned} \tilde{c}_1(f) &= \frac{\tilde{s}(f)\tilde{h}^*(f)}{S_n(f)} \\ \tilde{c}_2(f) &= \frac{\tilde{s}(f)\tilde{h}^*(f)e^{i\pi/2}}{S_n(f)} \end{aligned} \tag{107}$$

We do not, in practice, need to perform any additional operations to calculate $\tilde{c}_2(f)$ because $\tilde{h}^*(f)e^{i\pi/2}$ is identical to switching the real and imaginary parts of h, and changing the sign of the (now) real components.

We then inverse Fourier transform $\tilde{c}_1$ and $\tilde{c}_2$ into the "time lag" domain via

$$c_i(\tau) = 2\int_0^\infty \tilde{c}_i(f)e^{-2\pi i f\tau}df. \tag{108}$$

Then the correlation can be found from $c_1(\tau)$ and $c_2(\tau)$ (which for this example are strictly real) by

$$R(\tau) = \sqrt{c_1^2(\tau) + c_2^2(\tau)}. \tag{109}$$

With the correlation in hand we can determine the lag $\tau_{\max}$ which yields the maximum $R(\tau)$ and, by Parseval's theorem, $2R(\tau_{\max}) = (s|h)_{\max}$. We can now calculate the maximum likelihood for this template by differentiating the standard log-likelihood

$$\frac{1}{2}(s - Ah|s - Ah) \tag{110}$$

with respect to A and solving for the maximum via:

$$\begin{aligned} \partial_A(s - Ah|s - Ah) = (-h|s - h) &= 0 \\ -(h|s) + A(h|h) &= 0 \end{aligned} \tag{111}$$

then substituting $A = \frac{(s|h)}{(h|h)}$ back into Eq. 110 we arrive at the maximized likelihood:

$$p(s|\vec{\lambda})_{\max} = C\exp\left\{-\frac{1}{2}\left((s|s) - \frac{(s|h)^2_{\max}}{(h|h)}\right)\right\} \tag{112}$$

To determine the parameters which yield the maximum-likelihood template we must know the corresponding lag $\tau_{\max}$. The time to coalescence by which the template h must be shifted is then $t_c = t_c' + \Delta t$ where $\Delta t = \tau_{\max}$. The corresponding phase shift is calculated by

$$\Delta\varphi_c = \arctan\left(\frac{c_2(\tau_{\max})}{c_1(\tau_{\max})}\right) \tag{113}$$

and we rescale the luminosity distance by shifting $\ln D_L{}'$ by

$$\Delta \ln D_L = -\ln A \tag{114}$$

as the distance scales inversely to the amplitude.

To illustrate this process we will work through a toy example where the signal is noise-free and contains a gravitational wave $s(t) = A\cos(2\pi f_0 t + \phi)$. We will propose as a matched filter the trial template $h(t) = \cos(2\pi f_0 t)$. In this example we can apply either a time- or phase-shift, as for a continuous and monochromatic signal these two procedures are degenerate, and we elect to illustrate the former. Following the above

prescription using a discrete Fourier transform and assuming that f_0 is an integer multiple of $1/T$ with $n = f_0/T$ we have:

$$\begin{aligned} \tilde{s}_m &= \frac{A}{2}\sum_m \delta_{m,n}\left(e^{-i\phi} + e^{i\phi}\right) \\ \tilde{h}_m &= \frac{1}{2}\sum_m \delta_{m,n} \end{aligned} \tag{115}$$

and

$$\begin{aligned} \tilde{c}_{1,m} &= \frac{A}{2}\sum_m \delta_{m,n}\left(e^{-i\phi} + e^{i\phi}\right) \\ \tilde{c}_{2,m} &= i\frac{A}{2}\sum_m \delta_{m,n}\left(e^{-i\phi} + e^{i\phi}\right) \end{aligned} \tag{116}$$

then by inverse Fourier transforming we see

$$\begin{aligned} c_1(\tau) &= \frac{A}{2}\cos(2\pi f_0\tau + \phi) \\ c_2(\tau) &= \frac{A}{2}\sin(2\pi f_0\tau + \phi) \end{aligned} \tag{117}$$

Because we have chosen to solve for the phase shift between the signal and the template there is no maximization needed over τ. For this example the correlation $R(\tau) = A$ and the phase shift

$$\Delta\phi = \arctan\left(\frac{\sin(2\pi f_0\tau + \phi)}{\cos(2\pi f_0\tau + \phi)}\right) = \phi \tag{118}$$

as anticipated.

Because these maximization techniques must be done in the Fourier domain the search is most optimally performed in frequency space and therefore in the absence of noise fitting (which, for models robust against noise transients, is best performed in the wavelet domain). Consequently, the search-phase inner products use the initial LIGO and Virgo theoretical noise PSD (see Eq. 62 in chapter 3). The subsequent chain samples are biased by the maximization of the likelihood and thus can not be

used to make inferences about the posterior distribution function or evidence for the signal model. The benefit, however, of rapid convergence to the high-density regions of the posterior is ample compensation. These samples will even be used *to our advantage* in the strictly Markovian characterization phase of the algorithm.

Characterization and Evaluation

To characterize the signal model and evaluate the evidence we must now repeat the analysis with the full implementation of parallel tempering and noise fitting while preserving the Markovian nature of the chains. We can *greatly* expedite this phase of the analysis by using the (biased) samples from the search phase to drive MCMC jumps. Recall that the proposal distribution, by definition of the Hastings ratio, can not bias the chain's sampling of the target distribution. Even though the search-chain samples were obtained "illegally', they *can* be used to construct proposal distributions much in the way that the RJMCMC proposals were built out of the fixed dimension chains in chapter 5. As long as the N-dimensional histogram built out of the search-chain samples is normalized (and allows access to the entire prior volume) it is a perfectly valid proposal distribution from which we can draw new parameter combinations (taking care to include the $q(x|y)/q(x|y)$ term in the Hastings ratio). This proposal distribution, although biased and poorly sampled, is a sufficiently accurate approximation to the true PDF to *greatly* enhances the mixing of the chains during the characterization phase. The use of parallel tempering, coupled with the pilot exploration of the posterior, leads to rapid (less than 10^3 iterations) convergence and supplies further assurance that the chain is globally sampling the posterior.

As in the galactic binary study, we integrate $\langle \ln p(s|\vec{\theta}_i, \mathcal{M}_i)\rangle_\beta$ over β to calculate the evidence for $\mathcal{M}_i$ and from here acquire the Bayes factor, or odds (given our uninformative model prior ratio), for "detection". In this test case we do not

include a RJMCMC Bayes factor verification step. The increased dimensionality of this problem (one additional signal parameter, 64 additional noise parameters) could potentially warrant some finesse in tuning the trans-dimensional jumps. However, since we found the two methods to be in such excellent agreement in our previous example we have no hesitation towards relying on the thermodynamic integration approach and remain confident that, given adequate testing, such consistency between methods would continue in this application.

Priors

We use uninformative (flat) priors on all of the signal model parameters with angular variables taking the full allowed range. We allow for the individual constituents of the binary to have masses ranging from 1 $M_\odot$ to 30 $M_\odot$. The distance to the binary is constrained between 10^{-1} and 10^3 Mpc. Because the masses and the distance are scale parameters we use the log of these quantities in the chains, as well as for the prior ranges. The time to coalescence (t_c) is constrained to occur after the first second and before the end of the 16 second data segment under scrutiny. This should not be interpreted as a restricted prior due to some candidate detection. It is, in fact, poor practice to adjust the prior distributions based on the results of any analysis performed on the data under scrutiny and doing so leads to over-estimation of the confidence in a detection or the constraints on the model parameters. We instead envision performing this analysis on segments of data of a certain duration, avoiding times when the detectors were experiencing excess noise levels and have chosen $T_{\rm obs} = 16$ seconds somewhat arbitrarily.

We assume the noise in each detector to be normally distributed and, in the fourier domain, to have power spectral density $S_n(f)$. We want to construct our prior distributions such that the mean is centered at the expected noise level in each block

of wavelet pixels. The mean squared value of the noise in cell j of wavelet layer i (S_n^{ij}) can be calculated by filtering the noise power spectral density with $P_w^{ij}(f) = |\tilde{\psi}_{ij}(f)|^2$ via

$$S_n^{ij} = \int_0^\infty P_w^{ij}(f) S_n(f) df. \tag{119}$$

The power spectrum for any wavelet of the same layer, as well as the power spectral density for stationary noise, is independent of time. However, our inclusion of the time index j on all of the wavelet-related terms, while redundant, is done for clarity. We will again parameterize the noise with a multiplier η to S_n. We require one η_{ij}^k for each noise block (indices i and j) of each interferometer's data (index k).

Each block of wavelet pixels for which a noise parameter is assigned contains $N_{TFV} = 128$ wavelet pixels – each (in the absence of a GW) a single draw from some normal noise distribution. We therefore have N_{TFV} independent samples of a normal distribution in each noise block and should therefore expect the noise level in each block to be χ^2-distributed with N_{TFV} degrees of freedom – very well approximated by a normal distribution with variance $\sigma_n^2 = 1/N_{TFV}$. Thus our noise prior for the k^{th} detector and frequency layer i is

$$p(s|\eta_{ij}^k) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(\eta_{ij}^k - 1)^2}{2\sigma_n^2}}. \tag{120}$$

Because of the large time-frequency volume of each cell this turns out to be a substantially restrictive prior on the noise level, which is obviously what we wish to avoid if the noise were to contain glitches. We will specifically study the performance of the noise model when it is analyzing data with significant power due to gravitational waves to see how these priors effect it's ability to accommodate that excess power.

Results

Search

Two test sources with nearly identical parameters were used to study the performance of the enhanced PTMCMC detection algorithm in this new scenario. The two sources differed in their sky locations and the time to coalescence. Each binary system had a total mass of 15 $M_\odot$ and we adjust the luminosity distance for each injection to create signals with different SNR. The complete set of static signal parameters can be seen in Table 6.

Source	$m_1/M_\odot$	$m_2/M_\odot$	t_c/s	$\sin\delta$	α/rad	$\cos\theta_L$	ϕ_L/rad	φ_0/rad
1	10.0	5.0	6.0	0.05	4.2	-0.5	2.6	2.4
2	10.0	5.0	12.0	0.8	0.2	-0.5	2.6	2.4

Table 6: Injected black hole binary parameters.

The chains were initialized randomly and managed to begin sampling from the vicinity of the injected parameters within, at worst, a few hundred iterations making this a very successful search technique. The parallel tempering scheme involved 5 chains with a maximum temperature of 25, each spaced geometrically. Figure 30 shows, for an SNR of 20 version of Source 3, how the chain rapidly collapses to the injected value of 12 s.

After 10^4 iterations we declared the burn-in complete and the subsequent samples were stored and later used to populate the proposal distribution histogram for the characterization phase. The distribution of the post burn in samples is worth discussion, particularly for the sky parameters. Figure 27 shows the marginalized pseudo-PDFs for the sky location parameters δ and α. The location of each interferometer is projected onto the sky and labeled by an identifying letter (e.g. "L" for the Livingston

observatory, etc.). The white circles represent the rings of degenerate (to the injected black hole waveform) time delays for each pair of detectors. If the sky location of a signal is solely determined by the time delay between a pair of interferometers it is confined to exist along the corresponding ring on the sky. The intersection points of the three rings mark the degenerate locations for the LIGO/Virgo network and are highlighted in the figure by white squares - the double square indicating the injected sky location.

Notice how the chains do not sample both sky locations. This is a result of the phase information encoded in the waveform breaking the sky location degeneracy allowing the network to systematically approve of the appropriate region on the sky. Furthermore, we were encouraged to see the chains effectively sampling portions of the ring on the sky corresponding to the Virgo-LHO and Virgo-LLO combinations. This signal happened to be at a location in the sky which was particularly favorable for Virgo, with its response in that interferometer about a factor of 10 greater than in LHO and LLO. Not only does this serve as an informative example of the effectiveness of our search method, but also highlights the value of the global network of detectors, as just the LIGO observatories would have been hard pressed to make a firm detection of this signal.

Characterization

Upon completion of the search phase the analysis must be performed in the wavelet domain, once for each model, complete with a more thorough suite of parallel tempering and noise fitting.

The search chains were sorted into the 9-dimensional histogram in addition to a uniform contribution across the prior volume with grid spacings in each parameter determined by the Fisher approximation to the standard deviation of the PDF for

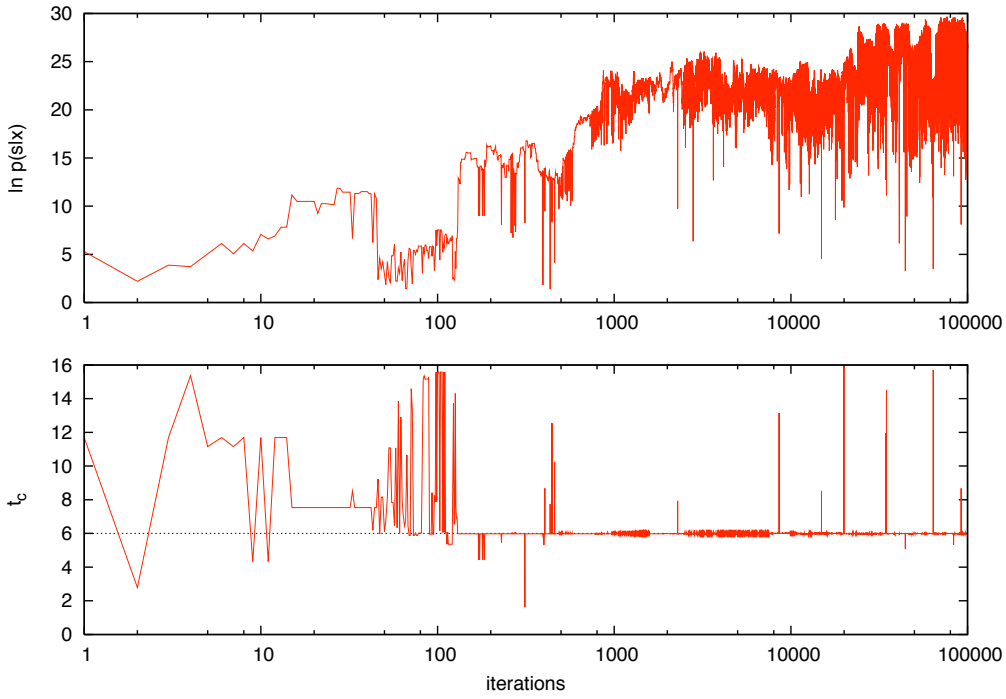


Figure 26: Search chain for an SNR = 7 signal with t_c injected at 6 s. The top panel shows the log-likelihood of the chain. The first 10,000 steps were the burn-in phase of the chain, after which the chain stabilized around the maximum likelihood. The bottom panel shows the time to coalescence chain. The expedited search performed very well, locating the injected signal within the burn-in time of the chain. Interestingly, the Bayes factor for this data was near one, revealing the signal models indistinguishability from the noise-only model despite the algorithm's success in locating the injection.

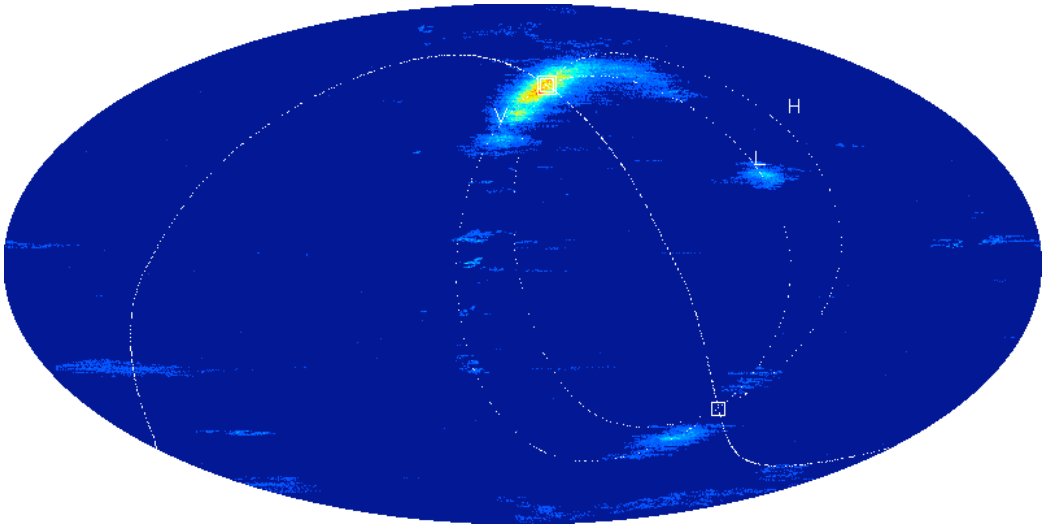


Figure 27: 2-D histogram of the sky location during the search phase. The location of each interferometer, projected onto the sky, is marked. Also indicated are the locations on the sky with equal time delays between a pair of interferometers (rings) or the entire network (double square for injected, single square for degenerate, sky location).

the MAP parameters recovered from the search phase. This distribution was used to drive new proposals for the characterization chains in addition to the standard Fisher proposals. The chains which were initialized randomly and the burn-in time set at 10^4 iterations as well, however inspection of the chains indicated that they had reached a stationary solution well before hand. We used 20 chains with a maximum temperature of 100. The marginalized PDFs for the sky location and the chirp mass - time to coalescence plane of Source 1 can be seen in Figure 28. As one would expect, the region of interest becomes more tightly confined around the injected parameters (double square for the sky location, white "+" for the mass-time plane) as the SNR increases.

Similar to the white dwarf analysis of chapter 6 we see an alternative sky location emerge as the best fit for a signal with a mid-range SNR. Close inspection of the mass-time plots shows that this erroneous sky location correlates with a slightly premature

time of coalescence – all the more support for the parallel tempering approach to signal characterization.

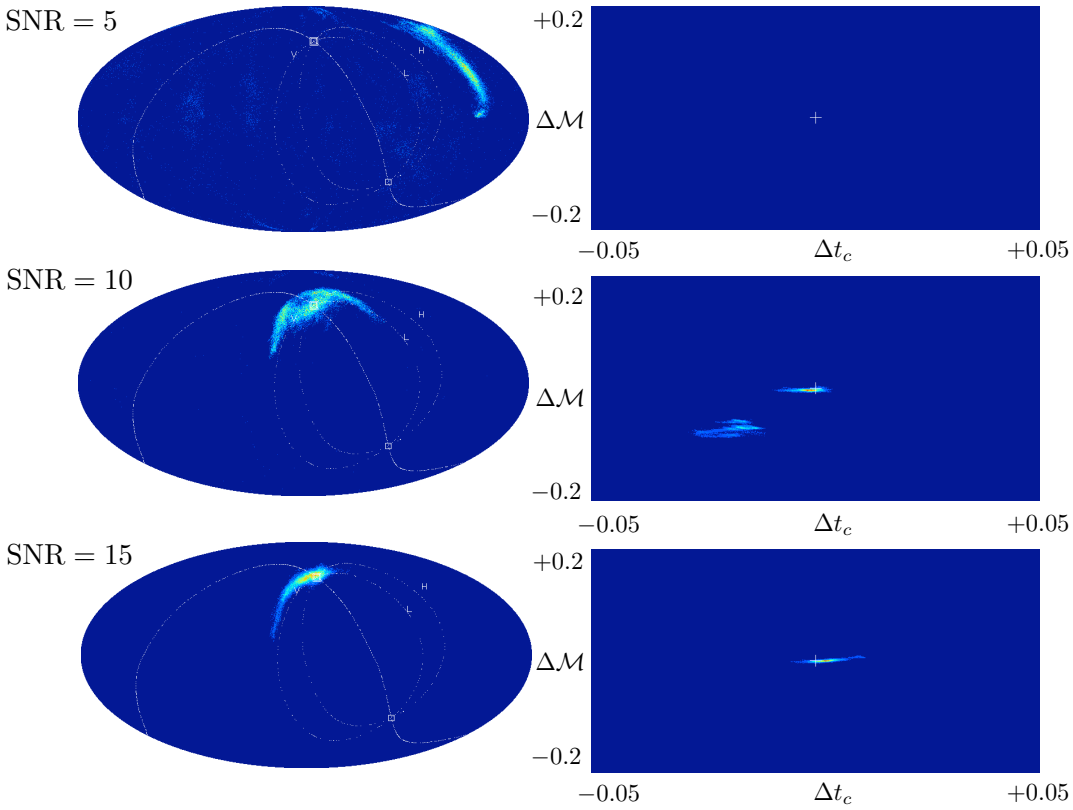


Figure 28: Marginalized PDFs of the $\delta - \alpha$ (left) and the $\ln \mathcal{M} - t_c$ (right) plane for data with injected SNRs indicated. The $\delta - \alpha$ PDF is shown in a Molweide projection. The location of the LIGO and Virgo detectors projected on to the sky are indicated, as are the sky locations which result in identical time delays for each pair of detectors, and the network as a whole. The $\log \mathcal{M} - t_c$ plane is for ± 0.2 (for the chirp mass) and ± 0.05 seconds (for t_c) of the injected parameter values.

Only the noise parameter that was in the wavelet pixel corresponding the black hole coalescence showed any significant elevation (in the noise-only model) most likely due to the rather tightly constraining prior that we have placed on the noise level in each block. Figure 29 shows the noise parameter distributions for the relevant noise block with a GW signal injected at SNR = 40. The power in the noise block which contained the majority of the signal power was $\sim$ 4 times that of the theoretical

noise level yet the noise-only model only elevated the noise parameter by ~ 1.4. To validate that it was the prior suppressing the chain from assigning the existing amount of power to the noise, we repeated the analysis with uniform noise priors. We see in Fig. 29 that the noise distribution in that block shows a definite preference for a value of approximately four – in perfect accordance with our prediction from the excess power level in that wavelet layer. This hints at the fact that the noise prior will need to be significantly relaxed in order to allow the noise model sufficient flexibility to account for glitches.

Evaluation

As previously described the expectation value of the likelihood is integrated over the inverse temperature β to yield the evidence for each model and ultimately the resultant Bayes factor. The results of this calculation can be seen in Figure 30. We find the signal model to transition more rapidly for Source 1 than for Source 2 – Source 2 being the signal with the sky location which was only clearly "audible" to a single detector in the network, thus hindering that signals distinguishability from the noise model.

While our exploration of this scenario has not yet been as thorough as the LISA white dwarf binary case, the early indication is that our approach has smoothly translated to the more challenging application. The primary features under test were that of the enhanced search and the effectiveness of its subsequent proposal distribution. Compared to the galactic binary searches, the relatively short burn-in times alone indicate that the addition of these new features yielded enormous returns regarding the efficiency of the algorithm. All of the black hole analysis was done with a rather small number of iterations compared to a typical MCMC analysis and yet we see excellent indications that we are reaching a convergent answer nonetheless.

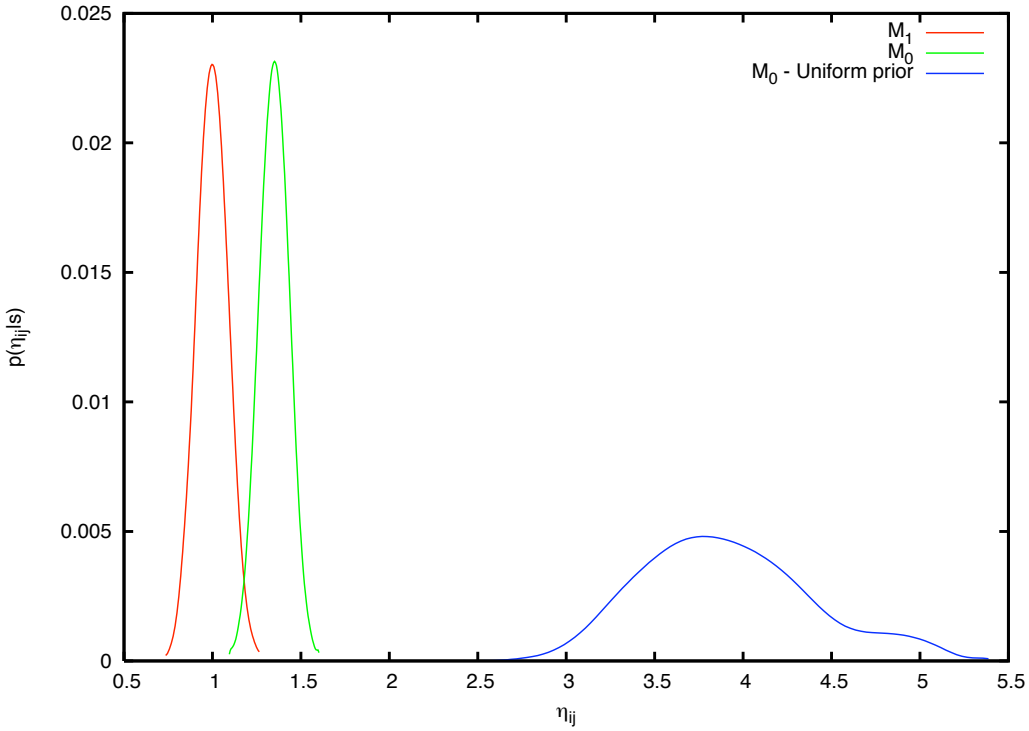


Figure 29: Marginalized PDFs of the LHO noise parameters corresponding to the "merger cell" of the black hole binary. $\mathcal{M}_0$ attempted to elevate the noise parameter from the expectation value for S_n^{ij} in order to account for the excess power in that cell but was restricted by the strong priors on the noise. $\mathcal{M}_1$ sufficiently accounted for that power allowing the noise parameter to remain at the appropriate level. We see that when uniform noise priors were adopted the noise preferred to be ~ 4 times the expected noise level, which we calculate to be approximately the amount of excess power in that block due to the GW signal.

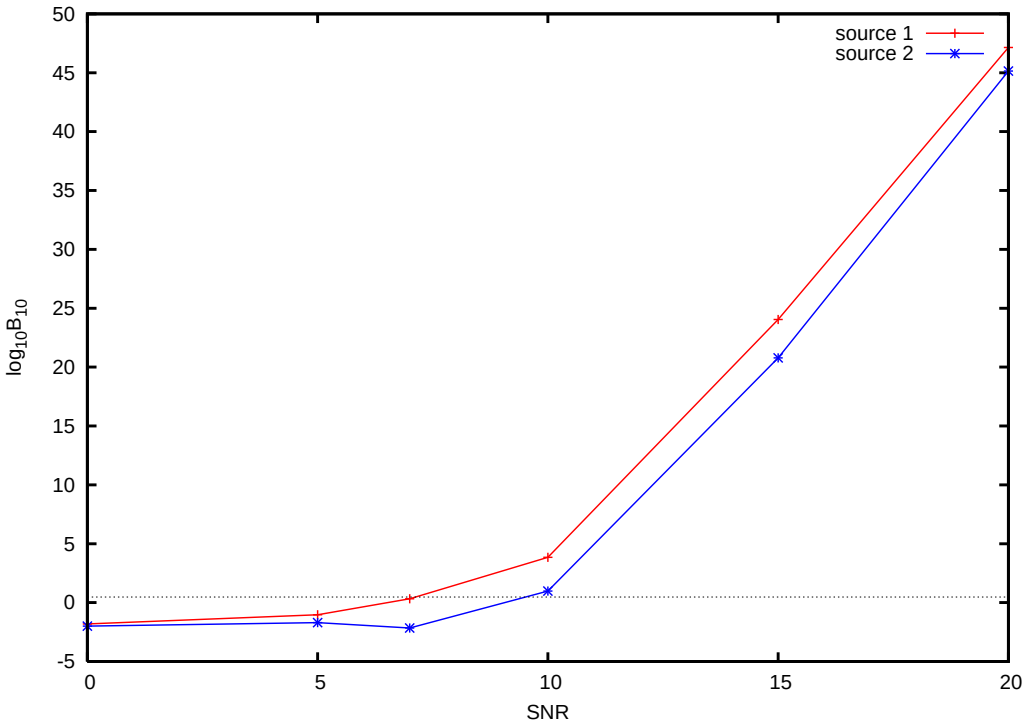


Figure 30: Bayes Factor as a function of increasing SNR for Source 1 [red] and Source 2 [blue]. Similar to the galactic binary example we find the transition point to be different for the two signals. The primary difference between the two was the poor sky location for Source 1, with only one of the observatories (Virgo) picking up a significant amount of power, underscoring the importance of using a cooperative global network of gravitational wave detectors.

This is extremely encouraging as we prepare for the next phases of the algorithm's enhancement.

CONCLUSIONS

The goal of this work was to develop a generic, comprehensive means for solving the gravitational wave detection problem. We consider a solution to be the fully resolved posterior distribution functions for each model in contention and the odds ratio between the models. As described in chapter 5, the detection problem can be separated into the search, characterization, and evaluation phases. We have devised a Bayesian approach which, for a given model and with minimal tuning, reliably resolves each phase. The development of our approach has centered around the "noise" or "noise plus a gravitational wave signal" manifestation of the problem and we have effectively demonstrated the algorithm's ability in both ground-based and space-based applications.

The simulated data are searched using a Metropolis-Hastings algorithm with parallel tempering which maximizes over phase, amplitude, and where applicable, time of arrival of the peak power. The chains explore the entire parameter space using a mixture of proposal distributions including moves along the approximated eigenvalues of the inverse covariance matrix, parameter jumps scaled by the estimated parameter uncertainties, and uniform draws on the full prior volume. The chains rapidly locate the regions of parameter space with high posterior probability and efficiently sample the surrounding regions resolving any secondary structure of the likelihood surface with significant posterior weight.

Although the resultant chains do not sample from the true PDF, they do map a qualitatively similar distribution which can be used as a highly efficient proposal distribution for chains which properly evaluate the likelihood and Hastings ratio, and whose samples therefore accurately reflect the target PDF. It is in this characterization

phase where the parallel tempering is extended to large "temperature" in order to accurately resolve the trend followed by the log-likelihood as a function of increasing heat. These MCMC chains use a similar set of proposal distributions as are employed in the search phase, with the added benefit of using the search-phase posterior as a proposal distribution. The subsequent efficient, global mixing allows for accurate inference of PDFs for a reasonable amount of computational effort.

The evaluation is performed by Thermodynamic Integration of the evidence, which integrates the expectation value of the log-likelihood from each of the different temperature chains from the characterization phase over the inverse-temperature of the chains. If only two models are under consideration, the Bayes factor, or evidence ratio, can be calculated revealing which (if either) of the models should be favored.

We have verified the accuracy of the Bayes factor calculations by comparing the results to the "gold standard" RJMCMC model selection method and found the two to be in excellent agreement. Other convergence tests, such as using different random number seeds for the chains – producing different "chain realizations" – showed the desired consistency.

The test cases for these techniques were more than just platforms for demonstrative purposes. During the development of this novel approach to gravitational wave detection some discoveries were made which have helped foster our understanding of the performance of the detectors.

In preparation for these computationally intensive methods we formulated a method for simulating LISA's response to gravitational radiation from a galactic compact binary system outlined in chapter 2. This software has been a vital tool towards the effectiveness of the Mock LISA Data Challenges as it has been the backbone for simulating the galactic foreground. It has also served as the waveform generator for the Blocked Annealed Metropolis-Hastings (BAM) algorithm [32] which has achieved

unparalleled performance in solving arguably the most important LISA data analysis challenge: That of globally fitting and regressing the resolvable galactic binaries allowing access to quieter sources with very little degradation to the residual.

In chapter 4 we performed the first detailed study of the detectability of the frequency evolution of compact galactic binaries. Assuming population synthesis models of the galaxy are valid we found that the majority of galactic binaries will not have measurable frequency evolution which is distinguishable from the prior predictions. If nature agrees with these predictions we can safely exclude the $\dot{f}$ parameter from the modeling of many sources. Alternatively, we can use the measured distribution of chirp frequencies to serve as evidence for, or against, the population synthesis models.

Chapter 6 tested the first generation of the PTMCMC detection algorithm on simulated LISA data with a single galactic binary injected at different amplitude to study when the source became detectable. The extensive use of parallel tempering revealed more complicated structure to the signal's posterior distribution function than was previously anticipated with several comparatively weighted regions in the PDF. This serves as an excellent cautionary tale against being solely interested in the parameters at the highest peak of the distribution, as other locations in parameter space can be as significant while going completely unnoticed by search techniques which are not able to efficiently explore the entire prior volume. We also clearly demonstrated how the noise realization affects the detectability of the source even when the noise is of the ideal stationary and gaussian variety. This is yet more support for the Bayesian approach to model selection as it clearly identifies the subtlety of how the signal interacts with the particular noise realization. This sensitivity is completely missed by setting thresholds based on monte carlo simulations of data, as is done by the ground-based analysis pipelines.

In chapter 7 we arrived at the current generation of the PTMCMC detection algorithm – the enhancements being the use of a maximized likelihood to accelerate the search phase and the subsequent utilization of that information to efficiently sample from the target distribution. The characterization phase was also performed in the wavelet domain, preparing the algorithm for the analysis of realistic data containing noise fluctuations.

This approach to the detection problem is now in a mature enough stage to begin addressing more complicated model selection problems. After further development we hope that it may serve as a foundation from which we can pursue the compelling questions in physics, astronomy, astrophysics, and cosmology which presumably encode their solutions in gravitational waves.

FUTURE EXTENSIONS OF THIS WORK

While we have demonstrated the successes of the PTMCMC detection algorithm there is still development work to be done. Our top priority is to implement the robust noise modeling discussed in chapter 7. This is the only blockade from seeing our approach take part in blind analysis of actual data. The structure is in place to make the appropriate adaptations and after sufficient testing on simulated "glitchy" noise we will be prepared to analyze LIGO S5 hardware and software injections in detector data.

In addition to more realistic noise realizations we also are looking to apply these techniques towards un-modeled burst sources. In chapter 2 we mention the limitless discovery potential of gravitational wave detectors, however all of the analyses discussed in this work have used analytic waveform templates. Although there is ample new information to be learned from these presupposed sources the detection and characterization of signals which have not been predicted is an exciting prospect. Detecting sources of this nature requires excess power searches, instead of matched filtering, and it is vital to such techniques that they are robust against noise excursions. Our ambition is to lean heavily on the pre-existing LIGO search pipelines and to devise a candidate detection "follow-up" which will further constrain the reconstructed waveform, in addition to calculating the evidence ratio for detection.

Another ongoing avenue of development is to apply this method to the more realistic LISA scenario where, instead of testing the "noise or noise plus GW" hypotheses, we must infer if there are N or $N+1$ sources in the data segment. As the number of signals in a scrutinized segment of data increases our brute-force method of searching for galactic binaries will announce its inefficiency. However, it should be a straight-

forward exercise to "back port" the search enhancements developed for the black hole case and we expect significant improvement from doing so.

While we do not anticipate nearly the number of transient noise events for LISA as are seen by LIGO and Virgo, there is still one outstanding problem for LISA noise modeling which may be tenable by our construction for modeling non-stationarities in LIGO data. The confusion-limited background of irresolvable galactic binaries is a non-stationary noise source for LISA, as the doppler and amplitude modulation of the source resultant from LISA's orbital motion depends on the sky location of each contributing binary. It is feasible that the wavelet approach to noise modeling might be best suited for handling the time varying confusion galactic foreground.

Most importantly, though, we are increasingly prepared for the first detection of gravitational waves. Detectors are soon to be working at unprecedented sensitivities and the GW analysis community now has many of the tools in place to respond to the exciting questions which are only tenable by way of the never-before heard messengers that are gravitational waves. It is an exciting time to be contributing to this field of research, and the best is yet to come.

REFERENCES CITED

[1] A. Raftery, *Practical Markov Chain Monte Carlo* (Chapman and Hall, London, 1996).

[2] Abramovici *et al.*, Science **256**, 325 (1992).

[3] B. Barish and R. Weiss, Phys. Today **52**, 44 (1999).

[4] B. Abbott *et al.*, arXiv:0810.0283v1 [gr-qc] (2008).

[5] B. Abbott *et al.*, arXiv:0901.0302v1 [gr-qc] (2009).

[6] `http://www.ego-gw.it/`

[7] P. Bender *et al.*, LISA Pre-Phase A Report (1998).

[8] `http://www.rssd.esa.int/index.php?project=LISAPATHFINDER`

[9] R. Umstatter *et al.*, Class. Quant. Grav. **21**, S1655 (2004).

[10] N. J. Cornish and J. Crowder, Phys. Rev. **D72**, 043005 (2005).

[11] P. C. Gregory, Mon. Not. Roy. Astron. Soc. **374**, 1321 (2007).

[12] P. Mukherjee, D. Parkinson, and A. R. Liddle, ApJ. **638**, L51 (2006).

[13] R. Trotta, astro-ph/0504022 (2005).

[14] R. Trotta, Contemp. Phys. **49**, 71 (2008).

[15] N. J. Cornish and T. B. Littenberg, Phys. Rev. **D76**, 083006 (2007).

[16] T. B. Littenberg and N. J. Cornish, arXiv:0902.0368 [gr-qc] (2009).

[17] T. B. Littenberg and N. J. Cornish, in preperation (2009).

[18] K. Thorne, in *"Gravitational Radiation"* in *300 Years of Gravitation*, edited by S. Hawking and W. Israel (Cambridge University Press, 1987), pp 330-458.

[19] C. Will, *Was Einstein Right?: Putting General Relativity to the Test*, 2nd ed. (Basic Books, New York, U.S.A., 1993).

[20] R. Hulse and J. Taylor, Astrophys. J. **195**, (1975).

[21] R. Hulse, Rev. Mod. Phys. **66**, 699 (1994).

[22] C. Cutler and K. S. Thorne, arXiv:0204.090 [gr-qc] (2002).

[23] http://www.ligo.org

[24] http://geo600.aei.mpg.de/

[25] http://tamago.mtk.nao.ac.jp/

[26] http://astrogravs.nasa.gov/docs/mldc

[27] M. Maggiore, *Gravitational Waves* (Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, 2008), Vol. 1: Theory and Experiments.

[28] G. Nelemans, L. Yungelson, and S. Portegies Zwart, Astron. Astrophys. **375**, 890 (2001).

[29] J. Crowder and N. J. Cornish, Phys. Rev. **D70**, 082004 (2004).

[30] R. Umstatter *et al.*, Phys. Rev. **D72**, 022001 (2005).

[31] S. E. Timpano, L. J. Rubbo, and N. J. Cornish, Phys. Rev. **D73**, (2006).

[32] J. Crowder and N. Cornish, Phys. Rev. **D75**, 043008 (2007).

[33] L. J. Rubbo, N. J. Cornish, and O. Poujade, Phys. Rev. **D69**, 082003 (2004).

[34] N. J. Cornish and L. J. Rubbo, Phys. Rev. **D67**, 022001 (2003).

[35] T. A. Prince, M. Tinto, S. L. Larson, and J. W. Armstrong, Phys. Rev. **D66**, 122002 (2002).

[36] D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial* (Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, 2006), No. ISBN 0 19 856831 2.

[37] P. C. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge University Press, The Edinburgh Building, Cambridge CB2 2RU, UK, 2006), No. ISBN 0 521 84150 X.

[38] G. Schilling, *Flash! The hunt for the biggest explosions in the universe* (Cambridge University Press, The Edinburgh Building, Cambridge CB2 2RU, UK, 2002), No. ISBN 0 521 80053 6.

[39] B. Abbott and f. t. L. S. Collaboration, Astrophys. J. **681**, 1419 (2008).

[40] A. C. Searle, P. J. Sutton, and M. Tinto, arXiv:0809.2809 [gr-qc] (2008).

[41] P. Green, Biometrika **82**, 711 (1995).

[42] A. R. Liddle *et al.*, astro-ph/0703285 (2007).

[43] P. Goggans and Y. Chi, in *Bayesian Inference and Methods in Science and Engineering*, edited by J. Rychert, G. Erickson, and C. Smith (American Institute of Physics, USA, 2004).

[44] J. Skilling and D. MacKay, `http://www.inference.phy.cam.ac.uk/bayessys/`

[45] N. Metropolis, A. Rosenbluth, R. M.N., A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[46] W. Hastings, Biometrika **57**, 97 (1970).

[47] D. Gamerman, *Markov Chain Monte Carlo: Stocastic Simulation of Bayesian Inference* (Chapman and Hall, London, 1997).

[48] N. J. Cornish and E. K. Porter, Class. Quant. Grav. **23**, S761 (2006).

[49] M. van der Sluys *et al.*, Class. Quant. Grav. **25**, 184011 (2008).

[50] N. J. Cornish and E. K. Porter, Class. Quant. Grav. **24**, S501 (2007).

[51] N. J. Cornish, arXiv:0804.3323 [gr-qc] (2008).

[52] J. Gair, E. K. Porter, S. Babak, and L. Barack, Class. Quant. Grav. **25**, (2008).

[53] C. Cutler and E. E. Flanagan, Phys. Rev. **D49**, 2658 (1994).

[54] M. Sambridge, K. Gallagher, A. Jackson, and P. Rickwood, Geophys. J. Int. **167**, 528 (2006).

[55] H. Lopes and M. West, Statistica Sinica **14**, 41 (2004).

[56] K. van der Linde, MVE: Minimum Volume Ellipsoid estimation for robust outlier detection in multivariate space. `http://www.kimbdlinde.com/professional/mve.html`

[57] J. Dickey, Ann. Math. Stat. **24**, 204 (1971).

[58] G. Schwarz, Ann. Stats. **5**, 461 (1978).

[59] K. A. Arnaud *et al.*, Class. Quant. Grav. **24**, S551 (2007).

[60] S. Babak *et al.*, Class. Quant. Grav. **25**, 114037 (2008).

[61] N. Seto, Mon. Not. R. Astron. Soc. **333**, 469 (2002).

[62] D. Lindley, Biometrica **44**, 187 (1957).

[63] J. Veitch and A. Vecchio, Phys. Rev. **D78**, 022001 (2008).

[64] J. Veitch and A. Vecchio, Class. Quant. Grav. **25**, 184010 (2008).

[65] R. Swendsen and J. Wang, Phys. Rev. Lett. **57**, 2607 (1986).

[66] P. Green, in *Highly Structured Stochastic Systems*, edited by P. Green, N. Hjort, and S. Richardson (Oxford University Press, 2003).

[67] C. Andrieu and A. Doucet, IEEE Trans. Signal Process. **47**, 2667 (1999).

[68] R. Umstatter *et al.*, Class. Quant. Grav. **22**, S901 (2005).

[69] M. Adams and N. J. Cornish, in preperation (2009).

[70] F. Feroz and M. Hobson, Mon. Not. R. Astron. Soc. **384**, 449 (2008).

[71] B. Allen *et al.*, Phys. Rev. **D67**, 122002 (2003).

[72] S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher, Class. Quant. Grav. **25**, 114029 (2008).

[73] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C, The art of Scientific Computing*, 2 ed. (Cambridge University Press, 1993).

[74] B. Allen, J. D. E. Creighton, E. E. Flanagan, and J. D. Romano, Phys. Rev. **D67**, 122002 (2003).

[75] P. C. Peters and J. Mathews, Phys. Rev. **131**, 435 (1963).

[76] R. N. Lang and S. A. Hughes, astro-ph/0710.3795 (2007).

[77] E. K. Porter and N. J. Cornish, arXiv:0804.0332 [gr-qc] (2008).

[78] L. Blanchet, Phys. Rev. D **51**, 2559 (1995).

[79] N. J. Cornish and E. K. Porter, Class. Quant. Grav. **24**, 5729 (2007).

[80] J. S. Key and N. J. Cornish, Phys. Rev. **D79**, 043014 (2009).