Developing a Stochastic Gravitational Wave Pipeline

Eli Wiston

May 2021

Abstract

With the detection of many individual gravitational wave signals, researchers have begun searching for a stochastic gravitational wave background. This superposition of weak, unresolved gravitational wave signals could hold a wealth of astrophysical and cosmological information. Studying the background with current detector sensitives could provide a measure of matter distributions in the universe. Eventually these searches could provide concrete evidence of inflation and act as a primordial CMB. This project will focus on developing time series data folding for the stochastic gravitational wave background analysis pipeline. With the implementation of data folding, anisotropic directional searches can be carried out far more efficiently. This proposal will detail the underlying physics and overall direction of the project.

1 Introduction

In 2015, LIGO made the first direct detection of a gravitational wave (GW) signal. Since then, interferometers have measured many more signals from black hole and neutron star binaries. These binaries must either have very high mass or be very compact in order to be detected by current ground-based detectors. However, the sky is filled with gravitational wave signals below detection thresholds that, when analyzed as a whole, contain a great deal of both astrophysical and cosmological information. These signals are unresolved, numerous, and best described according to probability distributions, hence they are known as the stochastic gravitational wave background (SGWB).

SGWB searches can be performed as either all-sky or directional searches. All-sky, isotropic searches can be used to characterize the average GW signal in the universe. Directional searches can be used to map anisotropies in the GW distribution. Due to the Earth's rotation, ground-based detectors are able to measure a signal from each part of the sky once per sidereal day.

This project is focused on developing a key component of LIGO's SGWB pipeline: data folding. Because directional searches rely on data that is periodic over one day, we can compress the time series information we measure. By folding gravitational wave data over one sidereal day, we can vastly improve the efficiency of current and future directional stochastic searches.

The contents of this proposal will be presented as follows. In Section 1, I present a brief overview of the motivations for stochastic gravitational wave searches. In Section 2, I provide the necessary background, starting with gravitational waves in general and then going over stochastic signals and their measurement and analysis. Section 3 presents the goals and objectives of this specific project. Section 4 shows the data folding approach, modeled after the work in [1]. Finally, in Section 5, I detail the work plan for the summer.

2 Motivations

Gravitational waves allow researchers to probe the Universe without relying on electromagnetic signals. This can be incredibly useful, providing independent measurements of electromagnetic sources and new measurements of GW sources. High signal-to-noise measurements can provide insight into individual events, but the stochastic gravitational wave background can provide information about large scale structure and cosmology.

The earliest electromagnetic signals come from the time of last scattering, at a redshift of around $z = 1100$, and comprise the Cosmic Microwave Background (CMB) [2]. Before then, the universe was too opaque for photons to travel very far. However, gravitational waves were able to propagate all the way back in the early moments of the universe. Eventually, stochastic gravitational wave searches may be able to find direct evidence of inflation and provide information about early universe phase transitions.

Current detectors lack the sensitivity to measure the comparatively weak signals from these cosmological background events, but they can be used to study lower redshift astrophysical sources. These sources are expected to be distributed somewhat anisotrophically. A directional search looking at these anisotropies in the SGWB can probe at the universe's underlying mass distribution. In particular, these searches can provide strong tests of the expected distribution of compact binary coalescences (CBCs)[3].

3 Background

3.1 Gravitational Waves

Gravitational waves manifest as strains, or changes in length per unit length. They arise when the quadrapole mass moments of objects, I_{uv} have a time dependence [4]. This is why the direct detections already made involve compact mass objects inspiraling. In the context of general relativity, gravitational waves can be thought of as linear perturbations of the background metric g_{uv}. Assuming that the gravitational field is weak and non-stationary, one can show that the solution to the Einstein field equations for such a perturbation, can be constructed as a plane wave, propagating at the speed of light [5].

Currently, the primary method for detecting gravitational waves is ground-based interferometry. The basic setup is that of a Michelson interferometer. A laser beam is split along two long arms and reflected off of mirrors, ending up in two detectors. Gravitational waves strain the travel distance between the detectors, manifesting in a change in light intensity at the frequency of the wave, which one can directly measure. From the measurements, one may be able to determine the frequency, amplitude, direction, and polarization of the wave. Gravitational wave strains are incredibly small, so interferometers have to be extremely sensitive to detect them. There are many sources of noise that also make detection difficult, including seismic activity and Brownian motion of the detector mirrors [4].

3.2 Stochastic Signals

Due to the low signal-to-noise nature of gravitational waves, only the most extreme GW events can be directly detected. However, these types of events constitute a tiny fraction of all gravitational wave signals; the rest comprise the stochastic gravitational wave background. These stochastic signals are weak, independent, random, and unresolved. The distinction between a stochastic and resolvable signal can be unclear, as it may depend on modelling decisions or the precision of a detector. A signal can be operationally defined as stochastic if a Bayesian model selection calculation prefers a stochastic signal model over any deterministic signal model [2]. There are two broad categories of stochastic GW signals, based on the nature of the GW source: astrophysical and cosmological. Astrophysical signals occur at low redshift and are stochastic in the the limit that number of sources N is very high. They are mainly comprised of compact binary systems. Cosmological signals arise from processes in the early Universe. They can be described stochastically as a result of the assumed homogeneity and isotropy of the universe. All inflationary models have some gravitational wave byproducts and phase transitions are also predicted to produce detectable signals [6]. LIGO does not have the sensitivity to measure weak cosmological signals, so this analysis will be aimed at measuring the astrophysical foreground.

A key parameter of interest in SGWB searches is Ω_{gw}, the fractional energy density of gravitational waves in the universe. The parameter can be expressed as $\Omega_{gw}(f, \hat{n})$, where f is the wave frequency and $\hat{n}$ is the direction [2]. Searches performed on LIGO's first three runs have not detected a stochastic background, but have set upper limits on Ω_{gw}. These limits fall in line with predictions based on the expected distribution of compact binary systems [7, 3].

3.3 Measurement and Analysis

The stochastic signal h_{ab} can be expressed as a superposition of sine waves as follows:

$$h_{ab}(t, \vec{x}) = \int_{-\infty}^{+\infty} df \int d^2\Omega_{\hat{n}} h_{ab}(t, \hat{n}) e^{i2\pi f(t+\hat{n}\cdot\vec{x}/c)}$$

where $h_{ab}(t, \hat{n})$ are the random variable Fourier coefficients that can be used to statistically describe the background. We can assume the background has zero mean, so $< h_{ab} >= 0$. For Gaussian sources, the signal is entirely characterized by its second order moment. These quadratic expectation values can be defined in terms of the strain density power spectrum S_h. From S_h, $\Omega_{gw}(f)$ can be found through a simple relation:

$$S_h(f, \hat{n}) = \frac{3H_0^2}{8\pi^3} \frac{\Omega_{gw}(f, \hat{n})}{f^3}$$

The signal-to-noise of the stochastic background is far too low to extract any meaningful information from a detector. However, by cross-correlating the strain data between multiple detectors, the stochastic signal can be found. The detectors will be measuring the same true signal, so those will add coherently. The noise in each detector, on the other hand, is independent and will not add coherently. Given a Gaussian approximation, the noise will be averaged down as $\frac{1}{\sqrt{time}}$, while the signal will be remain unsuppressed. The signal cross-correlation is directly related to key parameters, including S_h. By performing maximum likelihood analyses, one can calculate S_h from the observed cross-correlated data [2].

The longer the observation time being analyzed, the more the noise is suppressed. However, dealing with long periods of time is computationally demanding, both in terms of processing power and storage. This issue can be confronted by folding the strain data. We fold over one sidereal day so anisotropies in the same region of the sky can add coherently. Ain, Dalvi, and Mitra developed the algebra and algorithm for such data folding [1]. Testing on LIGO S5 data, they found very significant decreases in computation time. An analysis of the full S5 data on folded data was faster than the same analysis of unfolded data by a factor of 300. Furthermore, the data quality was virtually unchanged; the differences between folded and unfolded maps were orders of magnitude smaller than the values themselves. The folding increases efficiency, portability, and convenience, facilitating more analyses of strain data, carried out at faster rates.

4 Objectives

The goal of this project is to implement a data folding algorithm like [1] in Python for use in the LIGO stochastic gravitational wave pipeline. This folding should be performed as efficiently as possible, with negligible loss in data quality. By the end of the project, we hope to conduct a proof of principle test with mock data.

The data folding should be fixed to one sidereal day, as this will add isotropic signals coherently for any Earth-based detector. However, the code should be flexible to allow for the addition of more detector data. At any given time, a detector is only sensitive to certain regions of the sky. The sensitivity of a set of detectors is given by the overlap function. As seen in Figure 1 from [6], the overlap function for the two LIGO detectors has large areas of low sensitivity. While most of the sky will be covered as the Earth rotates, the sampling will be uneven. By designing our data folding code with the flexibility to allow the addition of more detectors, in new locations, we can enable a broader sampling of the sky.

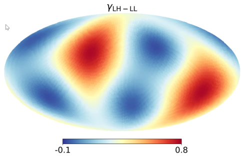

Figure 1: Instantaneous overlap function for pairing of LIGO Hanford and LIGO Livingston detectors in galactic coordinates [6]

5 Approach

The overall goal of a directional search is to estimate the amplitude of the SGWB power spectra density (PSD) as a function of position in the sky. For current searches, the shape of the PSD as a function of frequency is assumed. For the CBC dominated background in this project, this shape is well known.

The time series data from a baseline of two detectors, $s(t)$ is the sum of the stochastic signal and detector noise. Following the approach in [1], it is convenient to divide the data for each baseline $\mathcal{I}$ into short time segments of length τ. A Fourier transform is then performed on each of these segments as follows:

$$\tilde{s}_{\mathcal{I}}(t;f) = \int_{t-\tau/2}^{t+\tau/2} dt' s(t') e^{-i2\pi f t'}$$

The maximum likelihood solution for the coefficients of the SGWB skymap can be calculated using two matrix quantities, the dirty map X and the Fisher information matrix Γ:

$$X = \frac{4}{\tau} \sum_{Ift} \frac{H(f)\gamma^{I*}_{ft,\alpha}}{P_{I_1}P_{I_2}} \tilde{s}_{\mathcal{I}_1}(t;f)\tilde{s}_{\mathcal{I}_1}(t;f)$$

$$\Gamma = 4 \sum_{Ift} \frac{H^2(f)\gamma^{I*}_{ft,\alpha}}{P_{I_1}P_{I_2}} \tilde{s}_{\mathcal{I}_1}(t;f)\tilde{s}_{\mathcal{I}_1}(t;f)$$

$H(f)$ is the expected shape of the stochastic background's frequency power spectral density. $P_{I_{1,2}}$ is the one-sided power spectral density of the noise for a segment of time. Since the noise dominates over the signal for short time segments, this quantity can be accurately estimated from the data. $\gamma^{I*}_{ft,\alpha}$ contains all the specific information about the detectors' antenna pattern functions, baseline separations, and polarization basis.

Crucially, both quantities involve summations over all time segments. The time t can be re-expressed as $t = i_{day} \times T_s + t_s$, where i_{day} is the index of the sidereal day, T_s is the duration of a sidereal day, and t_s is the remaining time within a day. The summations over time can therefore be broken down into two parts $\sum_{i_{day}}$ and $\sum_{t_s}$. Performing the first sum folds the data, with the information from months or years compressed into one sidereal day. This process is shown below in Figure 2, visualized by [1].

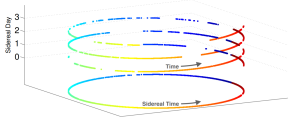

Figure 2: Folding process visualized for 3 days of LIGO S5 data. The three top rings are projected onto the ring below, representing the folding data. Gaps in the rings represent missing data. [1]

The data folding is somewhat complicated by the common application of window functions to the data. These functions help reduce spectral line leakage, but lead to an effective loss of data. To prevent this data loss, we use 50% overlapping windows in SGWB analysis. These overlaps lead to some additional complications in the data folding algebra, which manifest as corrections to the X and Γ, but do not impede our ability to fold the data [1].

6 Work Plan

This project can be broken down into three distinct parts:

1. Working with a simple SGWB pipeline in order to better understand both the data and the underlying astrophysics
2. Developing the data folding code for the working pipeline and testing to ensure it works properly
3. Using the new pipeline to carry out analysis with the folded data

Part 1, familiarizing myself with a simple pipeline will take place in the first 3 weeks of the program. The next 4 weeks will be devoted to Part 2, developing the data folding code. Part 3, using the new pipeline, will be carried out in the final 3 weeks. Due to the nature of code development, there is some uncertainty as to how long Part 2 will take. Part 3 is not necessary to the success of the project, so if development takes longer than expected, I can always cut down on the additional analysis. That being said, there is no shortage of interesting simulations and analysis to perform with the pipeline. If development is quicker than expected, Part 3 can also be extended to include more.

Presented below is a rough timeline of the entire program:

Date	Event
June 15	Program begins, start working with simple SGWB pipeline
July 6	Start development of data folding code
July 5-11	Submit interim report 1
July 26-August 1	Submit interim report 2 and abstract
August 3	Start analysis using new pipeline
August 18-20	Final presentation
September 24	Submit final report

References

[1] Ain, A., *et al.* (2015). ”Fast Gravitational Wave Radiometry using Data Folding.” arXiv:1504.01714v1

[2] Romano, J. & Cornish, N. (2017). ”Detection methods for stochastic gravitational-wave backgrounds: a unified treatment.” arXiv:1608.06889v2

[3] LIGO Collaboration (2021). ”Upper Limits on the Isotropic Gravitational-Wave Background from Advanced LIGO’s and Advanced Virgo’s Third Observing Run.” arXiv:2101.12130v1

[4] Grote, H. & Reitze, D. H. ”First Generation Interferometric Gravitational Wave Detectors.”

[5] Schutz, B. F. (2011). A First Course in General Relativity. Cambridge: Cambridge University Press.

[6] Renzini, A. (2020). Mapping the Gravitational Wave Background.

[7] LIGO Collaboration (2019). ”Directional limits on persistent gravitational waves using data from Advanced LIGO’s first two observing runs.” arXiv:1903.08844v3