Data Folding for the Stochastic Gravitational Wave Pipeline: Interim Report #1

Eli Wiston
University of Pennsylvania

Mentors: Arianna Renzini and Colm Talbot
California Institute of Technology

(Dated: August 19, 2021)

While many individual gravitational wave signals have been detected, researchers are still searching for a stochastic gravitational wave background. This superposition of weak, unresolved gravitational-wave signals could hold a wealth of both astrophysical and cosmological information. Studying both the isotropic and anisotropic components of the background at current detector sensitives could provide a measure of matter distributions and large-scale structure in the Universe. Eventually these searches may provide concrete evidence of inflation and act as a primordial analog to the Cosmic Microwave Background. This project will focus on developing a data folding algorithm 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 report will detail the underlying physics, overall direction of the project, and progress so far.

I. 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 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. Isotropic searches model the background with no directional dependence and can be used to characterize the average GW signal in the universe. Directional searches account for potential variation across the sky and can be used to map anisotropies in the GW distribution. Due to the Earth's rotation, ground-based detectors measure a signal from each part of the sky over a sidereal day.

This project is focused on developing a key component of LIGO's SGWB pipeline: data folding. Because LIGO data 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 2, I present a brief overview of the motivations for stochastic gravitational wave searches. In Section 3, I provide the necessary background, starting with gravitational waves in general and then going over stochastic signals and their measurement and analysis. Section 4 presents the goals and objectives of this specific project. Section 5 shows the data folding approach, modeled after the work in [1]. Section 6 details my progress in the first 3 weeks of the program. Finally, in Section 7, I describe the challenges I've faced so far and what challenges I expect moving forward.

II. 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 anisotropically. 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].

III. BACKGROUND

A. 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 nonstationary, 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, perpendicular arms and reflected off of mirrors, combining again at a photodetector. Gravitational waves strain the travel distance along the arms, creating an optical phase difference between the two beams. With a new phase difference, the electromagnetic laser waves interfere slightly differently, manifesting in a change in light intensity at the frequency of the wave, which one can directly measure. From these 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].

B. 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 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. Early universe phase transitions are also predicted to produce detectable signals [6]. LIGO does not currently 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 [3, 7].

C. Measurement and Analysis

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

$$h_{ab}(t, \vec{x}) = \int_{-\infty}^{+\infty} df \int d^2\Omega_{\hat{n}} h_{ab}(t, \hat{n}) e^{i2\pi f(t+\hat{n}\cdot\vec{x}/c)} \quad (1)$$

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 therefore 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:

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

The signal-to-noise of the stochastic background is far too low to extract any meaningful information from a single 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.

IV. 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 anisotropic signals coherently for any Earth-based detector. However, the code should be flexible to allow for the addition of more detectors' 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 FIG. 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.

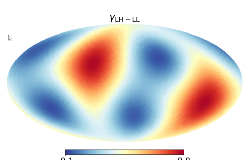

FIG. 1. Instantaneous overlap function for pairing of LIGO Hanford and LIGO Livingston detectors in galactic coordinates [6]

V. 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. This assumption will work well for this project, since the PSD shape for the CBC dominated background 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:

$$\tilde{s}_{\mathcal{I}}(t;f) = \int_{t-\tau/2}^{t+\tau/2} dt' s(t') e^{-i2\pi f t'} \tag{3}$$

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

$$\hat{P} = \Gamma^{-1} \cdot X \tag{4}$$

where,

$$X = \frac{4}{\tau} \sum_{Ift} \frac{H(f)\gamma^{I*}_{ft,\alpha}}{P_{I_1} P_{I_2}} \tilde{s}_{\mathcal{I}_1}(t;f) \tilde{s}_{\mathcal{I}_1}(t;f) \tag{5}$$

$$\Gamma = 4 \sum_{Ift} \frac{H^2(f)\gamma^{I*}_{ft,\alpha}}{P_{I_1} P_{I_2}} \tilde{s}_{\mathcal{I}_1}(t;f) \tilde{s}_{\mathcal{I}_1}(t;f) \tag{6}$$

$H(f)$ 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. $\gamma^{I*}_{ft,\alpha}$ is the overlap function, containing 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 FIG. 2, visualized by [1].

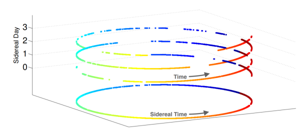

FIG. 2. Folding process visualized for 3 days of LIGO S5 data. The three top rings are projected onto the ring below, representing the folding data. Gaps in the rings represent missing data. [1]

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].

VI. PROGRESS FROM FIRST 3 WEEKS

A. Working with Overlap Functions

As seen in the previous section, the overlap function γ is an essential quantity for SGWB calculations. γ describes the relationship between the power of the SGWB and the cross-correlated response of a baseline of two detectors [8]. Essentially, it is a measure of the baseline's sensitivity to different parts of the sky, which varies as a function of frequency and time.

To begin, I focused on visualizing these overlap functions as skymaps and seeing how they evolve in time. As seen in FIG. 3, the function has two large areas of maximum sensitivity and a band of minimum sensitivity separating them. Over the course of the day, the overlap function rotates along with the rotation of the Earth, completing one rotation each sidereal day. This periodicity is what allows my project to work. If the sensitivity is periodic over one sidereal day and the SGWB is relatively stationary, the response from each baseline should also be periodic each day. This is what enables us to fold a year's worth of data down to one day without loss of information.

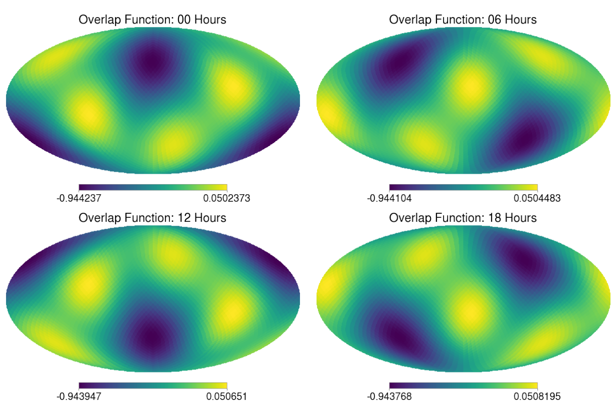

FIG. 3. The time-evolved overlap function for the baseline of the LIGO Hanford and LIGO Livingston detectors.

Next, I looked at how these skymaps evolved with frequency. The frequency of the signal enters the overlap calculation through a factor of $e^{2\pi i f \Delta t}$, where Δt is the time delay between waves reaching both detectors. In FIG. 4, we see that as frequency increases, the overlap function develops more maxima and minima. With a higher frequency, there are more points where the signals arriving at each detector can go in and out of phase.

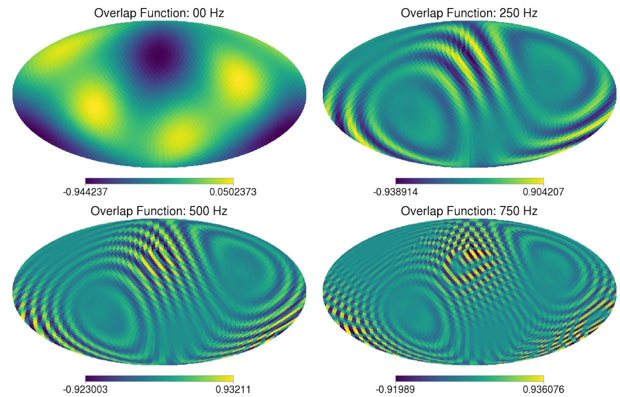

FIG. 4. The frequency-evolved overlap function for the baseline of the LIGO Hanford and LIGO Livingston detectors.

Alternatively, these skymaps can be visualized as "peanut plots," shown in FIG. 5. In these plots, the sky is represented 3-dimensionally. The radius of each point on the sphere represents the value of the overlap function at that point. As these plots evolve in time, they spin in 3-dimensional space. As they evolve in frequency, the additional maxima and minima appear as spikes in the plot.

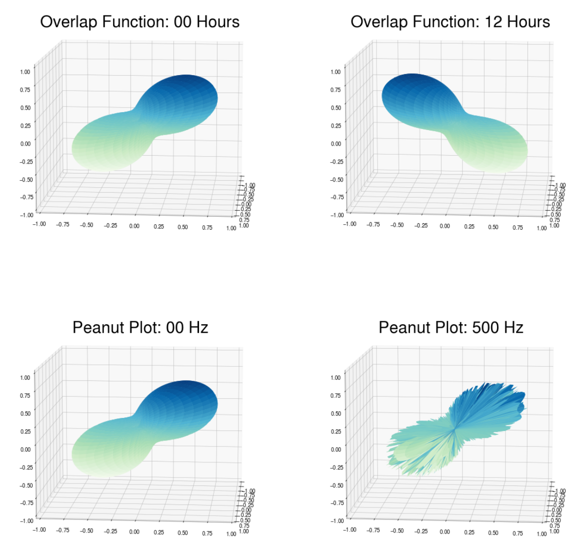

FIG. 5. Alternate visualization of how the overlap function evolves in frequency and time. The value of the function at any given point corresponds to the radius.

B. Spherical Harmonic Transformations

To gain further insight into these overlap functions, I worked on re-expressing them in terms of spherical harmonics. The spherical harmonic functions, Y_m^l, form a complete, orthonormal basis and, therefore, any function defined on the surface of a sphere can be re-written as a summation. For the overlap function:

$$\gamma(\theta, \phi) = \sum_{l,m} \gamma_{lm} Y^l_m(\theta, \phi) \tag{7}$$

The order of the harmonic l denotes the moment (monopole, dipole, etc.). The specific mode m corresponds to the frequency of oscillations. Each moment l has modes ranging from -l to +l. Each (l,m) mode can be found by integrating the overlap function multiplied by the spherical harmonic over the sky.

$$\gamma_{lm} = \int d\Omega \gamma(\theta, \phi) Y^l_m(\theta, \phi) \tag{8}$$

In FIG. 6, this transformation is applied to the first 3 moments of the overlap function. As expected, the monopole term is constant in time. Due to the azimuthal symmetry of the system, all 3 dipole terms are equal to 0. The quadrupole terms illustrate the mode m corresponds to the frequency of oscillations. The $m = 0$ term is constant in time. The $m = -1, +1$ terms go through one period over the course of a day, while the $m = -2, +2$ terms go through two periods. A factor of $(-1)^m$ accounts for the reflection in the $-1, +1$ modes.

The power spectra for each moment can be calculated by summing over all the m modes in the following equation:

$$\Gamma_l = \frac{1}{2l+1} \sum_m |\gamma_{lm}|^2$$

The power for each moment at fixed time is shown in FIG. 7. Here we see that the azimuthal symmetry ensures that all odd moments drop to 0.

VII. CHALLENGES

The main challenges I've faced so far were adjusting to the practices of new Python modules and functions. For example, SciPy uses a different convention for defining θ and ϕ than both Bilby and HealPy. When calculating the spherical harmonic transformations, I was faced with long computation times on my integrations. I fixed this by adjusting from an integration method that recalculated the overlap each time to one that utilized a fixed array of overlap function values.

My next steps on the project will be beginning to work on the data folding algorithm. I expect there will be more challenges working with new Python modules. Since I'm working on something that will be integrated into a broader pipeline, I will also need to learn the conventions of the group to ensure my code is readable and easy for others to work with.

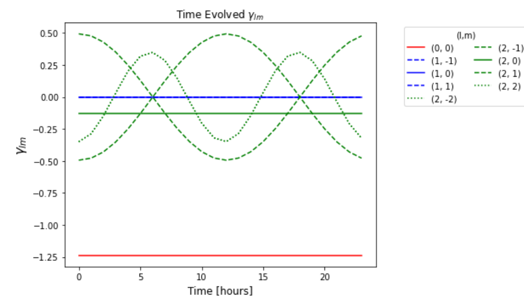

FIG. 6. First 3 γ_{lm} moments as a function of time. Monopole terms are shown in red, dipole in blue, and quadrupole in green.

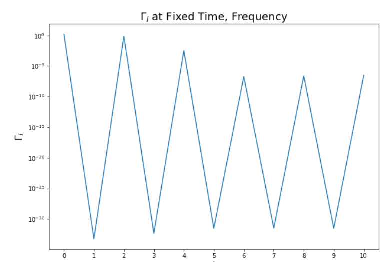

FIG. 7. Γ_l for the first 10 moments of the overlap function.

[1] A. Ain, P. Dalvi, and S. Mitra, Fast gravitational wave radiometry using data folding, Phys. Rev. D **92**, 022003 (2015).

[2] J. Romano and N. Cornish, Detection methods for stochastic gravitational-wave backgrounds: A unified treatment, Living Reviews in Relativity **20** (2017).

[3] T. L. S. Collaboration, T. V. Collaboration, *et al.*, Upper limits on the isotropic gravitational-wave background from advanced ligo's and advanced virgo's third observing run (2021).

[4] H. Grote and D. H. Reitze, First-Generation Interferometric Gravitational-Wave Detectors, in *46th Rencontres de Moriond on Gravitational Waves and Experimental Gravity* (Moriond, Paris, France, 2011).

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

[6] A. Renzini, *Mapping the gravitational-wave background*, Ph.D. thesis, Imperial College London (2020).

[7] B. Abbott *et al.* (The LIGO Scientific Collaboration and the Virgo Collaboration), Directional limits on persistent gravitational waves using data from advanced ligo's first two observing runs, Phys. Rev. D **100**, 062001 (2019).

[8] L. S. Finn, S. L. Larson, and J. D. Romano, Detecting a stochastic gravitational-wave background: The overlap reduction function, Phys. Rev. D **79**, 062003 (2009).