Improvements in Computational Efficiency for CBC Analysis of Advanced Gravitational Wave Detector Data

James Michael Bell[1, 2]

[1] *Millsaps College 1701 N. State Street Jackson, MS 39210,*
[2] *Nikhef, Science Park 105, 1098 XG Amsterdam, The Netherlands*

(Dated: August 14, 2013)

Bayesian inference algorithms have been adopted to perform parameter estimation and model selection for gravitational wave data from compact binary coalescence (CBC). However, their high precision has a high computational cost, which is especially apparent when analyzing signals from low mass systems. Here we show that parallelizing the Nested Sampling algorithm is a viable means of reducing the required computational time and discuss the optimal settings necessary to do so. We also produce an algorithm that variably resolves the frequency domain of an inspiral signal, reducing the time required to produce the waveform by an order of magnitude, and explore its applicability to the time domain. Both methods produced promising preliminary results that will lead to further optimization and implementation into the analysis software for advanced detector gravitational wave data.

I. INTRODUCTION

Mankind has observed the cosmos and sought to understand the underlying principles and mechanisms that govern the universe. Until now our observations have been of the electromagnetic spectrum, conventionally classified by names such as radio, microwaves, infrared light, visible light, ultraviolet light, x-rays and gamma rays. Each of these forms of electromagnetic radiation have led to the discovery of previously unknown phenomena with consequences reaching far beyond the physics community, such as the development of communication and medical technology. However each of these forms of radiation come from the same spectrum. A remaining consequence of General Relativity is expected to open a new window on the universe in the form of a completely separate specturm of radiation which will forever change the way in which it is observed−gravitational waves. The first gravitational wave detection will undoubtedly prove equal in historical significance to the ivention of the telescope and produce many unexpected results, furthering our understanding of the nature of the universe.

In the early twentieth century, physicist Albert Einstein developed the Theory of Relativity. Until then space, time, matter and energy were all considered separate entities. The theory suggests instead that space and time are manifestations of the four-dimensional reality in which we live known now as "spacetime" and that mass and energy are equivalent notions. Relativity arises from the principle that the speed of light is constant and concludes that the gravitational interactions we experience are not due to an independent force, but rather the curved geometry of spacetime. The geometry of spacetime, far from static, is capable of fluctuating in a manner similar to the surface of water. Just as ripples are produced when a stone is thrown into a pond, some physical systems are capable of sending ripples in spacetime across the universe, which we denote gravitational waves. Physicists from the Virgo and LIGO (Laser Interferometeric Gravitational Wave Observatory) collaborations have produced specialized Michelson interferometers designed to detect the path length difference of their two detector arms by using high-power lasers. The premise behind these massive rulers is that a passing gravitational wave will cause a fluctuation in the spacetime metric, which will change the length of one detector arm with respect to the other and produce recognizable patterns in the output. Currently, these devices are undergoing an upgrade to their second generation, "advanced" configuration, and it is expected that the resulting boost in sensitivity will enable physicists to measure the first ever gravitational wave signals and open this new window on the universe.

The first detected signals in ground based detectors are expected to come from coalescing compact binaries consisting of black holes (BBH), neutron stars (BNS) or both (NSBH). The reason for this is due to the strong quadrupole moment possessed by such systems, giving off more gravitational radiation than any other source. The detection rate for such interactions with second generation detectors is estimated at 40 events per year; however, the possible range is from 0.04 to 400 per year [2]. As the distances over which we may observe gravitational radiation increase, these numbers will grow. Also as the detectors improve, it is expected that we will eventually be capable of observing burst signals from supernovae, various stellar dynamical processes, and gain access to new standard candles for measuring cosmological distances more accurately. Initial LIGO and Virgo detectors were capable of measuring a signal in the visible band for approximately 30 seconds, while the advanced configurations will allow a signal to be visible for over 3 minutes. This is due to the detection of lower frequencies through improved detector technology, thus increasing the chirp time [9] of the wave. The subsequent data available from longer waveforms will allow for more accurate parameter estimation. However, the added cost associated with the increased volume of data is the increase in time required

to analyze it.

The framework currently used to analyze the data from gravitational wave signals is known as the LIGO Algorithm Library suite, or LALsuite, which houses Bayesian inference software called LALInference for signal analysis among other things. Bayesian inference is a mathematically straightforward approach to performing data analysis and has been implemented to determine the viability of various hypotheses and perform parameter estimation by returning probability density functions representing the likelihood of various values. Examples of parameters in gravitational wave astronomy include sky position, masses, distance to the source, and the magnitude and direction of the associated spin for each object in the system. For every unique combination of these and other parameters, a unique gravitational wave can be produced. The downside to Bayesian inference is the expensive computational cost of integrating over a high-dimensional parameter space. To address this problem in gravitational wave analysis, a method known as Nested Sampling [7] has been implemented. It takes as inputs the likelihood and prior density functions, maps the parameter space to one dimension, returns the evidence integral and samples from the posterior distribution using Markov-Chain Monte Carlo (MCMC) integration methods (see [8]). For an overview of this method, see [4], and for details on the current implementation of this procedure in gravitational wave detection, see [6].

Further boosts in the computational efficiency of LALInference implementation of Nested Sampling could improve the rate at which analysis of detector data is performed. This paper investigates two separate efficiency boosts to the LALInference software currently used by the Compact Binary Coalescence (CBC) group and data analysis of inspiral signals. First, we investigated an optimal parallelization of the implementation of the nested sampling algorithm and also confirmed the accuracy of the results given further parallelization. We also achieved an increase in efficiency through the variable resolution of the frequency domain. Current algorithms sample the frequency domain waveform at a regular frequency, meaning that at the high-frequencies, the waveform is oversampled. We present an algorithm to break up the frequency domain of waveforms and sample each broken interval at its Nyquist time to decrease the number of points to be processed. We also briefly discuss the theory, expected efficiency boost and potential sources of error of applying a similar approach to the time domain.

The paper is organized as follows: in Section II we provide a review of concepts pertinent to gravitational wave physics and binary inspiral signals; in Section III we describe Bayesian inference and how it is used to extract and analyze data; Section IV describes the Nested Sampling algorithm and elaborates on its implementation in LALInference; Section V presents the procedure and results of the parallelization performed; and Section VI contains a detailed description of the variable resolution algorithm as well as some preliminary results.

II. GRAVITATIONAL THEORY

A. General Relativity

General Relativity is a metric theory of gravity of that expresses gravitational interactions through the geometric curvature of spacetime. In flat, Minkowski spacetime, we can denote a line element representing the displacement between two events in spacetime by the following

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2. \tag{1}$$

This can be generalized to systems of arbitrary curvature described by any particular coordinate system through the introduction of the metric tensor, $g_{\mu\nu}$. The components of this tensor define the relationship between ds^2 and the coordinate directions dx^μ and dx^ν:

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu. \tag{2}$$

From this relationship, it is clear that the metric tensor defines lengths in arbitrary coordinate systems.

General Relativity relates the stress-energy tensor $T_{\mu\nu}$ to the Einstein tensor, $G_{\mu\nu}$, and thus the lengths measured in an arbitrary coordinate system, through a system of second order differential equations called the Einstein field equations, which are

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G_N}{c^4}T_{\mu\nu}, \tag{3}$$

where the Ricci tensor, $R_{\mu\nu}$, and the Ricci scalar, R, are contractions of the Riemann tensor. From this we can see that the Einstein tensor contains information about the curvature of a system as well as the lengths described by the metric tensor. Therefore, the physical significance of the field equations is that they connect the stress-energy density to the curvature of spacetime produced by an object. In short, the more massive an object, the greater curvature it will produce and the stronger a gravitational interaction it will have with other objects.

B. Gravitational Radiation

The Einstein field equations can be linearized by considering a metric perturbation, $h_{\mu\nu}$, on a Minkowski metric

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu},\ h_{\mu\nu} << 1. \tag{4}$$

Writing the linearized Einstein field equations in terms of the d'Alembertian operator, given by $\Box \equiv \partial_\mu\partial^\mu$, we find

$$\Box\bar{h}_{\mu\nu}+\eta_{\mu\nu}\partial^\sigma\partial^\rho\bar{h}_{\rho\sigma}-\partial^\rho\partial_\nu\bar{h}_{\mu\rho}-\partial^\rho\partial_\mu\bar{h}_{\nu\rho} = \frac{-16\pi G_N}{c^4}T_{\mu\nu}. \tag{5}$$

From here, it is convenient to use harmonic gauge such that $\partial^\nu\bar{h}_{\mu\nu} = 0$, which simplifies the expression above to the form

$$\Box\bar{h}_{\mu\nu} = \frac{-16\pi G_N}{c^4}T_{\mu\nu}. \tag{6}$$

Outside the source where $T_{\mu\nu} = 0$, the field equations simplify even further to

$$\Box \bar{h}_{\mu\nu} = 0, \tag{7}$$

which is immediately recognized as the wave equation

$$\left(-\frac{1}{c^2}\frac{\partial}{\partial t^2} + \nabla^2\right)\bar{h}_{\mu\nu} = 0. \tag{8}$$

Therefore, we see that Einstein's field equations give rise to waves which travel at the speed of light, and it is this wavelike nature that gives rise to the notion of gravitational waves.

1. CBC Inspiral Signals

Binary coalescence is theorized to be the first astrophysical source observed by advanced configuration gravitational wave detectors, and for this reason there is an enormous effort in the field of General Relativity to better understand their theoretical behavior. In this section, we focus on deriving the form of a binary inspiral signal from the Newtonian order approximation and present a formula for the frequency of a gravitational wave as a function of time for later use.

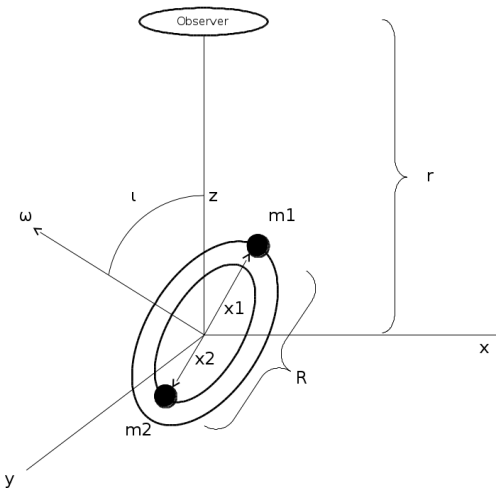

FIG. 1: A circularly orbiting binary system

Observing the system described in figure 1, we see that it consists of two masses, m_1, and m_2, separated by a distance R, orbiting their common center of mass in the counterclockwise direction. The orbital plane intersects the x-axis, making an angle ι with the observer's line of sight along the z-axis. Additionally, we see that the objects are not rotating themselves, leaving us with a total of nine parameters. Spinning objects would require 3 parameters each to describe their rotations, namely a magnitude and two angles describing orintation, raising the total number of parameters in such cases to 15. The positions of the masses as a function of time, denoted $\vec{x}_1(t)$, $\vec{x}_2(t)$, are given by

$$\vec{x}_1 = \frac{m_2}{m_1+m_2}R\hat{e}(t) = \frac{\mu}{m_1}R\hat{e}(t) \tag{9}$$

$$\vec{x}_2 = \frac{-m_1}{m_1+m_2}R\hat{e}(t) = \frac{\mu}{m_2}R\hat{e}(t) \tag{10}$$

where $\hat{e}(t) = (\cos(\omega t), \sin(\omega t)\cos(\iota), \sin(\omega t)\sin(\iota))$ and $\mu = m_1 m_2/(m_1+m_2)$ is the reduced mass.

The spatial tensor, representing the second moment of the mass distribution, or quadrupole moment, and can be written as

$$M^{ij} = \frac{1}{c^2}\int T^{00}(t,\vec{x})x^i x^j d^3x = \int \rho(t,\vec{x})x^i x^j d^3x, \tag{11}$$

where we rewrite T^{00}/c^2 as the mass distribution $\rho(t,\vec{x})$, which can be expressed by

$$\rho(t,\vec{x}) = m_1\delta^3\left(\vec{x} - \frac{\mu}{m_1}R\hat{e}(t)\right) + m_2\delta^3\left(\vec{x} - \frac{\mu}{m_2}R\hat{e}(t)\right). \tag{12}$$

Likewise, the strain amplitude of a gravitational wave measured in the approximately flat geometry far from the source [1] is

$$h_{ij} = \frac{2}{r}\frac{d^2M^{ij}(t)}{dt^2}. \tag{13}$$

Since the system system studied here is positioned such that the observer is along the z axis, the spatial components of the amplitude tensor are given by

$$h_{ij}^{TT} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{14}$$

where h_+ and $h_\times$ represent the two independent polarizations. In terms of the spatial tensor, the +-polarization, h_+, and $\times$-polarization M_{ij} are given by

$$h_+ = \frac{1}{r}\frac{G_N}{c^4}(\ddot{M}_{11} - \ddot{M}_{22}) \tag{15}$$

$$h_\times = \frac{2}{r}\frac{G_N}{c^4}\ddot{M}_{12}. \tag{16}$$

Therefore the relevant components of the spatial tensor are given by

$$M_{11} = \mu R^2\cos^2(\omega t) \tag{17}$$

$$M_{22} = \mu R^2\sin^2(\omega t)\cos^2(\iota) \tag{18}$$

$$M_{12} = \mu R^2\cos(\omega t)\sin(\omega t)\cos(\iota). \tag{19}$$

Substituting their second derivatives

$$\ddot{M}_{11} = -2\mu R^2\omega^2\sin(2\omega t)\cos(\iota) \tag{20}$$

$$\ddot{M}_{22} = -2\mu R^2\omega^2\cos(2\omega t) \tag{21}$$

$$\ddot{M}_{12} = 2\mu R^2\omega^2\cos(2\omega t)\cos^2(\iota) \tag{22}$$

into the polarization equations from equation 16, we find that the polarizations obtain the following form

$$h_+ = -\frac{4}{r}\frac{G_N}{c^4}\mu R^2\omega^2\cos(2\omega t)\frac{1+\cos^2(\iota)}{2} \quad (23)$$

$$h_\times = -\frac{4}{r}\frac{G_N}{c^4}\mu R^2\omega^2\cos(\iota)\sin(2\omega t). \quad (24)$$

By introducing the chirp mass, given by

$$\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}, \quad (25)$$

and defining $R = (G_N M_{tot})^{1/3}\omega^{-2/3}$ from the relationship between the Newtonian gravitational interaction and the angular acceleration of the system, we find that

$$h_+ = -\frac{4}{r}\frac{(G_N\mathcal{M}_c)^{5/3}}{c^4}\omega^{10/3}\cos(2\omega t)\frac{1+\cos^2(\iota)}{2} \quad (26)$$

$$h_\times = -\frac{4}{r}\frac{(G_N\mathcal{M}_c)^{5/3}}{c^4}\omega^{10/3}\cos(\iota)\sin(2\omega t). \quad (27)$$

From these polarizations, it is possible to use the Newtonian order approximation to derive the functional dependence of gravitational wave frequency on time and vice versa.

Due to the nature of the rotating system, the components of the quadrupole tensor complete a cycle every half a period. This means that the gravitational wave frequency is given by

$$f_{gw} = 2f_{orbit} = \frac{\omega_{orbit}}{\pi}. \quad (28)$$

We will define the characteristic radius, $R_c = 2G\mathcal{M}_c/c^2$, and wavelength $\lambda = c/f_{gw}$. Writing the polarizations in terms of R_c and λ, including an arbitrary phase factor, 2ϕ, and evaluating at the retarded time, t_{ret}, we have [3][1]

$$h_+ = \mathcal{A}\frac{1+\cos^2(\iota)}{2}\cos(2\pi f_{gw}t_{ret}+2\phi) \quad (29)$$

$$h_\times = \mathcal{A}\cos(\iota)\sin(2\pi f_{gw}t_{ret}+2\phi), \quad (30)$$

where

$$\mathcal{A} = \left(\frac{2\pi}{sqrt2}\right)^{2/3}\left(\frac{R_c}{r}\right)\left(\frac{R_c}{\lambda}\right)^{2/3} \quad (31)$$

An analysis of the radiated power from this system will lead toward the time-frequency functions we seek. First note that the power emitted per unit solid angle is given by

$$\frac{dP_{gw}}{d\Omega} = \frac{r^2c^3}{16\pi G_N}\left\langle \dot{h}_+^2 + \dot{h}_\times^2 \right\rangle. \quad (32)$$

Inserting the derived expressions for gravitational wave polarizations, we find

$$\frac{dP_{gw}}{d\Omega} = \frac{2}{\pi}\frac{c^5}{G_N}\left(\frac{G_N\mathcal{M}_c\pi f_{gw}}{c^3}\right)^{10/3} g(\iota), \quad (33)$$

where $g(\iota) = (1+\cos^2(\iota)/2)^2 + \cos^2(\iota)$. And by integrating over the sphere, we produce

$$P_{gw} = \int \frac{dP_{gw}}{d\Omega}d\Omega = \frac{32}{5}\frac{c^5}{G_N}\left(\frac{G\mathcal{M}_c\pi f_{gw}}{c^3}\right)^{10/3}. \quad (34)$$

Under the assumption that the binaries are on fixed orbital paths, the non-relativistic energy of the system is simply the sum of kinetic and potential energy, $E = E_k + E_p = -G_N m_1 m_2/2R$. The energy radiated by gravitational waves requires energy and will force the orbital radius to shrink as it is radiated. Noting that $\omega_{orbit}^2 = G_N M/R^3$, the R in the orbital energy expression may be eliminated

$$E_{orbit} = -\left(\frac{G_N^2\mathcal{M}_c^5\pi^2 f_{gw}^2}{8}\right)^{1/3}. \quad (35)$$

Therefore the loss in orbital energy per unit time is given by the energy flux

$$-\frac{dE_{orbit}}{dt} = P_{gw}, \quad (36)$$

which produces the following expression for the time evolution of gravitational wave frequency.

$$\dot{f}_{gw} = \frac{96}{5}\pi^{8/3}\left(\frac{G_N\mathcal{M}_c}{c^3}\right)^{5/3} f_{gw}^{11/3}. \quad (37)$$

Integrating this expression, we find that in terms of the time before coalescence, $\tau_{obs} = t_{coal} - t$, the gravitational wave frequency is given by

$$f_{gw}(\tau_{obs}) = \frac{1}{\pi}\left(\frac{5}{256}\frac{1}{\tau_{obs}}\right)^{\frac{3}{8}}\left(\frac{G\mathcal{M}_c}{c^3}\right)^{\frac{-5}{8}}, \quad (38)$$

For implementation in the frequency domain, however, we require the inverse of this function, $t_{obs}(f_{gw})$, which is:

$$t_{obs}(f_{gw}) = \frac{5}{256}\left((\pi f_{gw})\left(\frac{G\mathcal{M}_c}{c^3}\right)^{\frac{5}{8}}\right)^{\frac{8}{3}}. \quad (39)$$

These will be utilized later to determine the optimal sampling frequency for systems of arbitrary mass given by the Nyquist time and Nyquist frequency in the frequency and time domains respectively. A plot of equation 38 for given mass pairs are shown in figure 2.

III. BAYESIAN INFERENCE

Data analysis viewed as statistical inference consists of two closely related goals: model selection and parameter estimation. The former seeks to decide which available model best matches collected data, while the latter focuses on the determination of the governing parameters on which a model depends. Bayesian inference offers

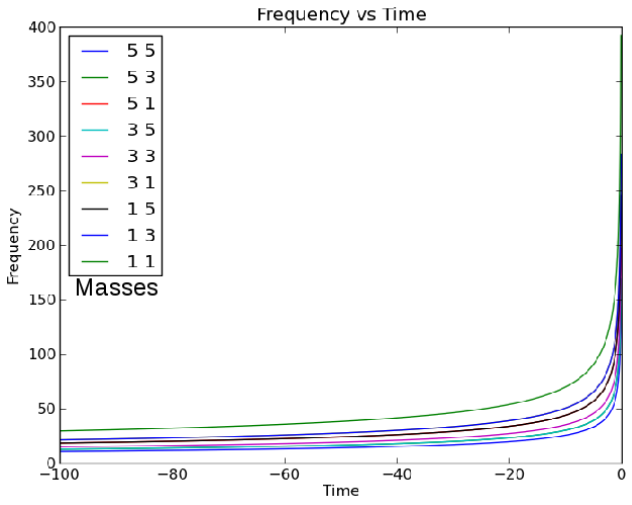

FIG. 2: The frequency dependence on time for various pairs of masses in units of solar mass. The second is a log-log scaled plot of the inverse function. The binary system described by these plotted functions evolves slowly at early times, but the evolution accelerates dramatically as the system approaches ISCO (Innermost Stable Circular Orbit).

a straightforward method to achieve both by exploring each system parameter and hypothesized model as separate dimensions in the parameter space and returning a probability density function (PDF) associated with each. In the case of gravitational wave signals produced by coalescing compact binaries, the analysis of gravitational wave signals with Bayesian inference requires the exploration of a minimum of nine dimensional parameters that is performed at an exceedingly high computational cost. However, the cost is driven down over time by new algorithms and faster computers, increasing the practicality of the application of the Bayesian framework in a growing number of fields.

To derive Bayes theorem, the principle underlying all of Bayesian inference, consider a set of hypotheses denoted

$$H = \{\mathcal{H}_i | i = 1, ..., N\}, \tag{40}$$

a vector of collected data, $\vec{d}$, and prior information, I. From the concept of conditional probability and by the law of total probability, we may construct Bayes theorem, which states:

$$P(\mathcal{H}_i|\vec{d}, I) = \frac{P(\mathcal{H}_i|I)P(\vec{d}|\mathcal{H}_i, I)}{P(\vec{d}|I)}, \tag{41}$$

where we read "$P(\mathcal{H}_i|\vec{d})$" as the probability of hypothesis "i" given collected data. By convention $P(\mathcal{H}_i|I)$ is called the prior probability, which encodes our confidence in a hypothesis. The likelihood function, $P(\vec{d}|\mathcal{H}_i, I)$, denotes how probable the data is given what we suspect theoretically, and the $P(\vec{d}|I)$ is called the evidence, which represents a sum over each hypothesis

$$P(\vec{d}|I) = \sum_i P(\vec{d}|\mathcal{H}_i, I), \tag{42}$$

and behaves as a normalization factor. The posterior probability on the left side of equation 41 may therefore be calculated by comparing the prior and likelihood distributions against the evidence.

A. Model Selection

The Bayesian framework permits the comparison of the viability of various hypotheses in the form of an odds ratio. The procedure to determine this ratio is denoted model selection or hypothesis testing. Consider Bayes theorem applied to the posterior distributions of two separate hypotheses, $\mathcal{H}_i$, and $\mathcal{H}_j$. Comparing these two posterior PDFs, we find

$$\frac{P(\mathcal{H}_i|\vec{d}, I)}{P(\mathcal{H}_j|\vec{d}, I)} = \frac{P(\mathcal{H}_i|I)P(\vec{d}|\mathcal{H}_i, I)}{P(\mathcal{H}_j|I)P(\vec{d}|\mathcal{H}_j, I)} = \frac{P(\mathcal{H}_i|I)}{P(\mathcal{H}_j|I)}\mathbf{B}_{ij}, \tag{43}$$

where $\mathbf{B}_{ij}$ is called the Bayes factor and the ratio of priors preceeding this term is typically set to unity unless there is reason to favor a particular hypothesis. The more the data supports the $\mathcal{H}_i$, the larger the Bayes factor and vice versa for $\mathcal{H}_j$. Convention suggests that values over 100 represent definitive of the relative validity of one hypothesis versus another.

The computation of the Bayes factor is simple in the case of hypotheses with no free parameters; however, in the case of gravitational wave analysis, the set of variable parameters makes this computation exceedingly difficult. To determine the Bayes factor in such a situation, the likelihood must be marginalized over each of the parameters weighted by the prior, which produces what is known as the evidence, Z, given by

$$Z = P(\vec{d}|\mathcal{H}_i, I) = \int_{\vec{\theta}\in\Theta} p(\vec{d}|\vec{\theta}, \mathcal{H}_i, I)p(\vec{\theta}|\mathcal{H}_i, I)d\vec{\theta} \tag{44}$$

In all but the most trivial cases, this integral is exceedingly difficult to compute due to the high dimensionality of the parameter space and the large intervals of values that can be taken by the parameters themselves. This computational challenge has been addressed in a paper by Veitch and Vecchio, and further boosts to their methods are described here.

B. Parameter Estimation

As mentioned in our derivation of equations 38 and 39, parameters associated with producing a unique gravitational waveform are the two masses, time, sky position, distance, phase of coalescence and three orientation

angles, which we express as part of a nine dimensional parameter space

$$\Theta = \{\mathcal{M}, \nu, t_0, \phi_0, D_L, \alpha, \delta, \psi, \iota\}. \tag{45}$$

Other parameters such as the 6 spin components of a [10] binary system exist as we remove some of our simplifying assumptions. The maximum number of parameters for a coalescing compact binary is therefore 15, and though finite, it nonetheless presents a computationally costly challenge.

To determine the distribution of each of these parameters, a procedure known as marginalization is performed. In the case of parameter estimation, the process begins by considering a subset of parameters, say $\vec{\theta_A}$ in the parameter space Θ. The marginalized distribution for one parameter may be calculated by "integrating out" the other parameters

$$p(\vec{\theta}_A|\vec{d}, \mathcal{H}, I) = \int_{\Theta_B} p(\vec{\theta}_A|\vec{d}, \mathcal{H}, I) d\vec{\theta}_B. \tag{46}$$

Then the expectation of a parameter may be determined by computing the weighted average over its marginalized distribution

$$\langle \vec{\theta}_A \rangle = \int_{\Theta_A} \vec{\theta}_A p(\vec{\theta}_A|\vec{d}, \mathcal{H}, I) d\vec{\theta}_A, \tag{47}$$

and the variance of the data is described by

$$\sigma^2 = \langle \vec{\theta}_A^{\,2} \rangle - \langle \vec{\theta}_A \rangle^2. \tag{48}$$

IV. NESTED SAMPLING

In 2004, Skilling developed the Nested Sampling algorithm [7], which represented a novel shift the conventional calculations of Bayesian inference. It it capable of performing both model selection and parameter estimation by taking as inputs the likelihood and prior distributions, sampling the parameter space using Markov-Chain Monte Carlo (MCMC) methods, and returning the computed evidence integral and samples from the posterior distribution. We begin our explanation of the algorithm by multiplying either side of bayes theorem (equation **??**) by the evidence,$P(\vec{d}|I)$, and establishing for each distribution the following nomenclature:

$$\begin{aligned} P(\vec{d}|\theta, I)P(\theta|I) &= P(\vec{d}|I)P(\theta|\vec{d}, I) \\ \text{Likelihood} \times \text{Prior} &= \text{Evidence} \times \text{Posterior} \\ L(\Theta) \times \pi(\Theta) &= \text{Z} \times P(\Theta). \end{aligned} \tag{49}$$

A. The Procedure

The algorithm begins by mapping the parameter space to a one dimensional line. To perform this mapping, the program simply extracts the likelihood of each value in the prior and orders these values from greatest likelihood to least over the interval $(0, 1)$. Next N_{live} samples, or live points in the conventional nested sampling jargon, are drawn uniformly from the prior and their likelihoods are calculated.

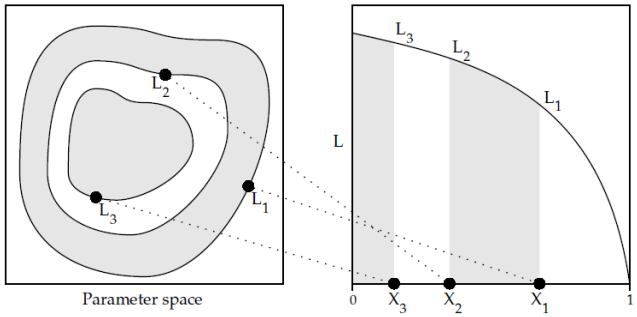

FIG. 3: This diagram depicts samples drawn from Area(Z) which are equivalent to those drawn from the posterior distribution.

The next step is for the algorithm to remove the live point farthest from 0, though it is stored along with its likelihood for later use. In its place another is resampled uniformly over the interval from 0 to the location of the removed point. This process is repeated until the sampled region has shrunk to very near 0 where the likelihoods are very large. Each contour of equal likelihood shrinks by a constant factor so that $\log(X)_{i+1} = \log(X)_i - 1/N_{\text{live}}$, so for fewer live points, an instance of this procedure will take less time than when utilizing more.

Once the algorithm explores the parameter space sufficiently by resampling many times and reaching a termination condition specified by the programmer, the algorithm calculates the evidence integral

$$Z = \int_{\Theta} L(\Theta)\pi(\Theta) d\Theta. \tag{50}$$

by defining the prior mass

$$X(\lambda) = \int_{L(\Theta)>\lambda} \pi(\Theta) d\Theta, \tag{51}$$

and performing the summation

$$Z = \int_0^1 L(X) dX \tag{52}$$

$$\approx \sum_{i=1}^{N} L(X_i)\Delta X_i \quad , \tag{53}$$

where N is the number of samples collected and $L(X(\lambda)) = \lambda$.

Next the area beneath Z is sampled which produces samples from $P(x) = L(x)/Z$ and thus from the posterior $P(\vec{x}|\vec{d}, I)$. By simultaneously sampling the posterior distribution and calculating the evidence integral, nested

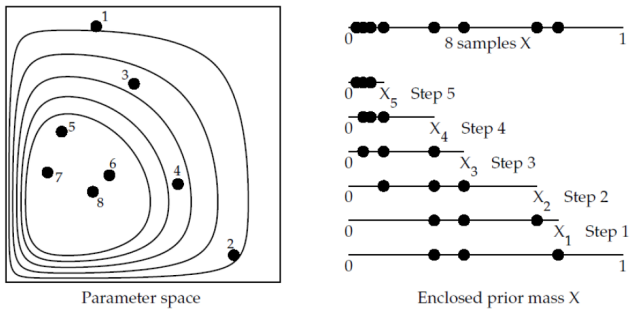

FIG. 4: A depiction of the nature of the repeated sampling and movement of the samples to regions of increasing likelihood.

sampling is highly efficient, greatly reducing the computational overhead necessary in the exploration of high dimensional parameter spaces. However, there remains a great deal to be desired in terms of the computational time , as data analyses of gravitational wave systems can take two months or more.

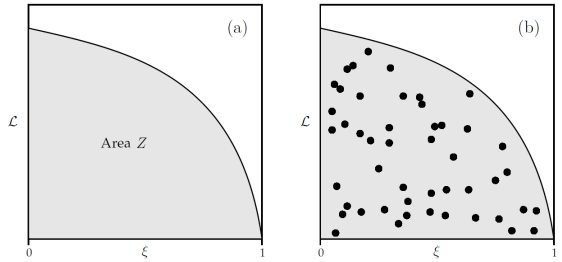

FIG. 5: A representation of the evidence integral and how the posterior is sampled by sampling the area below the calculated evidence integral at a negligible additional computational cost..

V. GAINING EFFICIENCY THROUGH PARALLELIZATION

For a given number of live points, the nested sampling procedure will converge on a maximum likelihood by a time dependent on the rate at which the contours of equal likelihood shrink. This means a computation performed with a lower number of live points, though less accurate, will be faster than for a larger number. We intend to reduce the number of iterations necessary to complete the algorithm by running multiple instances of the algorithm in parallel with a lower number of live points and analyze the effect of this reduction. The multiple runs will encourage statistical accuracy while reducing the overall computational cost. Our goal is to optimize this procedure while maintaining a sufficient level accuracy by finding the most effective numbers of live points to utilize.

A. Method

We first utilized a random number generator which provided each run with a different random seed to induce differences between each output. An integer with a large number of factors was chosen arbitrarily to provide the upper end of the test. We ran one instance at 1024, two at 512, four at 256, and so on to a minimum number of 16 live points where the algorithm no longer produced sufficiently accurate results. Instances using factors of 1200 live points to a minimum of 16 and 256 to a minimum of 64 were also performed following the same scheme. Next, we organized the posterior samples from each parallel run by weighting and resampling within each run. Once collected, we merged each of the parallel runs by weighting each run according to their own evidence estimates and drew posterior samples from all runs by drawing from the collated weighted parallel posterior samples.

The post-processing procedure for the parallelized instances generated plots of the probability distributions of each parameter and the covariance matrix of the parameters in each run individually. It then merged the data from each instance using a common number of live points and ploted the data together for comparison. The posterior samples from each of the runs were then combined to generate a figure of merit for the relative accuracy given various numbers of N_{live}. The merged files were be compared against each other to determine which numbers of live points required the least computational time. Once generated, the efficiency and accuracy figures of merit were compared, and an optimal range of live points determined.

B. Results

Observing the relative rates at which the nested sampling algorithm converges on the most likely parameter values produces a glempse of the efficiency boost gained by decreasing the number of live points. Figures 6 and 7 show how the reduction in live points more sparsely samples the parameter space and generate a mental picture of how nested sampling discovers and explores regions of increasing likelihood. The parameter shown here is the chirp mass, which is highly sensitive to changes in the waveform generation inputs. Besides the pictorial representation, these images provide a quantitative into the reduction in computational cost. Noting the horizontal axes which represent the number of iterations undergone by the algorithm up to that sample, computations with lower numbers of live points do indeed require far fewer computations and thus take less time.

The initial parallelization containing 1024 live points and factors thereof produced the following curve describing the number of posterior samples versus the number of live points per run. This plot also includes a more highly resolved segment where we suspected the algorithm to perform most effificnely. This is explained in the stas-

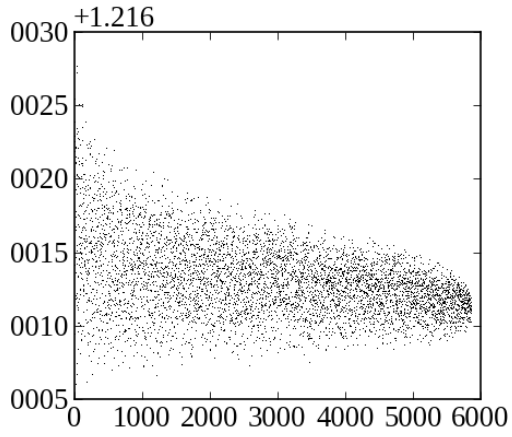

FIG. 6: Samples drawn during nested sampling instance run with 1200 live points. Here the x-axis denotes the number of iterations performed by the algorithm, while the y-axis describes the parameter estimate.

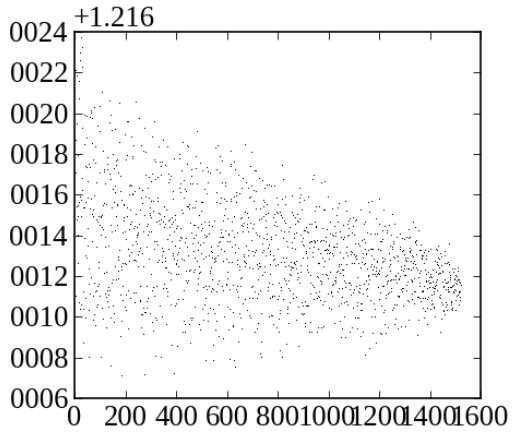

FIG. 7: Samples drawn during nested sampling instance run with 300 live points. Here the x-axis denotes the number of iterations performed by the algorithm, while the y-axis describes the parameter estimate.

tical variance of the time to perform the likelihood and thus maintains the increasing trend as N_{live} grows.

From the sampling, is clear that an efficiency boost can be generated by cutting the number of live points. However, such a procedure will not produce viable results when run with fewer than 64 live points. This is seen in the drastic decrease in posterior samples, and thus accuracy, with a relatively small gain in efficiency. This is in part because the amount of time to complete a run with half the number of live points does not correspond to half the time to complete that of the larger number. Instead instances run with fewer live points require roughly 75% of the computational time of the instances containing twice as many live points, which is seen in figure 8. Since accuracy is lost as the number of live points is decreased, the standard deviation of our parameter esti-

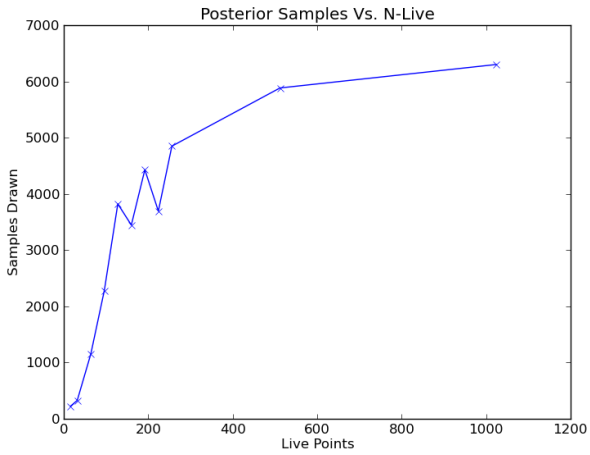

FIG. 8: A plot of posterior samples versus the number of live points used to perform the nested sampling algorithm. For each 50% reduction in N_{live}, approximately 75% of the posterior samples are retained.

mates also increases as can be observed in figures 11, 12, 9, and 10. At a sufficiently small number of live points, accurate physics can no longer be performed.

Continuing to analyze the accuracy of the parameter estimates as we decrease the number of live points, it is beneficial to analyze the distributions of each of the merged parallel runs. Similarity between the histograms from figures 9 and 10 shows that parallelization maintained a significant amount of accuracy in the waveform parameters estimated by the algorithm. Figures 9 and 10 depict the probability distributions for the values of the chirp mass. It was expected that as this number diminished, the variance would increase dramatically; however that was not observed. Though the spread of the data for relatively small values was larger than for large numbers of live points, the rate at which the spread increased was slower than expected.

Figures 11 12 depict the log Bayes factors calculated by the parallel instances containing the same numbers of live points. The range of the log Bayes factor and the accuracy of parameter estimates are highly correlated. At lower numbers of live points, the variance of the log Bayes factors becomes too large, meaning the algorithm will not converge on as consistent a quantity. However, these histograms show that there is a range inside 512 and 128 which produced sufficient accuracy to extract physics from gravitational wave signals.

Comparing the cumulative distributions of the chirp mass calculated with varying numbers of live points, we produced figure 13 and 14, the latter of which is a range of live points within the larger range. As is visible in the plots, there is little difference in the distributions calculated for different numbers of live points. Each run is slightly different; however, merging all of the output files for common numbers of live points, we see that the

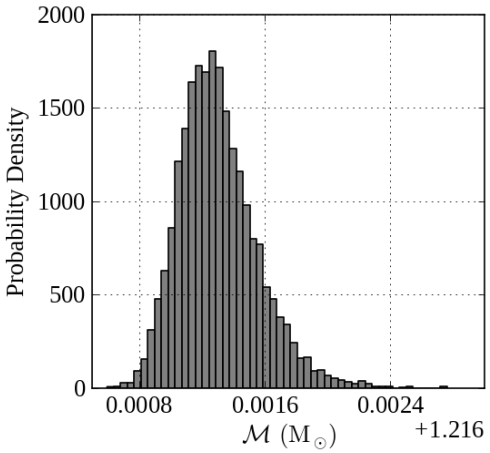

FIG. 9: A probability density function for chirp mass parameter calculated using 1200 live points.

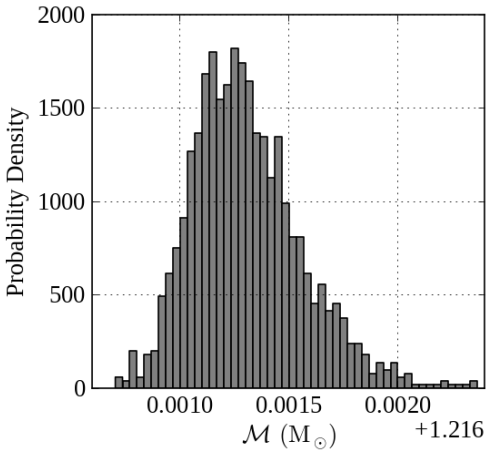

FIG. 10: A probability density function for chirp mass parameter calculated using 300 live points.

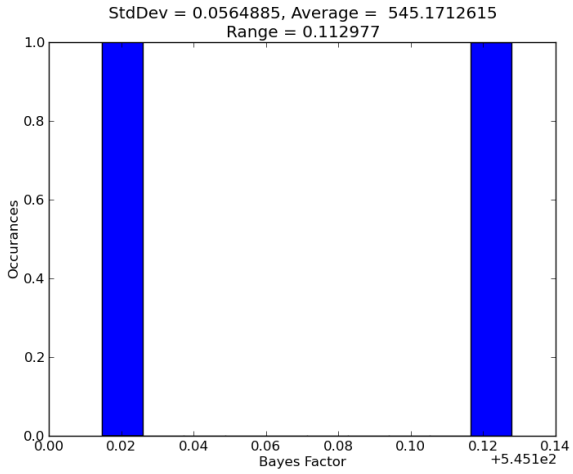

FIG. 11: A histogram of the 2 Instances with 512 Live Points. Here the log Bayes factor was 545.17 ± 0.06.

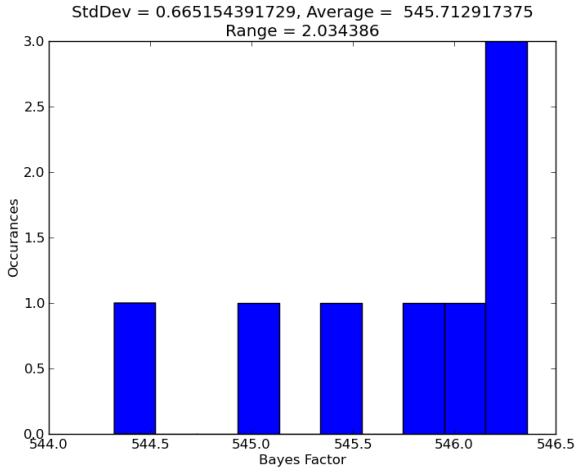

FIG. 12: A histogram of the 8 Instances with 128 Live Points. Here the log Bayes factor was 545.71 ± 0.67.

average of each has a nice overlap. This is precisely what we desired when the problem was defined.

Figure 15 describes the computational time versus the number of live points for this set of runs. There is a clear trend seen here; however, the greatest improvement relative to the preceding higher number of live points was found in the transition between 300 and 200. Based on the number of posterior samples and the efficiency boost in computational time, the optimal range of live point numbers falls between 250 and 300. Here it was determined that we produced the fastest result while maintaining sufficient accuracy to produce viable parameter estimates and allow us to maintain confidence in future applicatins of this approach.

VI. VARIABLE RESOLUTION OF THE FREQUENCY DOMAIN

At the most basic level, any attempt to better resolve of a set of data will cost computational time, but produce improved accuracy. The opposite is true for lower resolutions. There is a limit to how much or little we may resolve a function, given by the Nyquist frequency. The ideal resolution is achieved when a function is sampled at this rate, whereas sampling above this rate will produce no improvement in accuracy and a loss in efficiency and below it will produce effects known as aliasing where a waveform can no longer be uniquely determined.

When we refer to sampling in this section, we mean the variety performed in the reconstruction of a bandlimited signal given by the Sampling Theorem. The Nyquist frequency, coming directly from the this theo-

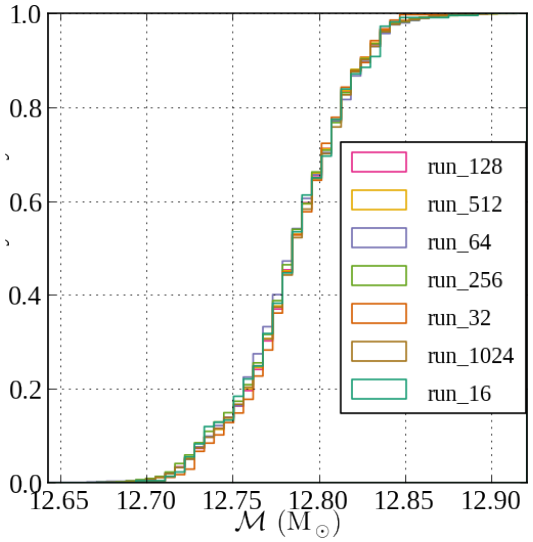

FIG. 13: Cumulative distribution for chirp mass parameter given merged parallel instances from a large range of live points

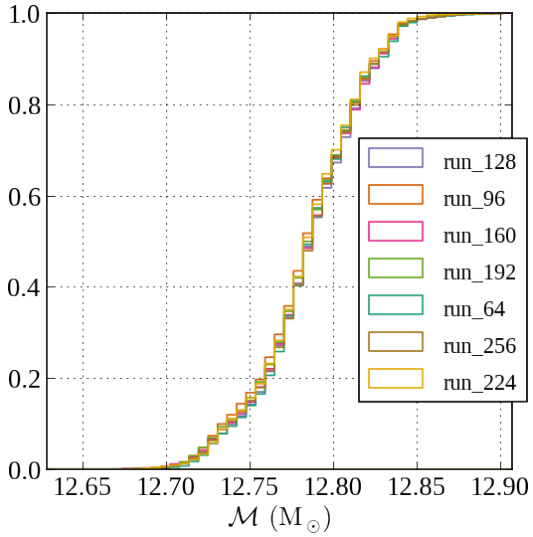

FIG. 14: Cumulative distribution for chirp mass parameter given merged parallel instances from a small range of live points

rem, states that a signal, $t(f)$, bandlimited by some frequency, F, may be completely determined by a series of its ordinates spaced at a frequency of $F/2$ Hz. The rate at which these ordinates occur is called the sampling frequency, and the exact frequency at which a signal is completely determined is called the Nyquist frequency. A subsequent result of this theorem is that a frequency domain waveform containing no amplitudes greater than T is completely determined by giving its ordinates at a series of abscissas spaced $1/(2T) = f_{\text{Nyquist}}$ Hz apart. Since this is a frequency domain waveform, and we have used the inverse of the function $t(f)$, we could think of this sampling frequency as the Nyquist time instead of

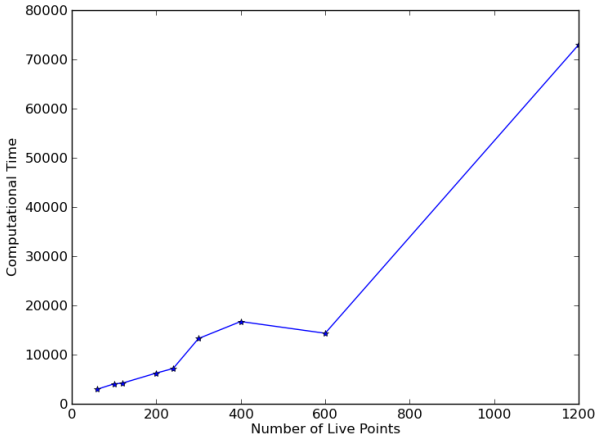

FIG. 15: A plot of the total computational time at various numbers of live points. There is an enormous redution produced by the first reduction of 50%. Further reductions produce relatively smaller reduction in computational time.

frequency; however, we have opted to utilize a notation mostly consistent signal processing conventions (see [5]).

The current algorithm utilizes constant resolution function stationed well above the Nyquist frequency for all but the lowest frequency. Instead, using a function which returns the Nyquist frequency for each point on the waveform we wish to generate and smapling at this rate would be more efficient. Our goal is thus to exploit the monotonically decreasing nature of the Nyquist frequency of an inspiral signal and optimize this aspect of waveform generation. This will result in sampling at a high rate at low frequencies and down-sampling gradually as we reach higher frequencies.

A. Method

The first step toward optimizing the resolution of the frequency domain waveform through variable resolution is to determine the Nyquist frequency at every point in a theoretical frequency domain waveform. We do this by introducing the functional dependence of gravitational wave frequency on time observed which is given by equation 38. Given particular systems of BNS, BBH and NSBH, we may establish scenarios for different mass ratios which result in a different time dependence on frequency and vice versa. The largest amplitude for this function occurs when the smallest chirp mass is implemented, which for the sake of gravitational wave analysis is a binary system of neutron stars, each weighing $1M_\odot$. Such a system establishes a worst-case scenario for sampling rates and was used in our subsequent calculations. The plot shown in figure 39 describe the functional dependence of time on frequency and vice versa given nine different mass combinations listed in solar masses. By

using this plotted equation we may determine the corresponding Nyquist sampling frequency by calculating

$$f_{nyq} = \frac{1}{2}\frac{1}{t_{obs}(f_{gw})}. \tag{54}$$

and thereby optimize the sampling rate for arbitrary mass CBC systems. Graphically this is shown in figure 16, where we can see that at each frequency at which we change sampling rates, we may simply take twice the value of the function, the inverse of which is the ideal sampling rate.

We have chosen to break frequency domain waveforms according to an optimization that minimizes the number of samples as a function of the frequencies at which the function is broken. This means that for any chosen number of breaks, $f_1, f_2, ..., f_M$, we may find the optimal frequencies at which to change the sampling rate and perform calculations with optimal efficiency. For this number of breaks, the function of the number of samples we will minimize is given by

$$\begin{aligned} N(f_1, f_2, ..., f_M) \;=\; & \frac{f_1 - f_{min}}{f_{nyq}(f_{min})} + \frac{f_1 - f_2}{f_{nyq}(f_1)} + ... \\ & + \frac{f_N - f_{N-1}}{t_{nyq}(f_{N-1})} + \frac{f_{max} - f_M}{f_{nyq}(f_M)}, \end{aligned}$$

where $_{nyq} = \Delta f_{nyq} = 1/(2\tau(f_{gw}) + 1)$ is the nyquist frequency given f_{gw} and utilizes equation 39. In words the length of each band divided by the rate at which it is sampled, produces the number of samples for each band. Summing this over each band produces the number of samples required to construct the waveform, so the minimum of this function returns the minimum number of samples required.

In the continuous case and with arbitrary computing power, it would be advisable to make an enormous number of breaks in the frequency domain so as to always sample waveform at an optimal rate. However, with computational limitations, it is more efficient to break a relatively small number of times because the time required to recompose the bands becomes a significant as the number of breaks increases.

Next we recompose the waveform by placing the bands side by side. This is achievable due to the analytic nature of the frequency domain, theoretically guaranteed to produce a perfect match, though this was not observed in our results. Here we have utilized linear interpolation to produce a sufficient match against a waveform sampled at a constant rate. Alternatives such as sinc and spline interpolation methods were considered; however, the nature of this problem led naturally to a linear procedure. The fully composed waveform was then compared against one at a constant sampling rate. Results of this implementation follow.

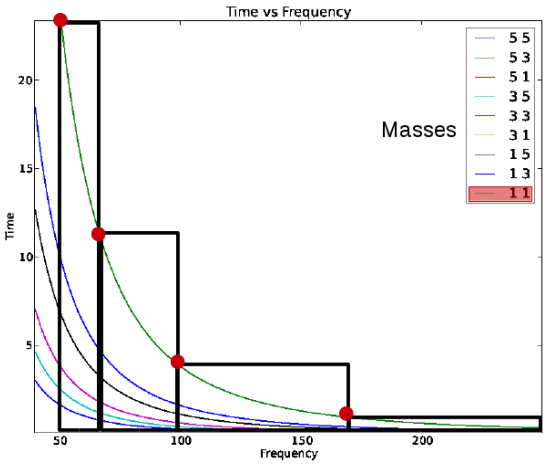

FIG. 16: An example of how a frequency domain waveform is broken and sampled. The location of each break is determined by the minimization of the number of samples as a function of frequency breaks 1 through 5 shown in equation 55, and the sampling rate for each band is determined by the nyquist frequency given by the inverse of twice the maximum amplitude of equation 39.

B. Results

Using linear interpolation between the reference waveform at a constant sampling rate and our broken waveforms with various numbers of bands, the following waveform matches were calculated. Note that the one band case is not a perfect match. This is due to error induced by the interpolation method itself, which propagates through the other matches as well. Despite this small error, the matches remain excellent for the purposes of signal analysis.

Bands	Percent Match	Time (d+h:m:s)
1	99.9997	$\approx 4 + 09:00:00$
2	99.9541	$3 + 16:28:54$
3	99.7589	$2 + 23:22:12$
4	99.5351	$2 + 18:19:39$
5	99.3865	$2 + 17:30:19$

FIG. 17: Preliminary results showing the greatly diminished computational time at little cost to accuracy

In terms of computational efficiency gain, it is clear that variable resolution will prove useful in future gravitational wave analysis. The results shown here suggest that a frequency domain broken into 5 bands will produce a nearly 50% reduction in computational time. Since these results are for only short wavefoms, it is expected that this reduction will be further amplified as the waveform length increases. A beneficial test for the future would be to test a waveform of the same length we expect our first detection to be. This would allow us to

understand just how long a detection will take to analyze and ultimately publicize.

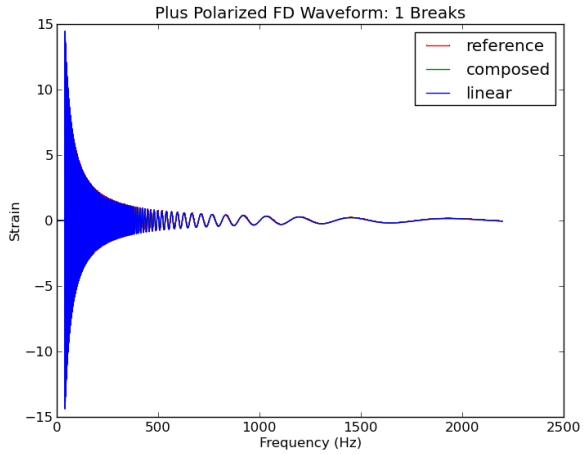

FIG. 18: Three overlayed frequency domain waveforms. One is sampled at a constant rate, another is composed without interpolation, and another uses linear interpolation to reconnect each of the bands. The linear interpolated match is seen in VI B.

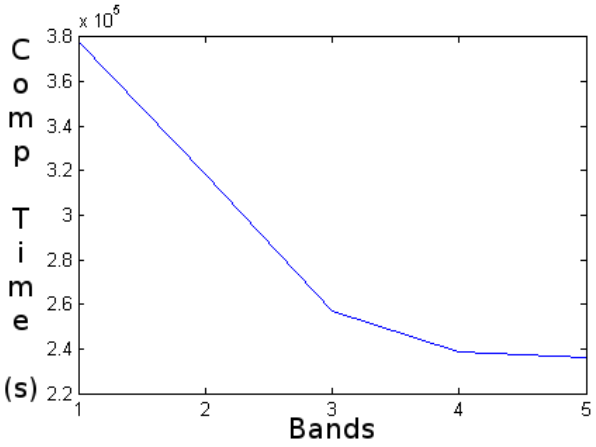

FIG. 19: Time required to generate the frequency domain waveform

VII. TOWARDS A TIME DOMAIN IMPLENTATION

A supplementary goal is to apply the variable resolution approach to the time domain. A large portion of the computational time required to complete a nested sampling instance is invested in the generation of waveforms. The most time consuming waves to generate are time domain waveforms because they are not analytic. They present a greater computational challenge because they require numerically solving a system of highly-coupled second-order differential equations and require the matching the broken waveforms in a consinuous and differentiable manner in order to reduce inconsistencies brought about by transforming between the time and frequency domains. While the former explains why the matches are imperfect, the latter expresses why this is a hurdle which must be overcome to produce accurate results.

In principle, the time domain waveform is very similar to the frequency domain except it is reversed and increases in amplitude from some initial time to ISCO. This means that the strategy initially implemented to break the frequency domain into multiple intervals must be reversed so the minimization of the new function $N(t)$ produces the optimal times at which the time domain should be broken from greatest to least. We may write this function as follows:

$$N(t_1, t_2, ..., t_N) = \frac{t_{ISCO} - t_N}{f_{nyq}(t_N)} + \frac{t_N - t_{N-1}}{f_{nyq}(t_{N-1})} + ... + \frac{t_2 - t_1}{f_{nyq}(t_2)} + \frac{t_1 - t_{min}}{f_{nyq}(t_1)}.$$

With the function $N(t_1, t_2, ..., t_N)$ established, it is easy to consider an implementation in the time domain using the same framework developed for the frequency domain. First we would minimize this function for the number of time domain breaks we desire, sample at the Nyquist frequency (instead of Nyquist time) for the largest time amplitude inside the band, and then compose the resulting pieces of the waveform.

However, the current code has the following limitation which will)have to be addressed before this concept is used in LALInference. The waveform generator produces waves from some initial frequency all the way to ISCO. This is not a problem, except that it results in a waste of computational resources when using a multi-band framework. Instead, we suggest the implementation of code which allows for a certain maximum frequency to be specified and the waveform calculated only for that region.

Once this is addressed, we expect there will be a problem matching the broken waveforms. The frequency domain is analytic, meaning that producing a match from the broken waveforms is relatively easy and, in fact, should be perfect except for interpolation generated errors. The time domain waveform, on the other hand, is

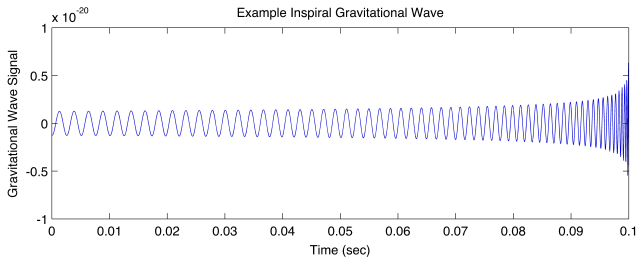

FIG. 20: An example time domain imspiral waveform.

found by solving a system of coupled differential equations. Changing boundary conditions such as the endpoints of the interval over which you seek a solution can cause changes in the endpoints. Even if only a slight difference between the end of one band and the beginning of the next is produced, this will produce what is known as Gibbs phenomenon when the function undergoes a fast Fourier transform. To reduce this effect, we suggest the eventual implementation of a convolution of a small overlap of adjacent waveform bands. This would smooth the function, thereby encouraging continuity and making the function analytic at points which would not otherwise be differentiable.

VIII. CONCLUSIONS

The initial goals set for this project were to analyze parallel instances of the nested sampling to obtain an optimal configuration and to study and implement a variable resolution algorithm for more efficient waveform generation. From the procedure desceibed here, we determined that the nested sampling algorithm could indeed be made more efficient by the implementation of parallel processing as well as by variable resolution of the frequency domain. There was also progress made toward the implementation of the algorithm in the time domain; however, those results will be saved for a future paper combining the results from this research and a more detailed analysis of the behavior of the multiple band approach in the time domain.

Regarding the parallelization performed, we determined that the optimal range of live points is between 250 to 300 with a minimum of 4 instances of nested sampling should be implemented with these values to perform meaningful statistics. There is too significant a loss in accuracy when the number is decreased below this value. Even if a large number of instances were run simultaneously, the results would be less meaningful than a smaller but more accurately computed group. Since the number of posterior samples drawn from runs containing 200 to 256 live points is approximately 50% of the number drawn from a run containing 1000 to 1024, this means that running, say, 4 instances with 250 live points will produce a 2.25% increase in the total number of posterior samples, meaning a more accurate result can be produced at a lower computatonal cost.

Similarly, results from the variable resolution algorithm indicate that it is possible to gain efficiency by breaking up the frequency domain waveform to perform analysis. The results shown above are only preliminary, as they utilize short waveforms for the sake of determining results quickly. To test the full potential of this approach, it will be necessary to test longer waveforms and also determine the effect of louder signals with higher signal to noise ratios. If the pattern found in table VI B remains consistant for longer and louder waveforms, it is highly likely that the algorithm will cut the time required to perform nested sampling in half. We feel that the development of a similar algorithm for the time domain would come full circle in boosting efficiency under by means of this framework.

A combination of parallelization and variable resolution would clearly be a beneficial application of these newly developed tools. A test of these in unison will help determie a new time frame within which we will be able to publish results from the first ever gravitational wave detection expected from the advanced-configuration detectors. With these procedures in hand, it is hoped that we will become closer to the goal time frame of a publication three months after the detection is made.

IX. ACKNOWLEDGEMENTS

The author would like to thank the gravitational physics group at NIKHEF for their encouragement, patience and advice throughout the research process. He would like to extend special thanks to his mentor John Veitch for his support and dedication. Additionally he would like to thank the University of Florida Department of Physics for the opportunity to perform this research via the NSF IREU program. It was a wonderful experience that will have benefits far beyond the scope the research performed and physics and computational methods learned.

[1] Gravitational Waves, Maggiore
[2] http://arxiv.org/pdf/1003.2480v2.pdf
[3] Schutz
[4] Data Analysis: A Bayesian Tutorial
[5] Communication in the Presence of Noise
[6] Bayesian Coherent Analysis of Advanced Detector Data
[7] http://www.inference.phy.cam.ac.uk/bayesys/nest.pdf
[8] http://web.mit.edu/ wingated/www/introductions/
[9] As a gravitational wave passes, the frequency and amplitude increase in the audible band, making a chirping noise, so the chirp time is in reference to the time the wave is within a detector's visible band.
[10] The spin of an object is described of a magnitude and two orientations, so for a binary system, a total of 6 parameters completely determine the spins.