Abstract: The Einstein Telescope (ET) is a proposed third generation gravitational wave interferometer with a 10 fold increase in sensitivity over advanced detectors and a detection band from 1 Hz to 10 kHz. For binary neutron star inspirals (NS-NS), population density studies suggest anywhere from $10^5$ to $10^6$ sources per year within ET’s detection horizon of $z \approx 1$. That, combined with an average chirp time of 6 days for the proposed detection band, suggests an average overlap of over 17,000 simultaneous signals. To test if current data analysis techniques can handle such significant numbers of signals, we design preliminary stages of a NS-NS inspiral data simulation similar to the LISA Mock Data Challenge. Using the restricted post-Newtonian waveform with a post-Newtonian phase out to 3.5, we generate two mock data sets using the proposed ET sensitivity curve with lower frequency cutoffs of 10 and 5 Hz, and the above population density. The details of the simulation will be discussed, as well as further improvements for the simulation.

I. INTRODUCTION

Gravitational Wave interferometry is a relatively new method for understanding the universe via the observation of propagating ripples in spacetime. By measuring the interference from two phase locked detector arms, the strain produced on the detector by an incident gravitational wave can be measured, providing an entirely new look into the state and structure of astrophysical phenomena. As gravitational waves are not absorbed or reflected, they can propagate without hindrance throughout the entire universe, yielding a view of objects previously unavailable to conventional electromagnetic astronomy [1].

There are currently four gravitational wave interferometers in operation: the two detectors of the Laser Interferometer Gravitational Observatory (LIGO), operating in Livingston, Louisiana and Hanford, Washington, the VIRGO detector, operating outside Pisa, Italy, the GEO 600 detector, operating outside Hannover, Germany, and the relatively new TAMA 300 detector, operating near Mitaka, Japan. These detectors are currently in their initial stage, with plans to upgrade LIGO and VIRGO to more advanced, second generation technology by the end of 2014. Once completed, Advanced LIGO and VIRGO will have sufficient range to detect several events per year, from binary neutron star inspirals (NS-NS) to gamma ray bursts, supernovae, black hole mergers and ring-downs, etc.

With advanced detectors already slated for upgrades, plans have begun on the third generation of gravitational wave observatories. These detectors will use cutting edge technology, combined with instrumentation and data analysis techniques learned from first and second generation detectors to push to sensitivity levels far beyond current planned detectors. One such detector, the Einstein Telescope (ET), is currently being investigated by the European Commission, for a possible completion date around 2020. Current design study targets have ET operating at 10 times the sensitivity of advanced generation detectors, with a detector bandwidth of 1Hz to 10kHz. This massive increase in the number of observable sources promises to significantly push forward current astrophysical and cosmological understanding; however, such a powerful increase in detector range brings with it a huge number of overlapping signals. Given current predictions of source population and the planned ET sensitivity curve [4, 5], signals for (NS-NS) inspirals could remain in the detector for up to 9 days at a time, causing a signal overlap of over 17,000 inspiral signals at any given time.

In this paper, we describe the initial stages of an ET Mock Data Challenge, designed to simulate the expected data produced by a third wave gravitational detector with such an increased detection horizon. Using well known Post-Newtonian waveforms, a simulation was written to generate inspiral data for the region of space ET is expected to operate in. The operation of the program is discussed, as well as plans for future improvement and current analysis via the iHope pipeline.

II. BACKGROUND

A. Theoretical Waveforms

When Coulomb’s theory of electrostatics was replaced by the more complete Maxwell’s theory of electrodynamics, solutions were introduced which did not vanish at infinity, describing electromagnetic radiation which traveled freely throughout space. Similarly, when general relativity replaced Newtonian gravitation, solutions were found which allowed curved ripples in spacetime to propagate outward from a source. By examining the Einstein Equations...
\[ G^{\alpha\beta} = \frac{8\pi G}{c^4} T^{\alpha\beta} \]  

\[ h^{\alpha\beta} = \sqrt{-g} g^{\alpha\beta} - \tau^{\alpha\beta} \]  

By a careful choice of coordinates, the Einstein equations can be switched to the harmonic (or de Donder) gauge, defined by the requirement

\[ \partial_{\nu} h^{\alpha\beta} = 0 \]  

This gauge simultaneously solves four of the Einstein equations related to the matter stress-energy tensor (via the contracted Bianchi identity) and provides a simplified expression of the remaining equations in terms of a tensorial d’Alembertian operator (\( \Box = \mu^{\nu} \)), such that

\[ \Box h^{\alpha\beta} = \frac{16\pi G}{c^4} \tau^{\alpha\beta} \]  

where the source term, \( \tau \alpha^{\alpha\beta} \), is a pseudo-tensor described by the standard energy-momentum tensor and the gravitational field source term, \( \Lambda^{\alpha\beta} \), as

\[ \tau^{\alpha\beta} = |g| T^{\alpha\beta} + \frac{c^4}{16\pi G} \Lambda^{\alpha\beta} \]  

with the non-linearities of the source term accounting for the higher order Post-Newtonian (PN) terms described here; however, at locations far from the source, the effects of the source term vanish, leaving only the standard homogeneous wave equation

\[ \Box h^{\alpha\beta} = 0 \]  

\[ h_{\alpha\beta}(t, z) = \left( \begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_0 & 0 \\ 0 & h_0 & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) e^{i\omega(t-z/c)} \]  

This describes the simplest possible gravitational wave traveling in the \( z \) direction through empty spacetime. The two polarizations of the gravitational wave are described here as the amplitudes \( h_{\text{plus}} \) and \( h_{\times} \). In general, for non-zero source terms, the two amplitudes are independent functions of time and position (as well as any relevant source parameters). Much work has been done to describe the evolution and gravitational wave emission of compact binary coalescence (CBC) systems via the post-Newtonian expansion. A post-Newtonian source is defined as a source which is both slowly moving and weakly stressed. Current approaches to theoretical waveforms employ the following procedure: a source is approximated as post-Newtonian and the appropriate multipole expansion for the waveforms are calculated; however, as the PN approximation is only valid in the source region and fails at infinity, the expansion is truncated outside the source region, and the waveform is asymptotically matched to a second approximation known as the post-Minkowskian expansion. Unlike the PN expansion, the post-Minkowskian takes into account sources with high velocities and is valid over all space; however, the closed-form solutions for the near source region cannot be calculated, requiring the generation of matched waveforms. For a more complete description, see Blanchet 2006 [3].

The current PN expansion for CBC’s is known out to 3.5 terms in phase and 2.5 terms in amplitude (ignoring spin effects and coupling). The phase is described in terms of a dimensionless time variable

\[ \Theta = \frac{\nu c^3}{5G\mu} (t_c - c) \]  

where \( t_c \) is the time of coalescence, and \( \nu \) is the symmetric mass ratio. In terms of \( \Theta \), the orbital phase out to 2 PN terms is given by

\[ \phi = -\frac{1}{\nu} \Theta^{5/8} \left( 1 + \left( \begin{array}{c} 3715 \\ 8064 \end{array} \right) + \left( \begin{array}{c} 55 \\ 96 \end{array} \right) \right) \Theta^{-1/4} - \frac{3}{4} \pi \Theta^{-3/8} + \left( \begin{array}{c} 9275495 \\ 14450688 \end{array} \right) + \left( \begin{array}{c} 284875 \\ 258048 \end{array} \right) \right) \Theta^{-1/2} \ldots O \left( \frac{1}{c^8} \right) \]  

See [2, 3] for the full 3.5 term PN phase. As for the PN amplitude, the full 2.5 terms are ordered in terms of a frequency related dimensionless variable, \( x \), as shown below:
\begin{align*}
  h_{+,x} &= \frac{2G\mu x}{c^3 R} \left( H_{+,x}^{(0)} + x^{1/2} H_{+,x}^{(1/2)} + x H_{+,x}^{(1)} ight) \\
  &+ x^{3/2} H_{+,x}^{(3/2)} + x^2 H_{+,x}^{(2)} + x^{5/2} H_{+,x}^{(5/2)} + O\left(\frac{1}{x}\right)
\end{align*}

(10)

The first terms are given by

\begin{align*}
  H_{+}^{(0)} &= -(1 + \cos^2(i)) \cos(2\psi) \\
  H_{x}^{(0)} &= -(2 \cos^2(i)) \sin(2\psi)
\end{align*}

(11, 12)

where \(i\) is the angle of incidence of the plane of the binary, and \(\psi\) is the auxiliary phase defined by

\[\psi = \phi - \frac{2GM\omega}{c^3} \ln\left(\frac{\omega}{\omega_0}\right)\]

(13)

where \(M\) is the ADM mass (the total mass with relativistic corrections), \(\omega\) is the current wave frequency, and \(\omega_0\) is the starting frequency for the detector.

The use of equations 11 and 12 with the full PN phase out to 3.5 terms is termed the “restricted waveform”. It is the dominant term of the expansion, and is the primary theoretical tool for the mock data challenge.

B. iHope Pipeline

Developed for the current LIGO science runs, the iHope pipeline is one of the standard data analysis packages currently used in the analysis of interferometer output. The pipeline begins by taking output data (from the two Hanford and one Livingston interferometers), and passing each set of data through a matched filtering search in which the data is searched for signals similar to the theoretically predicted waveforms mentioned above. Potential triggers are identified using a pre-set signal to noise (SNR) ratio. Once the list of signals has been identified, the triggers are compared via a coincidence test between the three detectors. The resulting triggers are then clustered together, and subjected to a \(\chi^2\) test to select the most likely candidates. Afterwards, the surviving triggers are passed through a time-shifted filter in order to estimate the likelihood of such a trigger being randomly generated from known sources of detector noise. The details of the pipeline have been described extensively elsewhere [6].

C. Einstein Telescope

The Einstein Telescope (ET) is a proposed third generation gravitational wave interferometer currently in the design study phase. The current and proposed technologies to be employed in the detector will give it a factor of 10 increase in sensitivity over advanced detectors, with a detection range from 1 Hz to 10 kHz. See Fig. 1.

The enhanced sensitivity will extend the detection horizon of ET significantly beyond that of advanced detectors. Specifically, for the case of compact binary inspirals, the detection range increases by orders of magnitude over current detectors; see Table I for details [5].

![FIG. 1: The proposed ET sensitivity curve, with starting frequency of 1 Hz. Also shown are the sensitivity curves for VIRGO, Initial and Advanced LIGO. Image taken from https://workarea.et-gw.eu/et/WG4-Astrophysics/basesensitivity/et-sensitivity.png/view](https://workarea.et-gw.eu/et/WG4-Astrophysics/basesensitivity/et-sensitivity.png/view)

<table>
<thead>
<tr>
<th>Detector</th>
<th>NS-NS Range</th>
<th>NS-BH Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial LIGO</td>
<td>(z = 0.0035 ) (15 Mpc)</td>
<td>(z = 0.007 ) (30 Mpc)</td>
</tr>
<tr>
<td>VIRGO</td>
<td>(z = 0.0025 ) (11 Mpc)</td>
<td>(z = 0.0055 ) (23 Mpc)</td>
</tr>
<tr>
<td>Adv. LIGO</td>
<td>(z = 0.045 ) (200 Mpc)</td>
<td>(z = 0.09 ) (420 Mpc)</td>
</tr>
<tr>
<td>ET</td>
<td>(z \simeq 1 ) (3.3 Gpc)</td>
<td>(z \simeq 2 ) (5.2 Gpc)</td>
</tr>
</tbody>
</table>

TABLE I: Comparison of detection horizons for neutron star (NS-NS) and neutron star/black hole (NS-BH) for Initial LIGO, VIRGO, and Advanced LIGO. Given distances are co-moving. Data taken from Regimbau and Hughes, 2009 [4].

However with the massive increase in detection range comes a large increase in the number of available sources. While local population studies have generated a large range of predicted CBC rates, the rate of NS-NS inspirals per year in Advanced LIGO’s detection range is expected to be anywhere from 1 to 30; however, with ET’s huge event horizon, the rate jumps to anywhere from \(O(10^5)\) to \(O(10^6)\). This is the primary source of the data analysis challenge for ET. The length of time a signal spends in the detector, called the Newtonian chirp time \((\tau_0)\), is a function of the masses and the starting frequency for the detector \((f_L)\).
TABLE II: Newtonian chirp times (equation 14) for NS-NS inspirals at various starting detector frequencies proposed for ET. The average chirp times are given for NS-NS systems at 1.4$M_{\odot}$ each, and the maximum possible chirp times are given for systems at solar masses. The average signal overlap is calculated by assuming a uniform distribution of $10^5$ signals over the course of a month.

<table>
<thead>
<tr>
<th>Starting Frequency</th>
<th>Chirp Time (1.4 $M_{\odot}$)</th>
<th>Chirp Time (1 $M_{\odot}$)</th>
<th>Average Signal Overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Hz</td>
<td>16.2 min</td>
<td>29.3 min</td>
<td>37.6</td>
</tr>
<tr>
<td>5 Hz</td>
<td>1.7 hours</td>
<td>3.1 hours</td>
<td>238.5</td>
</tr>
<tr>
<td>2 Hz</td>
<td>19.8 hours</td>
<td>1.5 days</td>
<td>2745</td>
</tr>
<tr>
<td>1 Hz</td>
<td>5.6 days</td>
<td>9.4 days</td>
<td>17,431</td>
</tr>
</tbody>
</table>

$$\tau_0 = \frac{5}{256} \left( \frac{1}{\pi \nu f_L} \right) \left( \frac{G}{c^3} M \pi f_L \right)^{-5/3}$$

As expected, the starting frequency for the detector is inversely proportional to the chirp time. The lower the seismic wall, the longer a signal will remain in the detector. Table II shows the average and maximum length a signal will remain in the detector for several proposed lower frequency cutoffs.

Even at Advanced LIGO’s starting frequency of 10 Hz, ET will detect on average 38 NS-NS inspirals at any given time. When the frequency is decreased to the expected 1 Hz, the number jumps to over 17,000. It is unclear if current data analysis techniques developed for LIGO will be successful at extracting this number of simultaneous signals. This mirrors a similar problem faced by the LISA collaboration; the increased range and non-existent seismic wall for the lower frequency lead to an equally high number of predicted simultaneous signals. The solution was to test the data analysis software using generating theoretical waveform injections in colored Gaussian noise generated from the proposed sensitivity curve. Here, we pursue a similar approach to test the potential application of current ground based data analysis techniques for a terrestrial detector with similar source simultaneity.

D. iHope Pipeline

III. MOCK DATA GENERATION

A. Inspiral Generation

In order to simulate a universe populated with random binary inspirals, the following assumptions were made:

- **Input Parameters** - the complete restricted waveform (equations 11 and 12) with all terms includes 17 input parameters (see Table III). For the sake of simplicity, the 6 components of the NS spin are ignored, as well as the eccentricity of their orbits (a reasonable assumption, as binary orbits tend towards circular as the system approaches the final inspiral phase).

- **Homogeneous Population Density** - the simulation assumed an isotropic and homogeneously populated universe (out to $z \simeq 1$). The inspirals were generated in uniformly in distance ignoring the effects of placement in a co-moving volume.

- **Detector Design** - the pattern functions used assume two interferometers at the same location with a 60°. The rotation of the Earth with respect to the source is neglected.

For each inspiral injection, a random number generator from the GNU Scientific Library (GSL) was used to generate random NS-NS inspirals within the detection range for ET. The input variables fall under three categories: masses, position, and wave parameters. These were generated using either uniform or Gaussian random distributions, as described in Table III. The masses were selected from a bounded Gaussian distribution, that is the random number generator was run until the value of the masses was within the given range (one and three solar masses). The position was defined by four variables: the position in the sky, given by the standard polar coordinates $\theta$ and $\phi$, the angle of incidence of the plane of inspiral, $i$, and the luminosity distance of the source, $D_L$. It should be noted that the luminosity distance to which ET is effective is far greater than the co-moving distance $E_G z \simeq 1$ implies $D_L \simeq 17$ Gpc.

A Poisson distribution via the so called ‘Poisson process’ was used to determine the point of injection. The process involved randomly calling a number from a Poisson distribution with a lambda equal to the expected rate of inspirals in a given data set, IE $\lambda = \text{Rate/Sample frequency}$. For the majority of sample bins, the distribution would return zero; however, occasionally the result would be one. At this point, a signal would be injected using the restricted waveform for each of the two detectors. The detector pattern functions were used to keep track of the difference between detector responses, IE to generate the linear combination of detector responses in terms of the two wave polarizations.
\[
H_1(\theta, \varphi, \psi; t) = \frac{3}{2\sqrt{2}} (F_+(\theta, \varphi, \psi)h_+(t) + F_\times(\theta, \varphi, \psi)h_\times(t))
\]

\[
H_2(\theta, \varphi, \psi; t) = \frac{3}{2\sqrt{2}} (F_+(\theta, \varphi + \pi/4, \psi)h_+(t) + F_\times(\theta, \varphi + \pi/4, \psi)h_\times(t))
\]

where

\[
F_+(\theta, \varphi, \psi) = \frac{1}{2} (1 + \cos^2(\theta)) \cos(2\varphi) \cos(2\psi) - \cos(\theta) \sin(2\varphi) \sin(2\psi)
\]

\[
F_\times(\theta, \varphi, \psi) = \frac{1}{2} (1 + \cos^2(\theta)) \cos(2\varphi) \sin(2\psi) - \cos(\theta) \sin(2\varphi) \cos(2\psi)
\]

After the injections were complete, Gaussian colored noise would be generated based on the ET sensitivity curve 1. At each point along the curve, a random Gaussian number with standard deviation of one and mean of zero would be generated, creating a white noise data set. The noise was then colored by multiplying by the strain sensitivity at that frequency. Once completed, the noise data was then transformed into the time domain via an inverse fast Fourier transform (IFFT) and added to the injection data.

### B. Program Design

The program was written in C with all code being written in house (with the exception of GSL). The primary structure consisted of 5 separate .c files and one header file. Control was handled directly from the command line, taking the sample frequency, the duration of the data run, the inspiral rate, and the detector starting frequency as inputs. After the run, the memory was allocated for the run from the outer program, main(). The Poisson process was started, with a Poisson random number being generated for each sample bin. However, rather than start at \(t_{\text{start}} = 0\), the outer program first calculated the maximum possible chirp time given the NS mass boundaries and the starting frequency of the detector. This maximum possible chirp time, called "\(t_{\text{maxchirp}}\)", was the starting point for the Poisson process, \(t_{\text{start}} = -t_{\text{maxchirp}}\). This took into account the possibility of injections which entered the detector before the simulation start time, and coalesced within the generated data. An if loop were then used to ensure that injections coalescing before \(t_{\text{start}} = 0\) were not calculated. Whenever the Poisson process determined an injection should occur, the number of injections at that sample bin (almost always one) was then passed to a second program, rng_shell, which generated random numbers according to the physical distributions described above. Once the input parameters for the injection were generated, they were then passed to two separate functions, amplitude() and PN_phase, which calculated two arrays describing each polarization of the injection in terms of the detector geometry (via the pattern functions, equations 15 and 16). These arrays were each summed into two total data arrays for each detector until the injection process was completed. Afterwards, the function noise was called, generating the Gaussian colored noise for the detectors. The function was called twice, in order to generate independent noise for each individual detector. Once this process was completed, the data was then printed to three .dat files, one for each detector, and one metadata file, listing all of the injections and their respective parameters.

## IV. INITIAL RESULTS

### A. Data Generation

The program proved capable of inserting random injections over a given duration following the given input at the command line. The time of simulation depended on several factors, including the sample frequency of the data run, the starting frequency of the detector (the lower the frequency, the greater number of simultaneous signals, and the longer computational time), and the duration. On average, a sample at 10Hz starting frequency, with the expected rate for ET, a sample rate of 4096Hz, and a duration of 16384 seconds took approximately 4 hours to complete on a single CPU machine. Figure 3 and 4 shows the result of a simulation at 10Hz starting frequency with ET’s projected source rate. Two tests were performed to test the effectiveness of the data simulation. First, a Fourier transform of the output with
FIG. 2: Flowchart describing the program flow and control for the ET mock data challenge. The data inputs for the program are entered at the command line, allowing the program to be run via a cluster.

FIG. 3: 1024 seconds of ETMDC without noise showing 10 inspirals. The characteristic chirp, as well as interference patterns are clearly visible.

FIG. 4: 1024 seconds of ETMDC with noise showing the 70 inspirals expected from the ET inspiral rate. The black is the noise plus signals, with the inspirals highlighted in red. The mixing and overlap of the signals is clearly visible, due to the average overlap of nearly 38 simultaneous signals.

no noise was performed. Since the frequency for the restricted waveform follows a specific power law over time (increasing until coalescence), a Fourier transform of the data with no noise plotted on a log-log plot should reveal the $-8/3$ power law. See 5.

FIG. 5: The fast Fourier transform of the data set (injections only) shown in figure 4. Since the lower frequency cutoff is 10Hz, the data set starts there, then follows a $-8/3$ power law as expected theoretically.

The second test of the data simulation was to compare the generated Gaussian noise with the original ET sensitivity curve. Generating the data set again without injections should produce colored Gaussian noise in the time domain. Performing an FFT on the data set should return a copy of the design study sensitivity curve, abet with Gaussian fluctuations. See 6.

FIG. 6: The fast Fourier transform of the Gaussian colored noise generator without injections. The resultant graph in the frequency domain should match the original ET sensitivity curve 1. A quick comparison of the two shows them to be easily comparable.

These tests suggest that the mock data challenge is
generating the correct theoretical output.

B. iHope Results

Two data sets have been generated so far using the ETMDC, with starting frequencies of 10 and 5Hz, durations of 16384 seconds, and sampling rates of 4096. After conversion to the appropriate frame file formats, the 10Hz data segment was run on the iHope Pipeline. The lower threshold for the detection pipeline was set at 40Hz (initial LIGO setting). After the initial run of the inspiral detection algorithm (LALapps Inspiral), the pipeline reported over 60,000 possible triggers; however, most of these were found to have an SNR of around 5. To correct this issue, the minimum SNR for the pipeline was increased to 7, and the run was restarted. Despite this, the pipeline has yet to produce output data, and is still performing the second run of the inspiral detections.

V. CONCLUSION AND FURTHER WORK

Several directions still exist for improvement in the ET mock data challenge. From a theoretical standpoint, the first necessity would be to embed the sources in a comoving volume, providing a more realistic source population at cosmological distances, especially those expected by ET. Secondly, the higher order harmonic terms from the PN expansion should be included in order to prove a more realistic theoretical model. Thirdly, the PN terms for NS-NS angular momentum and the orbital eccentricity should be included.

From a software standpoint, a large problem with the injection procedure is the sudden spike caused by an injection that begins at full amplitude when it enters the detector, a decidedly unphysical situation. This sudden injection forms essentially a step function, potentially generating analysis problems for any parts of the pipeline dependent on the frequency domain of the data. A solution would be to have the waveforms taper up from zero. Secondly, the current simulation prints output data in dat files; however, as the iHope Pipeline takes frame files as input, which are decidedly less taxing on memory space, the program should be rewritten to generate frame file output.

Once the necessary changes to the program are completed, there are several additional improvements that can be implemented. Currently only NS-NS binaries in the one to three solar mass range are being simulated. A first improvement would be to include the more massive binary black hole inspirals. This would require a change to the Poisson process to require the arrival times to be random in coalescence time, not detector entry time. Also, NS-BH inspirals should be added, requiring the use of more advanced orbital models describing the motion of two highly uneven masses. After that, the simulation could be expanded to include phenomena other than CBCs: gamma ray bursts, supernovae, stochastic background, etc.

Instrumentally speaking, the program should also be upgraded to include random glitches in the data, simulating problems in the data stream that actual detectors suffer from. Additionally, the motion of the Earth with respect to an incoming signal should be taken into account (via a simple addition of an increasing term in the ϕ component of the detector pattern functions).

VI. REFERENCES