Characterization of Hardware Injections in LIGO Data

Shannon Wang
LIGO SURF 2014
California Institute of Technology

Advisor–Dr. Alan Weinstein
Department of Physics
California Institute of Technology

Advisor–Dr. Jonah Kanner
LIGO Laboratory
California Institute of Technology
(Dated: October 6, 2014)

There are simulated astrophysical signals in LIGO data that were injected into the detectors by moving the test masses. The simulated signals should appear in the data more or less exactly at the labeled times as real signals. The objective of this study is to retrieve every compact binary coalescence hardware injection by implementing a matched filter search for the signals in the data, and comparing the results with the expectations from a list of attempted injections. The templates used to produce the hardware injections are reproduced for the matched filter search; there is one template for every injection in science mode data. Our basic method expects fifty seconds of science mode data on both sides of the merger. Methods for recovering injections when there aren't enough data are developed and tested in this paper. The matched filter calculates both the expected signal to noise ratio (SNR) and the recovered SNR; if the recovered SNR matches the expected SNR and is recovered at approximately the time of the merger, the injection is said to be recovered. All but ten of the injections that were expected have been recovered; two of those injections never made it into the data, and the signals from the other eight injections were overwhelmed by glitches.

1. BACKGROUND

LIGO, short for Laser Interferometer Gravitational-Wave Observatory, is a physics experiment that uses three Michelson interferometers (henceforce denoted as H1, H2, and L1) to detect gravitational waves. Currently, only the H1 and L1 detectors are in use; the H1 detector is located in Hanford, Washington, and the L1 detector is located in Livingston, Louisiana. The detectors use laser interferometry as their core technology and have mirrors that also serve as gravitational test masses. A gravitational wave passing through the interferometer will cause a phase modulation on the light in the interferometer's arms; there will be a relative phase shift between the arms. Fig. 1 demonstrates how passing gravitational waves produce a signal in the output photodetector..

The interferometers (IFO) are designed so that the unmodulated light will interfere destructively while the phase modulated light will interfere constructively and produce a signal that is proportional to the gravitational wave strain. The initial detectors were designed to be sensitive to strains with amplitudes as small as 10^{-21}. The arm length change resulting from such a strain is 10^{-18} m; thus, the detectors are built using highly stable lasers, vibration isolation, and interferometers with very long arms to increase signals from gravitational wave strains. [1] The size of the detectors can be observed from Fig. 2 and Fig. 3.

LIGO's data potentially contain gravitational waves from many sources, so LIGO analyses search for these

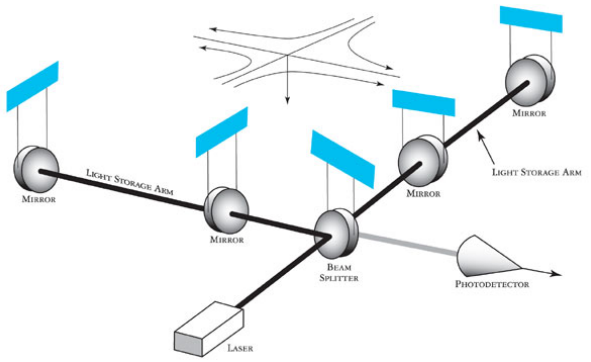

FIG. 1: The beam from the laser is sent to the beam splitter, where it is then split into two beams and sent down the two arms of the interferometer and then recombined before the photodetector. Gravitational waves passing through interferometer cause oscillating distortions, which in turn cause modulations in the phase shifts of the light in each arm. Adapted from [2].

sources in the data. Each source requires different analysis techniques, which depend on whether the gravitational waveforms are well-modeled, or whether only spectral characterizations can be produced. There are four types of signals LIGO searches for: transient, modeled waveforms from compact binary coalescences; transient, unmodeled waveforms from gravitational-wave bursts; continuous narrow-band waveforms from continuous wave sources; and continuous broad-band waveforms

FIG. 2: An aerial view of the LIGO observatory at Hanford, Washington is shown above. Evacuated beam tubes extend from the building, in which lasers and optics are kept. The full length of one H1 arm is shown; the arms of the H1 detector each extend 4 km. They are perpendicular to each other and enclosed in concrete. The 2km-long H2 detector also occupies the beam tubes. That detector was not in operation during the LIGO S6 run. Adapted from [1].

FIG. 3: An aerial view of the LIGO observatory at Livingston, Louisiana is shown above. As in the Hanford observatory, the beam tubes are at right angles to each other and extend 4 km in each direction. Adapted from [1].

from stochastic gravitational wave backgrounds.

This project focuses mainly on the compact binary coalescence search. Binary coalescences are systems in which the compact objects (neutron stars and/or black holes) are spiraling toward each other. After the distance between the objects decreases below the smallest stable circular orbit, or after the inspiral phase, the binary system becomes unstable and the objects merge. The system, now a single highly-perturbed black hole, will then relax through damped sinuosoidal oscillations during the ringdown phase. The inspiral, merger, and ringdown phases of the system are collectively called a compact binary coalescence (CBC). CBCs provide an environment for testing the general relativity theory in strong fields. The components of the binaries are black holes and neutron stars; thus, if gravitational waves are discovered from compact binary coalescences, they will serve as unequivocal evidence for the existence of black holes and provide information on the properties of black holes, the population of binary systems in the universe, and the nuclear equation of state in conditions such as the neutron stars. [1]

The LIGO Open Science Center (LOSC) is preparing to release its archived data from the S6 run to the public. Enhanced LIGO, or the S6 run, started on July 7, 2009 and ended on October 20, 2010, so the Global Positioning System (GPS) times range from 930960015 to 971654415. The S6 detector data have been downsampled from 16384 Hz to 4096 Hz and are stored in hdf5 files labeled by IFO and GPS time. The data contain simulated signals from CBCs, which must be documented before the data are released. The signals are chirp waveforms like those produced by CBC inspirals. Chirp waveforms are determined by the masses of the two objects; they take the form of sinusoids that increase in frequency and amplitude until the merger phase, as can be seen in Fig. 4, and can be modeled using Post Newtonian approximations. In order to locate all of the simulated signals, compact

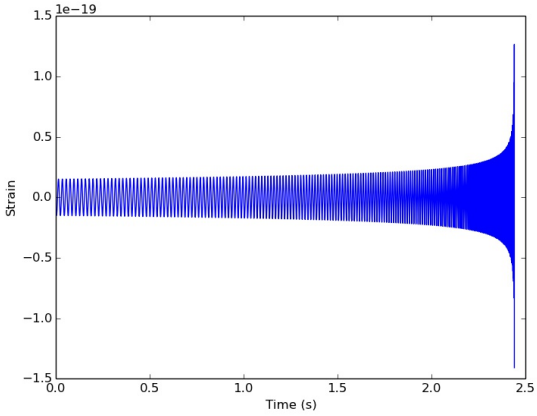

FIG. 4: This waveform was determined by an binary system with components that have masses of $7.60M_{\odot}$ and $2.85M_{\odot}$ and is shown here in the time domain. The strain detected oscillates between greater amplitudes until the merger phase. The inspiral phase lasts approximately 2.5 seconds.

binary coalescence templates are generated and used to perform optimal matched-filter searches for the hardware injections.

2. MATCHED FILTERING

The optimal detection method for the hardware injections is the matched filter search. A simple matched filter search consists of constructing frequency-domain templates that serve as approximations of gravitational wave signals and applying them to detector data that

have been windowed and undergone Fast Fourier Transforms. Complications that may arise in the search for gravitational waves, such as discrepancies between gravitational wave signals and templates that use parameters, i.e. the binaries' masses, distances from detectors, spin to approximate the signals, will not occur in the matched filter search for the hardware injections, because the templates used to recover the injections are taken from the GWF files used to produce the injections. Unknown parameters, such as phase, are accounted for by using the quadrature sum of orthogonal matched filter outputs to compute the recovered signal to noise (SNR). The matched filter search is best suited to detecting signals in stationary Gaussian noise. While non-Gaussian noise affects the detector data, the matched filter is robust enough to recover certain injections, provided that the mergers and part of the inspirals are in the data recorded when the detectors were in science mode and no big glitches occurred at the times of the mergers. [3]

There are two lists of planned binary inspiral injections (biinjlists) that contain the injections' start times, file names, and statuses. It is assumed that the biinjlists contain lists of every possible injection time. A few exceptions to this rule, especially blind injections, will be dealt with at a later date. H1biinjlist contains the H1 injections; L1 biinjlist contains the L1 injections. Likewise, there are two parameters lists; the H1 and L1 parameter lists differ in only the merger times. Coherence of the H1, L1, and V1 detectors was practiced in the implementation of the S6 hardware injections so that the signal would arrive at the detectors at slightly different times. The V1 detector belongs to VIRGO, an interferometer for detecting gravitational waves in Italy. There is no V1biinjlist.

An analysis pipeline was constructed to read in the biinjlists, identify the CBC injections, read in the relevant data, and perform the matched filter analysis on the injections. The analysis procedure identifies the CBC injections in the biinjlists, whose labels either start with CBC or HWINJ. The injections' start times are then matched to the merger times in the parameter lists; the difference between the start times and the corresponding merger times varies between ninety-five and ninety-six seconds. The matched filter takes both the start time and the merger time and then reads in the corresponding data file(s).

The optimal lengths for data segments undergoing Fast Fourier Transforms are powers of two, but because the templates used in the matched filter are one hundred seconds long, the matched filter takes one hundred seconds of data to analyze and eight hundred seconds of data to calculate the power spectral density. The data for the power spectral density is sliced so that the time of the merger is in the middle of the eight hundred seconds. If there aren't eight hundred seconds of data, the matched filter uses however much data there are in the segment identified by the start and merger times.

The Fast Fourier Transform (FFT) is an algorithm that implements the discrete Fourier transform (DFT). Let the time domain signal be defined as $x[n]$ on interval $0 \leq n < N$, where N is the number of samples in the domain. Then the DFT is defined as such [4]:

$$X_D(f_k) = \sum_{n=0}^{N-1} x[n] e^{-j2\pi n f_k} \tag{1}$$

The frequency is sampled at f_k; k denotes the sample's index. The DFT reproduces the discrete time Fourier transform at frequencies $f = \left(\frac{k}{2N}\right)$. The DFT does not perfectly execute the Fourier transform. Because the DFT is discrete, there may be aliasing, which will cause leakage at the intermediate frequencies. Discontinuities in the time domain data and waveforms will also cause leakage after the FFT. Therefore, the detector data and templates must be windowed while they are still in the time domain so as to avoid leakage that might overwhelm the real signal. Windowing increases spectral leakage, but it distributes the leakage to areas that cause the least harm; thus, it improves the signal to noise ratio. [5]

The templates contain the strains, the magnitude and phase, of the hardware injections in the data. XML files containing the parameters of the injections are taken from [6–11]. These XML files are extracted as parameter lists that can be later used in the matched filter. They can also be fed into coinj, a script in the LIGO Scientific Collaboration Algorithm Library (LALSuite) that produces injections using the parameters from the files [12]. Coinj uses the XML files, the GPS injection start times, and the GPS injection end times to produce the templates used to make the injections.

The templates produced by coinj are time domain templates and must be Fast Fourier Transformed into frequency domain templates. Coinj yields tapered templates, but because the data was originally recorded at 16384 Hz, the templates are naturally produced at 16384 Hz and must be downsampled to 4096 Hz. The Fast Fourier Transformed templates are then applied to the windowed data. The matched-filter output is computed as follows [3]:

$$z(t) = 4 \int_0^\infty \frac{\tilde{s}(f)\tilde{h}^*_{template}(f)}{S_n(f)} e^{2\pi i f t} df \tag{2}$$

In Equation (2), $z(t)$ is the complex matched filter output, $\tilde{s}(f)$ is the data stream in the frequency domain, and $\tilde{h}^*_{template}(f)$ is the complex conjugate of the template. $S_n(f)$ is the estimated detector noise power, which can be calculated by finding the power spectral density. The character n denotes the stationary Gaussian noise process assumed to calculate the power spectral density. The magnitude of the complex matched filter output allows the data to be efficiently searched for all arrival times, but because the coalescence phase is unknown, the complex matched filter output is used to efficiently search

the unknown phase. The square of the complex matched filter output is the quadrature sum of the two orthogonal matched filters. [3]

Three waveforms are used in the S6 hardware injections, so the following three types of templates must be constructed: EOBNRpseudoFourPN, GeneratePPNtwoPN, and SpinTaylorT4threePointFivePN. These respectively are the names of the effective-one-body (EOBNR), post-Newtonian (PN), and spin-Taylor waveforms used by coinj. The masses, effective distances, and other parameters of the injections are known, so once a template is constructed, all that remains to be done is writing a script that will feed the parameters into the template and calculate the signal to noise ratio. In order to calculate the signal to noise ratio, the normalization constant must be computed. It can be calculated with only the template and the power spectral density [3]:

$$\sigma_m^2 = 4\int_0^\infty \frac{|\tilde{h}_{D_{eff}Mpc,m}(f)|^2}{S_n(f)}df \tag{3}$$

The normalization constant represents the expected value for the matched filter output when there is a signal in the injection segment at the specified effective distance, denoted in Equation 3 as D_{eff}.

The signal to noise ratio can be computed by cross-correlating the template and the data [3]:

$$\rho_m(t) = \frac{|z_m(t)|}{\sigma_m} \tag{4}$$

The recovered SNRs show up as peaks in plots of SNR vs. time. The times at which the maximum SNRs are recovered are recorded in the log files produced by the analysis procedure. Because a template is a hundred seconds long and the merger usually occurs at around ninety-five seconds, there is usually a five second offset between the merger time and the time of recovery. The ringdown phase lasts less than 0.1 seconds, so the expected offset between the two times can be calculated by subtracting the merger time from the end of an injection, which is a hundred seconds from the injection time. If the discrepancy between the time of recovery and the merger time differs more than 0.2 seconds from the expected offset, then the signal has not been successfully retrieved. For a signal to be successfully retrieved, the time of recovery must match the merger time and the recovered SNR must match the expected SNR.

3. METHODS AND UNREFINED RESULTS

3.1. Approaches

Only four seconds of data were used for recovering each injection in the first matched filter run, for fear that there wouldn't be enough science mode data (defined as stable data taken when the detector was in lock) surrounding the mergers. The first matched filter run yielded disappointing results.

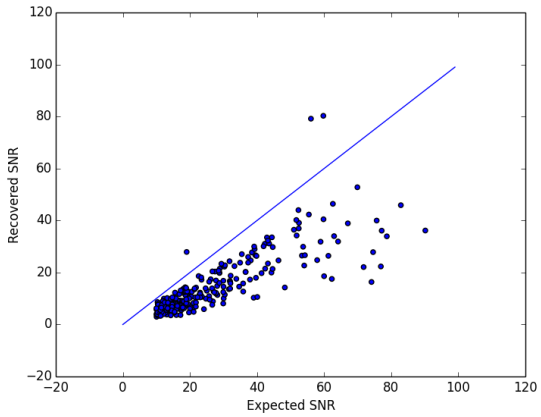

FIG. 5: This plot shows the recovered SNR plotted against the expected SNR for the L1 detector. The FindCHIRP PN frequency domain template was applied to four seconds of data, or 16384 samples, to produce these results. Injections with expected SNRs under ten are not included in this plot.

Fig. 5 demonstrates how most of the recovered injections are recovered with SNRs far below their expected values. While the poor results may be due to the usage of incorrect templates, they could also be attributed to narrow window of data. The chirp times are defined as contributions at different post-Newtonian orders to a gravitational wave signal duration; the start of a chirp is defined as the time when the signal is at f_L, the lower limit on the frequency integral in Equation 2, and the end is when the system coalesces. τ_0 and τ_3 are defined as such [13]:

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

$$\tau_3 = \frac{1}{8\eta f_L}\left(\frac{G\pi M f_L}{c^3}\right)^{-2/3} \tag{6}$$

In Equation (5) and Equation (6), M represents the total mass of the binary and η is defined as the symmetric mass ratio. Fig. 6 plots the chirp time τ_0 vs. the post-Newtonian contribution to the chirp time at 1.5 post-Newtonian order, τ_3. SI units were used in the calculations. It can be seen that many of the injections have long inspiral phases, and limiting the amount of data analyzed to four seconds causes losses in SNRs. Increasing the amount of data to thirty-two seconds greatly improved the results, as can be seen in Fig. 7; the remaining loss of SNR can be attributed to the template used.

The FindCHIRP PN (post-Newtonian) frequency domain template, taken from [3], was employed in the first

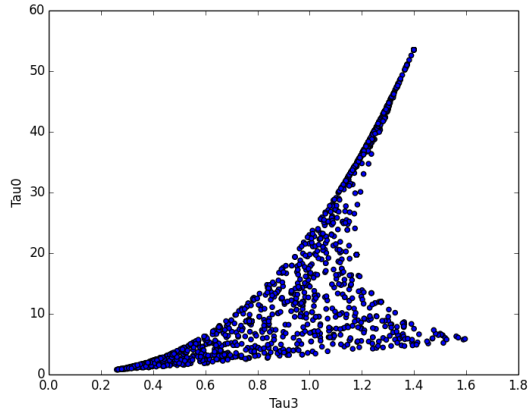

FIG. 6: The chirp time τ_0 plotted against the 1.5 post-Newtonian correction τ_3 gives the distribution of the injections' inspiral lengths. It can be seen that the majority of the injections have inspiral lengths longer than ten seconds, which is why using four seconds of data for the matched filter yields poor results. The loss of data also results in the loss of SNR.

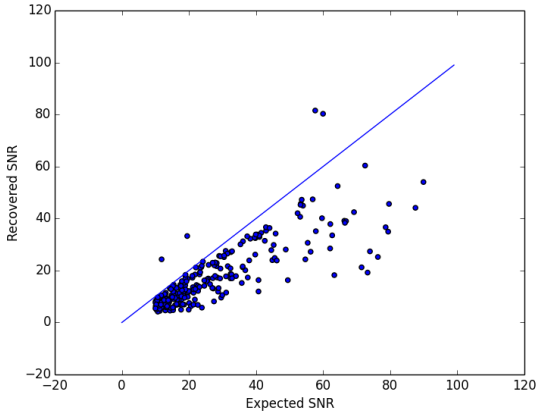

FIG. 7: This plot shows the recovered SNR plotted against the expected SNR for the L1 detector. The FindCHIRP PN frequency domain template was applied to thirty-two seconds of data, or 131072 samples, to produce these results. The injections that hover close to $y = x$ are GeneratePPNtwoPN injections. Injections with expected SNRs under ten are not included in this plot.

two matched filter runs; the FindCHIRP PN frequency domain template approximates a parameterized post-Newtonian (PPN) waveform of second post-Newtonian order, so it recovered the majority of the SNRs for the GeneratePPNtwoPN injections. However, because the FindCHIRP template is a PN template, it only contains the inspiral phase and stops at f_{ISCO}. [3] The EOBNR calibrates the Effective One Body approximation with Numerical Relativity simulations and includes the merger and ringdown phases. [14] Thus, the FindCHIRP PN template is a poor approximant for EOBNR waveforms.

LALSuite was employed in an attempt to use the parameters to produce templates as the matched filter ran. There was difficulty locating the proper implementation that would produce an EOBNRpseudoFourPN waveform; the approximant EOBNRv2 produces EOBNRv2pseudoFourPN waveforms that unfortunately don't match the EOBNRpseudoFourPN waveforms in phase. While the FindCHIRP PN frequency domain template performed admirably in recovering the GeneratePPNtwoPN injections, the approximant approach was abandoned because the EOBNR injections make up the bulk of the injections. LALApps, also part of LALSuite, is a collection of gravitational wave pipelines and data analysis codes that use the algorithm libraries to simulate gravitational waves. LALApps was used to reproduce the GWF files of the hardware injections. A template maker was then written in Python to locate the templates' GWF files by detector and injection start time. The script would then lift the template's strain from the GWF files, downsample it, and return the Fast Fourier Transformed results along with a frequency vector.

Once LALApps' coinj was chosen as the template producer, the amount of data needed for analysis was increased to a hundred seconds, and the amount of data needed for calculating the power spectral density became eight hundred seconds. This method allows the majority of the injections to be successfully recovered. In consideration of outliers, the analysis procedure stores the output in a log file that contains the injection times, the merger times, the parameters, the statuses, the expected SNRs, the recovered SNRs, the times of recovery, and the xml files containing the parameters. The injections that remained a matter of concern could only be resolved after a thorough analysis of the data and the log file the analysis procedure produced.

3.2. Initial Results

The matched filter run using the templates produced by coinj yielded reasonable results. Fig. 8 displays the loglog plots and truncated scatter plots of the recovered SNR vs. the expected SNR for the H1 and L1 injections labeled successful. Fig. 9 displays the log plots and truncated scatter plots of the recovered SNR vs. the expected SNR for the H1 and L1 injections labeled otherwise. It may seem odd that most of the H1 injections were recovered with SNRs higher than the expected values whereas most of the L1 injections were recovered with approximately a ten percent loss in SNR, but these discrepancies are most likely due to calibration uncertainty.

The injections that may be causes for worry are the outliers shown in the log plots. There are several explanations for these outliers: it could be that the injections were never injected, or that the mergers were overwhelmed by glitches or spectral leakage. In order to distinguish the outliers from the successful injections and understand the reasons behind the discrepancies, a script

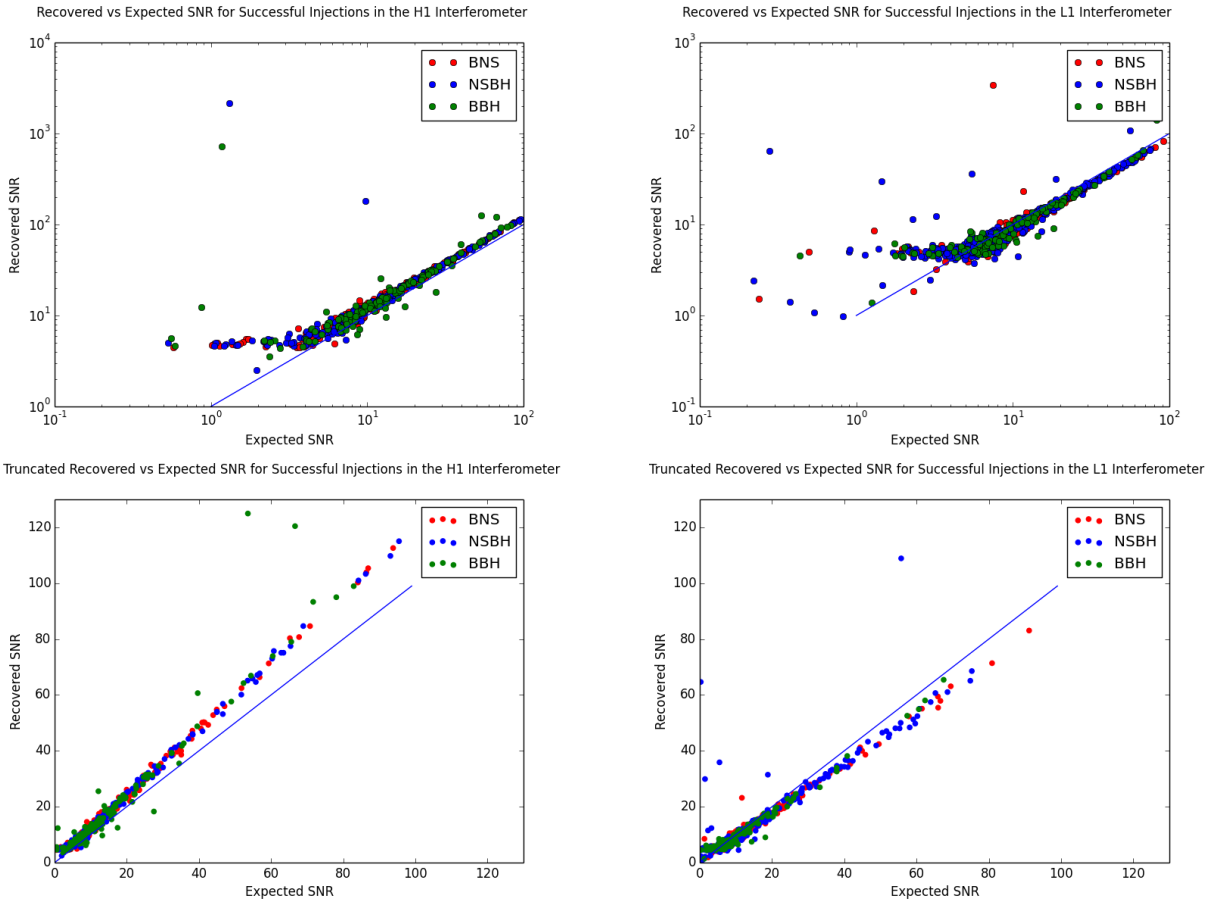

FIG. 8: The plots above show the recovered SNR vs the expected SNR in the (left) H1 and (right) L1 detectors. The top plots have log-log scale, and the bottom plots have linear-linear scale. The injections above are labeled successful in the biinjlists; these results were produced with coinj templates and a hundred seconds of data. It can be observed that there is a consistent loss in SNR for the L1 injections and a consistent gain in SNR for the H1 injections as a result of calibration issues. The points are color coded by binary type. Signals from binary neutron stars are labeled BNS, neutron-star black-hole binaries are labeled NSBH, and binary black holes are labeled BBH.

called `injection_checker.py` was written to separate the outliers into four categories: unexpected, anomalies, shorties, and deceptive.

There are injections in the log files with the following statuses: not in science mode, GRB Alert, injection compromised, and operator override. Most of these injections should not be in the data, but some of these injections have recovered SNRs that are high enough to warrant double checking. These injections make up the unexpected lists. Then there are the injections that are marked as successful in the biinjlists but may not be, because there are large discrepancies between the expected and recovered SNRs. As a result of calibration problems, it's normal that there exists a twenty percent difference between the recovered and expected SNR. In order to account for these differences, the successful injections were filtered so that only injections whose recovered SNR to expected SNR ratios are greater than 2 or smaller than 0.5 remained; the fifty percent difference was chosen so as to be conservative. Injections whose expected SNRs are lower than 6 were also eliminated, unless their recovered SNRs were more than eight times the expected SNRs, because injections with SNRs under 6 are easily overwhelmed by noise. The remaining injections make up the lists of the anomalies. Most of the anomalies are caused by glitches, but a few are caused by calibration problems and spectral leakage.

In order to determine which anomalies are truly anomalies, and in order to find the injections that don't fall within the definition of anomalies but are still a far cry from successful, lists of injections that are labeled successful but are recovered at times that differ from the merger times by more than .02 seconds were produced. These are the deceptive lists. Once again, injections whose expected SNRs are lower than six were filtered out. It's obvious that several injections in both the anomalies and deceptive lists don't have fifty seconds of data on both sides of the merger. Thus, all such injections, regardless of their statuses, are gathered into the shorties list.

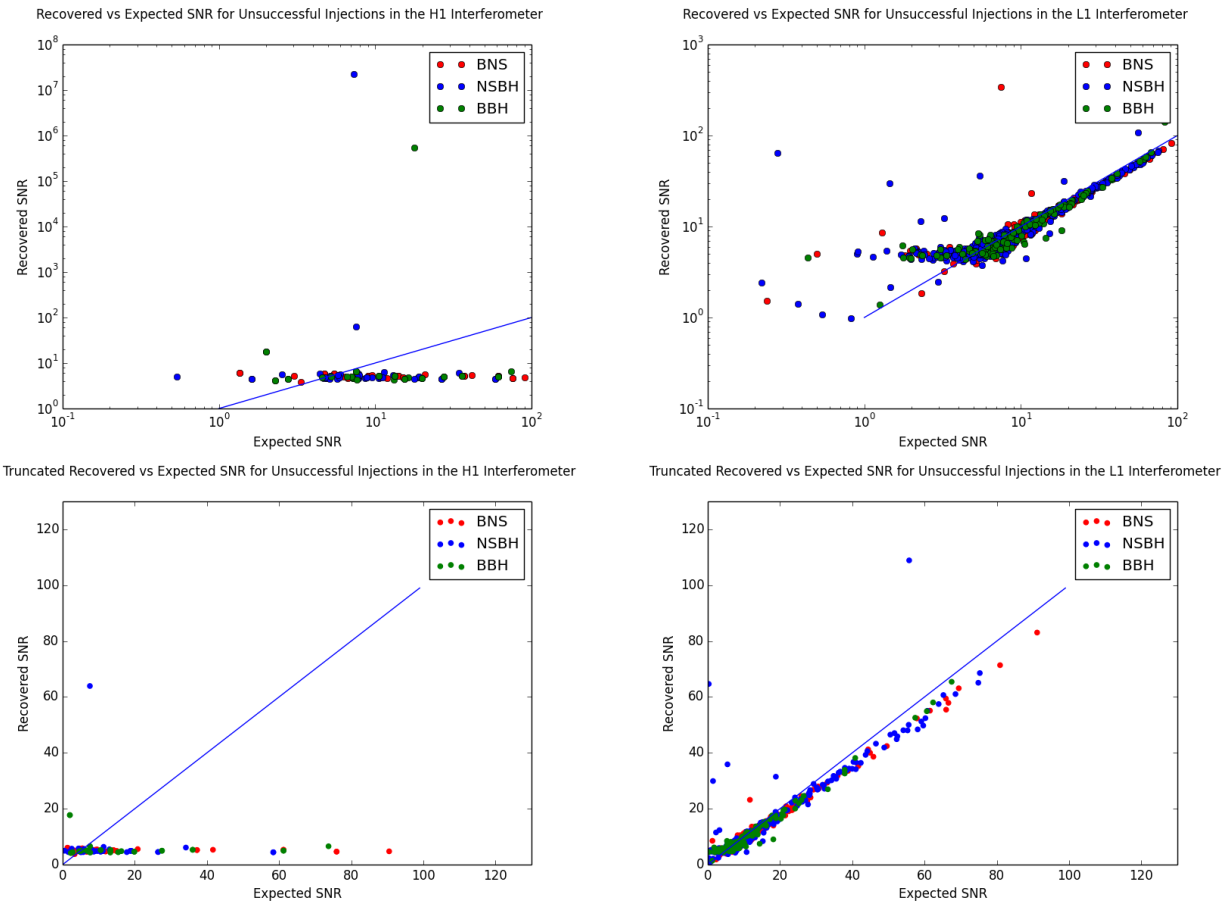

FIG. 9: The injections above are labeled other than successful in the biinjlists; these results were produced with coinj templates and a hundred seconds of data. The points are color coded as in Fig. 8.

The merger times of all the injections in these lists were then run through omega scan to ascertain their presences. Fig. 10 shows what a scan of a present injection looks like.

Most unexpected injections are absent from the data; the injections that never entered the data stream are now denoted as absent on the lists. The recovery times of all the injections were also run through omega scan to search for glitches. Fig. 11 demonstrates what a scan of a glitch looks like, and Fig. 12 demonstrates what a scan of a present and recoverable injection looks like.

Some scans revealed glitches; however, since most injections are located in science mode data, the glitches should not overwhelm the injections to the point that they are entirely undetectable, provided that the glitches do not happen too close to the merger and that the glitches are not excessively loud. Therefore, the presence of a glitch only explains why the injections cannot be recovered with 100 seconds of data and a Blackman window; for most of the injections, it is not an excuse for failing to recover the injections. There are scans that reveal nothing: there are no glitches to overwhelm the mergers and the injections are obviously present. Those discrepancies can be attributed to spectral leakage in some cases and loud noise in other cases.

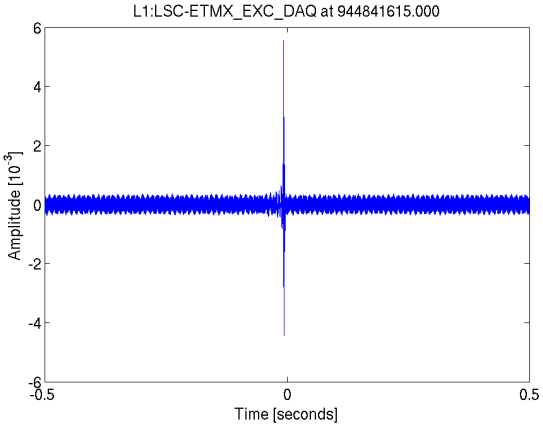

FIG. 10: Time series scan of a present injection. This L1 injection's merger takes place at GPS 944841615. There is an obvious spike in the injection channel data. If the spike is absent, then the injection never entered the data stream, which renders the matched filter result meaningless.

Most of the recoverable injections that are seemingly

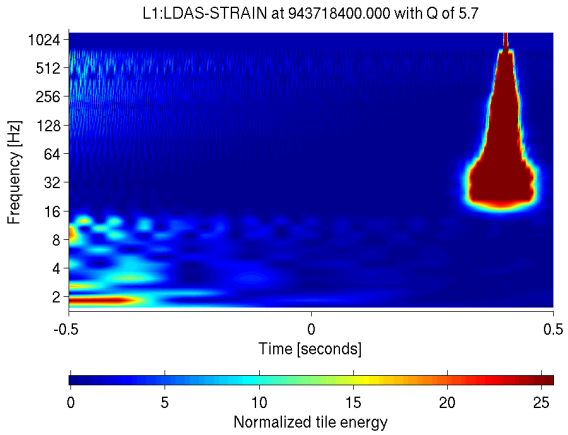

FIG. 11: Omega scan of a glitch. The red blotch signifies a glitch, and the magnitude of the blotch indicates the size of the glitch. A glitch can overwhelm an injection. The effects of a glitch can be mitigated by taking only four seconds of data, windowing the four seconds, and zero-padding the data segment so that it matches the length of the template.

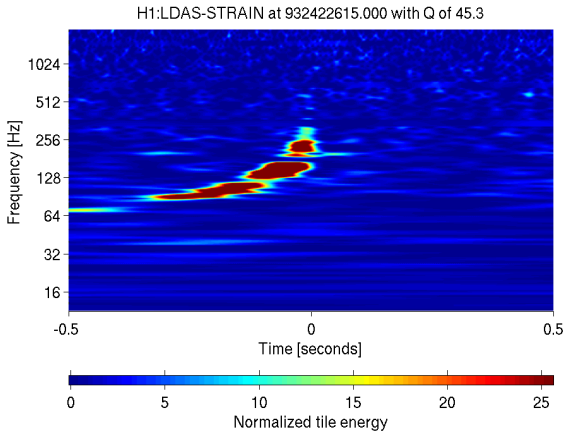

FIG. 12: Omega scan of a present and recoverable injection. The chirp waveform can be observed as the red curve in the blue background. The inspiral phase is the part of the curve that increases slowly and the merger and the ringdown phases can be observed from the rapidly increasing tail of the curve.

overwhelmed by glitches are shorties; of course, there are a handful of anomalies, and a few overlaps between the shorties and the anomalies. Since omega scans of shorties and anomalies do not give enough information about data quality, another method of analysis was used to determine whether the glitches are truly large enough to overwhelm the injections. Loglog plots of the windowed and Fast Fourier Transformed data, the Fast Fourier Transformed templates, and the power spectral densities were made with `data_plotter.py`. Fig. 13 shows a plot of an injection that can be recovered; Fig. 14 shows a plot of an injection that cannot be recovered because it has been overwhelmed by a glitch. However, the accuracy of `data_plotter.py` relies on the amount of data used and the window function chosen, so the methods of analyzing shorties and anomalies must be determined before claims of glitches can be made.

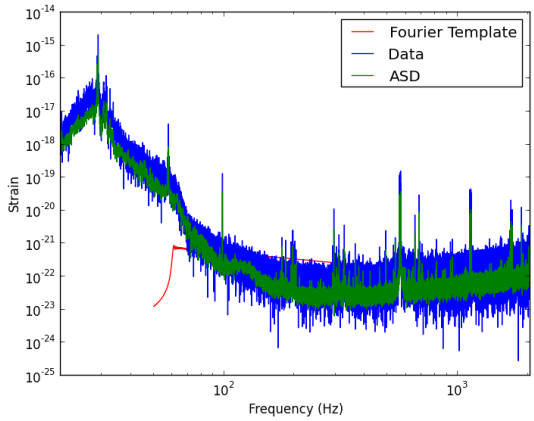

FIG. 13: The loglog plots of the windowed and Fast Fourier Transformed data, template, and power spectral density appear normal. The power spectral density is denoted as ASD, the amplitude spectral density. Plots of successfuly recovered injections look like this. This injection was injected at time 946771120; there aren't fifty seconds of data following its merger, but this plot demonstrates that this injection should be recoverable.

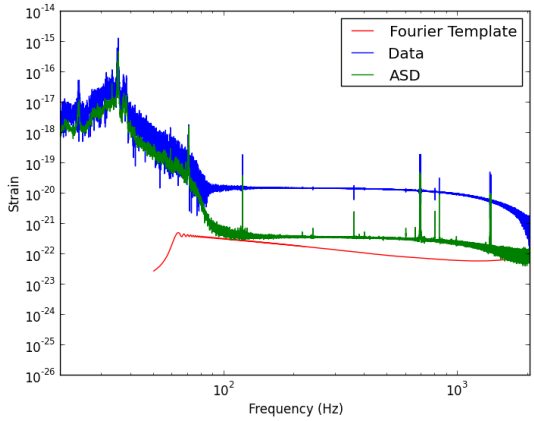

FIG. 14: The loglog plots of the data and power spectral density peter out midway. The power spectral density is denoted as ASD, the acceleration spectral density. This H1 injection, injected at time 943718320, could not be successfully recovered, despite being labeled successful. One possible explanation is that there are two data files for this injection, which might have led to the corruption of the data.

4. DATA AND ANALYSIS

Four methods of dealing with shorties were experimented with. In the first method, the first priority is centering the merger in the middle of the data, so the number of seconds between the merger and the cutoff time determines how much data was taken. If the detector went into lock thirty seconds before the merger, then sixty seconds of data will be used. If the detector went out of lock twenty-two seconds after the merger, then forty-four seconds of data will be taken. The template is first Fourier transformed and then sliced; because the frequency vector and the template vector are the same length, the frequency vector can be used to determine the template values at certain frequencies. Since f_{isco} is the frequency at which the merger occurs, the template function must be written so that the sliced template will always contain the strain at ISCO frequency. However, the power spectral density and the increment of frequency are affected by the length of the sliced template; thus, both expected and recovered SNR values are adversely affected when sliced templates are employed.

It is desirable to use a hundred seconds of data, because the template is a hundred seconds long and the first method has demonstrated that the templates should not be sliced. Thus, the second method of dealing with the template still takes a hundred seconds of data to analyze. Because there will be a shortage of data within fifty seconds either before or after the merger, the analysis procedure will read in however many seconds of data there are, and if it reaches the end of the science mode segment before it hits the fifty second limit, it will calculate how many more seconds of data are needed and take the required data from the other end of the merger, e. g. if the data segment ends at thirty seconds after the merger, the analysis procedure will use the thirty seconds of data following the merger and the seventy seconds of data prior to the merger. Because the merger is no longer located in the middle of the hundred-second segment, the Tukey window is used instead of the Blackman window. Fig. 15 shows the Blackman window, which is used to recover the injections that are not shorties.

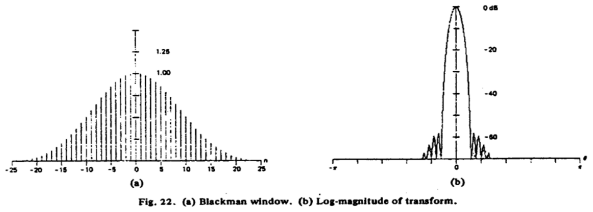

Fig. 22. (a) Blackman window. (b) Log-magnitude of transform.

FIG. 15: This plot demonstrates how the Blackman window tapers a waveform. The Blackman window is close to optimal and minimizes leakage. Adapated from [5].

A Blackman window sharply dampens the strain's amplitude for all values in the vector it is applied to except for the center; thus, the merger must be centered in the middle of the data if a Blackman window is used. However, if a Tukey window is used, then the injection can still be successfully recovered, even if the merger isn't centered in the data vector. The scope of the Tukey window can be seen in Fig. 16. The Tukey window is defined as [5]:

$$w(n) = \begin{cases} \frac{1}{2}\left[1+\cos\left(\pi\left(\frac{2n}{\alpha(N-1)}-1\right)\right)\right], & \text{if } 0 \le n \le \frac{\alpha(N-1)}{2} \\ 1, & \text{if } \frac{\alpha(N-1)}{2} \le n \le (N-1)\left(1-\frac{\alpha}{2}\right) \\ \frac{1}{2}\left[1+\cos\left(\pi\left(\frac{2n}{\alpha(N-1)}-\frac{2}{\alpha}+1\right)\right)\right], & \text{if } (N-1)\left(1-\frac{\alpha}{2}\right) \le n \le (N-1) \end{cases}$$

The alpha value determines the width of the Tukey window. A wider window allows more flexibility in the positioning of the merger. It is best to use a Tukey window with an alpha value of 0.5 for injections that have more data preceding and following the merger and a Tukey window with an alpha value of 0.1 for injections that have less data. The ratios of the recovered SNR to the expected SNR of the shorties recovered using the second method are recorded in Fig. 17.

While the second method recovers most of the recoverable injections, there may be one or two injections that have mergers too close to the edge of the 100 seconds; the Tukey window doesn't yield desirable results, but these injections are still recoverable, so a third method was developed. The core of the third method is the same as the first: the merger is centered in the middle of the data, so the amount of data taken is double the number of seconds from the merger to the cutoff time. A Blackman window is then applied to the data, and then the windowed data is zero padded on both sides until the vector is 409600 samples long. The vector is Fast Fourier Transformed before the template is applied to it. A shorter data segment allows less room for glitches and noise spikes, which increases the probability of successfully recovering the injection. All of the injections recovered by the second method were also found using the third method. Zero padding in the time domain is equivalent to interpolation, or upsampling, in the frequency domain, and when the

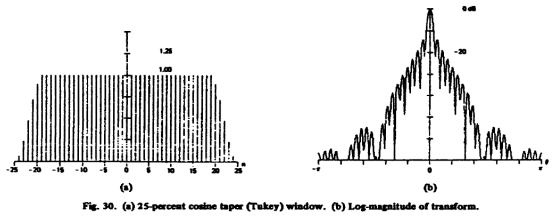

FIG. 16: This plot demonstrates how the Tukey window tapers a waveform. A larger alpha value gives a narrower window, whereas a smaller value gives a wider window. An alpha value of 0 renders the Tukey window a rectangular window. The alpha value chosen in the picture above is 0.25. Adapted from [5].

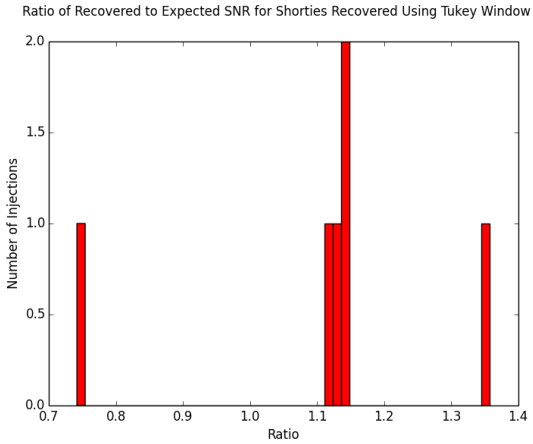

FIG. 17: The plot above is a histogram of the ratio of the recovered SNR to expected SNR for the H1 and L1 shorties when they are recovered using a Tukey window. Because only one L1 shorty was successfully recovered, this histogram combines both H1 and L1 recovered shorties. There are fifty bins in this histogram for finer resolution.

amount of data is little compared to zero-padding, upsampling becomes oversampling. Oversampling increases the SNR when there is enough inspiral phase data, so some of the injections the second method recovered were found by the third method at slightly higher SNRs [4]. The results from the third method are recorded in Fig. 18. Comparing Fig. 17 and Fig. 18 shows that except for the rightmost injection in the histogram, the second method yields lower recovered SNR to expected SNR ratios than the third method does.

Because the second and third methods deliver similar results, one final method must be developed to recover the anomalies and the few shorties left. In this last method, only four seconds of data are used. Once again, the merger is centered in the data vector; a Blackman window is applied to the data and then the vector is zero-padded to 409600 samples before undergoing the Fast Fourier Transform. This method guarantees the recovery of the anomalies and shorties that the second and third methods cannot recover. Nevertheless, because this

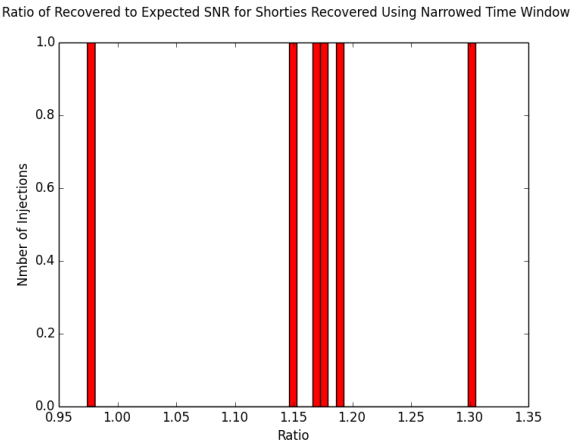

FIG. 18: The plot above is a histogram of the ratio of the recovered SNR to expected SNR for the H1 and L1 shorties when they are recovered using a narrowed time window and a Blackman window. As in Fig. 17, this histogram combines both H1 and L1 recovered shorties. There are fifty bins in this histogram.

method uses a very small segment of the data and neglects most of the inspiral phase, it is recommended that this method only be used on the injections otherwise unrecoverable. Fig. 19 and Fig. 20 show the results of a matched filter that only employs four seconds of data.

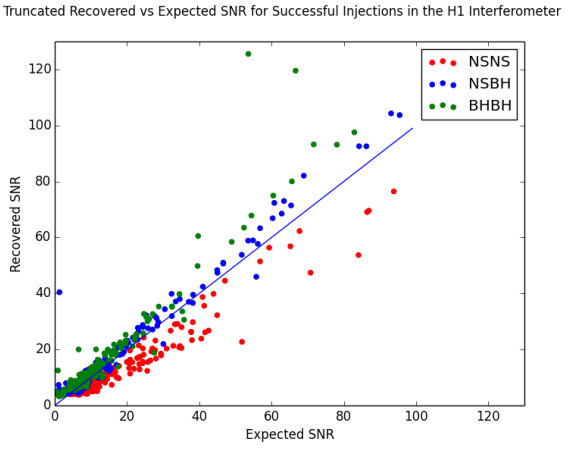

FIG. 19: The injections above are labeled successful in the H1biinjlist; these results were produced with coinj templates and four seconds of data. The losses in SNR can be attributed to insufficient data. The scattergrams are truncated so that extreme outliers are not displayed.

The most effective method would be to combine methods two and four. If inspiral phase data must be lost to recover the signal, then the matched filter might as well simply employ four seconds of data. This combination ensures that all recoverable injections may be recovered without the worry that the SNR might be too high as a result of oversampling or too low as a result of insufficient

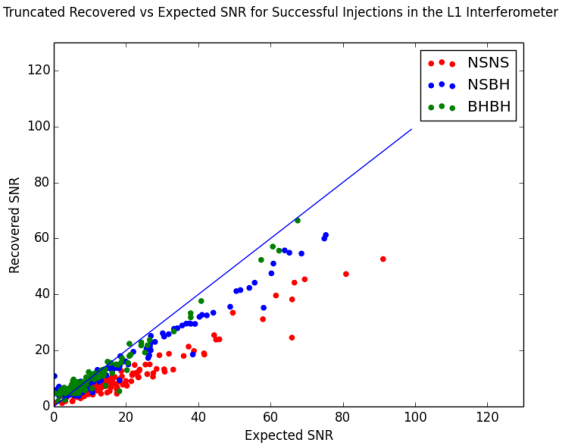

FIG. 20: The injections above are labeled successful in the L1biinjlist; these results were produced with coinj templates and four seconds of data. The losses in SNR can be attributed to insufficient data. Any gains due to oversampling are not obvious because of calibration issues. The scattergrams are truncated so that extreme outliers are not displayed.

data. The remaining discrepancies can be understood through the use of omega scan and `data_plotter.py`.

5. RESULTS

The analysis procedure was modified so that all shorties would be automatically directed to the Tukey window and the other injections would be passed through the Blackman window. The remaining shorties that couldn't be recovered were examined once more using a four-second window; if employing a four-second window, windowing, and then zero-padding the data still didn't yield results, then the injections were declared unrecovered. The results are stored in `H1_final_cbc_list.txt` and `L1_final_cbc_list.txt`, and are displayed in Fig. 21 and Fig. 22.

There are 1481 distinct injections in the H1biinjlist. 664 of those injections took place when the detectors were out of lock. One injection has part of its inspiral phase in science mode data, but because its merger isn't in science mode data, the injection is still unrecoverable. There are 722 injections labeled as successful in the H1biinjlist, and out of those injections, there are 569 injections with expected SNRs greater than six. Out of those 569 injections, two injections are overwhelmed by glitches; the rest have been successfully recovered. In addition to the 567 injections recovered, two injections labeled Injection Compromised have also been recovered. Thus, 569 out of 571 injections above the noise threshold were recovered for the H1 detector, yielding a recovery rate of 99.65%.

There are 1545 distinct injections in the L1biinjlist. 788 of those injections took place when the detectors were out of lock. There is also one injection, labeled Injection Compromised, that has part of its inspiral phase in science mode data but is unrecoverable because its merger isn't in science mode data. There are 658 injections marked as successful in the L1biinjlist, and out of those injections, there are 458 injections that have expected SNRs greater than six. Out of the 458 injections, there are two injections that are absent from the data and six of them that are overwhelmed by noise. The remaining injections have been recovered, as well as one injection labeled as Injection Compromised. Thus, 451 out of 459 injections above the noise threshold were recovered for the L1 detector, yielding a recovery rate of 98.26%. Table I presents a summary of these figures.

The recovery rates imply that the tools exist to recover a gravitational wave signal as long as the templates bear similarity to the signal and there are at least four seconds of science mode data. Future research should emphasize on recovering the burst hardware injections and developing methods to recover signals with imperfect templates. The EOBNRv2 template should be cross-correlated with the EOBNR template to determine the lowest degree of similarity that still allows for signals to be recovered, as should the FindCHIRP Pn frequency domain template. Because the EOBNRv2 template performed worse than the FindCHIRP template did in recovering the hardware injections, similarity might not be the only factor affecting the recovery of signals.

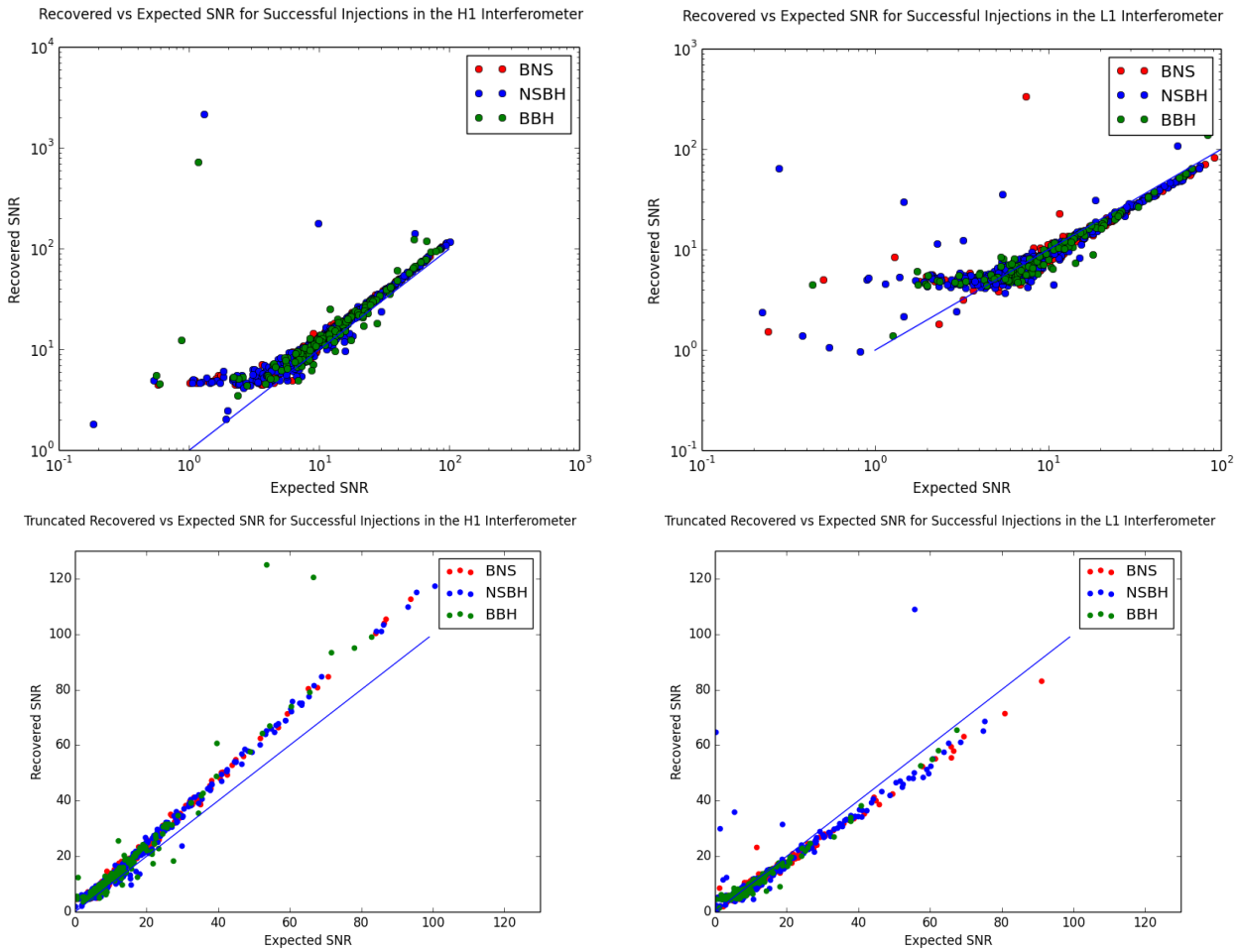

FIG. 21: The plots above show the recovered SNR vs the expected SNR in the (left) H1 and (right) L1 detectors. The top plots have log-log scale, and the bottom plots have linear-linear scale. The injections above are labeled successful in the biinjlists; these results were produced with coinj templates and a hundred seconds of data. These plots show the final results; the four-second window corrections have been applied to the initial results. The points are color coded by binary type, as in Fig. 8

TABLE I: Table of Results. All of the H1 and L1 injections are accounted for in this table. They have been grouped into injections labeled unsuccessful and in non science mode data, injections labeled successful but still in non science mode data, injections labeled unsuccessful in science mode data that haven't been recovered, injections labeled unsuccessful in science mode data but have been recovered, injections labeled successful in science mode data with SNRs under six, injections labeled successful in science mode data with SNRs over six that can't be recovered, and injections labeled successful in science mode data with SNRs over six that have been recovered.

	H1	L1
Injections Labeled Unsuccessful without Data	658	784
Injections Labeled Successful without Data	6	4
Unrecovered Injections Labeled Unsuccessful in Science Mode Data	93	98
Recovered Injections Labeled Unsuccessful in Science Mode Data	2	1
Injections Labeled Successful in Science Mode Data with SNR Under 6	153	200
Unrecovered Injections Labeled Successful in Science Mode Data with SNR Over 6	2	8
Recovered Injections Labeled Successful in Science Mode Data with SNR Over 6	567	450
Total Injections	1481	1545

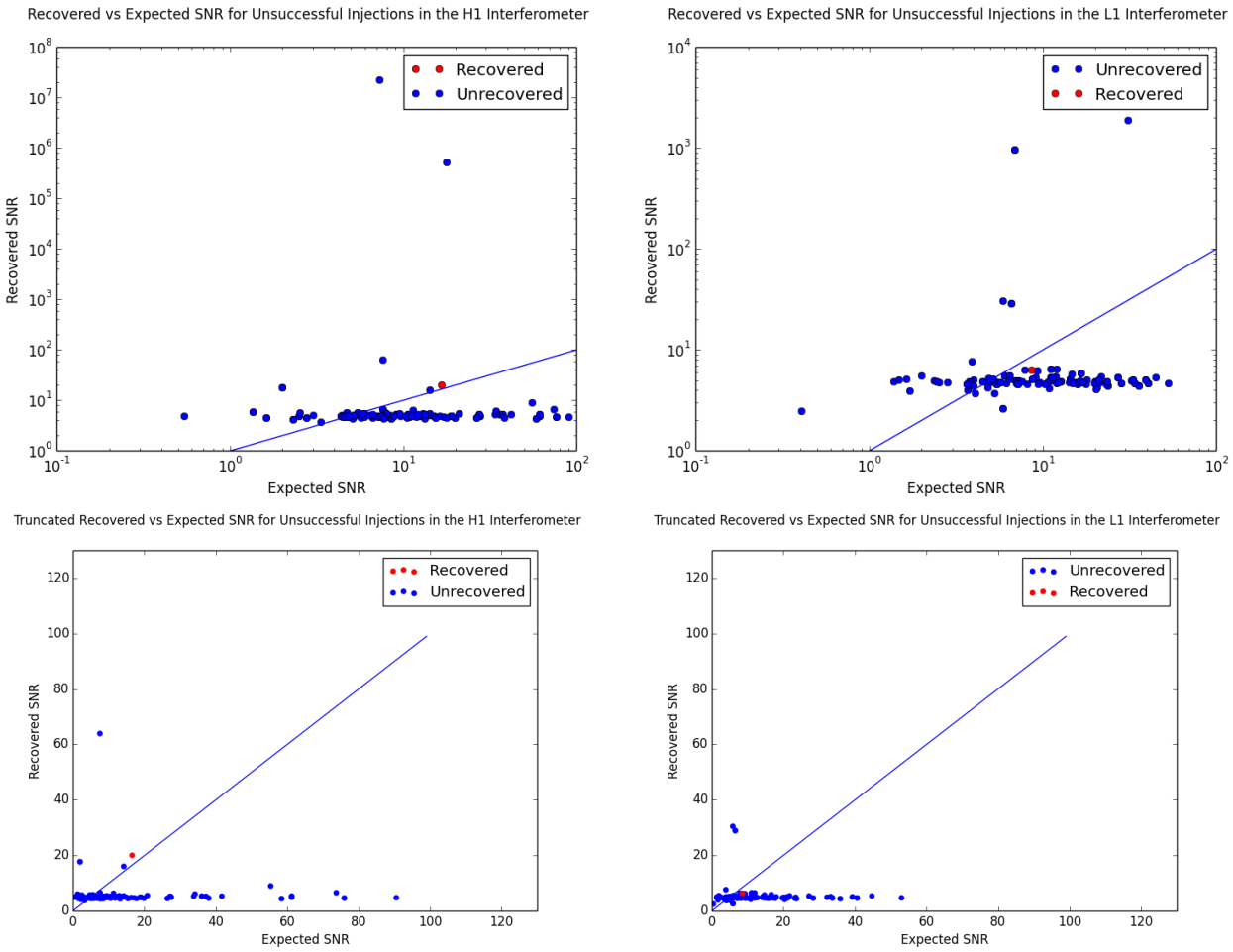

FIG. 22: The injections above are labeled other than successful in the biinjlists; these results were produced with coinj templates and a hundred seconds of data. These plots show the final results, after the four-second window corrections have been applied to the initial results. The recovered injections are colored red, and the unrecovered injections are colored blue.

6. THE BIBLIOGRAPHY

[1] The LIGO Scientific Collaboration, 2009, LIGO: The Laser Interferometer Gravitational-Wave Observatory, LIGO-P070082-v4, http://arxiv.org/abs/0711.3041.

[2] Introduction to LIGO and Gravitational Waves. LIGO Scientific Collaboration. Available: http://www.ligo.org/science/GW-IFO.php.

[3] B. Allen, 2012, FINDCHIRP: an algorithm for detection of gravitational waves from inspiraling compact binaries, Phys. Rev. D 85 122006.

[4] R.J. Marks, Handbook of Fourier Analysis and Its Applications, Oxford University Press, [2009].

[5] F. J. Harris, "On the Use of Windows for Harmonic Analysis with Discrete Fourier Transform," Proceedings of the IEEE, 66(1), 1978 pp. 5183.

[6] J. Veitch, S6 Injections, retrieved from `https://ldas-jobs.ligo.caltech.edu/~jveitch/S6inj/`

[7] J. Veitch, S6 Injections, retrieved from `https://atlas.atlas.aei.uni-hannover.de/~jveitch/931564743-931651143/hardware_inj/`

[8] J. Veitch, S6 Injections, retrieved from `http://www.sr.bham.ac.uk/~jveitch/E14/`

[9] J. Veitch, S6 Injections, retrieved from `http://www.sr.bham.ac.uk/~jveitch/HWINJ/`

[10] J. Veitch, S6 Injections, retrieved from `https://ldas-jobs.ligo.caltech.edu/~jveitch/S6inj/S6_endrun/`

[11] J. Veitch, S6 Injections, retrieved from `https://ldas-jobs.ligo.caltech.edu/~jveitch/S6inj/endofrun/`

[12] Retrieved from `https://www.lsc-group.phys.uwm.edu/daswg/projects/lalsuite.html`

[13] B.S. Sathyaprakash, B.F. Schutz, 2009, Physics, Astro-

physics and Cosmology with Gravitational Waves, Living Rev. Relativity 12.
[14] R. J. E. Smith, I. Mandel, A. Vecchio, 2014 Studies of waveform requirements for intermediate mass-ratio coalescence searches with advanced gravitational-wave detectors, (Preprint arXiv:1302.6049[astro-ph.HE]).

Acknowledgments

I would like to thank Dr. Jonah Kanner for his patient explanations on the intricacies of a matched filter search, his prompt replies to all of my questions, and his constant encouragement that gave me the courage to explore many different methods of signal recovery. I would also like to thank Dr. Alan Weinstein for giving me this wonderful opportunity and mentoring me through the writing of this paper.

Appendix A: Tools Used

The tools and scripts used to produce this paper are listed below in alphabetical order. They are:

1. Compare.py
2. Compare_lists.py
3. determiner.py
4. documented_matched_filter.py
5. documented_template.py
6. final_plotter.py
7. hopeful_matched_filter.py
8. hopeful_template.py
9. injection_checker.py
10. new_data_plotter.py
11. parser.py
12. readligo.py
13. template_producer.py
14. tukey.py
15. xml_extractor.py

The xml extractor was adapated from make_cbcLog.py, which can be found at https://ldas-jobs.ligo.caltech.edu/~jkanner/s6inj/make_cbcLog.py/. These tools can be found at https://ldas-jobs.ligo.caltech.edu/~shannon.wang/share/.

Appendix B: List of Files

The files used to produce the results mentioned in this paper are listed below:

Produced using the xml extractor and xml files from Dr. Veitch's directories:

1. H1_parameters.txt
2. L1_parameters.txt

Copied from Dr. Kanner's directory: https://ldas-jobs.ligo.caltech.edu/~jkanner/s6inj/

1. H1biinjlist.txt
2. L1biinjlist.txt

Output from matched filter and lalapp -coinj command:

1. H1_final_cbc_list.txt
2. L1_final_cbc_list.txt

Produced using the injection checker:

1. H1_anomalies_list.txt
2. H1_deceptive_list.txt
3. H1_shorties_list.txt
4. H1_unexpected_list.txt
5. L1_anomalies_list.txt
6. L1_deceptive_list.txt
7. L1_shorties_list.txt
8. L1_unexpected_list.txt

Copied from the following directories:

https://ldas-jobs.ligo.caltech.edu/~jveitch/S6inj/
https://atlas.atlas.aei.uni-hannover.de/~jveitch/931564743-931651143/hardware_inj/
http://www.sr.bham.ac.uk/~jveitch/E14/
http://www.sr.bham.ac.uk/~jveitch/HWINJ/
https://ldas-jobs.ligo.caltech.edu/~jveitch/S6inj/S6_endrun/
https://ldas-jobs.ligo.caltech.edu/~jveitch/S6inj/endofrun/

1. all_possible_endofrun.xml
2. all.xml
3. CBC_BLINDINJ_968654558_adj.xml
4. H1_all_successful.xml
5. h1_injections.xml
6. H1L1_all_successful.xml
7. H1L1V1_all_successful.xml
8. H1V1_all_successful.xml
9. HL-INJECTIONS_1-924652815-10000.xml
10. HL-INJECTIONS_1-928875615-244800.xml
11. HL-INJECTIONS_1-930493015-5305400_adj_err.xml
12. HL-INJECTIONS_1-930493015-5305400_adj.xml
13. HL-INJECTIONS_1-930493015-5305400.xml
14. HL-INJECTIONS_1-935798415-5270400_adj.xml

15. HL-INJECTIONS_1-941068815-5184000_adj.xml
16. HL-INJECTIONS_1_946339215-5097600_adj.xml
17. HL-INJECTIONS_1_951436815-5270400_adj.xml
18. HL-INJECTIONS_1_957052815-4924800_adj.xml
19. HL-INJECTIONS_1_961977615-5270400_adj.xml
20. HL-INJECTIONS_1_961977615-5270400_consolidated.xml
21. HL-INJECTIONS_1_961977615-5270400_EMtest_actuallydone.xml
22. HL-INJECTIONS_1_961977615-5270400_EMtest.xml
23. HL-INJECTIONS_1_961977615-5270400_preEMtest_actuallydone.xml
24. HL-INJECTIONS_1_967593615-4924800_adj.xml
25. HL-INJECTIONS_S6_ALL.xml
26. HLV-H_INJECTIONS_S6-930493015-31455814.xml
27. HLV-H_INJECTIONS_S6-930493015-36755000.xml
28. HLV-H_INJECTIONS_S6-930493015-37098117.xml
29. HLV-H_INJECTIONS_S6-930493015-42025400_original.xml
30. HLV-H_INJECTIONS_S6-930493015-42025400.xml
31. HLV_INJECTIONS_S6END_971049615_604800.xml
32. HLV-INJECTIONS-TRIPLE.xml
33. inj.xml
34. L1_all_successful.xml
35. L1V1_all_successful.xml
36. V1_all_successful.xml

The files can be located at `https://ldas-jobs.ligo.caltech.edu/~shannon.wang/` and `https://ldas-jobs.ligo.caltech.edu/~shannon.wang/share`.

For the sake of brevity, the templates and the hdf5 files are not listed here. The templates can be located on the LOSC cluster at UWM: `losc.phys.uwm.edu:/home/shannon.wang`. The hdf5 files can also be found at `losc.phys.uwm.edu:/losc/archive/strain-hdf/S6prototypeV1`.