AM3064 Project Report

Signal Processing in Interferometry

Cian Roche - 115535053
Department of Applied Mathematics
University College Cork

ABSTRACT. This document aims to provide an overview of basic signal processing techniques in the context of a real world example. This example is the 5σ confidence detection of a binary black hole merger via gravitational wave observatory advanced LIGO on 14/09/15. The techniques used in this relatively simple introduction include the discrete Fourier transform, fast Fourier transform, power spectral density, amplitude spectral density, whitening and visualisation techniques such as the creation of spectrograms.

KEYWORDS: signal processing, interferometry, discrete Fourier transform, DFT, fast Fourier transform, FFT, whitening, spectrogram, GW150914

Typeset using LaTeX

Preface

In this report, ideas, quotations and results are referenced using footnotes as opposed to a bibliograpy due to the relatively informal nature of many of the sources. A detailed list of all figures contained within the document and their origins (original or otherwise) can be found at the end of the report. Note also that all referencing is hyperlinked in the digital version, thus by clicking on a reference number most PDF readers will bring the referenced item to the top of the screen. This is not necessary for comprehension, but can be a useful tool for quick navigation.

1. Motivation

A reasonable question to ask is "Why are signal processing techniques important?". The answer to this is simple: the techniques of signal processing are of sufficient generality that their applications are plentiful and highly variable. The most basic (or conventional) application is to simply obtain information about data read by some instrument as a function of time. We do this by examining the components of this signal via various techniques outlined in this report, however these techniques have no restriction to time domain data, nor the dimension of said data. The "signal" to which we apply these techniques can be a signal in a more abstract sense, for example in the application of signal processing techniques as a tool to solve systems of partial differential equations (PDEs).

Applications of these techniques include speech recognition, weather, economic and seismoligical forecasting, image processing (including medical imaging), computer graphics and many areas of experimental physics. Their importance is illustrated explicitly in the example presented here; a signal seeming to be composed of mostly noise and of little interest is transformed into one of the most important physical detections of the 21st century.

2. Physical Preliminary

In order to understand the example, it is important to first have a basic understanding of the physical background and of the measurements being taken. A basic overview of the structure, effects and origins of gravitational waves along with the interferometric techniques used to observe them is presented thus.

2.1. Gravitational Waves

Gravitational waves were first predicted by Henri Poincaré in 1905, demonstrating that the well-established Lorentz transformations require an emission of some gravitational radiation from any accelerating mass[1]. Upon development of his general theory of relativity (published in November 1915), Albert Einstein and his colleagues realised that it too predicted this same gravitational radiation. This idea is in some ways analogous to that of acceleration of an electric charge, which results in the emission of electromagnetic radiation. Einstein was resistant to the idea however, as attempts to express the equations of general relativity in a form similar to that of Maxwell's equations and the wave equation (Eq. 1) required him to make many assumptions, due to the complicated nature of general relativistic calculations[2].

$$\nabla^2 \psi(r,t) = \frac{1}{v^2} \frac{\partial^2}{\partial t^2} \psi(r,t) \tag{1}$$

It emerged later that his efforts were unsuccessful because he was using a coordinate system not entirely suited to the calculations, which lead to singularities that raised doubts in Einstein's mind. After many more years of indecision, Einstein eventually accepted their existence with some "encouragement" from his colleagues, with one very important detail: gravitational waves, if they did exist, would have incredibly small amplitudes and may be impossible to detect. It is for this reason that it took exactly 100 years to progress from the conceptualisation of this phenomenon to it's measurement.

---

[1] "Sur la dynamique d llectron" - Comptes Rendus of the French Academy of Sciences, 5 July 1905

[2] A Brief History of Gravitational Waves - arXiv:1609.09400

Gravitational waves are therefore a type of radiation emitted upon acceleration of mass, which travel at the speed of light and do carry some energy (as was argued by Feynman with his "sticky bead" argument). Their effect is to stretch and squeeze the space through which they pass, the amount by which the space is stretched being proportional to the amplitude of the gravitational wave. The oscillatory stretching and compressing of space takes place orthogonal to the direction of propagation of the wave (like the E and B fields of an electromagnetic wave), and we define the strain $h(t)$ as the amount by which the length of an object differs from it's "natural length" l_0 at a given time t. This can be expressed as in Equation 2.

$$l(t) = [1 + h(t)]\, l_0 \tag{2}$$

2.2. Interferometry

The gravitational wave strains $h(t)$ typically observed due to massive astronomical events such as binary black hole or neutron star mergers at the position of the Earth are on the order of 10^{-21}. To put this in perspective, measuring a strain this small on a rod of length 4km would change the length of this rod by 10^{-4} times the diameter of a proton. Measuring this is equivalent to changing the distance to the nearest star to our solar system, Proxima Centauri, by the width of one human hair and then measuring this change in distance confidently[3]. In order to measure a strain like this, a highly precise and clever approach is needed. This approach is interferometry.

A Michelson interferometer is shown in Figure 1, which is an apparatus that operates on the principle of wave interference. If the beam which is split 50/50 at the mirror interface travels different distances along each of the arms of the interferometer, then the recombined beam will be of smaller amplitude than the input beam due to destructive wave interference (provided the path difference is not an integer multiple of λ, the wavelength of the light). Naïvely speaking, the amount by which this is smaller gives an indication of the path length difference, allowing for variations in space in the plane of the interferometer to be measured.

[3] LIGO press conference 11 February 2016

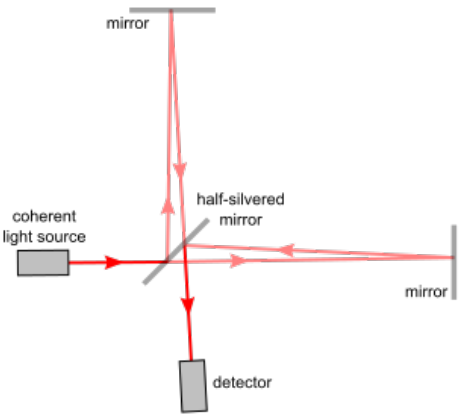


FIG. 1: *A simplified look at the Michelson interferometer*

Advanced LIGO consists of two highly specialised versions of the Michelson interferometer located in Livingston and Hanford, North America. They are the products of decades of work by some of the greatest minds currently in physics, and so it would be very difficult to explain here their function in it's entirety. Instead, it is simply important to know that via this interference method, a strain as a function of time is measured. All data collected by advanced LIGO must be released to the public with the tools to interpret this data, leading to the publicly available jupyter notebook on which the data analysis in this report is based[4].

The strain data ± 5 seconds around gravitational wave event GW150914 as measured by both advanced LIGO detectors is shown in figure 2. This data does not appear to contain any gravitational wave signal that would be expected from a binary black hole merger whatsoever; the expected signal is a regular oscillation of increasing frequency, building up to a "chirp" followed by a ringdown, corresponding to the increase in orbital frequency of a binary black hole system as they approach merger (via loss of energy through gravitational radiation itself) until they merge and no more gravitational radiation is observed. A single

[4] Full tutorial at https://losc.ligo.org including data releases

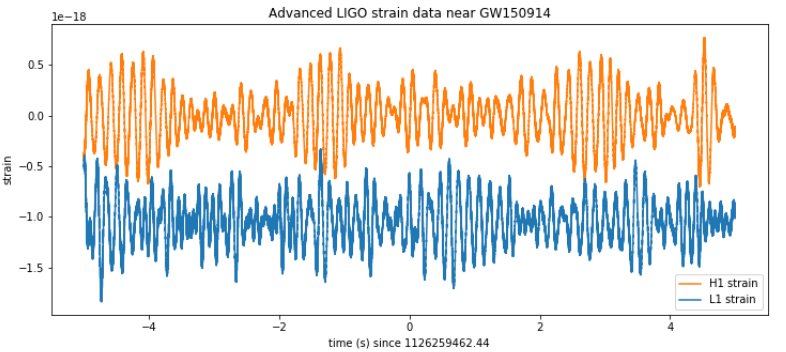


FIG. 2: *Note that the time displayed on the x axis is GPS time, ie. the number of seconds since Jan 6, 1980 GMT. Also note that the L1 strain is shifted from zero mean for presentation purposes.*

rotating body in space is not accelerating like two components of a binary are, and is thus invisible to a gravitational wave observatory. This is where the all important signal processing techniques come in, as this very signal contains evidence of that exact scenario: two black holes (one about 30 solar masses and the other about 40) spiraling towards their center of mass and merging in under one tenth of a second[5].

3. The Discrete Fourier Transform (DFT)

Recall the definition of the continuous Fourier transform to frequency space for a function of time $x(t)$ as in Equation 3. This can be visualised as examining the frequencies of the sinusoids required to "construct" this $x(t)$ as in Figure 3. Of course angular frequency ω could be used in place of f, however in these applications it is often easier to interpret results in linear frequency space. Also note that there is of course a reverse Fourier transform, and the normalisation factor of $1/2\pi$ can be included in the Fourier transform or its inverse (or perhaps even $1/\sqrt{2\pi}$ in each) but this is simply a matter of convention. It will be included in

[5] B.P. Abbott et al. hPhys. Rev. Lett. 116, 061102 Published 11 February 2016

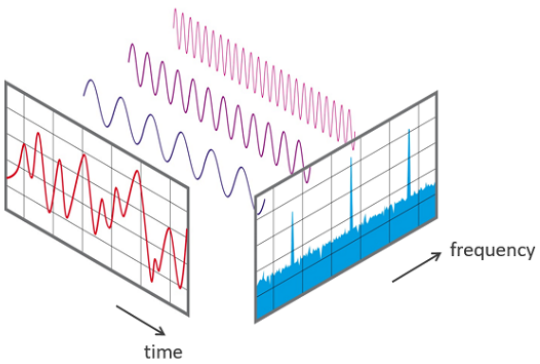


FIG. 3: *A visualisation of the Fourier transform from the time domain to the frequency domain*

the reverse Fourier transform throughout the remainder of this report.

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-2\pi i f t} dt \tag{3}$$

Of course we do not measure a continuous signal, we measure discrete data in time "bins" and thus a Fourier transform cannot be calculated in this way. We instead use the discrete Fourier transform (DFT) which can be thought of as a function which takes a series of data (N points) in the time domain and returns N complex numbers, the magnitudes and phases of which represent the frequency content on the time domain signal.

$$x_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i \frac{k}{N} n} \tag{4}$$

The formal expression of the DFT (Eq. 4) is analagous to the continuous case, with the function $x(t)$ replaced by the n^{th} data point x_n, the time t replaced by the summation index n and the frequency replaced by the index of the frequency bin being calculated, k, normalised to the total size of the dataset N. Note the summation is from $n = 0$ to $N-1$, yielding exactly N complex values in the frequency domain. A very simple example of the DFT is shown in figure 4, where the frequency content of a simple sinusoid of frequency 1 Hz and amplitude 1 is examined.

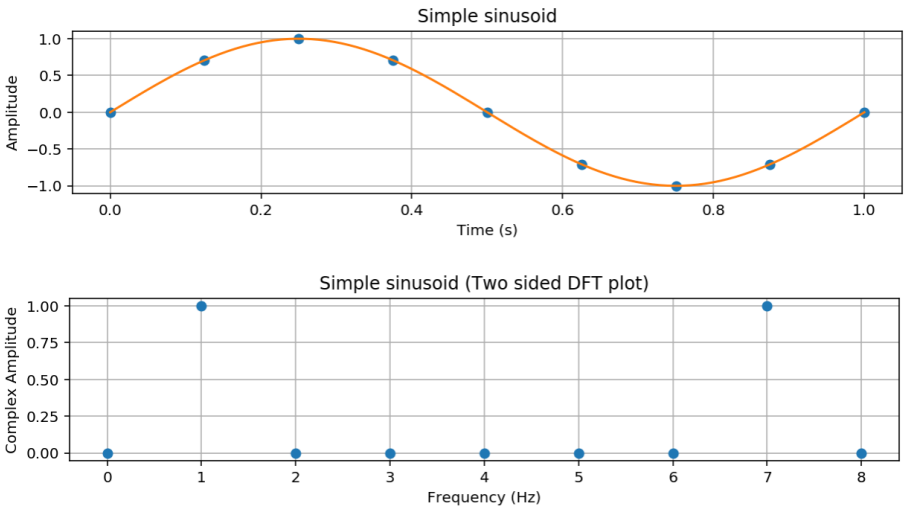


FIG. 4: *DFT of simple sinusoid on a two-sided DFT plot. Orange line is the theoretical signal and blue dots represents measured data points.*

Note that in the DFT complex amplitude plot, peaks are observed at both 1 Hz and 7 Hz, despite only the frequency 1Hz being present[6] in the data. The reason for this is the phenomenon of aliasing. The maximum measurable frequency in any data series is defined by the Nyquist limit, which is simply half the sampling rate. A useful visualisation of aliasing is shown in Figure 5, in which it is clear that the apparent frequency components of a signal are essentially "mirrored" about the Nyquist frequency, as past this value the apparent frequency decreases.

The second half of the double sided DFT plot can then be discounted, as it is past the Nyquist limit of 4 Hz for this example (Note the plot is symmetric about this point). In practice we double the amplitude of the first half and normalise to the total number of samples to obtain meaningful results.

[6] Of course mathematically, the 7 Hz signal is "present", but this is a result of the two-sidedness and symmetry of the DFT plot

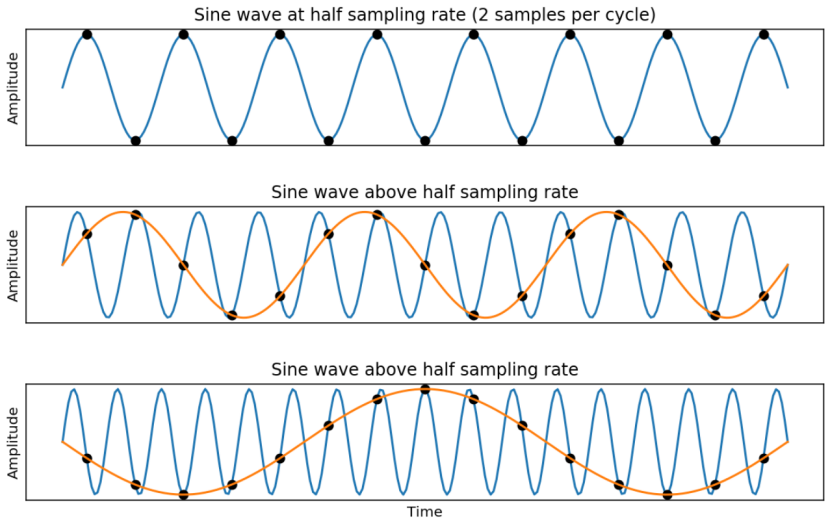


FIG. 5: *A demonstration of aliasing for constant sampling rate. An increasing frequency signal (blue) past the Nyquist limit is read as a decreasing frequency sinusoid.*

4. The Fast Fourier Transform

The FFT is simply an algorithm for the computation of the discrete Fourier transform which reduces the order of complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log n)$. It accomplishes this via a divide and conquer algorithm which takes 1 signal of n points and divides it into two signals of $n/2$ points. This can be repeated for each of these two new signals and so on until n signals of 1 point are obtained. At this point the summations involved in the computation of the DFT are simply one term, greatly reducing computational load. The "splitting" of the signal's DFT takes place via the summation index, as in figure 6.

Simply splitting the FFT into two does not reduce computational load, to do this we employ the so-called "symmetry identity" of Equation 5 in which m is a positive integer. This identity tells us that the exponential present in both summations repeats itself after the frequency index k reaches $N/2$, meaning we need only compute half the X_k.

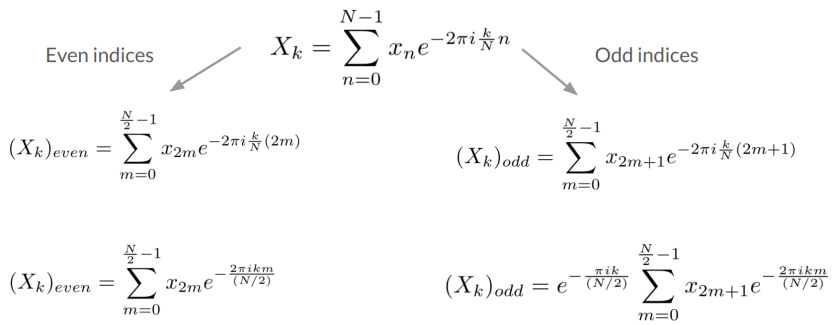


FIG. 6: *The splitting of a DFT by even and odd index for use in the FFT*

$$e^{-\frac{2\pi i(k)m}{(N/2)}} = e^{-\frac{2\pi i\left(k+\frac{N}{2}\right)m}{(N/2)}} \tag{5}$$

This is the power of the FFT algorithm; through divide and conquer methods the total number of operations can be greatly reduced, which lead to it being called "the most important numerical algorithm of our lifetime"[7].

5. Power Spectral Density & Amplitude Spectral Density

As was mentioned in Section 3, the DFT yields N complex numbers X_k which represent the amplitude and phase of the k^{th} frequency bin (Frequency bin and not specific value because the resolution in frequency space is simply the sampling frequency divided by the total number of samples N). In many physical applications, the phases are of relatively low importance, and it can be very useful to examine only the power and its variance with frequency. This leads to the concept of the "power spectral density", which describes the distribution of the power content of the signal over the frequencies present. The computation of the power spectral density follows simply from the DFT components as in Equation 6. The notation used for the PSD here is $\phi(f)$ in line with that employed by P Stoica & R Moses in "Spectral Analysis of Signals" (2005).

[7] Strang, Gilbert (MayJune 1994). "Wavelets". American Scientist. 82 (3): 250255

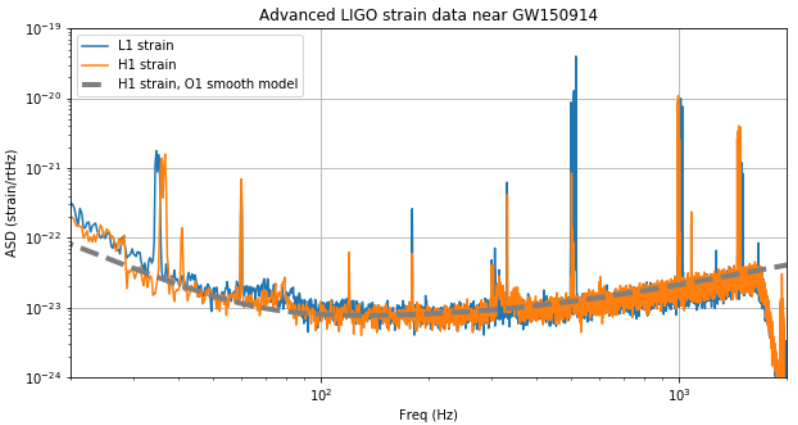


FIG. 7: *ASD for GW150914 from both detectors. The limits 20 Hz and 2000 Hz are such because advanced LIGO is not sensitive to frequencies below 20 Hz and the sampling rate is 4096 Hz, thus 2000 Hz is just below the Nyquist limit.*

$$\phi(f) = \left\langle |X_k|^2 \right\rangle \tag{6}$$

Noting that the power spectral density contains information about the square of the magnitude, we introduce the "amplitude spectral density", which is simply the square root of the PSD. Shown in Figure 7 is the ASD of the signals from both detectors and an approximation with an analytic expression for use in later signal processing. The ASD is an approximation of the strain-equivalent noise as a function of frequency, and thus to obtain the RMS noise for a given frequency band, the ASD need only be integrated over that band.

There are many sharp peaks in the ASD, many of which are at harmonic frequencies. This is no coincidence, as these lines correspond to the noise due to the modes of vibration of the cables suspending the test masses of the interferometer. It is also clear that in the frequency band of roughly 80 to 300 Hz the ASD undergoes a minimum, and therefore this is the most sensitive band. We now use the ASDs to more accurately examine the signal content that is not attributable to noise.

6. Whitening

"Whitening" a signal refers to transforming the signal such that it has a relatively flat PSD (and ASD) as is the case with purely white noise. This is accomplished by dividing the signal by its ASD in the Fourier domain, which has the effect of suppressing the extra noise at low frequencies and at the peaks so we can better see the signal in the most sensitive band. One of the greatest strengths of whitening is that it depends only on the ASD of data, which depends only on the data itself. No knowledge of what is being measured is required (although it helps to know what you're doing) and these processes can always be carried out.

In order to visualise the frequency content of the data over some time interval before and after whitening, we create a *spectrogram*; a 3D plot where the third dimension (either a third axis or a colour spectrum) represents the "amount" of a certain frequency which is present at a certain time. In order to create a spectrogram, we first divide the time signal into an appropriate number of section, then apply a windowing function to each of these sections (such as the Hamming window, which would be suitable here). We apply a windowing function to account for fact that the DFT assumes a signal repeats periodically to infinity, and so if the start and end of a particular section do not match up, a discontinuity in the repeated signal will cause the DFT to find many undesired high frequency signals.. The choice of windowing function is important, as the goal is to minimise the effect of *spectral leakage*, which is discussed in detail in by P Stoica R Moses in "Spectral Analysis of Signals" (2005).

With sectioned, windowed data, the FFT of each data segment is taken and this frequency content of each "time slice" is plotted together, resulting in a plot such as that of Figure 8 for the unwhitened data from GW150914. The spectrogram of data from both detectors is almost identical, and so only the Hanford detector data is shown here. Note the harmonic frequencies present at all times, which again correspond to the vibrational modes of the suspension cables.

By whitening the signal and focusing on the small irregularity around $t = 0$ in the frequency band between 0 and 300 Hz (which is our GW event), we obtain Figure 9. This "chirp", named so because the dominant frequency increases with time, has been identified as interesting and must now be examined.

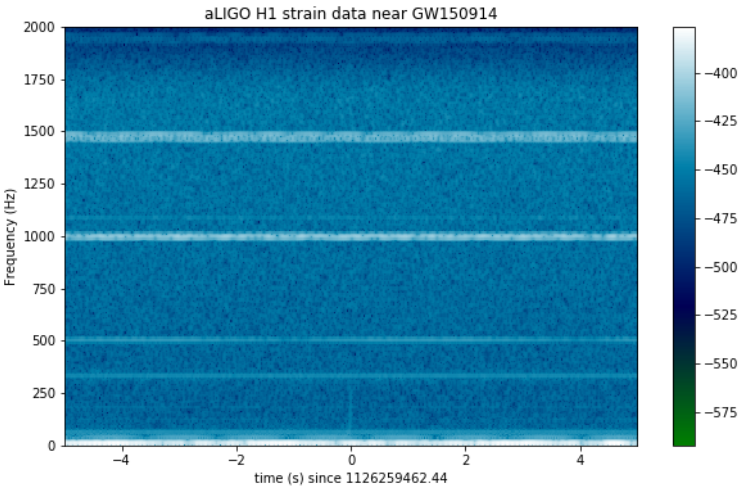


FIG. 8: *Unwhitened spectrogram of Hanford detector data*

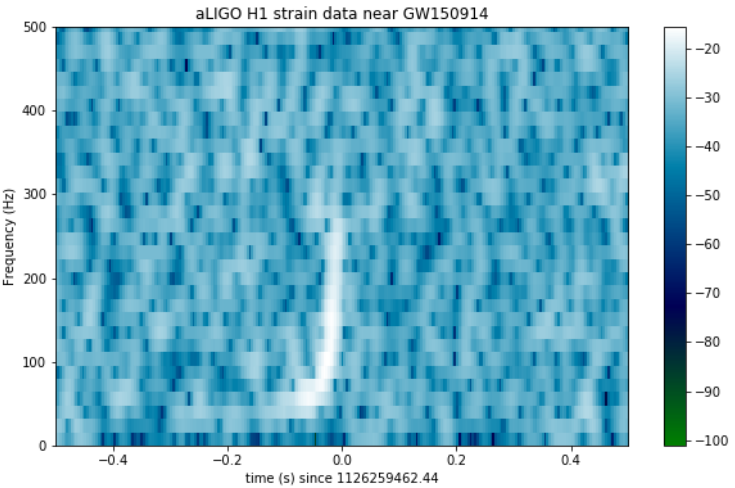


FIG. 9: *Irregularity in whitened Hanford detector signal*

7. Comparison to Theory

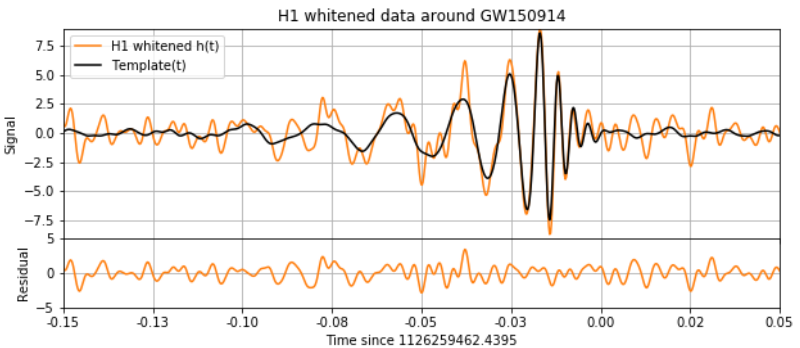
We must now examine if the frequency content of this signal matches what we would expect from general relativistic models. As mentioned in Section 2.1, these equations are very complicated and thus can only be solved numerically in the most symmetric cases, meaning that to model the GW signal we expect to see due to a specific event we employ numerical methods. Numerical relativity is a huge field which will not be discussed in detail here, rather it's results relevant to advanced LIGO; thousands of waveform templates which describe the GW signal we would measure on Earth for different astronomical events. In the advanced LIGO "data pipeline" after generic waveform searches identify candidates for GW events, the signal is compared to all of these templates with different parameters which is incredibly computationally intensive. This cannot be achieved on a consumer-grade computer and so we skip many of the subtleties and use the waveform template identified to be a match by this pipeline[8].

To match this template waveform to our signal, we must first whiten the template using our ASD calculated from the raw data. In essence, this approximates what would be read by this detector if the exact template waveform passed through it. After whitening the template, it can be superimposed on the event and the residual between the two calculated as in Figures 10a and 10b. Evidently, this was repeated for the Livingston detector and an equally close relationship was obtained.

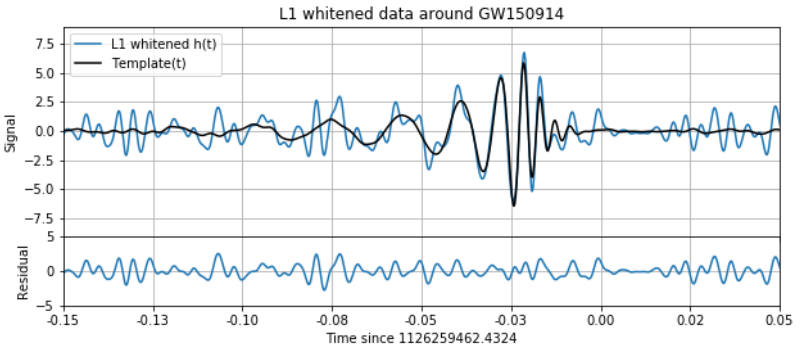
With this relatively simple analysis we obtain a close fit to the data around the time of maximum amplitude $t_{\max}$ in the range $t_{\max}{}^{+0.015}_{-0.05}$ seconds, which is the region in which the signal became of high enough amplitude to be detectable (ie. a sufficient signal to noise ratio). It is noteworthy that the event was not detected at the same time by both detectors, but in fact roughly 8 ms apart. This is because the distance between the detectors is such that a wave propagating at the speed of light takes 8 ms to travel from Livingston to Hanford[9]. Also note

[8] The data pipeline is described in full and GW analysis packages are publicly available from LIGO, the packages which were actually used to create the results used in LIGO publications.

[9] Provided the Earth is transparent to such a wave



(A) *Hanford detector data and whitened numerical relativistic template*



(B) *Livingston detector data and whitened numerical relativistic template*

FIG. 10: *Note the difference in scale of the residual and signal (roughly a factor of 2) which makes the residual appear slightly smaller than it is. The residual is calculated as the template subtracted from the measured data. The signal and residual axes are presented in units of noise standard deviations.*

that the amplitude of the event is smaller for the Livingston detector than that for the Hanford detector. The reason for this is simple, both detectors do not have the same orientation and thus the same gravitational wave signal will be at a different angle to each, thus exerting a different strain on the arms of the interferometer.

This is not a 5σ confidence detection as the result of LIGO publications is. Their data pipeline can produce very confident results through techniques like waveform template mixing which were not investigated here. However, by comparing the initial data to the whitened signal the power of signal processing becomes clear. Even by looking at an unwhitened spectrogram with appropriate time resolution and windowing provides insight into the event that took place. This simple analysis reasonably recreates one of the most important detections of the 21^{st} century, with nothing more than basic signal processing and common, publicly available python packages.

List of Figures

Originally Produced Figures

Public Domain Figures

[10] GW detector data obtained from LIGO public data release, figure produced using code based largely on the tutorial also distributed by LIGO at https://losc.ligo.org