J.N. Andrews
Honors Program
Andrews University

Proposal for Senior Honors Thesis

HONS 497 Senior Honors Thesis **Credits**____ 2______ (2 minimum required)

Directions: Please return signed proposal to the Honors Office **at least one week prior to your scheduled meeting with the Honors Council**. This proposal must be accepted by Honors Council the semester before presentation.

Student's Name: Kelsey Rook

Primary Advisor: Dr. Tiffany Summerscales

Secondary Advisor:

Thesis Title: Finding Optimal Input Parameters for BayesWave

Local phone: (607) 349-4781 Email: rook@andrews.edu

Expected date of Graduation: May 2020

1. Provide goals and brief description of your project or research.
2. Outline your methodology. **Please be specific**. How does this achieve your goals and how reliable is it?
3. Explain in what sense your project is original, unique, or beyond normal senior expectations. How does it relate to current knowledge in the discipline?
4. Include a substantive annotated bibliography of similar or related work.
5. Provide a statement of progress to date and list the research methods coursework completed.

Department Chair Approval

- **This student's performance in his/her major field is acceptable.**
- **He/she has completed the requisite research methods coursework for the research to be pursued.**
- **I understand that he/she plans to graduate with Honors.**

William D. Wolfe

Department Chair (signature)

Research Advisor Approval

I have read and support this proposal: Tiffany Summerscales Primary Advisor (signature)

I have read and support this proposal:______________________________ Secondary Advisor

(signature)

If human subjects or if live vertebrate animals are involved, evidence of approval from the Institutional Review Board or an Animal Use Committee is needed through the campus scholarly research offices (Ext. 6361).

Goals and Description

LIGO, which stands for the Laser Interferometer Gravitational-wave Observatory, consists of two large-scale detectors situated 3,002 kilometers apart in Livingston, Louisiana, and Hanford, Washington. This large-scale physics experiment measures the effects of gravitational waves from various astronomical events on beams of light passed through each facility's perpendicular 4-kilometer arms (Figure 1), revealing the nature of the phenomena originating the waves.

The measurement of gravitational waves by the LIGO observatories is made possible by our knowledge of general relativity, in which gravity is explained as the effect of the curvature of spacetime by different masses. The movement of massive objects causes changes in spacetime, which ripple outwards as gravitational waves. Because gravitational waves stretch and strain spacetime orthogonally to the direction in which they are propagating, their magnitude can be determined by measuring the difference between the lengths of two perpendicular, equally long objects. Each LIGO observatory monitors the minute changes in distance between massive mirrors suspended in both arms and uses the interference patterns of the light bounced back and forth between them to provide information on the direction and magnitude of gravitational waves (Abbott, 2009). Because the gravitational disturbances are so small in magnitude after travelling up to several billion lightyears to reach the Hanford and Livingston observatories, the instruments must be extremely sensitive in order to detect gravitational wave strains smaller than one part in 10^{21}.

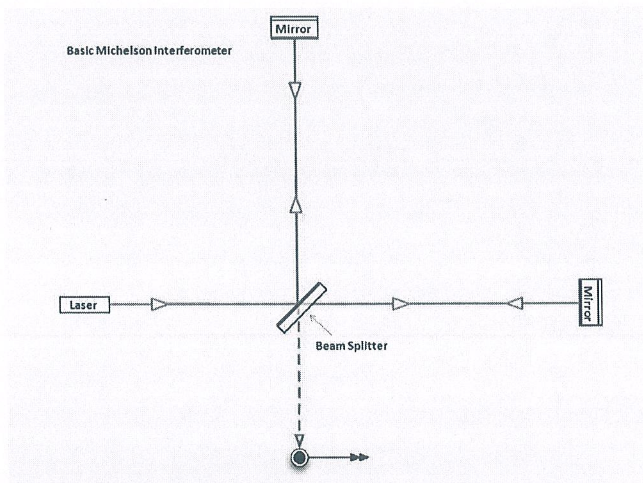

Figure 1: A basic interferometer [1]

While the Advanced LIGO detectors have utilized improvements such as stronger lasers and more sophisticated optics systems, heightening sensitivity enough to detect a change in arm length 1/10,000 the width of a proton (Abbott,), gravitational wave signals are still very faint compared to the environmental noise picked up from nearby traffic, environmental disturbances, and even ringing phones (Berger, 2018). As a result, developing the technology to unbury gravitational signals from the surrounding noise is just as important in characterizing wave sources as ensuring that the detectors themselves are sensitive enough to pick up the signals. Separating wave signals from noise is possible when well-defined signal models are available for the wave transient, but separating an un-modelled signal from noise necessitates being able to characterize the noise. The best understood sources, such as inspiraling neutron star pairs, have accurate computer models and predictable effects, allowing us to easily extract the GW signal, while less-understood sources such as supernovae explosions have only been modelled imperfectly, if at all. For this reason, BayesWave puts as much emphasis on modelling noise as modelling signals (Cornish, 2015).

Gaussian noise is relatively simple to model and isolate from other signals, as it is stationary and purely random. The noise picked up by the LIGO detectors, however, is difficult because of its non-Gaussian and non-stationary nature. Whereas

Gaussian noise would be purely random, detector noise is more structured and shows more variance in power for different frequencies, an example of which can be seen in Figure 2 below where actual LIGO noise data is represented by the irregular gray line. Glitches, which are short bursts of noise, introduce even more variations that also have to be included in the noise model. BayesWave is able to take detector data and express it as a wavelet transform, or sequence of wavelets, as represented by the solid black line in Figure 2. A wavelet is a mathematical function representing a wave-like oscillation that begins and ends at zero, with an average value of zero. Wavelets allow us to describe signals with a relatively small amount of information, since a signal can be expressed as a series of the same wavelet function with different coefficients that shift and scale each copy of the wavelet. BayesWave, in particular, models instrumental glitches and gravitational wave (GW) bursts as a series of Morlet-Gabor wavelets, a form of tapered sine wave (Figure 3). Ultimately, each signal will be modelled as a linear combination of a GW signal, Gaussian noise, and glitches.

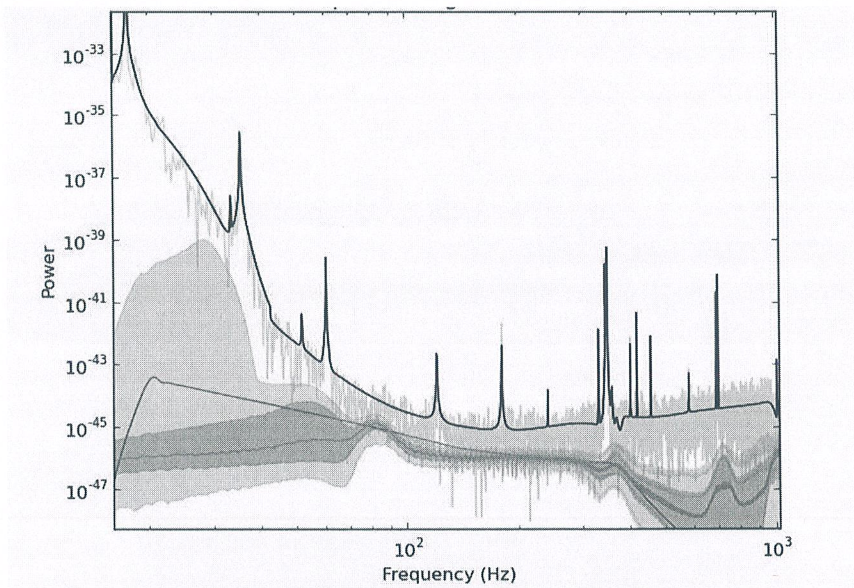

Figure 2: Sound curve for glitch model of LIGO data

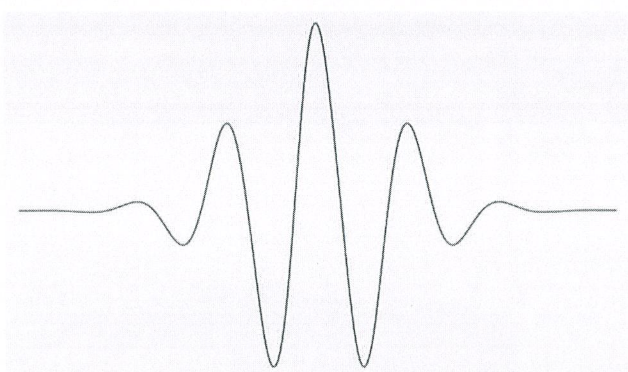

Figure 3: Morlet-Gabor wavelet

To test the probability of a signal being a gravitational wave or a glitch, we must establish a prior, or initial guess, of the amplitude of the wavelets, which doesn't take into consideration the values of our observed data. We can express the prior in terms of the signal-to-noise ratio (SNR) of the wavelet. From Cornish et. al. (2015) we see that BayesWave expresses the glitch amplitude prior as

$$p(\mathrm{SNR}) = \frac{\mathrm{SNR}}{\mathrm{SNR}_*^2}\, e^{-\mathrm{SNR}/\mathrm{SNR}_*} , \tag{1}$$

and the signal amplitude prior is chosen to be

$$p(\mathrm{SNR}) = \frac{3\,\mathrm{SNR}}{4\,\mathrm{SNR}_*^2\,(1 + \mathrm{SNR}/(4\,\mathrm{SNR}_*))^5} . \tag{2}$$

Both priors peak where SNR = SNR*, so changing the value of the parameter SNR* changes the value of the glitch and signal prior peaks.

Using the prior distributions and a likelihood function which expresses the likelihood of each parameter value for a given sample of data, BayesWave can compute the posterior distribution function, which is defined as

$$p(\mathbf{h}|\mathbf{s}, M) = \frac{p(\mathbf{h}|M)p(\mathbf{s}|\mathbf{h}, M)}{p(\mathbf{s}|M)} , \tag{3}$$

where $p(\mathbf{h}|\mathrm{s}, M)$ is the prior distribution for the signal or noise model M as defined in equations (1) and (2), $p(\mathrm{s}|\mathbf{h}, M)$ is the likelihood function, and $p(\mathrm{s}|M)$ is the evidence for the model M. BayesWave uses the posterior distribution to test modelled data samples against and calculate the probability of the data corresponding to one of the following three models: Gaussian noise alone, Gaussian noise with glitches, or Gaussian noise with a GW signal. BayesWave can then use those probabilities to guess whether each sample of data contains noise, glitches, or a gravitational signal from a particular source. BayesWave generates helpful visualizations of the signal model (Figure 4) and and signal evidence (Figure 5) for each sample of data, as well as a skymap of the signal's possible source location (Bécsy, 2017), as in Figure 6 below.

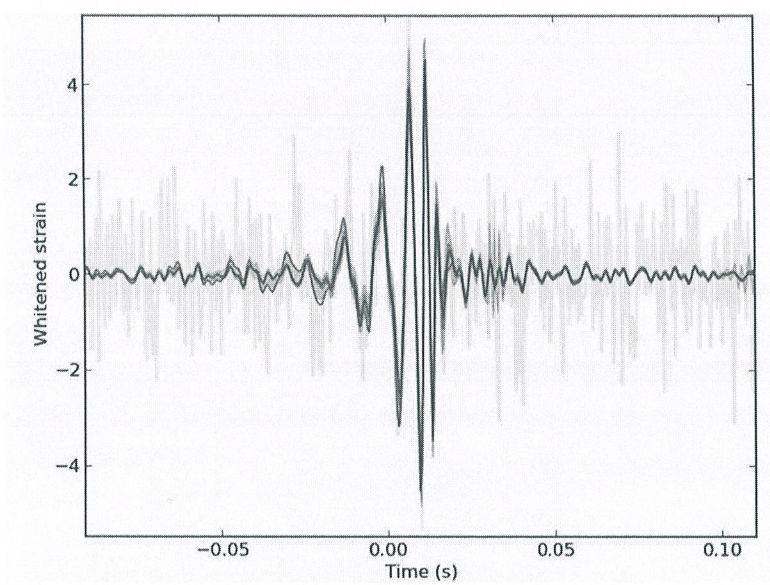

Figure 4: Reconstructed signal model from Hanford, WA observatory

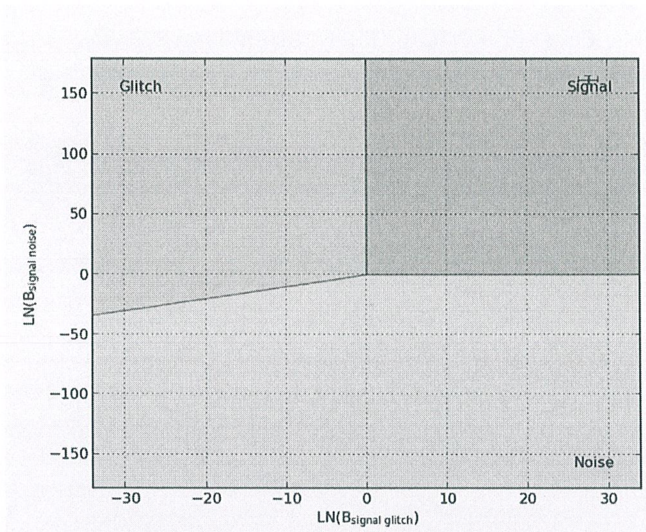

Figure 5: Graph of evidence for signal [3]

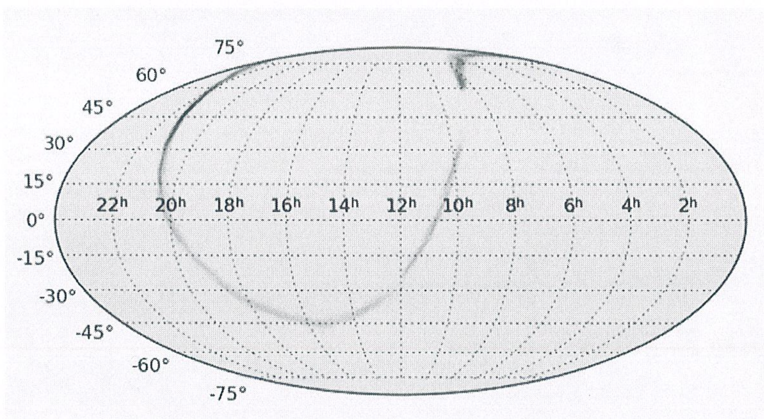

Figure 6: Probability density function of GW source location represented as a skymap

The goal of my research involves injecting binary black hole waveforms to LIGO noise and running BayesWave with different combinations of amplitude and signal prior peaks to determine what combination of parameters allows the algorithm to do the best job of separating noise from gravitational signal. By using various data collection, analysis, and statistical methods, I can determine which combination of parameters leads to the best results in the classification of LIGO data as noise or signal.

Methodology

To run the BayesWave pipeline on many signals at once, I use a workload management software system that allows me to queue jobs in batches, or large groups without having to submit each iteration of the algorithm by hand myself. For each parameter combination, BayesWave is run on two sets of detector data. The first set is LIGO noise injected with binary black hole (BBH) signals, and the second contains only noise or possibly a glitch, which refers to a short, distinct burst of noise. Once BayesWave finishes running on a job, it will create an output file containing values for predicted signal, glitch, and noise likelihoods. Because I know what sort of GW signal, noise, or glitches are in the signal data I am running BayesWave on, I can compare the actual contents of the signal to what BayesWave guesses they contain, and compare the accuracy of the BayesWave algorithm over each combination of amplitude and signal prior peaks using various statistical measures. Ideally, BayesWave will classify binary black hole data as containing a GW signal, and the data not containing a signal injection as a glitch or noise.

To determine the success of BayesWave in classifying data, I will first create a confusion matrix for each parameter combination. A confusion matrix is a table comparing how many samples are actually in each class with which samples the classifier predicted to be in each class— the classes, in this case, being signal, glitch, or noise. From the confusion matrix we can very quickly and visually assess the success of the classifier, BayesWave, and get counts for true positives, false positives, true negatives, and false negatives, measures which are foundational building blocks used to evaluate classification models, for each class. In a classifier where one class is defined as the positive class, a true positive is an outcome where the classifier predicts the positive class correctly, and a false positive is an outcome where the classifier predicts the positive class incorrectly. Similarly, the classifier correctly predicting that the positive class is not the case is a true negative, and predicting any class other than the positive class incorrectly is a false negative. For example, where the positive class is defined as data containing a GW signal, correctly identifying a signal would be a true positive, classifying noise as a signal would be a false positive, correctly identifying noise or glitches containing no signal would be a true negative, and classifying a signal as noise alone would be a false negative. Using the values in each confusion matrix, I will compute the F1 score from the results of each parameter combination. The F1 score is a metric that computes the success of a classifier with the formula

$$F1 = 2\left(\frac{Precision*Recall}{Precision+Recall}\right).$$

In other words, the F1 score is a balance between precision and recall scores for a classifier, where precision calculates the percentage of results the algorithm correctly classifies (see Equation 4), and recall calculates the percentage of relevant instances the algorithm is able to identify in a dataset (Equation 5).

$$precision = \frac{true\ positives}{true\ positives\ +\ false\ positives}$$

$$recall = \frac{true\ positives}{true\ positives\ +\ false\ negatives}$$

Relevance and Uniqueness

The LIGO Scientific Collaboration (LSC) seeks to detect and characterize very powerful events using gravitational waves rather than electromagnetic radiation, which has historically been the most used method of obtaining information about the cosmos. LIGO gives scientists a completely new, unique way of observing the universe's various phenomena, especially those that aren't easily studied using electromagnetic radiation, which includes visible light, X-rays, etc., particularly black holes. While Einstein's general theory of relativity was published in 1916, the detection of gravitational waves confirmed the last non-experimentally verified prediction of his theory. Parameter estimation and gravitational transient modelling is a logical next step after detecting and collecting data from LIGO, and the modelling and isolating of gravitational wave signals from glitches and noise is an important step in determining the characteristics of the astronomical events LIGO studies.

Annotated Bibliography

Abbott, Benjamin P., et al. "GW150914: The Advanced LIGO detectors in the era of first discoveries." *Physical review letters* 116.13 (2016): 131103.

On September 14, 2015, Advanced LIGO made the first detection of gravitational waves after being online for only two days. This paper by Abbott et. al. describes the technical upgrades of the Advanced LIGO detectors as well as its success in increasing its strain sensitivity. Engineering improvements included improved mirror suspensions to decrease the effect of underground vibrations and improved optical power to maximize the detector's sensitivity to changes in resonance. These and various other improvements allowed Advanced LIGO to outperform all previous LIGO runs within the first 16 days of observation.

Abbott, B. P., et al. "LIGO: the laser interferometer gravitational-wave observatory." *Reports on Progress in Physics* 72.7 (2009): 076901.

This paper by Abbott et. al. gives a relatively high-level description of the history, goals, and technical aspects of LIGO. Abbot et. al. give an overview of the science behind general relativity and gravitational waves, as well as an overview of the history of the worldwide detector network with a focus on LIGO. The paper goes on to provide a technical description of the observatory's configuration, including its optics, mirror suspensions, and sensing and controls, followed by the computing infrastructure and data analysis that complements the engineering. Next, the paper outlines LIGO's success in the search for four categories of GW sources: Transient modelled waveforms, which include the final stages of binary black hole or neutron star inspirals; transient unmodelled waveforms, including supernovae and black hole mergers; continuous narrow-band waveforms such as pulsars, and continuous broad-band waveforms, which refers to a random background of GWs which may have originated at the beginning of the universe. Finally, the paper provides a description of LIGO's future plans, including Advanced LIGO, which was, at the time, proposed to be constructed two years later and promised an improvement in sensitivity of a factor of ten..

Bécsy, Bence, et al. "Parameter estimation for gravitational-wave bursts with the BayesWave pipeline." *The Astrophysical Journal* 839.1 (2017): 15.

Finding the sky location of gravitational wave sources plays a key role in being able to conduct follow-up observations on GW detections. In this study, Bécsy et. al. tested the performance on Advanced LIGO data using similar methods and datasets to previous tests on WaveBurst (WB) and LALInferenceBurst (LAL), two other burst pipelines, with a particular focus on sky localization. Their results showed that BayesWave performed with comparable accuracy on Sine-Gaussian, Gaussian, white-noise bursts, and binary black hole injections, as well as performing comparably to the WB and LAL pipelines.

Berger, Beverly K. "Identification and mitigation of Advanced LIGO noise sources." *Journal of Physics: Conference Series*. Vol. 957. No. 1. IOP Publishing, 2018.

Efforts to identify and eliminate or account for environmental sources of glitches and noise in LIGO data have taken such a high priority that the LSC has assigned members to data quality (DQ) shifts, which involve monitoring detectors remotely or in person to locate glitches or noise that might be confused with astronomical events or indicate physical issues within the detector. LSC members assigned to DQ shifts are able to use various software tools to view glitch activity, spectrograms, and time series, as well as having access to advanced software that looks for statistically significant correlations between glitches in different channels. This article describes the basics of detector characterization (DetChar) and some of the ways this initiative has been able to find and resolve environmental sources of glitches as noise. In one case, a series of glitches was correlated with seismic activity in one arm of the Livingston, Lousisiana detector and traced back to an air compressor with defective vibration isolation feet. In another, a DQ member thought to create a sound file of a strange series of glitches, revealing a phone loudly ringing followed by an answering machine. The source of glitches and noise in each case were accounted for either by elimination, or by a strong correlation with a specific environmental event through which they could not be mistaken for an astrophysical event.

Cornish, Neil J., and Tyson B. Littenberg. "Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches." *Classical and Quantum Gravity* 32.13 (2015): 135012.

Identifying unmodelled gravitational wave signals amongst strong non-Gaussian noise is a current issue in Gravitational Study. In this paper, Cornish and Littenberg outline how the BayesWave analysis pipeline models and classifies LIGO data samples as glitches, noise, or gravitational wave signals through the process of Bayesian inference. The authors walk through the process of selecting appropriate likelihood functions and prior functions for both glitches and signal, describing the burst models using Morlet-Gabor wavelets, and producing posterior distributions for parameter estimation, and also provide examples of how BayesWave has successfully modelled and classified data samples from both LIGO and Virgo science runs.

Progress to date

While there are no specific research methods courses for Computer Science, I have taken Artificial Intelligence and Machine Learning, which both have given me the background to understand the BayesWave algorithm better as well as teaching me valuable data analysis and statistical tools. Additionally I am in the progress of taking Data Mining and Visualization, which has further emphasized data manipulation, analysis, and visualization techniques. I am also taking Mathematical Modelling in Biology, and I have completed Research Pro-Seminar which is required by Honors.

I began my research during the Fall semester of 2017, and up until Fall of 2018 I worked on running and troubleshooting BayesWave jobs. Ultimately I would like to run BayesWave on sixteen amplitude prior and signal prior peak combinations, but since BayesWave is computationally expensive and often requires troubleshooting, I was only been able to collect, analyze, and visualize preliminary SNR results on 8 parameter combinations from the beginning of my work to Fall of 2018. Since then I have updated to a newer version of BayesWave and have been rerunning jobs, while continuing to explore and perform analysis on the sets of data resulting from my work with the older version of BayesWave. This analysis has included creating graphs of the distribution and mean of signal-to-glitch ratio, as well as percentage of jobs that resulted in positive signal-to-glitch ratios, for each parameter combination. In addition to the computing part of my position, I presented an overview of my work with BayesWave at an American Physical Society poster session during their April 2018 meeting.