First Interim Report: Sifting Through Detector Noise Using Time Series Analysis

Gabriel Grant[1], **Jane Glanzer**[2],
[1]Minnesota State University Moorhead, Moorhead, MN 56563, USA
[2]LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

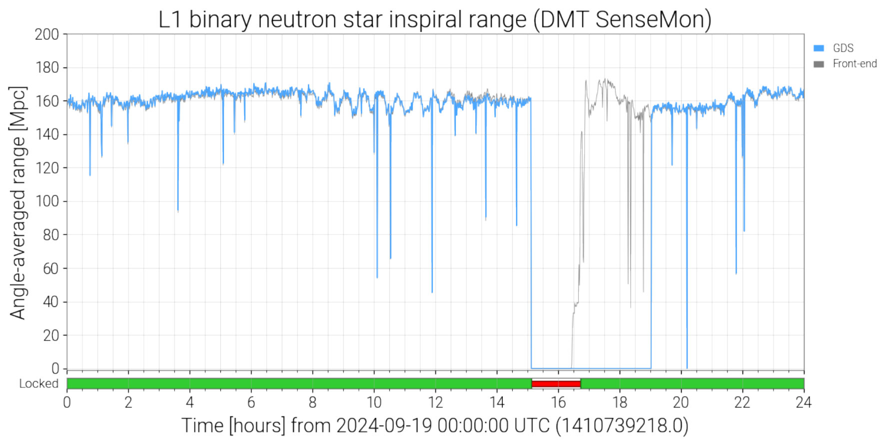
1. Introduction

Gravitational waves were first predicted by Albert Einstein in the early 1900s [1]. Sometimes thought of as "ripples in spacetime", gravitational waves are produced by extreme astrophysical events, like the merging of two black holes or neutron stars. It was not until recently that the first observation of a gravitational wave was made. On September 14, 2015, the Laser Interferometer Gravitational-Wave Observatory (LIGO) facilities in Livingston, Louisiana (LLO) and Hanford, Washington (LHO) observed a merger of two stellar mass black holes [2], which produced measurable gravitational waves.

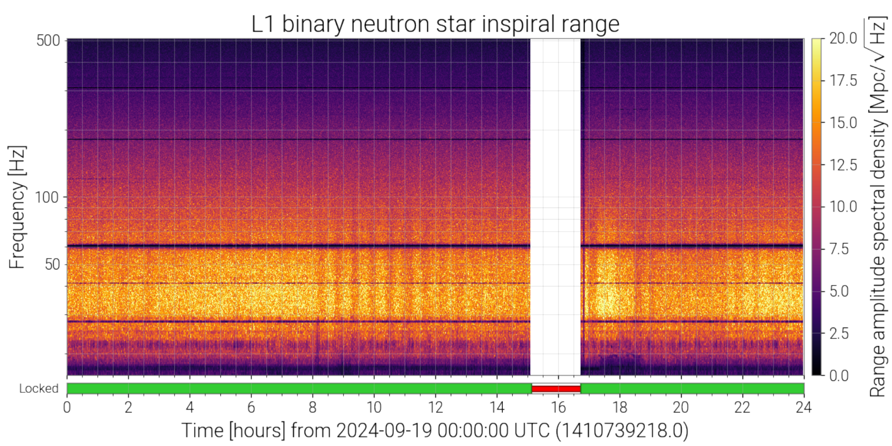
Gravitational waves are difficult to detect because they are incredibly weak. The gravitational wave amplitude, often referred to as the gravitational wave strain (GW), can be on the order of $10^{-21}\,\mathrm{Hz}^{-1/2}$ at 30 Hz [3]. Due to their elusive nature, detection requires extreme sensitivity. Both LIGO facilities operate as modified Michelson interferometers in order to achieve the sensitivity required to detect gravitational waves. Some of these modifications include increased arm lengths, optical cavities, and filter cavities, all of which allow us to reach the precision required to measure space-time fluctuations due to gravitational waves. With such large facilities and cutting-edge precision, the potential for noise is seemingly insurmountable.

The group known as detector characterization is the primary group responsible for understanding how noise can impact the interferometers. In general, detector characterization is the process by which noise sources are monitored, identified, and addressed. This includes studies of LIGO data quality to identify and mitigate noise sources to improve detector sensitivity [4]. There are hundreds of sensors, including accelerometers, microphones, temperature sensors, and seismometers, placed throughout the detector to help track potential noise sources. One of the ways that detector sensitivity is tracked is via the binary neutron star (BNS) range. The BNS range represents the distance at which a gravitational wave signal from the inspiral of two neutron stars with an SNR of 8 can be detected [5].

Since the beginning of the fourth observing run, the BNS range at LLO has had frequent oscillatory behavior with periods of roughly 30 minutes. Oscillations are also observed in many auxiliary channels [6]. The direct coupling of this noise has not yet been identified and that is the main objective of this project. This is important to address, as it significantly affects the range and precision with which we can measure gravitational events. During these oscillations, the BNS range can vary 5-15 megaparsecs in this 30 minute window [4]. These oscillations do not occur consistently, sometimes for part of a day or a whole day, usually disappearing from weeks to months at a time, but do appear most strongly within the 30-50 Hz frequency range. Due to the wide variety of potential noise sources, identifying the channels of these oscillations is difficult.



(a) top: Visualization of the BNS range on September 19th 2024. 30 minute oscillations are visible from roughly 8 UTC to 13:30 UTC on this day.



(b) bottom: Time-frequency spectrogram where the color represents the amplitude spectral density of the BNS range. During the oscillations, we can see ”stripes” corresponding to changes in noise from 30 Hz to 50 Hz.

Figure 1

2. Objectives

The 30-minute oscillations that appear in the main gravitational wave data at a frequency roughly between 30 and 50 Hz at the Livingston LIGO facility are the primary target of study. This project will analyze data over the entire span of the fourth LIGO observing run in which these oscillations are present. Identifying the source or sources of these oscillations requires exploration of many channels and how they interact with each other. Due to the elusiveness of the source, multiple auxiliary channels are believed to be contributing to the behavior. Identifying possible auxiliary channels, or at the minimum, ruling out channels, is the main objective. If significant progress is made, regression algorithms will be performed to model the noise to try to understand the behavior of this noise.

3. Methods

There are numerous auxiliary channels that appear to manifest some form of the half-hour oscillatory behavior. Due to this, it is unclear as to which channels or systems are the primary contributors to the noise that is seen in the main gravitational wave data. Correlation analysis will be used to help identify which channels are driving the noise behavior. Narrowing down the potential auxiliary channels is the goal.

We will identify which auxiliary channels are experiencing this noise, starting with data from a channel such as a temperature sensor or accelerometer, and the main gravitational wave data channel. We then generate a band-limited root mean value (BLRMS) plots of both to compare the auxiliary channel data to the GW data. This is used to visually identify auxiliary channels that mimic the GW data or exhibit oscillatory behavior. Using these, we can identify possible sensor groupings and locations that are experiencing this noise.

Once a collection of auxiliary channels has been created, we will develop a time-lag analysis algorithm, which can determine whether an auxiliary channel precedes or drags behind the noise in the gravitational wave channel by sliding their time series against each other for various time shifts to determine better fits. A preceding channel is more meaningful to us because it may be an initial driver of the noise that we see. If it drags, that means it may couple with another source of noise or witnesses the noise that contributes to the oscillations we see in the gravitational wave data.

Once we have identified a subset of interesting auxiliary channels, we will use regression algorithms to model the noise. Regression algorithms that have been used in the past in various detector characterization studies include Least Absolute Shrinkage and Selection Operator (LASSO) and Ridge regression [7]. It is possible that these regression methods may be insufficient, so other regression techniques may be explored to glean more connections between the channels.

4. Progress

The project initially started by combing through the O4a and O4b data and selecting several days that experience the oscillatory behavior. I settled with three days; August 24th, 2023, April 30th, 2024, and September 19th, 2024. These three days roughly span a year worth of time, and where this behavior was most prominent with some padding to give a baseline of the BNS range before or after the oscillations.

Using GWpy [8], a Python-based package used for GW analysis, I pulled the gravitational wave strain data and auxiliary channel data, in this case accelerometers [4] to begin with, for the same time period within that day. I then created band-limited root mean square plots (BLRMS) to get the data into a form that would make the two signals more easily comparable. BLRMS gives us the magnitude of the data, which is a time series, and also allows us to see if noise is appearing in a particular frequency range. This comparison was made to help identify specific auxiliary channels that mimicked or behaved in a similar manner to the gravitational wave strain. Although channels that mimicked the GW data were subjected to more scrutiny, any auxiliary channel that had periodic behavior at any of the frequencies was noted. Some notable channels include an accelerometer `L1:PEM-EX_ACC_EBAY_FLOOR_Z_DQ`, a temperature sensor `L0:FMC-EX_VEA_302B_TEMP`, and a seismometer `L1:ISI-ETMX_ST2_BLND_X_GS13_CUR_IN1_DQ`.

We computed BLRMS on each auxiliary channel in the 30-50 Hz range, and/or 0.10-0.25 Hz. An example comparison is shown in Figure 2. Low frequency noise was unexplored at the start of this project, but became important to study because it could somehow modulate into the high-frequency range and show up in the strain data. Accelerometers had shown some oscillations in the low-frequency range, but they were not made to accurately measure frequencies that low. To cover our bases, we checked a large list of seismometers in the low frequency range, as they are better geared for low frequencies. The reasoning for why we wanted to check the low frequency range was that there was potential for low frequency coupling. It is possible for low frequencies to upconvert into high frequency noise and we wanted to rule this out. The seismometers did not show much in the way of oscillations at either frequency range. Temperature sensors were also checked, but due to the sensors only being sampled at 16 Hz, an ordinary RMS was taken. Temperature sensors showed some significant oscillatory behavior that matched the strain data very closely. Most of these auxiliary channels that were studied were in the End-X station, as we think that this is where the source of the noise originates.

The next step will be to run a Python program that calculates the cross-correlation between each auxiliary channel and the gravitational wave strain data at different time lags. These time lags will give some insight into noise couplings and locations. The current program does not analyze whether the time lags, if present, are significant. The current idea to determine significance is to phase-randomize the strain data [9]. This creates surrogate strain data that we can cross-correlate with real auxiliary channel data. In this way, we can build a distribution to determine how significant the auxiliary channel correlates to the real strain data. After this script is developed, we will run it over many time periods in which this noise is present in O4.

The primary roadblock is the amount of data that I am needing to sift through and process. The set of accelerometers that I started working with is 20 long, while the temperature sensors have nearly 80, and now working with a list of over 100 seismometers. When starting with the accelerometers, generating all the BLRMS for an observing day at a given frequency range initially took 20 minutes. I had spent a day optimizing the code which brought this down to about 5 minutes. This does not include any additional scrutiny or tweaking of code parameters to get clearer pictures of the aux channel behaviors. Additionally, other noise sources do appear in the data, often making it difficult to work with or automate. When these noise transients appear I will often identify the channels and individually rerun the BLRMS with some code that have some additional clipping mechanisms. These plots are either saved or noted to not have any interesting behavior, which must be done manually to some extent. If more BLRMS are needed in the future, code may be developed to automate this more efficiently.

The current bottleneck is data acquisition. While I am studying the same three days, I have several hundred auxiliary channels that need data for each. I am also looking at several hours worth of data for each channel. Unfortunately, there is no way to speed up the download. The only reprieve is to keep track of what data has been downloaded and keeping the files organized for reuse. This will likely become more of a problem as more channels and days are evaluated. This will become more apparent as large sets of auxiliary channels are used in the cross-correlation techniques. Downloading channels ahead of time in anticipation of their analysis may prove to be a good strategy.

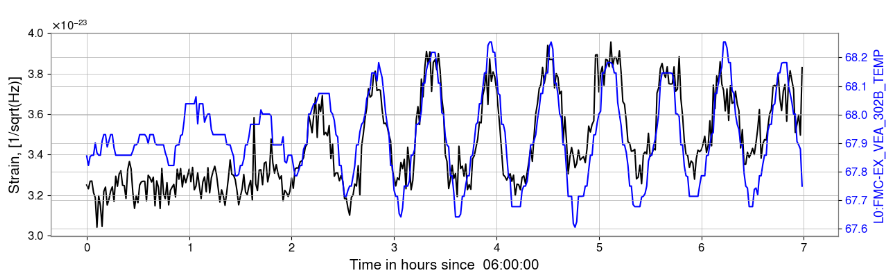


Figure 2: Comparison between a 30-50 Hz BLRMS of the strain data with an end-X temperature sensor.

References

[1] Einstein A 1916 *Annalen Phys.* **49** 769–822

[2] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Phys. Rev. Lett.* **116**(6) 061102 URL `https://link.aps.org/doi/10.1103/PhysRevLett.116.061102`

[3] Glanzer J 2024 *LSU Doctoral Dissertations* URL `https://repository.lsu.edu/gradschool_dissertations/6455`

[4] Soni S *et al.* (LIGO) 2025 *Class. Quant. Grav.* **42** 085016 (*Preprint* `2409.02831`)

[5] Finn L S and Chernoff D F 1993 **47** 2198–2219 (*Preprint* `gr-qc/9301003`)

[6] Glanzer J 2024 aLIGO LLO Logbook 70426

[7] Glanzer J, Soni S, Spoon J, Effler A and González G 2023 *Classical and Quantum Gravity* **40** 195015 (*Preprint* `2304.07477`)

[8] Macleod D M, Areeda J S, Coughlin S B, Massinger T J and Urban A L 2021 *SoftwareX* **13** 100657 ISSN 2352-7110 URL `https://www.sciencedirect.com/science/article/pii/S2352711021000029`

[9] Theiler J, Eubank S, Longtin A, Galdrikian B and Doyne Farmer J 1992 *Physica D Nonlinear Phenomena* **58** 77–94