

LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY  
- LIGO -  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

|   |                  |            |
|---|------------------|------------|
| Technical Note  | LIGO-T1900287-v1 | 2019/07/15 |
| <b>Data Clustering Techniques for the<br/>Correlation of Environmental<br/>Noise to Signals in LIGO Detectors</b> |                  |            |
| Jacob Bernhardt, Anamaria Effler, Rana Adhikari   |                  |            |

California Institute of Technology  
LIGO Project, MS 18-34  
Pasadena, CA 91125  
Phone (626) 395-2129  
Fax (626) 304-9834  
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology  
LIGO Project, Room NW22-295  
Cambridge, MA 02139  
Phone (617) 253-4824  
Fax (617) 253-7014  
E-mail: info@ligo.mit.edu

LIGO Hanford Observatory  
Route 10, Mile Marker 2  
Richland, WA 99352  
Phone (509) 372-8106  
Fax (509) 372-8137  
E-mail: info@ligo.caltech.edu

LIGO Livingston Observatory  
19100 LIGO Lane  
Livingston, LA 70754  
Phone (225) 686-3100  
Fax (225) 686-7189  
E-mail: info@ligo.caltech.edu

## 1 Introduction

The LIGO project uses laser interferometry to measure gravitational waves (GWs). LIGO interferometers transduce their relative arm length differences caused by GWs to a signal composed of optical power, known as DARM. Due to amplitude scales of astrophysical GWs, LIGO detectors have to operate at a very high sensitivity; the spectral density of a measurable length difference is as low as  $2 \times 10^{-20} \text{ m}/\sqrt{\text{Hz}}$  at 100 Hz. The design of earthbound LIGO is thus heavily focused on treatment of environmental noise.

To help identify and characterize environment-based noise, the LIGO detector has a Physical Environment Monitoring (PEM) system, a diverse array of environmental sensors positioned all over the facility. This is used for a multitude of purposes, including the data quality report (DQR) used for time segment vetoing, based on direct coherence of PEM channels to DARM. Supplementing coincidence analysis between the two detectors, DQR prevents GW-like noise transients from being falsely categorized as events. While vetoing and determining signal quality is useful, detector livetime can be increased by figuring out how to decouple environmental noise from DARM. Directly coupling noise, found by basic coherence, has been already addressed, but the complexity of the detector causes many noise sources to up- or down-convert. These require some more careful statistical correlation to identify, and are generally not well understood.

Separating noise sources out of a signal can be considered a clustering problem in a space covering different frequency bands in which noise appears. A previous LIGO SURF student has evaluated several data clustering algorithms with respect to their ability to properly sort out frequency elements of seismometer signals caused by specific earthquake events[1]. Both the  $k$ -means algorithm, which aims to make clusters with low standard deviation, and the DBSCAN algorithm, which minimizes overall inter-point distance in clusters, were evaluated using multiple methods, including the Calinsky-Harabaz index and direct comparison to earthquake times via time labeling of points, ultimately showing poor earthquake identification. A long short-term memory (LSTM) recurrent neural network (RNN) seemed to work much better, but due to small input sample size, this solution may have been plagued by over-fitting. Thus, it is imperative that a more robust frequency clustering mechanism be designed for the PEM system.

## 2 Objectives

- As a primary goal, **algorithms or clustering approaches which correctly identify known noise events need to be found**. As every algorithm has inbuilt assumptions about the dataset it is applied to, the results of an algorithm performance test on labeled data will yield information about the structure of the data. The general temporal non-stationarity of the DARM noise will need to be accounted for by varying testing time windows.
- The secondary goal is to **create a clustering approach to discover previously unknown noise correlations and possibly sources**. This is where the “detector characterization tool” that this project aims to advance will be functional—revealing new noise coupling pathways will help identify ways to improve the detector sensitivity.

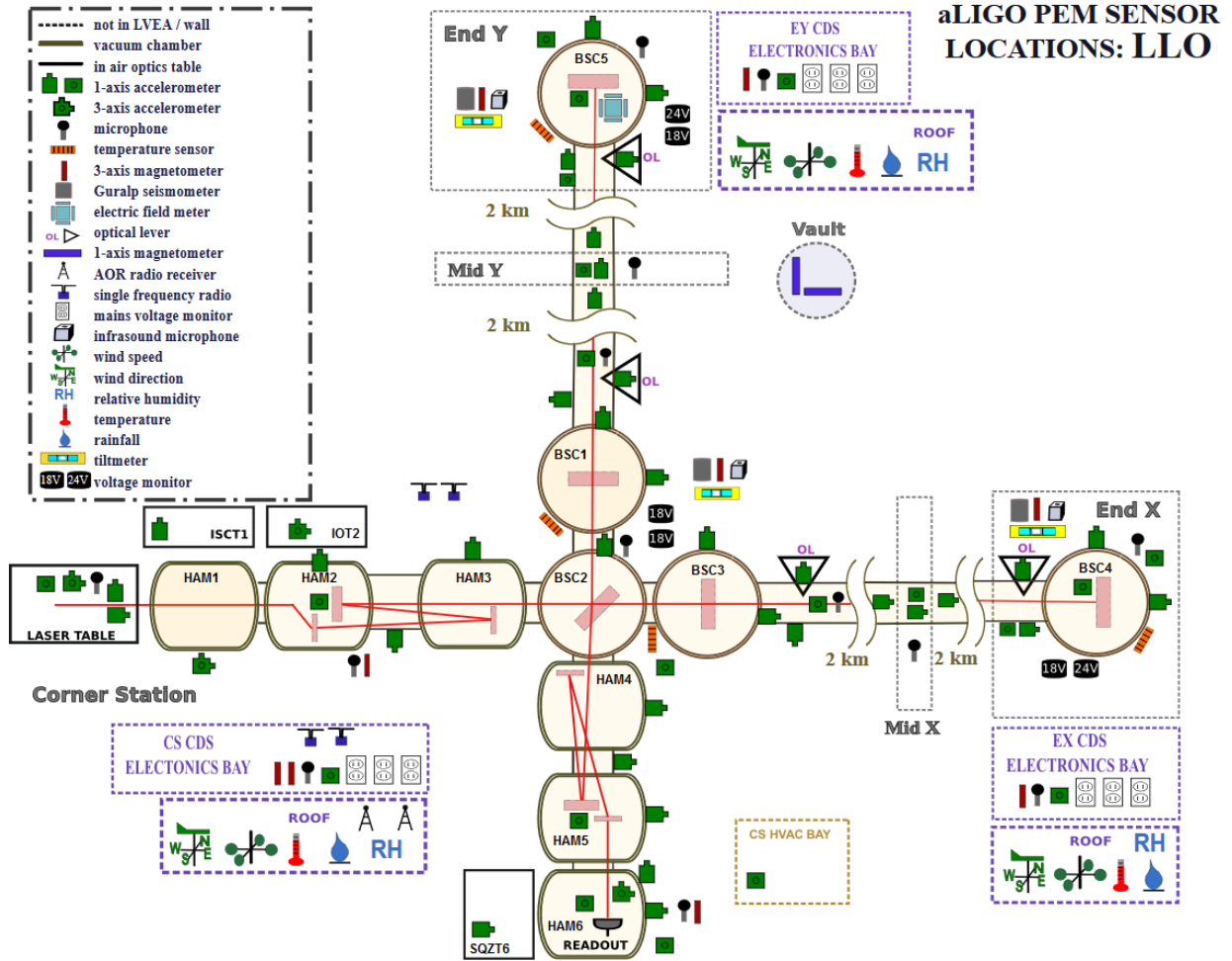


Figure 1: Schematic PEM map at the LIGO Livingston Observatory (L1). Shaded areas are in vacuum.

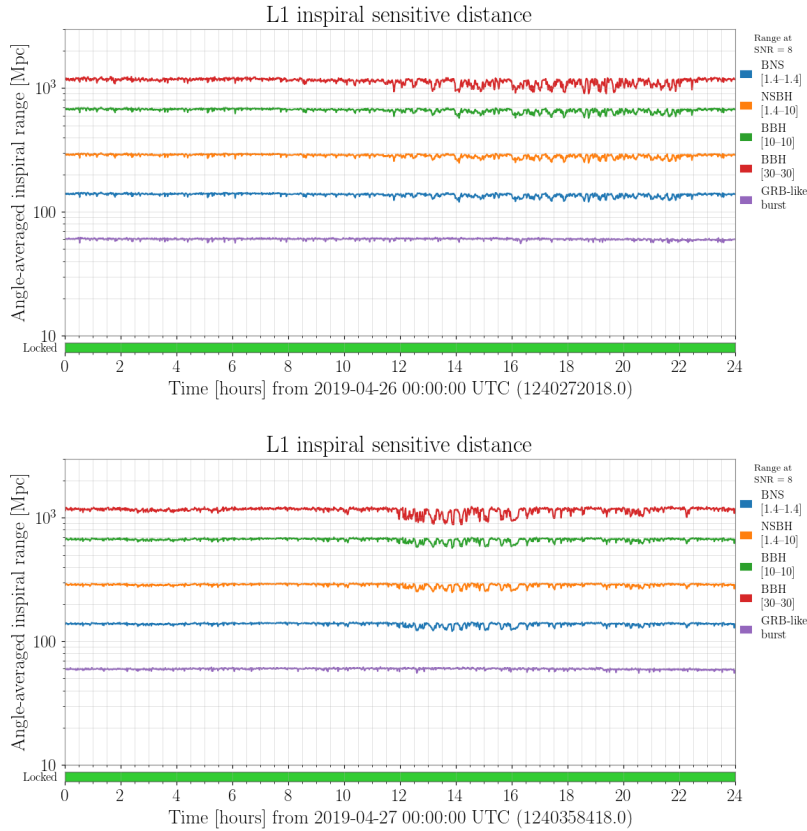


Figure 2: Detector range at L1 seems to consistently reduce during the day ( $\sim 6\text{am}-5\text{pm}$  CST). For large BBH in these plots, the reduction is about 300 Mpc. The source of this has been pinpointed to the Y end station, but the mechanism isn't fully clear.

### 3 Approach

Initially, a program will be written to take the spectral power of any PEM channel, in the form of band-limited RMS (BLRMS), likely using established methods like looping through a smoothed spectrogram of the channel[2].

To reach the first objective, a modular `python` testing suite will be written to probe the structure of the multidimensional frequency-domain sensor data. This will strategically implement `scikit-learn` clustering algorithms and classifiers with different optimal regimes of function or working assumptions and evaluate them using point labeling. This will require, additionally to researching clustering or unsupervised classification algorithms, thinking of as many variables which may affect the data structure (such as looking at different time windows) and intelligently testing them. Optimizations will need to be considered so that run times are reasonable.

The program tackling the second objective will use working clustering approaches identified in the first objective to find new noise correlations. In the event that no individual algorithm or technique outperforms the rest for all types of sensory data, the final program will use the modular programming environment created for the testing suite to match techniques to the regimes that they work in. The structure of the input data as determined by the first objective, including the dimensionality probed by the extra variables, may lend itself to additional algorithms that can be used to combine the target regimes. To this end, extra algorithm research will be conducted with specific consideration of the solved structure.

## 4 Project Schedule

### Before Arrival

- Research algorithms to use
- Develop glitch-resistant BLRMS-making code
- Continue familiarizing with `scikit-learn` by adapting old clustering code for beam-tube accelerometers

### Week 1-2

- Familiarize with types of PEM data that can have labeled events
- Complete the algorithm testing suite

### Week 3-4

- Develop ways to plug channels into the suite
- Run suite on determined variety of channel combinations and regimes

### Week 5-6

- Examine results, decide next steps
- Research new algorithms
- Adapt initial codebase for application

### Week 7-8

- Try to reveal new PEM-DARM coupling mechanisms and propose/implement solutions

### Week 9-10

- Finish up final analysis and writing final report.

## References

- [1] LIGO Document T1700198-v1
- [2] aLIGO LLO Logbook entry 45374 by Gabriele Vajente