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

Technical Note	LIGO-T20003903–v4	2020/12/02

Machine Learning Applications in Gravitational Wave Detection to Reduce Noise

Nadezhda Dimitrova

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

http://www.ligo.caltech.edu/

Abstract

The LIGO gravitational wave detector is contaminated in the sub-60 Hz range by noises caused by coupled auxiliary control channels. One phenomenon of interest is the bilinear angular noise, which is caused by coupling the high-frequency angular motion of the mirror and the low-frequency beam spot motion to create a length signal that mimics a gravitational wave. A dense neural network successfully identifies the low-frequency component and it serves as a critical first step of removing the bilinear noise from the real data. Minimizing the instrumental artifacts is crucial and paves the way toward a much deeper understanding of gravitational waves.

1 Introduction

1.1 Background

With two Michelson laser interferometers placed at the ends of 2.5 mile long arms, the laser interferometer gravitational-wave observatory (LIGO) is the largest earth-based gravitational wave detector. At the intersection between the perpendicular arms, a photo detector observes fluctuations in the induced differential arm length (DARM) signal [1]. These fluctuations are due to the strain, or bending, of space-time that creates optical paths of different lengths. The LIGO observatory analyzes these fluctuations to infer the nature of the physical phenomena causing them, such as colliding neutron stars or black holes [3]. The advanced LIGO (aLIGO) project aims to revise initial design and eliminate excess noise that pollutes observations.

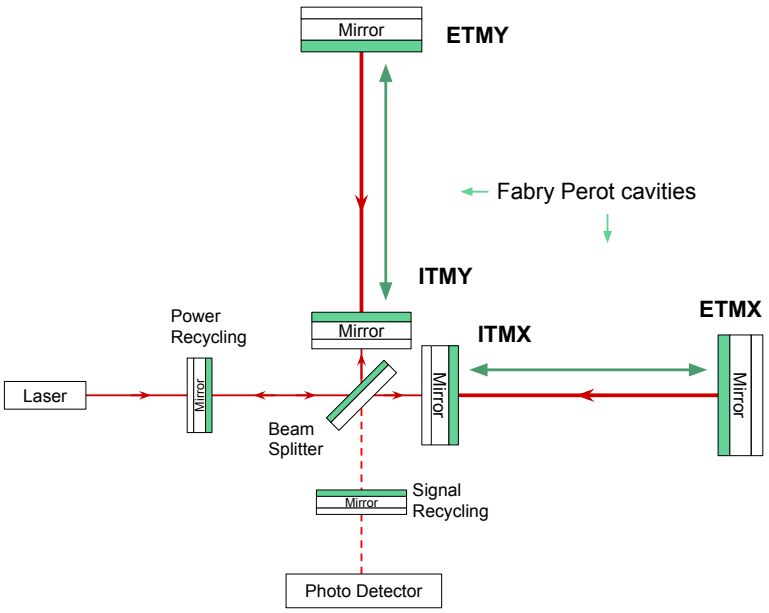


Figure 1: Schematic design of aLIGO optical system: a dual recycled Fabry-Perot Michelson interferometer

Invented in 1880s by American physicist Albert A. Michelson, the Michelson interferometer is at the core of LIGO's design. As illustrated in figure 1, the instrument splits a light beam into two parts and recombines them after travelling different optical paths between mirrors - input test masses (ITM) and end test masses (ETM) [6]. Each separated beam enters a Fabry Perot cavity and bounces between its two additional mirrors about 300 times before being merged with the beam from the other arm. This modification increases sensitivity to produce an interference pattern that allows the measurement of tiny displacements [7]. Power and signal amplification is created through recycling optical systems, while ground motion is reduced by suspending all instruments.

1.2 Motivation

In order improve aLIGO's sensitivity. Any disturbance that limits the sensitivity is defined as noise. Noise due to fundamental infrastructure limits, such as thermal and quantum fluctuations, is unsubtractable. The subtractable noise is caused by environmental perturbations and technical noise.

Despite the elaborate seismic isolations through mechanical pendulums and springs, the residual motion of the mirrors is large and a control system is implemented to manage resonance. However, this feedback loop can introduce sensor noise into the DARM. The problem is that feedback loops can be tricked to account for noise introduced by the sensors or actuators into the DARM main readout signal, as shown in figure 2. Often this contamination is nonlinear and hard to be removed by the traditional linear techniques [2]. That is why some gravitational wave emitting objects are still lying below the sensitivity limit of the detector. The physical phenomena that can be observed with a higher sensitivity range include higher mass binary black holes, early warnings of binary neutron star mergers, and cosmology with high red-shift sources.

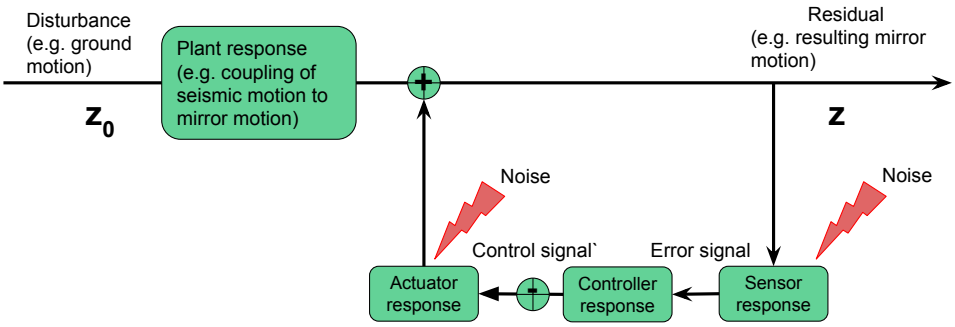


Figure 2: Noise coupling into the control feedback loops of aLIGO: Starting from the left, a ground motion disturbance couples with the motion of a mirror. The motion is observed by sensors that transmit the information to a controller, which in turn commands actuators. The goal is to reduce undesired noises by compensating for their effects on the residual signal.

2 Theory and Related Work

2.1 Feedback Control Loop

The control loops in aLIGO maximize accuracy through a complicated network of measurement tools shown in figure 3. The data from these sensors is then combined with the main readout through a control network illustrated in figure 4. To combine information and reduce error and noise, control loops transfer witnessed data between different instruments.

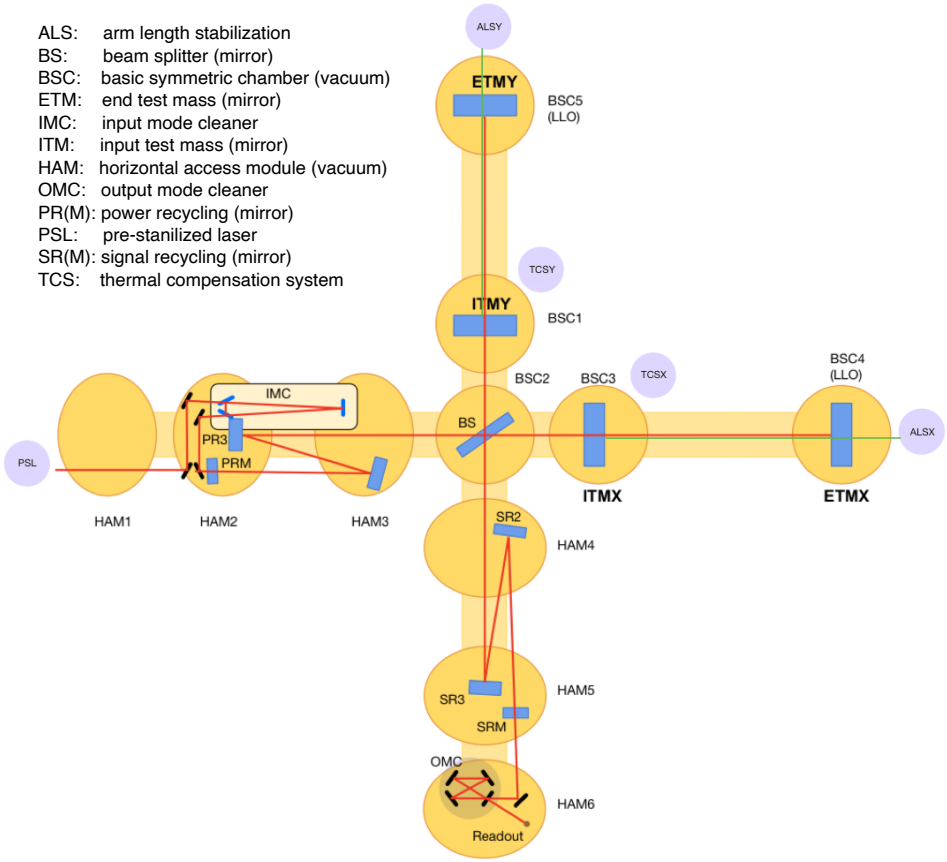


Figure 3: Overview of the aLIGO optical configuration and glossary: All vacuum chambers (yellow circles) contain active seismic isolation (SEI) platforms that support the optics (blue rectangles), their internal seismic isolation (ISI) and suspension (SUS) structures. The red lines represent the path of the primary laser beam, while the green lines represent the laser beams for the arm length stabilization (ALS) system. Illuminating the input test masses (ITMX and ITMY) is the thermal compensation system (TCS) [5].

2.2 Noise Sources and Current Efforts

The current regression method employed by aLIGO minimizes the squared error between the DARM main readout channel and the predicted noises from the physical environmental monitor (PEM) channels [4] - all the sensors. Nonetheless, this Wiener-Kolmogorov filter

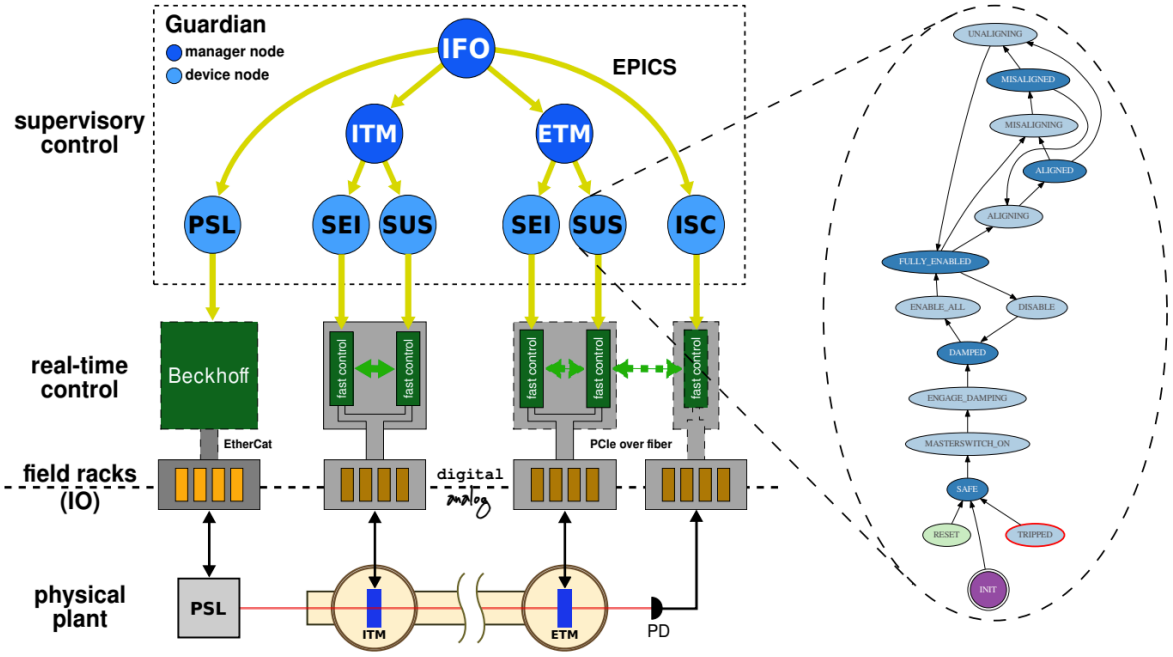


Figure 4: Digital control and supervision architecture: The LIGO interferometer is the physical plant that passes through an analogue-to-digital converter in the input/output layer and is then regulated in real-time by digital controls (gray boxes in the middle layer). Supervisory control is handled by Guardian, which is shown in the dashed box at the top. At the right is an example state graph for one of the nodes communicating between the front-end systems and Guardian [5].

performs regression analysis that does not capture non-linear noise effects well. As illustrated in figure 5, the advanced LIGO (aLIGO) design exhibits a curbed performance in lower frequency ranges, specifically sub-60 Hz, due to many noise sources. The subtractable noise is primarily introduced by interacting noisy signals from auxiliary control loops. This paper targets the bilinear angle-to-length coupling.

2.3 Signal Coupling

Subtractable noises coupled into signals sometimes interact in ways that can mimic real physical phenomena. Illustrated in figure 6, the angular rotation of the mirror can couple with the beam spot motion to create a cavity length signal that mimics a gravitational wave. For example, if the mirror is rotated by $\Delta\theta$ and the beam spot is displaced by Δy from the rotational pivot, the fluctuation in length Δl induced is [9]:

$$\Delta l(t) = \Delta y(t) \times \Delta\theta(t) \tag{1}$$

This bilinear coupling means that for a beam that is no longer centered in the mirror's rotational pivot, mirror rotation $\theta(t)$ and beam spot offset $y(t)$ affect the optical path $l(t)$

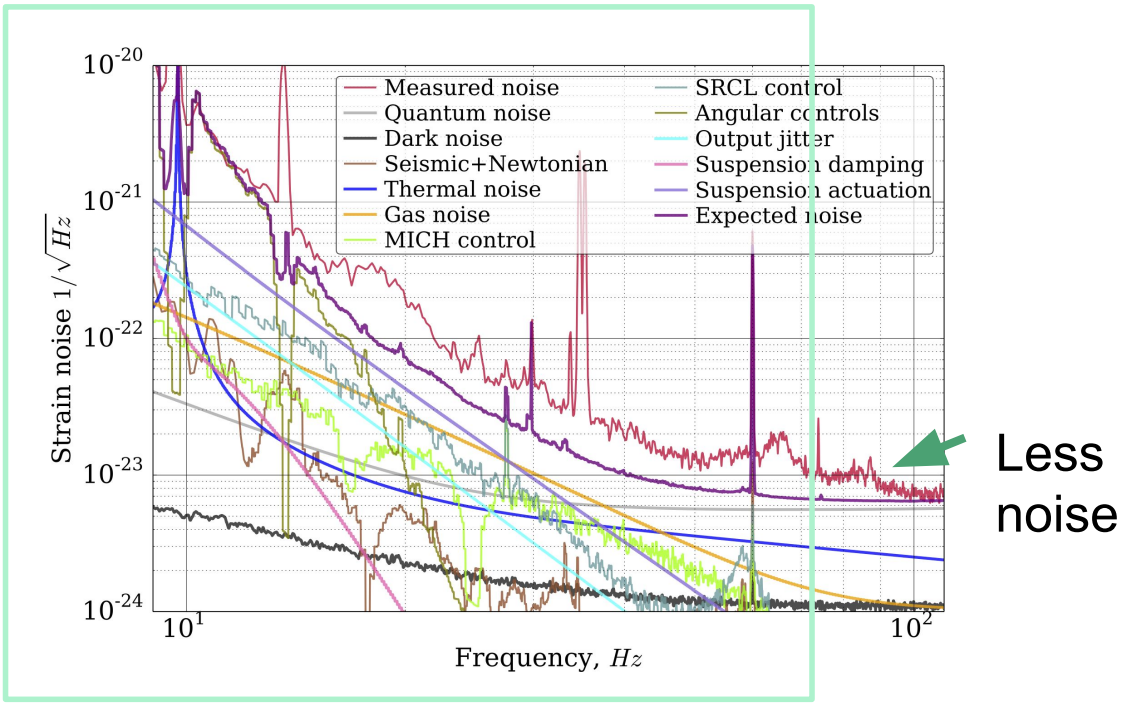


Figure 5: Limiting sensitivity noise budget: Livingston detector. The green rectangle encompasses the sub-60 Hz region, which is polluted by various noises shown in the legend. The angular controls are shown in olive green, and comprise the majority of the measured noises which are shown in red.

linearly given that the other is held constant. Converting to the frequency domain via a Fourier transform convolves the signals and approximates to the following equation:

$$\begin{aligned} \Delta \boldsymbol{L}(f) &= \Delta \boldsymbol{Y}(f) * \Delta \boldsymbol{\Theta}(f) \\ &\approx y^{\mathrm{RMS}} \times \Delta \boldsymbol{\Theta}(f) + \Delta \boldsymbol{Y}(f) \times \theta^{\mathrm{RMS}} \end{aligned} \tag{2}$$

2.4 Beam spot motion

The beam spot motion $\boldsymbol{y}$ on each mirror can be geometrically inferred directly from the mirror's angular motion. This is the theoretical expectation of a noiseless ADS channel measurement (refer to section 3.1). Starting from the residual pitch motions of the four test masses ETMX (end test mass, arm x), ITMX (input test mass, arm x), ETMY (end test mass, arm y), and ITMY (input test mass, arm y) respectively: $\boldsymbol{\theta} = [\theta_{\mathrm{ex}}, \theta_{\mathrm{ix}}, \theta_{\mathrm{ey}}, \theta_{\mathrm{iy}}]$, the

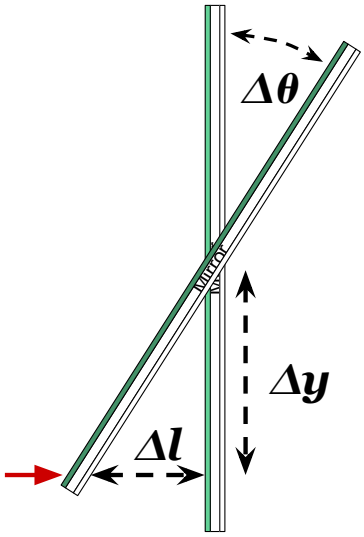


Figure 6: Beam displacement and mirror rotation create a length signal. To illustrate, this means that if holding one variable constant, the other linearly affects the length signal recorded. The paper focuses on the effect that beam displacement has on the induced length signal.

analytical spot position is:

$$\begin{aligned} \boldsymbol{y} = \begin{bmatrix} y_{\mathrm{ex}} \\ y_{\mathrm{ix}} \\ y_{\mathrm{ey}} \\ y_{\mathrm{iy}} \end{bmatrix} &= \frac{d\boldsymbol{Y}}{d\boldsymbol{\Theta}}\boldsymbol{\theta} \\ &= \frac{L}{1-g_{\mathrm{e}}g_{\mathrm{i}}} \begin{bmatrix} -g_{\mathrm{i}} & 1 & 0 & 0 \\ 1 & -g_{\mathrm{e}} & 0 & 0 \\ 0 & 0 & -g_{\mathrm{i}} & 1 \\ 0 & 0 & 1 & -g_{\mathrm{e}} \end{bmatrix} \begin{bmatrix} \theta_{\mathrm{ex}} \\ \theta_{\mathrm{ix}} \\ \theta_{\mathrm{ey}} \\ \theta_{\mathrm{iy}}. \end{bmatrix} \end{aligned} \tag{3}$$

where $L = 3995\,\mathrm{m}$ is the arm length, and $g_{\mathrm{e(i)}} \equiv 1 - L/\mathrm{RoC}_{\mathrm{e(i)}}$ captures the cavity geometry. Substituting the radius of curvature (RoC) of a mirror for aLIGO's input and end test masses gives the following constants: $g_{\mathrm{e}} = -0.78$ and $g_{\mathrm{i}} = -1.06$.

Since these relations are purely geometric, the coupling should be instantaneous. Thus, we can expect that the spot motion can be modeled only by data witnessed at the current time step. In other words, a purely dense network should be able to capture the geomtery equations.

2.5 Radiation pressure and Sidles-Sigg effect

Radiation pressure carried by the laser beam leads to optical torques on the test masses. This tilts and shifts the cavity axis (orange) with respect to the input axis (red) at the cavity's waist (see figure 7). Reducing such effects allows the axis to be fully overlapped, which also achieves the maximum power buildup inside the Fabry-Perot cavities. Depending

on these two degrees of freedom (DOFs), the combined motion of the ITM and the ETM generate a hard mode for a tilt (on top) and a soft mode for a shift (on bottom).

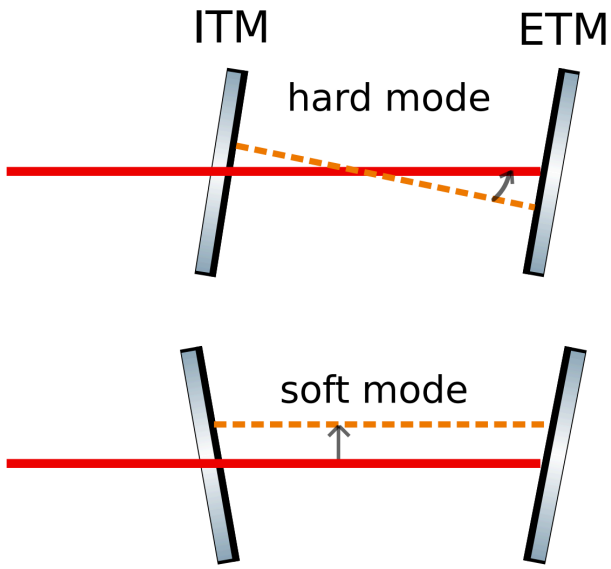


Figure 7: The hard mode "hardens" the optical spring by exerting a force that restores the mirror alignment. In contrast, a soft mode misalignment creates an optical torque that pushes the mirrors to be even more misaligned, and consequently "softens" the optical spring [9].

When the mirrors rotate about their rotational pivot due to radiation pressure, the coordinates of their centers of curvature change. The angular motion in turn changes the spot position on the test masses, and the spot motion couples with the circulating power in the arm cavity to create an optical torque that further perturbs the mirrors in angle creating a feedback loop. This effectively couples two test masses to form an "optical spring".

This effect is known as the Sidles-Sigg effect [10] and is crucial for aLIGO as it has to be properly controlled. This is accomplished by decomposing the individual mirror's motion into the Sidles-Sigg basis that diagonalizes the radiation-torque feedback. Such a basis corresponds to that formed by the 4 eigenvectors of the $d\boldsymbol{Y}/d\boldsymbol{\Theta}$ matrix. Changing basis from the hard/soft-mode angular motion to the common/differential arms numerically reads:

$$\begin{bmatrix} \theta_{\rm ch} \\ \theta_{\rm dh} \\ \theta_{\rm cs} \\ \theta_{\rm ds} \end{bmatrix} = \begin{bmatrix} 1 & 0.87 & 1 & 0.87 \\ 1 & 0.87 & -1 & -0.87 \\ -0.87 & 1 & -0.87 & 1 \\ -0.87 & 1 & 0.87 & -1 \end{bmatrix} \begin{bmatrix} \theta_{\rm ex} \\ \theta_{\rm ix} \\ \theta_{\rm ey} \\ \theta_{\rm iy}. \end{bmatrix} \tag{4}$$

A hard mode corresponds to an eigenvector associated with a positive eigenvalue of the $d\boldsymbol{Y}/d\boldsymbol{\Theta}$ matrix. This further means a positive spring constant that tends to restore the spring (i.e., the optical cavity) to its original location. On the other hand, a soft mode corresponds to an eigenvector associated with a negative eigenvalue and hence amplifies the perturbation.

For aLIGO, the eigenvalues are $d\boldsymbol{y}/d\boldsymbol{\theta}|_{\rm h} = +4.5\times10^4$ m/rad and $d\boldsymbol{y}/d\boldsymbol{\theta}|_{\rm s} = -2.1\times10^3$ m/rad, respectively. Thus, in principle no signal processing is required to simulate the spots on the mirrors from angular motions.

2.6 Machine learning network architecture

In our system, the input data undergoes pre-processing in order to improve numerical training. The input data is pre-processed through a standard scaling, as the numeric nature of the model means that normalized data will be easier to train computationally. This includes a scaling method that reduces the variance of the input dataset to unity. The mean is preserved, since it represents a DC offset, which is important for some of the witness channels, and is thus a valuable information to the network.

Machine learning (ML) neural networks (NN) perform a mapping of input to output data, without having to know much about the coupling in between, by employing an iterative regression model. The simplest network architecture, a dense model, takes input data and connects the values via layers with different weighs to produce a predicted output. The network aims to minimize the difference between its output and some target output. The target for this paper is to model the spot motion on the test masses. In our case, this implies that we can only combine data from the same time-step from different witnesses. This should be able to capture the physical equations theorized.

To ensure that the model focuses on capturing the desired frequency range, the model utilizes a custom bandwidth limited mean (BLM) loss function which computes a power spectral density (PSD) based loss within the 0.03 - 0.5 Hz frequency range.

Due to the purely geometrical conversions to obtain the spot motion from the witness channels, the subtraction can theoretically be performed using a dense network with linear activation (which can be thought of as representing geometric matrix operations). Introducing an alternative activation function can capture non-linear noise, yet introduces complexity which we want to minimize if possible. A convolutional layer can also be added, since it allows us to filter the inputs. This filtering can happen over time and thus prior inputs can help in the prediction of future signals. However, a network that requires a convolutional filter in order to reconstruct the spot motion implies that there are non-instantaneous couplings in the aLIGO control system that are currently not understood or accounted for.

2.7 MISO Linear Coherency Algorithm

To be able to analyze whether the data can model purely linear transfer functions, we can consider a coherence analysis with the witnesses inputted into the network as a multiple-input-single-output (MISO) formula. A key assumption is that we can extract all information possible from the multiple input channels and leverage it to reconstruct a desired target.

The linear coherency analysis estimates the magnitude-squared coherence function using Welch's overlapped averaged periodogram method. In other words, the function takes on values between 0 and 1 that indicate how well a combination of x inputs corresponds to a target output at each frequency. The magnitude-squared coherence is a function of the

power spectral densities, $P_{xx}(f)$ and $P_{yy}(f)$, and the cross power spectral density, $P_{xy}(f)$, of x and y:

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}P_{yy}} \tag{5}$$

3 Methodology

3.1 Witness channels

Though there are many sensors in aLIGO's digital control and supervision architecture (amounting to thousands of witness channels), we can expect the spot motion to be explained by only a few of those. From the physical properties of the system, there are a few witness channel groups (ADS, TR, C/DCHARD, OPLEV, ISI) shown in figure 8, which directly correlate to the spot motion. A complete list of the witness channels used is also shown in Appendix A.

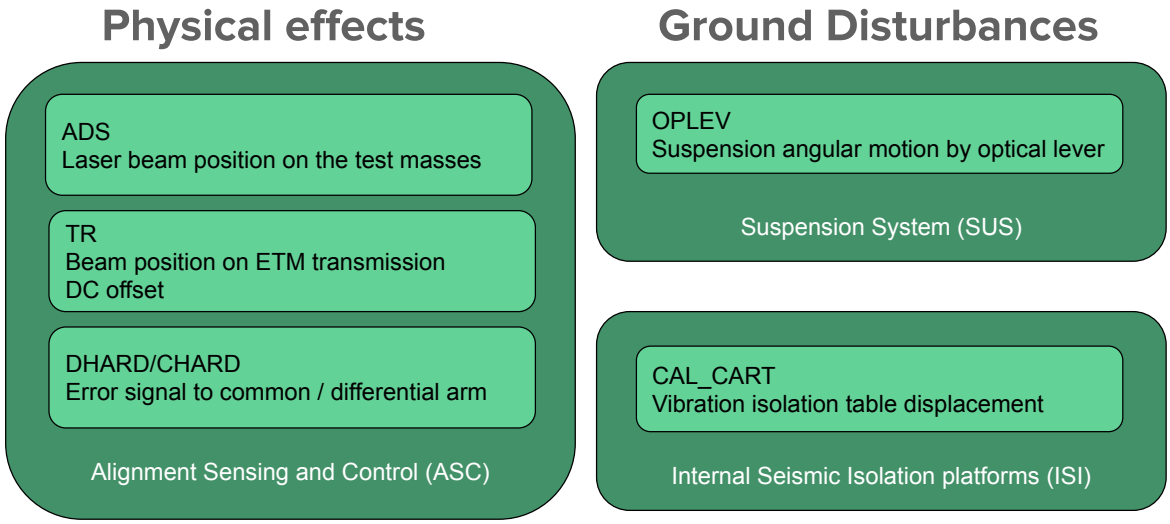


Figure 8: Witness groups correlated to spot motion: ASC, ISI and SUS. The ACS contains three relevant channel groups - ADS measures the laser beam position on the test masses, TRs measure the beam position on the ETM transmission with a DC offset, and CHARD/DHARD propagate an error signal to the common and differential arms. The OPLEVs measure suspension angular motion as a part of SUS, and the CAL CART measures displacement of the vibration isolation table as a part of ISI.

The first three sets of channels are a part of the alignment sensing and control system (ASC). This is important in order to account for physical effects such as radiation pressure. Note that all channels have different DC offsets because they have different reference points.

The first group in the ASC system are the ADS channels which measure directly the laser

beam position on the test masses relative to the suspension pivot with very low signal-to-noise ratio (SNR). This channel is currently used as a direct measurement of the spot motion that is coupled into the DARM, and thus is the target output of our network. The spot on the end test masses is also measured by more sensitive TR channels, the second group explored from the ASC system. However, the DC positions typically have offsets due to the telescope delivering the beam from the test mass transmission to the sensor. The motion of the sensor's mounting table also introduces AC noise. The third group of channels from ASC, the C/DHARD channels, feeds the common and differential arm the error signals of the angular motion relative to the input laser beam in the hard mode (high frequency).

Lastly, the beam motion is affected by ground disturbances, controlled by a suspension system (SUS) and internal seismic isolation system (ISI). The OPLEV channels are a part of the suspension system (SUS). They measure the angular motion of each individual mirror relative to the local chamber hosting the optics using an optical lever.

To capture more data, the ISI platforms measure the longitudinal motion at the top of the suspension chain. While this is the main source of mirror motion in the microseismic frequencies, the length-to-pitch transfer function rises as f^2 (two zeros at 0) at the range of interest, yet the phase advance is not exactly 180° because the resonances (poles) are not too far away from the micro-seismic band [8]. In other words, the ISI channel group has phase delay from the suspension transfer function that cannot be modeled via a dense model (because the coupling is not instantaneous). Thus, this channel group is not used in the model training illustrated in this paper. However, the information that the channel carries could prove useful in a convolutional regression model, if such is established in future efforts.

Before creating a model, it is helpful to understand what these channels actually look like in the time and in the frequency domain. The timeseries of the channels are not informative since the system convolves signals in the frequency domain. That is why the channels are mostly analyzed from spectral density vs frequency graphs. The input data consists of multiple witnesses and one target channel, each of duration of 2048 seconds and a sample rate of 16Hz.

3.2 Reconstructed Spot Motion

A technique called dither locking is used in aLIGO in order to use phase modulation in a control loop to reduce motion from residual degrees of freedom of the output mode cleaner (OMC). The phase modulation allows to measure the error in the two degrees of freedom caused by radiation pressure effects. This system couples the spot motion from the ADS into the DARM. To expose the signal, the feedback loop only cosiders the frequencies under 0.03 Hz because the dithering loops have low SNR but need long integration time (thus low bandwidth in frequency). This reduces injecting any sensing noise back into the system.

However, there is a nonlinearity involved in the problem because the dithering lines are very close to each other. For the Livingston detector those are only 0.3 Hz apart (at 7.4 Hz and 7.7 Hz), which means that the spot motion of ETMY at 0.1 Hz will be converted to 0.3 +- 0.1 Hz in the constructed ETMX's spot channel based on dithering-demodulation.

To reconstruct the data in a pseudo-ADS channel, some signal processing is required to

revert the signal coupling. The main GW readout $h(t)$, which theoretically observes the angular noise in the bilinear way $h(t) = spot(t) \times angle(t)$ is dithered to demodulate the signal and create an angular motion $angle(t) = cos(\omega_d t)$. Then this would up-convert the spot motion at ω_s to $(\omega_d \pm \omega_s)$ in the $h(t)$. The dithering frequency is known and can be used to multiply h(t) with cos($\omega_d t$). This would then down-convert the signal of $h(t)$ at $(\omega_d + -\omega_s)$ down to ω_s and $2\omega_d + -\omega_s$; to eliminate the latter, data is low-pass through a filter after demodulation. Demodulation can be performed with a cosine or a sine, resulting in a signal that would vary as cos($\omega_s t$) or sin($\omega_s t$). This requires the utilization of I and Q phases. When you demodulate a cos with a cos, as in I phase, you get a constant at DC. Whereas in Q phase, the demodulation of a cos with a sin results with 0 offset at DC. Conventionally, the demodulation phase is set so that the DC signal is maximized in I and zeroed in Q.

3.3 Using machine learning approach

The question in focus is whether a purely dense network can capture the effects as theorized, or will the noise coupling propagate through the signal-processing pipeline in unforeseen ways.

4 Data Analysis

4.1 Mock Data

To be able to reliably gauge whether a network is capable of capturing the complexity of a model, we perform the analysis on simulated data prior to real data. If the transfer function can be simulated on the mock data, then the real data will likely also be modeled by a similar network architecture. While the true coupling mechanisms remain unknown, the mock data should be similar enough to real data to capture its complexity.

Using this approach, the witness channels and coupling can be simulated through the physical and geometrical properties of aLIGO. This is illustrated in figure 9. The angular motion of test masses is simulated by filtering a randomized signal to mimic the test mass pitch motion. The signal is passed through a bandpass filter (microseismic region) and a lowpass filter, followed by emphasizing the micro-seismic peak and the main length-to-pitch suspension peak via introducing additional poles. The angular motion of the test masses $\boldsymbol{\theta}_{true}$ generated above is combined with the cavity geometry to give the true spot motion, as shown in equation 3:

$$\boldsymbol{y}_{\text{true spot}} = \mathbf{M}_{\text{a2s}} \boldsymbol{\theta}_{\text{true}} \tag{6}$$

This is the target channel for our network since it represents the theoretical spot location. Using the true spot as a target gauges whether a network infrastructure is well designed for the problem at hand. However, to truly represent real data, we need to introduce sensing noise which creates a noisy spot motion. In order to model the noise in the detector, the noise injected into the spot motion should match that of an ADS channel (which is currently

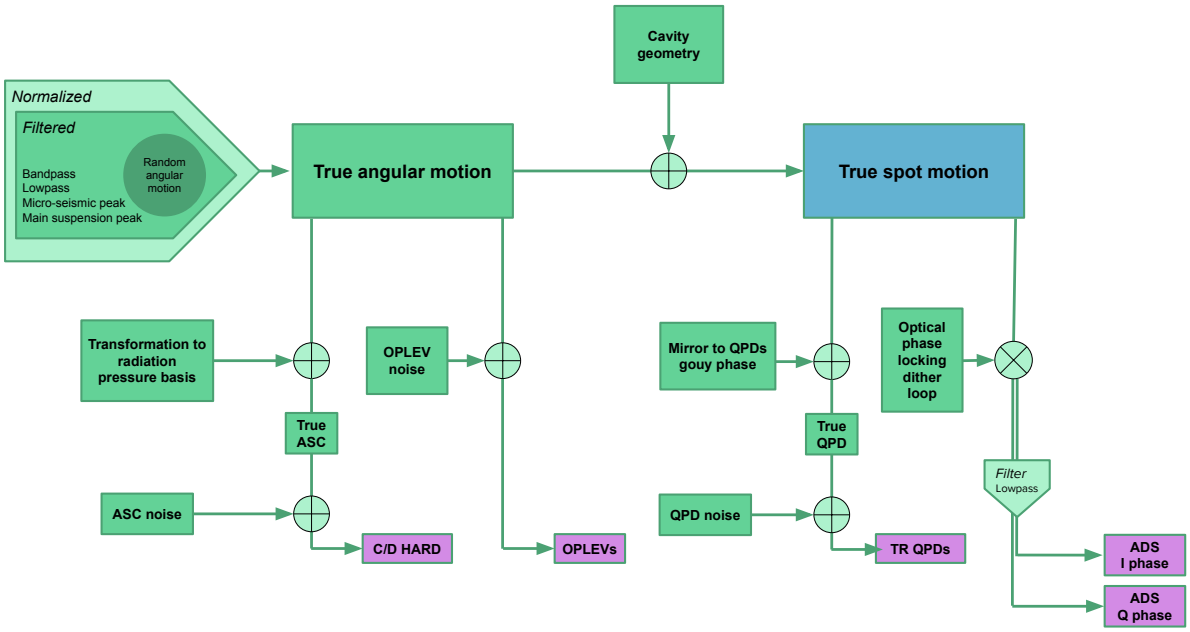


Figure 9: The physical geometry of the aLIGO system is used to generate mock data. Starting from random motion, we apply filters as shown on the top left. This results in a true angular motion which can be coupled with geometric relations to give two sets of witness channels (C/D HARD, OPLEVs), shown in purple. The target true spot motion shown in blue can also be derived directly from the angular motion. The rest of the purple witness channel groups (TR QPDs, ADS) can be generated by mimicking signal processing done on the true spot motion.

used as a spot motion witness fro the real detector, where we cannot measure the true spot motion directly). To achieve this, we can plot the spectra of an ADS channel and roughly match the SNR to that of a noisy true spot. This is done by adding sensing noise until the spectral density matches the 0.03 Hz level of the ADS channel, since the spot motion signal is at 0.12 Hz and at 0.03 Hz the ADS channel is limited by noise. The SNR value used is 1.

The witness channels are simulated in similar ways, starting from the generated true angular motion and introducing noise $\boldsymbol{N}_i$. The noise is normalized to have a standard deviation and RMS of unity for each channel set, then calibrated for the corresponding units and known SNR ratio (determined by the sensitivity of the sensors).

$$\mathbf{OPLEV} = \boldsymbol{\theta}_{\text{true}} + \mathbf{N}_{\text{oplev}} \tag{7}$$

$$\mathbf{C/DHARD} = \mathbf{ASC}_{\text{true}} + \mathbf{N}_{\text{ASC}} \tag{8}$$

where $\mathbf{ASC}_{\text{true}} = \mathbf{M}_{\text{m2r}}\boldsymbol{\theta}_{\text{true}}$, which is the angular motion in the radiation-pressure basis shown in equation 4

$$\mathbf{TR} = \mathbf{QPD}_{\text{true}} + \mathbf{N}_{\text{QPD}} \tag{9}$$

where $\mathbf{QPD}_{\text{true}}$ is the 0 DC offset pitch and yaw calibrated for the QPDs gouy phase

$$\mathbf{DARM} = \text{dither}(\boldsymbol{y}_{\text{true spot}}) \tag{10}$$

where the dithering is described in further detail in section 3.2.

Reconstructing the expected ADS channel allows us to compare theoretically whether the ADS channel introduces any additional noise due to the dithering non-linearity. The ADS channel can be reconstructed by demodulating the DARM, which is what happens in the real detector. We bandpass the DARM around 7.7 Hz using a 4th order butter filter with +- 0.03 Hz bandwidth (with a 2nd order Butterworth filter to get rid of phase delay), and demodulate the results at 7.7 Hz, assuming optimal demodulation phase.

In order to better understand the signals carried by each witness channels, we can also visualize them in the time domain. Figure 10 shows a snippet of the time-series of the mock witness channels used.

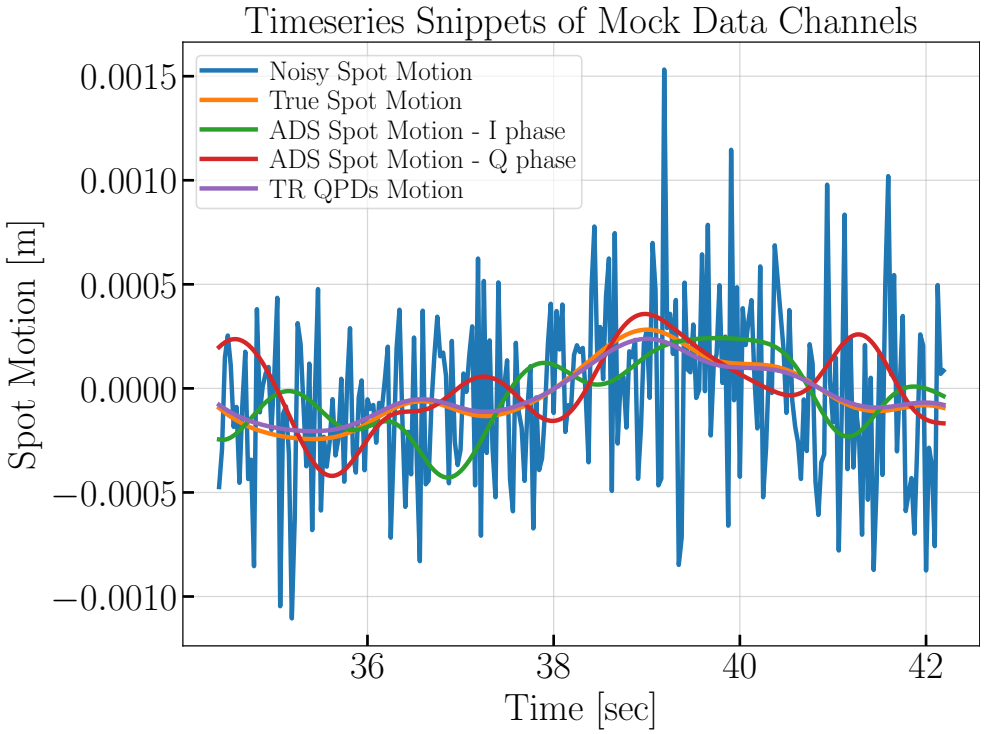


Figure 10: Snippets of mock data channel signals in the time domain: timeseries of channels tracking displacement [(noisy) spot motion, ADS, TR QPD]

In addition, the spectra of the mock data is shown in figure 11, where one channel from each group is plotted to compare them.

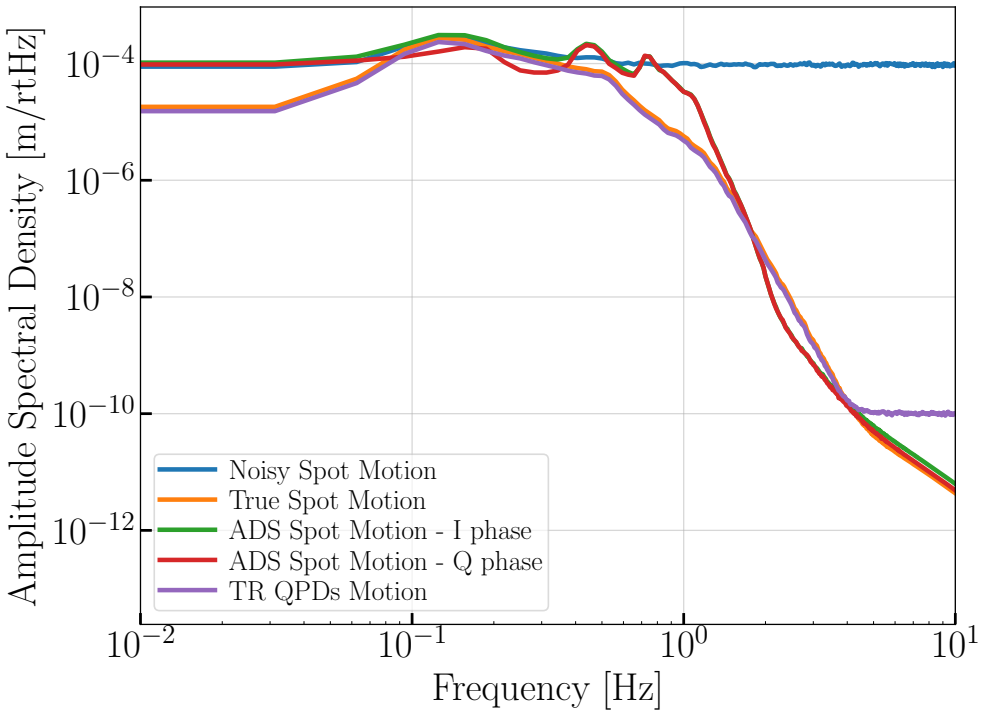


Figure 11: Mock data channel signals in the frequency domain: spectra of channels tracking displacement [(noisy) spot motion, ADS, TR QPD]

4.2 Real Data

Figure 12 shows a witness channel from each groups mentioned above as it varies its output over time. Compared to the spot motion, or target data, the witness channels fluctuate a lot over time. The spot motion is directly measured via the ADS channel, and then the signal is coupled into the DARM main readout. We can also reconstruct the ADS channel from the main readout by inverting the signal processing.

Naturally, the analysis continues by plotting the power spectra of the data, shown for each channel in figure 13. The measured spot motion varies its period on a time-scale of 5-20 seconds. This means that the frequency we will be interested in modelling will be in the 0.05-0.2 Hz band. Thus the peak around 0.12 Hz represents the motion over time-periods similar to that of the beam spot motion.

5 Mock Data Results

5.1 Linear coherency

The linear coherency between the true/noisy spot and the witness channels used (TR, OPLEV, C/D HARD) is shown in figure 14. From the figure, we can see that although the noiseless spot has a coherence of 1 with the witnesses in the low frequency range, which means that the true spot can almost entirely be reconstructed, the noise introduced highly obscures these relations.

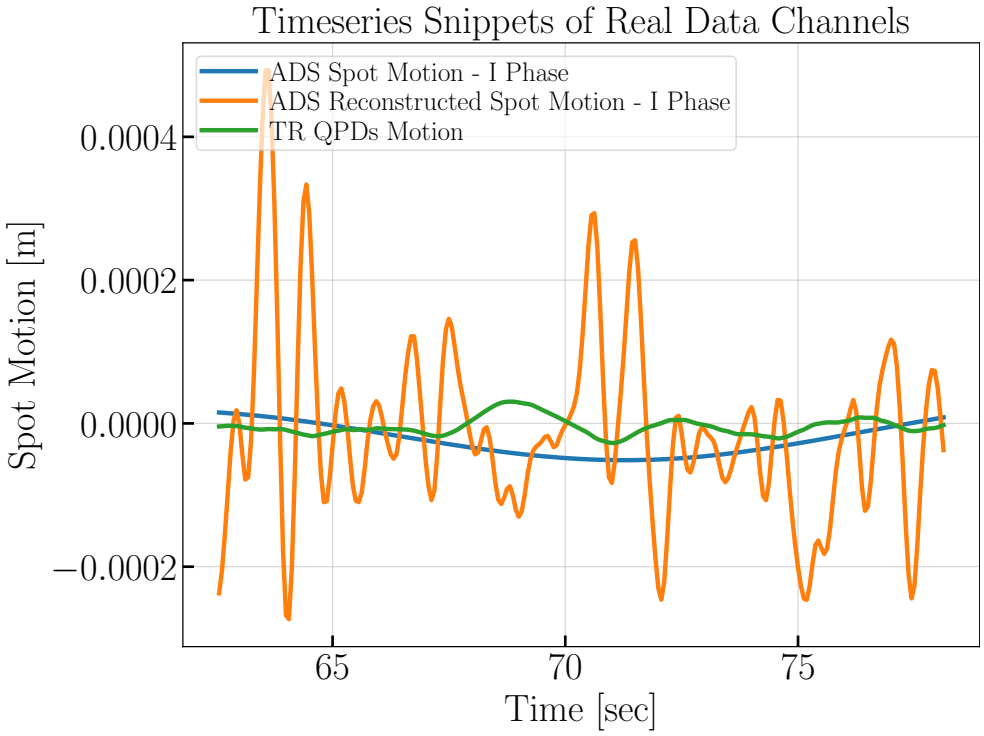


Figure 12: Snippets of real data channel signals in the time domain: timeseries of channels that track the displacement of the spot motion [ADS, Reconstructed ADS, TR QPD]

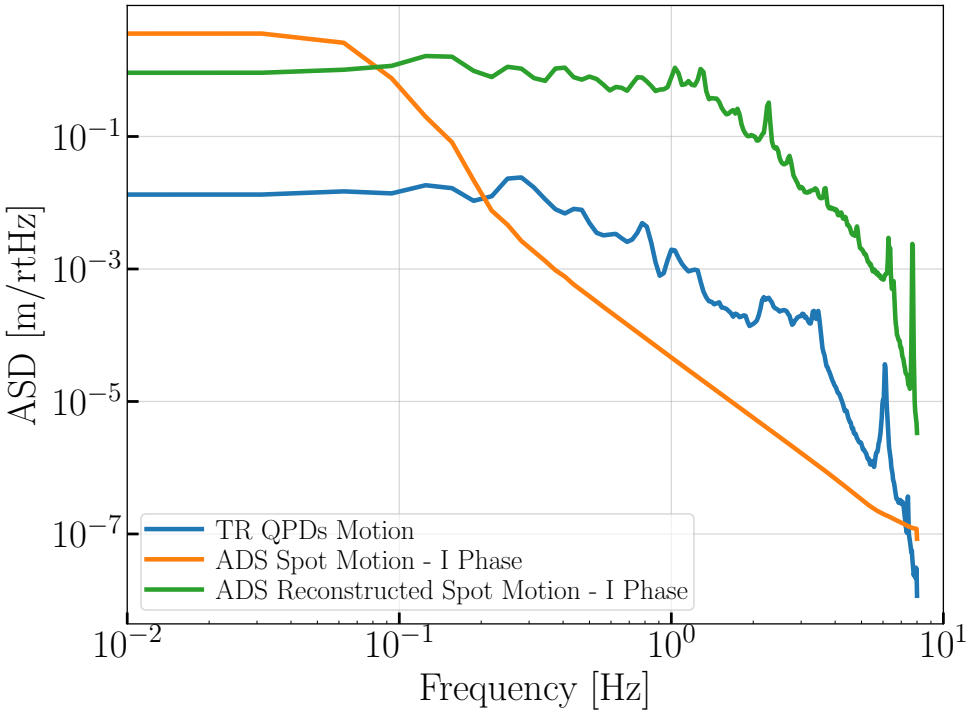


Figure 13: Real data channel signals in the frequency domain: spectra of channels that track the displacement of the spot motion [ADS, Reconstructed ADS, TR QPD]

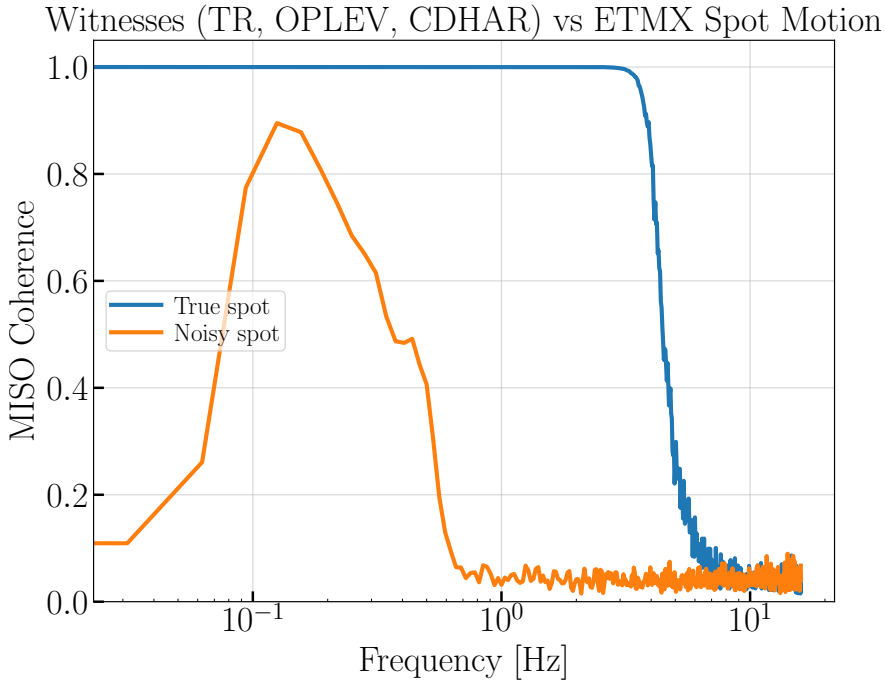


Figure 14: Coherency

5.2 Machine Learning Models

The target used in the neural network is the mock noisy spot motion that uses the true spot and white sensing noise model. The sensing noise is added so that the SNR is consistent with what an ADS channel would output, as shown in section 4.1. Results, including predicted timeseries and spectra, as well as loss during training and validation, from a fully dense network with linear activation functions are shown in appendix B. To better understand how the model performs, MISO linear coherence from spectra is also shown.

From this purely linear network, we can already see that even when using a noisy spot as a target, the network can leverage the dependencies to capture the true spot motion. It is important to note that the coherence between the network's prediction and the true spot motion is much higher than expected. We can deduce that the linear nature of the network puts emphasis on the linear geometric dependencies of the beam spot motion, despite high levels of random noise. This effect can be leveraged in the real aLIGO interferometer, since the true spot motion is not known, but rather inferred from noisy channels.

After proving that the spot can successfully be modeled through the witness channels with a purely dense network, the focus is to model the spot through the ADS channels. Although theoretically those should correspond to the noisy spot motion, the dithering lines cause the ADS channel to be polluted. The mock ADS data is generated in a way mimicking reality and contains all the contamination from adjacent dithering lines. The results are shown in appendix C, alongside the coherence between the prediction and the target true spot.

The results can be interpreted as implying that the dithering lines induce non-linearities in the ADS control loop. However, using a dense machine learning network focuses on the linear components only, and reverses the non-linearities to capture a much higher coherence

with the true spot in the low frequency range.

6 Conclusion

In an attempt to increase sensitivity, and hence number of events detected, LIGO has been working toward effectively implementing ML techniques in noise reduction. To capture the complicated coupling mechanisms between various auxiliary channels, neural networks are much more appropriate than standard linear models. Networks developed so far demonstrate an ability to model bilinear coupling given perfect witnesses of beam spot motion and angular motion. This report focused on modeling the beam spot motion in order to help with the reduction of the bilinearly coupled noise it induces. The results are promising, since using the ADS channel as a target of a dense neural network seems to reverse any non-linearities introduced by dithering lines and successfully models the true spot motion, which we cannot measure in the real system. Minimizing the instrumental artifacts is crucial and paves the way toward a much deeper understanding of gravitational waves.

7 Future Development

In order to be able to train the machine learning network better, some pre-processing techniques can be leverages to emphasize the microseismic band in the model through whitening which captures the coherency observed. Pre-conditioning the target channel to be flat in frequency (whitened) allows the lower frequencies to dominate the output, while punishing the higher frequencies. However, this is also a non-linear filter that might complicate the expected output. Further reddening the input leads to an unnecessarily dense matrix that attempts to eliminate higher frequencies (instead of keeping them random).

The beam misalignment with respect to the mirror at DC means that the AC angular motion will further interfere with the DC cavity field and cause power fluctuations at AC leading to an angle-to-power coupling. However, due to the finite mass of the mirrors, the power fluctuation will also create a length signal via the radiation pressure force. This angle-to-power-to-length coupling may confuse the geometrical angle-to-length coupling due to spot mis-centering, and consequently contaminate the error signal used for sensing a pointing degree of freedom (DOF). This effect can be significant when aLIGO reaches 100W of input power, and would thus deserve more detailed study in the future [9].

A hyper-parameter optimization can be developed to best capture the desired dependencies and further reduce noise in the seismic band. Future work should explore the performance of network parameters when capturing the spot motion model. Then, the overfitted networks can be reduced by choosing the optimal hyper-parameters (e.g. learning rate, hidden layers, features in each layer) that govern the learning process via an empirical hyper-parameter tuning in Neptune's Optuna [11].

Perhaps most importantly, the next step of this analysis is incorporating the dense network that models the angle-to-length coupling with a convolutional network that eliminates bilinear coupling. By implementing the two systems in tandem in the aLIGO pipeline, we will be

able to further analyze any effects that are unaccounted for and gain superior understanding of the detector and its limitations

8 Acknowledgments

I would like to thank my mentor Hang Yu for the constant support and guidance this summer with my project, to Rana Adhikari for always ensuring that I understand the physical meaning and motivation behind my work, and to Zachary Mark for introducing me to Neptune and its amazing organization. I would also lke to dedicate a thank you to the Caltech SURF program, the LIGO collaboration, and the National Science Foundation for making this research possible, as well as the MIT UROP office for making this research experience a collaboration between the two institutions - MIT and CALTECH. Finally, I am grateful for my MIT supervisor, Erik Katsavounidis, who ensured that I am welcome in the MIT LIGO noise reduction group.

Appendices

A List of all channels used in the models for real data

Input:

'ASC-CHARD_P_IN1_DQ',
'ASC-CHARD_Y_IN1_DQ',
'ASC-DHARD_P_IN1_DQ',
'ASC-DHARD_Y_IN1_DQ',

'ASC-X_TR_A_PIT_OUT_DQ',
'ASC-X_TR_A_YAW_OUT_DQ',
'ASC-X_TR_B_PIT_OUT_DQ',
'ASC-X_TR_B_YAW_OUT_DQ',
'ASC-Y_TR_A_PIT_OUT_DQ',
'ASC-Y_TR_A_YAW_OUT_DQ',
'ASC-Y_TR_B_PIT_OUT_DQ',
'ASC-Y_TR_B_YAW_OUT_DQ',

'SUS-ITMX_L3_OPLEV_PIT_OUT_DQ',
'SUS-ITMX_L3_OPLEV_YAW_OUT_DQ',
'SUS-ETMX_L3_OPLEV_PIT_OUT_DQ',
'SUS-ETMX_L3_OPLEV_YAW_OUT_DQ',
'SUS-ITMY_L3_OPLEV_PIT_OUT_DQ',
'SUS-ITMY_L3_OPLEV_YAW_OUT_DQ',
'SUS-ETMY_L3_OPLEV_PIT_OUT_DQ',
'SUS-ETMY_L3_OPLEV_YAW_OUT_DQ',

'ISI-ITMX_CAL_CART_X_OUT_DQ',
'ISI-ETMX_CAL_CART_X_OUT_DQ',
'ISI-ITMY_CAL_CART_Y_OUT_DQ',
'ISI-ETMY_CAL_CART_Y_OUT_DQ'

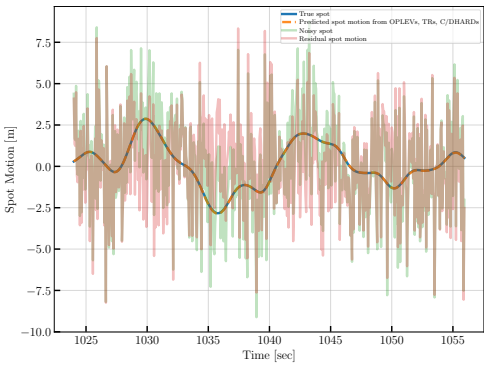
Target:

'ASC-ADS_PIT3_DEMOD_I_OUT_DQ',
'ASC-ADS_PIT4_DEMOD_I_OUT_DQ',
'ASC-ADS_YAW3_DEMOD_I_OUT_DQ',
'ASC-ADS_YAW4_DEMOD_I_OUT_DQ',

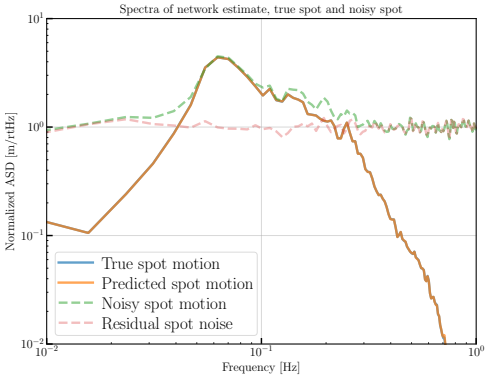
'ASC-ETMX_PIT_SPOT_I',
'ASC-ETMX_PIT_SPOT_Q',
'ASC-ETMY_PIT_SPOT_I',
'ASC-ETMY_PIT_SPOT_Q',
'ASC-ETMX_YAW_SPOT_I',
'ASC-ETMX_YAW_SPOT_Q',
'ASC-ETMY_YAW_SPOT_I',

B Mock data results - Noisy spot

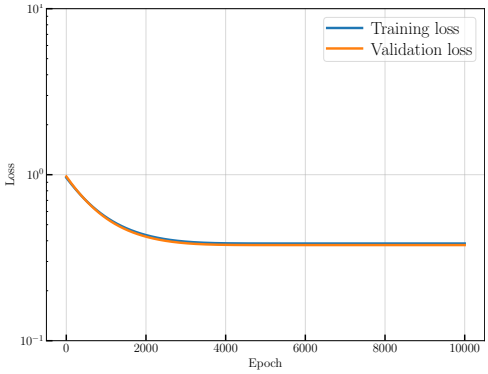
B.1 Network Results



(a) Timeseries

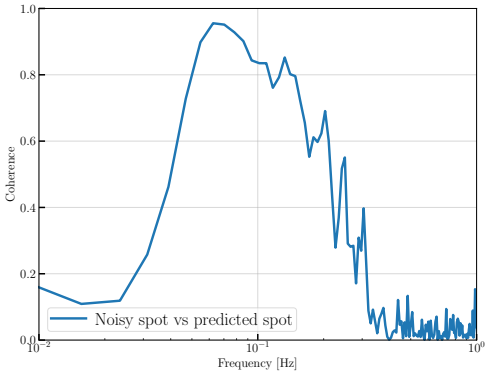


(b) Spectra

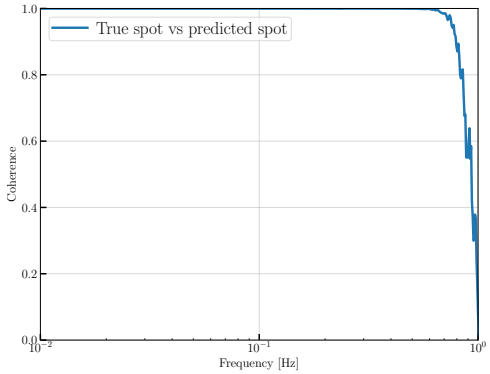


(c) Loss

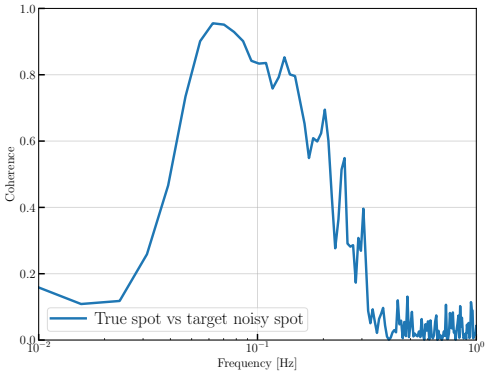
B.2 Predictions Accuracy



(a) Coherence between the target data (noisy spot) and the noisy spot model prediction



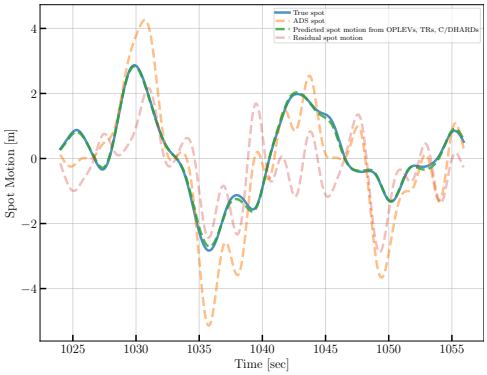
(b) Coherence between the true spot motion and the noisy spot model prediction



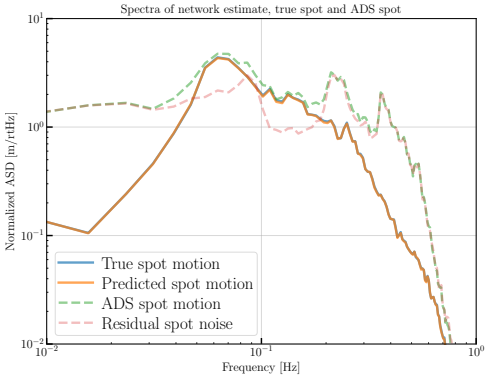
(c) Coherence between the true spot motion and the target data (noisy spot)

C Mock data results - ADS spot

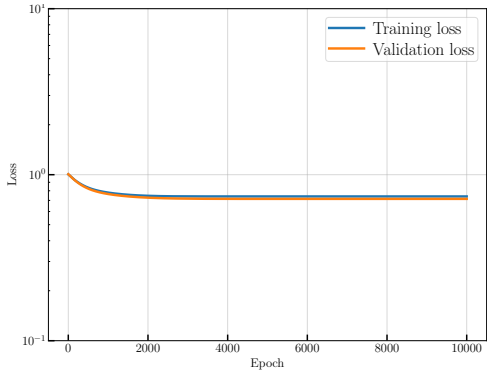
C.1 Network Results



(a) Timeseries

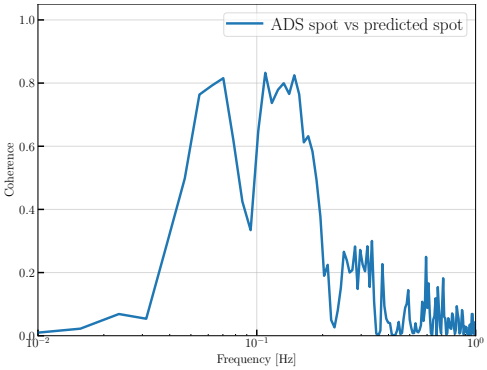


(b) Spectra

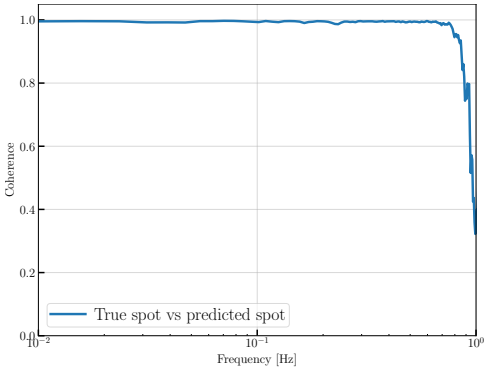


(c) Loss

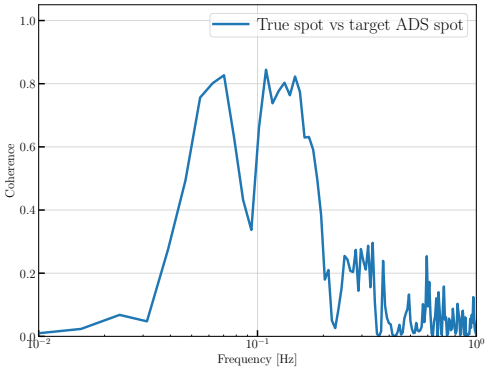
C.2 Predictions Accuracy



(a) Coherence between the target data (ADS spot) and the ADS spot model prediction



(b) Coherence between the true spot motion and the ADS spot model prediction



(c) Coherence between the true spot motion and the target data (ADS spot)

References

[1] Barish, Barry C. and Rainer Weiss, *LIGO and the Detection of Gravitational Waves.* Physics Today 52 (1999): 44-50 http://dx.doi.org/10.1063/1.882861

[2] Adhikari, Rana X *Gravitational Radiation Detection with Laser Interferometry* Reviews of Modern Physics 86.1 (2014): 121-151 https://arxiv.org/abs/1305.5188

[3] Abbott, B. P. et al *Observation of Gravitational Waves from a Binary Black Hole Merger* Phys. Rev. Lett. 116.061102 (2016) https://link.aps.org/doi/10.1103/PhysRevLett.116.061102

[4] Tiwari, Vaibhab et al. *Regression of Environmental Noise in LIGO Data* Class. Quant. Grav. 32 (2015) 165014

[5] Rollins, Jameson G. *Distributed State Machine Supervision for Long-baseline Gravitational-wave Detectors* arXiv 1604.01456 (2016)

[6] The LIGO Scientific Collaboration *Instrument paper on Advanced LIGO* Class. Quantum Grav. 32 (2015) 074001

[7] LIGO Laboratory *LIGO - A Gravitational Wave Interferometer* https://www.ligo.caltech.edu/page/ligos-ifo

[8] Massinger, Thomas James *Detector Characterization for Advanced LIGO* Dissertation: Syracuse University (2016) https://surface.syr.edu/cgi/viewcontent.cgi?article=1633&context=etd

[9] Yu, Hang *Astrophysical signatures of neutron stars in compact binaries and experimental improvements on gravitational-wave detectors* Thesis: Ph. D., Massachusetts Institute of Technology, Department of Physics (2019) https://dspace.mit.edu/bitstream/handle/1721.1/123343/1132722343-MIT.pdf?sequence=1&isAllowed=y

[10] Sidles, John and Daniel Sigg *Optical torques in suspended Fabry–Perot interferometers* Phys. Rev. Lett. 354.032905 (2006): 167-172

[11] Akiba, Takuya and Sano, Shotaro and Yanase, Toshihiko and Ohta, Takeru and Koyama, Masanoriv *Optuna: A Next-generation Hyperparameter Optimization Framework* Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019)

[12] Adhikari, Rana et al. *DC Readout Experiment at the Caltech 40m Prototype Interferometer* Class. Quant. Grav. 25 (2007) 114030