Using Gas Detection Values to Normalize Data

Marisol Guzman[1], David Schneider[2], Ryan Coffee[2]

[1]Intern

[2]Linac Coherent Light Source, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA.

[+]Contact: davidsch@slac.stanford.edu

Introduction

This project goal was to analyze data from Ryan Coffee's project which focused on being able to measure the reaction (lasing) of shooting two pulses at distinct energy values into a molecule. The measurements were predicted using a machine learning based model. The model predicts two things, first whether the beam produced by the accelerator had lasing in low energy light (E1) or high energy light (E2) or both (E3). Second, the actual energy values of the low or high lasing.

First we examined how the molecule reacted to different values for E1 and E2. Then, we looked for the E1 values produced the greatest reaction. Next, we looked at how the gas detection values were able to be used in order to normalize data. Gas Detection values measure gas disruptions that happen if there is lasing and electrons coming off of the molecule.

According to Ryan Coffee, we can normalize using gas detection values because the bigger the gas detection values, the more lasing there is. Thus if show that there was a linear relationship between the gas detection values and E1 values, then we could potentially show that normalizing by the gas detection values was possible.

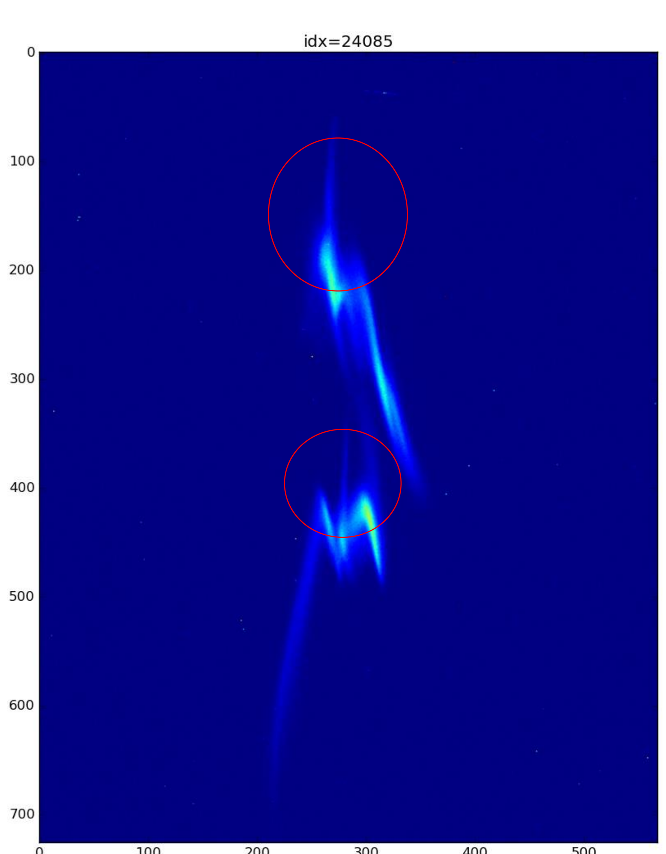

Fig 1. This is an example that shows both high energy and low energy lased (E3). Note that the image is flipped so the top finger (circled in red) is low energy E1 and the bottom finger is E2 (also circled in red).

Research

In order to find the largest amount of signal, we had to graph the mean energy of each E1 position in order to see which E1 position interval had the largest amount of mean energy. In the graph below, the interval with the positions between 4.5263 to 4.8783.

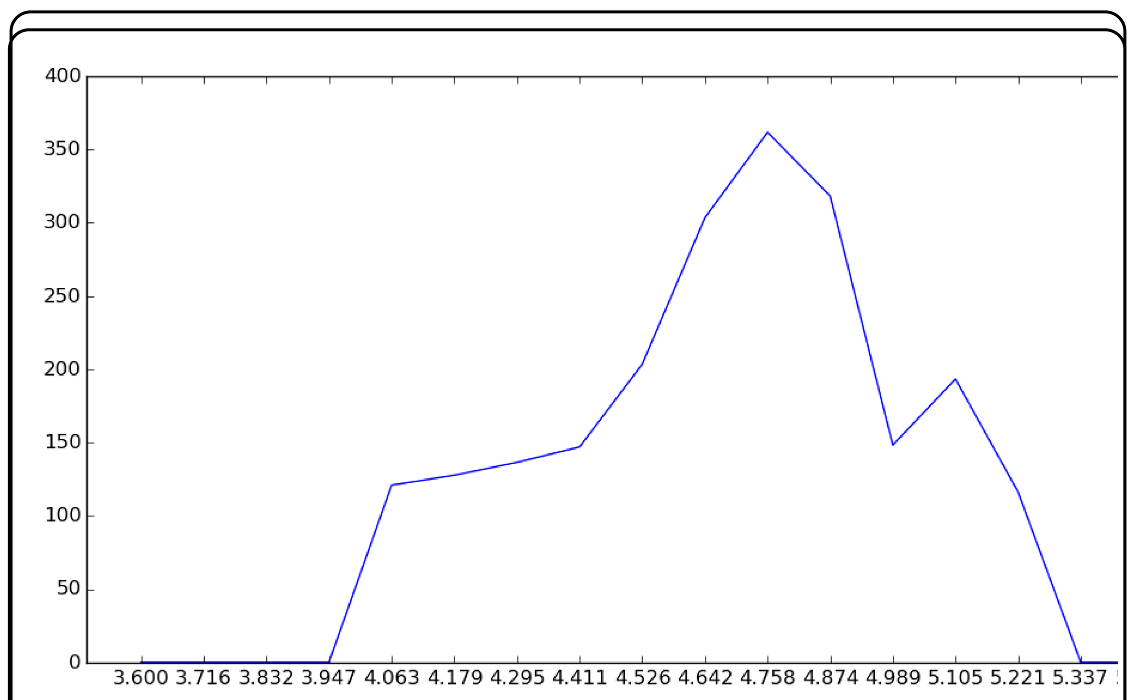

Fig 2. Finding the highest signal

Then, we graphed a histogram that contained all of the gas Detection values that were in were Class 1, where only the low energy lased, and that were part of run 72 (a training run). For the histogram we created 35 bins and selected the 3 bins that had the largest amount of items. In this case those gas detection values were 0.0921836 to 0.1106645.

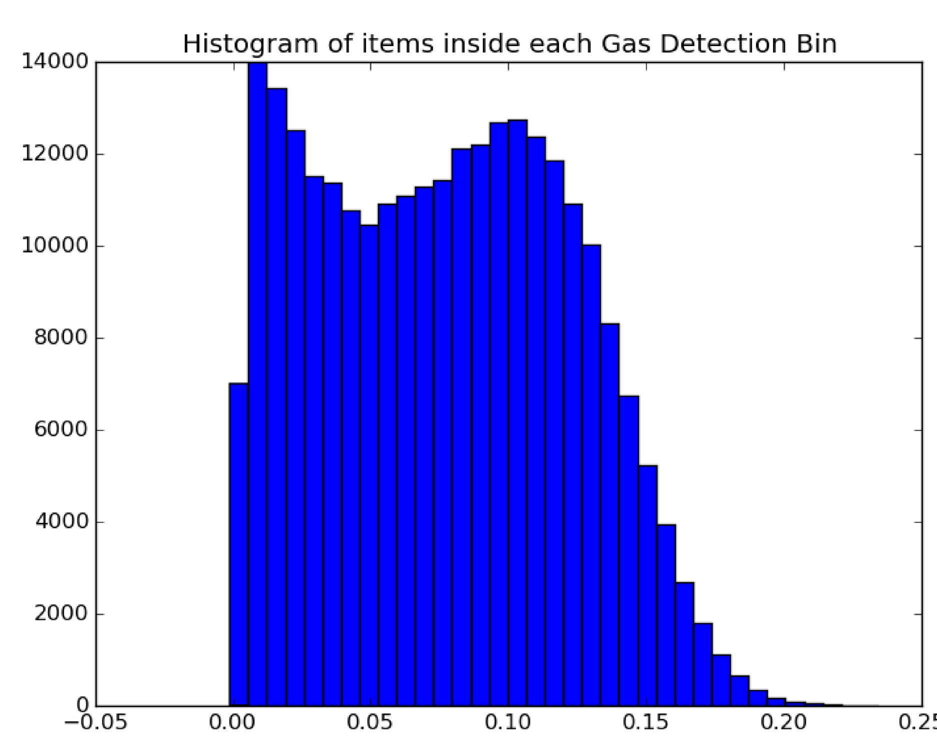

Fig 3. Histogram of Gas Detection Values. Note that the high bars toward the left were not considered because they had a gas detection value close to zero, which signals that there was no lasing.

After finding the boundaries of E1 positions for the highest signal and the largest amount of gas detection values, we used these numbers as constraints in order to build bins and scan over E1 values once again. In the end, we were able to build 11 bins and the bin with the highest mean energy was the bin was in between positions 4.5 and 4.6. With this information we were able to further to refine the constraint of E1 positions

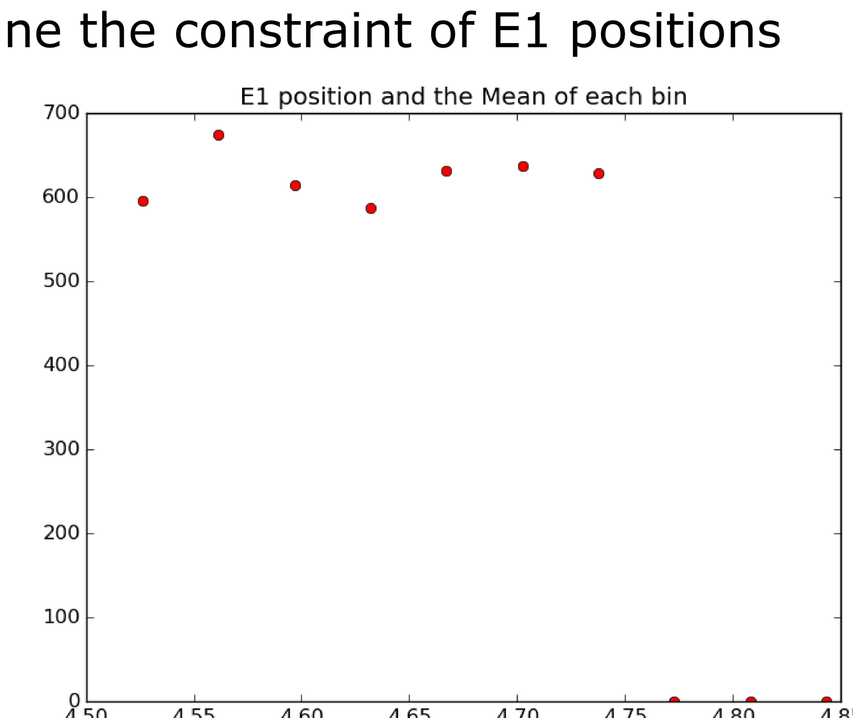

Fig 4. Fixing the gas detection values and scanning over the E1 values.

Finally, now that we had the exact constraints of the E1 positions we could scan over the gas detection values. By scanning the gas detection values, we could analyze whether there was some sort of linear relationship between E1 energy and gas detection values. Ryan Coffee's theory was that if these gas detection values created a line, then it could verify that the gas detection values could be used in order to normalize future graphs.

In the end, the linear regression revealed that:

- r value = 0.993285026751
- p value = 8.82371270884e-09
- standard deviation = 9.3758571527
- slope = 227.679022705
- intercept 132.830143591

The high r value reveals that there is a linear correlation that exists between the bins of gas detection values and the means of the E1 values.

Also, this additional information reveals that the relationship between gas detection values and means of E1 are in the form of y~ 228x + 133.

The information above reinforces the idea that gas detection values can be used as a better way to normalize E1 graphs.

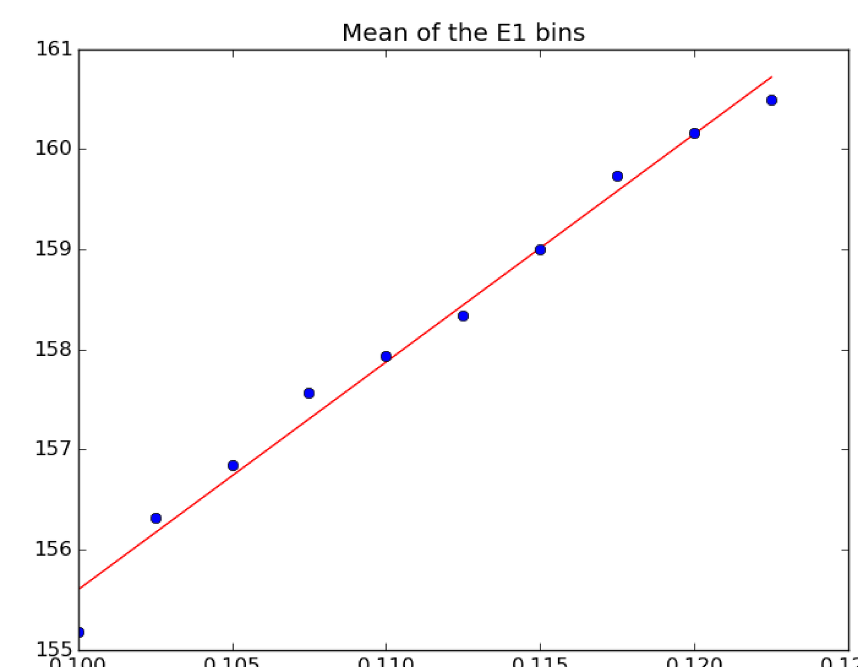

Fig 5. Linear Regression of E1 bins and gas detection values

Conclusions

This linear regression verifies that one can expect that when lasing increases, then so do the gas detection values. Also, normalizing by the gas detection values allows for scientists and the data analysis team to be able to accurately combine and analyze values.

Acknowledgments

Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515.

Date: 08/12/2016