

# A Compilation of Internship Reports

2021

Prepared for  
The Office of Educational Programs  
Brookhaven National Laboratory  
Office of Educational Programs

### **DISCLAIMER**

This work was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.



# Table of Contents

Numerical study and computational simulations of laser-plasma interactions.....	1
<i>James Adams and David Wan</i>	
Anomaly detection algorithm research and testing with benchmark datasets for the performance analysis of high-performance computing applications running at exascale .....	15
<i>Arnav Agrawal</i>	
Performance analysis of tomography algorithms.....	26
<i>Jabel Alcantara</i>	
Visualization of structured threat information expression for malware analysis .....	36
<i>Jeremy Anderson</i>	
In situ spectroscopy of metal oxide catalysis.....	49
<i>Lukas Baker</i>	
Comparison of scattering and extinction spectra by analytical and Finite Element Analysis methods .....	64
<i>Kevin Beckford</i>	
Inorganic chemistry of battery electrodes.....	69
<i>Melanie Brito</i>	
A feedback-controlled loop to regulate a power supply .....	88
<i>Isaiah Brown-Rodriguez</i>	
Energy extraction in a superconducting magnet, power supply and LP filter design.....	100
<i>James Bush</i>	
Design analysis and engineering improvements for a next-generation x-ray powder diffraction sample furnace.....	109
<i>Edwin Cardenas and Amanda Sirna</i>	
Super Pioneering High Energy Nuclear Interaction eXperiment infrastructure interlocks and monitoring systems.....	124
<i>Michael Charumaneeroj</i>	
Two-photon interferometry simulations .....	135
<i>Zhi Chen</i>	
Design NSLS-II PPS test fixtures .....	153
<i>Michael Chin</i>	
The redevelopment of the explosives safety program at Brookhaven National Laboratory.....	157
<i>Colin Clark</i>	
Studies of the mechanism of CO oxidation using transient experiments.....	168
<i>Gerardo Claudio-Serrano and Lisa Feder</i>	
Optimizing automation of neutron resonance parameterization.....	183
<i>Rose-Marie Crawford</i>	

Where's Waldo on the beam line: Using unsupervised learning to automate sample detection .....	194
<i>Kaleigh Curtis</i>	
High-level synthesis tools and methodologies for embedded AI application-specific integrated circuits .....	200
<i>Bshara Dababneh</i>	
EIC SRF cavity radiation potential and coupled-cell energy balance analyses .....	211
<i>William Daniels</i>	
Numerical simulation of impact ionization in low-gain-avalanche-diodes .....	220
<i>Nimal de Lanerolle</i>	
Electron-Ion Collider infrastructure cooling through thermal hydraulic modeling and analysis .....	228
<i>Joseph DeRienzi and Thea Vijaya Kumar</i>	
Optimizing neural networks for next-gen artificial intelligence .....	251
<i>Andrew Deutsch</i>	
Analysis of the sPHENIX cooling water piping system .....	258
<i>David Drysdale</i>	
Numerical simulations of small-scale junction field effect transistors .....	268
<i>Dominique Duliepre</i>	
Machine learning for classifying neutron resonances .....	287
<i>Mary Fucci</i>	
Automation and reporting of the deployment of multiple packages .....	
<i>John Gedell</i>	
Quantum simulation of spin entanglement in quantum chromodynamics strings using a Kondo-type spin model .....	293
<i>Wenjie Gong</i>	
Optical model for a vacuum transport upgrade .....	324
<i>Benjamin Herr</i>	
Experimental design for the measurement and analysis of vibration-rotation spectra of molecules relevant to extra-terrestrial bodies .....	333
<i>Joshua Heuvel-Horwitz</i>	
Variability in precipitable water vapor from the HoustonNet GPS Network .....	349
<i>Marcie Hogan</i>	
Charge sharing in pixelated sensors .....	361
<i>Andrea Houck</i>	
Text mining from the scientific literature with natural language processing methods .....	388
<i>Darrien Hunt</i>	
Neutronics modeling for intentional forensics .....	399
<i>Shia Jones, Eric Schess, and Caroline Sears</i>	
Simulation and analysis of neutrino fluxes from the Large Hadron Collider .....	413
<i>Fnu Karan Kumar</i>	



Converting National Synchrotron Light Source II accelerator status pages from HTML to PHOEBUS .....	419
<i>Stephanie Keyes</i>	
Variability in the areal characteristics of precipitation in the Houston, TX region .....	431
<i>Grace Kowalski</i>	
Design of high current electron transport system. ....	449
<i>Danielius Krivickas</i>	
Integrated Facility Management: Remote support for proactive building maintenance .....	454
<i>Angela Lagnese</i>	
Development of data analysis and automation routines for the Inner Shell Spectroscopy beamline.....	463
<i>John Lictro</i>	
The study of 2D ion crystals for quantum information processing.....	473
<i>Jian Jun Liu</i>	
Quantum-assisted telescopes .....	483
<i>Ryan Mahon</i>	
Cosmic rays in astronomical charge-coupled devices .....	497
<i>Jocelyn McMahon</i>	
Analysis and simulation of polarization control in quantum networks.....	512
<i>Liam McSweeney</i>	
Structural analysis of materials used for energy storage .....	527
<i>Marie Francoise Millares</i>	
Study of pad geometries in AC-coupled low gain avalanche diodes using numerical simulation .....	536
<i>Patricia Moore</i>	
Interaction energy analysis between diiron enzyme active sites and their natural substrates ..	543
<i>Ambika Natarajan</i>	
Design and implementation of novel applications for KBase .....	558
<i>B. J. Roxane Ouango</i>	
Transfer of NSLS-II conda packaging to conda-forge.....	
<i>Katherine Perez</i>	
Home-based citizen science program for middle school students to build skills necessary to conduct authentic research.....	560
<i>Alyssa Probst</i>	
A study of fail-slow fault tolerances in databases .....	574
<i>Andrei Racila</i>	
Validating the Bayesian Resonance Reclassifier with polarized-beam data .....	594
<i>Sergio Ruiz</i>	
Spin reassignment in the Atlas of Neutron Resonances.....	601
<i>Sergey Scoville</i>	

User-friendly interfaces for Jupyter Notebook data analysis scripts.....	610
<i>Ramon Sepulveda</i>	
Shielding for accelerator radiological operations .....	618
<i>Zaha Shahdad</i>	
Mechanical engineering and design for the Coherent Electron Cooling diagnostics beamline and the Electron-Ion Collider Superconducting Radio Frequency test facility upgrades.....	625
<i>Christopher Sherling</i>	
Spin matching for the Electron-Ion Collider electron storage ring .....	668
<i>Matthew Signorelli</i>	
Study charge carriers diffusion in field free silicon.....	685
<i>Mikhail Smirnov</i>	
Automated generation and machine learning analysis of a transition metal containing active sites relevant to synthetic and biological applications .....	700
<i>Robert Stahl</i>	
Anomaly detection algorithm research and testing with benchmark datasets for the performance analysis of high-performance computing applications running at exascale .....	713
<i>Julianne Starzee</i>	
Exploring geometric algebra for beamline operations.....	723
<i>Taylor Stephens</i>	
Statistical analysis of solar generation intermittency .....	738
<i>Justin Tondo</i>	
The application of current physics research in an educational setting .....	781
<i>Nkeiru Ubadike</i>	
Developing tools for analysis of the spectro-microscopic data obtained from 2D materials and surfaces.....	796
<i>Jeannet Vargas</i>	
Civil engineering system design for a superconducting RF support facility.....	812
<i>Oksana Vysochanska</i>	
Advancing computer aided design skillset to grow professionally and contribute to the National Synchrotron Light Source-II mechanical design team .....	832
<i>Matthew Warner</i>	
Environmental safety and health-business and environment.....	843
<i>Miniya Williams</i>	
Convective cloud core lifecycle over the Houston region using a cell tracking algorithm .....	853
<i>Brian Wittemann</i>	
Creating a software-defined emergency power off button .....	866
<i>Samuel Woronick</i>	
Electron-Ion Collider civil engineering requirement for the Linear Accelerator support building and tunnel .....	874
<i>Angela Zavala Yanes</i>	

Upgrading the Area Radiation Monitor infrastructure at the National Synchrotron Light Source-II .....	890
<i>Robert Zinser</i>	
Monte Carlo simulations of radiological issues for the Electron-Ion Collider around Interaction Point 12 of the Relativistic Heavy Ion Collider.....	897
<i>Jacob Zweifel</i>	
Deployment and repository deposition of multiple software packages using automation.....	911
<i>John Gedell</i>	

# **3D spatially modulated plasma dipole oscillation induced by a detuned counter-propagating laser pulse**

James Adams, David Wan,  
Electrical and Computer Engineering Technology,  
Farmingdale State College, Farmingdale, NY 11735

Paulo Castillo,  
Electrical and Computer Engineering Technology,  
Farmingdale State College, Farmingdale, NY 11735

Kwang Min Yu,  
Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973

**Abstract**

Dr. Kwangmin Yu from the Computational Science Initiative at Brookhaven National Laboratory and other researchers have developed a code called SPACE, which studies particle behavior by simulating laser wake-field plasma acceleration in plasma fields as well as ionized fields. We are using SPACE to simulate laser-plasma interactions based on counter-propagating detuned lasers and particle behavior. Previous simulations resulted in an efficient radiation source through plasma dipole oscillations where the coupling strength and pulse width of the laser electric field along preformed plasma, and a self-ionized wake-field target induced by the laser intensities were used as parameters. In our current study, we are focusing on a lower intensity by lowering the laser's pulse power to a strength still sufficient enough to ionize a neutral gas and to drive a wake-field associated with longitudinal fields in the order of GV/m (GigaVolts per meter). Under these conditions, we will observe the behaviors in the far-field radiation fields.

**Introduction**

The study of intense laser-plasma interactions has become an active and growing field of both theoretical and applied research encompassing a wide variety of different disciplines. With the growing development of high-intensity ultrashort lasers, a deeper and more detailed understanding of various novel mechanisms is critical to the development of practical high-energy particle accelerators and novel radiation sources. [1][2][3] The complex nonlinear time-varying processes that may be encountered in the interaction of laser irradiation and high concentration plasmas make analytical modeling very difficult and approximate, and most fundamental studies require sophisticated numerical modeling approaches to account for the complex interactions that charged particles carriers can have with an intense electromagnetic



(EM) pulse that can further be focused onto spatial targets. The development of numerical algorithms to model and simulate laser-plasma interactions provide a synthetic theoretical introduction of a wide range of phenomena within the interaction, including the excitation of plasma waves, plasma instabilities, wave-particle interactions, and linear and non-linear effects during laser-plasma coupling.

During this research, the simulation would be set by a pair of detuned ultra-sharp beam pulses of the length of 48  $\mu\text{m}$  between them both. The field at which the beam pulse would interact with would form a column shape in the trapezoidal domain full of plasma. This beam pulse interacting with the plasma would produce a disturbance on the charge distribution as a result of a dipole moment. At the same time, we would be starting to observe radiation from the formed dipole instance. However, after finalization from the collision, we could see the dipole oscillation radiating energy.

## **Methods**

### **A. PIC**

The particle-in-cell (PIC) method is the tool of choice for the investigation of non-linear processes in kinetic and collision-less plasma processes and laser beam physics related by Maxwell Equations. Specifically, we choose to use the electromagnetic approach to the PIC method over the electrostatic form even though both use similar Finite Difference Time Domain (FDTF) approaches for approximating Maxwell's equations. However, particle-in-cell electrostatic (PIC-ES) does not provide reasonable solutions at relativistic speeds and cannot compute for large amounts of radiation. Another reason for choosing the particle-in-cell electromagnetic (PIC-EM) method is the necessity to create a history of particle data, where the

PIC-ES method only has to update and calculate maxwell equations every unique time division, the PIC-EM method must store previous data and use those values for the next time division.

## B. SPACE

The SPACE code is a relativistic and parallel 3D electromagnetic code being developed at Brookhaven National Laboratory (BNL). We simulate the laser-plasma interactions by setting real-world parameters for simulations we need to obtain data from. Parameters we regulate include the domain size of the simulation, theoretical simulation time, number of grid divisions for the PIC method, amplitude, frequency, and focus radius for each beam. Using the SPACE code, we integrate the PIC method to numerically generate a grid that simulates these charged particles or plasma matter. The simulations aim to generate a standing localized plasma dipole oscillation using two detuned highly-intense ultrashort laser pulses in a domain of previously formed plasma. The two laser pulses are propagated towards each other in the grid domain. The dipole is formed upon the moment when these two laser pulses collide and radiate energy in levels of 9 GeV at frequencies within the terahertz range ( $10^{12}$ ).

### System

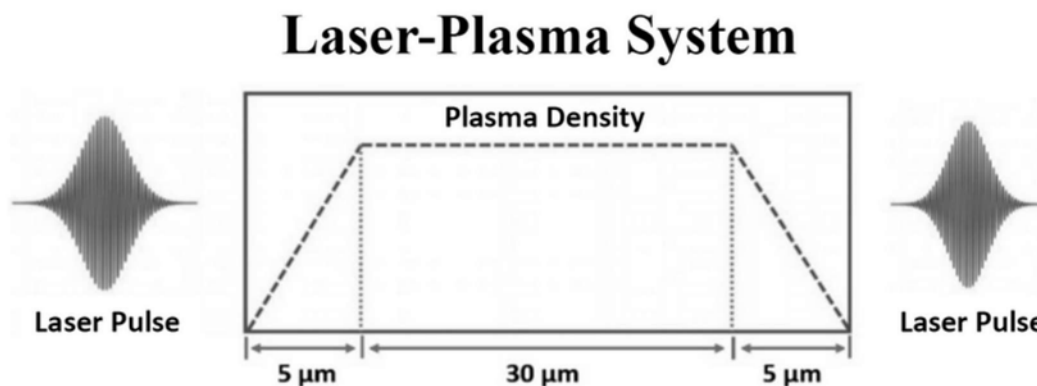


Figure 1: Schematic diagram of the laser pulse and plasma density profile along the pulse propagation direction. The preformed plasma has a cylindrical shape of  $40\ \mu\text{m}$  length and  $5\ \mu\text{m}$  radius. The plasma density ramps up linearly from both edges and the center part has constant density.

We set up a 3D cubical spatial domain of 48  $\mu\text{m}$  simulation length, corresponding to a theoretical beam propagation time from side to side of 160 fs. Within this domain, the first and last 4  $\mu\text{m}$  is set as vacuum, followed by the actual geometry of the plasma column, which has a trapezoidal shape corresponding to a linear ramp-up/down behavior of 5  $\mu\text{m}$  for the incoming (left side) and outgoing (right side) traveling beam, and a flat volumetric region of 30  $\mu\text{m}$  length of effective plasma channel as in Fig. 1.

Specifically, the vacuum and linear ramp model are established in the domain since first, the system has no physical boundaries, and second, when simulating the interaction volume, we use a gradually linearly increasing tapered edge to reduce any reflection artifacts that will occur when outgoing radiation modes are propagated from the induced plasma dipole sources. Therefore any reflection is considered numerical noise and not due to physical boundaries, which are undesired and unphysical.

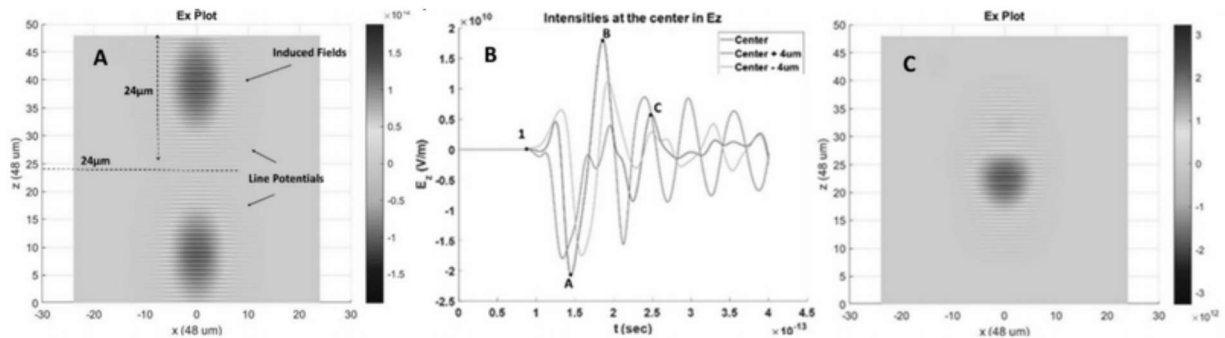


Figure 2: A) On-axis ( $x=0$ ) transverse induced fields prior to collision. B) Electric field as a function of propagation time along the longitudinal axis. C) Axial view of the plasma dipole formation during the laser's induced field collision.

Figure 2A shows the counter-propagating laser pulses parallel to the z-axis. Both laser's beam profiles include a beam waist radius of 5  $\mu\text{m}$  and beam duration of 30 fs. Gaussian profile short beams with a 5  $\mu\text{m}$  radius allow the formation of a cylindrical channel within the total plasma domain, and therefore the main slab to account for laser-plasma interaction. In addition,

the numerical setting for the plasma channel is modeled in a grid domain to account for the number plasma density of  $4.97 \times 10^{18} \text{ cm}^{-3}$  as in [1].

## Experimentation

A new test is being done with the simulation to better understand the laser-plasma interactions calculated by the SPACE code. By altering the frequency intensity setting of the laser pulse, the pattern of the electric field intensity over time no longer matches the expected results as seen in figure 2B. This inconsistency has resulted in an investigation where the data gathered from the supercomputer simulations are compared to previous results through the use of matrices. Another data analysis technique was through the creation of MATLAB programs to compare the simulation results to these theoretical models. The data matrix created by this MATLAB simulation showed where the point of maximum electric field occurs in the interference pattern. The values used for each laser pulse are shown in Table 1.

<b>Preformed Plasma</b>	<b>Ionized Plasma</b>
Laser Pulse (Gaussian Beam) Specs:	Laser Pulse (Gaussian Beam) Specs:
Left Laser	Right Laser
Frequency: 3.7474e14 Hz	Frequency: 3.9446e14 Hz
Beam Waist: 5e-6 m	Beam Waist: 5e-6 m
Amplitude: 2.41e10 V/m	Amplitude: 2.53e10 V/m
Duration: 3.53223e-14 s	Duration: 3.53223e-14 s

Table 1: Simulation settings from the recent simulation run

## Results

During the internship, the data collected from the simulations generated by the SPACE code were analyzed by comparing the collected data to previous simulation results and

theoretical models. Unfortunately, the analysis conducted over the course of this internship was not able to provide satisfactory results. The Center electric field results in the diagram (Figure 3) are shown below.

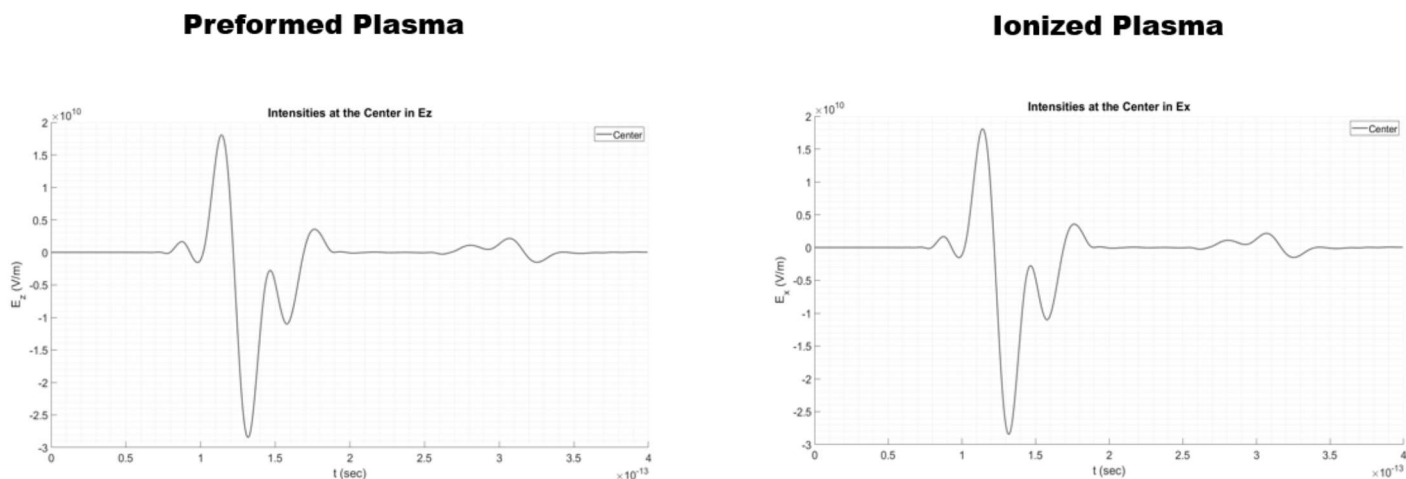


Figure 3) Electric field measurements at the center of the simulation plot

The results from Figure 3 yield a dipole formation starting at  $1e-13$ s; however, the relaxation is too short. This makes the dipole oscillation frequency difficult to determine. The main problem here is that the time step for the traveling laser pulses is 40fs. This means that values for the electric field are skipped over, for reality is continuous. More computational power is needed to properly study these lower intensity models.

### Future Plans

More research is needed to better understand the recent simulations. The resulting data does not match the results from previous simulation runs. The conclusions from this data need more time for analysis in order to determine what interactions have occurred. One way to better understand the data is by increasing the resolution of the time step. A smaller time step would increase the resolution. The increase in the resolution would allow for more data to work with.

Our future goal is to implement the non-linear effects of quantum electrodynamics (QED) on laser-plasma interactions over preformed or ionized targets. Research is being done to



understand quantum mechanics and quantum computing so that future simulations can use quantum computers to calculate the laser-plasma interactions. The computations done within a quantum computer would be more accurate compared to the results received from the current supercomputers. The randomness of the qubits used in a quantum computer would more accurately reflect the random nature of a real-life simulation using particles. The plan is to use the open-source software known as Qiskit as the new platform to run simulations. Currently, research is being done to translate the SPACE code into Qiskit.

### **Acknowledgments**

We would like to thank our mentors, Kwang min Yu and Paulo Castillo, for their kind support and their helpful guidance while working on this project. We would also like to thank the Office of Educational Programs (OEP). This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Visiting Faculty Program (VFP) and the Science Undergraduate Laboratory Internships Program (SULI). This research used the computational resources from the Stampede2 supercomputer with allocation by the Extreme Science and Engineering Discovery Environment (XSEDE) program and computational resources from the Seawulf supercomputer with allocation given by Stony Brook University.

**References**

- [1] K. Kwon et al., "High-Energy, Short-Duration Bursts of Coherent Terahertz Radiation from an Embedded Plasma Dipole," *Sci. Rep.*, vol. 8, 2018. doi:10.1038/s41598-017-18399-3
- [2] P. Castillo et al., "Numerical Study of Coherent Radiation from Induced Plasma Dipole Oscillation by Detuned Laser Pulses," 2019.
- [3] P. Castillo et al., "Numerical simulation of terahertz radiation by laser-driven plasma dipole oscillation," 2020. doi: 10.1117/12.257628

# Anomaly detection algorithm research and testing with benchmark datasets for the performance analysis of high-performance computing applications running at exascale

Arnav Agrawal, Department of Physics, Cornell University, Ithaca, NY, 14850  
, Dr. Line Pouchard, Computational Science Initiative, Brookhaven National Laboratory,  
Upton, NY, 11973

---

## Abstract

Simulations of physical phenomena, such as climate modeling, high energy physics, and galaxy interactions, are driven partly by high performance computing (HPC) systems run at exascale. These simulations can run for days, even weeks, producing a large volume of output data, called trace data, that cannot practically be stored and analyzed after the completion of the program. Instead, the trace data must be analyzed in situ, or in the memory of the computer. Efficiency of the simulations is pivotal, and for this, we must look at run times of individual functions within the program. This data allows researchers to detect and understand performance anomalies that can come about as a result of simulation parameters, workflow contention inside the computer, and slight inaccuracies in the underlying code. In this project, we study two anomaly detection (AD) algorithms, Histogram-based Outlier Score (HBOS) and Streaming Standard Deviation based anomaly detection (SSTD), and validate their behavior using various synthetic data sets created in Python. We use Chimbuko, a framework that analyzes the trace data of a program, to simulate running the data in streaming mode, that is, in batches instead of in bulk (1). We find that the HBOS algorithm surpasses every other algorithm per the AUC metric, especially after a small initialization step. As a result of this project, I have learned many new concepts in data science and machine learning. In addition, I have become a far more effective scientific communicator.

---

## 1. Introduction

Anomaly detection is the process from which we find data points that are different from the rest of the data set. This process of finding outliers is relevant to every field of study. In scientific fields in particular, with the steady rise of computing power, more and more scientists are using coding as a tool to drive forward their research. It therefore becomes more important every day for scientists to be confident that their programs are running efficiently and as intended. Many of these programs run for days, weeks, even months, as they simulate complicated phenomena such as the spread of Covid-19 across the globe. The trace data produced by these programs is far too voluminous to be analyzed in bulk, instead, it must be analyzed in situ, in batches. Chimbuko is a performance analysis framework that does just this. It uses anomaly detection algorithms to detect performance anomalies of a program in situ.

Our project focuses on testing two anomaly detection algorithms to determine which is better at detecting global outliers, and limiting the number of local outliers detected. Global outliers are anomalies that are different from the entire data set. Local outliers are anomalies that are different from the rest of their batch (keeping in mind that we are splitting up each larger data set into batches, the process of which I will describe later).

The two algorithms we focus on are

- HBOS [2], groups data points into various bins based on their value to create a histogram
- SSTD: Considers data points to be an outlier if they are greater than X standard deviations away from the mean of the data (in this case 12)

There are other algorithms we looked at, but these two were our main priorities.

In what follows, I will talk about preliminary research (Section 2), discuss methods (Section 3) and results (Section 4), and reflect back on the project as a whole (Section 5).

## 2. Preliminary Research

Before working with with the HBOS and SSTD algorithms, we tested other algorithms from the clustering library in NumPy. These included the DBSCAN and BIRCH algorithms. We also tested a couple of other well-known algorithms such as moving quartile, isolation forest, and local outlier factor.

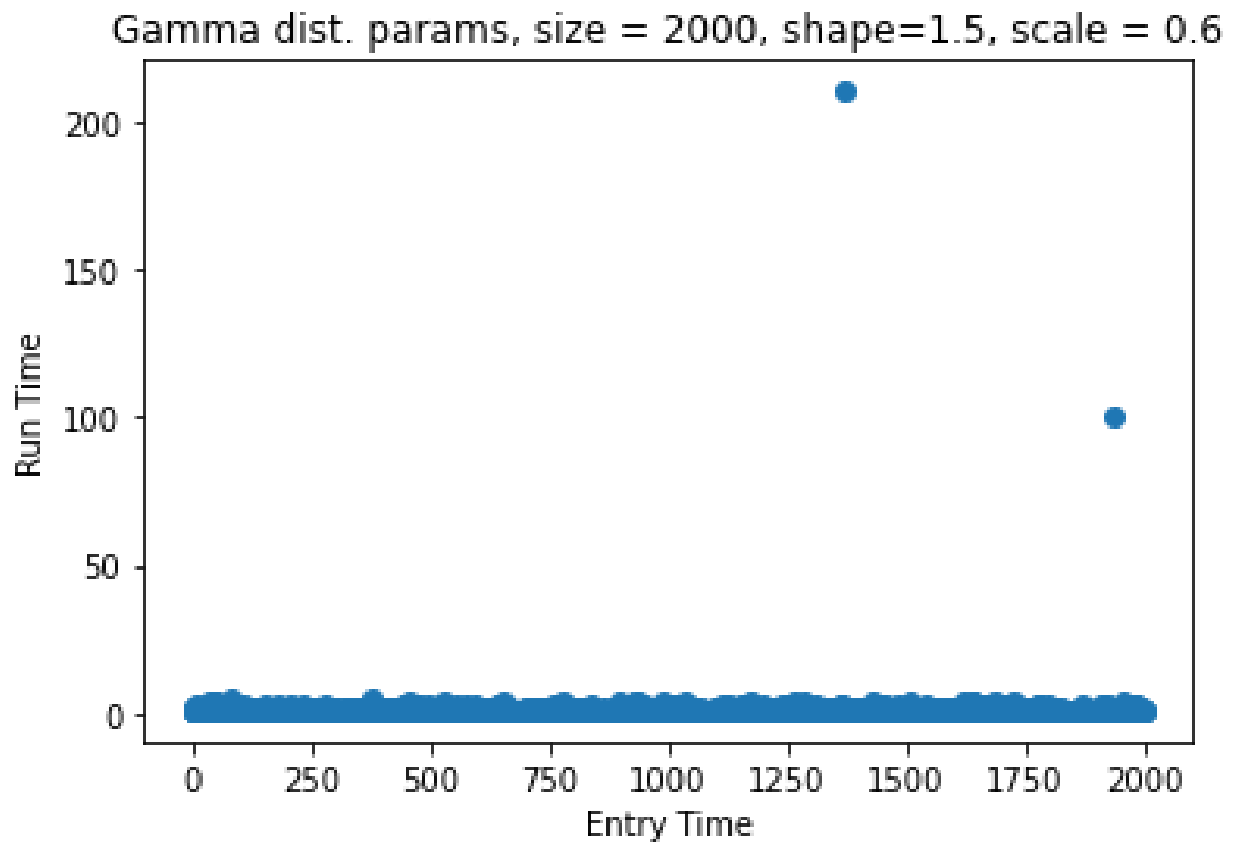
In order to test the efficacy of these algorithms, we created synthetic data sets in Python, an example data set of which is shown below in Figure 1. This data set was created using the `numpy.random` library, specifically the gamma distribution function. The two clear outliers were added manually, albeit in random locations in the data set. Imagining that we have plotted a time series of functions and their run times, the two axes are accordingly labeled "Entry Time", denoting the time stamp in milliseconds of when that function ran, and 'Run Time', denoting the run time of that function.

There was a varying level of success with which the aforementioned clustering algorithms were able to detect outliers in our data sets. The success could also be varied by changing the parameters of the algorithms, and *different* parameters would often result in *different*

algorithms performing well for *different* data sets. That is to say, the same algorithm would not necessarily perform well for various data sets, we would need to manually change its parameters to optimize performance for different data sets. While the specifics are not too important, the fact that these algorithms required user adjustment to optimize is a very important point to note.

The HBOS algorithm, on the other hand, does not require any user adjustment. Furthermore, it is an unsupervised algorithm, meaning that given an unlabeled training data set, it can pick out the anomalies and normal data points. A description of the HBOS algorithm, and a Python implementation of it, was given to us [1].

Our project focuses primarily on testing HBOS and SSTD, using a Chimbuko simulator, to determine which is better at detecting global outliers, and limiting local outliers.



**Figure 1:** Example time series synthetic data set, sampled from a gamma distribution. Notice the two clear outliers.

### 3. Methods

#### 3.1. Procedure

Concisely, the steps taken for the project were as follows:

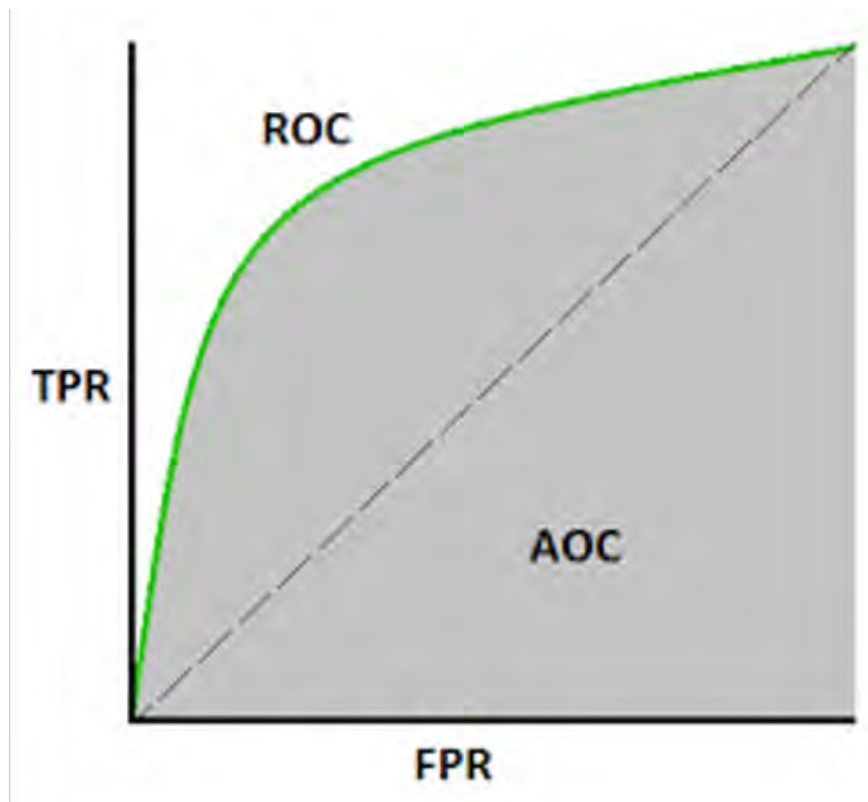
1. Understand how Chimbuko works by reviewing literature



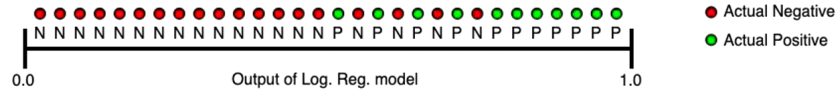
2. Look at clustering-based algorithms in the Sci-kit learn cluster module to understand why these algorithms are inferior for our purposes
3. Review the literature on HBOS to see why it is expected to detect more outliers on a greater array of data set
4. Create a variety of synthetic data sets in Python using the numpy.random library
  - (a) Done by sampling 2000 points from a distribution, such as a Gamma distribution, and manually adding global outliers to the data set
  - (b) Varied the number of outliers added, the ‘center’ of the distribution, the number of modes, and the type of distribution (gamma, normal, etc.)
5. Pass each data set into the Chimbuko simulator, running the HBOS and SSTD cases separately
6. Evaluate the run using the ROC-AUC metric, defined in the next subsection

### 3.2. ROC-AUC Evaluation Metric

We evaluated the performance of algorithms using the ROC-AUC metric, an industry standard. This metric is used to predict the probability that the anomaly detection model ranks a positive instance more highly than a negative instance, as pertaining to the anomaly score of the instance. The ROC is plotted using the true positive rate and false positive rate of the algorithm, and then the area under that curve is calculated, giving the AUC. A diagram is shown in Figure 2 and Figure 3.



**Figure 2:** Graph showing the calculation of AUC from ROC. Graph from <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>.

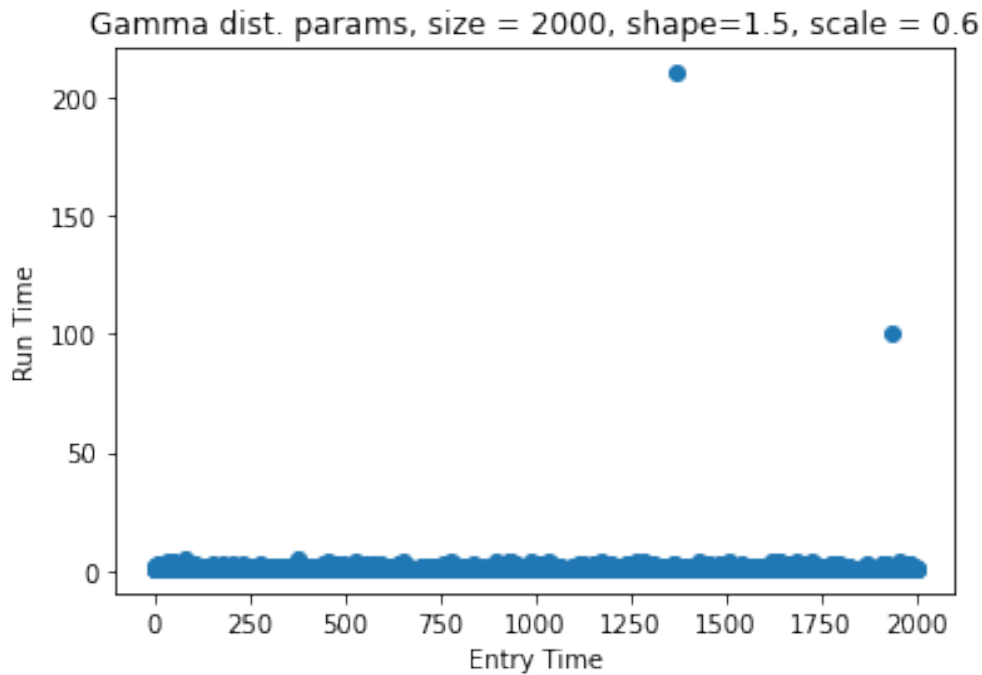


**Figure 3:** Graph showing examples of anomaly scores on a line plot. An AUC of 1.0 means that all the anomaly scores of actual anomalies are higher than the anomaly scores of normal data points. Graph from <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>.

#### 4. Results

After extensive testing, we found HBOS to be significantly better at detecting anomalies than SSTD, especially with multimodal data sets. We show AUC score of some sample data sets below, along with plots of some of the relevant runs, before and after using Chimbuko w/ the two AD algorithms.

Here are some example data sets, in blue, and the Chimbuko output, in black and red, with the red being outliers detected by the AD algorithm within Chimbuko.



**Figure 4:** Sample single modal data set

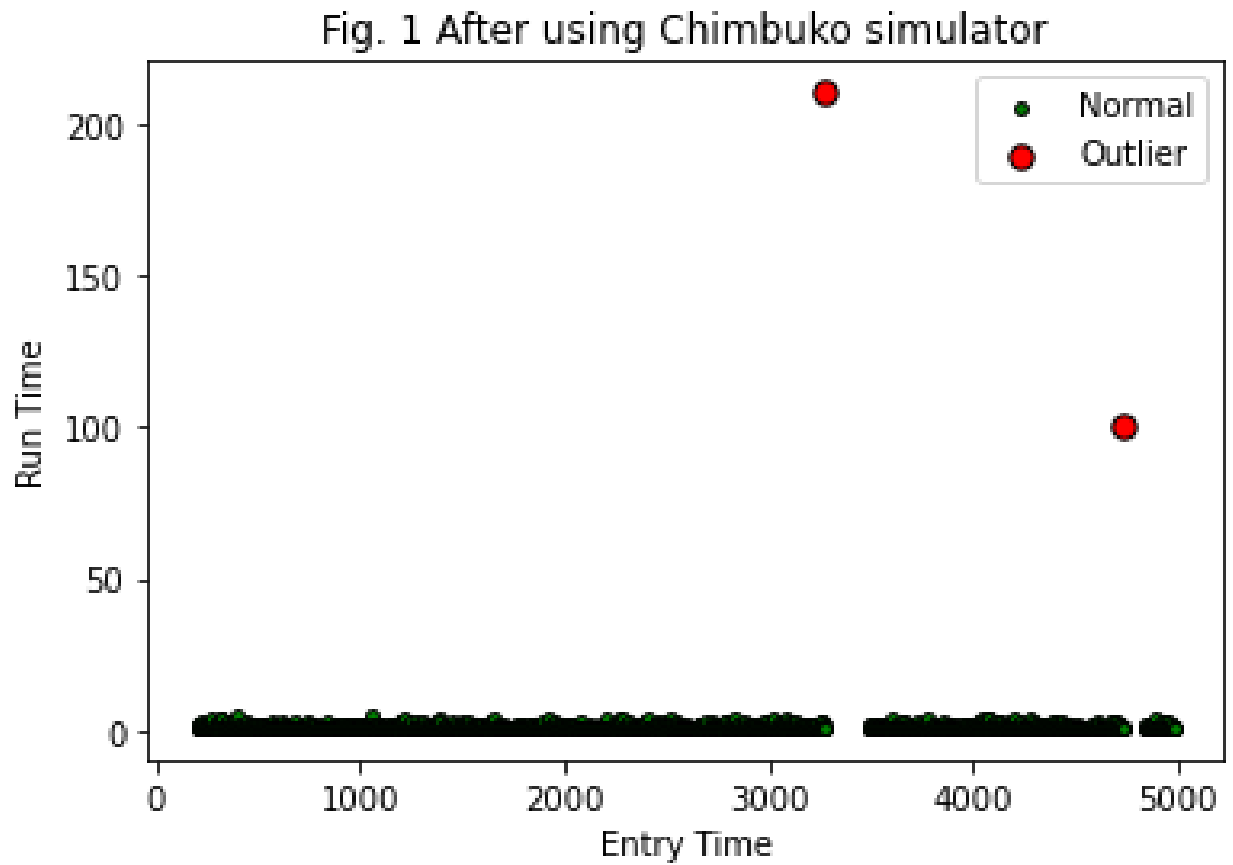


Figure 5: HBOS and SSTD output using Chimbuko of Figure 4

Gamma 1 params, size = 2000, shape=1.5, scale = 0.6  
Gamma 2 params, size = 2000, shape=500, scale = 0.6

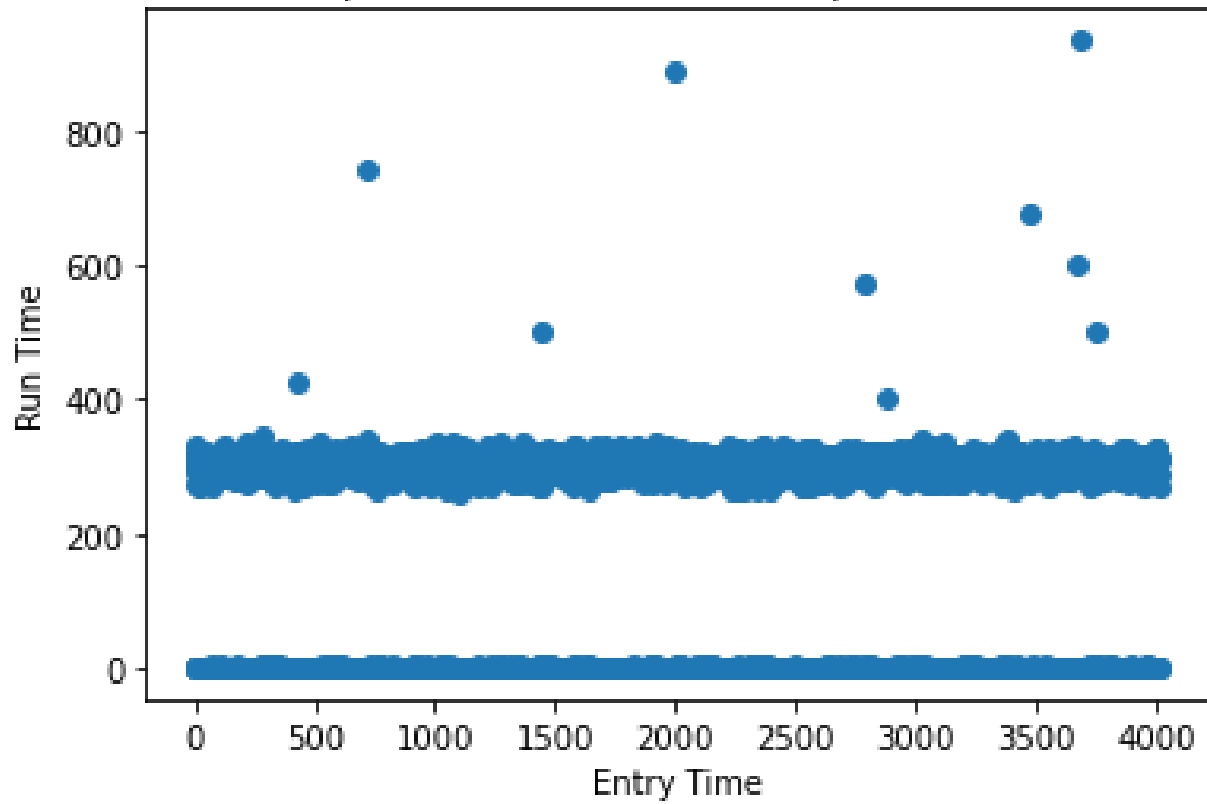


Figure 6: Sample two-mode data set

Fig. 1 After using Chimbuko simulator

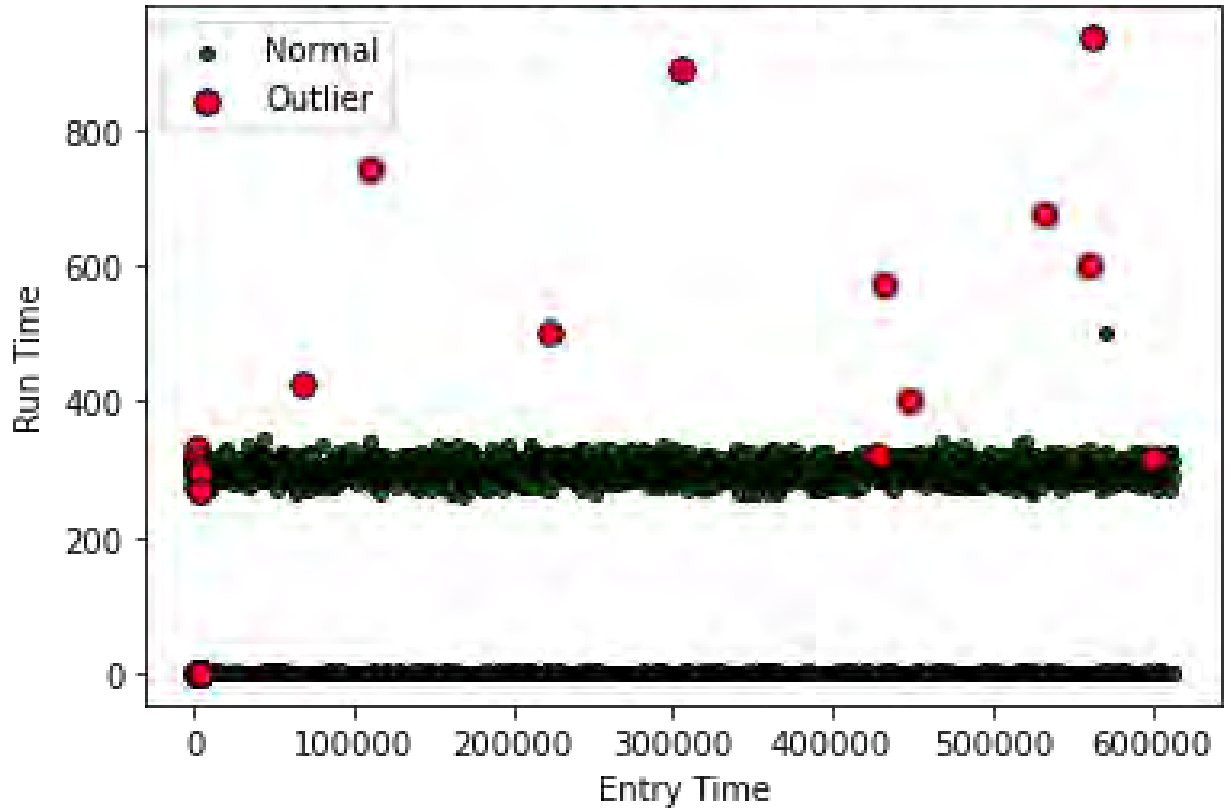


Figure 7: HBOS output using Chimbuko of Figure 6



Fig. 1 After using Chimbuko simulator

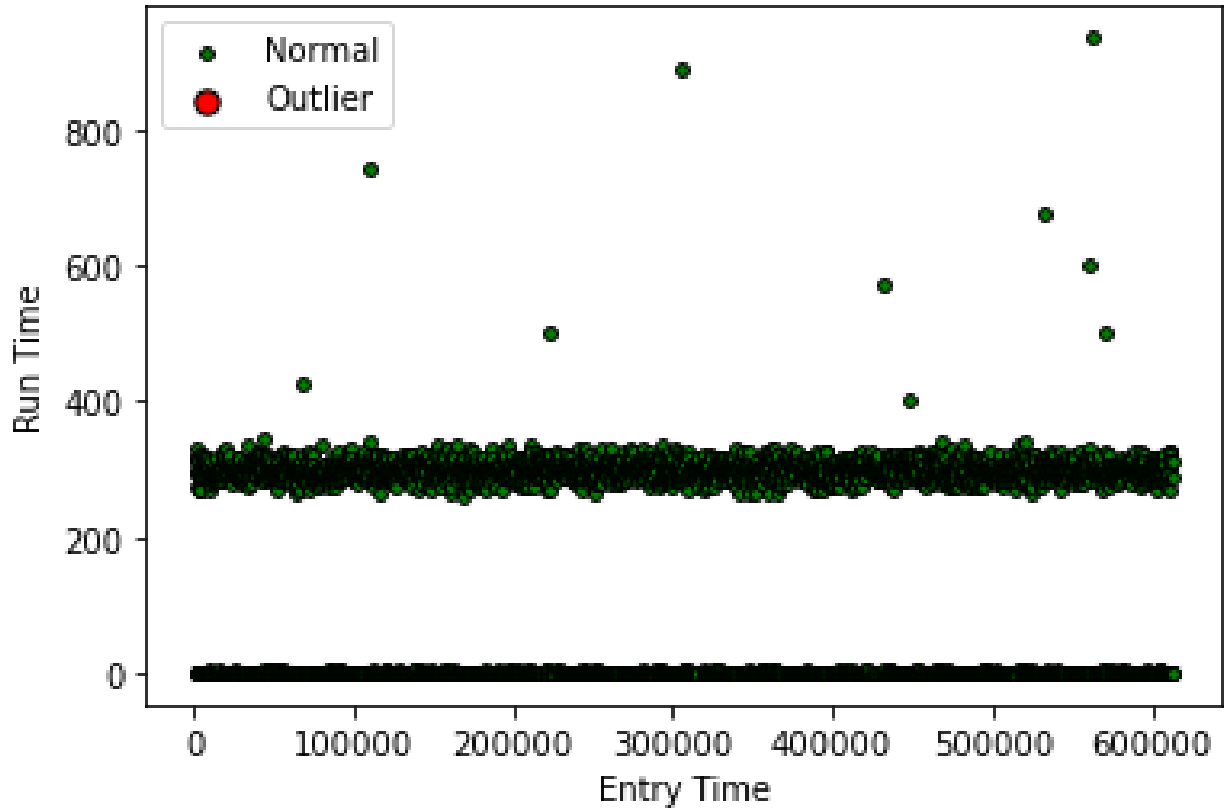
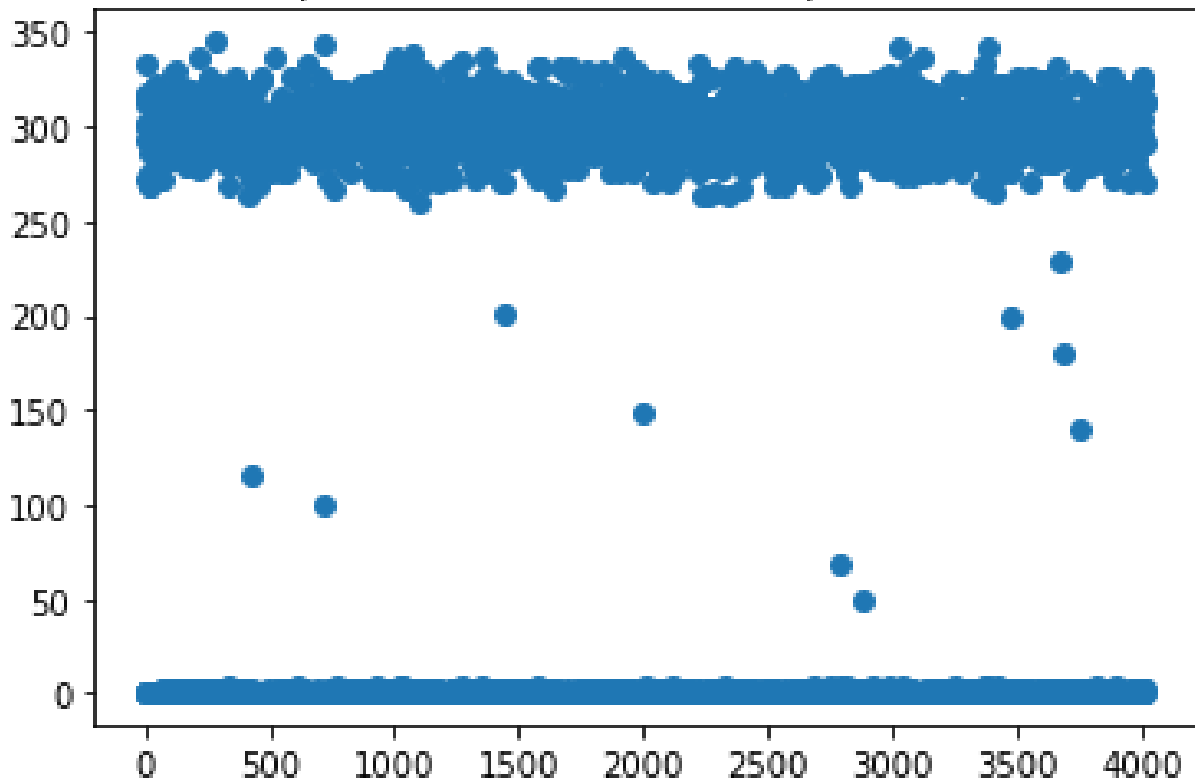


Figure 8: SSTD output using Chimbuko of Figure 6

Gamma 1 params, size = 2000, shape=1.5, scale = 0.6  
 Gamma 2 params, size = 2000, shape=500, scale = 0.6



**Figure 9:** Sample two-mode data set with outliers in between the distributions, as opposed to on the 'end'

This table summarizes the results well.

Description of Data Set	HBOS AUC	SSTD AUC	True Positive Rate (HBOS)	True Positive Rate (SSTD)	False Positive Rate (HBOS)	False Positive Rate (SSTD)
Single Modal Gamma distribution	1.0	1.0	1.0	1.0	0.0	0.0
Bimodal Gamma distribution, outliers at the end	0.946	0.50	0.9	0.0	0.007	0.0
Bimodal gamma distribution, outliers in the middle	0.497	0.50	0.0	0.0	0.007	0.0

**Figure 10:** Summarized results of the runs above. Notice how HBOS is much better at detecting anomalies in the multimodal case, but neither algorithm was able to handle the anomalies in the 'center'.

## 5. Conclusion

After extensive testing, we concluded that the HBOS algorithm is superior to the SSTD algorithm for anomaly detection, as expected. This is due to several reasons. Although the algorithms performed similarly in single modal testing, HBOS was far better at detecting anomalies in multimodal data sets, as seen in Figures 4 and 5 above. This came at the cost of slightly more false positives, as shown in the FPR column of the results table above. However, we noticed that the AUC metric placed far more emphasis on the TPR than the FPR, allowing HBOS to score a high 0.9 for this metric on the bimodal data set, compared to the 0.5 scored by SSTD. In addition to this quantitative analysis, there are a few perks of using HBOS that should be considered. First of all, HBOS is an unsupervised algorithm, meaning that it needs no training data set to be used. Further, the user is not required to manually set an algorithm parameter (e.g. the number of standard deviations away from the mean in SSTD) as in most clustering algorithms. We notice that HBOS struggled to detect outliers with values in between the two modes. Future studies should focus on this and determine what about the internal histogram structure is causing this.

### 5.1. Acknowledgments

I'd like to thank Brookhaven National Laboratory for this summer opportunity, as well as my mentors Line Pouchard, Sandeep Mittal, and Christopher Kelly for their help and guidance throughout this project.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Programs (SULI).

## References

- [1] Goldstein, M., & Dengel, A. 2012
- [2] Ha, S., Jeong, W., Matyasfalvi, G., et al. 2020, CoRR, abs/2008.13742.  
<https://arxiv.org/abs/2008.13742><https://arxiv.org/abs/2008.13742>

Performance analysis of tomography algorithms

Jabel Alcantara

Computer Information Systems, Kingsborough Community College, Brooklyn, NY 11220

Thomas Flynn

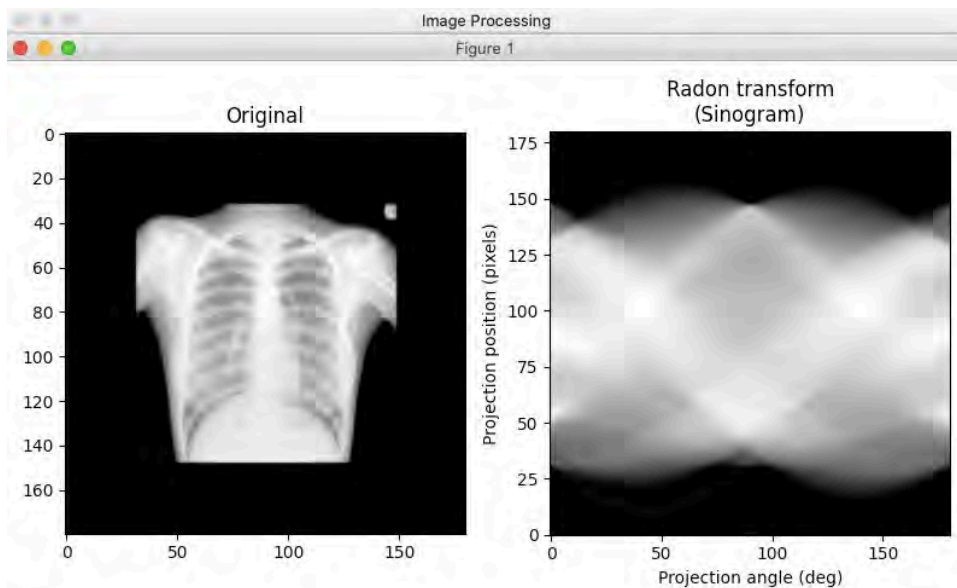
Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973

## **Abstract**

The process of analyzing the performance of tomography algorithms is based on reconstructing a 3D (or three dimensional) structure, from 2D projections. Tomography is a technique of displaying the representation of the interiors of the human body or other 3D objects, using x-rays or ultrasound. The process to create these images is based on mathematical reconstruction algorithms. We have written code in Python, a high-level programming language. It has built-in data structures to handle big data and perform complex mathematics. Moreover, we'll be using NumPy, Numerical Python, the core library in Python used for scientific computing. It allows for efficient operations on data structures used in machine learning, such as vectors and matrices. Since computers portray images as a sequence of numbers, 2D images can be represented as a matrix. That led to the start of implementing fundamental image processing techniques. I created a Graphical User Interface (GUI) application that allows a user to easily access different operations (Brighten, Darken, Invert, Blur) and perform those operations on an image of their choice. I've also placed buttons on the application that allow you to create the sinogram of an image, which is then used to create a tomographic image. We began analyzing the Filtered Back Projection (FBP) algorithm, a commonly used algorithm in Tomography. An important adjustable parameter for the FBP are the 5 filters. I collected a set of ten images and created a table to record the root mean square (RMS) error for each filter. RMS measures the accuracy of the reconstructed image. On average, the 'Ramp' filter contained the lowest RMS error, and the 'Hann' filter contained the highest. As a result of this summer, I've been exposed to many new features in Python as well as enhancing my programming skills.

## **Background**

Computed Tomography (CT) is a technique based on x-ray transmission through an image to create 3D images from a sequence of 2D projections. Tomography has many applications in the real world with ties to the medical and materials science fields, for instance. In 2D Tomography, the images of sections are formed by generating a set of parallel rays through a 2D image. In digital imaging, the smallest controllable element of a picture is a pixel. Essentially, each ray goes through a single pixel in the projection. The Radon transform is the creation of an image from the collection of several projections that are created by the rays. In other words, it is a linear transformation of the original image as shown in Figure 1.



**Figure 1.** *An original image and its Radon transform.*

There are two major categories of reconstruction methods that exist: analytical reconstruction and iterative reconstruction. This project focused on the analytical reconstruction of images. The Filtered Back Projection (FBP) algorithm is a classic technique for computerized tomographic imaging. It's well known in part because of its computational efficiency. The FBP algorithm requires choosing a filter, which can impact performance. I wrote code in python, which allows for efficient operations on data structures, to compare the different filters in the algorithm.

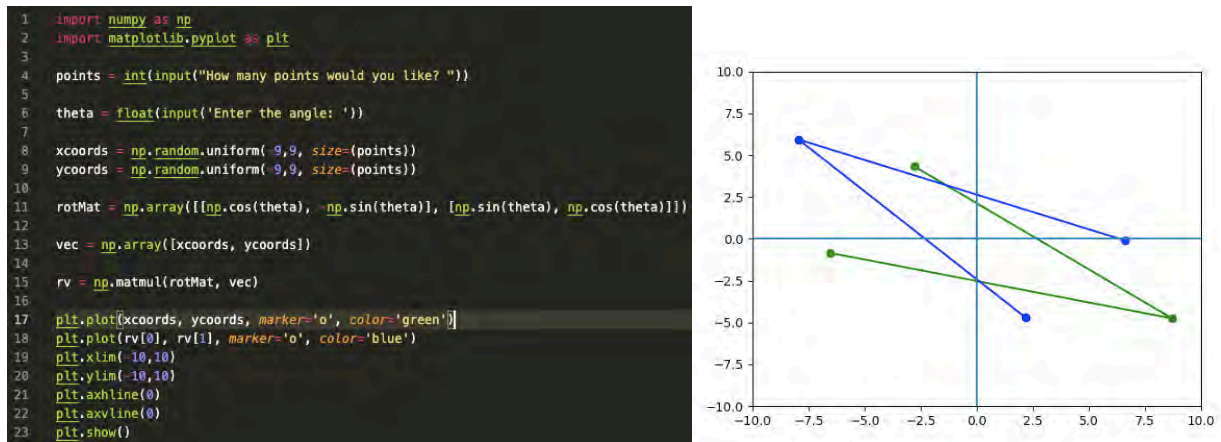
## Methods and Materials

At the start of my 2021 summer has internship, I was unfamiliar with most terms in relation to Tomography, but I was familiar with Python. Its high-level built-in data structures to handle big data and perform complex mathematics. In Python, we accessed the Numerical Python (NumPy) library which added support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. As previously stated, NumPy is the core library for scientific computing in Python. I had to install many libraries, along with NumPy, onto my laptop in order to access these functions. Some of these libraries included: Matplotlib, PIL, Skimage, and PyQt6. In order to create a program that allows you to process images, this was necessary. With the help of my mentor, this included writing commands in my computer's terminal to download these libraries.

I started off by practicing basic linear algebra computations, which consisted of writing programs to create a matrix and a vector and compute the matrix vector product. Matrix multiplication is a binary operation that creates a matrix from two matrices. This is relevant for Tomography because images can be represented as a matrix. In Figure 2.1, I performed basic operations on 2D vectors that focused on rotating points on a graph, by a given angle, using a built-in plotting library in Python called Matplotlib, to create animated and interactive visualizations. The rotation of a vector can be expressed using the following equation shown in figure 2.

$$Rv = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{bmatrix}.$$

**Figure 2.**  
*The Rotated Vector expression*



**Figure 2.1.** On the left, is the code written that produces the original points (green), then after the Rotated Vector expression, it plots the rotated points (blue).

I then had to practice solving systems of linear equations, solving  $Ax = b$ . That is, you have a matrix  $A$  and a vector  $b$  and wish to find a vector  $x$ , in which  $Ax = b$ . Although there are many iterative algorithms that solve this expression, I compared two algorithms that can approximately calculate a solution:

- For  $t = 1, 2, \dots, 1000$ :

$$x_{t+1} = x_t - u (A^T Ax - A^T b)$$

- For  $t = 1, 2, \dots, 1000$ :

$$x_{t+1} = x_t - u (A^T (Ax - b))$$

The top algorithm is the original, while the one under it is a simplified version that uses less computations. The code in Figure 3, represents the simplified version. So, what I was trying to find was the speed of both programs, or more precisely the speed of the calculations in the algorithms above. After running each program 10 times, I found that the original algorithm took 52.90 seconds longer on average. I used the ‘time.time()’ function in Python in order calculate the speed of the calculations.



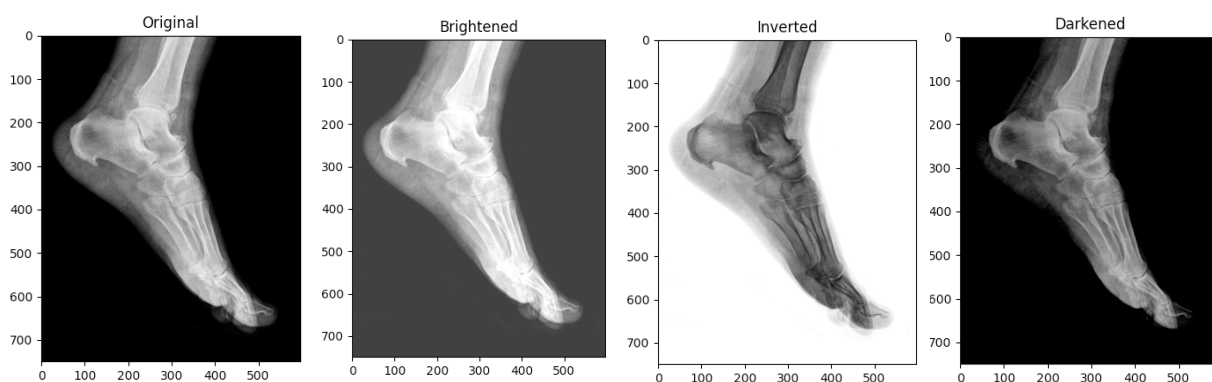
```

20 u = 1/(size * (np.max(np.abs(matA)) ** 2))
21
22 x0 = np.zeros(size)
23
24 print(x0)
25
26 beg = time.time()
27
28 for t in range(1, 1000):
29     x1 = x0 - u * (np.matmul(np.transpose(matA), (np.matmul(matA, x0) - vecB)))
30     print(x1)
31     x0 = x1
32
33 end = time.time()
34 speed = end - beg
35 print(speed)

```

**Figure 3.** The code written to time the simplified version of the original algorithm.

That led to the start of implementing basic image processing techniques. For the grayscale images we are using, the pixel values range from 0 (black) to 255 (white). In order to implement these image processing procedures, I used NumPy, Matplotlib and PIL (Python Imaging Library). I worked on four different operations: darkening an image, brightening an image, inverting an image and blurring an image. The results are shown below in Figure 4. Since we're dealing with the grayscale images, an approach to brightening an image is merely just making the pixel values larger. Similarly, to darken an image, you want to make the pixel values smaller. To invert an image, I was given the formula 'newPixel = 255 - oldPixel' in order to make the light pixels dark and the dark pixels light.



**Figure 4.** After processing the original image through the desired functions, this is the result.

The next steps were adding these functions to a Graphical User Interface (GUI) application (Figure 5) that lets a user do these operations. In order to do this, I had to download the PyQt6 library which basically sets up the application's framework. I created buttons on the application window that lets a user easily access these operations. Along with those buttons, I added a couple more; a button so that user is able to select an image, one to get the Sinogram of an image and a button for each of the 5 filters in the Filtered Back Projection algorithm. In order to get these operations to access the selected image by the user, I sent the array of the selected image to the desired function if the button was clicked. In order to keep that image, I had to replace the array of the selected image with the array of the new image that is being processed.



**Figure 5.** *The GUI application with a randomly selected image being displayed.*

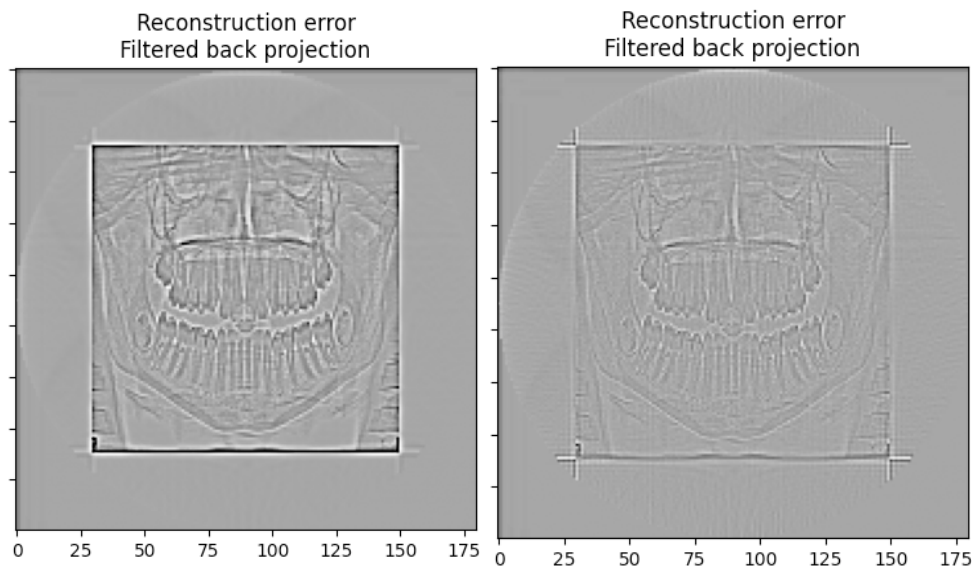
## Results and Discussion

I then began analyzing the FBP algorithm and its filters. The filters in this reconstruction algorithm are used to reduce high frequency noise in the projection images. Filtering is a technique for modifying or enhancing an image. Filters can very much improve the image

resolution, in fact, the goal of the filters is to compensate for loss of detail in the image while reducing the noise.

Ramp ( $\sigma$ )	Cosine ( $\sigma$ )	Shepp-logan ( $\sigma$ )	Hamming ( $\sigma$ )	Hann ( $\sigma$ )
.0208	.0305	.0234	.0305	.0365
.0126	.0191	.0141	.0222	.0233
.0157	.0205	.0172	.0229	.0237
.0185	.0289	.0210	.0335	.0351
.0198	.0320	.0299	.0375	.0395
.0152	.0203	.0167	.0230	.0238
.0133	.0180	.0145	.0203	.0211
.0198	.0304	.0227	.0353	.0369
.0299	.0529	.0360	.0627	.0661
.0112	.0133	.0118	.0147	.0150
<b>Avg = .0177 ± .005</b>	<b>Avg = .0266 ± .011</b>	<b>Avg = .0200 ± .008</b>	<b>Avg = .0307 ± .014</b>	<b>Avg = .0321 ± .016</b>

**Figure 6.** This table shows the recorded RMS error after reconstructing a set of ten images with each filter.



**Figure 7.** The reconstructed image on the left was created using the ‘Hann’ filter. The one on the right represents the reconstructed image with the ‘Ramp’ filter.

After collecting a set of ten images I created a table to record the RMS error for each reconstructed image after processing it through the FBP algorithm with each filter. I first downloaded example images from google, then I had to create the sinogram of each image. ‘Ramp’ as you can see in the table in Figure 6, had the lowest RMS error by average, as well as the lowest standard deviation, while ‘Hann’ has the highest. After creating the Sinogram, I was then able to do the reconstruction as shown in Figure 7. To calculate the standard deviation, I created a table using Excel and was able to insert the standard deviation function to calculate how dispersed the data is in relation to the mean. I could conclude that Ramp is the best filter to utilize in order to suppress high frequency noise in the reconstructed image. Although, ‘Hann’ had the highest RMS error, to the human eye it seems like it creates the better image, in my opinion.

All in all, I would like to take the time to thank my mentor, Thomas Flynn for his patience with me and his teaching ability. I wouldn’t have been able to complete this project without his assistance and as a result I was exposed to new programming skills, software, and techniques. This project was supported in part by the U.S. Department of Energy, Office of

Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program (CCI).

## References

1. Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. *scikit-image: Image processing in Python*. PeerJ (2014). URL <<https://scikit-image.org/>>
2. <https://numpy.org/doc/1.21/user/numpy-for-matlab-users.html>
3. [https://matplotlib.org/stable/tutorials/introductory/sample\\_plots.html](https://matplotlib.org/stable/tutorials/introductory/sample_plots.html)
4. <https://doc.qt.io/qt-6/qtwidgets-index.html>

Visualization of structured threat information expression for malware analysis

Jeremy N. Anderson, Department of Mathematics/Computer Science, Saint Joseph's College  
L.I., Patchogue, NY 11772

Meng Yue, Interdisciplinary Science Department, Brookhaven National Laboratory, Upton, NY  
11973

## I. Abstract

The study of cyber security incorporates visualization tools and techniques used to assist research malware analysis including threat information expressions. In our research, we use a virtual operating system as our primary environment to ensure safer and more secure malware analyses. The tools installed in this environment, STIG, Stix-Viz, and Stix elevator allow users to input a threat model file and visualize its contents and relationships. STIG also allows users to construct these relationships if missing. The Stix elevator allows its user to convert a .xml file into a .json through Python script. STIG only functions with .json file types and Stix-Viz only functions with .xml files. We first acquire the threat information from the CISA (Cybersecurity and Infrastructure Security Agency) website ([use-cert.cisa.gov](https://use-cert.cisa.gov)), and then convert this information into a format the tools can understand. We visualize and compare these threats with other threats to better understand their structure and implications. Results show similarities and differences which we can use to identify threat types and their source. By analyzing malware and visualizing structure threat information and expression, we can better share the threat information to understand how these threats work and create methods to defend against emerging malware and threat issues within cyber security. Through this research I can try new tools that allow me to work closely with old, new, and emerging data related to cyber security attacks.

## II. Introduction

A threat information visualizer assists its users by displaying threat information in varying formats such as graphs and charts. This helps to organize extensive code and provides a visual approach to understanding threat information associated with cyber-attacks using different techniques such as malware. These visualizers shorten the process of categorizing extensive code and overall reduces time spent during research and detection of new threats. By understanding threat information faster, researchers can develop better defense against these emerging cyber threats.

This experimental research focuses primarily on analyzing Structured Threat Information Expression (STIX) containing various indicators of compromise (IOC). Majority of IOCs files are in XML format. In recent years, JSON IOC files have become more popular because of advantages, e.g., JSON has a more compact style than XML, it is often more readable. Our subject will be an example of threat information from a malware referred to as ARTFUL-PIE.<sup>4</sup> This example threat information expression is from a Trojan created by the North Korean government under the name of HIDDEN COBRA.<sup>4</sup> We will be using three primary tools to carry out this research. These tools are StixViz, Stix Elevator, and STIG. In this experiment, we followed an order of steps that can be repeated on other threat information examples analogous to the same resources and tools. The research was conducted in a virtual machine environment, where the tools were downloaded and run.

The process goes as follows, acquire the appropriate threat information IOC files from a trustworthy source or website (usually in .xml format). After acquiring this information, import the xml file into the Stix-Viz. With this tool, we can begin the process of analyzing the threat



information visually. The information will be displayed in a visual format of objects, and relationships. Objects have unique images displayed as icons, and relationships are arrows that connect the different objects with one and other, creating a visual representation of the chosen threat information expression. Stix-Viz works only for XML IOC files and also, the relationships cannot be edited if needed. Therefore, we will convert the .xml file into a format that is quicker, more compact, and conventionally easier to read and comprehend. This format will be a .json file. This conversion step is essential to progress within this experiment since the tools only read and write in the formatted .json file type.

For this experiment, we will be using an open-source tool to efficiently convert the xml file to a json file. This tool is the Stix elevator. After the file is converted appropriately, we will then import the file into the Linux virtual machine where we can run the program STIG. STIG is a visualizer that reads the json file type. Additionally, the program allows users to manually create the relationships. One of the major issues we encountered is that, during the conversion step, relationships can go missing, leaving an incomplete threat expression. Nevertheless, this can be resolved through visually analyzing the json file, and then comparing the threat information with the image created using the original .xml file before it's conversion. After recreating these relationships, the threat information is now completed and converted into a versatile format.

### III. Methods

#### A. Obtaining Threat Information

The first step to this process is obtaining the threat information expression from a reputable source. For this research, the threat model was provided. The subject for this research involved the threat information on the ARTFULPIE malware. Threat information expressions can be

found on government websites that collect similar categories of information to share amongst a vast community. A reputable source to find a subject analogous to the one used in this research involves the website <https://us-cert.cisa.gov/>. This website is controlled by the United States' Cyber and Infrastructure Security Agency (CISA) under the Department of Homeland Security (DHS).<sup>1</sup> The website is frequently updated with emerging cyber related threats and updates. After visiting this website, you can search and select a document that discusses a chosen malware attack. Malware examples are denoted as MAR in the beginning of its file name, following a series of numbers. After selecting the appropriate document, the report will contain a downloadable link which ends with the file type .stix. After clicking the link, you will download and save the page or file as a .xml document. The threat information expression of the malware chosen is now suitable to be uploaded and run using the Stix Visualization tool.



Figure 1. This image displays the file type displayed on the us-cet.cisa.gov website

## B. Using the Stix-Viz

After obtaining the appropriate threat information example, the xml file is then uploaded into the pre-installed program. This program was installed through an open-source GitHub repository

called Stix Project. The Stix-Viz is a program developed to visualize Structure Threat Information eXpressions (STIX) documentation in the format of a node link-tree with its root at the top of the structure.<sup>6</sup> The Stix viewer only reads files in the xml format. The program provides three methods of visualizing the document. This includes graph view, timeline view, and tree view. For this research, the two methods we used were graph and tree view. Both views display the threat information into node object types represented as square icons and with relationships connecting the objects represented as arrows stemming from one object to the next.

Through this visual representation, the threat information expression can now be analyzed by moving the icons around, identifying which relationships they share amongst one another, along with their corresponding relationship types. The program breaks up code into various categories that allow the code to be read quicker. The program does not allow for the user to manually edit the relationships and objects within the program. Nevertheless, by viewing the information in a graph-like view, it is easier to understand the functionalities within the malware threat

information sample. The .xml file type has drawbacks which make it less versatile regarding its ease of use. In the .xml format, the code is significantly lengthier and complex.

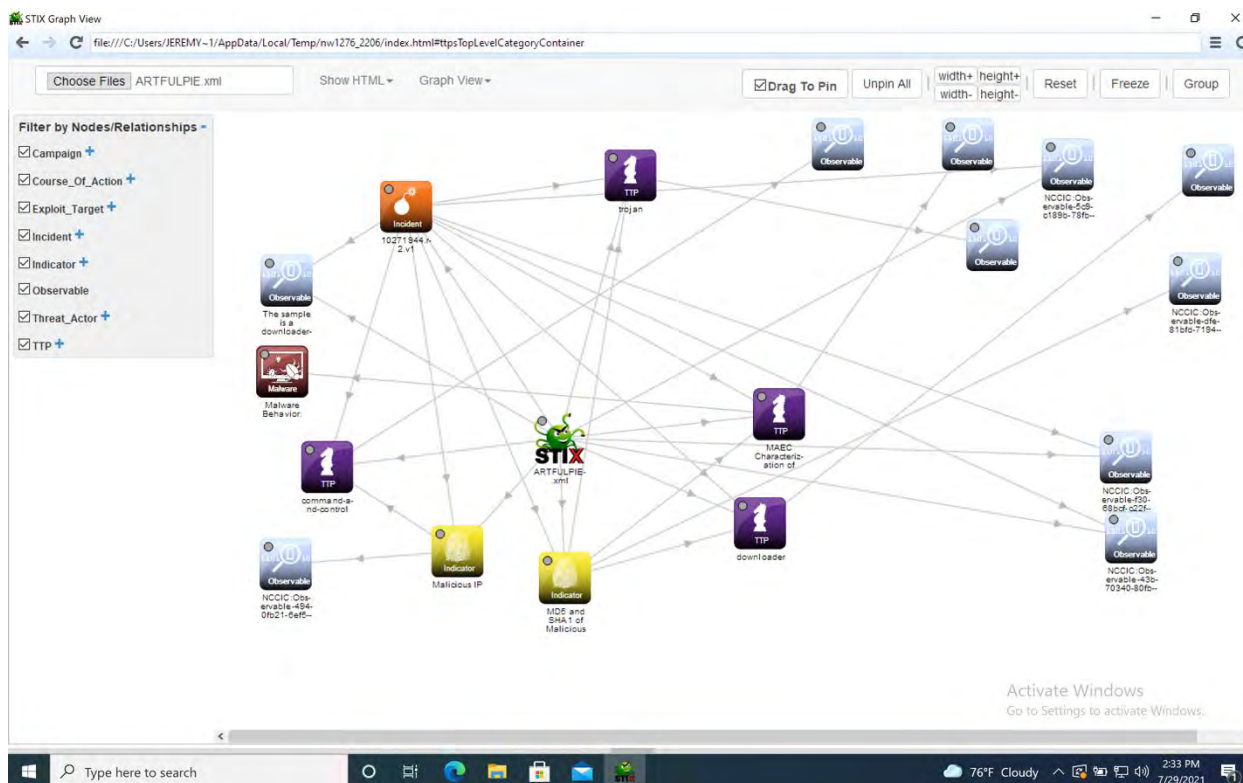


Figure 2. This image displays the ARTFULPIE.xml in the tool Stix-Viz

This research focuses on analyzing structured threat information in a versatile format. After carefully documenting and analyzing the information of the ARTFULPIE.xml malware example, the file will then be converted into a language better fit for the research, which adds more capability, is easier to understand its code, and faster to analyze. This format will be in the .json document type. A .xml file is an acronym for Extensible Markup Language which uses code and tags to describe a document. A .json file is short for JavaScript Object Notation, which is a lightweight and data-interchange format. It's easier for humans to read and write, and easier for machines to parse and generate.<sup>3</sup> This makes .json files more versatile for analyzing structured threat information expressions.

### C. Converting with the Stix Elevator

To successfully convert the .xml file into the desired .json format, we use another open-source tool called Stix Elevator. The Stix Elevator converts the STIX 1.x to STIX 2.0 or 2.1. The program was developed by Oasis-Open and can be installed from their public GitHub page. The program functions within the python programming language.<sup>5</sup> Furthermore, to convert the file we would run a python environment's terminal and use the commands stix2-elevator.py (file name)> outputtest.json to create a json file conversion of the ARTFULPIE.xml file we are using in this example. Once the file is converted, it can be observed in a text editor. There you can observe the changes made in the coding language. We will now need to use the text editor to manually remove comments and warnings created by the Stix elevator. This step is important because the next program will not run the new .json file type without the correct edit. After removing the undesired text, import the file into the chosen virtual machine. For this research, we used VMware to create a Linux (Ubuntu) virtual operating system capable of running the STIG program, which is another visualization tool used for its different functionalities.

```
usage: stix2_elevator [-h]
                    [--missing-policy {use-custom-properties,add-to-descriptio
n,ignore,use-extensions}]
                    [--custom-property-prefix CUSTOM_PROPERTY_PREFIX]
                    [--infrastructure] [--incidents]
                    [--package-created-by-id PACKAGE_CREATED_BY_ID]
                    [--default-timestamp DEFAULT_TIMESTAMP]
                    [--validator-args VALIDATOR_ARGS] [-e ENABLED]
                    [-d DISABLED] [-s]
                    [--message-log-directory MESSAGE_LOG_DIRECTORY]
                    [--log-level {DEBUG,INFO,WARN,ERROR,CRITICAL}]
                    [-m MARKINGS_ALLOWED] [-p {no_policy,strict_policy}]
                    [-v {2.0,2.1}]
                    file
stix2_elevator: error: unrecognized arguments: outputtest.json
(base) Jeremys-Air:~ jeremyanderson$ stix2_elevator BNLTest1.xml > outputtest.js
on
[stix2elevator.options] [214] [WARNING] [2021-06-17 16:24:03,794] incidents opti
on was not given, but it defaults to true for version 2.1
[stix2elevator.options] [214] [WARNING] [2021-06-17 16:24:03,795] infrastru
cture option was not given, but it defaults to true for version 2.1
[stix2elevator] [ERROR ] [2021-06-17 16:24:03,796] Error occurred: The file 'BN
LTest1.xml' was not found.
```

Figure 3. This image displays the Stix Elevator in python terminal.

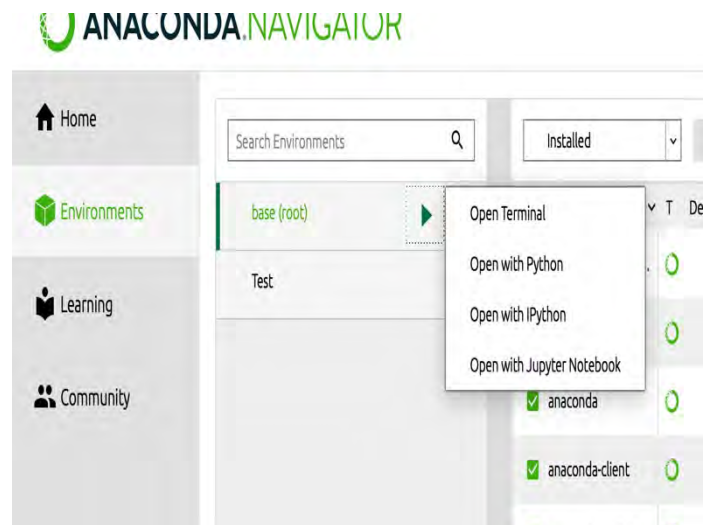


Figure 4. This image shows the python tool used.

## D. Using STIG

Once the .json file created by the ARTFULPIE.xml threat information is uploaded into the virtual machine, we then run and use the program STIG. The tool STIG is short for Structured Threat Intelligence Graph and was created by Idaho National Lab. The program can be installed from their GitHub page. This tool can create, edit, query, analyze and visualize threat intelligence.<sup>2</sup> To upload the program into STIG, we drag and drop the json file into the program. Nevertheless, the program will then generate a visual representation of the ARTFULPIE .json file. The visual representation should be similar to its .xml format previously displayed in the Stix-Viz. An issue that does arise within this research was missing relationships that were once connected in the file's original .xml version. After converting the file, multiple relationships connecting different objects are no longer present. This has been an open issue associated with STIX Elevator and remain unsolved so far.

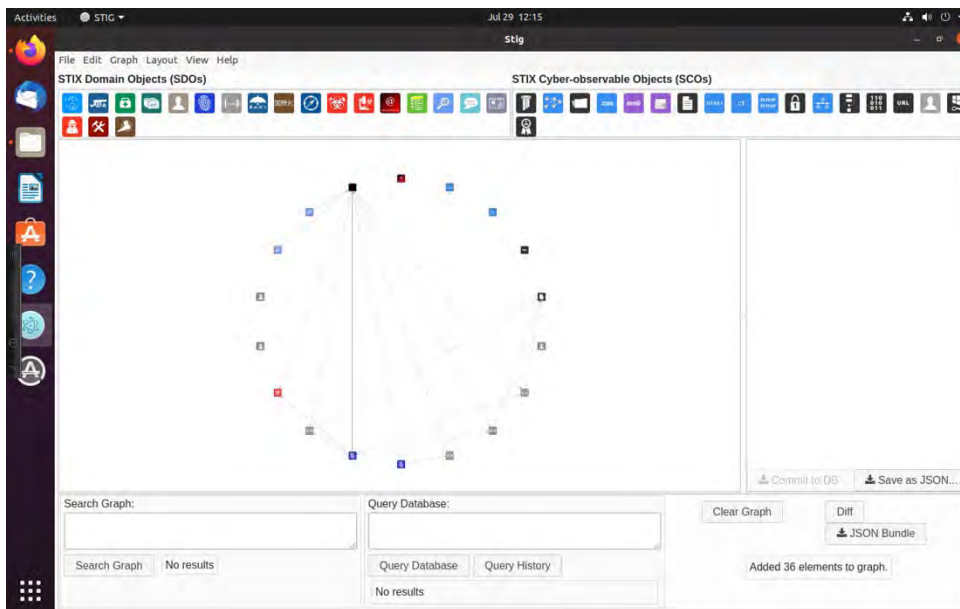


Figure 5. This image shows ARTFULPIE in the STIG tool.

## E. Missing Relationships

Since the subject is in its .json format, it is easier to find where the missing relationships occur. This can be identified in the code. STIG visually displays these missing relationships making the process less difficult to identify. Additionally, STIG has functionalities capable of correcting the missing relationships manually taking less time. To create a new relationship, the user clicks an object and drags the cursor to the object they wish to create a relationship with. They will then choose the relationship type in the information panel for the relationship. This process reduces the time it would take to carry out the same task by writing the code.

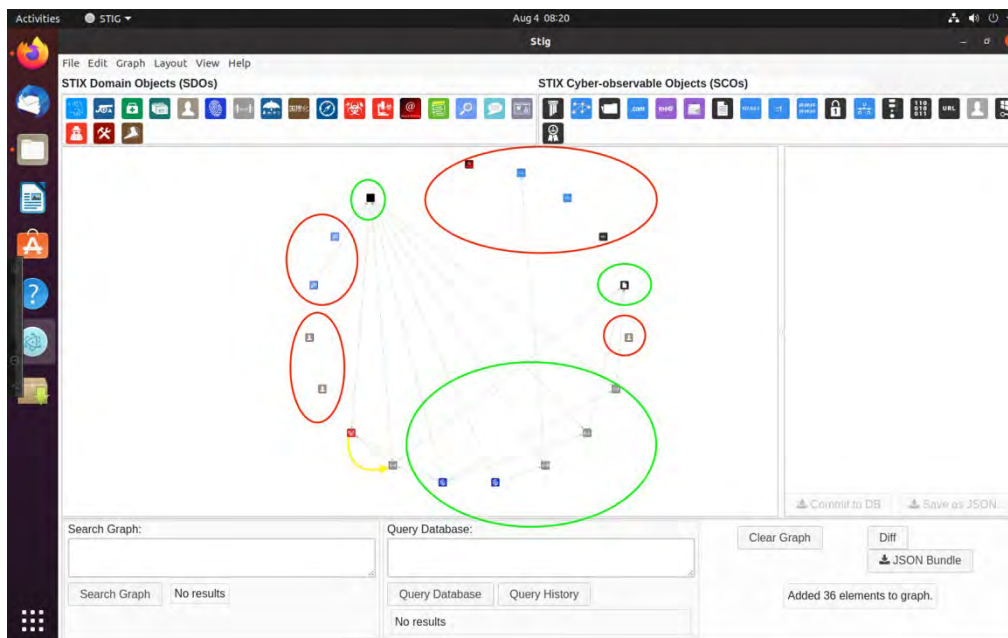


Figure 6. This image shows the missing relationships in the STIG tool. The red ovals show object with missing relationships, and the green shows objects with their correct relationships. The yellow arrow show relationship manually added.

Nevertheless, STIG corrects the problem with missing relationships. There are a multitude of reasons why relationships go missing during conversion. A reason why relationships can become missing is the Stix elevator's inability to correctly create the appropriate identity for objects because the program lacks source code to identify the specific object type. Since the field is frequently growing, new updates are always being made to improve its capabilities.

## F. Manually Reconstructing the Missing Relationships



To correctly reconstruct the missing relationships in STIG, we must first identify the object and relationship types within both tools' images. This process is done because Stix-Viz is the original and unaltered file of our ARTFULPIE.xml threat information. The .xml displays all the correct relationships before conversion. By comparing both visual representations, we use the .xml image to recreate the .json equivalent image in STIG. To match the correct objects, we look at the names and hashes of each object in Stix-Viz that correspond with the names in STIG. Then, we create the appropriate relationships that were no longer there by simply connecting arrows. Both programs use different naming conventions for their relationship and object types. To recreate the relationships correctly, I would first document all the relationships and create a chart which identifies each object and relationship type between both tools.

## IV. Results

After reconstructing the appropriate relationships and nodes using both tools, we can see that we were successfully able to visually analyze the threat information obtained and convert the .xml ARTFULPIE example into the .json file which is a more desirable format for analyzing threat information.

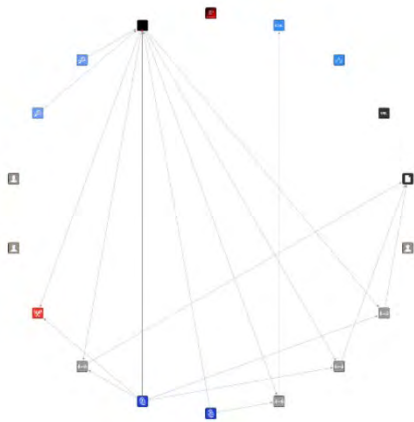


Figure 7. This image shows the ARTFULPIE example before adding the relationships

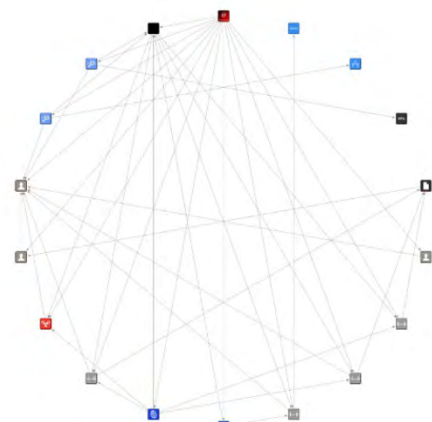


Figure 8. This image shows the ARTFULPIE example after adding the relationships



## V. Conclusion

Now that the file is in a versatile format, the structured threat information expression can be used to help others visually understand the threat information and create methods of defending and protecting against new and emerging threats. This information can be used in tools to create a database of threats that are capable of being recognized just by their visual structured threat information, disregarding the need to extensively analyze lengthy code. By improving this process, threat information can be shared faster, and individuals will be able to analyze the similarities and differences between malwares.

## VI. Acknowledgements

I would like to thank my mentor Meng Yue from the Interdisciplinary Department at Brookhaven National Laboratory (BNL) for choosing me to join this research, and Joshua Steir an intern from Stony Brook University who assisted me in this research. I would like to thank the Brookhaven National Laboratory for the (SULI) program and hosting this internship. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## VII. References

<sup>1</sup>*Homepage | CISA*. (n.d.). US-Cert.Cisa. Retrieved August 11, 2021, from <https://us-cert.cisa.gov/>

<sup>2</sup>I. (n.d.-a). *GitHub - idaholab/STIG: Structured Threat Intelligence Graph*. GitHub. Retrieved August 11, 2021, from <https://github.com/idaholab/STIG>

<sup>3</sup>*JSON*. (n.d.). JSON. Retrieved August 11, 2021, from <https://www.json.org/json-en.html>

<sup>4</sup>*MAR-10271944-2.v1 – North Korean Trojan: ARTFULPIE | CISA.* (n.d.). US-Cert.Cisa.

Retrieved August 11, 2021, from <https://us-cert.cisa.gov/ncas/analysis-reports/ar20-045e>

<sup>5</sup>O. (n.d.-b). *GitHub - oasis-open/cti-stix-elevator: OASIS Cyber Threat Intelligence (CTI) TC*

*Open Repository: Convert STIX 1.2 XML to STIX 2.x JSON.* GitHub. Retrieved August

11, 2021, from <https://github.com/oasis-open/cti-stix-elevator>

<sup>6</sup>S. (n.d.-c). *GitHub - STIXProject/stix-viz: STIX Visualization Tool.* GitHub. Retrieved August

11, 2021, from <https://github.com/STIXProject/stix-viz>

In situ spectroscopy of metal oxide catalysis

Lukas Baker, Department of Physics and Astronomy, Union College, Schenectady, NY 12308

Burcu Karagoz, Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton,  
NY 11973

Ashley Head, Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton,  
NY 11973

## **Abstract.**

Metal oxide catalysts are used in a wide variety of applications including gas masks to protect individuals from chemical agents. To better understand these catalysts chemical properties, and possibly find other catalysts that work better, analytic techniques such as Atmospheric Pressure X-Ray Photoelectron Spectroscopy (APXPS) and Infrared Reflection Absorption Spectroscopy (IRRAS) can be performed at Brookhaven National Lab's Center for Functional Nanomaterials to probe single atomic layers of these catalysts to better understand their surface chemistry. To gain useful information from these methods of spectroscopy, data analysis tools must be used as several hundred spectra can be generated in the relatively short period of time of a few hours. Thus, it was necessary to use software such as IGOR pro 8 to write functions to automate much of the data analysis so it can be performed quickly and efficiently. In this paper, we discuss two sets of functions written, one for APXPS data analysis and the other for IRRAS data analysis. These functions produce high-quality plots which allow for visualization of chemical processes occurring on a nanometer scale and are a tool that will be useful for future APXPS and IRRAS data analysis projects as well.

## **Introduction.**

Metal oxide catalysts are used in a wide variety of applications from gas masks to synthesis of petrochemicals. Specifically, chemical processes that occur on the surfaces metal oxide catalysts are of interest. While these catalysts are used in a variety of industries, better characterization of their chemical properties may uncover new possible applications. To study these catalysts in situ, two analytic techniques are widely used: Atmospheric Pressure X-Ray Photoelectron Spectroscopy (APXPS) and Infrared Reflection Absorption Spectroscopy (IRRAS). APXPS can be used in “snapshot” mode, which provides time resolved data over long periods of time by capturing spectra at regular intervals, for example spectra may be taken every 10 seconds for an hour. This method of data collection may generate hundreds of spectra which are impractical to analyze by hand. Specifically, the generation of two-dimensional color plots which show the change in spectra over time is quite tedious to perform by hand. Therefore, functions in IGOR pro 8 were written to represent this data in a useful way. These purposes of these functions include representing data in 2d image and 3d surface plots, signal averaging, integration, and displaying specific spectra from image plots. These functions greatly increase the capacity for data analysis of metal oxide catalysts using APXPS. IRAS is often performed at a variety of pressures. Often, reference spectra are taken at ultra-high vacuum (UHV) and then experimental spectra are taken at increasing pressures up to the millibar regime. To get chemical information from these spectra, it is often necessary to divide one spectrum by another or perform other mathematical operations between spectra to remove contributions from the gas phase and from the catalyst surface. To perform these operations quickly and efficiently, a spectrum calculator was developed.

## **Igor Functions.**

A series of functions were written to facilitate data processing, some of which were integrated into a window user interface. Below is a description of the tasks these functions perform.

## **APXPS Functions.**

### **Interpolating the Data.**

While Igor does have functions to load specific types of data, the data files produced by the APXPS software commonly used at Brookhaven's Center for Functional Nanomaterials are not supported in the prebuilt Igor Pro data loading functions. As a result, code had to be written in Igor to extract the data from the .xy files produced in the Prodigy XPS data collection software.

While this function had been written before summer 2020, there was an issue with the function specifically in loading the snapshot APXPS data. Specifically, the snapshot data loaded into two waves (Igor's version of an array), as opposed to a single array. Igor stores waves differently than many other languages. While python for example, would generally store arrays with N coordinate pairs as a list of N tuples, or coordinate pairs, Igor only stores the data to be plotted on the vertical axis. The horizontal axis data is stored in a "start" and a "delta". From these values, the x values are calculated on the spot each time the data is plotted. As a result of this, Igor implicitly assumes that the data being imported has a constant x axis interval.

One of the first problems encountered in loading the APXPS snapshot data into IGOR was circumnavigating this issue. Using the loading function written by colleagues, it was possible to load the APXPS .xy files into IGOR, but snapshot data was loaded in pairs of waves,

and it was unclear as to why this was the case. After loading the data into excel and manually combing through it, it was noted that the delta, or interval between points was not constant, which lead the function to classify the data differently and load the X axis data and Y axis data in as two separate waves. To make the data usable in functions however, it was much more convenient to have all the data in a single wave. This was achieved using Interpolation. A function was written which would take as input the X-axis (Binding Energy) and Y axis (intensity) waves, and interpolate values using a linear interpolation. This was then output as a single usable wave.

### **Making Two-Dimensional Color Plots.**

Two-dimensional color plots are simple ways to visualize large quantities of data. One use case of these types of plots, is displaying time resolved data. To display time resolved data using a two-dimensional plot, the spectra are plotted in two dimensions with a horizontal axis displaying the x axis (Binding Energy) data of a single wave and a vertical axis displaying time. The y-axis (intensity) data, in this case the counts or intensity, is displayed as a color at each point in the spectrum.

To generate these plots, Igor takes as input a 2-dimensional wave, or matrix containing all the waves that need to be plotted. The horizontal axis data is stored again in a start and delta value, meaning that all the waves that compose the image plot must have the same interval. This data structure, however, is not an issue as this is ensured to be the case in the interpolation function. The matrix is created simply by looping through the raw data waves, interpolating them, compiling the interpolated waves into a matrix, and then plotting that matrix. The column number of the matrix corresponds to the time at which the spectrum was taken. The row of the

matrix corresponds to the intensity of the output measurement, and the index of the matrix row corresponds to the energy. Which is calculated using the index, start, and delta values.

Once the two-dimensional (2D) color plots were made, making the 3-dimensional (3D) surface plots is quite simple as they also use the same matrix structure as the 2D plots. As a result, it was possible to add a single line of code to generate the 3D surface plots in a separate function. That said, it is also possible to generate these plots directly by making the 2D plot first and then converting it to a 3D plot.

### **Displaying Spectra with Cursors.**

Sometimes it is useful to see individual spectra at specific locations on the color plot. To do this, a function was written to take as input the position of cursors which can be placed on the graph and output the spectrum at their position.

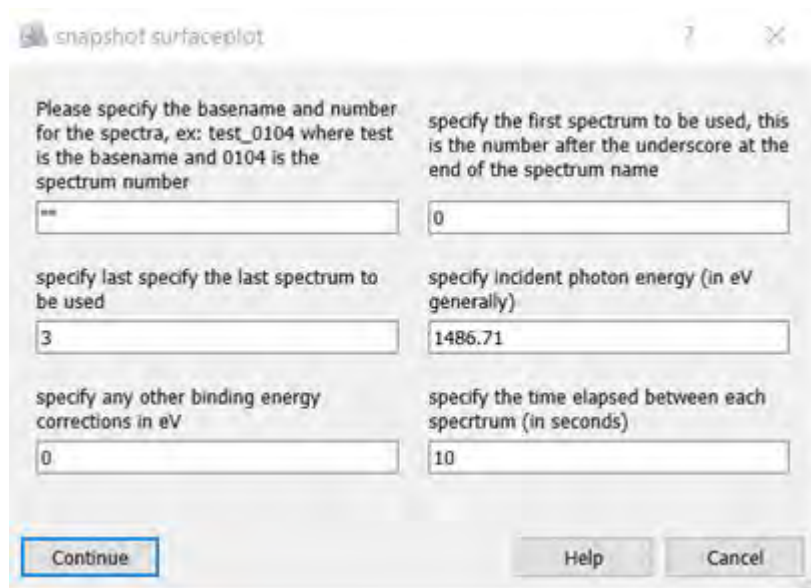
Igor has built in cursors which can be placed on a graph to mark a location. Using these cursors, a function was created which takes their position as an input, and outputs the spectrum selected by the cursor. To do this, first the function had to access the location of the cursor. Once the cursor had been identified, the vertical axis position corresponding to the time that the spectrum was taken in seconds is calculated. Then, the index of the spectrum can be calculated. Since the index is related to the spectrum number, the corresponding spectrum can be displayed.

### **Making A Spectra Display Window.**

Initially as a user interface, a dropdown menu was created where the user could input the base name (a global variable created when loading in the data), the start and end of the spectra to be plotted, the incident photon energy, the time elapsed between each measurement, and any

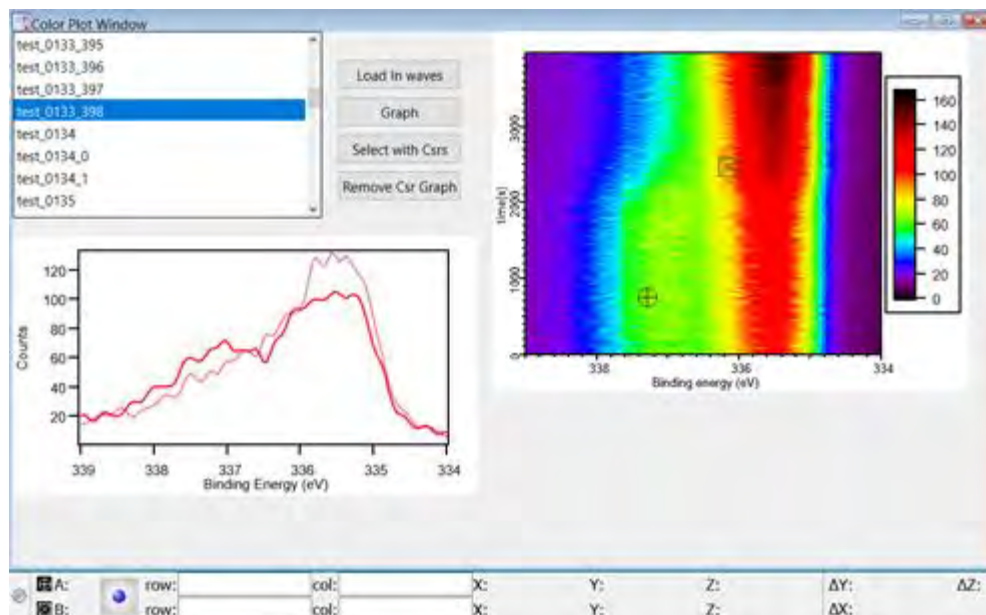


correction terms. This is shown in Figure 1. Upon pressing the “continue” button, the plot would be generated. This, however, was somewhat clunky and would be much nicer in a user interface.



**Figure 1.** Initial user interface for the 2D color plot graphing function.

To make an improved user interface, a window was created which can be accessed from the Igor taskbar. This window, as seen in figure 2, compiles all the previously written functionalities into a single window which allows a user to load in all of the spectra of interest, display them in a 2D graph, and use cursors to select the graph of interest and display it.



**Figure 2.** Window interface displaying a sample 2D color plot in the top right corner, the graph in the bottom left corner displays the two spectra selected with cursors from the image plot.

### **Integration.**

An integration function was written to integrate a curve given a start x axis position, end x axis position, and spectrum name. This integration can be done for not only a single spectrum, but an entire grouping of spectra. When the integration is carried through an entire group of spectra, the area under each curve is plotted as a function of the time at which the individual spectra were taken.

### **Spectrum Averaging.**

A simple spectrum averaging function was also written. The function takes as input a start and end spectrum, and outputs a wave and plot containing the average of all the waves.

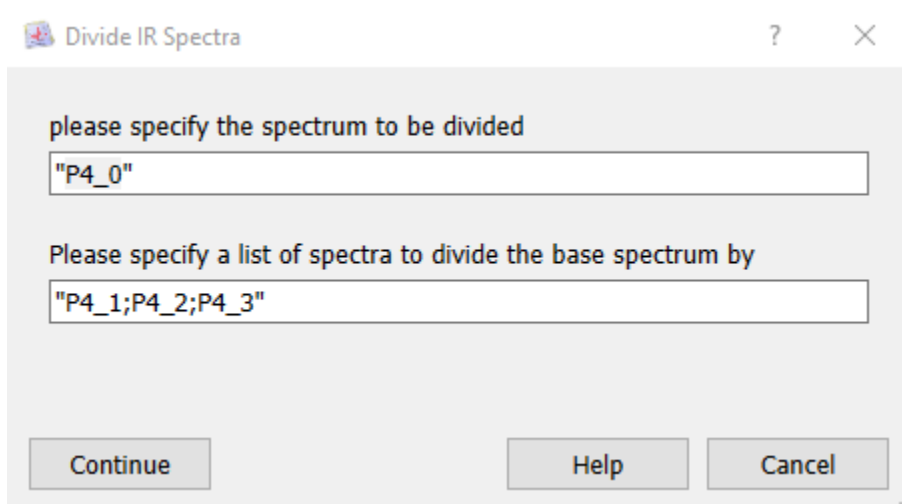
## IRRAS Igor Functions.

### Loading Folders of Data.

Similarly, to the APXPS data, the IRRAS data was also contained in formats that are not easily loaded into Igor. While the files are supported, the IRRAS data is created, each spectrum is loaded into a separate file in a folder. While Igor can load individual files, it can not load entire folders easily. As a result, it was necessary to loop through all the files the folder of interest and load them individually into waves.

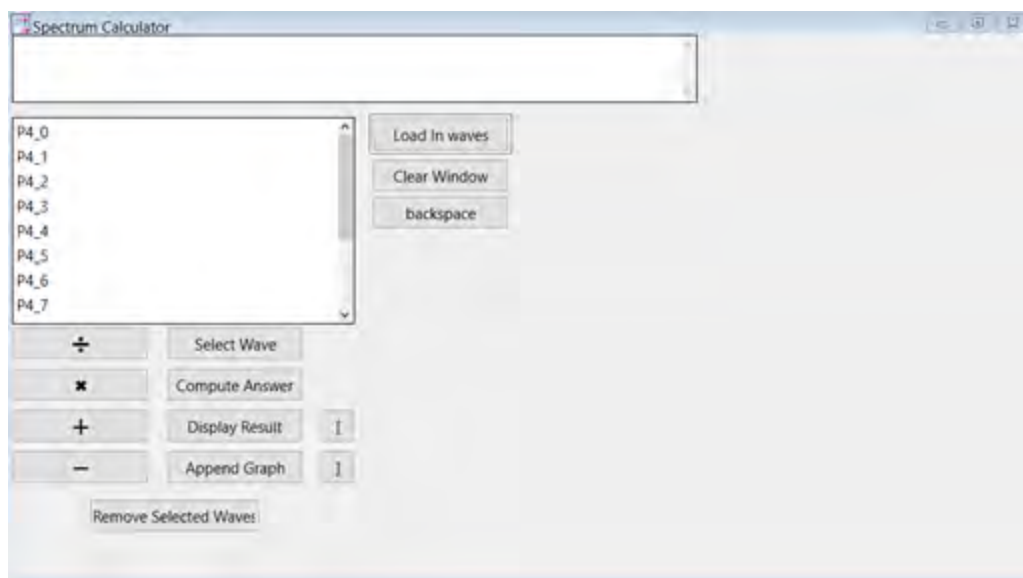
### Operations on Waves.

IRRAS spectra can be taken at various pressures and conditions. The user can then add, subtract, multiply, or divide spectra by one other to gain various types of information. As a result, four functions were written to add, subtract, multiply and divide spectra. Again, similarly to the APXPS functions this was initially constructed in a dropdown menu, and later refined to a window. The initial dropdown menu is shown below in figure 3.



**Figure 3.** IRRAS initial dropdown sample menu for division.

To make this more user friendly, a calculator window was constructed, which allows the user to select spectra to operate on, just like a normal four function calculator. This is depicted in figure 4.



**Figure 4.** IRRAS spectrum calculator window.

## **Experimental Methods.**

The scripts were used to plot and analyze a test data set that has been collected previously using the conditions outlined below.

### **APXPS Data Collection.**

APXPS spectra of PdO were captured in snapshot mode and fixed analyzer transmission mode (FAT) as a calibration for the apparatus. Snapshot mode spectra were taken every 10 seconds during both the oxidation and reduction of PdO. The oxidation was carried out at 500 C, and the reduction was carried out at 400 C in the presence of CO at  $10^{-5}$  mbar.

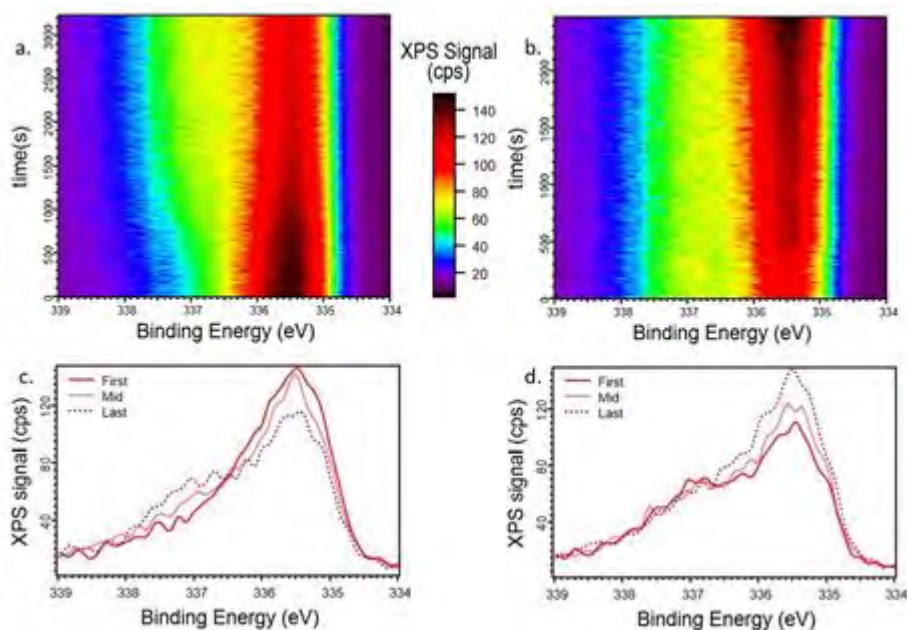
## IR Data Collection.

IRRAS spectra of CO adsorption on a clean Cu<sub>2</sub>O (111) crystal were taken at various pressures. Spectra were taken at UHV, 10<sup>-3</sup> mbar, 10<sup>-2</sup> mbar, 10<sup>-1</sup> mbar and 1 mbar.

## Results and Discussion.

### APXPS of Pd and PdO.

The APXPS snapshot data of the oxidation of Pd and reduction of PdO produced the expected behavior, although further data collection may be beneficial. The reduction of PdO was not complete, as it was stopped once it was clear reduction was occurring.



**Figure 5.** (a) 2d image plot of APXPS data showing the oxidation of Pd to PdO. (b) 2d image plot of APXPS data showing the partial reduction of PdO to Pd. (c) First, middle and last spectra in the above color plot showing the spectra of Pd, partially oxidized Pd, and fully oxidized PdO respectively. (d) First middle and last spectra in the above color plot showing PdO, PdO at an earlier stage of reduction, and PdO at the latest stage of reduction reached respectively.

As shown in Figure 5c, during the oxidation the first peak is relatively narrow and centered around 335.5 eV. This peak corresponds to the Pd metal bulk material.<sup>1</sup> Since it has not been oxidized yet, the shoulder corresponding to the PdO is small. The mid peak exhibits a shoulder around 336.7 eV, and these trends continue in the last spectrum, which has even further reduced Pd peak size and increased PdO shoulder size. It is noteworthy that while some other studies may show a peak corresponding to the oxide instead of a shoulder, this is due to the probing depth of the APXPS used. The apparatus used in this experiment probed 2-4 nm into the sample, meaning that even after a lot of oxidation had taken place at the surface, there was still significant amounts of bulk material to show a Pd metal signal. It is also noteworthy that the oxidation is a gradual process happening over a course of about 1500 s, and then leveling off. This can be seen in Figure 5a, where after 1500 s, the spectra are relatively constant, and no major changes occur.

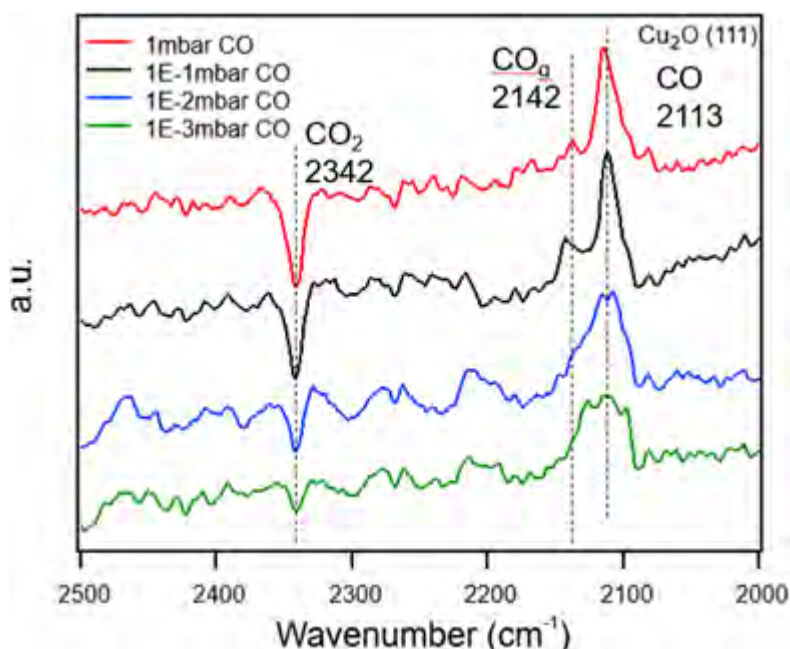
The reduction exhibits opposite behavior as the metal oxide is reduced over time. For the reduction, the first peak (figure 5d) exhibits a strong shoulder corresponding to a significant amount of PdO present in the sample while also showing a relatively weak peak corresponding to Pd. As reduction continues, PdO is reduced to Pd which is shown in the spectra as the Pd metal peak grows, and the shoulder decreases slightly in size. Since reduction was not carried out until completion, the PdO shoulder does not completely disappear. Again, like the oxidation, the reduction is a gradual process occurring over a period of around 2000 s.

## IRRAS of Cu<sub>2</sub>O.

IRRAS spectra taken at various pressures were divided by one another using the following spectrum division expression where “I” is intensity, the light polarizations are indicated in superscripts, and pressures are indicated in subscripts:

$$\frac{\frac{I_{@pressure}^P}{I_{@pressure}^S}}{\frac{I_{UHV}^P}{I_{UHV}^S}}$$

Once this division is carried out, it is possible to compare the resultant spectra. Four spectra taken at various pressures are depicted below in figure 6.



**Figure 6.** IRRAS spectra taken of Cu<sub>2</sub>O at various pressures.

There are multiple noteworthy features of these spectra. The Peaks at 2113 and 2142 cm<sup>-1</sup> correspond to adsorbed and gas-phase CO respectively.<sup>2</sup> An interesting feature of the spectra, however, is the negative CO<sub>2</sub> peak that arises at 2342 cm<sup>-1</sup>. Currently, it is speculated that this

CO<sub>2</sub> arises because of reactions of CO catalyzed by the Cu<sub>2</sub>O. To test this, future experiments will be conducted using mass-labelled CO to determine the origin of the CO<sub>2</sub>. The other interesting feature of this peak is that it is negative, whereas the carbon monoxide peaks are positive. The reason for this is likely due to the nature of the interaction between the P-polarized light and the bonds in CO as opposed to CO<sub>2</sub>. In CO, adsorption is normal to the plane of the metal and therefore, the perpendicular (z) component of the P polarized wave is absorbed. This corresponds to a positive signal.<sup>3</sup> The CO<sub>2</sub> signal on the other hand is adsorbed parallel to the plane of the metal and therefore the parallel (x) component of the P polarized wave is absorbed. This corresponds to a negative signal.<sup>3</sup> For those familiar with IRRAS, this may seem odd, because metals do not exhibit this complicated interaction with the P-polarized beam. This is because metal oxides behave differently to traditional metals, as they are dielectric materials which therefore interact differently with electric fields.<sup>3</sup>

## **Conclusion.**

Functions in IGOR Pro 8 were written to process large quantities of APXPS snapshot and IRRAS data. Using these functions, APXPS data of Pd oxidation and PdO reduction was analyzed to monitor the conditions in situ. This data was used to check whether the reaction was reaching the desired temperature and pressure to ensure future reactions are actually at the desired conditions. IRRAS spectra were taken of CO adsorption to Cu<sub>2</sub>O (111) at a range of pressures. The various spectra all exhibited the expected peaks for CO as well as a peak for CO<sub>2</sub>, which may arise due to catalytic activity of the Cu<sub>2</sub>O. To test whether this is the case, further experimentation will be done using isotopically labeled CO.



## **Acknowledgements.**

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). I would also like to thank Ashley Head and Burcu Karagoz for their continued support throughout this internship, as well as BNL and DOE for the opportunity.

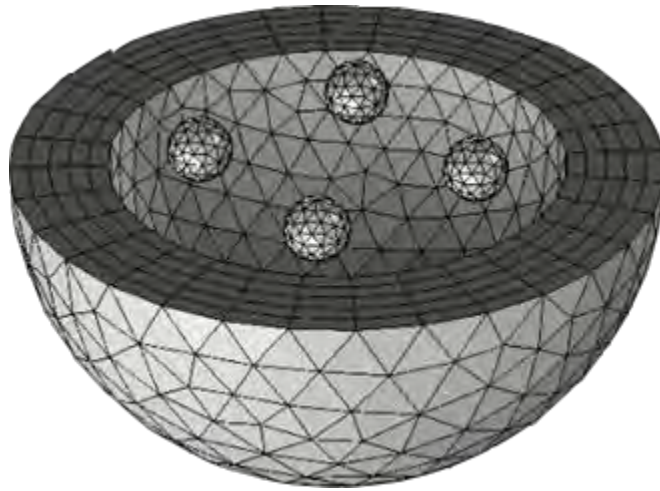
## **References.**

- [1]: Toyoshima, R., Yoshida, M., Monya, Y., Kousa, Y., Suzuki, K., Abe, H., Mun, B. S., Mase, K., Amemiya, K., & Kondoh, H. (2012). In situ ambient Pressure XPS study of CO oxidation reaction on Pd(111) Surfaces. *The Journal of Physical Chemistry C*, 116(35), 18691–18697. <https://doi.org/10.1021/jp301636u>
- [2]: Boyd, D. A., Hess, F. M., & Hess, G. B. (2002). Infrared absorption study of physisorbed carbon monoxide on graphite. *Surface Science*, 519(1-2), 125–138. [https://doi.org/10.1016/s0039-6028\(02\)02207-0](https://doi.org/10.1016/s0039-6028(02)02207-0)
- [3]: Wang, Y., & Wöll, C. (2017). Ir spectroscopic investigations of chemical and photochemical reactions on metal oxides: Bridging the materials gap. *Chemical Society Reviews*, 46(7), 1875–1932. <https://doi.org/10.1039/c6cs00914j>

Comparison of scattering and extinction spectra by analytical and finite element analysis methods

Kevin Beckford, Farmingdale State college, Farmingdale, NY 11735

Noel Blackburn, Office of Educational Programs, Brookhaven National Laboratory, Upton, NY 11973



## Abstract

It is conceivable that a material can be constructed whose permittivity and permeability values may be designed to vary independently and arbitrarily throughout a material, taking positive or negative values as desired. This opens a domain of activity which enables controlling a variety of electromagnetic applications. This can be accomplished with transformation optics and its core idea is the correspondence between coordinate transformation and material implementations. The design of a material can constitute a time-consuming process when following a trial-and-error procedure. An alternative is to let a neural network perform the decision-making process. For setting up a neural network one must train it for learning the right decisions. One of the inputs is a method that enables the neural network to calculate the way electromagnetic waves are traversing an arbitrary arrangement of scatterers in a medium. Such methods can be provided by finite element methods (FEM) calculations and an analytical computational technique of light scattering with the so-called T-Matrix method. The method used in this project is the T-matrix method.

## Background

To analytically calculate electromagnetic scattering off multiple nanoparticles we needed a code that implemented the T-matrix method. The code that we discovered was the Multiple Sphere T-matrix (MSTM) FORTRAN 90 code. This code allows for the calculation of time-harmonic electromagnetic scattering properties of a group of spheres. This program has many uses such as modelling both plane wave and Gaussian profile incident beam, generating maps of the electric field distributions along any plane, etc. However, the functionality that is needed for this project is the calculation of cross sections and far-field scattering matrix elements for both fixed and random orientation. What we're looking for is the scattering and extinction cross section values. Now, it will be important to discuss the equations that are used in the code so that we may arrive at our results. This first equation is the extinction cross section, and it represents the rate at which the sphere removes energy from the incident light.

$$C_{ext} = \sum_{i \in \mathcal{N}_{int}^0} C_{ext,i}$$

The second equation is the independent scattering cross section. The usefulness of this sphere is that it provides in combination with the absorption and extinction cross section, a relative measure of the scattered field intensity.

$$C_{i-sca,i} = \frac{2\pi}{k^2} \sum_{n=1}^{L_i} \sum_{m=-n}^n \sum_{s=1}^2 |\tilde{a}_{mns}^i|^2$$

The third equation describes the energy balance at a given sphere surface in which it shows the dependent cross section of the sphere.

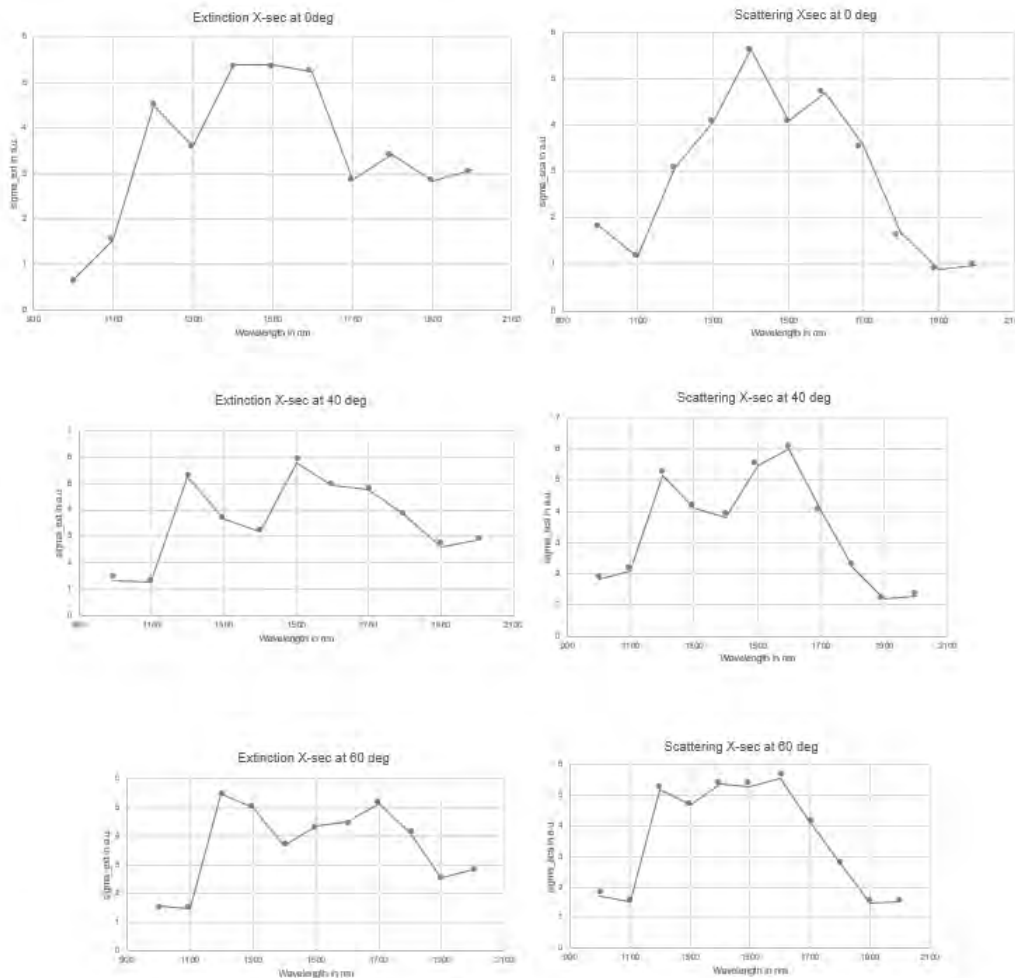
$$C_{ext,i} - C_{i-sca,i} - C_{abs,i} = C_{d-sca,i}$$

The last equation that will be discussed in this paper shows how the T-matrix method is calculated.

$$a_{mnp}^0 = \sum_{l=1}^{L_0} \sum_{k=-l}^l \sum_{q=1}^2 T_{mnp\,klq} f_{klq}^0$$

## Results

In this project we were working with a fixed four-sphere nanoparticles that were with an incident light of different wavelengths. We set our individual nanoparticle radii to 230 nanometers and manually set the wavelength by using the length scale factor. The length scale factor is defined as  $\frac{2\pi}{\text{Wavelength}}$ . The wavelength that we were working with ranged from 1000nm-2000nm. Additionally, we set incident light to scatter off the nanoparticle at different polar angle. This resulted in different scattering and extinction cross section. I gathered the values of all four spheres than I plotted them on excel so that I can observe the results. In this case I will only show the results for one out of the four spheres because I didn't find it necessary to show all. The graphs demonstrate that at different incident polar angles both the scattering and extinction values will be drastically different throughout the range of angles we used. The graphs are shown below.



## Conclusion

In conclusion, These T-matrix results are preliminary and need to be cross-checked with the COMSOL model. This model is done by my colleague Othman Alrawi. The COMSOL method is an FEM which is a more accurate way of calculating the scattering cross section and extinction but takes some time. Additionally, more complex geometrical orientation should be explored like multi-layered sphere, random oriented multi-sphere etc. Finally, more complex shapes should be explored as well so that we can see if this method is valid for those situations. Both the T-matrix method and the FEM calculations could potentially be an excellent way to cut time on designing a material that can be used in the real-world.

## References

- [1] C. F. Bohren, D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, Wiley, 1983.
- [2] P. Flatau, Fast solvers for one dimensional light scattering in the discrete dipole approximation, *Opt. Express* 12 (2004) 3149–3155.
- [3] D. W. Mackowski, M. I. Mishchenko, Calculation of the T matrix and the scattering matrix for ensembles of spheres, *J. Opt. Soc. Amer. A* 13 (1996) 2266–2278.
- [4] K. A. Fuller, D. W. Mackowski, Electromagnetic scattering by compounded spherical particles, in: M. I. Mishchenko, J. W. Hovenier, L. D. Travis (Eds.), *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications*, San Diego: Academic Press, 2000, Ch. 8, pp. 225–272.
- [5] M. I. Mishchenko, L. D. Travis, A. A. Lacis, *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering*, Cambridge, UK: Cambridge University Press, 2006.
- [6] A. Doicu, T. Wriedt, Computation of the beam–shape coefficients in the generalized Lorenz–Mie theory by using the translational addition theorem for spherical vector wave functions, *Appl. Opt.* 13 (1997) 2971–2978.
- [7] A. Doicu, T. Wriedt, Plane wave spectrum of electromagnetic beams, *Opt. Comm.* 136 (1997) 114–124.
- [8] D. W. Mackowski, Calculation of total cross sections of multiple sphere clusters, *J. Opt. Soc. Amer. A* 11 (1994) 2851–2861.
- [9] D. W. Mackowski, M. I. Mishchenko, Direct simulation of multiple scattering by discrete random media illuminated by Gaussian beams, *Phys. Rev. A* 83 (1) (2011) 013804–+. doi:10.1103/PhysRevA.83.013804.
- [10] [www.eng.auburn.edu/admin/ens/hpcc/](http://www.eng.auburn.edu/admin/ens/hpcc/).

My colleague Owen Awari worked on the FEM calculations. We had to discuss specific input parameters that we wanted to include. We were working with a silicon nanoparticle with a radius of 230 nm, nanoparticle separation of 2R,4R,6R, and the wavelength is 1000-2000 nm at 20 nm steps. The preliminary results that I got need to be compared with colleague to verify my results.

Inorganic chemistry of battery electrodes

Melanie Brito, Department of Materials Science and Chemical  
Engineering, Stony Brook University, Stony Brook, NY, 11794

Esther Takeuchi, Energy Sciences Directorate, Brookhaven National Lab,  
Upton, NY, 11973

## Abstract

The NSLS II at Brookhaven National Lab offers beamlines and the labs of the Takeuchi Research group at Stony Brook University provide instruments that can be used to conduct battery research. An area in which future batteries will be needed in which our group focuses is in portable devices that require fast charging times. Magnetite is a desirable replacement for graphite battery anodes because of its high energy density and low toxicity. Accurate analysis techniques like X-ray diffraction, XRD, and X-ray absorption, XAS, are crucial for material understanding. During XRD, x-rays were diffracted by the different planes of magnetite powder and electrode samples to obtain a pattern that after peakfitting and Rietveld refinement, showed that crystallite sizes were as expected from the synthesis even after the material was put into an electrode. In XAS, the difference in x-ray intensity before and after hitting a sample was measured and, after plotting and analysis, the XANES region gave information about the oxidation state of magnetite used in a pouch cell. Operando data plots showed the reduction of magnetite as having 3 regions, lithium intercalation, SEI formation, and conversion to FeO then Fe metal. Electrochemistry tests like cyclic voltammetry, electrochemical impedance testing, and rate capability can also give insight into how the magnetite performs in a cell. This project impacts BNL and society because as there is a need for more environmentally friendly transportation options, magnetite can be used in electric vehicles in a compact, high capacity, non-toxic battery but in order to do this, the material must be well understood, and this starts with characterization techniques.



# Introduction

The study of batteries is important in tackling current energy problems. With wind and solar renewable energy, weather inconsistency causes the need for energy storage and the solution is large, stationary grid batteries with ample storage capacity. The implementation of these systems will decrease fossil fuel reliance. Batteries are also important in portable devices, namely in electric vehicles, as this would also decrease fossil fuel reliance and CO<sub>2</sub> emissions. Electric vehicles, EVs, need batteries that will allow them to travel long distances on each charge and charge quickly.

A material of interest for portable battery applications is magnetite, especially for use in the anode. Most commonly in lithium-ion batteries, a graphite anode is used but has limitations like lithium dendrite formation occurring on the surface of the anode that impedes proper battery function.<sup>1</sup> Most metal oxides are not very conductive but magnetite is an exception and can be used in battery anodes.<sup>2</sup> Magnetite is also appealing because of its low cost, relative non-toxicity and earth abundance.<sup>1,2</sup> Furthermore, magnetite is a conversion material, meaning that as it is reduced, magnetite becomes iron metal which is desirable because multiple electron transfers make a higher capacity material.<sup>4</sup> There still exist some challenges with magnetite; iron nanocrystals tend to form upon cycling which causes volume fluctuations and electrode cracking.<sup>1</sup> Additionally, degradation takes place at high charge densities.<sup>1</sup> Many studies have looked at ways of improving magnetite functionality. One study looked at the impact of surface treatment on the electrochemistry of magnetite nanoparticles and found that different ligand molecules used in treatment caused improvement in different aspects of the

magnetite<sup>1</sup>. Another study looked at using different polymers as binder in electrodes and a third looked at using magnetite nanorods vs. nanoparticles in electrodes.<sup>2,3</sup> One study focused on magnetite as a conversion material, using XAS and isothermal microcalorimetry to find that while the iron is reduced, an electrolyte reduction is happening as well.<sup>4</sup>

The use of effective analysis techniques is crucial to the study and improvement of magnetite. This project highlights several techniques that can be used to get information from X-ray diffraction patterns from both pristine magnetite powder and pristine magnetite electrodes and operando XAS spectra from a cycled magnetite electrode. During x-ray diffraction, x-rays are shot at a sample, powder or electrode. The different planes of the material cause the rays to be diffracted and sensed by a detector. These rays give the diffraction peaks observed in the pattern because rays that come from the same plane on the material constructively interfere, so each peak observed represents a particular plane.<sup>5</sup> These peaks can be analyzed to give structural information about the material. VESTA was used to view the structure of the magnetite. Peakfit software was used to obtain the crystallite size. Rietveld refinement was also done on the patterns. During XAS, x-rays are shot at the sample and the difference in intensity in the incident and transmission ion chambers is measured.<sup>6</sup> This difference in intensity is plotted versus the energy, the main areas of the plot being the XANES and EXAFS regions. Operando XAS spectra were taken as the magnetite electrode, in a pouch cell, underwent the first discharge, or lithiation. Athena was used to normalize the spectra to see the changing oxidation state and conversion of the material in the XANES region.

Several different electrochemical techniques can be used to analyze magnetite such as cyclic voltammetry (CV), rotating disk electrode cyclic voltammetry (RDE), electrochemical

impedance spectroscopy (EIS) and rate capability testing. These tests can show how magnetite functions in an electrode or in a cell. During cyclic voltammetry, an electric potential is applied to a working and counter electrode and the current is measured and plotted vs. the voltage. The procedure can be performed at different scan rates to find how the scan rate affects peak current and the diffusion layer thickness. Rotating disk electrode cyclic voltammetry uses a similar concept except that the voltage scan rate remains the same and the rotation rate of the working electrode changes; once again, a current vs. voltage plot can be made. During electrochemical impedance spectroscopy, the potential is varied slightly at different frequencies and the impedance is measured. A Nyquist plot of the imaginary vs. the real part of the impedance can be constructed to show the different areas of resistance and capacity. Rate capability testing allows the cell to be cycled at different current densities while the voltage is measured. From the applied current values and the active mass, the specific capacitance can be calculated, and rate capability plots and voltage profiles can be made to show how the cycling rate and number of cycles affect the performance of the cell.

# Research Methods

## Sample Preparation and Measurement

Two different forms of the magnetite system were investigated in the present project. The first form is pristine magnetite powder. X-ray diffraction patterns of this material were taken on the Rigaku SmartLab X-ray diffractometer in the Takeuchi Research Lab at Stony Brook University. Two different crystallite sizes of particles were analyzed, 10 nm particles and 30 nm particles. The second form of magnetite analyzed was magnetite electrodes, one made from the 10 nm powder, and one from the 30 nm powder; the XRD patterns of these samples were taken at the NSLS II at Brookhaven Lab. Electrodes are most often synthesized using a slurry method in which the active material, in this case the magnetite, a carbon additive, and a polymeric binder are mixed, and this slurry is spread out onto a current collector, a metal foil. A different set of magnetite electrodes were used for the operando XAS analysis, also conducted at the NSLS II. Pouch type cells were constructed with a magnetite cathode, a separator, a lithium metal anode, and a liquid electrolyte. The casing for the pouch cell is a flexible plastic metalized foil. 27 scans were taken at different electron equivalencies as the pouch cell discharged and the magnetite underwent lithiation and conversion. Three standards were also measured, pristine magnetite  $\text{Fe}_3\text{O}_4$ ,  $\text{Fe}_2\text{O}_3$ , and  $\text{FeO}$ , and the Fe metal reference foil as well.

## X-Ray diffraction analysis methods

For the 2 powder patterns, an analysis was done on Peakfit. The pattern was loaded into the program and the desired peak isolated by cutting out the other peaks. The background was accounted for, and the peak was fit using a two-point linear method and a Gaussian fit. When using Peakfit, the goal is to get the  $R^2$  value, the error between the fit and the observed pattern, as close to 1 as possible. The rule of thumb used was values of above 0.99 were considered adequate. This same peak fitting technique was repeated for all the significant peaks in the pattern. After fitting, Peakfit gave the full width at half max values, and this was used in the Scherrer equation to get the crystallite size based on each peak in the pattern.

The next step in the XRD pattern analysis was to perform a Rietveld refinement on all 4 patterns. This technique uses phases of the elements and compounds present in the sample and matches a calculated pattern as closely as possible to the observed pattern. This method of fitting is more accurate than fitting each peak individually because it is important to consider the effect of the other peaks, which come from the unique planes in the material, when fitting a specific peak. The program used to perform the Rietveld refinements was GSAS II. This program accepts data files or tiff images from which the program does the 2D to 1D reduction. The first step in fitting is to fit the background with an 8<sup>th</sup> order Chebychev-1 fit to give the program a baseline for the pattern. After the background fit is at a good starting point, the phases are loaded in in the form of CIF files. For the present project, a pristine magnetite phase was used for the powder and electrode patterns. The program was run for 20 cycles to refine the zero and lattice cell parameters. This gave the program freedom to make the adjustments necessary

to get a good initial fit. The cycles were turned back to 3 and the other parameters were refined. Rietveld refinement is tedious and can be tricky so the parameters for isotropic crystallite size, isotropic micro strain, and the atomic positions can be modified in many different orders leading to much trial and error to figure out what worked best. The background was also refined again after the other parameters. Once the fit was a good visual match and the  $R_{wp}$ , the combined error value, was below 6%, the rule of thumb for this project, the refinement was complete. The structure of the finished refinement was exported from GSAS II and VESTA was used to verify the structure.

## **X-ray absorption analysis methods**

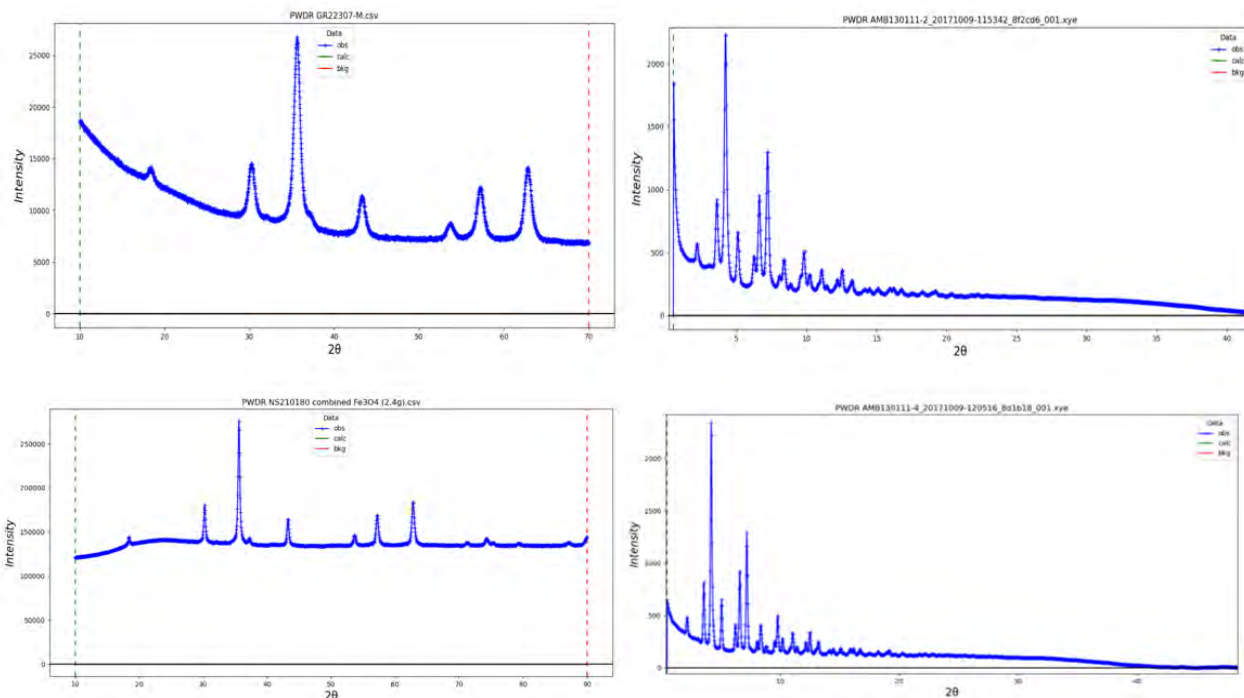
The first step in the analysis of the operando XAS data was the focus on the XANES region. When XAS data is plotted as the normalized absorption coefficient vs. energy, there are 3 main regions of the plot, the pre-edge region, the XANES region, and the EXAFS region. The XANES region gives information about the oxidation state of the element of interest. XAS is element specific so by choosing the energy of the scan, one can choose which element to analyze. Athena was used to normalize the XANES spectra. First the standards,  $Fe_3O_4$ ,  $Fe_2O_3$ ,  $FeO$ , and their references were loaded in. The standards were used to have something with a known oxidation state to serve as a benchmark for the experimental data. References are used to track any changes in the intensity of the beam so that the spectra can be calibrated correctly. The references were calibrated to 7112 eV, the edge energy of iron, and the edge energy of each of the standards was chosen based on the maximum of the first derivative and then, finally, the background points were normalized. Next, the operando spectra were loaded in

with the singular foil reference. The reference was not collected with each scan to reduce the noise in the patterns. The reference foil was calibrated to 7112 eV and the energy shift was adjusted to be the same for all the operando spectra. The maximum of the first derivative was used to find the edge energy points of all the spectra. Once the background points were normalized, the analysis in Athena was done and the data was exported. The energy(eV) and the normalized absorption coefficient were plotted in excel to allow for the comparison of the normalized spectra.

## Results

### X-Ray diffraction analysis

The XRD patterns of the two powder samples had many similarities, as did the two beamline electrode patterns. For both pairs, the peaks were found at about the same angles. The two beamline electrode patterns had peaks at much lower two theta angles as compared to the two powder patterns due to the difference in wavelength of the two instruments. The Rigaku SmartLab diffractometer uses a higher wavelength incident beam than the beamlines at the NSLS II at Brookhaven, leading to the difference in the angles of the observed peaks. It was also observed that the peaks are wider on the 10 nm patterns than the 30 nm which makes sense because smaller crystallite size leads to broader peaks on an XRD pattern. The four XRD patterns are shown in Figure 1.



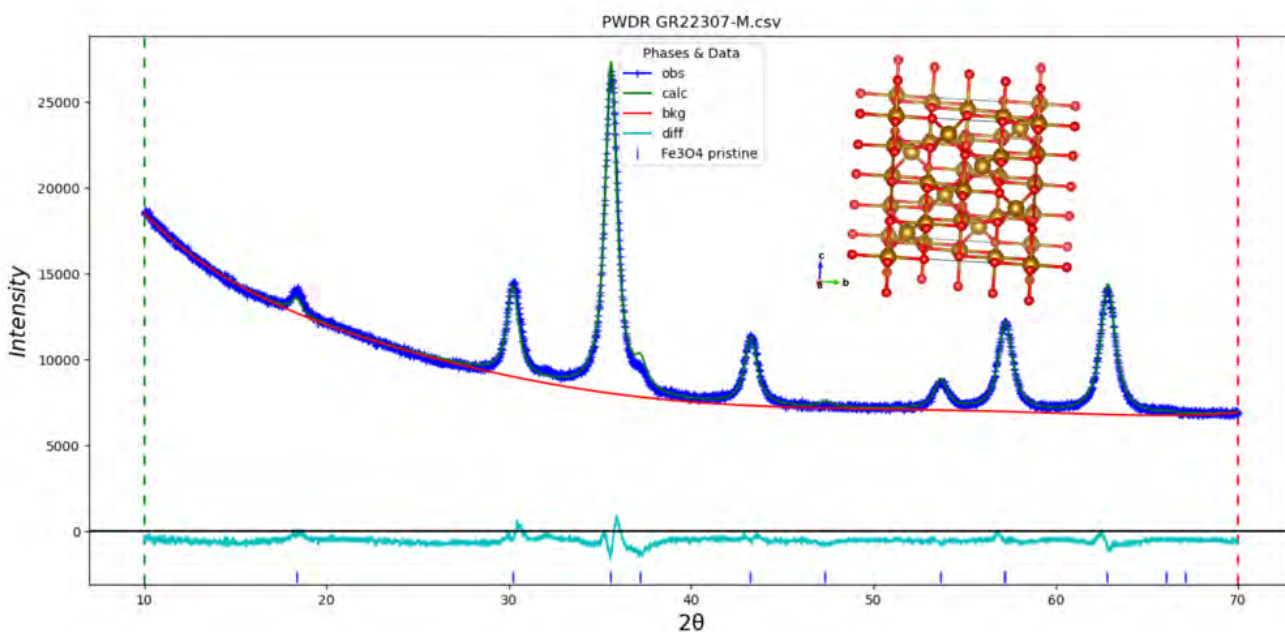
**Figure 1:** Top left is 10 nm powder XRD pattern. Top right is 10 nm electrode. Bottom left is 30 nm powder. Bottom right is 30 nm electrode.

Peakfit analysis of the 10 nm powder XRD pattern, the peaks analyzed being those at angles 30.2551, 35.617, and 43.280, gave average crystallite size 10.71 nm, which is close to the expected 10 nm. Similarly, the peaks that were analyzed on the 30 nm powder pattern, 30.229, 35.598, 43.257, gave average crystallite size 29.987 nm, almost exactly the expected 30 nm. These findings are a testament to the accuracy of the analysis. Figure 1A in the Appendix displays the fitted peaks and crystallite sizes for the XRD powder patterns.

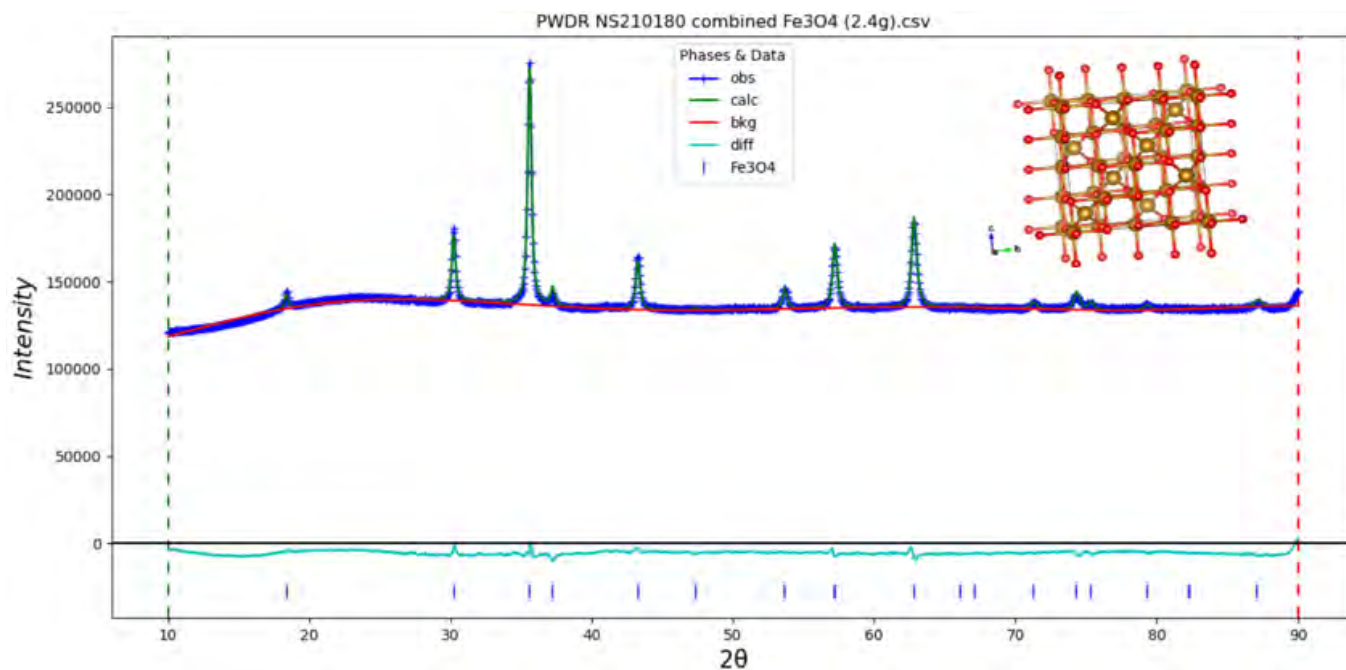
The Rietveld refinements of the powder samples gave accurate and expected results. The 10 nm powder pattern yielded a crystallite size of 13.9 nm with an error of 0.2 nm, close to the expected 10 nm. The micro strain was quite low, 0.42(4), and is expected when considering the structure in VESTA because the polyhedra are not stretched out. The lattice parameter was



8.3623 Å with error 0.0003 Å and the fit had very little error with a total  $R_{wp}$  of 1.9%. The 30 nm powder pattern yielded a crystallite size of 40 nm with error 1 nm. This was a bit further from the expected 30 nm but still within reason. The micro strain was low at 0.16(3) and the lattice parameter was 8.3668 Å with error 0.0002 Å. The pattern was very well fit in terms of intensity, peak broadness and position with an  $R_{wp}$  of 0.80%. The 10 and 30 nm powder post-refinement patterns and structures as well as their results are shown in Figures 2 and 3 below and Table 1A in the Appendix, respectively.



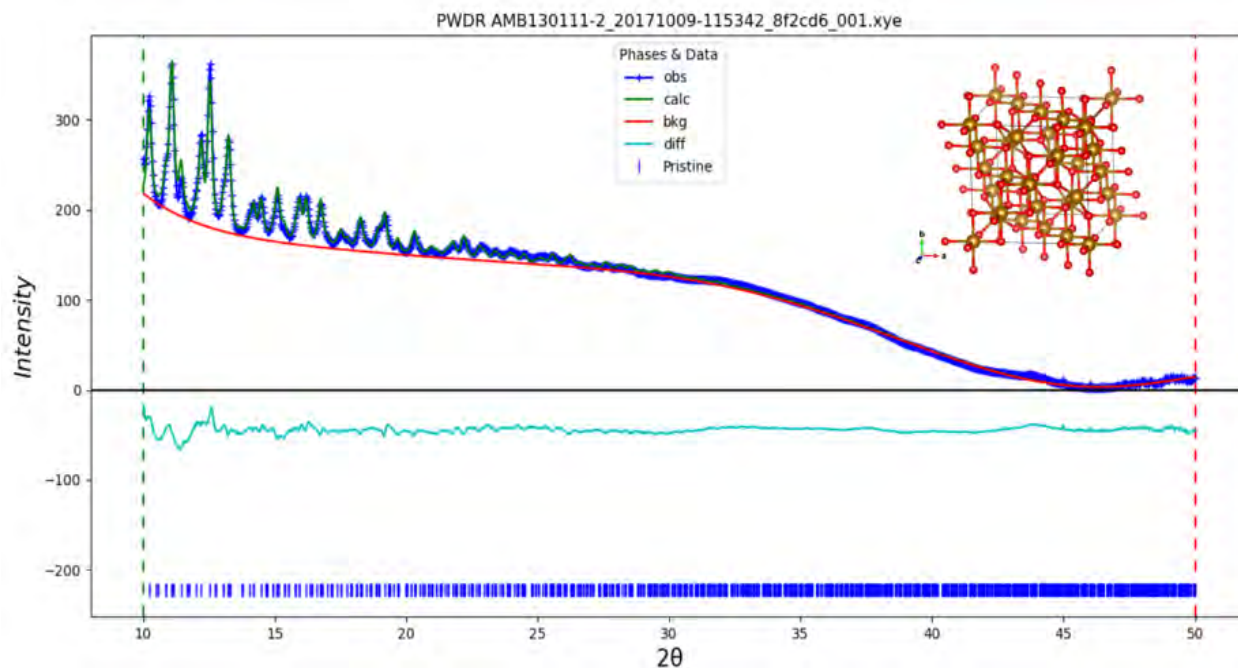
**Figure 2:** Refined XRD pattern for the 10 nm magnetite powder.



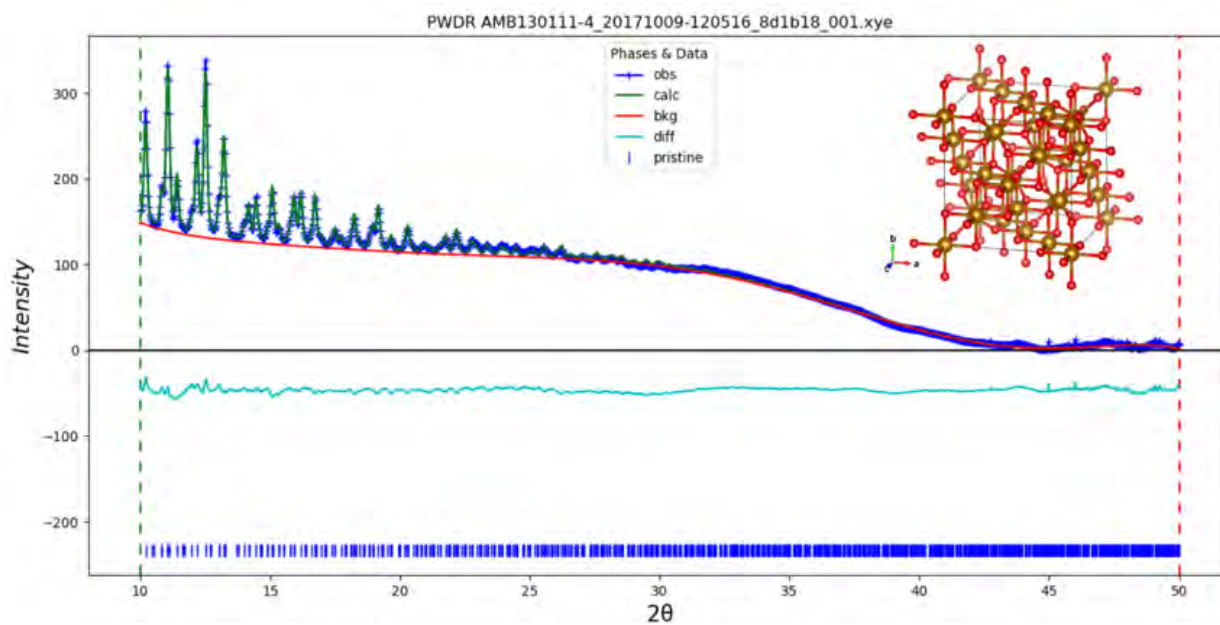
**Figure 3:** Refined XRD pattern for the 30 nm magnetite powder.

The two magnetite beamline patterns from electrodes were a bit more difficult to refine. One problem that arose was the presence of a Teflon background in these patterns so the few Teflon peaks that showed up hindered the fitting of the magnetite phase. The solution was to only fit the later portion of the pattern, angles above 10 degrees, in which the Teflon peaks did not show up. The 10 nm electrode gave crystallite size 15.3 nm with error 2.9 nm, close to the expected 10 nm. This fit had a low micro strain of 0.3(2) and a lattice parameter of 8.396 Å with error 0.0001 Å. All the peaks matched up in position and intensity giving a total  $R_{wp}$  of 4.08%. There were similar findings for the 30 nm magnetite electrode. The crystallite size was close to expected at 31.4 nm but with a slightly larger error of 9 nm. Small micro strain was seen again with the value being 0.12 with error 0.14. The lattice parameter was 8.3942 Å with error 0.0008 Å and the peaks were well fit with a total  $R_{wp}$  of 5.57%. The 10 and 30 nm

electrode post-refinement patterns and structure as well as their results are shown in Figures 4 and 5 below and Table 1A in the Appendix, respectively.



**Figure 4:** Refined XRD pattern for the 10 nm magnetite electrode.



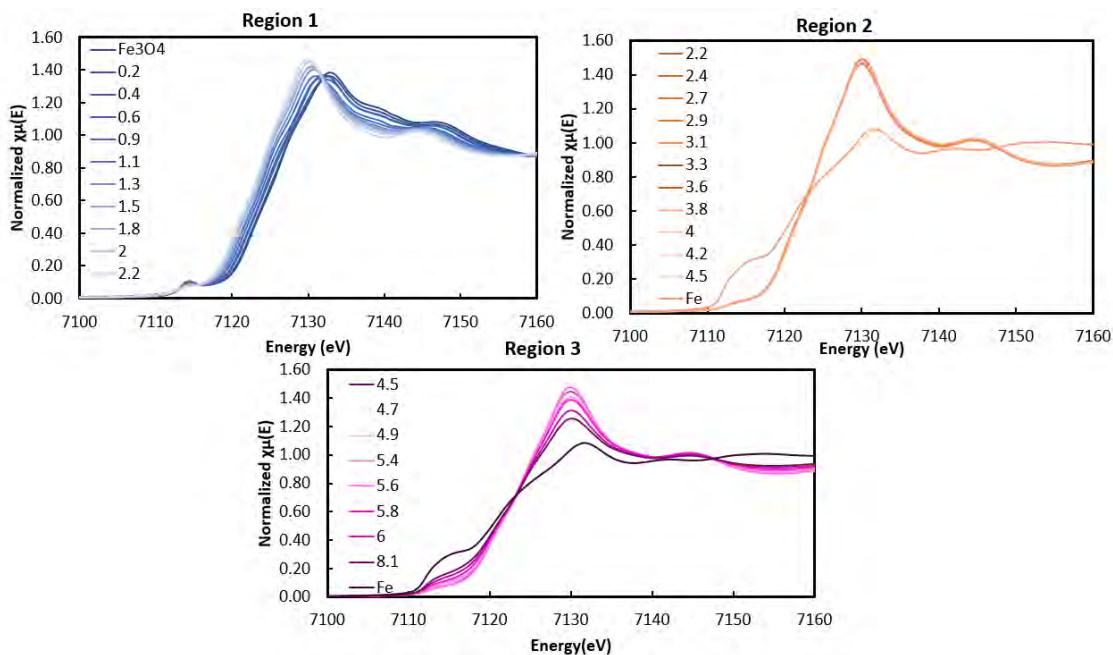
**Figure 5:** Refined XRD pattern for the 30 nm magnetite electrode.

The low error values and expected values and structures indicate that the use of Rietveld refinement to obtain these parameters is accurate and reliable.

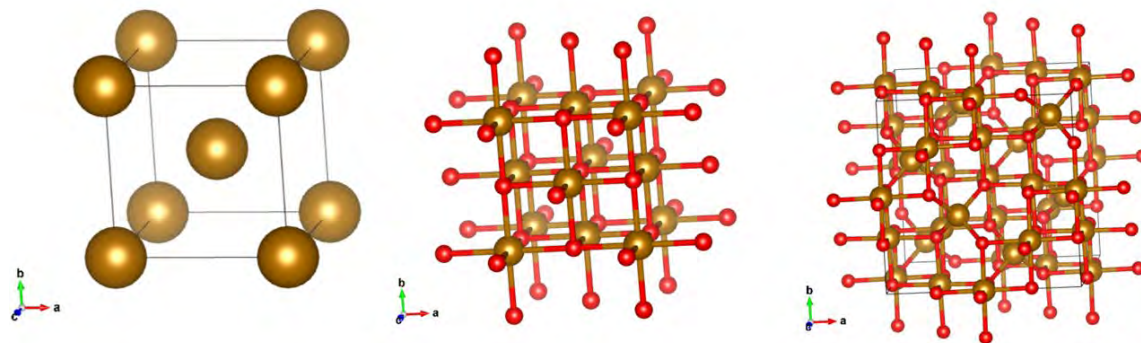
## **X-Ray Absorption Spectroscopy**

Upon plotting the operando XAS-XANES data from the cycled electrode in Microsoft Excel, 3 distinct groups could be seen in the spectra. The spectra from the pristine electrode to 2.2 electron equivalents showed a linear trend of decreasing edge energy as electron equivalents increased. This indicated a reduction of  $\text{Fe}_3\text{O}_4$  spinel to a FeO rocksalt like structure and the intercalation of lithium ions into the magnetite. As an element is reduced, the nucleus becomes less positive, making it easier for a core electron to be kicked out of the atom, thus the reason for lower edge energy for a more reduced metal atom.<sup>6</sup> The spectra from 2.2 to 4.5 electron equivalents were almost all identical, indicating almost no reduction of iron during that time. The continued presence of capacity indicated another reduction taking place which was most likely an electrolyte reduction. The liquid electrolyte in the pouch cell reacts with lithium ions to produce a solid electrolyte interphase; this interphase acts as a protective barrier for the electrode but when too thick, can impede ion diffusion.<sup>4</sup> From 4.5 to 8.1 electron equivalents, it became apparent that there was a change in the iron taking place. In these spectra, there was not a huge energy shift but rather the difference noted was in the white line intensity. As the electron equivalencies went up, the white line intensity went down and there was a more prominent presence of a peak around 7112 eV, the edge energy of the iron metal. This indicates that there is a structural change and conversion from the FeO like structure to iron metal, thus the indication of the full reduction of magnetite at the end of the discharge. The

XANES plots of the 3 regions of spectra are shown in Figure 6 below. The Fe metal foil reference spectra are shown on the plots for regions 2 and 3 for comparison.



**Figure 6:** Operando XANES plots of the magnetite electrode. Top left is region 1 from the pristine electrode to 2.2 electron equivalents. Top right is region 2, from 2.2 to 4.5 electron equivalents. Bottom is region 3, from 4.5 to 8.1 electron equivalents.



**Figure 7:** VESTA structures of Iron metal (left), FeO (center), and Magnetite (right)

## Conclusion

During the study and characterization of materials, accurate and reliable analysis techniques are crucial for material and mechanism understanding. XRD can be used to identify crystallite sizes which can provide important insight into a material's structure. The present project showed the reliability of this technique from magnetite peakfitting and Rietveld refinements that gave crystallite sizes exactly as expected for both powder and electrode samples, a testament to the rangeability of the technique. Tracking the reduction of an electrode in a battery is crucial to understanding the mechanism and tracking its progress. XAS, specifically the XANES region, is very useful for this. Operando data of a magnetite electrode during discharge was seen in the present project and gave information about SEI formation and conversion of the material to iron metal, something that would have not been possible without XAS. Magnetite is a promising anode material and its use in portable devices, namely electric vehicles, can help to create more sustainable power solutions and the study of this material using versatile and accurate analysis techniques is extremely important.

## Acknowledgements

I would like to thank my mentor Dr. Esther Takeuchi for all her guidance and support. I would also like to thank Dr. Kenneth Takeuchi, Dr. Amy Marschilok, Dr. Chavis Stackhouse, Dr. Lisa Housel, Edelmy Marin-Bernardez, Steven Tyler King, Gurpreet Singh, Nahian Sadique, Christopher Tang, Cynthia Huang, and Jason Kuang for their help and guidance. Additionally, I would like to thank Olivia Bennett and Marie Millares for their collaboration this summer.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTs) under the Science Undergraduate Laboratory Internships Program (SULI).

## Appendix

**Table 1A:** Refinement data for 10 and 30 nm magnetite powder and electrodes

Sample	10 nm magnetite powder	30 nm magnetite powder	10 nm magnetite electrode	30 nm magnetite electrode
Phase	Pristine Fe <sub>3</sub> O <sub>4</sub>	Pristine Fe <sub>3</sub> O <sub>4</sub>	Pristine Fe <sub>3</sub> O <sub>4</sub>	Pristine Fe <sub>3</sub> O <sub>4</sub>
Space Group	<i>F d-3 m</i>	<i>F d-3 m</i>	<i>F d-3 m</i>	<i>F d-3 m</i>
a,b,c (Å)	8.3623(3)	8.3668(2)	8.396(1)	8.3942(8)
α, β, γ (deg.)	90	90	90	90
C.S (nm)	13.9(2)	40(1)	15.3(2.9)	31.4(9.0)
Microstrain (%)	0.42(4)	0.16(3)	0.3(2)	0.12(0.14)
R <sub>wp</sub> (%)	1.90	0.80	4.08	5.57



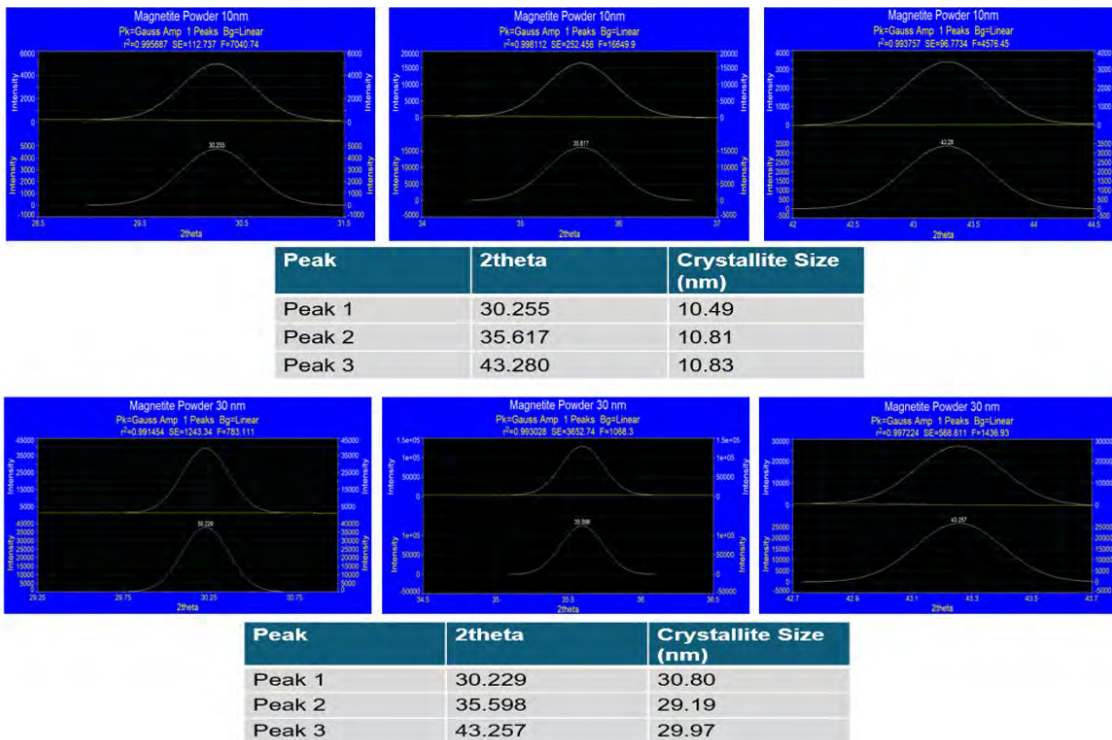


Figure 1A: Fitted peaks and sizes for the 10 nm (top) and 30 nm (bottom) powder patterns.



## References

<sup>1</sup>Wang, L.; Housel, L. M.; Bock, D. C.; Abraham, A.; Dunkin, M.; McCarthy, A.; Wu, Q.; Kiss, A.; Thieme, J.; Takeuchi, E. S.; Marschilok, A. C.; Takeuchi, K. J. Deliberate Modification of Fe<sub>3</sub>O<sub>4</sub> Anode Surface Chemistry: Impact on Electrochemistry, *ACS Appl. Mater. Interfaces*, 11, 19920-19932 (2019). [DOI: [10.1021/acsaami.8b21273](https://doi.org/10.1021/acsaami.8b21273)]

<sup>2</sup>Minnici, K.; Kwon, Y. H.; Housel, L. M.; Renderos, G. D.; Ponder, J. F.; Buckley, C.; Reynolds, J. R.; Takeuchi, K. J.; Takeuchi, E. S.; Marschilok, A. C.; Reichmanis, E. Tuning Conjugated Polymers for Binder Applications in High Capacity Magnetite Anodes, *ACS Appl. Energy Mater.*, 2, 7584-7593 (2019). [DOI: [10.1021/acsaem.9b01580](https://doi.org/10.1021/acsaem.9b01580)]

<sup>3</sup>McBean, C. L.; Wang, L.; Moronta, D.; Scida, A.; Li, L.; Takeuchi, E. S.; Takeuchi, K. J.; Marschilok, A. C.; Wong, S. S. Examining the Role of Anisotropic Morphology: Comparison of FreeStanding Magnetite Nanorods versus Spherical Magnetite Nanoparticles for Electrochemical Lithium-Ion Storage, *ACS Appl. Energy Mater.*, 2, 4801-4812 (2019). [DOI: [10.1021/acsaem.9b00456](https://doi.org/10.1021/acsaem.9b00456)]

<sup>4</sup>Huie, M.M.; Bock, D.C.; Wang, L.; Marschilok, A.C.; Takeuchi, K.J.; Takeuchi, E.S., Lithiation of Magnetite (Fe<sub>3</sub>O<sub>4</sub>): Analysis Using Isothermal Microcalorimetry and Operando X-ray Absorption Spectroscopy, *J. Phys. Chem. C*, 122, 10316–10326 (2018). [DOI: [10.1021/acs.jpcc.8b01681](https://doi.org/10.1021/acs.jpcc.8b01681)]

<sup>5</sup>Cullity, Bernard Dennis., and Stuart R. Stock. *Elements of X-Ray Diffraction*. Pearson Education Limited, 2014.

<sup>6</sup>Calvin, Scott. *XAFS for Everyone*. Taylor & Francis, 2013.

**Isaiah Brown Rodriguez**  
**James Bush**

# Energy Extraction in a Superconducting Magnet

Suffolk County Community College, Ammerman Campus, 533 College Rd, Selden, NY 11784  
Superconducting Magnet Division, Brookhaven National Laboratory, Upton, NY 11973



**Abstract:**

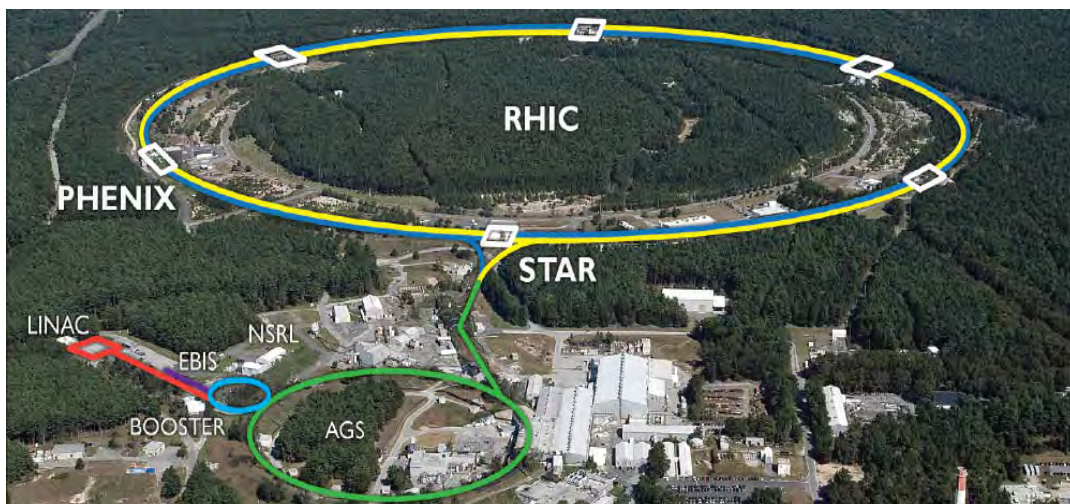
At Brookhaven National Laboratory (BNL), my partner and I simulated and calculated data for a couple of quench detection circuits used while testing superconducting magnets. Brookhaven National Laboratory designs, tests, and builds magnets for particle accelerator applications. Currently, they are testing magnets that will be used for the Large Hadron Collider (LHC) upgrade. BNL is also home to the Relativistic Heavy Ion Collider (RHIC) consisting of 1740 superconducting magnetic elements built by the Superconducting Magnet Division (SMD).

During our internship at BNL, we researched how we can dissipate high energy in the circuit without compromising the integrity of the superconducting magnet during a quench. A superconductor is an ideal inductor that is cooled with liquid helium to temperatures near absolute zero to achieve no electrical resistance. This allows a large current to flow freely through the superconductor without power losses. However, if for any reason during testing, any of the magnet's voltage tap signals reach a specified threshold for a certain amount of time, this will represent a rise in temperature at a particular location, the system's quench detection systems will quickly activate and rapidly interrupt the energy flowing through the coil and allow it to be quickly extracted and discharged through the dump circuit.

During our time at BNL, we simulated multiple quench protection systems. One, in particular, was in series with a dual 5KA power supply system. We simulated a couple of different energy extraction and power supply systems. The magnet under test determines the setup used. The types of quench protection systems we simulated were an SCR switch network, which utilized an LC pulse forming network and an Insulated-Gate Bipolar Resistor (IGBT) switch.

## Introduction:

Brookhaven National Lab is a United States Department of Energy (DOE) national laboratory that specializes in nuclear and high energy physics research. Brookhaven Lab (BNL) has housed numerous one-of-a-kind particle accelerators, one of them being the famous Relativistic Heavy Ion Collider, also known as RHIC. RHIC is an atomic collider consisting of two separate rings approximately 2 miles long that accelerate heavy ions at the speed of light in opposite directions and collide at various intersection points. Scientists study these collisions in hopes of understanding the creation of the universe also known as the Big Bang Theory. To achieve this, superconducting magnets are used to steer the beam of ions around this ring. A superconductor allows direct current to flow without electrical resistance. To achieve this, they are cooled with superfluid helium (He) at temperatures close to absolute zero, hence why we need cryogenic cooling systems. The superconducting magnet division is responsible for creating these specialized magnets that steer the heavy ion beams into collisions, and not just for BNL, but for other national and international particle colliders .



**Figure 1:** This shows the basic outline of the Relativistic Heavy Ion Collider.

Some of the magnets being tested at BNL are going to be used for the Large Hadron Collider (LHC). The LHC is a Ion collider that accelerates protons or ions to near the speed of light. It consists of a 27-kilometer ring of superconducting magnets

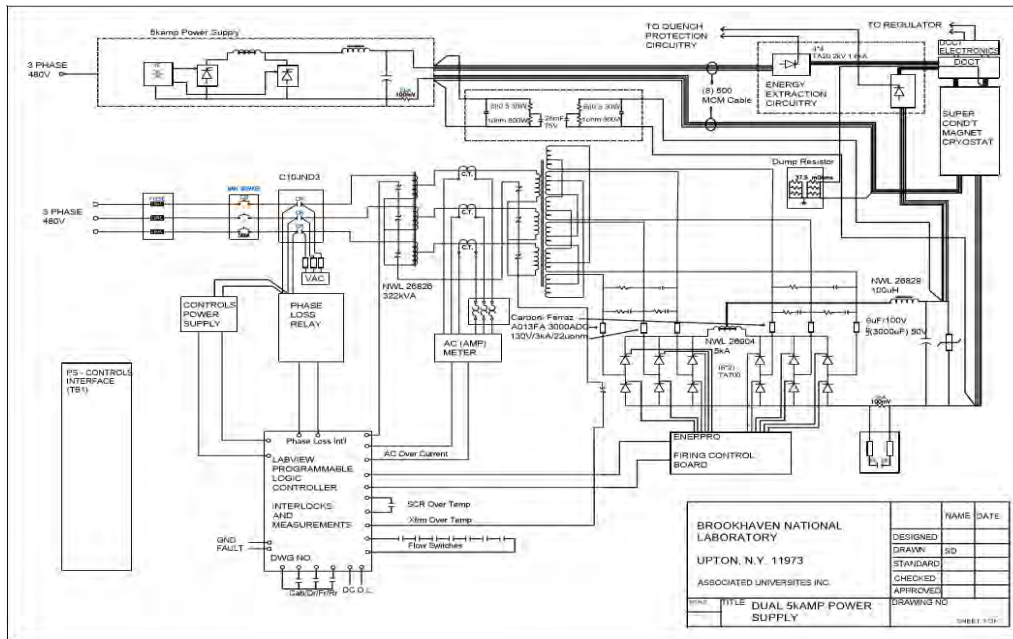
with several accelerating structures that boost the energy of the particles along the way. Located in Geneva, Switzerland it first ran on September 10, 2008, and is expected to run over the next 20 years with several stops scheduled for upgrades and maintenance work. The LHC has had many achievements in the past years including world records and being the first to have the ions collide at 8Tev. The LHC costs about 80% of the CERN annual budget for operation, maintenance, technical stops, repairs, and consolidation work in personnel and materials (for machines, injectors, computing, experiments). Our testing will help lower the cost in operations, maintenance, technical stops, and repairs as we make testing a more effective and efficient quench protection system that quickly stops the flow of current and extracts the energy.



**Figure 2:** This shows a superconducting magnet used for the LHC upgrade.

### **Power Supply:**

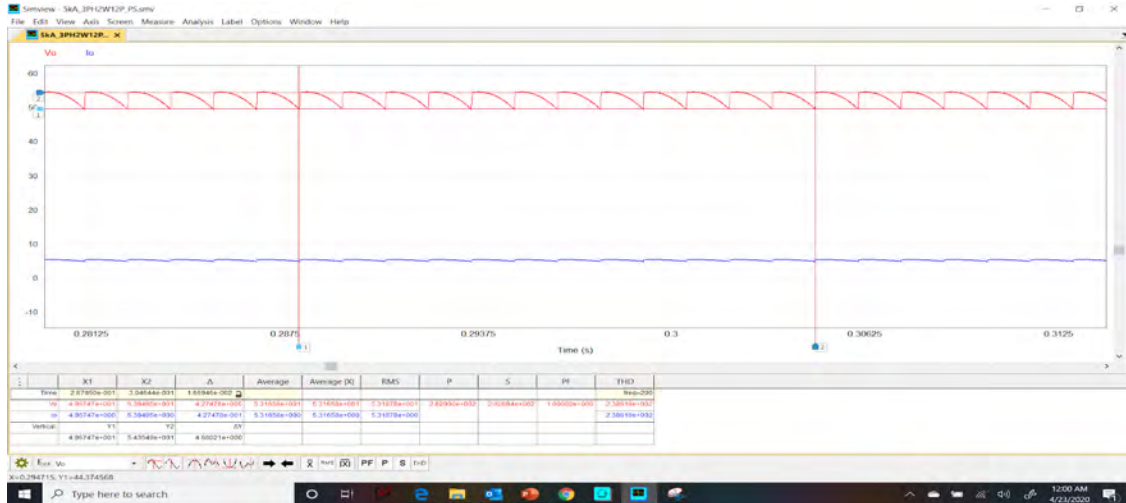
In our study, we examined the use of a 5K amp power supply that has an output of 50VDC and 5kA. The output is then passed through a low pass filter resulting in a <2% ripple. In Figure 3, the power supply is made up of various stages from a line input of (3phase) 480V delta/star step down transformation configuration, to a complex dual bridge SCR rectification system, then through a low pass filter to achieve the desired output of 50VDC at 5kA with <2% ripple.



**Figure 3:** This shows two 5KA power supplies ran in parallel

The 3 phase 480VAC input delta/star configuration introduces a 30 degrees phase shift necessary for operations. A phase loss relay is used to make sure all three phases are present for PS protections. Next, a transformer is used to lower the voltage to a more usable value. The AC then gets rectified via 12 SCR switches that fire at the perfect timing to create 50VDC with a little ripple. The more stages can reduce the output ripple even more, but at a higher cost and greater complexities. Connected to this sequence is an Interphase transformer that is used when two rectifier bridges are used in parallel to balance the voltages between them the rectified AC then passes through a low pass filter to make the output a little cleaner. In Figure 4, you can see the output around 50VDC with a little ripple in it [3].





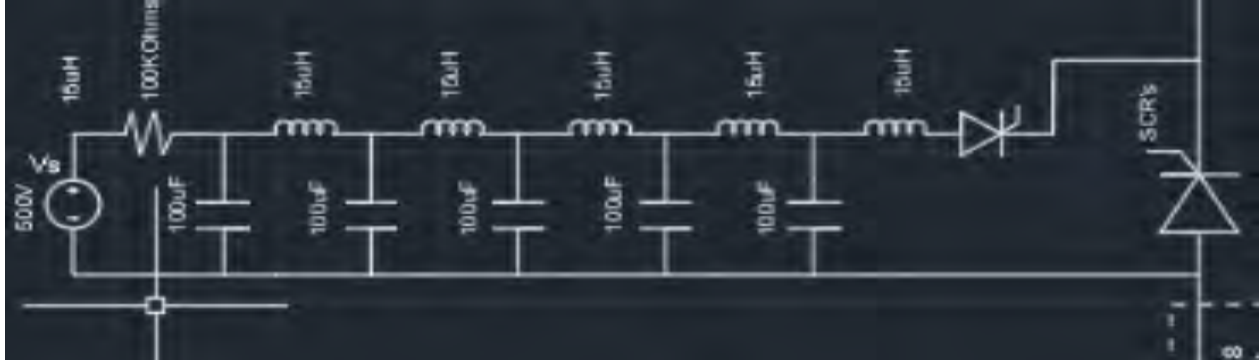
**Figure 4:** shows the output of the 5KA power supply at 50 Volts DC with a little ripple.

### Quench Protection System:

A critical component of any superconducting system is its energy extraction circuit. Energy extraction is especially important because once the power source connected to this magnet is shut-off, the residual energy needs to be extracted as quickly as possible to prevent damage to the superconducting material. The main purpose of these two circuits is to safely and efficiently remove the remaining energy that is stored within the magnet.

### SCR Switch

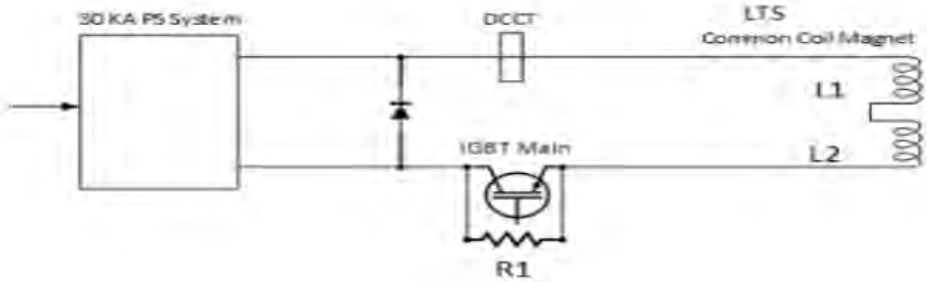
The LC discharge network will be engaged during a quench to quickly stop the current flowing from the magnet, and then allow the stored energy to dissipate across the dump resistor. A quick pulse of equal but opposite current would allow the SCR switch to open, stopping the current source to the magnet. A trigger from the quench detection system allows the LC network to discharge its stored energy.



**Figure 5:** This shows an SCR quench protection system.

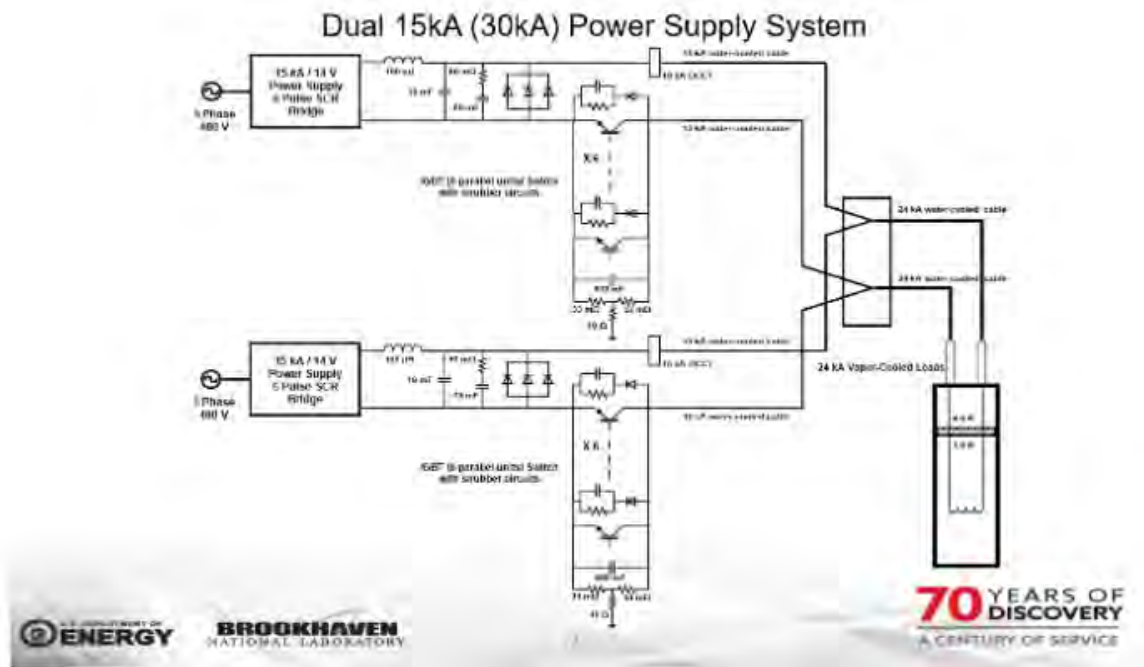
**IGBT Switch**

This is a quite simple circuit at a glance however it is a necessity because the energy within a superconducting magnet does not just dissipate instantly when the power supply is shut off. There is a lot of energy that is stored within the inductor’s magnetic field which needs to be extracted out of the magnet before or as it comes out of the superconducting stage. A fast energy extraction system is crucial for this part and is achieved by using high-power IGBT switches. When a quench is detected, a trigger is sent to the IGBT which opens the flow of current from the source. As seen in Figure 7 below, energy flows from the inductors are now passed through the IGBT when the gate opens which allows all remnant energy to be fed to the dump circuit. This resistor is sized appropriately to handle the circuit’s copious amounts of energy. This is important because if this resistor is not sized appropriately to handle the vast amounts of energy, it can fail to cause catastrophic damage to the system or magnet.



**Figure 6:** This is a block diagram of a typical energy extraction circuit currently used for LHC upgrade magnet testing.





**Figure 7:** The Circuit shown here is a superconducting magnet submerged in coolant, connected to the power supply and energy extraction systems through 24kA leads.

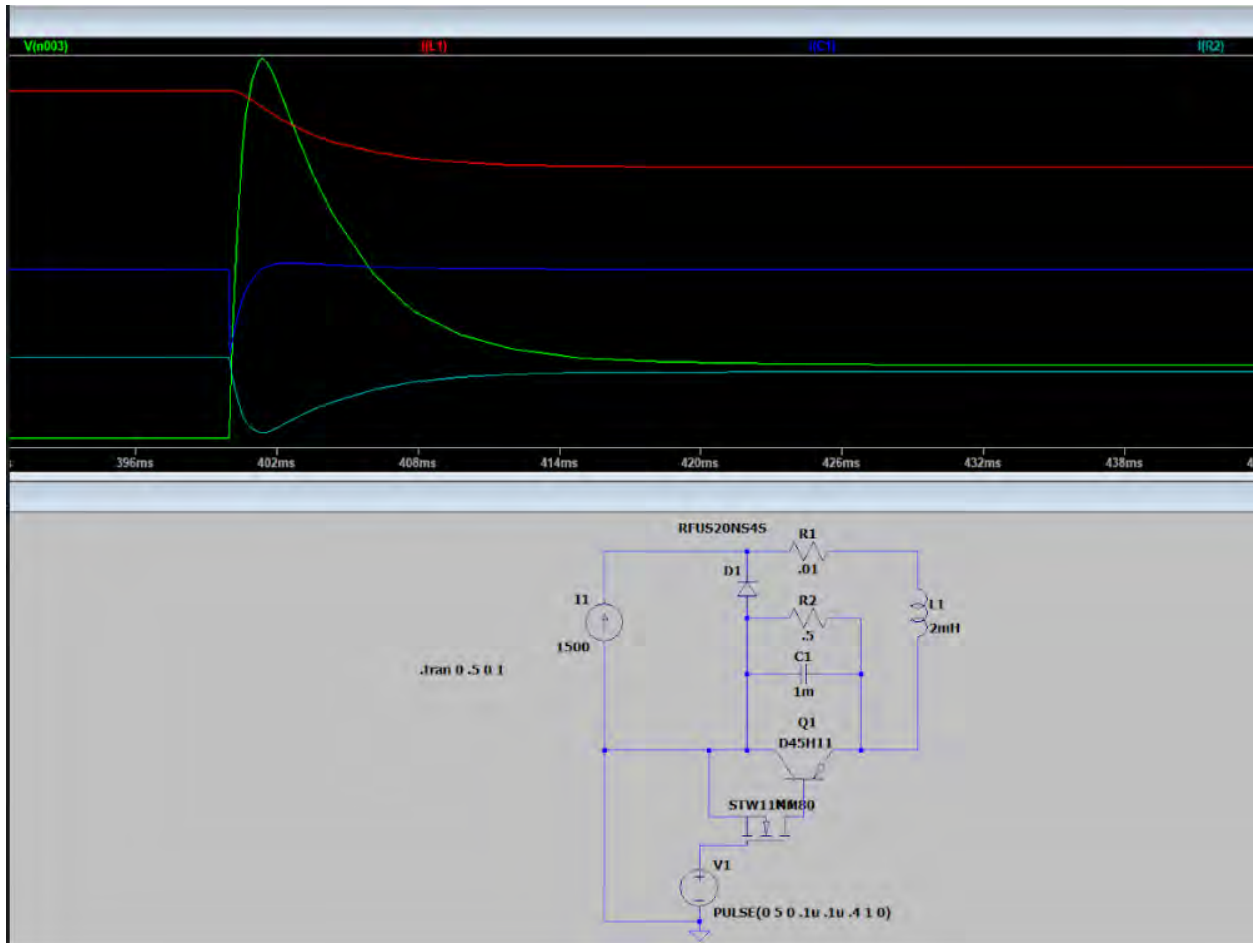
**Methods/ Formulas:**

- We had meetings with our mentors and our professors at SCCC.
- Reviewed any report/PowerPoint given to us.

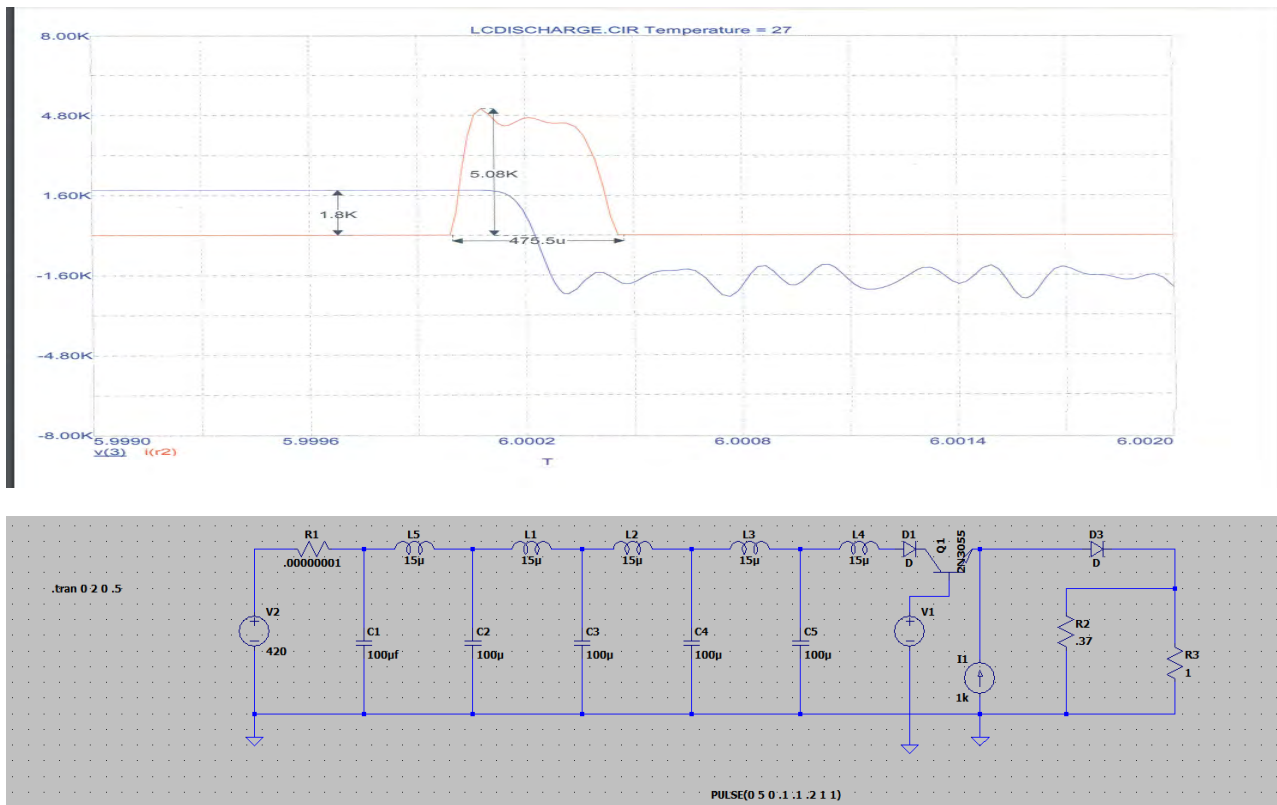
**Programs:**

During the internship, we used programs such as LTSpice, Multisim, and AutoCAD. LTSpice and Multisim were used to simulate circuits and collect all our data. These programs were helpful and allowed us to collect data even though we never worked in person with the circuit. AutoCAD was used to draw up technical drawings that we could use whenever we would make a report.

## Results:



**Figure 8:** The circuit above is a simulation of an IGBT dump circuit made on LTSpice. When the simulation is running, it shows the rapid response of the circuit's components during a potential quench. However, we substituted the IGBT with a transistor with equivalent values to effectively represent real-life circumstances. This process is shown through the graph, you can see that the voltage (green) and current (red) rapidly change throughout the graph. The superconducting magnet is essentially an ideal inductor, this energy extraction circuit will rapidly dissipate its energy over the course of about 15ms. The calculated results of the inductive voltage kickback that occurs during a quench, which appears across the dump resistor (r2) matches with the simulated values shown above. Showing that this energy extraction circuit is effective during a quench.



**Figure 9:** The circuit above is a LC pulse forming network and was created on LTSpice to simulate an existing circuit our mentor provided us with. This circuit shows a rapid 400us pulse from the LC pulse forming network that will inject an equal but opposite current into the circuit. The pulse has to be 400us because there is just enough time to make sure all the SCRs turn off during a process called commutation. This is necessary because during a quench, it is important to stop the large incoming current from the power supply from continuing into the superconducting magnet. Our purpose for this circuit is to stop the energy from continuously entering the circuit during a quench.

### Future Work:

We are currently using simulation software to get our data and conduct research. Hopefully, in the summer internship, we will be working in person, get hands-on experience and get used to the systems used in superconducting magnet testing. With this experience, we hope to get a job and implement everything we have learned.

**Conclusion:**

In conclusion, we were able to build a simple energy extraction circuit on LT spice, simulate and calculate results in our efforts to understand this important concept of energy extraction in superconducting systems. The IGBT quench protection system was simulated with LTSpice and was able to simulate how it would react to a potential quench. The LC pulse forming network was also simulated on LTSpice. Doing this allowed us to manipulate the values of components and change voltages and timings that would provide us with the results that we needed. Essentially the proper time, voltage, and current that we needed to negate the incoming current during a quench. These simulations prove that when made properly an energy extraction system will allow superconducting magnets to be used safely with minimal risk in current and future accelerator projects such as RHIC, EIC, and LHC.

**Acknowledgments:**

I would like to express our appreciation to our mentors, Piyush Joshi and Suresh Deonarine, for the time and effort they put into making our internship a valuable and rewarding experience for us. Suresh was always readily available, knowledgeable, kept communication open, and we are extremely grateful. I would also like to express our appreciation for Melvin Morris and Karl Clarke, the OEP team leaders, for always being readily available and setting up every meeting and interview needed for this project. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program (CCI).

**References:**

Joshi, P. (n.d.). *LTS-HTS Hybrid Dipole Magnet Quench Protection System*. Brookhaven Laboratory.

*LHC Facts and Figures*. (2021). Retrieved March 8, 2021, from CERN Accelerating science:

[https://home.cern/sites/home.web.cern.ch/files/2018-07/CERN-Brochure-2017-002-Eng\\_0.pdf](https://home.cern/sites/home.web.cern.ch/files/2018-07/CERN-Brochure-2017-002-Eng_0.pdf)

*LHC Facts and Figures*. (2021). Retrieved March 8, 2021, from CERN Accelerating science: [https://home.cern/sites/home.web.cern.ch/files/2018-07/factsandfigures-en\\_0.pdf](https://home.cern/sites/home.web.cern.ch/files/2018-07/factsandfigures-en_0.pdf)

Velavan, R., & Myvizhi, P. (2018). Superconductors and its Applications. *International Journal of Pure and Applied Mathematics*, 119(12), 7377-7386.

[Magnet-Division-Cryogenic-Facility\[2\].docx](#) Piyush Joshi, Ph.D. Kathleen Amm, Ph.D. -BNL

[LTS-HTS Hybrid Dipole Magnet Quench Protection System](#). Piyush Joshi, Ph.D.- BNL

## Appendix:

Participants:

Name	Institution	Role
Piyush Joshi	Brookhaven National Laboratory	Mentor, project lead.
Suresh Deonarine	Brookhaven National Laboratory	Mentor, Provided leadership
Peter Maritato	Suffolk Community College	College Faculty, provided qualitative information
Richard Biscardi	Suffolk Community College	College Faculty provided qualitative information
Melvin Morris	Brookhaven National Laboratory	Faculty participant provided qualitative information
Karl Clarke	Brookhaven National Laboratory	Faculty participant provided qualitative information
Noel Blackburn	Brookhaven National Laboratory	Faculty participant provided qualitative information

**James Bush**

# Feedback Control Loop for a Power Supply

Suffolk County Community College, Ammerman Campus, 533 College Rd, Selden, NY 11784

Superconducting Magnet Division, Brookhaven National Laboratory, Upton, NY 11973



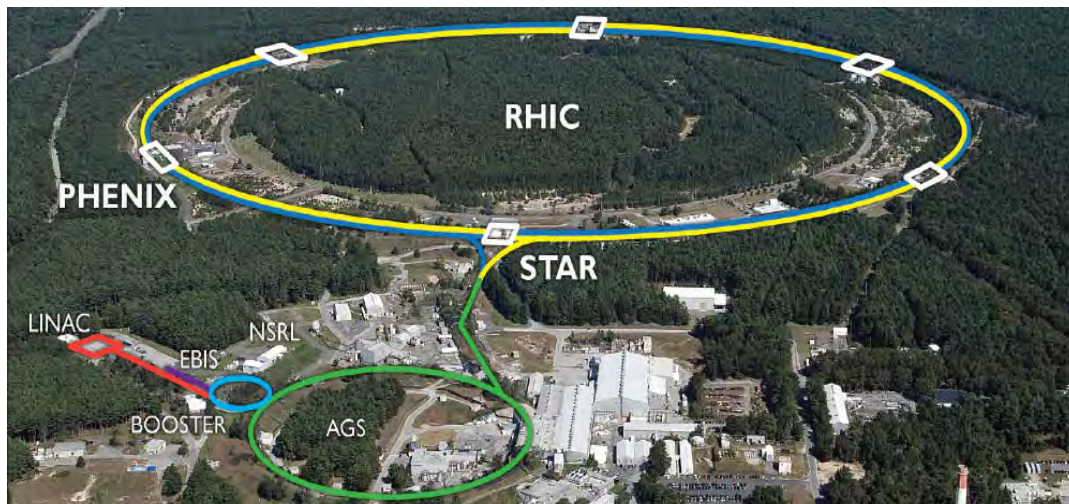
**Abstract:**

At Brookhaven National Laboratory (BNL), my partner and I are learning and simulating a feedback control loop for a power supply. Brookhaven National Laboratory designs, tests, and builds magnets for particle accelerator applications. Currently, they are testing magnets that will be used for the Large Hadron Collider (LHC) upgrade. In this process, We have to make sure the power supplies are within the parameters and if it isn't, we have to shut it down. There are so many different parameters that the power supply has to be in such as voltage, current, heat, time and so many more. This research will help understand how feedback loops function and how they are used. We researched and got started on building and simulating a feedback control loop. This feedback control loop is to protect the power supply and make sure it is putting out the correct values. If the power supply puts out the wrong output and the feedback control loop catches it, it will correct itself. During our time at BNL, we simulated multiple circuits in preparation for the feedback control loop such as a basic power supply, low pass filter and used various tools such as a spectrum analyzer and an oscilloscope. Depending on the frequency, we decide which low pass filter to use.

**Introduction:**

Brookhaven National Lab is a United States Department of Energy (DOE) national laboratory that specializes in nuclear and high energy physics research. Brookhaven Lab (BNL) has housed numerous one-of-a-kind particle accelerators, one of them being the famous Relativistic Heavy Ion Collider, also known as RHIC. RHIC is an atomic collider consisting of two separate rings approximately 2 miles long that accelerate heavy ions at the speed of light in opposite directions and collide at various intersection points. Scientists study these collisions in hopes of understanding the creation of the universe also known as the Big Bang Theory. To achieve this, superconducting magnets are used to steer the beam of ions around this ring. A superconductor allows direct current to flow without electrical resistance. To achieve this, they are cooled with superfluid helium (He) at temperatures close to absolute zero, hence why we need cryogenic cooling systems. The superconducting magnet division

is responsible for creating these specialized magnets that steer the heavy ion beams into collisions, and not just for BNL, but for other national and international particle colliders .



**Figure 1:** This shows the basic outline of the Relativistic Heavy Ion Collider.

Some of the magnets being tested at BNL are going to be used for the Large Hadron Collider (LHC). The LHC is a Ion collider that accelerates protons or ions to near the speed of light. It consists of a 27-kilometer ring of superconducting magnets with several accelerating structures that boost the energy of the particles along the way. Located in Geneva, Switzerland it first ran on September 10, 2008, and is expected to run over the next 20 years with several stops scheduled for upgrades and maintenance work. The LHC has had many achievements in the past years including world records and being the first to have the ions collide at 8Tev. The LHC costs about 80% of the CERN annual budget for operation, maintenance, technical stops, repairs, and consolidation work in personnel and materials (for machines, injectors, computing, experiments). Our testing will help lower the cost in operations, maintenance, technical stops, and repairs as we are making sure the correct outputs are being put out.

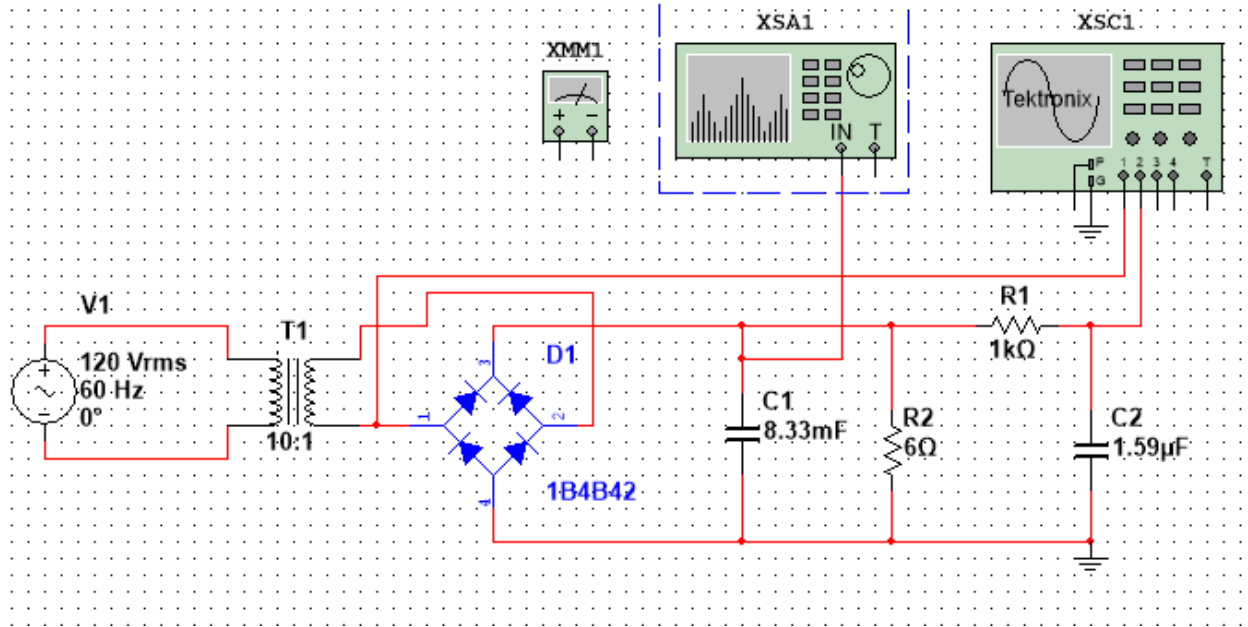




**Figure 2:** This shows a superconducting magnet used for the LHC upgrade.

### **Power Supply:**

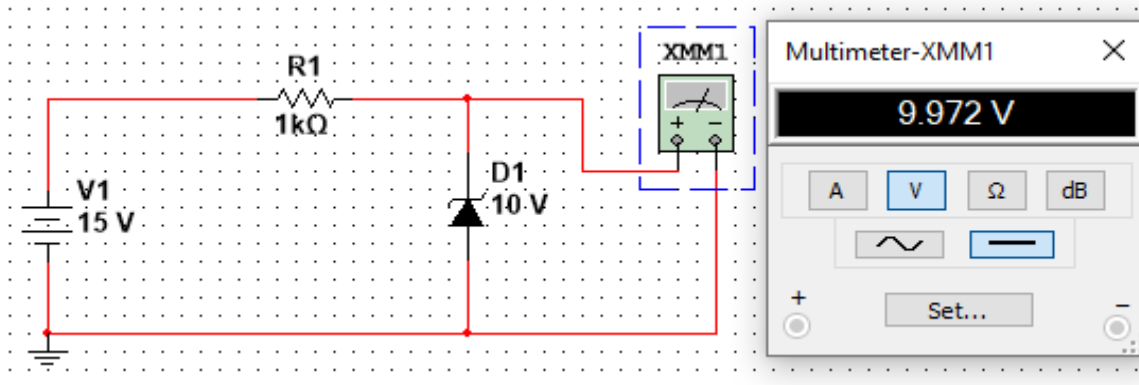
In our study, we examined the use of a basic DC power supply. In order to make this we used an AC power supply, transformer, bridge rectifier, and a low pass filter. First we have the AC power supply with 120 Volts at 60 Hertz, then it goes to a transformer with a 10 to 1 ratio. This turns the 120 volts AC to 12 volts AC and now since it is not smaller, it is more easy to manipulate. The 12 Volts now goes to a bridge rectifier which then turns the AC into DC and with this the DC voltage is more easy to control. After the bridge rectifier, there is a resistor and a capacitor that controls the Dc signal and how big it can get. The signal isn't perfect after the Bridge rectifier so it needs some extra components to smooth out the signal. Last but not least is the low pass filter, which only allows a certain frequency through and at this point we get a nice smooth signal that won't change too much.



**Figure 3:** This shows a basic DC power supply with a low pass filter

### Zener Diode:

For the next step in our project I researched zener diodes. A zener diode is a silicon semiconductor that allows current to flow in either a forward or reverse direction. This diode has a special, heavily doped p-n junction that is designed to conduct in the reverse direction up to a certain point. In the figure below is a basic 10 volt regulator, no matter how high you raise the voltage on the power supply, it will only allow 10 volts to go in the reverse direction.



**Figure 4:** This shows a Basic 10-volt Regulator

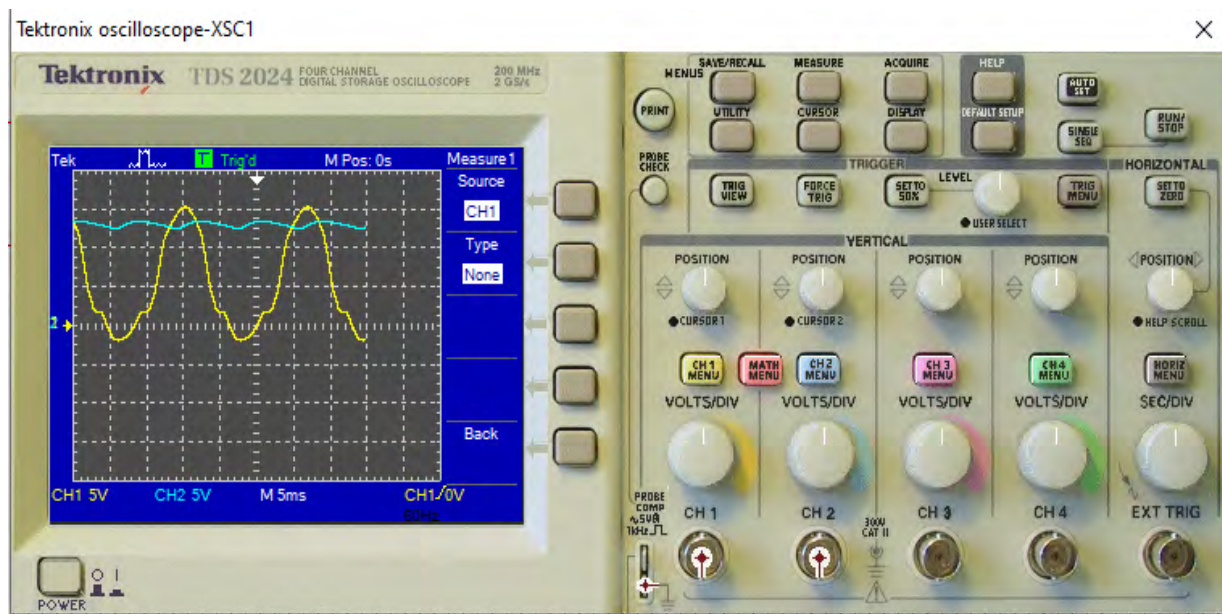
## Methods/ Formulas:

- We had meetings with our mentors and our professors at SCCC.
- Reviewed any report/PowerPoint given to us.

## Programs:

During the internship, we used programs such as LTSpice, Multisim, and AutoCAD. LTSpice and Multisim were used to simulate circuits and collect all our data. These programs were helpful and allowed us to collect data even though we never worked in person with the circuit. AutoCAD was used to draw up technical drawings that we could use whenever we would make a report.

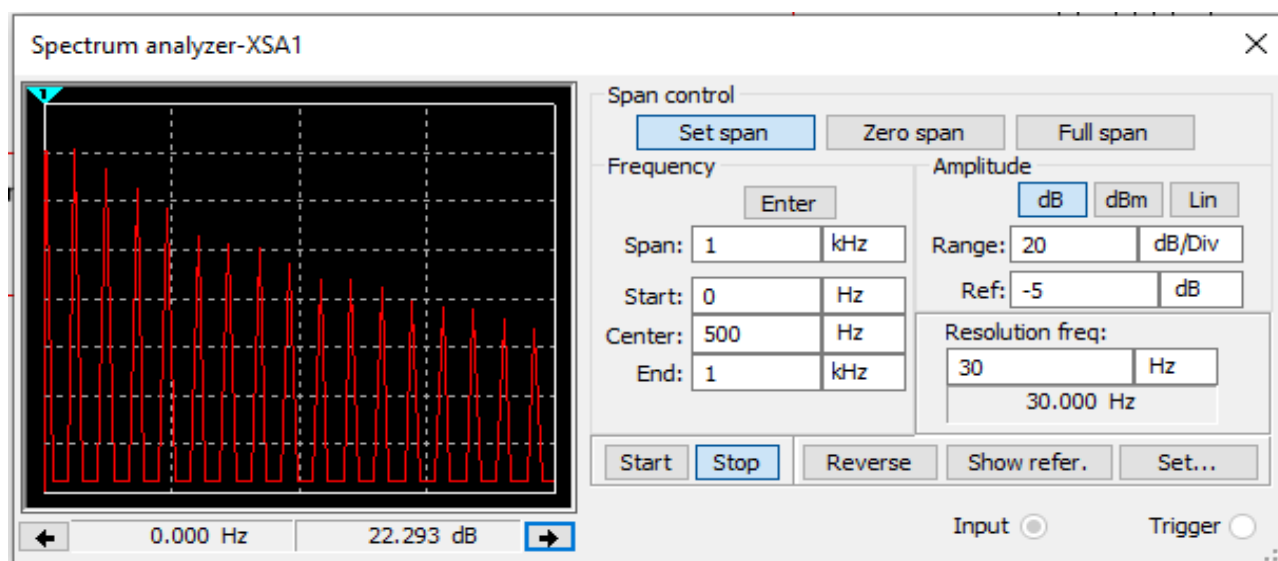
## Results:



**Figure 5:**

This is an oscilloscope that was simulated on Multisim and shows the input and output on Figure 3. The yellow line shows the voltage after the transformer when it's 12 volts AC. The line is a sine wave and is fluctuating up and down, not keeping the same

value so this wouldn't be very good to use if we wanted a constant value. The blue line is the output after the low pass filter and you can see it isn't perfect, but it is very close to the same value and barely changing at all. This is something we can use to keep constant power going without any worries.



**Figure 6:**

This is a Spectrum analyzer that was simulated in Multisim to show the decibels at the certain frequencies. This allows us to measure the power of the spectrum of known and unknown signals. I looked at 120 Hertz and saw it was 20 decibels, so I knew if I went to 1.2KHz it should be around 40dB. When I looked at 1.2KHz it wasn't around 40dB but was at 60dB and I couldn't understand why. Come to find out the reason why it was 60dB and not 40dB was because of the low pass filter at the end of the circuit, but that just proves that it was doing the right thing and the values were correct.

### **Future Work:**

We are currently using simulation software to get our data and conduct research. Hopefully, after the summer internship, I hope to be working in person, get hands-on experience and get used to the systems used in superconducting magnet testing. With this experience, I hope to get a job and implement everything I have learned.

**Conclusion:**

In conclusion, I was able to build and simulate a basic power supply with a basic voltage regulator, but wasn't able to finish the feedback control loop. I did research into Feedback control loops, but was not able to simulate it, but learned more in depth about zener diodes, cap capacitors and low pass filters. The results for the oscilloscope were simulated in multisim and prove that the low pass filter is working and the voltage is constant and barely changes. The results for the spectrum analyzer were also simulated in multisim and showed the power throughout the circuit. This research with the feedback control loop is important because we are able to see what values are coming through the output and if it isn't the value we want, then we can save and protect the rest of the circuit.

**Acknowledgments:**

I would like to express our appreciation to my mentors, Piyush Joshi and Suresh Deonarine, for the time and effort they put into making our internship a valuable and rewarding experience for me. Suresh was always readily available, knowledgeable, kept communication open, and I am extremely grateful. I would also like to express my appreciation for Amy Engel and Kenneth White, for always being readily available and setting up every meeting and interview needed for this project. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program (CCI).

**References:**

Joshi, P. (n.d.). *LTS-HTS Hybrid Dipole Magnet Quench Protection System*. Brookhaven Laboratory.

*LHC Facts and Figures*. (2021). Retrieved March 8, 2021, from CERN Accelerating science: [https://home.cern/sites/home.web.cern.ch/files/2018-07/CERN-Brochure-2017-002-Eng\\_0.pdf](https://home.cern/sites/home.web.cern.ch/files/2018-07/CERN-Brochure-2017-002-Eng_0.pdf)

*LHC Facts and Figures*. (2021). Retrieved March 8, 2021, from CERN Accelerating science: [https://home.cern/sites/home.web.cern.ch/files/2018-07/factsandfigures-en\\_0.pdf](https://home.cern/sites/home.web.cern.ch/files/2018-07/factsandfigures-en_0.pdf)

Velavan, R., & Myvizhi, P. (2018). Superconductors and its Applications. *International Journal of Pure and Applied Mathematics*, 119(12), 7377-7386.

[Magnet-Division-Cryogenic-Facility\[2\].docx](#) Piyush Joshi, Ph.D. Kathleen Amm, Ph.D. -BNL

LTS-HTS Hybrid Dipole Magnet Quench Protection System. Piyush Joshi, Ph.D.- BNL

## Appendix:

Participants:

Name	Institution	Role
Piyush Joshi	Brookhaven National Laboratory	Mentor, project lead.
Suresh Deonarane	Brookhaven National Laboratory	Mentor, Provided leadership
Peter Maritato	Suffolk Community College	College Faculty, provided qualitative information
Richard Biscardi	Suffolk Community College	College Faculty provided qualitative information
Amy Engel	Brookhaven National Laboratory	Faculty participant provided qualitative information
Kenneth White	Brookhaven National Laboratory	Faculty participant provided qualitative information
Noel Blackburn	Brookhaven National Laboratory	Faculty participant provided qualitative information

Design analysis and engineering improvements for a next-generation X-ray powder diffraction sample furnace

Edwin Cardenas, Department of Mechanical Engineering, NYIT, New York, NY 10023

Amanda Sirna, Department of Mechanical Engineering, Stony Brook University, Stony Brook, NY 11794

Edwin Haas, National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY 11973

## **Abstract**

A compact sample furnace was designed to heat samples to temperatures of 2000 - 2300°C at the X-ray Powder Diffraction (XPD) beamline at NSLS-II. This furnace allows the XPD photon beam to pass through with adequate downstream opening to collect diffraction data for high-temperature materials research. Since the XPD samples did not reach the desired temperatures several areas of improvement were proposed and researched to improve the heat transfer efficiency to the sample. These approaches included the use of inert gases, high temperature coatings, infrared lenses, spherical reflectors, CO2 lasers and a redesign of the sample holder. Engineering studies and simulations were done to decide which of the above outlined paths would produce the best results at the lowest cost. The three final paths chosen to experiment with were high temperature coatings, inert gases, and a redesign of the sample holder. These three paths proved to increase radiation absorption, heat transfer efficiency, and absorptivity of samples inside the furnace. The work I did for this hexapole light furnace has helped me improve my heat transfer knowledge and capability and better understand all of the details that go into a project such as this.



## Introduction

The XPD Beamline at NSLS-II does materials research at temperatures ranging from cryogenic to  $\sim 1700^{\circ}\text{C}$ , and efforts to conduct materials research at higher temperatures are ongoing. The present XPD sample furnace uses infrared lamps with gold-plated reflectors to focus infrared rays at a single sample. This furnace has been operating in air, and high-temperature oxidation has been destroying thermocouples. A plan therefore was developed to investigate the following potential areas of improvement:

- a. The use of IR lenses to focus forward-directed IR rays otherwise lost as halo.
- b. Improved sample tube holders.
- c. The use of inert gas shielding to prevent oxidation.
- d. The use of high-temperature coatings.
- e. The use of a custom-designed secondary reflector to redirect forward-directed IR halo.
- f. The use of a CO<sub>2</sub> laser as a supplementary (or primary) sample heat source
- g. Optimization of heat lamps and focal distance
- h. A comparison of actual lamp reflector geometry to ideal geometry

This paper briefly discusses each of the above options and explains why some of the above items were pursued further and others rejected. Considerations of material properties at high temperatures are also included.

## Sample and Furnace Conditions

Powder samples contained in sapphire tubes are held by ceramic holders for X-Ray Powder Diffraction research in air at atmospheric pressure. Sapphire is used for its' high-temperature and optical properties. A computer graphic (CG) image and photo of a sapphire tube in its' ceramic tube holder is shown in Fig. 1 along with an image of one infrared heat flux pattern. This heat flux for a single ellipsoidal reflector indicates loss of some flux past the sample holder tube.

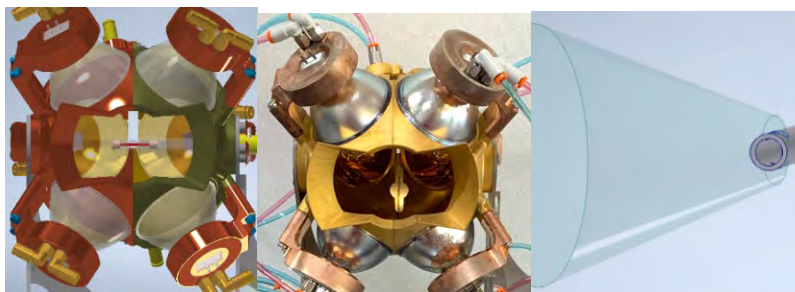


Figure 1: XPD Sample Furnace looking upstream: (a) CG image (b) photo (c) heat flux pattern

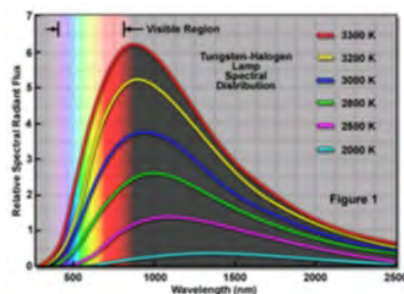


Figure 2: IR lamp flux spectrum

Infrared lamps direct 150W each of heat flux at a ceramic sample tube holder in the bandwidth shown in Fig. 2. Six infrared lamps each have ellipsoidal reflectors and water-cooled jackets to minimize external temperatures for safety and to extend lamp life. Cooling water at 17-20°C from

a chiller is circulated in parallel paths through all lamp cooling jackets and through two water-cooled outer shells.

### Engineering Investigations

- a. The use of plano-convex infrared lenses as a primary means to focus forward-directed rays was considered, but not pursued as available infrared lens geometries and IR source distance limitations meant only a small portion of forward-directed rays were properly focused. When IR lenses were used to supplement ellipsoidal reflectors, many reflected rays that would have been focused on the sample center were misdirected. IR lenses therefore were ruled out as a viable solution. Two IR lens configurations are shown in Fig. 3.

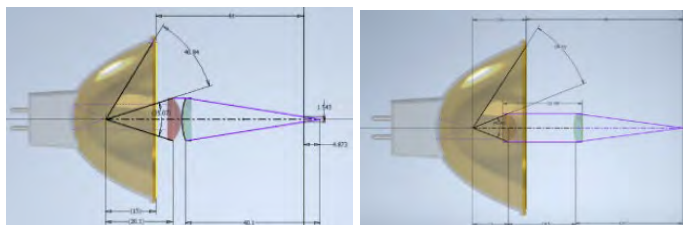


Figure 3: IR lens designs; focal point (a) behind sample, and (b) at sample center

- b. Iterative design and thermal analyses were used to optimize the design of the sample tube holder. The external surface area was increased to capture more infrared rays while eliminating an internal cavity. Thus, convective cooling was reduced, and the conduction path to the sample itself shortened to transfer more heat to the sample by direct conduction to the sapphire tube. Figure 4 below shows the improved sample tube holder. Its features include a small hole in the front to allow the XPD beamline in, and a  $\pm 45^\circ$  downstream opening to allow X-ray diffraction data to be collected.

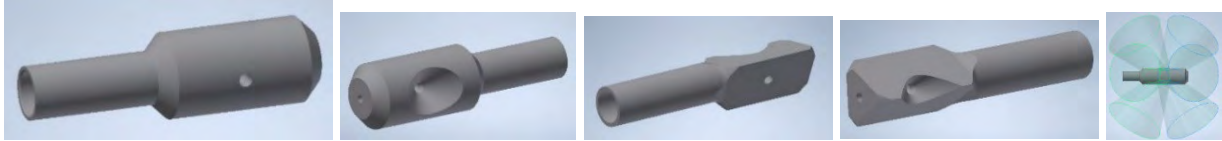


Figure 4: New sample holder designs and heat flux pattern impinging on a redesigned sample holder

c. Inert gas shielding - Oxidation within the furnace increases non-linearly as the temperature inside the furnace increases. This may be mitigated by introducing an inert gas to shield materials from oxidation. Three inert gases and their specific heat capacities are shown in Table 1. Using the heat capacities tabulated and the equation written below, it was determined that the gas with the lowest specific heat would absorb the least amount of heat. By absorbing the least amount of heat, the majority of the heat trays would be left to be absorbed by the sample. Based on gas specific heat capacity and cost effectiveness, Argon is the best inert gas for this application as it will carry away the least heat via convection. This theory is proved in Table 5 where various gases were tabulated with their highest temperature achieved.

$$Q=mc\Delta T \text{ (1)}$$

Table 1: Specific Heat Capacities for Helium Argon, and Nitrogen Gases [1]

Gas	Specific Heat Capacity (KJ/kg K)
Helium	5.1926
Argon	0.5203
Nitrogen	1.039
Air	1.005

Two high-temperature coatings were tested to improve heat transfer to high heat-flux areas of the sample holder and insulate areas outside the heat flux area. Important properties include coating adherence and high-temperature resistance in air. YSZ coating from Zypcoatings is a zirconia-

based coating that can reach temperatures  $>2000^{\circ}\text{C}$  without oxidizing per the manufacturer's specifications. This coating reduced the maximum internal temperature in initial tests and is being considered as an insulator outside the sample holder high-heat flux area. In other tests, a silicon carbide high-temperature coating significantly increased heat transfer to increase heat absorption are ongoing. Thermal properties of three sample holder materials are presented in Table 2.

Table 2: Thermal Properties of Tubing Materials [2-4]

Properties	Alumina	Zirconia	Sapphire	Silicon Carbide
Thermal Conductivity (W/m K)	2.2-4.3	14-30	34.6	60-120
Thermal Expansion ( $\mu\text{m}/\text{m K}$ )	8.8-9.5	6.7-8.2	4.5	4.0
Max. Temp ( $^{\circ}\text{C}$ )	1330	1730	2053	1650
Specific Heat Capacity (J/KgK)	400-510	870-940	750	750

d. The use of a custom designed secondary reflector was investigated to recover some of the forward directed IR rays otherwise lost as halo. The shape and overall design of the reflector is shown in the Fig. 5. This shape and location were determined by radiation heat transfer view-factor calculations. The reflector needs the lowest view factor possible to reduce the number of rays blocked. Figure 5 (a) represents the View Factor Geometry. Spherical reflectors were not pursued further due to decreased lamp life and loss of flux reflected by the ellipsoidal reflector.

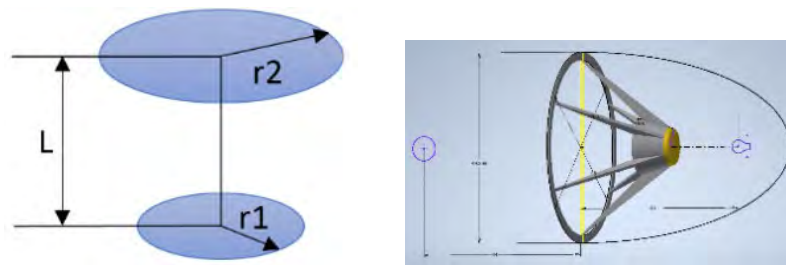


Figure 5: (a) Reflector View Factor Geometry and (b) Proposed Spherical Reflector Design

- e. Small cross-sectional beam (i.e.  $\leq 1.5$  mm diameter) CO<sub>2</sub> lasers operating at a wavelength between 9.5 and 10.5 $\mu$ m can apply IR heat energy directly to the sample at a wavelength where the heat energy will be absorbed more efficiently by the sapphire sample tube. Due to a limited budget, this area was not pursued yet, but future laser power versus sample temperature tests are being considered.
- f. IR heat lamps tests indicated as much as 28°C difference when moving a single heat lamp 1 mm away from its' maximum temperature point. Further tests using shims to optimize the focus of heat lamps are ongoing.
- g. The heat lamp selection and positioning were optimized to ensure the maximum amount of radiation produced by the lamps was absorbed by the sample. Tests were done to determine optimum working distance between lamp and sample as well as confirm the shape of the reflectors used.
- h. The ideal shape for a reflector is half an ellipsoid; that is, a full ellipsoid bisected equally at a point equidistance from the two foci that form it. An ellipsoid is formed by tracing a point on a plane at a constant combined distance from two focal points (called foci) and then revolving this traced shape about the axis between the two foci. This shape will reflect all the rays emanating from one focal point and focus them at the other. In this application, the sample center should be located at one focal point and the infrared bulb filament should be centered at the other focal point. A reflector having any other shape will not focus all the rays from one point onto another.

## Results

### *Material Analyses:*

The following two graphs were produced by focusing an ILT (model L6409G) lamp at ¼" diameter Alumina and Zirconia ceramic tubes with, and without silicon carbide and type YSZ high-temperature coatings. In all tests, lamp voltage, tube position, and thermocouple position were held constant, and the same size tube was used. The heating and cooling curves in Figures 6 and 7 show the results of the tube materials and the effects of the coatings. The highest temperature was reached by a silicon carbide coated zirconium tube, and the YSZ coating showed some value as an insulator.

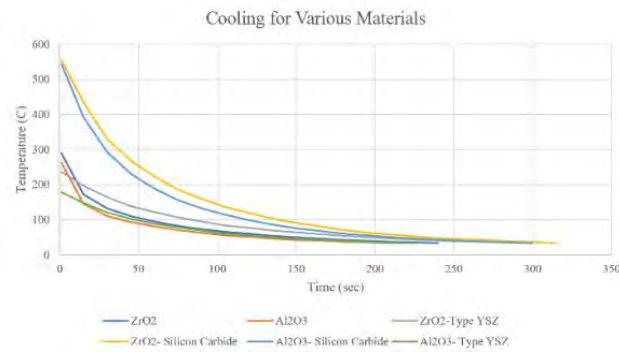


Figure 6: Cooling Rate for Selected Ceramic Materials and Coatings

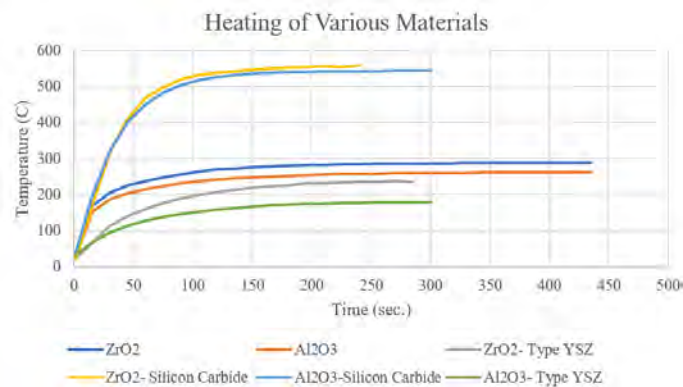


Figure 7: Heating Rate for Selected Ceramic Materials and Coatings

Based on Figures 6 and 7, we can derive heating and cooling curves for each material and coating type. The heating and cooling curves take the form of the equation shown below with the equation constants tabulated in Tables 3 and 4. The tables confirm that the material with the best heat absorption is Zirconium and the coating that improves absorption the best is Silicon Carbide. The Type YSZ coating proves to make heat absorption difficult. However, this means the Type YSZ can be used as an insulator as it has the slowest cooling curve. Therefore, the improved design ceramic tubing should be made of Zirconium with the area of maximum heat flux be coated in Silicon Carbide.

$$y = a * \ln(x) + b \quad (2)$$

Table 3: Heating Curves for the Materials

<b>Material</b>	<b>a</b>	<b>b</b>
ZrO3	40.551	60.129
Al2O3	36.967	52.822
ZrO3 Type YSZ	44.37	-11.637
ZrO3 Silicon Carbide	110.16	-15.332
Al2O3 Type YSZ	29.802	11.723
Al2O3 Silicon Carbide	98.644	19.151

Table 4: Cooling Curves for the Materials

<b>Material</b>	<b>a</b>	<b>b</b>
ZrO3	-48.01	292.69
Al2O3	-43.26	260.28
ZrO3 Type YSZ	-41.66	276.98
ZrO3 Silicon Carbide	-104.6	630.9
Al2O3 Type YSZ	-30.4	204.43
Al2O3 Silicon Carbide	-99.8	591.15



### *Inert Gas Analyses:*

The effects of various gases were also explored and tabulated in Table 5. It was found that the use of an inert gas can improve the efficiency of radiation transfer from lamp to inside ceramic tubing while also shielding everything inside the furnace from oxidation. The inert gases with a smaller specific heat capacity absorb less of the radiative heat and leave more of it to be absorbed by sample. The specific heats and maximum temperatures reached confirm the theory discussed in the Engineering investigations section of this paper. The tests proved that Argon gas achieved the highest temperature.

Table 5: Single lamp Max Temps for Noble Gases

<b>Gas</b>	<b>Specific Heat Capacity (KJ/kg K)</b>	<b>Max Temp Achieved (°C)</b>
Helium	5.1926	185
Argon	0.5203	258
Air	1.039	242
Nitrogen	1.005	243

### *Heat Lamp and Focal Distance Analyses:*

The reflectors used in the HLF are from Osram and ILT and should be elliptical in shape. This elliptical shape is crucial to directing rays towards the sample. If these reflectors are not elliptical in shape, this expected phenomenon will not be as efficient, and a portion of the produced radiation would be lost to the environment. Because of this, studies were done to confirm the elliptical nature of the reflectors. Using a coordinate measurer, points along the inner curvature of the ellipse were taken and plotted in Figure 8. In addition to the data plotted, a circle and ellipse were plotted to determine which shape best represents the data collected.

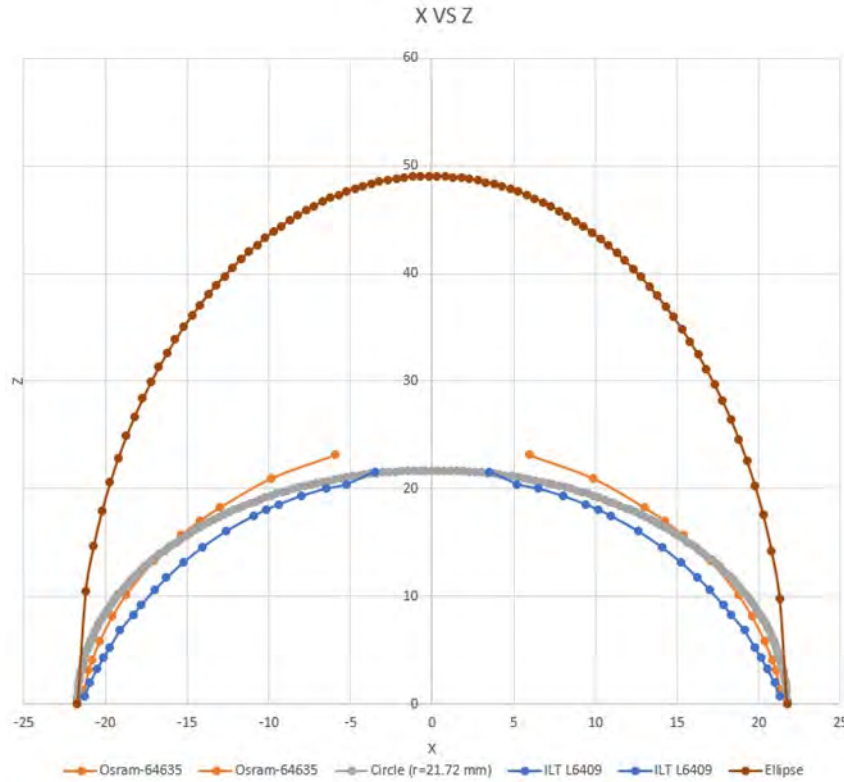


Figure 8: Plot of recorded point data for ILT and Osram reflectors

Based on the figure above, it appears both the Osram and the ILT reflectors do not have the desired ellipsoid shape. This is one reason for the loss of radiation from source to sample. A way to improve this would be to make or use a commercially available reflector with a more elliptical shape, if such a reflector exists. TO date a reflector that meets our needs has not been found.

## Conclusions

Materials research at temperatures above 1750°C in air requires careful sample holder material selection and an understanding of material properties at high temperatures. Refractory metals undergo oxidation, and high-temperature ceramics are very difficult to machine. Additionally, thermal absorptivity and emissivity are wavelength dependent when using infrared heat sources. Some materials (e.g. sapphire) are mostly transparent to near infrared rays, and many ceramics are poor heat conductors. Only a limited number of materials were found that can perform acceptably in air at temperatures above 2000°C. Cost is also a consideration. Due in part to cost and time considerations, the solutions chosen to improve the performance of the XPD Infrared Sample Furnace include the selective use of infrared lamp position adjustments and an improved sample holder. The new sample holder maximizes heat absorption, provides a short heat conduction path to samples, and optimizes external size to reduce convective heat transfer losses. The use of high-temperature coatings is promising, but coating adherence and maximum coating temperature need further improvement. The authors were not able to improve performance with infrared lenses, and secondary reflectors will diminish lamp life with insufficient thermal gain. Lastly, CO<sub>2</sub> lasers (~10μm wavelength) hold significant promise as a primary or secondary heat source, but tests were not undertaken in time for publication of this paper.

## **Acknowledgements**

The authors thank Jianming Bai, PhD (XPD scientist and project leader) and Sanjit Ghose, PhD (XPD Lead Beamline Scientist) for project guidance, Hui Zhong (SUNY SB liaison engineer) for thermal data, G. Lawrence ('Larry') Carr, PhD (FIS/MET Lead Beamline Scientist) for infrared guidance, Leo Reffi for design drawings, and John Trunk (HXSS program chief technician) for thermal experiment support. Also, thanks go to Eric Dooryhee, PhD (HXSS program manager), and Lisa Miller, PhD (NSLS-II educational program coordinator). This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## References

- [1] Bhattacharjee, S. *Properties of Various Ideal Gases (at 300 K)*,  
[www.ohio.edu/mechanical/thermo/property\\_tables/gas/idealGas.html](http://www.ohio.edu/mechanical/thermo/property_tables/gas/idealGas.html).
- [2] “Oxide-Based Engineering Ceramic.” *MakeItFrom.com*, 2020,  
[www.makeitfrom.com/compare/Alumina-Aluminum-Oxide-Al2O3/Zirconia-Zirconium-Dioxide-ZrO2](http://www.makeitfrom.com/compare/Alumina-Aluminum-Oxide-Al2O3/Zirconia-Zirconium-Dioxide-ZrO2).
- [3] “Properties: Supplier Data - Sapphire Single Crystal (Alumina 99.9%) - (Goodfellow).”  
*AZoM.com*, 9 July 2021, [www.azom.com/properties.aspx?ArticleID=1721](http://www.azom.com/properties.aspx?ArticleID=1721).
- [4] “Silicon Carbide, SiC Ceramic Properties.” *Silicon Carbide SiC Material Properties*,  
[accuratus.com/silicar.html](http://accuratus.com/silicar.html).

Super Pioneering High Energy Nuclear Interaction eXperiment infrastructure interlocks & monitoring systems  
Michael Charumaneeroj, Stony Brook University, Deer Park NY 11729  
Joel Vasquez, Physics Department, Brookhaven National Laboratory, Upton, NY 11973

## Abstract

The super Pioneering High Energy Nuclear Interaction eXperiment (sPHENIX) will be an upgrade to the PHENIX detector, located along the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory. This detector, expected to be completed in 2023, will collect information about Quark Gluon Plasma. This information will help us understand more about the initial conditions of the universe, directly following The Big Bang. The sPHENIX detector will reside within Building 1008, the sPHENIX Experiment Hall. The previous detector was monitored through a hardware implementation using physical racks within the control room. These racks displayed pertinent information to operators through indicators, pushbuttons and window boxes.

The new sPHENIX facility will be monitored via software on a Human-Machine Interface (HMI) display and control logic will be implemented through a Programmable Logic Controller (PLC), thus eliminating many of the current racks. A set of HMI Displays has been created that will allow operators view the status of the area, monitor different alarms and perform various actions, such as acknowledging alarms, tripping breakers, and setting bypass signals. HMI displays are created by creating polygons, text boxes and indicators that will all interact with the logic from the PLC, which acts as the brains for the displays. Creation of HMI Displays requires great attention to detail, as well as a good sense of industry standards, for example, color codes and operator tendencies.

## **I. Introduction**

Perhaps the most important thing in a facility besides the data collection is the interlocking and monitoring system. In the past, the facility was monitored via the PHENIX Control Room (PCR) Racks. These racks displayed facility monitoring information, which was implemented through relays. While they are good at displaying information for operators, they take up a lot of physical space. Also, there is only so much that can fit on each rack before technicians need to bring in a new one. A proposed solution to these issues is a shift from hardware monitoring to software monitoring, where the PCR Racks are condensed into two main HMI Displays. These displays would show information sectioned off by location, system and subsystem for diagnostic information, and an alarm history page for operator awareness and bookkeeping of alarms.

## **II. The Project**

### **A. HMI Display**

The HMI Design portion of the project was the first priority of the summer. An HMI program is a set of screens (commonly referred to as “displays”) that can be programmed to display custom information, normally on a tablet-like screen or desktop. Each of the HMI displays were composed of combinations of polygons, text boxes and built in features such as push buttons and indicators. The task was to design these displays for the stakeholders who would approve or deny the proposed software implementation of the monitoring system. Knowledge of industry standards and operator tendencies is imperative to create a good display and indicator. For example, grey backgrounds are normal in displays, as they are not flashy and will not take the operator’s eyes away from the important information on the screen. For the



indicators, a green indicator symbolizes a “good” state, where nothing is wrong. Red, yellow and orange indicators often show “bad” states, where something is occurring that requires attention.

Another important aspect in designing these displays is an eye for detail. Since these displays begin as blank white screens, the designer must create the display in a way that is pleasing to the eye, and effective in portraying the correct information to operators. In this project, the approach was to create displays with columns of panels, categorized by subsystem within each location (Figure 1).

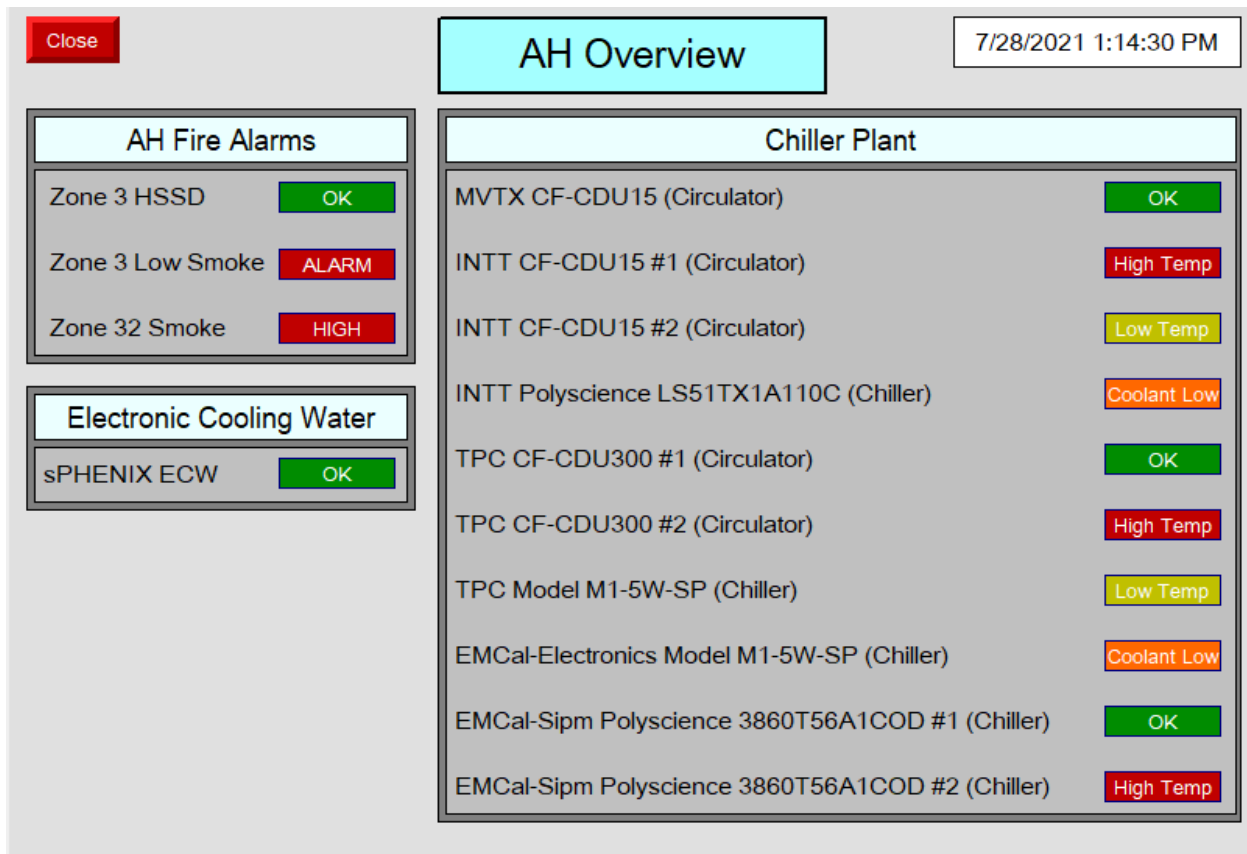


Figure 1 shows an example of an HMI Display. Panels are divided by subsystem.

In order to make each panel appear uniform, the sizing of each polygon is calculated dependent on the number of alarms that need to be monitored per subsystem; each panel requires 2 calculations, one for the light grey polygon and one for the dark grey polygon. The formula for the size of the light grey polygon is  $36 + 42 * (N - 1)$  and the formula for the size of the dark

grey polygon is  $38 + 42 * N$ , where N is the number of alarms in each subsystem. Using these formulas I was able to calculate the size of each panel that would leave a boarder of 5 pixels on every side from the dark grey polygon and the light blue / light grey polygons. This size also allowed for proper spacing between each of the indicators and the text of each alarm.

The status indicators for each subsystem panel adjust according to the state they are in, which is determined by ladder logic. Each indicator has at least two generic states: OK and Alarm/Trouble. Depending on the specific subsystem and alarm, the states can change to something more specific, like “High Temp” or “Smoke”. Indicators can also have more than one state if necessary; the implementation of these “multistate indicators” allows for less space to be used on each panel. For example, the Electronic Cooling Water panel on the left of Figure 1 is currently designed to have multiple states in each indicator. Likewise, the sPHENIX indicator has multiple states; OK, High Temp, Low Flow and Makeup Flow to name a few. Figure 2 below shows an example how a state is programmed on the HMI side.

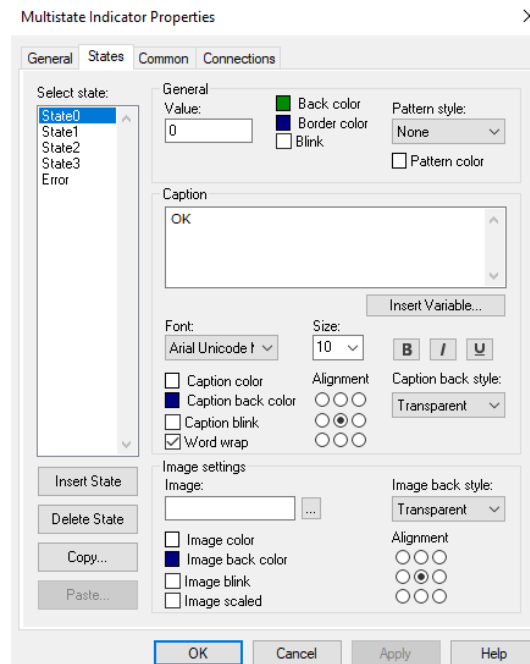


Figure 2 shows an example of a state in a multistate indicator.

## B. Ladder Logic

After the meeting with our shareholders was held and the project concept was approved, the next step was to begin programming the PLC. While the HMI Display is what an operator would be able to interact with, the PLC is what would act as the brains for the displays. The program is “written” using Ladder Logic. Ladder Logic is different from text based programming languages. A user will create “rungs” that connect metaphorical power to a metaphorical ground in a graphical interface. Programmers design these rungs with different conditions that will allow the “power” to flow from one side to the next.

The main logic is comprised of two contacts. The first is a Normally Open (NO) contact, and the second is a Normally Closed (NC) contact. For a NO contact, if the condition is true, then the contact will close and power will be able to flow in the rung. A NC contact is the opposite, where if a condition is true then the contact will open and power will stop flowing in a rung. Another important concept of ladder logic is a “latch”. A latch is when an output signal is used as a second NO input to the rung. This allows the code to “store memory” of outputs. When a signal is latched, it’s important to include a “clear” signal. This clear signal is a NC contact that allows the latched rung to become de-energized, no matter what the inputs are. These three concepts come together to create some of the main logic rungs for many of the system’s alarms, as seen in Figure 3.

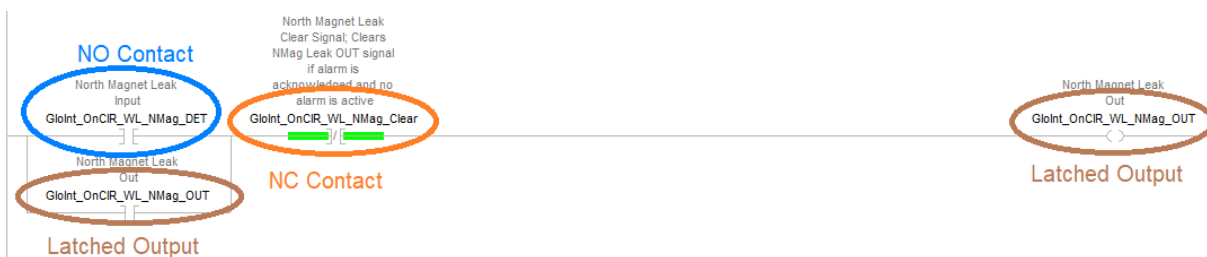


Figure 3 shows a logic rung. Depicted are NO and NC contacts, as well as a latched output.

In this example, we can see all of the above explained concepts; the “GloInt\_OnCIR\_WL\_NMag\_DET” tag (variables that handle signals can be referred to as “tags” in Ladder Logic) is our NO Contact. If there is a leak detected, then this tag will have the value of “1” and the contact will close, allowing power to flow. The next contact we see to the right of this is our NC Contact, “GloInt\_OnCIR\_WL\_NMag\_Clear”. If this tag receives the value of “1”, then the contact will open and power will not be able to flow. When this happens, the rung will be de-energized, regardless of the states of the other signals. The final contact we see is the output on the right of the rung, which is also a NO contact below the first signal discussed. This is the latch that was discussed previously; once the rung is energized initially, the output will be stored and the rung will be energized until the clear signal is active.

### III. Product Operation

The two programs work together to create the Global Interlocks monitoring system, which protects experimental equipment from facility and utility failures or alarms. Using a truth table (excerpt seen in Figure 4) the PLC detects alarms that could damage the detector or supporting machinery and sends a trip signal to de-energize the corresponding area or equipment.

Output Element Response	Experimental Power, Utility Power, & Gas House Shunted Power (DP 2,5,8 & Gas house shunt trip)	Solenoid magnet power supply (Relay)	Gas mixing house supply valves (Relay KGT)	Gas house shunt trip/shunt trip panel (Relay CR7C Coil)	Counting house breaker (Relay CR8C Coil)	sPHENIX Rack Room HVAC Units
19 Smoke Inside North / South Solenoid (VESDA)	On	On	On	On	On	On
20 Water Leak Detection Bottom of the North/South Solenoid	On	On	On	On	On	On
21 Electronic Cooling Water high temperature	On	On	On	On	On	On
22 Electronic Cooling Water Low Flow	On	On	On	On	On	On

Figure 4 shows an excerpt of the Global Interlocks Truth Table

We can understand the operation of this project through an example scenario. A technician would sit in the Control Room that overlooks the sPHENIX facility where they are able to see the two screens depicted in Figures 5 and 6.

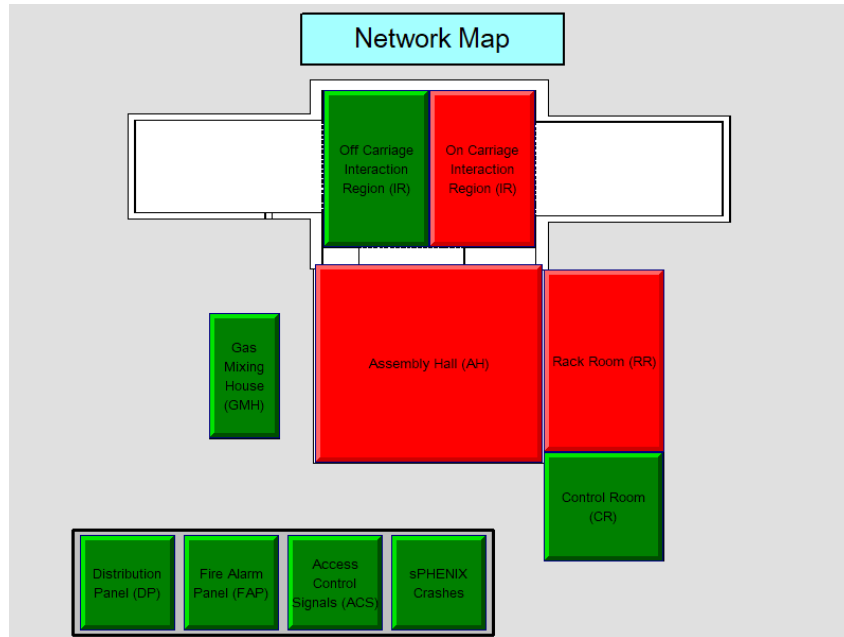


Figure 5 shows the proposed left HMI screen, depicting a network map.

Alarm Status		
Alarm time	Acknowledge time	Message
! 6/28/2021 12:30:37 PM		On Carriage IR S Mag Leak
! 6/28/2021 11:42:11 AM		RR High Temp
! A 6/28/2021 11:40:52 AM	6/28/2021 11:45:44 AM	AH Low Coolant Level

Ack Alarm	▲	▲	Go To Alarm History	Close
Ack All	▼	▼		

Figure 6 shows the proposed right HMI screen, displaying the Alarm Status. Example display shows a few different states of alarms.

On the left screen (Figure 5), there would be a map showing an overview of the system, complete with indicators that show a location's status; each location is navigable via pushbutton to view the location's subsystems. The indicators reflect the subsystem statuses within each

location. On the right screen (Figure 6), an Alarm Status page where the operator can see a history of currently active alarms. The page includes important information for operators such as alarm active time, acknowledgement time and a description of each alarm.

It's seen on Figure 5 that there is trouble in the On Carriage Interaction Region. This is mirrored on the Alarm Status page (the selected alarm). The operator would navigate to the On Carriage Interaction Region on the screen and see the screen depicted on Figure 7. They would see that there is a leak at the South Magnet, indicated on the top right panel.

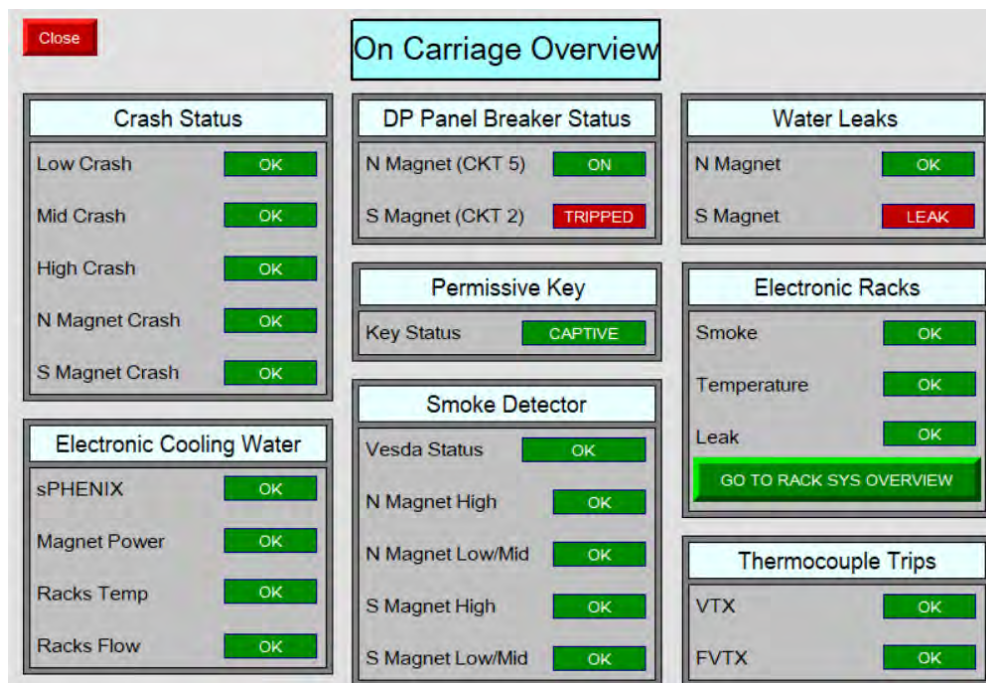


Figure 7 shows the On Carriage Interaction Region Overview screen under the example from figure 5 and 6 where there is a leak on the South Magnet.

Since a leak in the South Magnet of the Interaction Region can cause damage to the detector and surrounding equipment, a global interlock is asserted, as described in Figure 4. The detected alarm will trigger the corresponding Distribution Panel Circuit Breaker (CKT 2) to trip power via the PLC. The process by which this is done is shown in Figure 8 and described below.

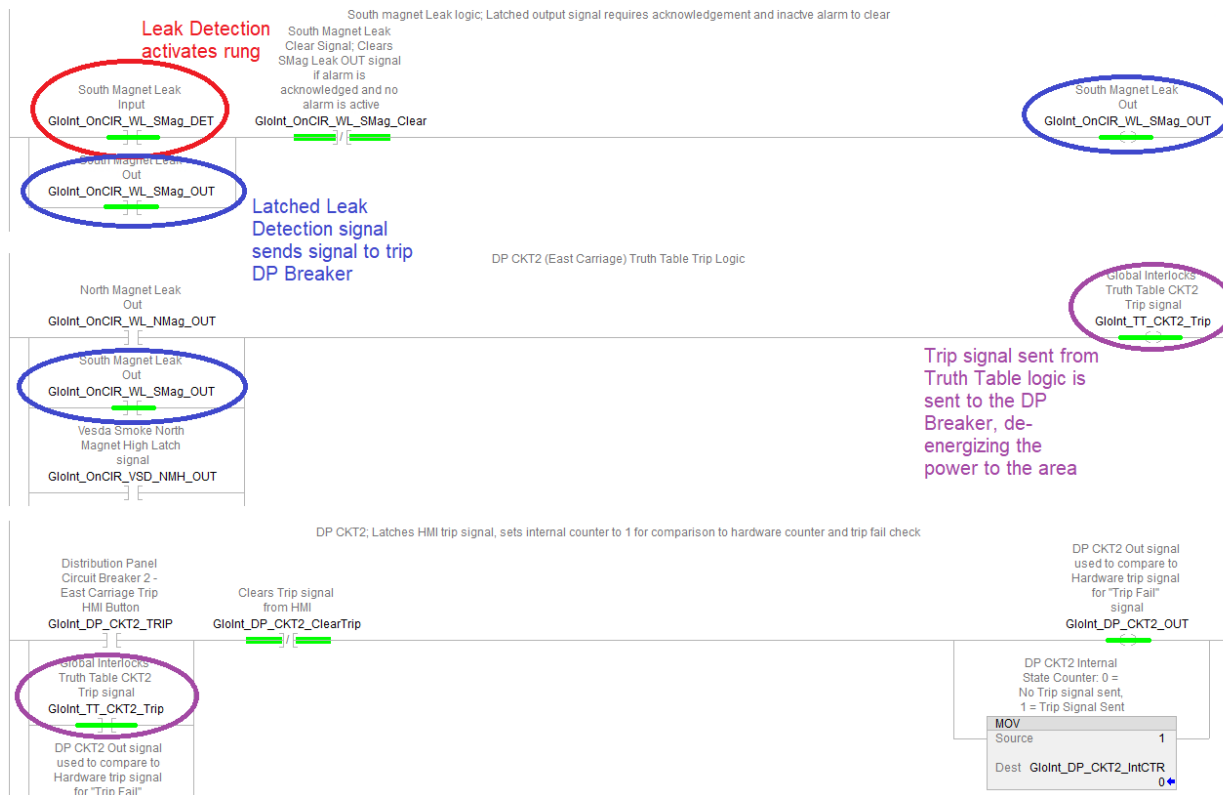


Figure 8 shows the process by which a Global Interlock is activated through the PLC’s logic

Once an alarm is detected, the corresponding “DET” tag becomes active. This enables the corresponding “OUT” tag. OUT signals defined by the truth table are monitored in parallel so that if one becomes active, the PLC will know to send the corresponding trip signal. The HMI can then display the correct state of the breaker, depicted in the top panel of Figure 7.

With the indicators and alarms shown, the operator would then be able to address the situation according to the operator’s manual.

#### IV. Conclusion

The product of this summer’s internship is a foundation for the replacement of hardware based monitoring racks in the control room. The proposed solution will use HMI displays that work with Ladder Logic from a PLC in order to allow operators to monitor the sPHENIX facility with ease. As this internship was virtual, it was important that there was a way for me to access

the equipment remotely. With the help of Martin Purschke, I was able to remote in and access the necessary programs for the summer. Without his help, the foundation for this project would not have been as clear as it is now for future designers to work on.

Further developments to this project include linkage to a database for convenient on screen operator information, as well as greater development of sPHENIX related alarms. This project has the ability to be changed according to requests from different operators and technicians.

## **V. Acknowledgements**

I would like to extend my gratitude to my mentor Joel Vasquez for his help and guidance throughout the summer. I would also like to acknowledge Lee Flader, who I had the pleasure of working with this summer as I developed the foundation for the HMI Displays and Ladder Logic. I would also like to thank Martin Purschke, the sPHENIX team and the SULI team for their additional help and support throughout the summer.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).



# Two-photon interferometry simulations

Zhi Chen, Physics, Stony Brook University, Stony Brook, NY 11794

Andrei Nomerotski, Physics, Brookhaven National Laboratory, Upton, NY 11973

Summer 2021

## Abstract

Two-photon interferometry is an innovative quantum astrometry technique based on quantum mechanics that gives more precise measurements compared with classical one-photon interferometry. Two photons from independent light sources are interfered by using a beam splitter and thus become quantum entangled. In theory, measurements in the number of photons in different channels of the telescopes give information about the relative separation between the two light sources. The two-photon interferometry has many potential applications such as direct imaging of black hole accretion discs, mapping microlensing events, directly resolving the star-planet binary systems, etc. My project involves studying the response of the photon coincidence rate due to Earth's rotation for arbitrary telescope placements and source positions in the sky by developing a simulator. I worked with the bright star catalog to select out target pairs of stars based on their properties such as position and brightness, then the selected star pairs become the input data for the simulator to analyze. Lastly, the simulator is used to generate simulated data which can be used for estimating the precision between the theoretical results and the real data in the future.

## 1 Introduction

For thousands of years, mankind has held curiosities toward the distant stars in the night sky. Four hundred years ago, Galileo was able to discover four moons of Jupiter using the simple telescope that he made. Telescopes have been improved and modified since then so that we can see further and clearer images of the sky. Large-size telescopes are capable of capturing sharper and detailed images of astronomical events compared to small-size telescopes since the angular resolution of images is proportional to the size of the telescopes. There is a clear limitation on the angular resolution of the telescope because the size of the telescope cannot be infinitely large due to space limitations, cost, and engineering issues. However, scientists discovered that interfering with the data from multiple telescopes to ob-

serve the same light source can mimic the result of a large size telescope. The current theory behind modern astronomical interferometry is based on classical physics, which interferes with the results of many telescopes for a single light source.

Due to the rise of quantum mechanics, astronomers are seeking methods of producing more precise measurements of the sky by using quantum mechanics. Two-photon interferometry is one of the ideas that use the idea of quantum entanglement between two light sources to do precise dynamic astrometry without having an optical path for interferometers. Gottesman suggested [1] that if the entangled Bell states between the two telescope stations can be provided and the photons from two light sources have interfered locally, then the phase difference between the arrival times of two photons is related to the separation distance between the two sources. Since the number of photons in each station carries the information of the separation distance between the two light sources, interferometry is established.

In this report, I give a short summary on the theoretical paper[2] which I used as the basis of my work. Then I will present my work on the development of the two-photon interferometry simulator. This simulator aims to produce the theoretical number of photons observed in the two telescope stations by given an arbitrary position of the telescopes and the two light sources. This simulator is capable of presenting a visual intuition on the data collection for real experiments in the future. We can also select the ideal star pairs that are suitable for future experiments based on their properties such as brightness, positions, etc.

## 2 Objectives & Scopes

There are four big steps for this summer project:

- Development of a code that would simulate the response of the star pairs for arbitrary star pair positions and telescope placements.
- Processing the bright star catalog so that it filters out most of the stars based on

some criteria such as star brightness, star pair separation distance, relative declination between stars and the telescopes, and the right ascension of stars for different time of year. The processed data can then be fed into the simulator for analysis.

- Analysis of the results including Fourier transform to see how oscillation frequency varies, the relationship between oscillation frequency and angular separation in different directions, etc.
- Generating simulated data which can be used for predicting the data collection of real experiment in the future.

## **3 Methods and Results**

### **3.1 Initial Development of Simulator**

#### **3.1.1 Background Information**

The main idea behind this simulator is based on a theoretical paper [2] that describes the application of two-photon interferometry. See [2] for a more detailed explanation. Figure 1 [2] shows the general telescope setup for observing a particular star pair in the sky. There should be two light sources in sky that are close together in opening angular angles. There are two telescopes, L and R, where each of them have two input channels for receiving the photons. The photons then go through the beam splitter to become interfered, so the chances for photons to land in the two receiving channels are purely probabilistic.

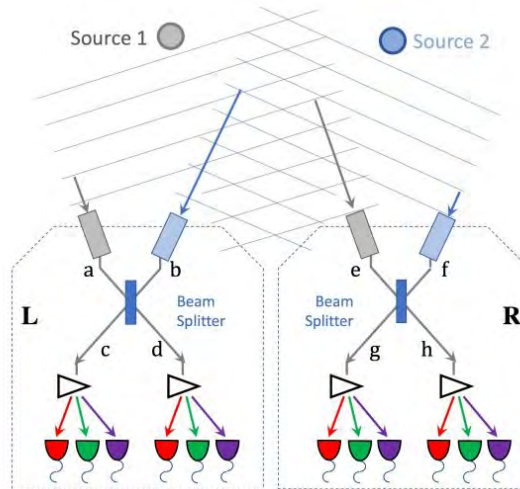


Figure 1: This figure [2] shows the general setup for the two telescopes and two sources in the sky. Since photons appear to travel in plane waves if the sources are far away, the arrival time for photons from a source is different for telescope L and R due to the path difference. Therefore, the path difference of photons traveling to the two telescope leads to phase shift of  $\delta_1$  for source 1 and  $\delta_2$  for source 2. Once the photon enters one of the input channels  $a$ ,  $b$ ,  $e$ , or  $f$ , photons are interfered using the beam splitter and thus become quantum entangled. If the two photons come from different sources, then the probability of observing the photon at different channels is related to the difference of the phase shift, or  $\delta_1 - \delta_2$ . Since the amount of phase shift depends on the relative position of the two sources on the sky, the relative position of the pair source can be deduced by observing the number of photons in different channels.

There are two possible initial states for the arrival of two photons. One of initial states of the two photons is when two photons are coming from the same source, and the probability of them arriving at different channels  $c$ ,  $d$ ,  $g$ , and  $h$  is described by Equation 1 [2]. When the two photons both origin from the same source, the probability of observing them in different channels pairs are all constant, which is not very interesting. The other initial state is when the two photons are coming from different sources, and the probability of the two photons arriving at different channels is described by Equation 2 [2]. Even though the probability of two photons entering the same channel is constant, the probability of two photons entering different channels depend the difference in the phase shift of the photon between the two telescopes. Moreover, the probability of receiving photons in channel  $ch$  and  $dg$  will be anticorrelated with photons in channel  $ch$  and  $dg$  since they have opposite signs in front of the cosine function. There is 0% for combination  $cd$  and  $gh$  due to Hong–Ou–Mandel

effect[2]. The overall dependence on  $\delta$  is interesting since it entails the relative position of the two light sources.

$$\begin{aligned}
 P_{11}(cc) &= P_{11}(dd) = P_{11}(gg) = P_{11}(hh) = 1/16 \\
 P_{11}(cd) &= P_{11}(cg) = P_{11}(ch) = P_{11}(dg) = P_{11}(dh) = P_{11}(gh) = 1/8
 \end{aligned}
 \tag{1}$$

$$\begin{aligned}
 P_{12}(cc) &= P_{12}(dd) = P_{12}(gg) = P_{12}(hh) = 1/8 \\
 P_{12}(cg) &= P_{12}(dh) = (1/8)(1 + \cos(\delta_1 - \delta_2)) \\
 P_{12}(ch) &= P_{12}(dg) = (1/8)(1 - \cos(\delta_1 - \delta_2)) \\
 P_{12}(cd) &= P_{12}(gh) = 0
 \end{aligned}
 \tag{2}$$

The average number of photons observed in different channels can be formulated if only the combinations with different channels are considered. Equation 3 [2] describes the average number of photons for different channels combinations, where + is for  $cg$  and  $dh$  combinations and - is for  $ch$  and  $dg$ .  $S_1$  and  $S_2$  represent the flux densities for each source.  $V_{2PS}$  is the fringe visibility defined in Equation 4 [2].  $k$  is a constant defined in Equation 5 [2], where  $\tau$  is the width of the time bin for correlation,  $\Delta t$  is time step for observation,  $\Delta\nu$  is the detector's bandwidth,  $\nu$  is the frequency of the light that we want to observe, and  $A$  is the effective collecting area of each telescope.

$$\langle N(xy) \rangle = \frac{k(S_1 + S_2)^2}{8} [1 \pm V_{2PS} \cos(\delta_1 - \delta_2)]
 \tag{3}$$

$$V_{2PS} = \frac{2S_1S_2}{(S_1 + S_2)^2}
 \tag{4}$$

$$k = \tau \Delta t (A \Delta \nu / h \nu)^2
 \tag{5}$$

The simulator aims to use Equation 3 to produce the average coincidence rate, which is defined as dividing Equation 3 by  $\Delta t$ , for arbitrary telescope placement and source positions in the sky. Since all the terms are constant, the only complicated variable is the phase shift between the two telescope for each source. Equation 6 describes the difference in the phase shift for the two sources, where  $\lambda$  is the wavelength of the light,  $\Delta L$  is some constant instrumental path length difference between the two telescopes,  $\vec{B}$  is the baseline of the telescope, or the vector pointing from one telescope to the other, and  $\hat{s}$  is the unit vector point from the source to the telescope.

$$\Delta\delta = \delta_1 - \delta_2 = \frac{2\pi}{\lambda}\vec{B} \cdot (\hat{s}_1 - \hat{s}_2) + \frac{2\pi\Delta L}{\lambda} \quad (6)$$

The average coincidence rate will have a sinusoidal behavior since  $\delta$  varies with time as Earth rotates at an angular velocity of  $7.3 \times 10^{-5}$  rad/s. Therefore, the oscillation angular frequency,  $\omega_f = 2\pi f$ , of this sinusoidal behavior is determined by the rate of change for  $\Delta\delta$ . We can study the general behavior of the oscillation frequency by considering the extreme cases of telescope and pair source orientation. The first situation is when the baseline vector of the telescopes is pointing in the east-west direction, then the change in dot product between the baseline vector and the difference of the source unit vectors is very minimum if the source separation is also in the east-west direction. The two vectors become parallel when sources are right on top of the telescopes and perpendicular when the sources are on the horizon of the telescope. Due to the nature of dot product,  $\Delta\delta$  changes from maximum to minimum with time. On the other hand, if the source separation is in the north-south direction, the baseline vector is always perpendicular with the source separation unit vector, so there are no significant changes in  $\Delta\delta$ . The same idea applies when the baseline is in the north-south direction or perpendicular to the equator. The baseline will be always perpendicular to the separation unit vector between the two sources regardless of them being separated in the north-south or east-west direction.

### 3.1.2 Simulator Construction & Results

The main functionality of this simulator is to produce Equation 3 by generating an appropriate environment and supplying it with positions of the two telescopes on Earth and stars position in right ascension and declination. The simulator calculates  $\Delta\delta$  over the period of observation for a small time step  $\Delta t$ , where  $\Delta t \ll \frac{1}{\omega_f}$ . The default parameters of the telescope for this simulator are shown in Table 1.

Parameters	Values
Effective Collecting Area ( $A$ )	1[m <sup>2</sup> ]
Time bin width ( $\tau$ )	0.15[ns]
Detector Bandwidth ( $\Delta\nu$ )	1[GHz]
Instrumental Path length difference ( $\Delta L$ )	0
Wavelength ( $\lambda$ )	0.55 [ $\mu\text{m}$ ]

Table 1: This table shows the default parameters for the telescopes.

In order to test the validity of code, one quick way is to check the oscillation frequency behavior of the coincidence pair rate for baseline orienting in the east-west and north-south direction, which are shown in Figure 2. The results match with the theoretical prediction, where the east-west baseline and east-west source separation have a fast oscillating signal while the others don't.

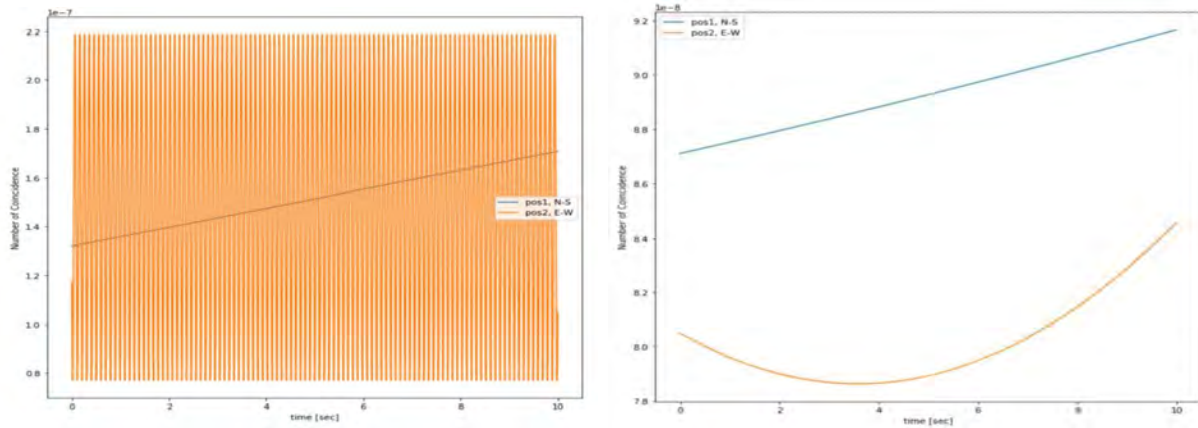




Figure 2: The two figures shows the number of coincidences as a function of time for channel combination  $cg$  and  $dh$  when the baseline is oriented in the east-west direction (left figure) and north-south (right figure). Left: The orange curve shows the oscillation pattern for a star pair separated in the east-west direction, while the blue curve shows the oscillation pattern for a star pair separated in the north-south direction. As expected with the theory, the orange curve is oscillating at much faster frequency, while the blue curve is oscillating very slowly. Right: The orange curve shows the oscillation pattern for a star pair separated in the east-west direction, while the blue curve shows the oscillation pattern for a star pair separated in the north-south direction. As expected with the theory, both of them are oscillating very slowly.

### 3.2 Bright Star Catalogue Processor

The bright star catalogue, or BSC, is a star catalogue that lists around 9000 stars of stellar magnitude 6.5 or brighter, which is roughly every star visible to the naked eye from Earth. This catalogue is useful at providing information about all the possible stars for observation in real experiments in the future. The catalogue contains a bunch of properties for each star, so it is necessary to extract out the relevant ones including right ascension, declination, and spectral flux density. Since the catalogue only has star magnitude in different filters, the magnitude in visible filter is converted to flux density in Jansky using Equation 7.

$$F[\text{Jy}] = 3640[\text{Jy}] \times 10^{\frac{M_V}{-2.5}} \quad (7)$$

The entire bright star catalogue is processed into an array of shape  $(N, 4)$ , which contains  $N$  stars and the corresponding star number, right ascension, declination, and flux density. Then the array is filtered to extract out the appropriate star pairs for real observations based on different criteria. Currently, the conditions are set to:

- Stars whose relative declination between the telescope declination needs to be within 30 Degrees.
- The spectral flux density of stars should have at least 50 [Jy].
- The angular separation distance between the two stars in a pair must be less than 0.01

rad.

- Stars should be 180 Degrees (12 Hours) to 225 Degrees (15 Hours) behind the sun so that they can be seen in the night sky.

After eliminating stars based on their basic properties such as flux density and declination, the rest of the stars need to pair with other stars to account for all possible star pair combinations. If there are M stars after the initial filtering, then there will be  $\frac{M(M-1)}{2}$  unique pairs of stars. The last filtering will be done by requiring the separation within each pair to be less than 0.01 rad.

The telescopes are chosen to be in New York and placed in a way that the baseline is parallel to the Equator. There are 18 unique pairs of stars if we ignore the last condition regarding the position of the stars relative to the sun. A quick visual representation of the 18 pairs of stars is shown in Figure 3, and the coincidence rate curve for all 18 pairs is shown in Figure 4.

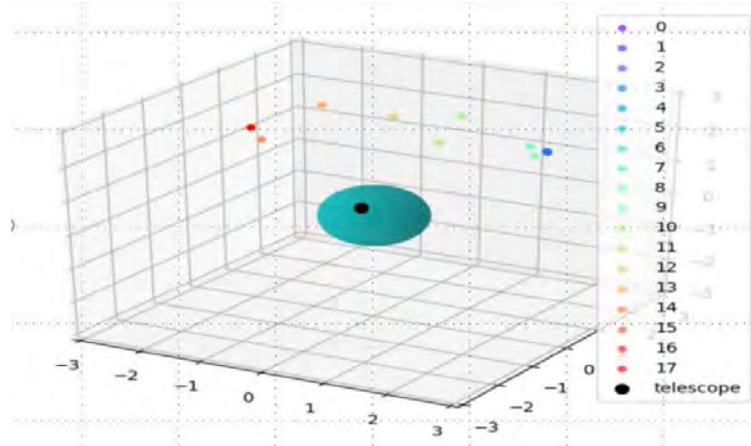


Figure 3: This figure shows a simple representation of 18 unique pairs of stars. The blue big sphere represents the Earth, while the black dot represents the position of the telescope. There are no 18 unique stars because many pairs of stars overlap with each other.

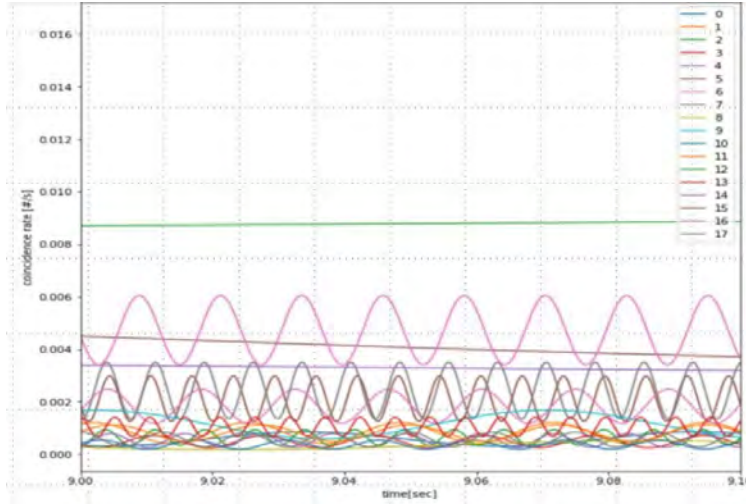


Figure 4: This figure shows the coincidence rate for 18 star pairs as a function of time. These coincidence rate curves are for channel combination  $cg$  and  $dh$  where the telescope is placed in the east-west direction at New York.

### 3.3 BSC Result Analysis

In order to see the behavior of the change in oscillation frequency for the coincidence rate curve as a function of time, the period of observation is increased to a much longer period than 10 [sec] in Figure 4. Figure 5 shows analysis plots for pair #16 with observation period = 3000[sec]. The upper left panel is the regular coincidence pair rate as a function of time, while the bottom panel shows the amplitude plot of the Fourier transform of the upper left panel. Clearly, the sinusoidal shape has different frequencies, which suggests the oscillation frequency changes over time. The panel on the upper right shows the oscillation frequency as a function of time, which is found by doing  $\frac{d\Delta\delta}{dt}/(2\pi)$ . The maximum and minimum frequency determined by both methods give the same results, which is also a quick sanity check on these analysis.

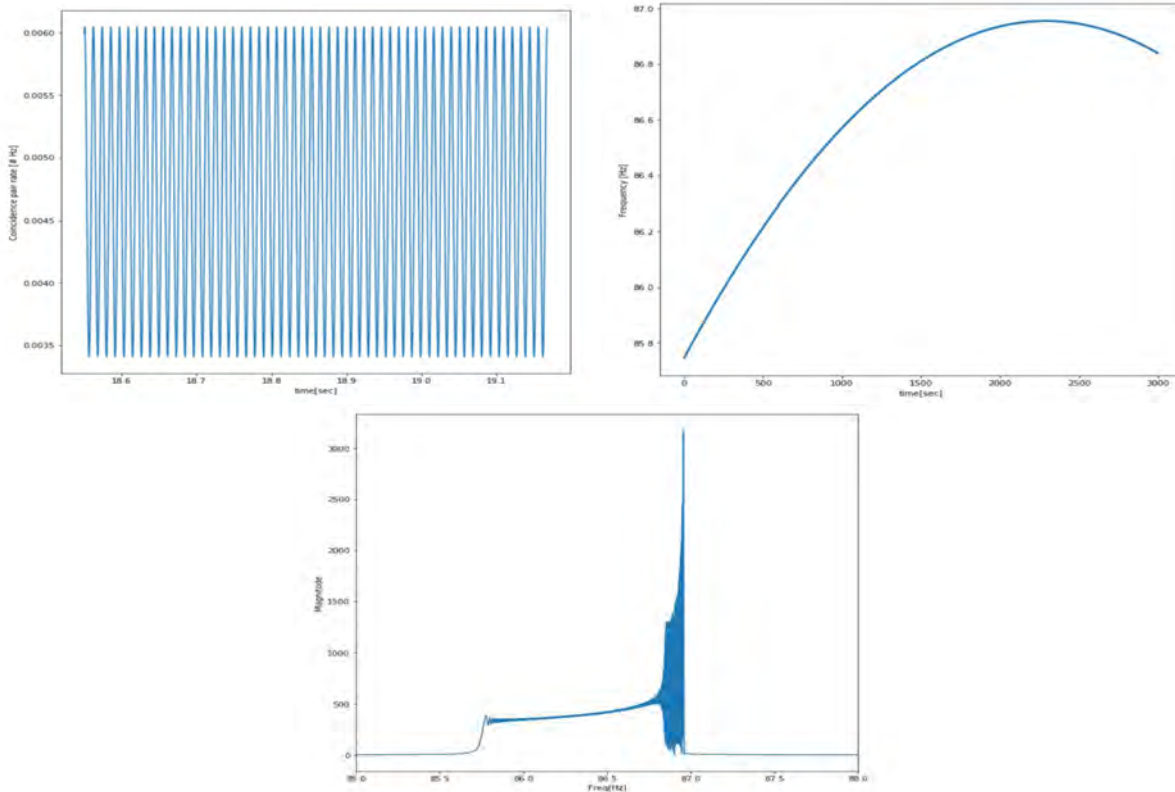


Figure 5: This figure shows the analysis for pair #16 with observation period = 3000[sec]. The upper left panel shows the standard coincidence pair rate as a function of time. The bottom panel shows the amplitude plot of the Fourier transform of the upper left panel. The panel on the right shows the oscillation frequency as a function of time. From both oscillation frequency analysis, the frequency components varies from 85.8[Hz] to 87[Hz] over the course of 3000[sec].

If the observation period is increased to a day, we see that the change in the oscillation frequency appears to have a sinusoidal shape as a function of time as shown in Figure 6. There are positive and negative frequencies, which represent the alignment of baseline vector of the telescopes and the separation unit vector between the two sources. Even though only the relative longitudinal direction between the baseline and the separation unit vector is changing due to Earth rotation, declination also seems to play a role at when will the oscillation frequency be maximized.

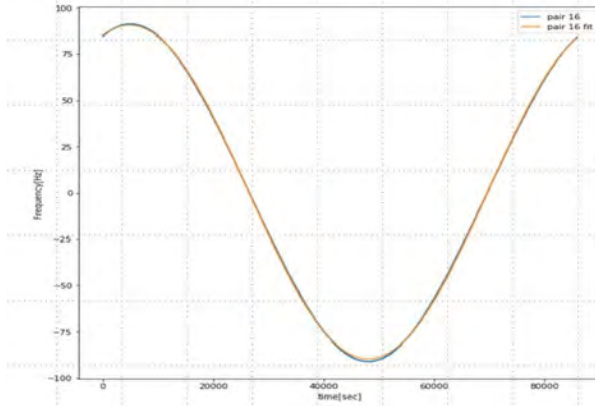


Figure 6: This blue curve shows the theoretical oscillation frequency of pair 16 as a function of time over the course of a whole day, while the orange curve shows a simple fitting of a cosine function to this curve. The positive and negative frequencies simply mean that the baseline and the separation unit vector of the two sources are parallel and anti-parallel with each other. There seems to not have a simple relationship at what right ascension the maximum frequency, or the largest change in  $\Delta\delta$ , is going to happen, rather it depends on both declination and right ascension.

Figure 7 shows the maximum oscillation vs. the average coincidence rate for each pair of stars. The two panels of figure 8 shows the relationship between maximum oscillation frequency vs. the angular separation in right ascension and declination of each star pair. In general, we see a strong correlation for separation in right ascension with Pearson's R coefficient of 0.94 while a very weak correlation for separation in declination with Pearson's R coefficient of 0.42. These results are expected because for the baseline oriented in the east-west direction, the pair separated in the east-west should have a higher oscillation frequency in theory, while the pair separated in the north-south direction should have a small oscillation frequency.

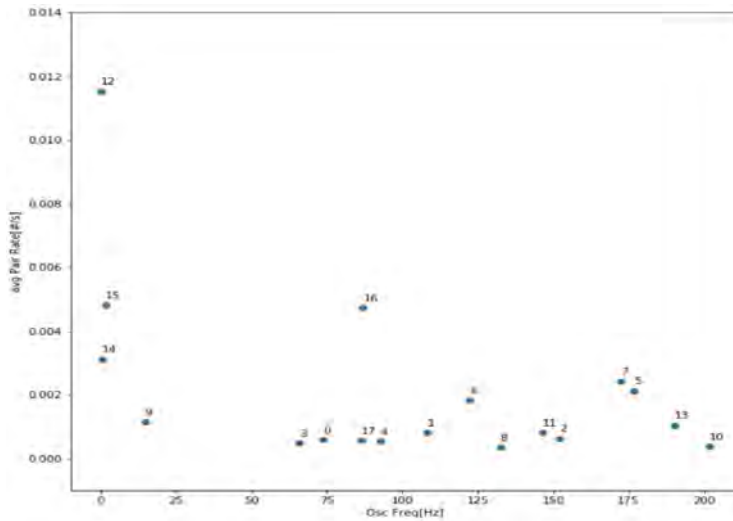


Figure 7: This figure shows the average coincidence rate of each star pair vs. the maximum oscillation frequency with star pair # labeled on the side.

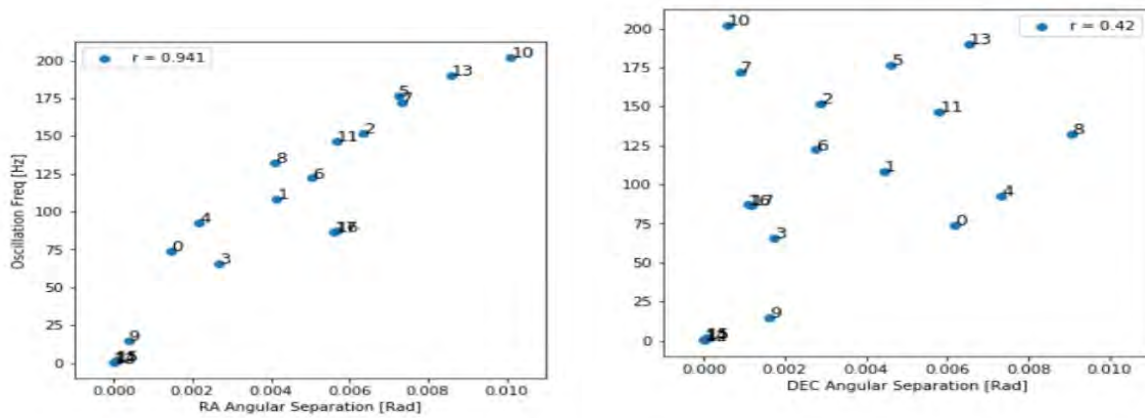


Figure 8: The scatter plot on the left shows the maximum oscillation frequency vs. the difference in right ascension between the two sources, which has a Pearson's R coefficient of 0.94. The scatter plot on the right shows the maximum oscillation frequency vs. difference in declination between the two sources, which has a Pearson's R coefficient of 0.42. The strong correlation in the angular separation in right ascension is reasonable since the oscillation frequency is a lot higher for sources placed in the east-west direction as we saw before. On the other hand, there is very little correlation for the maximum frequency vs. the angular separation in declination because the oscillation frequency is not sensitive to the separation in the north-south direction.

## 3.4 Simulated Data

### 3.4.1 Procedure

Once we have the theoretical distribution of the number of coincidences for different channel combinations, producing simulated data through Poisson process allows comparison between the theoretical probability density function and the maximum likelihood estimation of the sample data. Poisson process is the general idea of finding the coincidence counts for each coincidence cycle. Poisson distribution requires the average number of coincidence counts, which is equal to the average coincidence rate,  $\bar{n}$ , multiplied by the period of the cycle,  $T$ . After obtaining the theoretical  $\bar{n}$  and  $T$  from the simulator to get the number of occurrence of events for each cycle, then we determine the phase of the cycle at which there will be a coincidence count. The probability density function, or PDF, of different phase is hinted by Equation 3, which is shown in Equation 8. The cumulative probability density function is simply the area under the PDF or integrating the PDF from  $-\pi$  to  $\phi$ , which is shown in Equation 9. Clearly, there will be a higher probability of receiving a coincidence when the phase is equal to multiple of  $2\pi$ . The CDF gives a relationship of the corresponding possibility of finding a certain phase,  $\phi$ .

$$PDF(x) = \frac{1 \pm V \cos(x)}{2\pi} \quad x \in [-\pi, \pi] \quad (8)$$

$$CDF(\phi) = \frac{\phi \pm V \sin(\phi) + \pi}{2\pi} \quad \phi \in [-\pi, \pi] \quad (9)$$

After feeding a random probability from 0 to 1 to the CDF, the phase at which the coincidence count is recorded,  $\phi$ , is calculated through the bisection method after inverting the CDF. Lastly, the theoretical coincidence rate curve is used to find the corresponding timestamp for the phase in each cycle.

### 3.4.2 Sample Data

The sample trial data shown under this section is specifically for star pair # 16 for a total observation period of 3000[sec], and the wavelength of observation is changed to 1[ $\mu\text{m}$ ] from 0.55[ $\mu\text{m}$ ]. The left panel of Figure 9 shows the histogram of the simulated phases along with the theoretical PDF. For about 130,000 cycles, a total of 50 event occurrence happened based on Poisson process, and for cycles that happen to have a count, the phase is calculated using the inverse CDF method. The right panel of Figure 9 shows the scatter plot after feeding the simulated phase into Equation 3 and plotting it against the corresponding timestamp of each simulated phase.

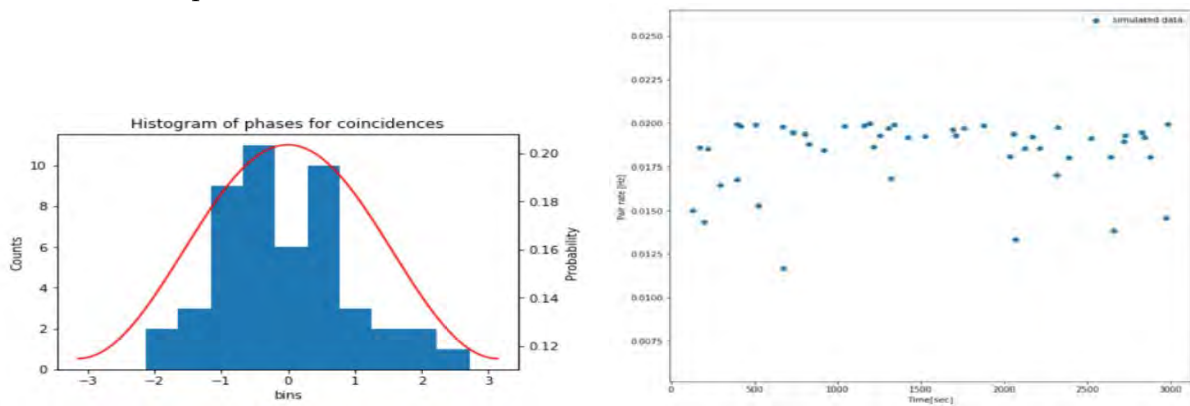


Figure 9: Left: It shows the simulated phase for star pair # 16 observing at wavelength 1 [ $\mu\text{m}$ ]. There are about 130,000 total number of cycles with a total of simulated event occurrence of about 50 only. The simulated phase are put into a total of 10 bins with uniform bin width and the number of counts for each bin on the left side of the y-axis. The theoretical probability density is also shown in the red curve with probability shown on the right side of the y-axis. Right: It shows the scatter plot coincidence pair rate as a function of time after feeding the simulated phase into Equation 3 and plotting it against the corresponding timestamp. Clearly, most of the data points are bunched on the top as expected.

### 3.4.3 Future Works

In the future, the simulated data can analyzed calculating the maximum likelihood estimation of the simulated phase at which an coincidence count occurs. The maximum likelihood estimation can then be compared with the theoretical probability density function to estimate the precision of the data set. We can also crosscheck the precision on the different



parameters such as the oscillation frequency and the angular separation between the two sources with the fisher's results in paper [2] since we know the theoretical results from the simulator. In summary, a careful analysis of these simulated data will be conducted in the future.

## 4 Conclusion

This paper discussed the general procedure of constructing a simple simulator for producing theoretical coincidence pair rate as a function of time for a source pair in the sky based on this paper [2]. The simulator requires information about telescope placements, position of the source pairs, and their brightness in order to generate the coincidence rate curve. Then the next step is to use the bright star catalogue as the source input for the simulator, which contains information about all stars that are bright enough for the human eyes. Many filtering conditions were applied to narrow down the selection of ideal star pairs for the real observation in the future. There are 18 pairs of stars after all the filtering if the condition of the right ascension is ignored. A simple analysis of the results is conducted, which include the Fourier components of the coincidence rate curve, the relationship of oscillation frequency vs. time, and the relationship of maximum oscillation frequency vs. angular separation between the two sources. Lastly, simulated data were generated using Poisson process, which can be analyzed and compared with the theoretical results in the future.

## 5 Acknowledgement

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

I want to thank my mentor, Andrei Nomerotski, and other staff members, specifically Anze Slosar and Paul Stankus for their teaching and guidance on the project.

## References

- [1] Daniel Gottesman, Thomas Jennewein, and Sarah Croke. “Longer-Baseline Telescopes Using Quantum Repeaters”. In: *Physical Review Letters* 109.7 (Aug. 2012). ISSN: 1079-7114. DOI: 10.1103/physrevlett.109.070503. URL: <http://dx.doi.org/10.1103/PhysRevLett.109.070503>.
- [2] Paul Stankus et al. *Two-photon amplitude interferometry for precision astrometry*. 2021. arXiv: 2010.09100 [astro-ph.IM].

Michael Chin

## Design NSLS-II PPS Test Fixtures

For my project I will be designing a NSLS-II PPS Test fixture. You may be asking yourself what is a PPS test fixture? Well first let's define what PPS stands for. PPS stands for Personnel Protective System. This PPS is a system that Brookhaven National Laboratory uses to ensure safety to its employees. And a test fixture is a mechanical structure that helps the PPS system do its job. I will be working on the booster gate test fixture. The test fixture I am working on will make it so a group called Flowco can come into the lab and easily recertify the door switches using the test fixture.

Before I started to work on the test fixture, I first had to gain a basic understanding of mechanical engineering terms. Without the basic understanding I would be lost in the wind. The terms and knowledge I had to learn before working with the actual test fixture are terms like screw sizing's, dimensions, tolerances, etc. I had to go over decimal charts and study the conversions just so that I can get used to what the numbers look like and how big or small those decimal charts make when looking at a part.

After I started learning about decimal charts and how the numbers in each decimal place change depending on part sizing, I then learned about screws. Before I began this program, I thought that a screw was just a screw, but no, there is a lot of precision that goes into screws. Such precision includes pitch, inner and outer, threads per inch, and the actual type of screw, sinkhole, or countersink. There are tons of different screw options and I familiarized myself with each one and learned why one screw would be better for the job than another.

I then learned about tolerances. Tolerances are basically the amount of leeway you can give when designing a part. Let's say we must add two holes to a metal plate. And say we must put the holes two inches apart and a half inch from the starting left side. Well, when dimensioning the part in the software we would dimension the two holes first, and then dimension those two holes a half inch from the side. That is because those two holes have relation to each other. The holes must be two inches apart and dimensioning a part this way is more accurate. That is because the tolerance, or room for error is less when dimensioning this way

Once I learned all the basic mechanical engineering terms that I had to learn, now It was time to delve into Inventor and Vault software. The first thing I had to learn was file management. Inventor and Vault work with each other. Although they are different software's, they depend on each other. Inventor and Vault are not like Microsoft word, you can't just open a file last used in Inventor like you can in Microsoft word. There are specific steps that you must take for the process to work. If I missed a step I would be disconnected from the VPN and would have to take time reconnecting. Vault works like a check in check out system. To work on a drawing in Inventor, first you must check the part out of Vault. Once the part is checked out of Vault, then you can work on the design in Inventor. Then if you are finished working on a part in Inventor, you can either check the part into Vault, or keep the part checked out so that you can work on it at a different time.

Once I got out all of the basics out of the way, then It was time to start working on the Booster Gate Test Fixture. First, we started going over what the test fixture was for and why it is so important to have in the first place. At BNL there is a powerful beamline that goes through certain hutches. For safety reasons no personnel are allowed to be in one of those hutches when

the beam is running. In order to combat that, BNL has implemented a redundant gate system that ensures a door is closed when it should be closed and vice versa. The door switches at BNL have four devices called rotary switches. When the door is closed, each rotary switch closes, which sends a signal back to the control room to let the operator know if the door is closed. But these four switches are not the same. They have redundancy towards them. The top two switches are created by the company called Simens, and the bottom two are created by the company called Alan Bradley. This is so that if one of the switches were to go bust, then the other three can show an accurate decision. The test fixture I will be working on will make it easier for the group called FLowco (which is the group that recertifies at BNL) to recertify the booster gate, which is basically the four switches.

While working on the booster gate test fixture there are a couple of things that I worked on. The first thing that I did was adding eight holes into slots. This is where all the dimensioning work that I did in the beginning came into play. First, I had to carefully measure the parts on each side just to make sure that the hole widths and dimensions were correct. Then I had to add the holes. Adding the hole and choosing the correct hole was easy because I learned about holes in the beginning, and which one was better for which type of situation. Then I had to constrain parts together. There are moving parts on the test fixture so I had to measure those parts and make sure that in real life they would not interfere with each other, and that the movement would work smoothly.

Then I had to add screws to the holes by doing constraints. There are different constraints, and each have their specific purpose. To constrain a part to another part, you have to have a clear understanding on how the part works. You have to constrain the screws to specific

parts that are already there or else the screw will move and won't be in the right place, which will cause confusion on the future.

This internship was extremely insightful. Even though I only scratched the surface with knowledge and just started my journey, I am ready to begin. Now that I got my toes wet this is where my career starts. The knowledge I gained in the internship is perfect to get my foot into the door. I would like to thank my mentors Gregory Fries and Scott Orban, as well as my team leader Laurie Clark for this wonderful experience.

The redevelopment of the explosives safety program at Brookhaven National  
Laboratory

Colin B. Clark and Michael F. Clancy Jr.

Safety & Health Services Division, Brookhaven National Laboratory, Upton, New  
York 11973-5000

Colin B. Clark, Chemistry Department, Binghamton University, Binghamton, NY 13902-  
6000

Michael F. Clancy Jr., Safety & Health Services Division, Brookhaven National  
Laboratory, Upton, NY 11973-5000

## **Abstract**

The Explosives Safety Program (ESP) at Brookhaven National Laboratory (BNL) is integral not only to the safety of personnel working directly with explosives, but for all those at BNL. While BNL does not physically modify explosives, the Laboratory is responsible for operations involving the procurement, storage, handling, analysis, and disposal of explosives. There is great emphasis on safety at BNL because safety makes science possible. My project has been to create a document, known as the Explosives Safety Manual (ESM), that will act as the program description for the ESP at BNL. The ESM is to be used by subject matter experts (SMEs) in the field and to lay out a pathway for the potential expansion of the ESP past de minimis quantities of explosives. Throughout the course of writing the ESM, I have had to cross reference documents and seek out the appropriate personnel to gather information where it was not documented. The current ESP was rather limited and required a lot of attention and detail to ensure it is compliant with the Department of Energy Explosives Safety Technical Standard (DOE-STD-1212-2019). The lack of readily available information is a flaw of the ESP because when pertaining to explosives, there should be no lapse in knowledge; no difficulty to ensure a safe workplace by having everyone aware of all safety measures. Any problems I noticed were addressed and brought to the attention of my mentor to either change or implement into the ESM as well as the subject area website used by all personnel. I have learned the difficulty in making certain all standards and regulations are followed. This internship allowed me to experience working in a professional setting with one project consistently and communicating with some of the leading experts in the field at BNL.



## I. Introduction

When I was first given my original project title, Explosives Safety, on my Intern Request Form, I was unsure what to expect. This was my first experience working fully online, let alone my first internship. Once I met with my mentor, Michael Clancy Jr., he informed me about what he saw as the objectives of my internship. This internship focused on comparing the Department of Energy Explosives Safety Technical Standard (DOE-STD-1212-21019) to Brookhaven National Laboratory (BNL)'s Explosive Safety Program (ESP) and seeing where it was compliant and where it needed improvement. I was given the Standard and access to the Explosives Safety subject area, which is within the BNL Standards-Based Management System (SBMS). The SBMS is the system used for all employees at BNL and serves as the requirements for all work performed in the Laboratory. It is organized into specific subject areas as to easily be navigated by employees working within them. Upon reading the Explosives Safety subject area, I learned that the scope of what BNL does with explosives differed greatly from my expectations.

Normally, when one hears the word “explosives”, the most common thing to think about is “explosions”. At BNL, this is not the case. There is no physical modification of explosives happening on-site; no synthesis, development, testing, or disarming of explosives. Instead, the subject area provides guidance on the procurement, storage, handling, analysis, and disposal of explosives. Unfortunately the subject area was the only resource available for explosives safety and lacked a great deal of information.

This was the moment where I began to understand how exactly my project would take shape.

The ESP at BNL is limited to explosives of de minimis quantities. De minimis quantities are quantities of explosives that are considered to be non-detonable by the environment or any other abnormal stimuli. Numerically, that is 1 mg of primary explosives and 10 mg of secondary explosives within the entire Laboratory. In August 2014, an explosives site safety plan (ESSP) was approved by DOE that would allow BNL to surpass de minimis quantities but never received the funding to successfully be implemented. Speaking with my mentor and explosives safety SMEs, there still appeared to be a great interest in expanding the ESP. Regardless of what is wanted to be done, there still lies the problem of what needs to be done. A stronger foundation for the documentation of explosives safety would not only be more informative for the personnel at BNL; it would also increase awareness for interest in explosives research. I have decided to create the Explosives Safety Manual (ESM) as a way to accomplish this.

## **II. Initial Research**

I began my research by reading through each part of the Explosives Safety subject area to understand the exact standards, methods and procedures for the ESP. It was provided in a convenient step-by-step method yet it lacked the details as I would have expected with explosives. The information present was compared to the requirements of DOE-STD-1212-2019. Afterwards I was faced with many holes in my findings. The lack of information required me to navigate through the SBMS and

approach my mentor with questions. If there were questions that he was unable to answer, I sought out the appropriate personnel to gather more information. DOE-STD-1212-2019 is very detailed and specific with its regulations and there was no resource from BNL where I could verify if all the applicable regulations were being followed, coupled with the inability to view the facilities on-site.

Of the information that was missing from the Explosives Safety subject area, several topics stood out to me the most. Information about the Chemical Management System (CMS) was limited and was not included as a resource, despite the CMS Team being responsible for providing safety data sheets (SDSs) and classifying, storing, and keeping inventory of explosives. Concerning the transportation of hazardous materials, where explosives are classified by the Department of Transportation (DOT) as a Hazard Class I; it is not mentioned once in the subject area it is while being a closely related and intersecting subject area. Another glaring problem with the subject area was a lack of training requirements. Despite stating that all users and staff handling explosives are properly trained and qualified, there is no documentation to verify this. The subject area itself states there are no training requirements. Speaking with my mentor and SMEs I learned that SMEs have a training course at the Pantex Plant in Amarillo, Texas. However, for all other users on-site, the explosives are in de minimis quantities and training was deemed unnecessary.

This is when I understood that my project would not be fixing the subject area but to create something else entirely. The vast amount of information related to explosives safety and a need for detailed documents is what led me to write the ESM, which I

based off of the Hazardous Material Transportation Manual (HMTM). The HMTM is a program description used by Transportation Safety SMEs and provides guidance and procedures for all divisions and departments under it. I used the applicable chapters of DOE-STD-1212-2019 and the format of the HMTM as an outline for the ESM. Once I had an idea about how I wanted to present my information, it was a matter of finding and organizing it.

### **III. Writing the ESM**

Unlike subject areas, which are meant for a more general audience, program descriptions are read by SMEs and can be more detailed and involved with the scientific language used. I began by creating a table of contents based on my outline to ensure I addressed all parts of the ESP. The first chapter of the ESM explains the purpose, scope, and applicability of the ESM. Without a purpose in mind, it is just a collection of information. It explains what will be covered in the manual. This is especially important for the Explosives Safety SMEs as to be able to quickly navigate through it and to make changes if anything is missing.

Chapter 2 goes through each role at BNL and all of the responsibilities associated with it. While at first glance each role appears to not be involved with explosives safety, reading through their responsibilities leads to the conclusion that they are and each plays an integral part in the ESP. When the responsibilities are divided and organized, you know who to contact depending on your question. Appendices are attached for organizational charts of the Laboratory, Department, and the ESP.

Chapter 3 addresses all of the external regulations that BNL is compliant to; which depends if the explosives operations are on-site or off-site. At BNL, the only off-site activities taking place are the shipment or disposal of explosives. All other laboratory operations occur on-site.

Prior to writing the ESM, I found many other subject areas that had answers to the questions I was asking. To ensure all personnel is aware of these resources, Chapter 4 exhibits all of the site-specific literature pertaining to explosives safety. This has a wider scope than the explosives safety subject area where the subject area has no reference to the many other subject areas that are associated with and support the ESP. Examples include the Transportation of Hazardous and Radiological Materials Off-site subject area, Movement by Vehicle of Hazardous and Radiological Materials On-site subject area, Shipping Process Subject Area, Waste Subject Area, and several others. If I had not have been searching for more information to determine compliancy with the Technical Standard, I would have not have found these resources otherwise. The ESM serves to display the connections between disciplines and how the ESP is more robust than it initially appears. The ESM provides a brief description for each piece of literature and a hyperlink for direct access if more information is desired.

The very first responsibility of the ESP is to approve explosives operations before they are purchased and brought on-site for research purposes. This is done through submitting forms known as the BNL Explosives Request Authorization Form (ERAF) and the Experiment Safety Review (ESR) which are reviewed and approved by the Explosives Safety Committee (ESC). A major component of these forms is an in-depth

safety assessment and hazard analysis. As previously stated, explosives are a Class Hazard I by the DOT. However, Class Hazard I has six subdivisions with each containing their own classifications and hazard controls. On top of this, these hazard divisions only apply for the transportation of explosives. When storing explosives, different groupings are assigned based on storage compatibility and when performing explosives operations, there are six hazard classes depending on the nature of the operation, type of explosive, and quantity used. To assist principal investigators (PI) and the ESC, the Chapter 5 of the ESM provides a detailed explanation of each hazard categorization as well as a step-by-step guide for how to fill out the BNL ERAF and ESR. This is done to ensure that no hazards are being overlooked and all are properly controlled. There are also several other forms that need to be filled in once explosives are on site to keep track of inventory and location as to not surpass the de minimis quantities given by DOE standards.

Chapter 6 is dedicated to operational safety. It addresses what activities are permitted with explosives and additional safety measures to be made to the facility. A clear list of exemptions for facilities working with de minimis quantities of explosives is provided and the ESM can be updated accordingly if a new explosives safety site plan (ESSP) is approved in the future. Along with this is addressing several other safety subject areas that can be utilized to prevent accidents from happening. With chemical safety, it is important to be able to understand how to read an SDS and the location of any emergency equipment in the facility. Electrical safety is important because the DOE-STD-1212-2019 has regulations involving electrical equipment, permanent wiring, and lightning protection. Also, some explosives at the Laboratory are sensitive to static

electricity, where a buildup could cause a potential explosion. If an explosion were to occur and cause a fire, then fire safety becomes relevant, including an egress plan to successfully evacuate all personnel. However, the most important aspect of safety is prevention so that emergencies do not occur. That is why it is important to have a resource where all personnel, directly and indirectly involved with explosives, are made aware of the hazards they possess.

The ESM serves to establish training requirements for all users working with explosives to ensure they are qualified for the tasks being performed. Prior to the ESM, there was no records of training or qualification. While the ESM does not have any specifics for courses personnel would have to take, Chapter 7 is outlined for the SMEs to fill in at a later date once the training is formalized. Not all workers require the same training depending on their tasks so the ESM divides each course into levels that correlate to specific personnel. It will serve as a guide for initial and recurrent training and accurately document resources where training records can be found.

Coupled with the idea with training records, it is important for information to be documented and told where it can be found. Doing combats the problems of confusion and searching for the right person to ask. Chapter 8 of the ESM presents a list of all documents that support the ESP at BNL, along with who is responsible for maintaining specific records and where they can be found. By having a more organized program, it minimizes the amount of questions that can be raised.

Following this is emergency response and incident reporting. While accidents and incidents are to be prevented via safety assessments and hazard analysis, there

needs to be preparation for when these methods fail. Chapter 9 is dedicated to providing procedures that are to be followed in the event of an emergency. Based on previous history from within the current ESP, procedures can be written for emergencies that have an increased likelihood to occur, if there have been any. BNL has what is known as the Laboratory Protection Division which is dedicated to promoting the safety of personnel at BNL in times of crisis.

Chapter 10 is dedicated to the expansion of the ESP by providing guidance on how to submit a new ESSP to DOE. The information is based on the previous ESSP submitted in August 2014 that was approved but not funded. The chapter contains a checklist of requirements that need to be acknowledged and addressed to ensure compliance with DOE standards when operating with explosives above de minimis quantities. If there are plans to expand the research, this resource will provide all of the necessary information.

The final two chapter are included to make the ESM more accessible to anyone who may choose to read it. Included are acronyms and definitions in alphabetical order of words used in the ESM. There are several appendices that are included to increase the reader's understanding of several key factors of the ESP. These include mandatory forms used by the ESP, a BNL site map, organizational charts, an SDS outline, explosives storage compatibility, and hazardous materials vehicle operations.

#### **IV. Conclusion**

The ESM has been a daunting task from the beginning, but its importance is what has kept me driven throughout the internship. I have learned how federal



regulations are written and how to navigate the SBMS to find the information I needed, improving my research and scientific writing skills. This project allowed me to utilize my communication and collaborative skills before I begin my PhD in the Fall by being able to present the research I have worked on. As someone who has never worked here or with explosives before, I had no reference to how activities are carried out at BNL. I gained a great amount of knowledge of the ESP and recognized the various shortcomings, dedicating my time to ensure compliancy to DOE-STD-1212-2019 so the ESP can reach the high caliber expected at BNL. It is my wishes that the ESM can be successfully published as a program description for explosives safety and lay out a pathway for the redevelopment and expansion of the ESP at BNL. I am grateful for the opportunity to have been able to work alongside some of the leading experts in the field and hope that I have made valuable contributions to furthering the research of explosives at BNL.

## **V. Acknowledgements**

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). I would also like to thank the following people for their contributions to this project and for providing me with their time, advice, and assistance: Michael Clancy Jr., Dr. Steven Coleman, Dr. Janice Epstein, Cassandra Asselta, Fredrick Horn, and Dr. Joe Pavlak. This project would not have been possible without their help.

## **Studies of the mechanism of CO oxidation using transient experiments**

Gerardo J. Claudio-Serrano

School of Science and Technology, Universidad Ana G. Mendez –  
Gurabo Campus, Gurabo, PR 00725

Lisa Feder

The Grove School of Engineering at The City College of New York, New York, NY  
10031

José A. Rodríguez

Chemistry Division, Brookhaven National Laboratory, Upton, NY 11973

Jorge Moncada

Chemistry Division, Brookhaven National Laboratory, Upton, NY 11973

## Abstract

This project in the Chemistry Division at Brookhaven National Laboratory (BNL) is a part of a major effort that seeks to understand and control chemical systems to develop clean energy solutions for national needs. The present study is focused on understanding the mechanism for the oxidation of CO (a major air pollutant) on metal/oxide-catalysts through a fundamental understanding of the catalytic process. Cerium oxide ( $\text{CeO}_2$ ) based materials have been under constant experimentation in recent years as potential catalysts with promising results for CO oxidation when paired with non-expensive materials such as copper oxide ( $\text{CuO}$ ). However, a fundamental understanding of the mechanism of this catalytic process remains difficult, because of the complex nature of phenomena that occur dynamically between reactants and polycrystalline catalyst materials. In these experiments the redox chemistry and CO oxidation ( $2\text{CO} + \text{O}_2 \rightarrow 2\text{CO}_2$ ) activity of an inverted  $\text{CeO}_2/\text{CuO}$  catalyst was investigated by using advanced transient characterization techniques such as time-resolved mass spectroscopy under operando conditions.  $\text{CO}_2$  was found to be over produced, with the amount of  $\text{CO}_2$  that evolved into the gas phase being substantially larger than the amount of CO consumed. This could be due to an evolution in the composition of the catalyst as a result of the decomposition of carbonates present on the catalyst surface. Different types of characterization techniques available at BNL will provide an opportunity for a multiangle approach toward understanding the reaction mechanism and the intrinsic nature of the catalyst active phases in the coming months. As a result of our internship this summer, we have developed our skills in data analysis using the software Originlab, as well as added basic knowledge of techniques for transient studies such as: mass spectrometry (MS), time-resolved X-ray absorption near-edge spectroscopy (XANES), diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS).

## **Introduction**

CO is a toxic gas that is often called the silent killer since it does not have taste, color, or smell and generally results from the incomplete combustion of fossil fuels. A small exposure to CO can be fatal since it has a high affinity to replace oxygen and binds to hemoglobin in blood cells. CO not only affects human beings but also vegetation and indirectly increases global warming<sup>[4,5,14]</sup>. The present study is focused on understanding the mechanism for the oxidation of CO (a major air pollutant) on metal/oxide-catalysts through a fundamental understanding of the catalytic process.

### Catalysts and chemical transformations

Catalysts play a big role in many chemical reactions in the laboratory and industry. A catalyst is substance or material that can speed up a reaction by providing an alternate pathway to the chemical process<sup>[1]</sup>. This alternate pathway requires less energy, therefore getting more product by spending less. Without a catalyst, a reaction can take days, weeks, months, and even years. But with a catalyst, the reaction can happen in a much lower time. It is important to note that a catalyst works as a facilitator of the reaction and it is not consumed during the catalytic process<sup>[1,2]</sup>. Therefore it appears in the steps of a reaction, but it will not appear as a reactant or product.

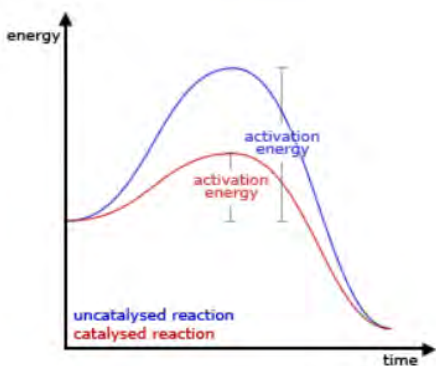


Figure 1. Activation energy diagram comparison of a catalyzed and uncatalyzed reaction.

During a reaction, chemical bonds between atoms break and form new bonds with different atoms. Breaking and forming bonds or how tight the molecules are holding on to each other is dependent on the bonding energy of the molecule. Bonding energy is the amount of energy it takes to break a chemical bond. Adding energy over a certain limit can break bonds and when new bonds are formed energy is released<sup>[1]</sup>. This means, lower energy bonds break more easily than higher energy bonds. In a reaction, higher energy bonds may take plenty of time before breaking and forming new ones because they are more stable. Bond angle, bond length, and other reaction coordinate parameters affect the rate at which a reaction occurs as well as other non-geometric parameters for more complex reactions such as bond order. A shorter bond requires more energy to break as opposed to a longer one. As for bond angle, breaking or forming new bonds can also be facilitated based on the position of the bond in a molecule.

A catalyst can make these molecules break or form more easily by lowering the activation energy of the reaction by getting involved in its transition states<sup>[1,2]</sup>. Activation energy is the energy needed for a reaction to occur. For a reaction to occur, existing bonds must break and new ones must form until the product is more stable than the reactant. In general, a catalyst alters the stages of a reaction to form intermediates which yield the same reaction products. For example, in figure 2, the

hydrogen lands on the surface and breaks its bond to form H atoms bonded to the surface (2). The double bond of the ethene is also broken and the two carbon atoms also bond to the surface (3). Then the H atoms can migrate until they collide with the bound carbon species and react (4) to form ethane which can then leave the surface (5). The Arrhenius equation states that the lower the activation energy, the higher the rate at which a reaction occurs. A reaction can also operate at lower temperature if a catalyst is present<sup>[15]</sup>.

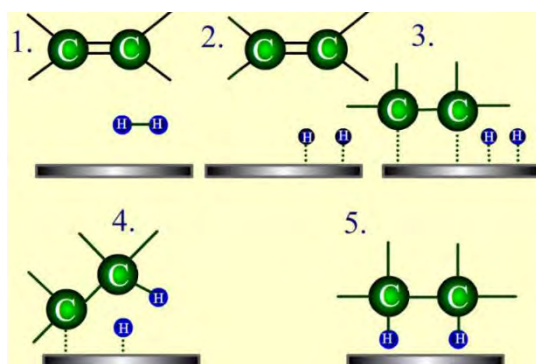


Figure 2. Cartoon depicting the reaction of ethene and hydrogen gas.

In general, there are two types of catalysts, homogeneous and heterogeneous<sup>[2]</sup>. Homogeneous catalysts exist in the same phase as the reactants in a reaction. A heterogeneous catalyst is not in the same phase as the reactants (usually a solid catalyst in a liquid or gas reaction mixture). Heterogeneous catalysts adsorb the reactants into sites on the surface of the catalyst, facilitating breaking of chemical bonds, and making it easier to convert to the product. It is important to note that heterogeneous catalysts have a limited number of active sites, limiting the rate of reaction.

Another important idea about catalysts is selectivity. A catalyst does not speed up all the reactions in a network. Instead, it only interacts with a specific reaction out of all the occurring reactions in a given space or network<sup>[2]</sup>. Choosing which catalyst to use is essential when performing catalyst research.

## CO-Oxidation

The Chemistry Division at Brookhaven National Laboratory (BNL) seeks to understand and control chemical systems to develop clean energy solutions for national needs. This involves studies on mechanisms of catalyst to meet demands for energy, synthesis of fuels for the future, and the remediation of environmental pollutants, including greenhouse gases such as carbon monoxide (CO). An oxidation-reduction reaction or redox reaction is a type of chemical reaction that involves the exchange of electrons between two chemical species. The oxidation state of an element is the number of electrons an atom loses or gain while joining other atoms in compounds<sup>[1,2]</sup>. When this oxidation number of a molecule, atom, or ion changes it is called a redox reaction. CO oxidation reaction is a redox reaction in where carbon monoxide gets oxidized, and oxygen gets reduced. In the case of the catalyst, as part of the intermediates it gets oxidized and reduced while oxidizing and reducing CO and O<sub>2</sub>. Although, CO oxidation in the presence of O<sub>2</sub> is considered a relatively simple reaction, researchers have proposed over 20 different mechanistic steps occurring during this reaction<sup>[1,16]</sup>.

Catalytic CO oxidation has gained increasing attention in recent years, primarily due to its demand in industrial processes, such as: pollution reduction in the auto industry, gaseous waste in petrochemical industries, synthesis of pure gases, ethanol or other fuel production, and pure hydrogen production. CO oxidation also has a wide range of diverse applications in long-life carbon dioxide (CO<sub>2</sub>) lasers, gas masks, catalytic converters, sensors, indoor air quality improvement, and more<sup>[6,7,8]</sup>. A high temperature operation catalyst is considered to function above 570 Kelvin. Hence, recent studies have targeted the development of new cost-effective and low-temperature catalysts<sup>[9]</sup>.

Transition metals are widely used to activate surface catalyzed reactions, such as CO oxidation. Transition metals are characterized by a specific energy level or half-filled d band which enables them to have the highest catalytic activity compared to other metals. They are efficient catalysts for CO oxidation due to their ability to dissociate molecular oxygen at a low temperature and to bind strongly with both atomic oxygen and CO. On the other hand, noble metal catalysts are known for their excellent activities, such as water tolerance, and low light-off temperatures<sup>[10]</sup>. Recent studies report that gold (Au) is highly active towards CO oxidation at low temperatures or close to ambient temperature<sup>[11]</sup>.

Noble metals such as Au can be costly, therefore, even more recent studies focus on reducing the high cost and improving the stability of catalysts, motivating the investigation for new materials such as Mn, Fe, Co, Ni, Cu or their combinations to find suitable substitutes for noble metals<sup>[12,13]</sup>. Copper being a transition metal is commonly used as a catalyst for numerous reactions, including CO oxidation. The unique ability to convert CO is the key property for supporting Cu over oxides. Cerium is considered one of the best catalysts for complete oxidation of CO, and its performance can be more superior by the inclusion of additional elements into the crystal lattice. Ceria has a huge oxygen storage capacity and redox properties; therefore, it can provide additional oxygen available for the CO oxidation reaction. Cerium oxide also acts as an oxide support, improving the catalytic performances via a metal-support interface and better diffusion of active metal components. Catalysts based on combinations between copper and cerium oxides constitute a more interesting alternative from an economical point of view and have also shown promising properties for the CO oxidation reaction.



## Transient Methods

Transient methods can be used to study the mechanisms of complex catalytic reactions. This is because transient techniques provide insight on reaction intermediates that occur from pulse responses, and the sequence of the reaction when there are multiple steps involved; two phenomena that can not be observed under steady state conditions. This is due to the transient nature of the intermediates as well as the influence of the reactants and product on the reactivity of adsorbates and catalyst surfaces.<sup>[20]</sup> In attempting to understand the overall reaction, looking at the elementary steps independently would not be sufficient. The surface of the catalyst is constantly changing during the working state, sometimes by adsorbates inducing surface restructuring, which in turn will modify the reactivity of the adsorbates.<sup>[21]</sup> While steady-state methods measure overall performance and give an integrated understanding of a reaction system, transient methods give information on individual steps in the reaction because of their operating in a millisecond time regime.

Due to the dependence of the reactivity of adsorbates on the structure of the catalyst surface, reactivity must be examined during the course of the reaction. Infrared (IR) spectroscopy is one important method that can be used to study adsorbates on catalysts.<sup>[21]</sup> Therefore, transient in situ IR experiments are widely used to study the dynamic behavior of adsorbates by applying a forcing function to the adsorbates on the catalyst in the IR cell to induce a transient state. In this experiment, the forcing function used was pulses that caused a change in the concentration of the inlet flow. Pulses are generated by varying valves.

Describing our reaction by elementary steps poses a problem because the relative rates of the steps will change as conditions change due to the transient nature of the experiment. The adsorption

capacity of the catalyst is influenced by the transient concentration.<sup>[20]</sup> Chemical analysis of the concentration of gases during transient experiments is best performed by mass spectrometry.

Ceria-supported CuO catalysts have attracted attention because of the unique properties of ceria for oxygen storage, ready reducibility, and its ability to firmly anchor Cu, reducing its tendency to sinter. Advanced transient characterization techniques such as time-resolved mass spectroscopy under operando conditions, will let us discern the redox behavior of CeO<sub>2</sub>/CuO under CO reduction, or constant-flow (steady-state) CO oxidation.

### Experimental procedure

CO and O<sub>2</sub> were fed to the reactor at a concentration of 10 and 5% respectively, using a total flowrate of 10 sccm. The amount and type of gases were measured during the reaction using time-resolved mass spectroscopy under operando or working conditions. With a differential reactor, the CO reactant gas enters at a constant flow rate, reacting at 250°C with the inverse catalyst (CeO<sub>2</sub>/CuO) and producing the CO<sub>2</sub> product. Parameters for the gas pulses calibrated loop (100μL) were set for the temperature at 30°C and 1atm for pressure. Figure 3. offers a visualization of a reactor modified for transient purposes such as the one used for this project. The pulse profile changes can be monitored using a Mass Spectrometer (MS) connected at the outlet of the reactor.

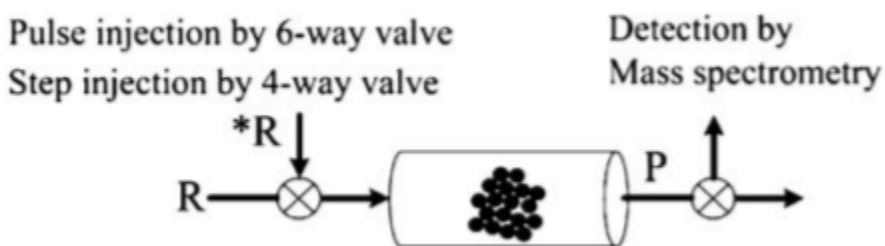


Figure 3. A differential reactor packed with catalyst particles.

The OriginPro software was used for data analysis. Methods of analysis included in the software such as an peak analyzer for integration, linear fitting, error bars for the calibration curve, and graphs were used. Microsoft Excel was used for calculation purposes and data correlation from data gathered in OrginPro. For calibration purposes, a series of 10 pulses with known concentrations (20, 40, 60, 80, 100%) were quantified. A typical calibration curve was built for every element tracked by the MS based on the known concentrations of the reactant. After a calibration curve was developed, the concentration and mass of both the reactant and product in the reaction were calculated using the ideal gas equation ( $PV=nRT$ ). The CO fragment produced during the ionization of  $CO_2$  in the MS was taken in account to precisely quantify the CO converted during the reaction. The response of these pulse changes was monitored by using the results from the MS. As an example of the tracking done by the MS, in figure 4 (a) we can see the series of 10 pulses used in the calibration curve for each concentration (20, 40, 60, 80, 100%) of  $O_2$ . In figure 4 (b), We can see the calibration curve built to find the moles of  $CO_2$  in the reaction.

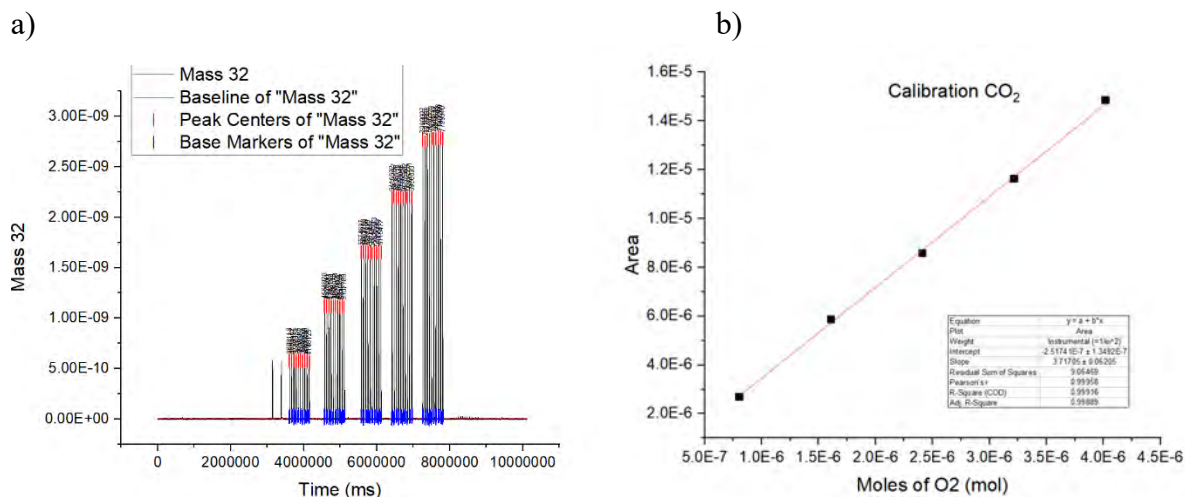


Figure 4. On the left (a) mass spectrometry data for the pulses of oxygen. On the right (b) calibration curve for  $CO_2$ .

## Results

As a result of this work, the reaction ( $2\text{CO} + \text{O}_2 \rightarrow 2\text{CO}_2$ ) on the surface of the catalyst was successfully tracked. MS measurements of CO ( $m/z = 28$ ),  $\text{O}_2$  ( $m/z = 32$ ), and  $\text{CO}_2$  ( $m/z = 44$ ) show the  $\text{CeO}_2/\text{CuO}$  reduction over 10 minutes with 30 pulses, see figure 5 and 6. Starting with CO consumption at 100%, with time, the system changes. The oxidation of CO, and reduction of  $\text{O}_2$ , decreased, and almost reached zero (Figure 5) to reach a steady-state around the 15<sup>th</sup> pulse which represents the point where the catalyst sites become less available on the surface of the catalyst. For  $\text{CO}_2$  not reaching zero or steady-state could indicate the composition of the catalyst is still evolving<sup>[17,18,19]</sup>. Figure 6, shows that the stoichiometry of the reaction for  $\text{CO}_2$  does not match with the expected value (1CO:1CO<sub>2</sub>) The total moles of  $\text{CO}_2$  produced were larger than CO being consumed. This could be due to the presence of carbonates decomposing on the surface of the catalyst during the oxidative step<sup>[17,18,19]</sup>. On the other hand, in Figure 6, we can see the different pulse times for  $\text{O}_2$  which follows the stoichiometry with respect to CO. This further correlates with the presence of carbonates on the surface of the catalyst<sup>[17,18,19]</sup>.

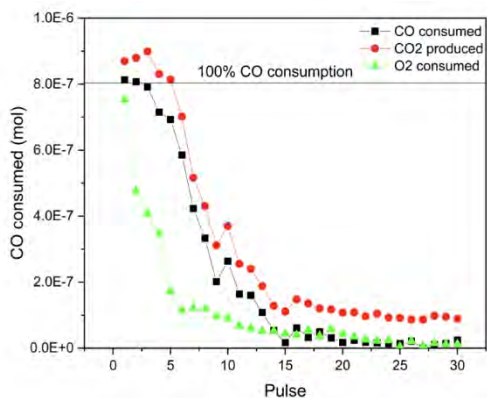


Figure 5. Amount of  $\text{O}_2$  consumed, CO consumed, and  $\text{CO}_2$  produced in pulses.

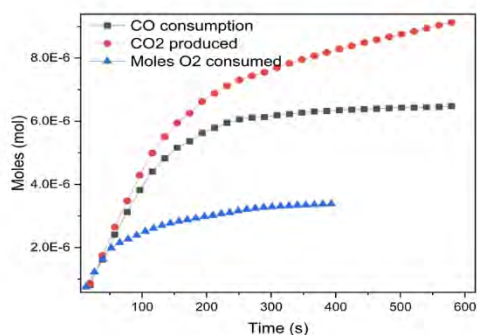


Figure 6. Amount of O<sub>2</sub> consumed, CO consumed, and CO<sub>2</sub> produced as a function of time.

## Discussion

As seen in previous experiments and literature, the moles of CO consumed, and CO<sub>2</sub> produced should be the same in this 1:1 reaction. However, our results showed substantially more CO<sub>2</sub> evolution than expected. This could be due to the composition of the catalyst still evolving with the decomposition of carbonates on the surface causing the overproduction of CO<sub>2</sub>. More characterization techniques along with sequential pulses are needed to further study the reactions on the catalyst surface and its intermediates. One such technique could be infrared spectroscopy examining the catalytic process. Another method for further research will be to vary temperature and pressure and use mass spectrometry along with the other experimentation methods to monitor how these changes affect the products generated as well as the catalyst itself. This research will be performed in the coming months by the team in the Chemistry Division at Brookhaven National Laboratory.

## Acknowledgements

We would like to thank our mentor José A. Rodriguez for recruiting us for this program, Jorge Moncada Vivas for his assistance throughout the project, and the Department of Education at Brookhaven National Lab for their guidance.

This project was supported in part by the U.S Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI)

## References

1. Hamers L. Explainer: What is a Catalyst? Science News for Students. Chemistry, 2017. <https://www.sciencenewsforstudents.org/article/explainer-catalyst-chemistry>.
2. Chemistry 302. University of Texas. Kinetics, Catalysts 2014. <http://ch302.cm.utexas.edu/kinetics/catalysts/catalysts-all.php>
3. <https://courses.lumenlearning.com/boundless-chemistry/chapter/catalysis/>
4. Al Soubaihi, R.M.; Saoud K.M.; Dutta J. Critical Review Of Low-Temperature CO Oxidation and Hysteresis Phenomenom on Heterogeneous Catalyst. Catalysts 2018, 8, 660. <https://doi.org/10.3390/catal8120660>.
5. Heck, R.M.; Farrauto, R.J.; Gulati, S.T. Catalytic Air Pollution Control; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2009. DOI:10.1002/9781118397749.
6. Wang, L.; Yang, L.; Zhang, Y.; Ding, W.; Chen, S.; Fang, W.; Yang, Y. Promoting effect of an aluminum emulsion on catalytic performance of Cu-based catalysts for methanol synthesis from syngas. Fuel Process. Technol. 2010, 91, 723, 728. <https://doi.org/10.1016/j.fuproc.2010.02.003>.
7. Grabow, L.C.; Gokhale, A.A.; Evans, S.T.; Dumesic, J.A.; Mavrikakis, M. Mechanism of the Water Gas Shift Reaction on Pt: First Principles, Experiments, and Microkinetic Modeling. J. Phys. Chem. C 2008, 112, 4608–4617. <https://doi.org/10.1021/jp7099702>.
8. Zhang, Y.; Wang, W. Recent advances in organocatalytic asymmetric Michael reactions. Catal. Sci. Technol. 2012, 2, 42–53. <https://doi.org/10.1039/C1CY00334H>
9. Valange, S.; Védrine, J. General and Prospective Views on Oxidation Reactions in Heterogeneous Catalysis. Catalysts 2018, 8, 483. <https://doi.org/10.3390/catal8100483>
10. Comotti, M.; Weidenthaler, C.; Li, W.-C.; Schüth, F. Comparison of gold supported catalysts obtained by using different allotropic forms of titanium dioxide. Top. Catal. 2007, 44, 275–284. <https://doi.org/10.1007/s11244-007-0300-1>
11. Oh, S.; Hoflund, G. Low-temperature catalytic carbon monoxide oxidation over hydrous and anhydrous palladium oxide powders. J. Catal. 2007, 245, 35–44. <https://doi.org/10.1016/j.jcat.2006.09.016>
12. Biabani-Ravandi, A.; Rezaei, M.; Fattah, Z. Study of Fe-Co mixed metal oxide nanoparticles in the catalytic low-temperature CO oxidation. Process Saf. Environ. Prot. 2013, 91, 489–494. <https://doi.org/10.1016/j.psep.2012.10.015>
13. Yu, H.; Zhong, S.; Zhu, B.; Huang, W.; Zhang, S. Synthesis and CO Oxidation Activity of 1D Mixed Binary Oxide CeO<sub>2</sub>-LaO<sub>x</sub> Supported Gold Catalysts. Nanoscale Res. Lett. 2017, 12, 579. <https://doi.org/10.1186/s11671-017-2352-x>
14. Prasad, R.; Singh, P. A Review on CO Oxidation Over Copper Chromite Catalyst. Catal. Rev. 2012, 54, 224–279. <https://doi.org/10.1080/01614940.2012.648494>.
15. Ashter, S.A. 6- Mechanics of Materials. Thermoforming of a Multilayer Laminates. Plastic Film Technologies, Testing, and Applications. 2014, 123-145. <https://doi.org/10.1016/B978-1-4557-3172-5.00006-2>.

16. Salomons, S.; Hayes, R.E.; Votsmeier, M.; Drochner, A.; Vogel, H.; Malmberg, S.; Gieshoff, J. On the use of mechanistic CO oxidation models with a platinum monolith catalyst. *Appl. Catal. B Environ.* 2007, 70, 305–313. <https://doi.org/10.1016/j.apcatb.2006.01.022>.
17. Dey, S.; Dhal, G.C. Cerium catalysts applications in carbon monoxide oxidations. *Materials Science for Energy Technologies.* 2020, 3, 6-24. <https://doi.org/10.1016/j.mset.2019.09.003>.
18. Yao, S.; Mudiyansele, K.; Xu, W.; Johnston-Peck, A.C.; Hanson, J.C.; Wu, T.; Stacchiola, D.; Rodriguez, J.A.; Zhao, H.; Beyer, K.A.; Chapman, K.W.; Chupas, P.J.; Martínez-Arias, A.; Si, R.; Bolin, T.B.; Liu, W.; Senanayake, S.D. Unraveling the Dynamic Nature of a CuO/CeO<sub>2</sub> Catalyst for CO Oxidation in Operando: A Combined Study of XANES (Fluorescence) and DRIFTS. *ACS Catal.* 2014, 4, 1650-1661. <https://doi.org/10.1021/cs500148e>.
19. Hornés, A.; Hungría, A.B.; Bera, P.; López Cámara A.; Fernández-García, M.; Martínez-Arias, A.; Barrio, L.; Estrella, M.; Zhou, G.; Fonseca, J.J.; Hanson, J.C.; Rodriguez, J.A. Inverse CeO<sub>2</sub>/CuO Catalyst As an Alternative to Classical Direct Configurations for Preferential Oxidation of CO in Hydrogen-Rich Stream. *J. AM. CHEM. SOC.* 2010, 132, 34-35. <https://doi.org/10.1021/ja9089846>.
20. Chuang, S.C.; Guzman, F.; Mechanistic Investigation of Heterogeneous Catalysis by Transient Infrared Methods. Springer. *Top Catal.* 2009, 52, 1448-1458. <https://doi.org/10.1007/s11244-009-9317-y>
21. Bennett, C. Understanding Heterogeneous Catalysis Through the Transient Method. ACS Symposium Series Vol. 178. *Catalyst Under Transient Conditions.* 1982, 1, 1-32.



# Optimizing automation of neutron resonance parameterization

Rose Marie Crawford<sup>1</sup>, Gustavo Nobre<sup>2</sup>, and David Brown<sup>2</sup>

<sup>1</sup>Department of Physics, Willamette University, Salem, OR 97301

<sup>2</sup>National Nuclear Data Center, Brookhaven National Laboratory,  
Upton, NY 11973

## Contents

<b>1 Abstract</b>	<b>1</b>
<b>2 Introduction</b>	<b>2</b>
<b>3 Motivation</b>	<b>3</b>
<b>4 Methods</b>	<b>3</b>
<b>5 Results</b>	<b>7</b>
<b>6 Conclusion</b>	<b>11</b>
<b>7 Acknowledgements</b>	<b>11</b>

## 1 Abstract

The National Nuclear Data Center (NNDC) at Brookhaven National Laboratory, through experimental processes, collects and evaluates nuclear data such as neutron resonance data compiled in the *Atlas of Neutron Resonances* [1]. Even the most precise methods lead to errors in the collection of resonance data, causing the misclassification of or gaps in resonance data. Imprecise or incomplete information about neutron resonances lead to average resonance spacings with unreliable values or unrealistic uncertainties; a problem since average spacings are important properties for many applications. To combat this, we combine nuclear physics theory with experimental data by developing an algorithm to extract resonance spacings and their uncertainties. We utilize a cumulative distribution plot of neutron resonances as a function of their energies and the linear

fit of this distribution to investigate how intentionally removed resonances impact the slope of the fit. We varied the number of resonances used to generate a fit both above and below an arbitrary cutoff resonance energy. Plotting the difference between the distribution and the list of the slopes above and below the cutoff energy showed a deviation from the calculated slope for the distribution. Since there is an inverse relationship between the average spacing and the slope of the fit, further investigation should reveal quantifiable patterns that may be used to write an algorithm to automate the detection of the location and impact of missing resonances. Availability of such a tool will allow scientists access to nuclear resonance data with increased accuracy and precision to be used in applications such as nuclear energy, astrophysics, and national security, among others. Aiding in the development of this tool has strengthened my programming skills and challenged me to be creative in the context of mathematically representing patterns and insights found in data sets.

## 2 Introduction

Nuclear data is a valuable tool used across a multitude of science disciplines, including astrophysics, nuclear power, and nuclear medicine. The nuclear data this project centers around is found in the *Atlas of Neutron Resonances* [1]: the most current compilation of nuclear resonances as well as resonance parameters and properties, including the property of resonance spacings.

A nuclear resonance is the extremely short-lived bound state of a neutron and a nucleus when the two interact or collide at a particular energy that is sufficiently close to excited states in the compound nucleus that is formed. This bound state leads to increased probability of interaction, seen graphically as cross-section peaks. The study of nuclear resonances is aided by the investigation of resonance properties such as resonance spacings. Nuclear resonance spacings are characterized as the distance between resonance peaks. Analysis of resonance spacings for a sequence of resonances as well as the average spacing of a sequence for an element provides valuable insight about the resonances themselves. Moreover, these insights expand our depth of understanding of nuclei and their properties.

However, the study of nuclear resonances is limited by much of the data in the *Atlas of Neutron Resonances* [1] being non-reproducible due to the loss of documentation on the experimental processes it sources from. One factor contributing to the non reproducibility of the data is the lack of or misclassification of resonances in many resonance sequences. Visually, these missing or misclassified resonances appear in step plots of the cumulative number of resonances as obscurely placed wider steps. Moreover, the linear fit to these plots - the inverse of the slope of which being the average spacing of a sequence - is rendered invalid due to the impact of these wider steps on its calculation.

Analyzing the cumulative distribution of resonance step plots and their linear fit lines may illuminate quantifiable relationships between missing or misclassified resonances and the calculation of resonance spacings. As such, this is

the angle of approach this project will undertake. Given the time allotted for this study, our analysis will be contained to the Chromium-52 distribution of resonances.

### 3 Motivation

The purpose of this investigation is to inform the construction of an automated algorithm to identify the location of and impact of missing or misclassified resonances, by yielding insights as to how missing resonances alter the calculation of the average resonance spacing for a given sequence.

### 4 Methods

To gain an understanding of the scope of the problem our project aims to resolve, we begin by creating a graphical reference point for all current and future work which revolves around the property of resonance spacings. This is achieved by writing a script to parse all information on mean s-wave, p-wave and d-wave level spacings for every element and isotope in the *Atlas of Neutron Resonances* [1]. The bulk of the previous work relevant to this project is embodied in gitlab in classes notation in python. Therefore, gaining familiarity with python and functions such as csv writer is necessary for the writing of this script. Once completed, we are able to use the script to generate plots of each type of level spacing as a function of their atomic number, as seen in Fig. 1. The significance of this plot lies in the fact that it was generated using the resonance data in the *Atlas of Neutron Resonances* [1], yet because this data is impacted by missing or misclassified resonances, this plot also contains these impacts or errors. Hence we have created a large-scale reference point for the problem that our subsequent work focuses on.

Next, we take on the core aspect of the project: using primarily plotting methods to explore how wider steps in a Chromium-52 cumulative distribution of resonances alter the slope of the linear fit line for the distribution. Chromium-52 is the selected distribution because of previously gained familiarity with its problems from a recently completed evaluation for its isotopes. Additionally, because it is nearly a closed shell, its level density is lower and thus it possesses an ideal number of resonances for the investigation.

The first set of plots we generate test whether manually removing resonances creates a visible change in the form of a wider step where the resonances were previously located in the distribution. We learn that removing resonances from distributions creates a wider step in the cumulative distribution plot as compared to steps that comprise the original distribution.

Success of this tactic prompts us to create two modified data sets: the first being the original sequence for Chromium-52 but with five resonances removed, and the second being with one resonance removed. We then use the original sequence and these two modified data sets to plot three cumulative distribution

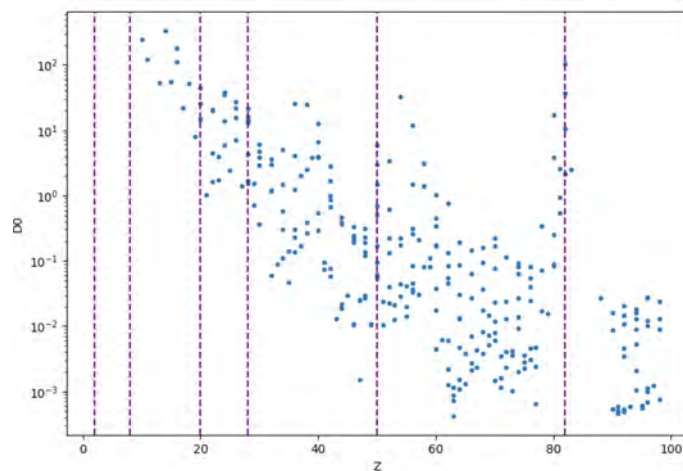


Figure 1: This plot shows the mean s-wave level spacings for  $L=0$  resonances for all nuclides, as function of atomic number. These data were collected from the *Atlas of Neutron Resonances* [1]. The dashed lines show where resonances increase at  $Z = 2, 8, 20, 28, 50, 82$  (the nuclear magic numbers where the number of protons or neutrons in a nucleus reach a closed shell).

plots with their linear fits. We can then take the difference between each distribution and its linear fit and form a superimposed plot that shows the three differences as a function of the resonance energies, as seen in Figures 2 and 3. We also plot the location of the manually removed resonances as points on the plot. We can magnify different regions in the figure to inspect any impacts the removed resonances have on the difference curves. We discover that there is a large downward spike in the curve created with the distribution that had five resonances removed, near the location where the resonances were removed. However, the difference curve made using the distribution with only one resonance removed does not exhibit similar behavior. Therefore we determine that a different approach may provide more insight into the behaviors seen in the curves.

We decide to create two new types of difference plots, made by looping through every resonance on each of the three cumulative distributions, and calculating two slopes at each resonance energy point: one slope that takes into account all of the resonances below the moving cutoff point, and one which factors in all of the resonances above. These slopes are then compiled into two separate lists of slopes, one list solely for the slopes generated above the cutoff point and one for slopes below. From here, six unique difference plots are made, where two separate difference curves are calculated for each data set: one which takes the difference between the cumulative distribution and the list of slopes below the moving cutoff point, and the other which does the same calculation

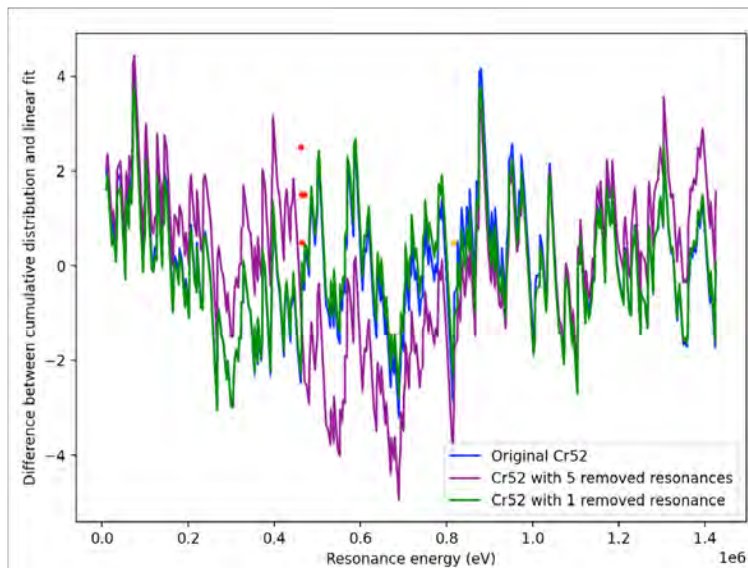


Figure 2: This plot shows the difference between the cumulative distribution and linear fit as a function of the resonance energies for all three data sets. The red dots indicate where five consecutive resonances were removed from the sequence, and the orange dots indicate where one resonance was removed. The purple curve has a large downward spike in region where five resonances were removed. In contrast the green curve shows no noticeable deviation in response to single removed resonance as compared to curves from other sequences.

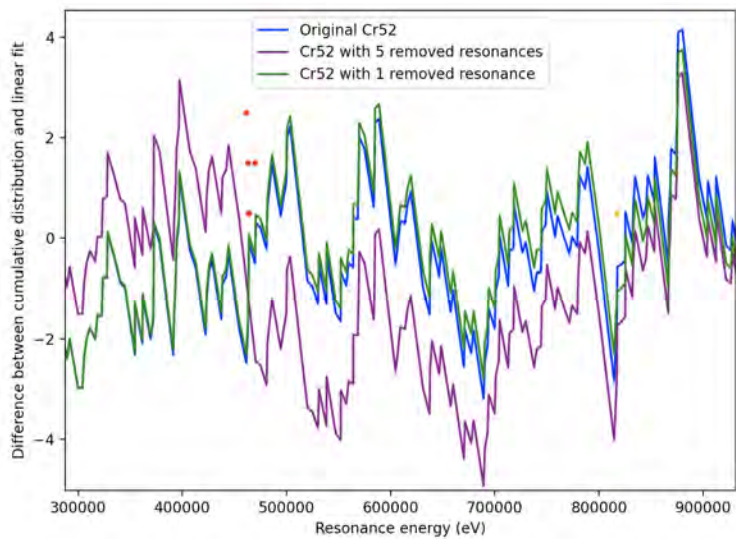


Figure 3: This plot contains the same curves from the previous figure, limited to the region around the locations where the resonances were removed in distributions such that the contrasting behavior of the purple and green curves can be seen.

with the list of slopes above the moving cutoff point. This plot can be seen in Figures 4 and 5. Finally, three curves are calculated for each data set by taking the difference between the two linear fits just calculated for each distribution, and these differences are then plotted as functions of the resonance energies, as shown in Figures 6 and 7. The location of the removed resonances are again labeled on both plots as points.

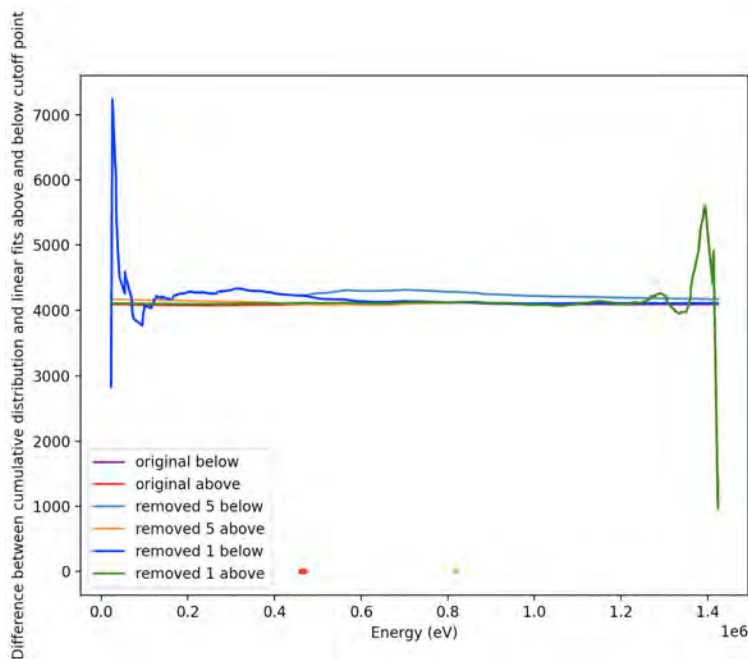


Figure 4: This plot shows six different curves, where there are two curves created using each distribution. The first curve was calculated by taking the difference between the cumulative distribution and its linear fit made from the list of slopes above the moving resonance energy point, and the second curve was made using the same calculation but with the linear fit created from the list of slopes below the moving resonance energy point.

## 5 Results

Regarding the difference plots made using slopes calculated from the entire distribution for each data set (Figures 2 and 3), we note an apparent downward spike in the curve calculated from the distribution with five resonances removed, likely due to the five resonances removed in the region. However, the curve calculated from the distribution with a single resonance removed does not exhibit similar behavior. Therefore we cannot deem, what is speculated to be a

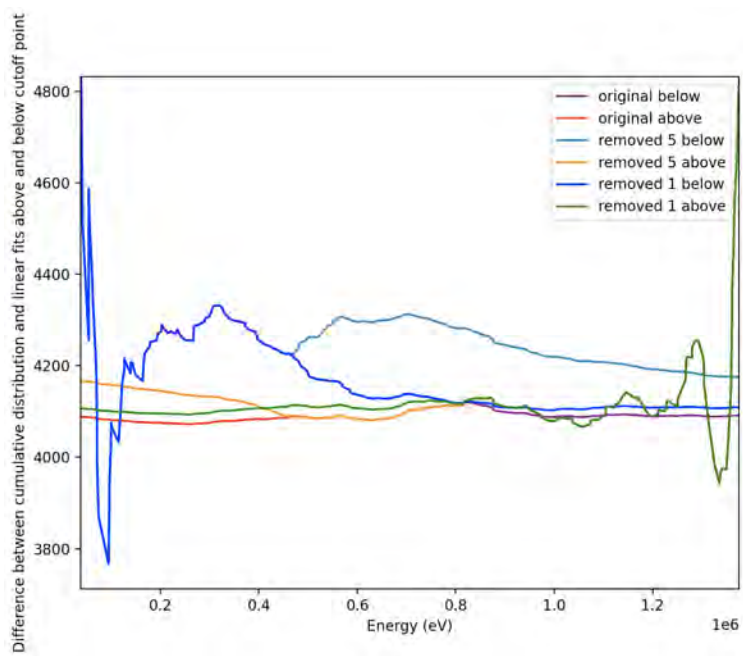


Figure 5: This figure is a closer view of Figure 4 such that the behavior of the six curves is more easily displayed.



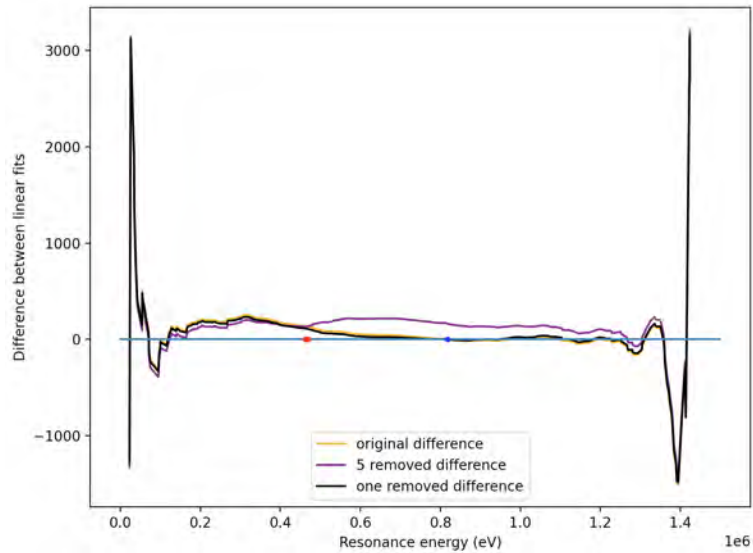


Figure 6: This plot shows three superimposed curves, one for each data set, which are the difference between the lists of linear fits above and below the continually changing cutoff point as a function of the resonance energies.

response of the distribution and linear fit to the five removed resonances, as a conclusive result.

Instead, we turn to the results seen in Figures 6 and 7. We are most interested in the curves shown in these plots because upon initial inspection, they provide more apparent evidence that manually removed resonances lead to a behavioral response graphically (despite amount of resonances removed or location of their removal), as compared to the difference plots previously made as well as compared to the plots in Figures 4 and 5. Yet we note that given more time, further analysis of the curves in Figures 4 and 5 may yield valuable insights.

In the view of the curves provided by Fig. 7, it is clear that the path of the curve made using the data with five removed resonances deviates significantly from the path of the other two curves at the location where the resonances were removed. This behavior is verified by similar behavior of the curve made using the data with a single resonance removed at the location where that resonance was removed, although that deviation is considerably smaller in magnitude. However, we realize that these results maintain a level of ambiguity due to the possible interference that unknown missing resonances located across the distribution may have on the calculation of our slopes at the regions where we have intentionally removed resonances.

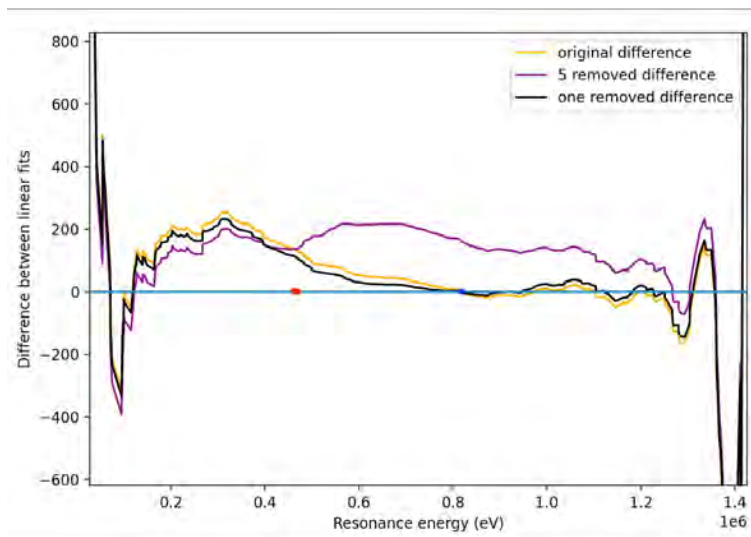


Figure 7: This plot is a magnified view of Figure 6 such that the behavior of the three curves is more visible. The purple curve shows a large deviation from other two curves at location where five resonances were removed. The black curve also shows slight deviation at location where the single resonance was removed, not dissimilar from deviation of purple curve.

## 6 Conclusion

We have investigated the systematic behavior of average spacings for neutron resonances, by compiling all information available in the *Atlas of Neutron Resonances* [1]. Additionally, we have explored multiple potential methods to identify the impact of missing resonances in cumulative distributions on the calculation of resonance spacings. This involved calculating global fits and running averages of the distributions below and above a moving resonance energy cutoff point. The analysis of the difference of such slopes provided many insights, even though not conclusive, relative to the possible ways of automating the search for missing resonance. Moving forward, we pursue similar analysis but with a more localized approach, calculating an averaging spacing for a distribution only in a region around the investigated range. Hopefully, this will lead to an automated, reproducible and more reliable way to obtain average spacings.

## 7 Acknowledgements

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). In addition I would like to thank my mentors Dr. Gustavo Nobre and Dr. David Brown, as well as Dr. Matteo Vorabbi, Dr. Amber Coles, for all of their help and support throughout the internship. I also want to extend my appreciation to my fellow interns Mary Fucci, Sergio Ruiz, and Sergey Scoville.

## References

- [1] S. F. Mughabghab, *Atlas of Neutron Resonances, Vol.1 and 2*. Elsevier, Amsterdam (2018).

# Where's Waldo on the beamline: Using machine learning to automate sample detection

Kaleigh Curtis\*, Thomas Caswell<sup>¶</sup>, and Daniel Olds<sup>¶</sup>

\* *Department of Materials Science and Engineering, Texas A&M University, College Station, TX 77843*  
<sup>¶</sup> *National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY 11973*

**Abstract.** The NSLS-II at Brookhaven National Laboratory supplies a broad spectrum of light that can be used to uncover the structure and properties of materials. Some of the beamlines in the NSLS-II specialize in x-ray diffraction for material characterization purposes, and due to the intense brightness of the beam, these measurements can occur much more quickly than on most diffraction instruments. However, presently these samples must be manually aligned and positioned at the beamline to be measured correctly. This alignment and position step bottlenecks how many samples can be processed on the beamline. This research attempts to solve this problem by implementing automated clustering algorithms so that after a fast, rough scan of all samples, the computer can decide what areas are important and can then remeasure the samples to get a much better picture of what the sample's diffraction pattern truly is. This would replace manual alignment with automation and would save hundreds of hours on the beamline, allowing for more research to be done in the same amount of time. Functions have been created that manipulate diffraction data into a readable format, group the diffraction patterns using clustering algorithms, and visualize the scanned data using intuitive, easy-to-use methods. Being able to find what should be scanned and recorded without human input can be used on many different projects within the NSLS-II and will be a valuable tool for all future users.

## I. INTRODUCTION

The NSLS-II is a synchrotron light source that is ten billion times brighter than the sun and it supplies wavelengths of light that can be used for a variety of research applications. Due to the brightness of the beam, the data acquisition time for these applications can be cut down drastically compared to other facilities and allows many more samples to be measured than at other light sources.

One application of the NSLS-II is x-ray diffraction, which is an important material characterization technique<sup>1</sup>. By scattering x-rays from a sample, we can measure the atomic and nanoscale structure of the material. X-ray diffraction can be used on many kinds of samples, including powders, liquids, and even solid disordered materials. Because of the range of samples that can be tested, and because of natural variation in sample preparation, each sample must be individually aligned so that the beam passes through the sample instead of the bracket that holds the samples or the sample container. Manual alignment is time consuming, as each sample can take several minutes to align and up to 200 can be loaded on the beamline at once. Thus, alignment takes hours, and it causes a bottleneck in data acquisition. Powder samples, as seen in Figure 1, are approximately 1 mm in diameter, and this size makes manual alignment even more taxing. A misaligned sample, scattering off either the sample holder or air, results in wasted beam time and errors in analysis if the invalid data is not identified and removed.

If you are measuring a powder sample packed vertically in a capillary, it is possible the powder has either crystallized or is packed differently in one spot, and that can also result in low quality data. A way to remedy this is to oscillate the bracket vertically during data collection and collect diffraction data through a range of positions on the sample, averaging the total measured spectras. However, this additional step can be time-consuming, as it further requires determining the vertical bounds of each individual sample within the bracket.

This project aims to create a model using Python that can find these sample bounds so a beamline scientist can collect more diffraction data more quickly, saving hours of time and eliminating this bottleneck in data acquisition.



FIGURE 1. Image of samples in a frame.

## II. METHODS

### A. Data Mapping

To begin, the input data must be transformed into a consistent format. The inputs that were used to test all code were a TAR file and a text file, and between these files, we had the x position, y position, q-space over which data was collected, and the intensity data for each datapoint. We found that Python's dictionary data structure was the most appropriate frame to use for this mapping because it was, in our opinion, the most compact way to connect a coordinate to a diffraction pattern.

We first collected all unique y-positions, and had those values be the keys for the outermost layer of the dictionary. For each unique y-position, we made each unique x-position that existed at that y-position a key, and then we made the value for that key the diffraction data at that coordinate. The result was a dictionary of dictionaries, layered to have y data be the outermost layer and x data be the innermost layer. This mapping process combined all the data into one object, which was easier to manipulate in code and allows for different kinds of input files to be used in the future if they are transformed into this dictionary as well. A visual representation of this mapping is shown in Figure 2.

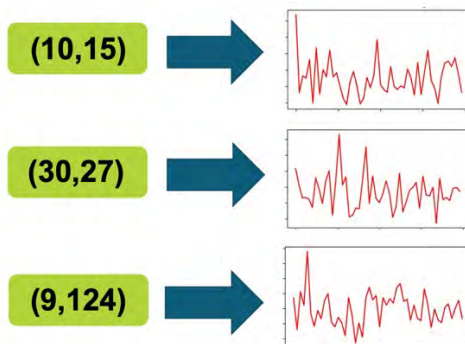


FIGURE 2. A visualization of data mapping, with each data point (shown as x, y coordinates) corresponding to a unique diffraction dataset. The example diffraction data, along with all other plots in this paper, were created using Matplotlib<sup>2</sup>.

### B. Clustering

In x-ray diffraction, data is collected over a set range of scattering angles, and these angles make up the x-axis of a diffraction dataset. We can convert these angles into a scalar<sup>3</sup>,  $q$ , from the scattering angle,  $2\theta$ , and the wavelength,  $\lambda$ , using Equation (1). Since the range of these  $q$ -values was the same for all datapoints, we did not include the  $q$ -values in the clustering dataset, since they would have no impact.

$$q = \frac{4\pi \sin(\theta)}{\lambda} \quad (1)$$

The first clustering algorithm we tried was K-Means. This algorithm is a method of vector quantization in which each observation is grouped with the cluster with the nearest mean, effectively replicating Voronoi cells. For each observation, the objective is to find the equality<sup>4</sup> in Equation (2).

$$\arg \min_s \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg \min_s \sum_{i=1}^k |S_i| \text{Var } S_i \quad (2)$$

Another algorithm we tested was Agglomerative Hierarchical Clustering. This bottom-up method starts with every observation being its own cluster with clusters merging as one moves up the hierarchy. This algorithm requires a distance metric and a linkage metric, and we chose the Euclidian distance<sup>5</sup> and Ward linkage<sup>6</sup> as our metrics. The equations for these metrics are shown in Equation (3) and Equation (4). Each algorithm we tested was available in scikit-learn's clustering package and we took additional steps to ensure that the output for every algorithm would be the same. When we tried both algorithms, we noticed that the K-Means algorithm in scikit-learn<sup>7</sup> had a preset number of clusters at 8, and Agglomerative Hierarchical Clustering had a preset number of clusters at 2. In our application, this meant that K-Means had a more fixed number of clusters while Agglomerative Hierarchical Clustering is more flexible.

$$\|a - b\|_2 = \sqrt{\sum_i (a_i - b_i)^2} \quad (3)$$

$$d_{ij} = \|X_i - X_j\|^2 \quad (4)$$

The variables  $a$  and  $b$  in Equation (3) represent two points in Euclidian space and  $d_{ij}$  in Equation (4) represents the distance between observations  $i$  and  $j$ .

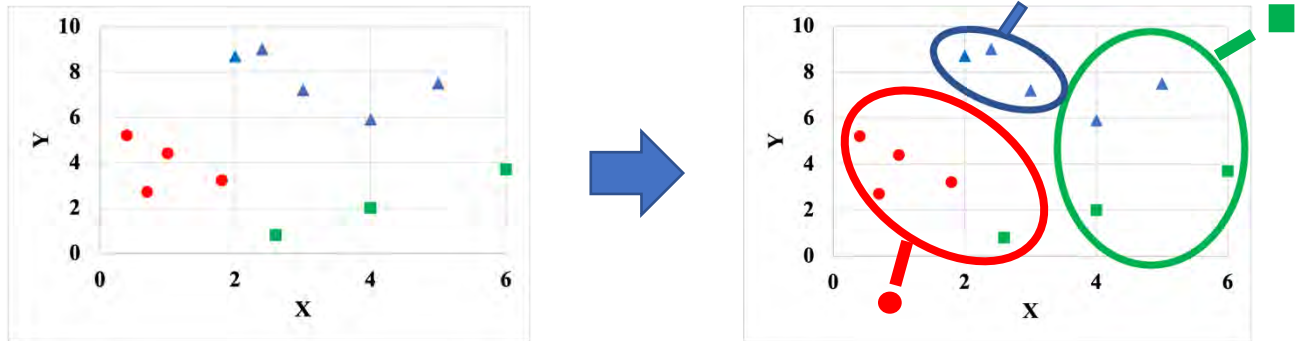
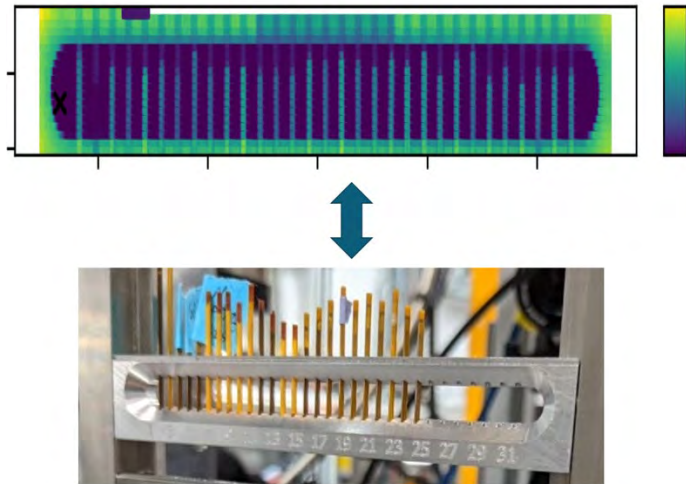


FIGURE 3. Example of K-Means clustering using randomly generated data.

### C. Visualizing

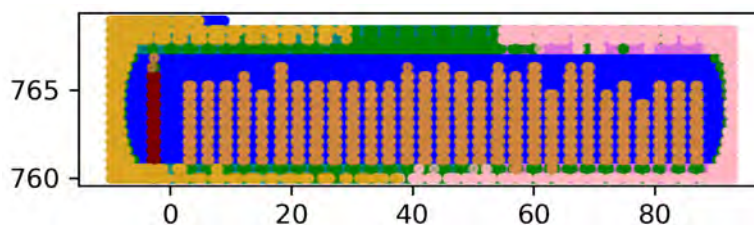
As this project is essentially moving the visualization of the data from human eyes to a computer's observations, we viewed it important to allow the user to view what the computer had done. By creating a program that can visualize how well a selected cluster compares with the dataset, one can see how well the clusters match the data and what parts of the dataset were grouped together. A visualization method that both shows the user how accurate the clustering algorithms were and replicates the samples that were measured is plotting residuals on a heatmap.



**FIGURE 4.** Example of visualization that shows how a residual plot compares to a real image of a similar dataset.

### III. RESULTS

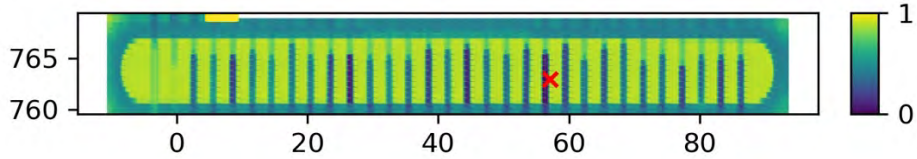
We found that the clustering algorithms found the samples within the dataset and also effectively classified different materials within the dataset. Observing previous research<sup>8</sup>, we saw that others implemented dissimilarity metrics so that they could run diffraction data through clustering algorithms, but we found that dissimilarity metrics were not necessary for this project. The data was already distance-based due to the nature of q-space, and by running both K-Means and Agglomerative Hierarchical Clustering on our test data, we saw that the algorithms clustered similar materials together even when dissimilarity metrics were not used, so we decided to eliminate the step. As our data was unlabeled, we were not able to compare the accuracy of using dissimilarity metrics versus not using these metrics, and further work will need to be done to validate the decision that we made.



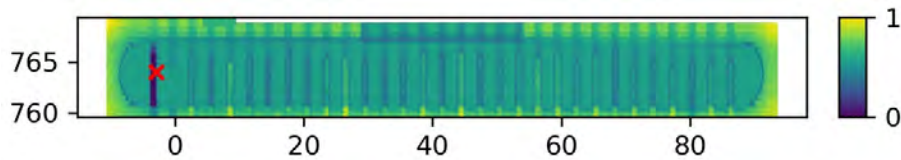
**FIGURE 5.** The result of running the dataset through the K-Means algorithm and plotting each cluster group. Each capillary sample can be clearly seen and distinguished from the background of air. Other features are visible in this plot, specifically areas where the aluminum frame holding the samples has areas of lower density.

By observing our residual plots, we were able to see that each cluster did find dissimilar materials within the dataset. Overall, we saw that our models were effective and could identify sample regions to a point where a later scan could view these regions in more detail.

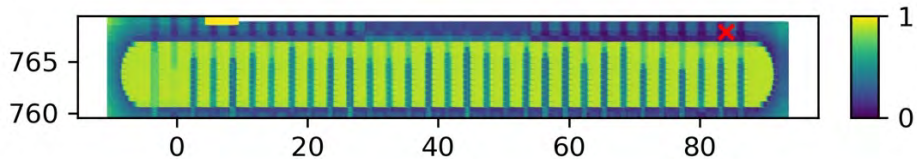
By using pick events in Matplotlib, which allows the user to click on a plot and returns the coordinates of where the user clicked, we were able to use the cluster data at those coordinates to create a residual plot. These residual plots are shown in Figure 6, Figure 7, and Figure 8. For each of these figures, the darker in color a point is, the closer the data at that point matches the data at the 'x,' which is where the user clicked. All three of these plots show that the material at the location that the user clicked is clustered with itself, that is, the clustering algorithm correctly distinguished each material. These plots were generated from the cluster data from the K-Means algorithm, but the same results were seen when using the Agglomerative Hierarchical Clustering algorithm as well.



**FIGURE 6.** A residual plot for the entire dataset when compared to the data point at the ‘x,’ which is a sample. The red ‘x’ in the figure is a point where the user clicked, and the color bar represents how closely the cluster assigned to the point at ‘x’ matches the data at all other points, with darker colors meaning the two points match more than lighter color points.



**FIGURE 7.** A residual plot for the entire dataset when compared to the point at ‘x,’ which is the control sample. The red ‘x’ in the figure is a point where the user clicked, and the color bar represents how closely the cluster assigned to the point at ‘x’ matches the data at all other points, with darker colors meaning the two points match more than lighter color points.



**FIGURE 8.** A residual plot for a point found on the frame that holds the samples versus the rest of the dataset. The red ‘x’ in the figure is a point where the user clicked, and the color bar represents how closely the cluster assigned to the point at ‘x’ matches the data at all other points, with darker colors meaning the two points match more than lighter color points.

All the code that was created to generate these results is available on GitHub and is cited in the references<sup>9</sup>.

## IV. CONCLUSION

Overall, we conclude that clustering algorithms such as K-Means and Agglomerative Hierarchical Clustering are effective in clustering diffraction data. Our overall process was to restructure the data into a map, cluster the data, and then visualize the results, all of which were successful. For future research and work, we recommend that an interface is created that makes running this code easier. Future work also needs to validate our choice to process the diffraction data without using any dissimilarity metrics. We also recommend that this code is run using labeled data in order to quantify how accurate these clustering methods are.

## V. ACKNOWLEDGEMENTS

Thanks to Dr. Daniel Olds and Dr. Thomas Caswell for their assistance and guidance through this project.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).



## VI. REFERENCES

1. Daniel Olds (2020) Synchrotron X-ray Diffraction for Energy and Environmental Materials: The Current Role and Future Directions of Total Scattering Beamlines in the Functional Material Scientific Ecosystem, *Synchrotron Radiation News*, 33:5, 4-10, DOI: [10.1080/08940886.2020.1812352](https://doi.org/10.1080/08940886.2020.1812352)
2. J. D. Hunter, "Matplotlib: A 2D Graphics Environment", *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-95, 2007.
3. Henry E. Fischer, "From Q-Space to R-Space: Introduction to PDF Analysis," October 4, 2016, Institut Laue-Langevin, Grenoble, [https://www.ill.eu/fileadmin/user\\_upload/ILL/6\\_Careers/1\\_All\\_our\\_vacancies/PhD\\_recruitment/Student\\_Seminars/2017/19-2017-05-09\\_Fischer\\_Cookies.pdf](https://www.ill.eu/fileadmin/user_upload/ILL/6_Careers/1_All_our_vacancies/PhD_recruitment/Student_Seminars/2017/19-2017-05-09_Fischer_Cookies.pdf)
4. Stuart P. Lloyd, "Least Squares Quantization in PCM," in *IEEE Transaction on Information Theory*, vol. IT-28, no. 2, pp. 129-137, March 1982.
5. Karl Smith, "Precalculus: A Functional Approach to Graphing and Problem Solving," Jones & Bartlett Publishers, 2013, pg. 8.
6. Joe H. Ward Jr. (1963) Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58:301, 236-244, DOI: [10.1080/01621459.1963.10500845](https://doi.org/10.1080/01621459.1963.10500845).
7. Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, *JMLR*12, pp. 2825-2830, 2011.
8. Iwasaki, Y., Kusne, A.G. & Takeuchi, I. Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries. *npj Comput Mater* **3**, 4 (2017). <https://doi.org/10.1038/s41524-017-0006-2>
9. K. Curtis, XRD-Clustering, 2021, GitHub repository, <https://github.com/kaleighacurtis/XRD-Clustering>

# High-level synthesis tools and methodologies for embedded AI application-specific integrated circuits

Bshara Dababneh<sup>1</sup> and Sandeep Miryala<sup>2</sup>

<sup>1</sup>*Department of Engineering, State University of New York at New Paltz, New Paltz, NY 12561*

<sup>2</sup>*Instrumentation Division, Brookhaven National Laboratory, Upton, NY 11973*

(Dated: August 11, 2021)

## Abstract

The Instrumentation Division (IO) of Brookhaven National Laboratory (BNL) is at the forefront of developing new design methodologies for mapping machine learning algorithms to the transistor level circuit implementations targeting edge computing in scientific applications. Machine learning developers currently use open-source frameworks such as PyTorch or TensorFlow based on high-level programming languages such as Python/C++ to design and train a neural network. BNL is investigating High-Level Synthesis (HLS) tools to transform Python/C++ based machine learning algorithms into Hardware Description Language (HDL). HLS provides capabilities to automatically convert C/C++/System C code into accurate and optimized HDL models. Designers specify the constraints such as clock frequencies so that the HLS tool optimizes the data path for minimal hardware resources. These new methodologies based on HLS tools reduce the overall time to design and fabricate Artificial Intelligence (AI) based Application-Specific Integrated Circuits (ASICs). I am a part of this research of a revolutionary project with a primary goal of developing HLS methodologies, or flows, to translate a neural network in C++ to HDL. Due to limitations in project duration, we investigated just the convolution layer of a Convolutional Neural Network (CNN) in this project. Our project contributes to the ongoing research by laying the foundation for developing HLS methodologies for hardware platforms: ASICs and Field Programmable Gate Array (FPGA). My engagements at BNL have increased my knowledge and experience in the design, implementation, and optimization of machine learning algorithms from a circuit design perspective. Further, it has reinforced the theoretical knowledge imparted in engineering classrooms. Overall, the project has presented me with the opportunity to participate in cutting-edge technology, as is evident in the accomplishments of this advanced research.

## I. INTRODUCTION

Artificial Intelligence (AI) has made significant strides in closing the gap between human and machine capabilities. Researchers and enthusiasts alike work on various aspects of the field to achieve powerful results [1]. The domain of computer vision is one of many such areas. Computer vision could be a thriving subset of the phony AI market [2]. It is attracting a significant amount of funding because it targets a wide range of practical applications. Deep Learning advancements in Computer Vision have been built and perfected over time, primarily through a single algorithm – a Convolutional Neural Network (CNN).

Furthermore, BNL's Instrumentation Division (IO) is working on new design methodologies translating machine learning algorithms to the transistor level implementation to enable edge computing in scientific applications. During this summer research program, we looked into commercial HLS tools from Siemens, Catapult. HLS shortens production cycles by introducing additional abstraction on top of Register Transfer Level (RTL), typically utilizing System C, C, or C++ to specify the project specifications before synthesis [2]. It is frequently beneficial due to a beneficial alignment between the hardware and recursive-style environments.

Machine learning developers prefer to write their algorithm code in C++ or scripting languages such as Python. They do not have to look into semantics associated with the register transfer language like Verilog, Very High-Speed Integrated Circuit (VHSIC), or Hardware Description Language (HDL). Some rule developers write their code in C++ and use HLS to decide how to vary the clock speed to address this limitation [3]. Once the clock constraint is changed, the HLS tool regenerates the RTL based on the new clock speed, with no changes to the C++ code.

The primary goal of utilizing an HLS tool like Catapult is to provide superior

capabilities to automatically convert code descriptions into accurate and optimized hardware language models. Furthermore, designers specify the constraints such as clock frequencies to optimize the HLS tool for the least amount of hardware resources [4]. As a result, these new methodologies shorten the time required to design and fabricate AI based ASICs.

While the rationale union employs affiliation, RTL depiction of the plan, significant level blend works to ensure a degree of reflection, beginning characterization during very issue-specific languages like System C, C, and C++. For the most part, the architect promotes module reasonableness and, as a result, interconnect convention. The undeniable level of union apparatuses handles the miniature engineering. It converts untimed or incompletely ordinary intentional code into completely regular RTL executions, precisely making cycle-by-cycle detail for equipment execution [5]. The RTL executions are then used directly during a ubiquitous rationale union stream to frame a door-level commission, as shown in Fig. 1.

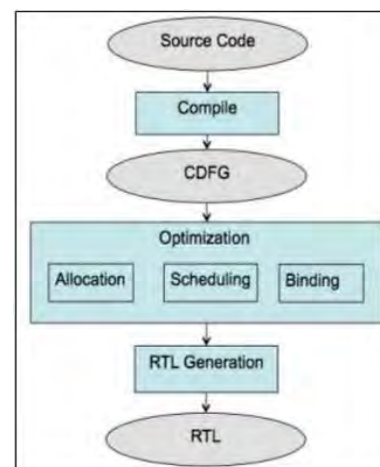


Fig. 1. RTL execution process

Autopilot is a contemporary HLS device built using AutoESL1 that accepts System C, C, and C++ input languages that may communicate with directives directing the equipment's overall execution. Autopilot supports a collection of C/C++ programming languages; the primary alternatives that are not supported are dynamic

memory allocation and discretionary indirection [6]. Autopilot supports both numerical and skimming reason data assortments, as well as subjective exactitude fixed-point data assortments. Autopilot extensively uses expected compiler improvements such as dead-code elimination, strength reduction, and execution inclination. After achieving these code changes, the mix is conducted on a work-by-work basis, generating RTL modules for each job. Each module's administrative logic is based on DataPath and Finite State Machine (FSM) [6]. Naturally, all data clusters have the intention to use local Block Random Access Memories (BRAMs); scalar factors use registers. Autopilot will improve five distinct categories of source code: correspondence interfaces, work calls, circles, information exhibitions, and named locations.

Autopilot performs targeted improvements, such as counting articulation smoothing, circle unrolling, circle smoothing, and apparent cluster splitting. In any event, Autopilot is cautious in applying these enhancements, allowing the user to customize the plan for space, clock speed, turnout, or a few combinations of these [7]. Each enhancement to Autopilot has documentation as `#pragma` comments and `combo` script commands. Following code changes, Autopilot practices the code to the appropriate execution stage using information about the execution stage. At that point, the equipment amalgamation method maps the streamlined code to the equipment, performing calculation programming, asset restriction, and pipelining [8]. Finally, Autopilot generates the interface code in this manner. Due to the immediate execution, the integrated code maintains a consistent communication interface. For circles are solid; naturally, they need attention to the first chance for asset sharing. Autopilot commands will indicate a whole or partially unrolled circular body, a mixture of many circles, or a combination of settled circles [6]. After obtaining information in exhibits, loop

unrolling is often used in conjunction with exhibit apportionment to enable various equal independent displays to be accessed, followed by creating a similitude possibility linked to pipelining potential. Additionally, the circle requirements, as shown, will demonstrate articulation parity for further developing fine-grained parallelism and computation pipelining inside a code region.

## II. METHODS

### A. CNN & ReLU Function

The convolutional layer is the fundamental building component of a CNN that requires a huge number of computations [9]. As shown in Fig. 2, a generic CNN is composed of three primary layers: Convolutional, Pooling, and Fully Connected (FC).

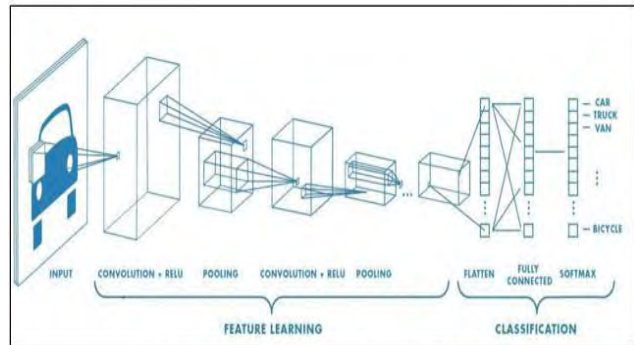


Fig. 2. Convolutional Neural Network (CNN)

Our job was to study the convolutional layer with the Rectified Linear Unit (ReLU) as an activation function. As shown in Fig. 3, ReLU is a nonlinear function that acts as a "filter." By using the ReLU function, the whole layer becomes dependent on the sign bit. If the input value is negative (less than zero), it will be snapped to zero. This number is "Dead Neurons." However, if the input value is positive (higher than zero), the value is returned unaltered. This number is "Active Neurons." We discovered that it is critical to use this nonlinear function since it does not stimulate all neurons concurrently and works well than other functions.

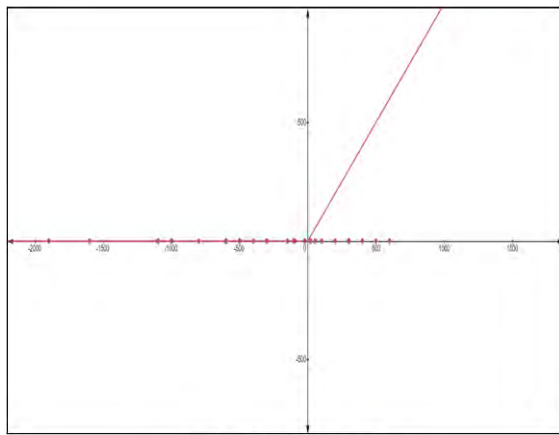


Fig. 3. ReLU function graph

### B. Testbench in Digital Design

Creating a testbench was a critical stage in our study. As shown in Fig. 4, the testbench is utilized in digital design simulations from the viewpoint of Test Engineers. The testbench is used to determine whether or not our Design under Test (DUT) module functions as expected [10].

The output checker compares DUT output with expected values to report PASS/FAIL. It was utilized to rapidly inform the user whether changes to design components would break the whole design before applying the changed component inside a design.

Although the testbench module is mainly developed and used in Verilog, our testbench module is created in C++. Thus, the testbench's compelling reasons for use are as follows: detecting human mistakes, showing data on a terminal, and confirming findings through a waveform.

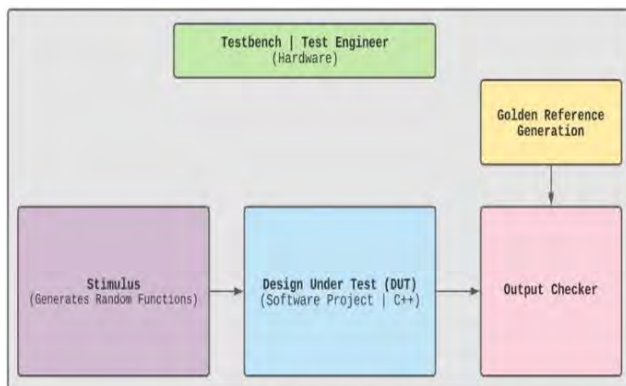


Fig. 4. Testbench block verification

## III. RESULTS & DISCUSSION

We needed to determine whether or not our convolutional layer functioned adequately at the software and hardware levels.

### A. Software Level

Three matrices were used in our convolutional layer: input image, kernel, and output image. Different dimensions were given to these matrices. For example, we set the input image to the [8x8] matrix and the kernel to the [3x3] matrix; we end up with an output image of the [6x6] matrix.

The software level was determined using the output of the GNU Compiler Collection (GCC). Random numbers between -100 and 100 were used to create the input image values. By contrast, the kernel was constructed using only positive numbers between 0 and 100. This is done to enhance the likelihood of obtaining negative values. As a consequence of the ReLU function, negative values return the final result as zero, as shown in Fig. 5.

```

Transcript
0 Errors 1 Warnings 45 Infos 240 Comments
Message
-----
Input Image
#
# -81 0 47 84 55 -3 -33 29
# 5 -93 -47 72 34 -15 43 0
# 98 -73 -9 -18 -44 67 86 -94
# -77 -59 82 -98 60 -21 81 29
# 80 -43 -38 -16 54 30 63 -42
# -13 -85 80 -80 -50 -27 -29 -53
# -50 62 -20 6 -21 -35 13 2
# -44 -56 12 66 23 -28 -5 2
#
-----
Kernel
#
# 32 5 15
# 58 77 72
# 95 8 38
#
-----
Final convolution of BOTH matrices
#
# -3758 -10025 5344 9116 4787 6096
# -6266 -16723 3640 -6227 18307 3070
# 5343 -11256 -578 -1851 14236 8489
# -1811 -19875 8586 -6531 6640 -1176
# -4778 -378 -8256 -11154 -6823 -16115
# -3379 -3748 1742 -1614 -3376 -5313
#
-----
Test Passed
-----
Output after applying ReLU Function
#
# 0 0 5344 9116 4787 6096
# 0 0 3640 0 18307 3070
# 5343 0 0 0 14236 8489
# 0 0 8586 0 6640 0
# 0 0 0 0 0
# 0 0 1742 0 0 0

```

Fig. 5. GCC Output – Software Level

As shown in Fig. 5, our design succeeded when we received the expected output "Test Passed," which led us to apply the ReLU function to our plan. All negative numbers were returned as zero, while positive values remained unchanged.



### B. Hardware Level

After translating the C++ code into Verilog HDL, the corresponding simulation results shown in Fig. 6, all values were allocated in hexadecimal to facilitate calculation.

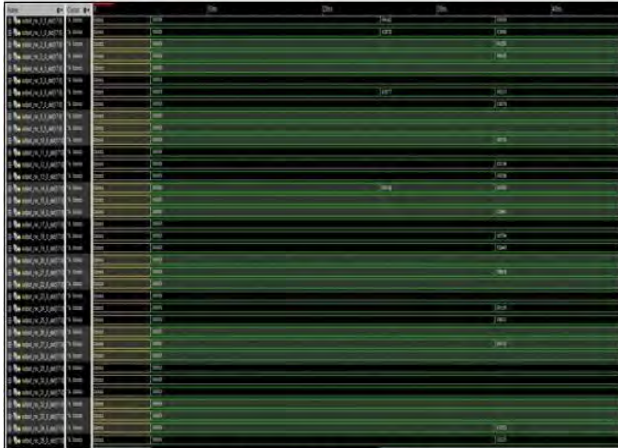


Fig. 6. Verilog's waveform – Hardware Level

Staring at the waveform was sort of hard to distinguish the different values. As a result, we set all the waveform values in the matrices block, as shown in Fig. 7.

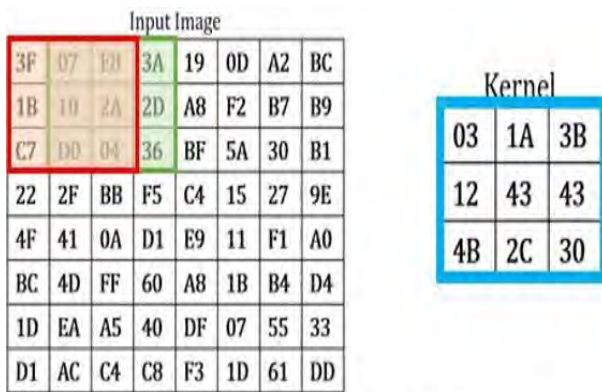


Fig. 7. Input Image [8x8] with Kernel [3x3] matrices

The large portion in the preceding example matches our [8x8] input image. The kernel is the element that performs the convolution operation in the first section of a convolutional layer. It can be seen on the right side with a [3x3] dimension.

When the filter is moved to the right by a specified Stride Value, the whole width is parsed. It then descends to the start (left) of the supplied picture. As a result, we get the following values, which match the red box in Fig. 8.

$$\begin{aligned}
 &(03)(3F) + (1A)(07) + (3B)(E8) + \\
 &(12)(1B) + (43)(10) + (43)(2A) + \\
 &(4B)(C7) + (2C)(D0) + (30)(04) \\
 &A6CC = -9,932 \\
 &00000
 \end{aligned}$$

Fig. 8. Returning zero value due to negative input value

Following the convolution process, we ended up with A6CC in hexadecimal, equivalent to -9,932 in unsigned numbers. As a result, in our Verilog's waveform, this value was returned as 00000.

On the other hand, once we moved the kernel to the right by one step, the green box shows. As shown in Fig. 9, we got a positive result, 01FA8 in hexadecimal. Since it is a positive number, this value was returned unchanged.

$$\begin{aligned}
 &(03)(07) + (1A)(E8) + (3B)(3A) + \\
 &(12)(10) + (43)(2A) + (43)(2D) + \\
 &(4B)(D0) + (2C)(04) + (30)(36) \\
 &01FA8
 \end{aligned}$$

Fig. 9. Returning the same value due to positive input value

## IV. DESIGN DIAGRAMS

### A. Schematic Design

After being synthesized using a Catapult HLS tool, the schematic design of the convolutional layer is illustrated below in Fig. 10. The schematic design aims to convert the convolutional layer into actual representations of space. The resultant of HLS tools is an HDL program (VHDL or Verilog). Both schematics that are presented here reflect the same schematic design. The only difference is that the full view of the design is on the left side, and on the right side is the zoom-in view. As can be seen, the computations inside a convolutional layer are comprised of two fundamental operations, which are addition and multiplication, also known as Vector Matrix Multiplication (VMM). Our schematic design contains red circles, green circles, and purple boxes. The red circles indicate the addition

operation, while the green circles represent the multiplication operation; lastly, the purple boxes represent flip-flops for storing intermediate signals.

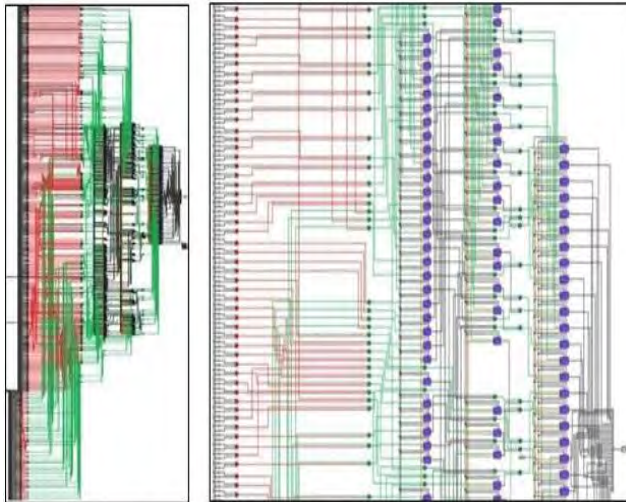


Fig. 10. Convolutional Layer Schematic Design

### B. RTL Schematic

Similarly, the RTL schematic shows our design’s inputs, outputs, clock, and reset ports, as shown in Fig. 11.

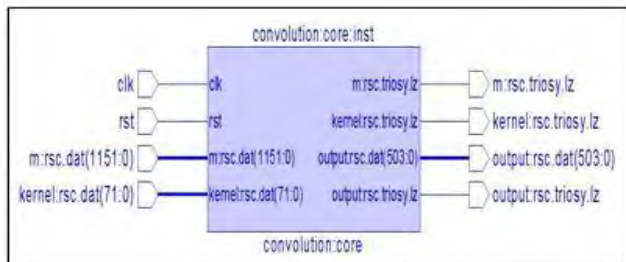


Fig. 11. RTL Schematic

## V. IMPROVING THE HARDWARE PERFORMANCE

Increasing the performance of the convolutional layer increases hardware productivity. We have two distinct kinds of loops in the Catapult HLS tool: Pipelining and Unrolling. As shown in Fig. 12, when the convolutional layer design is executed for the first time on its own using the Pipelining loop, it takes 36 clock cycles. In other words, our hardware design requires 360 nanoseconds to complete the circuit from the first input to the first output. As a result, we needed to allow both loops to communicate and operate together to enhance hardware performance.

Consequently, we increased hardware performance by decreasing clock cycles from 36 to 1 or 360 nanoseconds to 10 nanoseconds. This has improved the hardware performance by 97%, as shown in Fig. 12. As a result, the more complex the design project, the more loops will be used throughout the design implementation. As a result, the possibility of enhancing the hardware design is very significant.

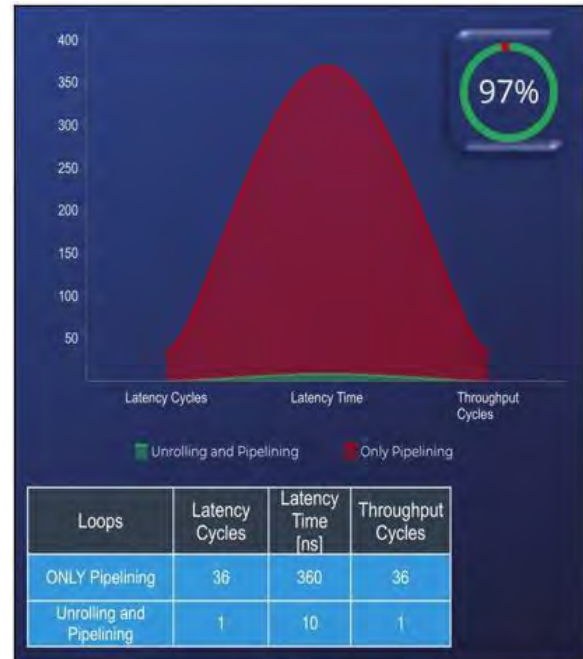



Fig. 12. Improving the hardware performance of the convolutional layer

## VI. ASIC Vs. FPGA

We have two hardware platforms in VeryLarge-Scale Integrated (VLSI): Application-Specific Integrated Circuit (ASIC) and Field Programmable Gate Array (FPGA). FPGAs may perform a Central Processing Unit (CPU) functions, an encryption unit, a graphics card, or even all three simultaneously [1]. On the other hand, ASICs, as the name suggests, are intended to perform a specific function and are not modifiable. The HLS synthesized HDL could be used for both FPGAs and ASICs. As a result, TABLE 1 summarizes the distinctions between ASIC chips and FPGA boards.



TABLE 1  
The difference between ASIC Vs. FPGA

Hardware Platform	ASIC Chip	FPGA Board
HLS Tool	CATAPULT	CATAPULT 
Flexibility	Manufactured for <b>ONE</b> specific application <b>WITHOUT</b> any modifications after it is produced	Manufactured to perform <b>numerous</b> functions after they have left the manufacture
Design Flow	Complex Fabrication	Simplistic Fabrication
Performance	Faster Speed	Slower Speed
Power Consumption	Less Power	More Power
Efficiency	Higher Level	Lower Level
Time to Market	Long Time	Short Time

Consequently, by using the Catapult HLS tool for ASIC and Xilinx Vivado HLS tool for FPGA, we determined that ASIC performance is considerably more excellent and more efficient than FPGA performance, as shown in Fig. 13. It is observed that ASIC outperforms FPGA in three different C++ projects: addition, multiplication, and convolution.

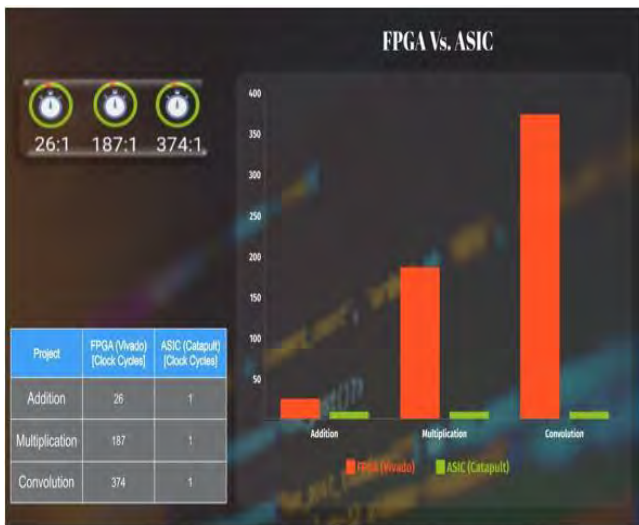


Fig. 13. Showing that ASIC is better than FPGA from a performance perspective

## VII. DIGITAL LOGIC-BASED DESIGN

After our project was successfully compiled without problems, we produced 2,372 lines of Verilog code. The convolutional layer's introductory module is shown in Fig. 14. We were unable to display the complete code owing to a large number of lines.

```

HLS HDL:      Verilog Netlist
HLS Version:  10.5c/898140  Production Release

Design Unit:  convolution - Design Under Test (DUT)

module convolution (
    clk, rst, m_rsc_dat, m_rsc_triosy_lz, kernel_rsc_dat,
    kernel_rsc_triosy_lz, output_rsc_dat, output_rsc_triosy_lz);
    input clk;
    input rst;
    input [2807:0] m_rsc_dat;
    output m_rsc_triosy_lz;
    input [287:0] kernel_rsc_dat;
    output kernel_rsc_triosy_lz;
    output [1151:0] output_rsc_dat;
    output output_rsc_triosy_lz;

    // Interconnect Declarations for Component Instantiations
    convolution_core convolution_core_inst (
        .clk(clk),
        .rst(rst),
        .m_rsc_dat(m_rsc_dat),
        .m_rsc_triosy_lz(m_rsc_triosy_lz),
        .kernel_rsc_dat(kernel_rsc_dat),
        .kernel_rsc_triosy_lz(kernel_rsc_triosy_lz),
        .output_rsc_dat(output_rsc_dat),
        .output_rsc_triosy_lz(output_rsc_triosy_lz)
    );
endmodule

```

Fig. 14. Convolutional Layer – Verilog code using Catapult HLS tool

Furthermore, in a digital ASIC design, the first stage is the logic synthesis that was carried out using Cadence Genus Synthesis Solution Tool. The synthesis tool translates Verilog code to a gate-level netlist, according to Fig. 15. All the schematics below reflect a similar schematic design. The difference is that on the right side is the zoom-in view, and on the left side is the full view.

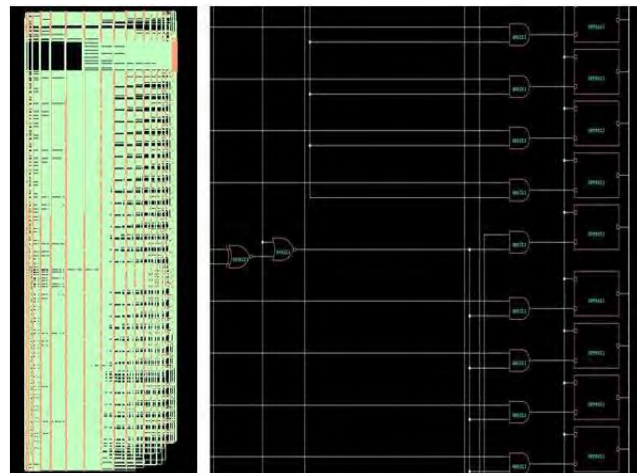


Fig. 15. Convolutional Layer in Digital Logic Based Design using Genus Tool

In the world of chip design, there is a constant reminder of improving the chip for better performance, improved area, and low power. Therefore, from Fig. 15., we were able to generate different logic gates like AND, NOR, and XNOR in replacement of multiplication and addition operations.



## VIII. CONCLUSION

To summarize, this sophisticated and challenging research endeavor synthesized all of the information gained from our work with CNN. We demonstrated using the HLS Catapult tool; we converted convolutional layer algorithms written in C++ to HDL. Additionally, we showed that HLS is better to automatically translate code descriptions to precise and efficient hardware models. Using this sophisticated approach, developers may significantly decrease the time required to design and fabricate AI based ASICs. Lastly, we also carried out logic synthesis to translate Verilog code to gate-level netlist using the Genus tool. When all factors are considered, this endeavor may be regarded as a success. Not only did this result in a more nuanced and developed knowledge of the convolutional layer, but it also resulted in the development of essential problem-solving abilities and advanced logical thinking. In any case, by starting from scratch, our efforts paid off. The difficult skills required for the convolutional layer were created, which will be essential to possess while pursuing careers in both Computer and Electrical Engineering.

## IX. ACKNOWLEDGMENTS

Special thanks to Dr. Miryala for all his knowledge and guidance on this advanced project. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). Many thanks to Siemens and Cadence companies for the excellent service and valuable seminars provided by their delightful employees. Finally, I genuinely express my appreciation to BNL for allowing me to be part of their Instrumentation Division team.

## X. REFERENCES

- [1] Singh, R., 2021. FPGA Vs. ASIC: Differences between them and which one to use. Numato Lab Help Center. [online] Available at: <https://numato.com/blog/differences-between-fpga-and-asics/>
- [2] BDTI. 2010. The AutoESL AutoPilot High-Level Synthesis Tool. [online] bdti.com. Available at: <https://www.bdti.com/MyBDTI/pubs/AutoPilot.pdf>
- [3] Education, I., 2021. What are Convolutional Neural Networks?. [online] Ibm.com. Available at: <https://www.ibm.com/cloud/learn/convolutional-neural-networks>
- [4] Moore, D., 2018. Testbenches. [online] Youtube.com. Available at: <https://www.youtube.com/watch?v=UGWZilMG8yY>
- [5] Star, Z., 2019. The Applications of Matrices. [online] Youtube.com. Available at: <https://www.youtube.com/watch?v=rowWM-MijXU>
- [6] High-Level Synthesis, C., 2020. Floating-point C++ Algorithm to Optimized RTL Implementation Using Catapult HLS. [online] Youtube.com. Available at: <https://www.youtube.com/watch?v=HkBF0xqcx3I>
- [7] XilinxInc. 2016. Vivado High-Level Synthesis. [online] Available at: <https://www.youtube.com/watch?v=hZ2RGwLmXc0&t=311s>
- [8] Deeplizard. 2017. Convolutional Neural Networks. [online] Available at: [https://www.youtube.com/watch?v=YRhxVdVk\\_sIs](https://www.youtube.com/watch?v=YRhxVdVk_sIs)
- [9] Saha, S., 2021. A Comprehensive Guide to Convolutional Neural Networks. [online] Medium. Available at: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
- [10] Coding, P., 2021. Purpose of Testbench in VHDL. [online] Forums.xilinx.com. Available at: <https://forums.xilinx.com/t5/Simulation-and-Verification/Purpose-of-Testbench-in-VHDL-coding/td-p/1085424>

## XI. APPENDIX

### ❖ The Convolutional Layer C++ code

```
//=====
// Name      : ConvoTest.cpp
// Author    : Bshara Dababneh & Sandeep Miryala
// Description : Testbench project for verification purposes – Convolutional Layer
//=====
#include<iostream>
#include "MatrixConvo.h"
#include<assert.h>
#include <cstdlib>
#include <mc_scverify.h>
#include<ac_int.h>

using namespace std;

//Generate random numbers for "Input Image"
void generateRandomMatrixInput(ac_int<8,true> m[8][8])
{
    for(int i=0; i<8; i++)
    {
        for(int j=0; j<8; j++)
            //random numbers between (-100, 100)
            int range = 100 - (-100) + 1;
            m[i][j] = rand() % range + (-100);
        }
    }

//Generate random numbers for "Kernel"
void generateRandomMatrixKernel(ac_int<8,true> k[3][3])
{
    for(int i=0; i<3; i++)
    {
        for(int j=0; j<3; j++)
            k[i][j] = rand() % 101; //random number between (0, 100)
        }
    }

//Through this method, the entire convolution process goes on
void testConvolution(ac_int<8,true> m[8][8], ac_int<8,true> kernel[3][3],
                    ac_int<18,true> output[6][6])
{
    for (int i=0; i<6; i++)
    {
        for (int j=0; j<6; j++)
        {
            output[i][j] = m[i][j]*kernel[0][0]      + m[i][j+1]*kernel[0][1]      +
                m[i][j+2]*kernel[0][2]      + m[i+1][j]*kernel[1][0]      +
                m[i+1][j+1]*kernel[1][1]      + m[i+1][j+2]*kernel[1][2]      +
                m[i+2][j]*kernel[2][0]      + m[i+2][j+1]*kernel[2][1]      +
                m[i+2][j+2]*kernel[2][2];
        }
    }
}

//Verify whether the test is passed or not
void test(ac_int<18,true> m1[6][6], ac_int<18,true> m2[6][6])
```

```

{
    for(int i=0; i<6; i++)
    {
        for(int j=0; j<6; j++)
            assert(m1[i][j] == m1[i][j]);
    }
    cout << "          Test Passed          " << endl;
    cout << "-----" << endl;
}

//Output method for the input image
void printInput(ac_int<8,true> a[8][8])
{
    for(int i=0; i<8; i++)
    {
        for(int j=0; j<8; j++)
            cout << a[i][j] << " ";
        cout << endl;
    }
}

//Output method for kernel
void printKernel(ac_int<8,true> b[3][3])
{
    for(int i=0; i<3; i++)
    {
        for(int j=0; j<3; j++)
            cout << b[i][j] << " ";
        cout << endl;
    }
}

//Output method for the final result
void printOutput(ac_int<18,true> convo[6][6])
{
    for(int i=0; i<6; i++)
    {
        for(int j=0; j<6; j++)
            cout << convo[i][j] << " ";
        cout << endl;
    }
    cout << "-----" << endl;
}

CCS_MAIN(int argv, char **argc)
{
    //Test
    ac_int<8,true> m1[8][8];
    ac_int<8,true> kernel[3][3];
    ac_int<18,true> output[6][6];
    ac_int<18,true> ref[6][6];
    for(int i=0; i<2; i++)
    {
        generateRandomMatrixInput(m1);
        cout << "-----" << endl;
        cout << "          Input Image          " << endl;
        cout << "-----" << endl;
        printInput(m1);
        generateRandomMatrixKernel(kernel);
        cout << "-----" << endl;
    }
}

```

```

        cout << "          Kernel          " << endl;
        cout << "-----" << endl;
        printKernel(kernel);
        convolution(m1, kernel, output);
        cout << "-----" << endl;
        cout << "Final convolution of BOTH matrices" << endl;
        cout << "-----" << endl;
        printOutput(output);
        testConvolution(m1, kernel, ref);
        test(output, ref);
    }
    return 0;
}

//=====
// Name      : MatrixConvo.cpp
// Description : Matrix file -- Design Under Test (DUT) | Convolution
//=====
#include "MatrixConvo.h"
#include <mc_scvrify.h>
#include<ac_int.h>

#pragma hls_design top
void CCS_BLOCK(convolution)(ac_int<8,true> m[8][8], ac_int<8,true> kernel[3][3],
                           ac_int<18,true> output[6][6])
{
    loop1:for (ac_int<4,true> i=0; i<6; i++)
    {
        loop2:for (ac_int<4,true> j=0; j<6; j++)
        {
            output[i][j] = m[i][j]*kernel[0][0] + m[i][j+1]*kernel[0][1] +
                m[i][j+2]*kernel[0][2] + m[i+1][j]*kernel[1][0] +
                m[i+1][j+1]*kernel[1][1] + m[i+1][j+2]*kernel[1][2] +
                m[i+2][j]*kernel[2][0] + m[i+2][j+1]*kernel[2][1] +
                m[i+2][j+2]*kernel[2][2];

            //Calling applyReLU function within convolution method
            output[i][j]=applyReLU(output[i][j]);
        }
    }
}

//Applying Activation function applyReLU
int applyReLU (ac_int<18,true> convo)
{
    if(convo < 0)
        return 0;
    else
        return convo;
}

//=====
// Name      : MatrixConvo.h
// Description : Header file
//=====
#ifndef matrix_h
#define matrix_h
#include<ac_int.h>

void convolution(ac_int<8,true> m[8][8], ac_int<8,true> kernel[3][3], ac_int<18,true> output[6][6])
int applyReLU(ac_int<18,true> output);

#endif

```

# Electron Ion Collider Superconducting Radio Frequency Cavity Radiation Potential and Coupled-Cell Energy Balance Analyses

William Daniels<sup>1, a)</sup> and Zachary Conway<sup>2, b)</sup>

<sup>1</sup> *Departments of Science and Engineering, Suffolk County Community College, Selden, NY 11784*

<sup>2</sup> *Electron Ion Collider Division, Brookhaven National Laboratory, Upton, NY 11973*

<sup>a)</sup> [wdaniels@bnl.gov](mailto:wdaniels@bnl.gov)

<sup>b)</sup> [zconway@bnl.gov](mailto:zconway@bnl.gov)

**Abstract.** Brookhaven National Laboratory (BNL) will be upgrading its Relativistic Heavy Ion Collider (RHIC) to the Electron Ion Collider (EIC) in the future. In such devices, Superconducting Radio Frequency (SRF) cavities are used to accelerate and guide beams of particles, driving collisions for experiments. Before being installed in the EIC, SRF cavities will be tested to ensure they meet the performance requirements for a successful upgrade. Since SRF cavities can cause dangerous levels of radiation while in operation, it is important to have appropriate protective shielding to protect BNL staff from the health detriments of radiation, which include but are not limited to an increased risk of most cancers. Additionally, SRF cavities must meet tight tolerances in their operating frequencies, necessitating a tuning process. During my term as an intern at BNL, I created preliminary estimates for radiation shielding requirements for testing EIC SRF 5-cell cavities and developed MATLAB code to facilitate the tuning process by automating necessary calculations.

## RADIATION POTENTIAL

SRF cavities are susceptible to the field emission of electrons, a process in which electrons are emitted from cavity surfaces due to the presence of particulates and impurities on the cavity surface under the influence of high electric fields. While testing cavities, the process will result in electrons being emitted into the accelerating fields generated by the cavities. This has the potential to bring the electrons to a high enough energy such that the electrons could interact with the cavity and test enclosure walls to generate potentially dangerous levels of radiation.

### Source of Radiation

Once accelerated by the cavity's electric field, the electrons will collide with either the cavity itself, or the walls of the test enclosure. In either case, the collision will result in radiation producing interactions. The three main radiation producing interactions are the Cherenkov effect, Bremsstrahlung effect, and characteristic emissions.

The Cherenkov effect is caused by charged particle exceeding the speed of light in that medium [1]. This is not in violation of special relativity because it requires that nothing with mass may exceed the speed of light in a vacuum; the optical properties of niobium and concrete (the anticipated materials of the cavity and test enclosure) include indices of refraction greater than one. In these materials, Cherenkov radiation will account for some of the lowest portions of the emission spectrum.

Bremsstrahlung (also called braking) radiation is caused by the deceleration of charged particles in matter. This phenomenon has both classical and quantum explanations, both of which rely on the principle of conservation of energy. In either case, the result is a continuous spectrum that starts and ends at the bounds of the accelerated electrons energy. Note that this puts an upper bound on the maximum energy of radiation that the SRF cavity testing setup can produce, the maximum energy of a field emitted electron.

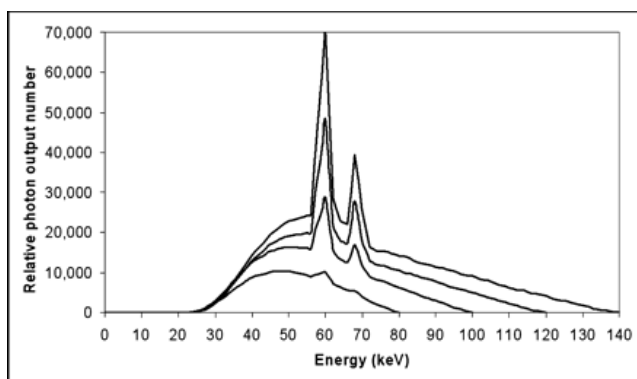


FIGURE 1. X-ray emission spectrum of a tungsten x-ray tube anode [2].

Finally, we may discuss characteristic emissions. These appear in emission spectrums as sharp peaks at specific energies. These emission lines are caused by the excitation of certain orbitals. As per quantum theory, electrons in atoms can only absorb certain amounts of energies that result in allowable changes in quantum numbers. Therefore, as shown in Fig. 1, the peak locations are invariable after changes beam energy.

### Maximum Energy of Field Emitted Electrons

Since higher energy x-rays propagate further through matter than lower energy ones, it follows that shielding from the highest energy radiation will in turn shield from all the radiation. Thus, in order to calculate the amount of shielding necessary for the SRF cavity testing enclosure, we must calculate the highest energy radiation. The highest energy radiation in any emission spectrum will be equal to the highest energy bremsstrahlung radiation, which is in turn equal to the highest energy a field emitted electron can achieve in the anticipated cavities.

TABLE 1. Important characteristics of anticipated EIC 5 cell SRF cavity

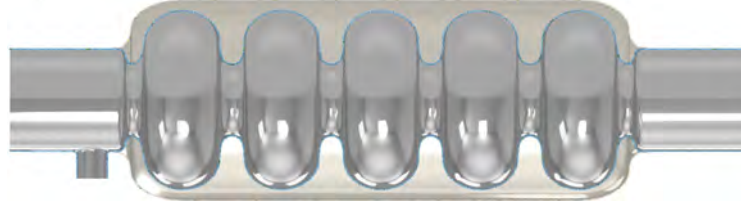
Accelerating Gradient, $E$	25 MV/m
RF Frequency, $f$	591 MHz
Number of Cells, $N$	5
Effective Length, $l_{eff}$	1.267 m

The cavity that stands out as having the greatest ability to accelerate field emitted electrons to high energy is a 5 cell 591 MHz niobium cavity. This is because its purpose will be to accelerate electron beams in the EIC at a high beta. To calculate the maximum energy a field emitted electron will reach in this cavity, we assume the following:

1. Electrons are emitted in phase with RF fields operating in the cavity's pi-mode,
2. Electron travel on axis, and
3. Electrons are emitted at the speed of light.

Assumption (1) is obvious enough not to require justification, but assumptions (2) and (3) are best justified. We begin by justifying assumption (2).

In SRF cavities, electric fields are strongest closest to cavity surfaces. So, it is reasonable to expect that an electron that spends as much time possible near cavity surfaces would gain the most energy. Due to the cavities magnetic fields, traveling near the surface will result in a deflection towards the cavity's center axis. This ultimately results in an electron traveling in a cyclic path. In this treatment, we assume the path is sinusoidal and lies entirely in a two-dimensional cross section of the cavity. The reason this electron trajectory will not result in a greater electron energy is due to the path length difference. For the on-axis path, the path length is 0.507 m and for the sinusoidal path it is 0.641 m in a single cell. For the case where the electron travels at the speed of light, the electron spends 1.69 ns in the call for a straight path and 2.14 ns for a sinusoidal path. The RF period of the 591 MHz cavity is 1.69 ns, so by traveling the sinusoidal path, the electron is decelerated by the cavity's fields for 21% of its path through the cavity. This discrepancy increases as the electron speed decreases. Thus, the assumption that the electron travels on axis will result in the highest energy electron.



**FIGURE 2.** A 3D model of an anticipated EIC 5 cell cavity with titanium jacket.

We now justify assumption (2), that an electron is emitted at the speed of light. To do so we will use an approach inspired by Delayen working on a similar problem [3]. We will demonstrate that the electron being emitted at the speed of light has only a marginal effect on its energy. To do so, we first calculate how much time it takes for the electron to reach a beta of 0.9.

$$\begin{aligned}
 mdv &= Fdt \\
 m \int_0^{0.9c} dv &= qE \int_0^t \sin(\omega t) dt \\
 0.9mc &= -\frac{qE}{\omega} (\cos(\omega t) - 1) \\
 -0.9 \frac{mc\omega}{qE} + 1 &= \cos(\omega t) \\
 \frac{1}{\omega} \arccos\left(-0.9 \frac{mc\omega}{qE} + 1\right) &= t \\
 t &= \frac{1}{2\pi \cdot 591 \times 10^6} \cdot \arccos\left(-0.9 \cdot \frac{9.11 \times 10^{-31} \cdot 3 \times 10^8 \cdot 2\pi \cdot 591 \times 10^6}{1.602 \times 10^{-19} \cdot 25 \times 10^6}\right) \\
 &= 0.186 \text{ ns}
 \end{aligned}$$

The time it takes for the electron to reach beta 0.9 is 0.159 ns, which is only 11% of the cavity's RF period. Over the five cells, this is only 2.20 % of the time spent in the cavity, which is an acceptably low correction for this calculation.

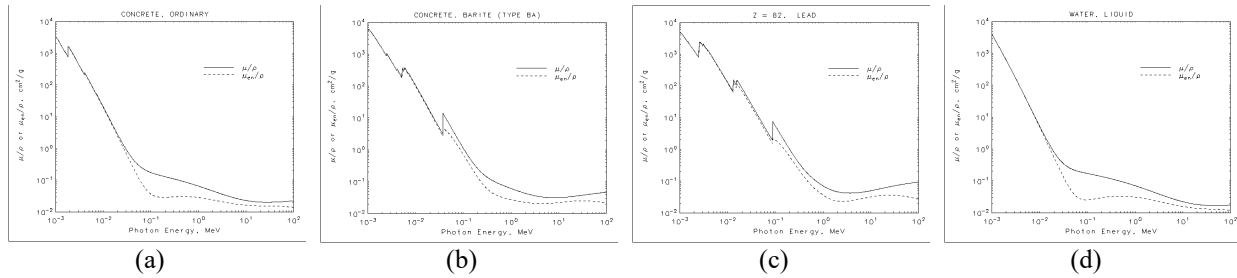
Finally, we can calculate the maximum kinetic energy an electron can achieve in the EIC 5 cell cavity using the work energy principle.

$$\begin{aligned}
 K_{max} &= NE \int_0^{l_{eff}} \sin\left(\frac{\omega x}{c}\right) dx \\
 &= -\frac{NEc}{\omega} \left(\cos\left(\frac{\omega l_{eff}}{c}\right) - 1\right) \\
 K_{max} &= -\frac{5 \cdot 25 \times 10^6 \cdot 3 \times 10^8}{2\pi \cdot 591 \times 10^6} \left(\cos\left(\frac{2\pi \cdot 591 \times 10^6 \cdot 1.267}{3 \times 10^8}\right) - 1\right) \\
 &= 20.194 \text{ MeV}
 \end{aligned}$$

We have approximated the maximum energy achievable by a field emitted electron to be 20.194. As a final check to the assumptions made, I also performed this analysis on the Jefferson National Accelerator Facility CEBAF upgrade [4]. It was found that my method applied to the CEBAF cryomodules produced an overestimation of about 10%.

## Radiation Shielding Requirements

With our estimate for the maximum energy a field emitted electron can obtain, we are able to calculate the amount of shielding necessary to protect workers from the cavity during testing. The attenuation of x-rays through matter is



**FIGURE 3.** Plots of mass attenuation coefficients for (a) ordinary concrete, (b) barite concrete, (c) lead, and (d) water with respect to photon energy [5].

dependent upon the physical properties of the matter. The relationship is directly proportional to the density and a quantity called the mass attenuation coefficient (depicted in Fig. 3 for two types of concrete), which described the probability of photon interactions in the matter [6]. Because field emission is (as far as we currently know) fundamentally random, the only way to approximate radiation dose from maximum energy and other cavity parameters is to extrapolate from measured dose rates of cavities already in existence. This will be a rather rough approximation because SRF cavities will vary widely in their number of cells, physical dimensions, and field profiles. The problem is enhanced by the novel nature of the EIC. Nonetheless, again using measured data from the CEBAF upgrade, the highest dose was caused by the cavity C100-1. It was a 7-cell cavity operating at 16 MV/m to produce a dose of 10,000 mrem/hour. Impact energies as high as 11 MeV were measured for the cavity at 16 MV/m.

**Table 2.** Comparison of anticipated EIC 5-cell cavity to CEBAF C100 cavities

Cavity Parameter	EIC 5-cell Cavity	CEBAF C100 Cavity
Accelerating Gradient	25 MV/m	16 MV/m
RF Frequency	591 MHz	1497 MHz
Number of Cells	5	7
Effective Length	1.267 m	0.701 m
Electron Impact Energy	20.194 MeV	11 MeV
Dose	<i>Extrapolate</i>	10,000 mrem/hr

Table 2 summarizes the differences between the cavities. I will reference these quantities in my justification for how I extrapolate the dose rate. The RIC cavity has a greater electric field strength than the CEBAF cavity, which should increase field emission. The EIC cavity is also physically larger, which means more surface area for field emitters and impurities to exist on, which should also increase field emission. The EIC cavity also has a higher maximum impact energy, which would enhance neutron production (niobium, the cavity material in both cavities, has a neutron production threshold of 8.33 MeV). Wanting for similar data for various cavities to attempt a regression analysis, I will assume that dose will scale linearly with accelerating gradient.

$$\begin{aligned} \frac{E_{EIC}}{d_{EIC}} &= \frac{E_{CEBAF}}{d_{CEBAF}} \\ d_{EIC} &= d_{CEBAF} \frac{E_{EIC}}{E_{CEBAF}} \\ d_{EIC} &= 10,000 \cdot \frac{25}{16} \\ d_{EIC} &= 15625 \text{ mrem/hr} \end{aligned}$$

With a dose rate estimated for the EIC cavity, we can now estimate the amount of shielding necessary to reduce that rate to the DOE/BNL acceptable level of 2 mrem/hr. The attenuation relationship is represented with a differential equation. It is a direct relationship between dose and the mass attenuation coefficient  $\mu/\rho$  and density.



$$\begin{aligned}
-\frac{dI}{I} &= \mu/\rho \rho dx \\
-\int_{I_0}^I \frac{dI}{I} &= \mu/\rho \rho \int_0^x dx \\
-\ln\left(\frac{I}{I_0}\right) &= \mu/\rho \rho x \\
x &= \frac{-\ln\left(\frac{I}{I_0}\right)}{\mu/\rho \rho}
\end{aligned}$$

At this point, it is useful to provide the necessary shielding thickness for multiple potential shielding materials. I will provide thickness for ordinary concrete, barite concrete, lead, and water; the first three materials for being common shielding materials and water due to the potential for a water door for the test enclosure.

Ordinary Concrete	Barite Concrete	Lead	Water
$x = \frac{-\ln\left(\frac{2}{15625}\right)}{2.030 \times 10^{-2} \cdot 2.4}$ $= 184 \text{ cm}$	$x = \frac{-\ln\left(\frac{2}{15625}\right)}{3.439 \times 10^{-2} \cdot 3.5}$ $= 74 \text{ cm}$	$x = \frac{-\ln\left(\frac{2}{15625}\right)}{6.206 \times 10^{-2} \cdot 11.3}$ $= 13 \text{ cm}$	$x = \frac{-\ln\left(\frac{2}{15625}\right)}{1.813 \times 10^{-2} \cdot 1.0}$ $= 494 \text{ cm}$

We see from the above calculations that anywhere from 13-184 centimeters of shielding will be necessary to reduce the anticipated dose rate to 2 mrem/hr if it is constructed from a combination of ordinary concrete, barite concrete, and lead. Our calculation indicates that 494 cm of water is necessary to attenuate the radiation to 2 mrem/hr. This does not invalidate the water door idea but does require that it exist in a corridor that itself is shielded from the cavity to use used.

## COUPLED CELL ENERGY BALANCE

For an upgrade to be successful, cavities must meet stringent performance specifications. One such parameter is the frequency of the cavity. This can be influenced by imperfections in the manufacturing process resulting in cavities with unideal physical dimensions, which results in an energy imbalance between the cells. The energy imbalance manifests as each cell having a different resonant frequency. To correct this and enhance the cavity, a tuning process is used in which cells are squeezed to decrease the frequency and pulled apart to increase it.

### Coupled Oscillators

In my studies before my term as an intern at BNL, I had not studied the coupled oscillator problem. In order to effectively work on cavity tuning, a good understanding of this problem is essential because multicell SRF cavities are coupled oscillators. So, I spent time studying the case of five pendulums coupled by springs shown schematically in Fig. 3. This section of my report is dedicated to my study of the 5 coupled pendulums, which can be broken down into

1. Determining the equations of motion of each pendulum,
2. Combining these equations into a single matrix equation, and
3. Applying spectral theory to solve for the normal modes of oscillation.

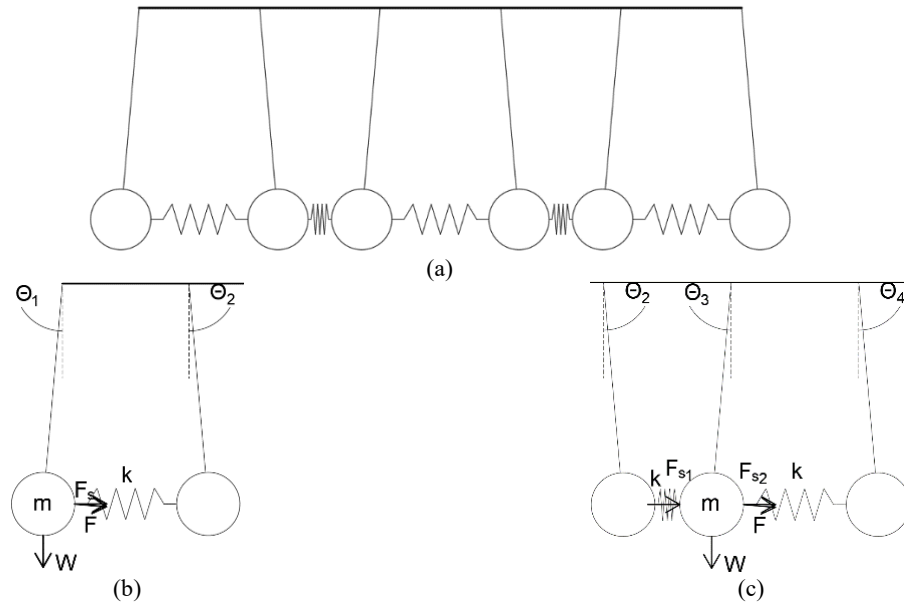
#### *Determining the equations of motion of each pendulum*

From Fig. 1b, we see that

$$\begin{aligned}
F &= -mg \sin \theta_1 \\
F_s &= -kl(\sin \theta_1 - \theta_2)
\end{aligned}$$

Summing these forces gets the equation of motion

$$\Sigma F = -mg \sin \theta_1 = -kl(\sin \theta_1 - \theta_2)$$



**FIGURE 4.** Schematic of five pendulums coupled by springs (a), and force diagrams for (b) the leftmost pendulum and (c) the center pendulum.

The equation of motion is more useful to us in its rotational form, so we invoke  $\tau = Fl = I\ddot{\theta}$  and get

$$ml^2\ddot{\theta}_1 = -mgl \sin \theta_1 - kl(\sin \theta_1 - \sin \theta_2)$$

$$\ddot{\theta}_1 = -\frac{g}{l} \sin \theta_1 - \frac{k}{l}(\sin \theta_1 - \sin \theta_2)$$

Now, we can use the small angle approximation and make the substitutions  $\omega_0 = \sqrt{g/l}$  and  $\omega' = \sqrt{k/l}$  to arrive at the final version of the equation of motion for the leftmost pendulum:

$$\ddot{\theta}_1 = -\omega_0^2 \theta_1 - \omega'(\theta_1 - \theta_2)$$

We will now apply the same analysis to the center pendulum depicted in Fig. 3c. From that diagram, we derive

$$F = -mg \sin \theta_3$$

$$F_{s1} = -kl(\sin \theta_3 - \sin \theta_2)$$

$$F_{s2} = -kl(\sin \theta_3 - \sin \theta_4)$$

We sum the forces and convert to rotational quantities

$$\Sigma F = -mg \sin \theta_3 - kl(\sin \theta_3 - \sin \theta_2) - kl(\sin \theta_3 - \sin \theta_4)$$

$$\Sigma F = -mg \sin \theta_3 - kl(-\sin \theta_2 + 2 \sin \theta_3 - \sin \theta_4)$$

$$ml\ddot{\theta}_3 = -mgl \sin \theta_3 - kl^2(-\sin \theta_2 + 2 \sin \theta_3 - \sin \theta_4)$$

$$\ddot{\theta}_3 = -\frac{g}{l} \sin \theta_3 - \frac{k}{l}(-\sin \theta_2 + 2 \sin \theta_3 - \sin \theta_4)$$

Using the same substitutions as before, we arrive at

$$\ddot{\theta}_3 = -\omega_0^2 \theta_3 - \omega'^2(-\theta_2 + 2\theta_3 - \theta_4)$$

If we are clever, there is no need to solve for the remaining pendulums. We avoid this by taking advantage of the symmetry of the problem. Pendulums with springs on each side will have equations of motion following the form

$$\ddot{\theta}_j = -\omega_0^2 \theta_j - \omega'^2 (-\theta_{j-1} + 2\theta_j - \theta_{j+1})$$

Where  $j$  is the pendulum number, and the end pendulums will follow the forms

$$\ddot{\theta}_1 = -\omega_0^2 \theta_1 - \omega'^2 (\theta_1 - \theta_2) \qquad \ddot{\theta}_N = -\omega_0^2 \theta_N - \omega'^2 (\theta_N - \theta_{N-1})$$

### *Equations of Motion as a Matrix Equation*

Although it is possible to decouple the equations of motion for each pendulum and solve for the equations of motion that way, this approach becomes increasingly impractical as the system grows. A better approach is to use linear algebra and spectral theory to find the eigenvalues and eigenvectors.

To convert our problem to linear algebra, we must define matrices and vectors. We require a square matrix to hold each angular acceleration and a column vector to hold each angle

$$\ddot{\boldsymbol{\theta}}_N = \begin{bmatrix} \ddot{\theta}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \ddot{\theta}_2 \end{bmatrix} \text{Type equation here.} \qquad \overrightarrow{\boldsymbol{\theta}}_N = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_N \end{bmatrix}$$

We also require a matrix that embodies the symmetry of the problem which will be multiplied by the scalar  $\omega'^2$ , which I will call matrix  $\mathbf{P}$ , for pendulum.

$$\mathbf{P} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}$$

With these three matrices, we can define a general matrix equation for the coupled pendulum problem

$$\ddot{\boldsymbol{\theta}}_N = \overrightarrow{\boldsymbol{\theta}}_N (-\omega_0^2 \mathbf{I} - \omega'^2 \mathbf{P})$$

### *Eigenmodes*

We are now able to solve for the eigenmodes of the equation. First, we must use complex math to get the equation in solvable form. Consider  $\overrightarrow{\boldsymbol{\theta}}_j = RE[Z_j]$  and  $Z \equiv \exp[i(\omega t + \phi)] \vec{A}$ , where  $\vec{A}$  is a column vector with elements  $a_1, a_2, \dots, a_N$  and  $\mathbf{A}$  is a matrix with diagonal elements  $a_1, a_2, \dots, a_N$ .

$$\begin{aligned} \ddot{\mathbf{Z}} &= \vec{Z} (-\omega_0^2 \mathbf{I} - \omega'^2 \mathbf{P}) \\ \omega^2 \mathbf{Z} &= \vec{Z} (-\omega_0^2 \mathbf{I} - \omega'^2 \mathbf{P}) \\ \omega^2 \mathbf{A} &= \vec{A} (-\omega_0^2 \mathbf{I} - \omega'^2 \mathbf{P}) \\ 0 &= \vec{A} (-\omega_0^2 \mathbf{I} - \omega'^2 \mathbf{P} - \lambda \mathbf{I}) \end{aligned}$$

The eigenvalues can then be found by solving  $\det[-\omega_0^2 \mathbf{I} - \omega'^2 \mathbf{P} - \lambda \mathbf{I}]$ , which I will perform for  $N = 5$ . Thankfully, since  $\omega_0^2$  we know will be a solution and factoring out  $\omega'^2$  will not affect the basis obtained because it would create a linear combination, this is equivalent to solving  $\det[\mathbf{P} - \lambda \mathbf{I}]$ . This is done on the following page.

$$\det[\mathbf{P} - \lambda \mathbf{I}] = 0$$

$$\det \begin{bmatrix} 1-\lambda & -1 & 0 & 0 & 0 \\ -1 & 2-\lambda & -1 & 0 & 0 \\ 0 & -1 & 2-\lambda & -1 & 0 \\ 0 & 0 & -1 & 2-\lambda & -1 \\ 0 & 0 & 0 & -1 & 1-\lambda \end{bmatrix} = 0$$

$$(1-\lambda) \begin{vmatrix} 2-\lambda & -1 & 0 & 0 \\ -1 & 2-\lambda & -1 & 0 \\ 0 & -1 & 2-\lambda & -1 \\ 0 & 0 & -1 & 1-\lambda \end{vmatrix} + \begin{vmatrix} -1 & -1 & 0 & 0 \\ 0 & 2-\lambda & -1 & 0 \\ 0 & -1 & 2-\lambda & -1 \\ 0 & 0 & -1 & 1-\lambda \end{vmatrix} = 0$$

$$(1-\lambda)(2-\lambda) \begin{vmatrix} 2-\lambda & -1 & 0 \\ -1 & 2-\lambda & -1 \\ 0 & -1 & 1-\lambda \end{vmatrix} + (1-\lambda) \begin{vmatrix} -1 & -1 & 0 \\ 0 & 2-\lambda & -1 \\ 0 & -1 & 1-\lambda \end{vmatrix} - \begin{vmatrix} 2-\lambda & -1 & 0 \\ -1 & 2-\lambda & -1 \\ 0 & -1 & 1-\lambda \end{vmatrix} = 0$$

$$(1-\lambda)(2-\lambda)\{(2-\lambda)[(2-\lambda)(1-\lambda) - 1] - (1-\lambda)\} - (2-\lambda)[(2-\lambda)(1-\lambda) - 1] + (1-\lambda) - (1-\lambda)[(2-\lambda)(1-\lambda)] + (1-\lambda) = 0$$

$$(1-\lambda)^2(2-\lambda)^3 - (1-\lambda)(2-\lambda)^2 - (1-\lambda)^2(2-\lambda) - (1-\lambda)(2-\lambda)^2 - (1-\lambda)^2(2-\lambda) + (2-\lambda) + 2(1-\lambda) = 0$$

$$(1-2\lambda+\lambda^2)(2-\lambda)(4-4\lambda+\lambda^2) - 2(4-8\lambda+5\lambda^2-\lambda^3) - 2(2-5\lambda+4\lambda^2-\lambda^3) - 3\lambda+4 = 0$$

$$8-28\lambda+38\lambda^2-25\lambda^3+8\lambda^4-\lambda^5-8+16\lambda-10\lambda^2+2\lambda^3-4+10\lambda-8\lambda^2+2\lambda^3-3\lambda+4 = 0$$

$$\lambda(\lambda^4-8\lambda^3+21\lambda^2-20\lambda+5) = 0$$

$$\lambda(\lambda^2-5\lambda+5)(\lambda^2-3\lambda+1) = 0$$

$$\lambda = \left\{ \frac{5}{2} \pm \frac{\sqrt{5}}{2}, \frac{3}{2} \pm \frac{\sqrt{3}}{2}, 0 \right\}$$

## Tuning Program

Now that I had developed a strong understanding of the coupled oscillator problem, I was prepared to work on a cavity tuning program. The purpose of the program is to automate tuning calculations from bead pull data. I wrote it in MATLAB based on a cavity tuning paper by Peter Schmuser [7]. The code used the measured pi and pi/N mode frequencies, frequency offsets of each cell operating in pi mode and offset caused by the bead to calculate how much tuning is needed per cavity in hertz.

## FUTURE WORK

My shielding requirement calculation is not sufficient justification to begin work on cavity testing. More sophisticated methods typically involving simulations run on supercomputers will be required, which is beyond the scope of an internship term. My calculations will serve as a guideline for the health physicists to which the project will be handed off to.

The work on the cavity tuning program could have proceeded to completion if the program was on site. What remains to be done is interfacing the MATLAB code with the bead pull apparatus and developing a tuning apparatus to safely manipulate the cavity.

## CONCLUSION

Preliminary estimates of shielding requirements for EIC SRF cavity testing have been made and a cavity tuning automation program has been started. The shielding requirements will be verified and refined by BNL health physicists and the program's development will continue on site, where it can be interfaced with hardware.

## ACKNOWLEDGEMENTS

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program (CCI). I would like to thank my mentor Zack for his guidance not only in the scope of my CCI appointment, but also in my continuing education and professional development.

## REFERENCES

- [1] P. A. Cherenkov, "Visible luminance of pure liquids under the action of gamma radiation," *Reports of the academy of sciences of the USSR*, vol. 2, no. 8, pp. 385-389, 1934.
- [2] J. A. Sibery, "X-ray imaging physics for nuclear medicine technologists. Part 1: Basic principles of x-ray production," *Journal of nuclear medicine technology*, vol. 3, no. 32, pp. 139-147, 2004.
- [3] J. Delayen, "Longitudinal transit time factor of short independently phased accelerating structures for low velocity ions," *Nuclear instruments and methods in physics research*, vol. A258, pp. 15-25, 1987.
- [4] F. M. E. Al, "Field emission and consequences as observed and simulated for CEBAF upgrade cryomodules," in *SRF 2013*, Paris, 2013.
- [5] J. H. a. S. Seltzer, "X-ray mass attenuation coefficients," *NIST standard reference database 126*, 2004.
- [6] A. T. E. Al., X-ray data booklet, Berkeley: LBNL center for x-ray optics and advanced light source, 2009.
- [7] P. Schmuser, "Tuning of multi-cell cavities using bead pull measurements," in *SRF920925-10*.

Numerical simulation of impact ionization in LGADs

Nimal de Lanerolle

Department of Mathematics, St. Joseph's College, Patchogue, NY 11772

Gabriele Giacomini, Instrumentation Division, Brookhaven National Laboratory, Upton, NY  
11973

## **Abstract**

Low-gain-avalanche-diodes (LGAD) are a recently-developed class of silicon sensor devices, characterized by an internal moderate gain that enhances the output signal amplitude. They feature fast signals and exhibit excellent timing performance. They are planned to be incorporated in timing detectors in High-Energy Physics experiments. Electronic devices are fabricated to perform within certain electrical specifications, such as the breakdown voltage. Simulation enables determining the device processing (fabricating) conditions without actual experimentation. Measured breakdown voltages on actual devices, dopant distribution in monitor wafers run parallel with the device wafers and the simulation package (Silvaco). Also, Silvaco offers empirical models that enable fine tuning the simulated output results to that measured. The goal was to obtain by simulation processing conditions that will achieve the measured breakdown voltage for the device. This was achieved by choosing different dopant (dopant) levels for n-type and p-type, choice of initial silicon wafer dopant concentration, silicon wafer thickness, and choice of theoretical models. Changing the dopant concentration within the device alters the electric field during operation, by a process called impact ionization, where extra electron-hole pairs are created when a moderate voltage is applied to the device. The empirical models attempt to predict this behavior by relating the electric field generated in the device by the electro-hole pairs to the breakdown voltage. The empirical Valdinoci model predicts that the measured breakdown voltage can be achieved if it is assumed that approximately only 96% of boron dopant is activated in the device.

## Introduction

Low-gain avalanche diodes are a recently developed class of silicon-based sensors that enhances the signal amplitude. Since they are built on thin silicon wafer substrates they also feature fast signals and exhibit fast collection performance. Therefore, they are seen to be an attractive alternative to the current detectors in High Energy Physics Experiments. This new technology also is creating interest in other areas of application such as, medical physics, imaging and photon science. The principle of operation is the concept of impact ionization, which is akin to that of billiard balls being impacted by a cue. The high energy particle in the physics experiments, causes a cascading effect of electron-hole pair creation by its impact on the lattice of the device material similar to that of the cue impacting on billiard balls. Electrons are scattered from the valence band into the conduction band, creating an electron-hole pairs. Since intrinsic silicon if impacted in the same fashion does not create sufficient electron-hole pairs it is imperative to introduce impurity atoms, called dopant atoms, of the n-type and p-type into the lattice, via the effect of implant ionization. If the impact ionization occurs in a region of high electric field it can result in avalanche breakdown, where an original signal is amplified before entering an external electronic circuit. This fact is exploited in LGADs.

The devices are usually fabricated on thin high-resistivity silicon p-type substrates, about 50  $\mu\text{m}$  thick and are based on the principle of p-n junction diodes. An additional p-type layer, called the multiplication or gain layer, is implanted under the shallow n-type layer. Typically, this dose is around  $10^{18} \text{ cm}^{-2}$  and extends into the substrate to a few microns deep<sup>1</sup>. This layer provides a region of electric field which is responsible for the avalanche impact ionization needed for operation of a LGAD<sup>2</sup>.



The breakdown voltage  $V_{BD}$  of a pn-junction diode depends on the doping concentration of the lowly doped side and can be expressed as:

$$V_{BD} = \frac{E_{max} \epsilon_0}{q N_A} \approx \frac{E_{max} \epsilon_0}{q N_A} \quad (1)$$

Where  $N_A$  is the doping concentration of the lowly doped side, in this case the p-type layer mentioned earlier,  $E_{max}$  the maximum field strength,  $\epsilon_0$  the permittivity of Si<sup>3</sup>. This relation indicates that the breakdown voltage is inversely related to the p-layer concentration, a fact alluded to later in Figure 2.

## Methods

The starting or reference points for the simulation were the n- and p- type dopant distribution profiles provided by SIMS measurement on the device and the breakdown voltages for different p-type deep implants. Simulations using the Silvaco Device Simulations package for different n-type (phosphorus) were made and compared with that of the SIMS profile. The best fit was then used for simulating the deep p-type (boron) implant, using different empirical models, that were provided by Silvaco. From this simulation it was concluded that the Valdinoci empirical model was the best fit for the measured breakdown voltage measurements. Taking the Valdinoci model, further simulations were made compare the effects of thinner 30  $\mu\text{m}$  silicon substrate lower initial substrate doping conditions, and boron doping and compared with measured breakdown voltage. The results are discussed below.

## Results and Discussion

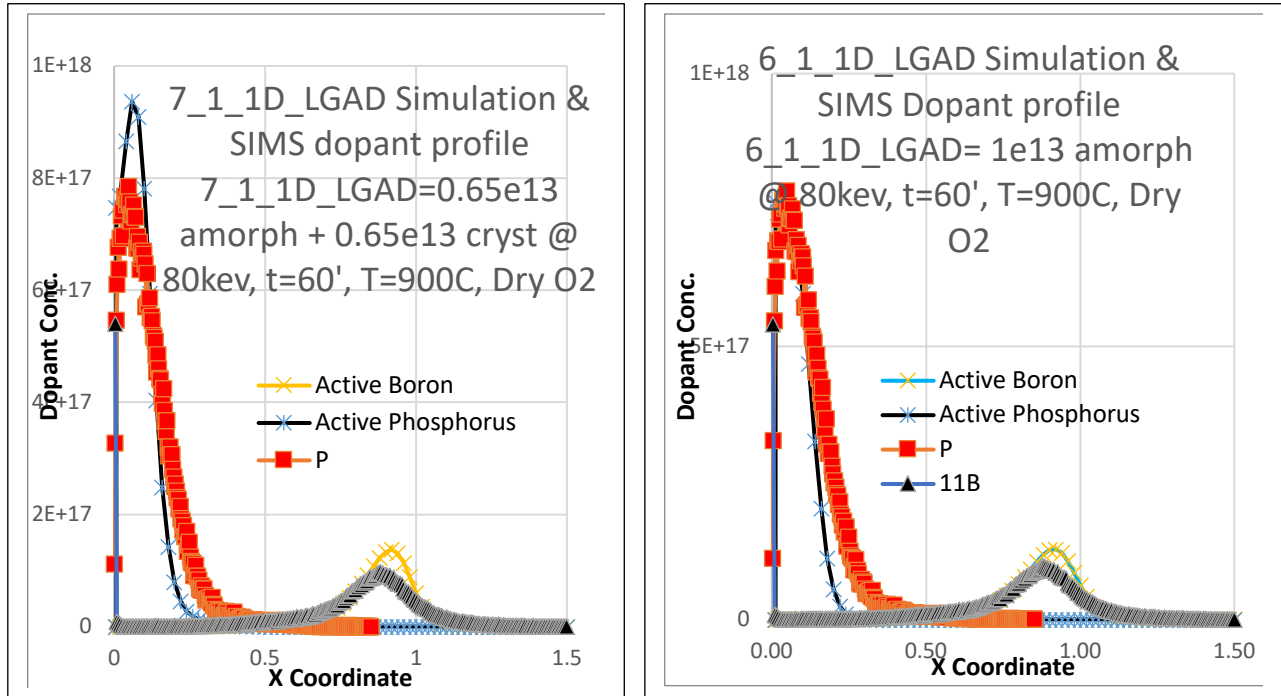
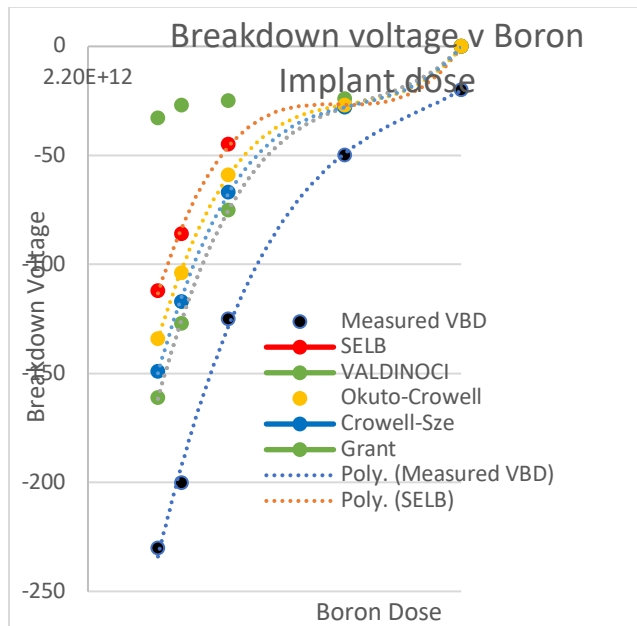


Figure 1 – Dopant distribution profile from simulation compared with measured SIMS distribution of dopant profile on actual device indicated by the “active” legend.



Measured VBD		VBD by simulation				
Boron dose	Meas. VBD	Valdinoci	Selb	Okuto	Crowell	Grant
2.35E12	-230	-161	-112	-134	-149	-33
2.4E12	-200	-127	-86	-104	-117	-27
2.5E12	-125	-75	-45	-59	-67	-25
2.75E12	-50	-20	-27	-27	-28	-24
3.0E12	-20	-0.125	-0.125	-0.125	-0.125	-0.125

Figure 2 – Measured breakdown voltage (VBD) compared with that predicted by several empirical models are depicted in above figure. The Valdinoci model is closest representation of measured VBD.

Several n-type phosphorus implant doses were simulated for an initial substrate doping level of  $7 \times 10^{13}$  p-type boron, and substrate crystalline conditions of amorphous or crystalline. Figure 1 shows one example of such a scheme of simulation. Since this is a simulation it was possible to choose two types of substrate crystal structure for 7\_1\_1D\_LGAD simulation, implanting one half the actual dose of  $1 \times 10^{13}$  phosphorus into amorphous and the other half into crystalline substrate. After implantation the devices get a heat treatment that is meant to activate the impurity dopants implanted. Figure 1 clearly shows that the activated distribution from simulation predicts a higher level of activated phosphorus compared to that of the measured from SIMS. This observation was shown by all the simulations. That is the measured phosphorus distribution in the actual device is lower than that predicted by simulation. In the literature this has been explained with phosphorus segregating into the silicon dioxide layer which grows during the activation heat treatment<sup>4</sup> [applied physics letters]. Further simulations were performed using the conditions represented by 7\_1\_1D\_LGAD. Subsequent simulations were aimed at predicting the breakdown voltage for different levels of the deep p-type boron implant. The actual and simulated implant energy for the boron was 380 keV, which would cause the boron, which is an interstitial diffuser to go deep into the substrate. A condition that is required as per the earlier discussion for the need to create the region for impact ionization. Table 1 shows the conditions used for the actual device, the breakdown voltage measured in the actual device under these conditions. As per equation (1) above the breakdown voltage is inversely related to the p-type dopant, and the measured breakdown voltage clearly shows that this relationship is followed in the actual device. Taking the same p-type dopant levels and the same implant energy, simulations were made using

several empirical models available in Silvaco. These models attempt to represent the field dependence from the electron-hole pair creation<sup>3</sup>. The field dependence is implanted by these models through a variation of the Chynoweth law:

$$\alpha = \alpha_1 e^{\frac{b}{r}}$$

Where  $\alpha_1$  and  $b$  have been experimentally determined by measurements on pn – diodes. As per the literature the models are a form of the above representation. Table 1 clearly indicates that the simulated breakdown voltages for the same p-type implant dose is numerically different from that measured. Figure 2 represents this behavior graphically. The figure shows that although the curves predicted by the different models follow the same type of measured behavior they do not match up. Furthermore, the model that represents the measured behavior is the Valdinoci model.

## Conclusions

The simulations clearly shows that the n-type phosphorus implant in the measurements made on the device using the SIMS technique is lower than that predicted by simulations. The simulations were made using different doping conditions, different substrate crystalline conditions. In all cases the predicted differed from the measured. The literature does try to explain this behavior as a segregation phenomena. In the creation of a LGAD device it is imperative to obtain a region that will induce impact ionization when the device is exposed to an electric field. As discussed above this is achieved by the deep p-type boron implant. The measured breakdown voltage created by this scheme is inversely related to the p-type dose as indicated by equation (1), and the predicted breakdown voltage by simulations also follows this behavior. However, the predicted values for the breakdown voltage is very different from the

measured values. Further simulations will be carried out to explore this behavior, and also to determine if it is possible to achieve the measured breakdown voltage with lower p-type dopant levels, and if there is a relationship to the breakdown voltage with substrate thickness of 30  $\mu\text{m}$ , which is less than the standard thickness of 50  $\mu\text{m}$  for LGAD devices, and a lower initial substrate doping level.

## References

<sup>1</sup> G. Giacomini, W. Chen, G. D'Amen, L. Lavitola and A. Tricoli, "Fabrication and Testing of AC-coupled Low-Gain Avalanche Diodes at BNL," 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2019, pp. 1-3, doi: 10.1109/NSS/MIC42101.2019.9059960.. ]

<sup>2</sup> M.Mandurrino, et al, " First demonstration of 200, 100, and 50  $\mu\text{m}$  pitch resistive ac-coupled silicon detectors with 100% fill-factor for 4D particle tracking", submitted to IEEE device letters, arXiv: 1907.03314v3 [physics.ins-det] 24 Sep 2019.

<sup>3</sup> [https://iis-people.ee.ethz.ch/~schenk/euban/English/avalanche\\_19.pdf](https://iis-people.ee.ethz.ch/~schenk/euban/English/avalanche_19.pdf)

Fluid dynamic simulation and analysis of water-cooling systems for the Electron-Ion Collider

Joseph DeRienzi

Department of Mechanical Engineering, North Carolina State University, NC 27607

Thea Vijaya Kumar

Department of Mechanical Engineering, Stony Brook University, Stony Brook, NY 11794

Ram Srinivasan

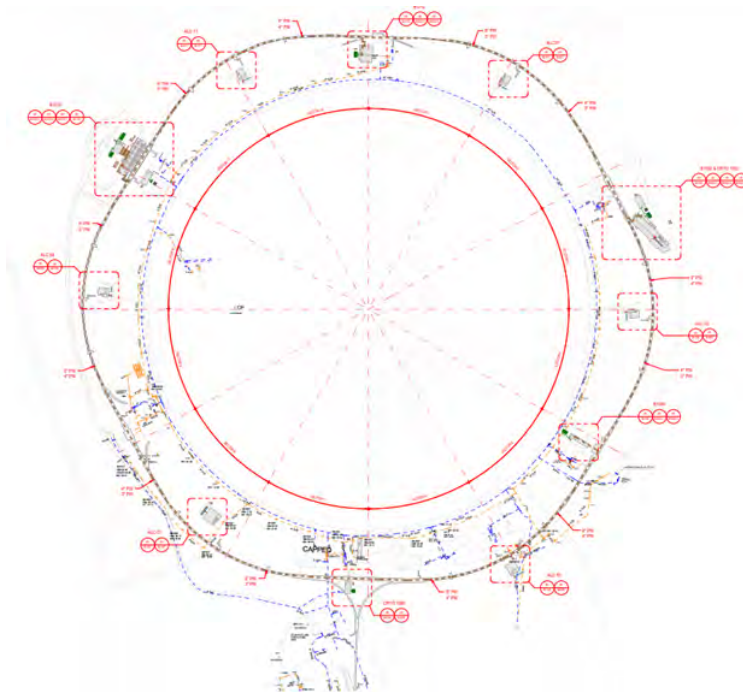
Electron-Ion Collider, Brookhaven National Laboratory, Upton, NY 11973

## **Abstract**

The Electron-Ion Collider is the newest large-scale project at Brookhaven National Laboratory. The collider's purpose is to provide further advancements in the knowledge of the universe's origin by accelerating particles near the speed of light. Our project for this 3.8 km ring was to create a thermal hydraulic steady-state simulation design of the water-cooling system to be cost-effective and energy efficient, as envisioned by Charlie Foltz, the EIC Infrastructure Division Director. The system would include a supply and return header, which cools several thousand components of the ring. The water would then be returned and cooled down using a system of cooling towers and plate and frame heat exchangers. Due to the size of the system and the complexity of the network analysis, a fluid dynamic simulation software, AFT Fathom, was used. Since previous methods of maintaining systems relied on building upon smaller real-life models and implementing empirical data, this flow model was unique and first of a kind in the domain of accelerator design, construction and operation. Therefore, our hydraulic team piloted a new method to perform network analysis on a large scale cooling system. We successfully created several test scenarios for system behavior in a shorter time compared to the method of performing hand calculations. Cooling specifications for heat rejection, pressure drop, flow rate, and pipe sizing were changed based on the individual systems of the vacuum, radio frequency (RF), magnet and power supply, and cryogenics sections. Finally, we used DOE guidelines to perform life-cycle cost analysis with net present value and carbon saving analysis on the systems where pipe size could be optimized.

## I. Introduction

The Electron-Ion Collider is a large model, which requires cooling systems for every component to ensure the temperatures are in safe ranges. A hydraulic model is extremely beneficial to assess the behavior of the system at different components, especially for something as complex as this. Charlie Foltz, the Infrastructure Division Director for the EIC, decided that the hydraulic model was the best approach. AFT Fathom, a cooling simulation software, was used to get a comprehensive understanding of network analysis with such a large system. The ring is broken down into various systems, which was worked on throughout the summer. The first section includes the vacuum ring, which follows the path of the outer ring in Figure 1.



**Figure 1:** EIC Layout with buildings for cooling

The next sections included the radio frequency (RF) systems. The cooling department worked closely with other interns who designed the layout of power amplifier units in Creo. We added



the magnets and power supplies in one system, and the final systems included the Cryogenics buildings from the 1002, 1006, and 1010 buildings.

We designed the layout of all buildings and systems and then used inputs to match cooling requirements of each individual component within the systems from a Master Spreadsheet. The team ultimately utilized a brand new method of cooling system design in a fraction of the time it would take for hand calculations.

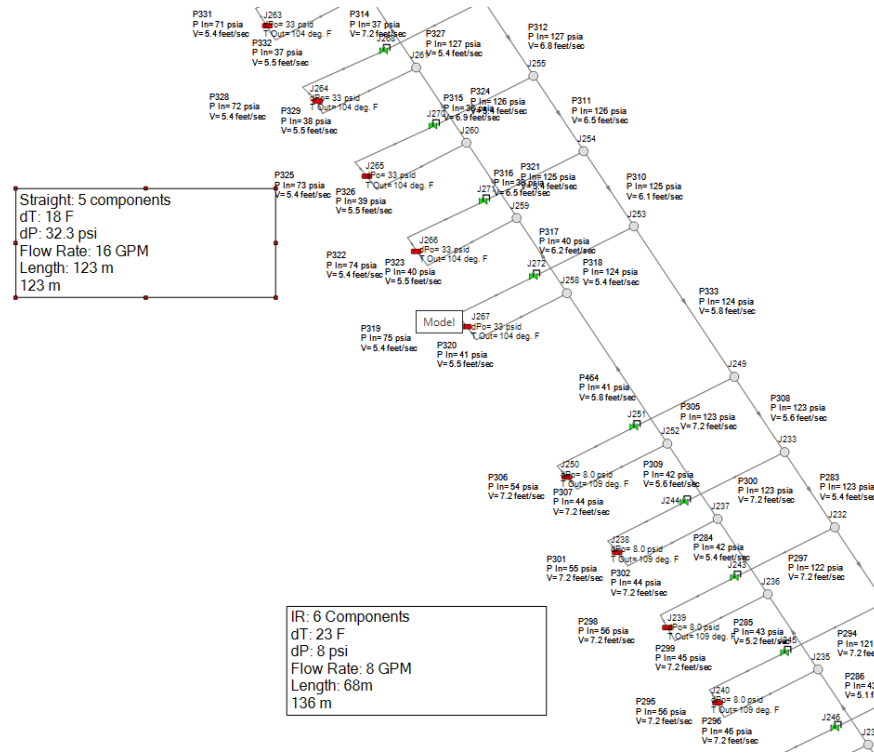
## II. Vacuum Ring

The vacuum ring follows the perimeter of the EIC. Since the ring is large and complex, it was divided in half between two interns. First, quadrants were easier to create to determine the proper layout and parameters. Each quadrant was composed of arcs, straight sections, and interaction regions (IRs). The arc sections have 16 heat rejection components, while the straights have 5 and the IRs have 6. Additionally, each of these sections had their own set of parameters that needed to be met. The data for each section is compiled in Table 1 below.

**Table 1:** Summary of the Vacuum System cooling specifications

Section	Heat Load (kW)	Pressure Drop, dP (psi)	Temperature Rise, dT (°F)	Velocity (ft/s)	Flow Rate (gpm)	Nominal Size (in)
Arc	65	28	27.7	5.4	16	0.75
Straight	42	32.3	18	5.4	16	1.00
IR	27	8	23	7.2	8	0.50

The model data is also displayed in the software to keep track of all required parameters, as seen in Figure 2. We first created each loop for the sections and then added all of the data values.



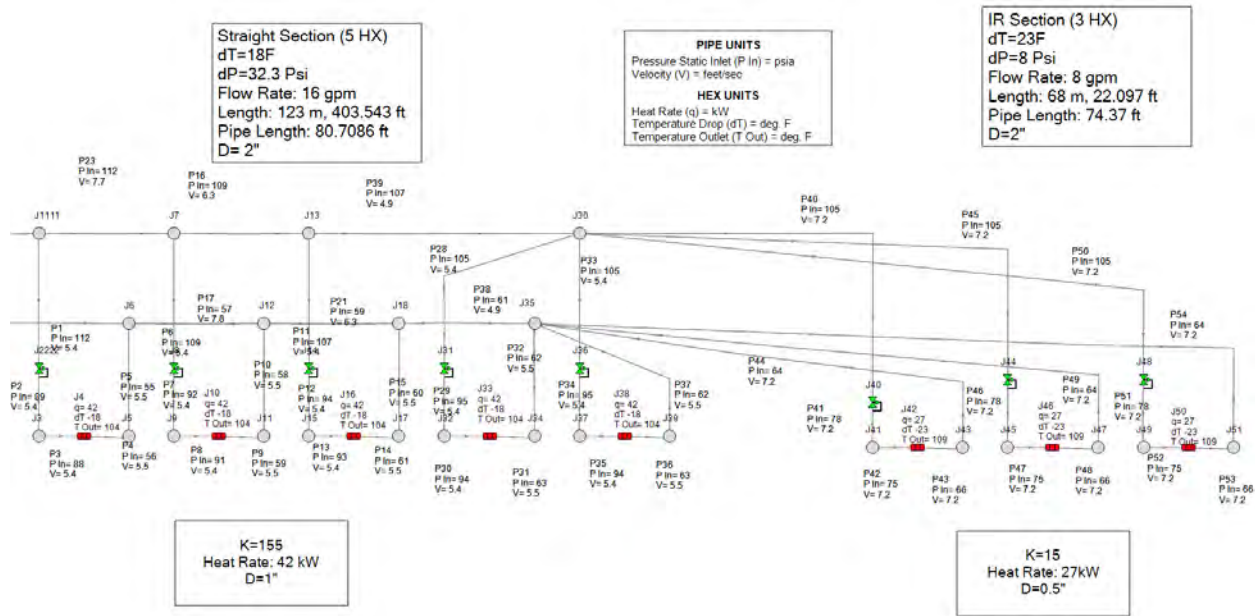
**Figure 2:** Straight and IR Sections of the Vacuum Ring

Along with the data provided in a master spreadsheet, loss values specifically associated with the heat rejection components were determined through the Darcy-Weisbach Equation. This equation is as follows:

$$h_f = K \frac{v^2}{2g}$$

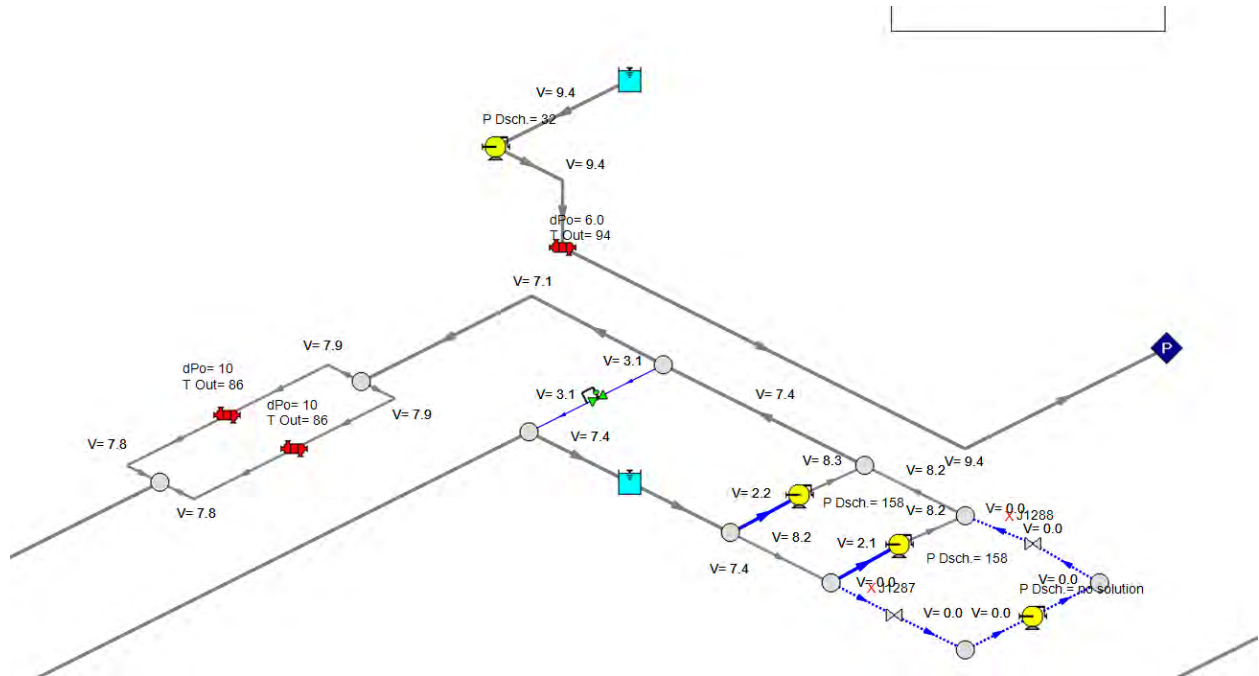
The velocity is in feet per second and head is in feet, while gravity is in feet per second squared. The loss factor, K, is unitless. The velocity is found through the simulation and varies depending on the section. Each section has its own loss factor value, which is detailed in information boxes

on the model, shown in Figure 3. These values set the required pressure drop across the heat rejection components.



**Figure 3:** Loss factors and cooling specifications detailed on straight and half IR sections

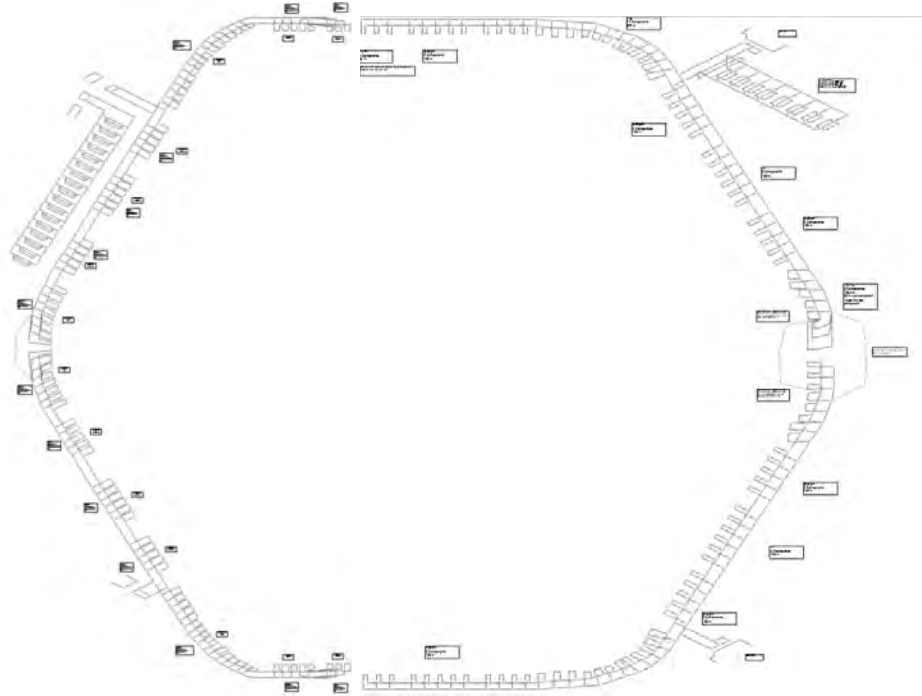
The pump systems for the vacuum system, as well as the rest of the systems are made up of two running pumps and one stand-by pump. The stand-by pump is meant to illustrate a situation in which one pump would be shut down or in maintenance. All pumps are set to run at half the required gpm of the system, and any pumps that do not have a real pump curve are assumed to operate at 70% efficiency. All systems also include two plate and frame heat exchangers that cool down the water to the required 86°F. A figure of the standard layout is displayed in Figure 4.



**Figure 4: Pumps and Heat Exchanger Layout**

Since only half of the ring needed to be designed, the northern halves were mirrored across a horizontal axis to create the bottom two quadrants of the ring. Also, since each quadrant is controlled by its own building, connections between the quadrants are made in the form of closed valves. This is to prevent mixed flow between the two sides, but in the event that one quadrant is down, it can get water from another quadrant.

Both halves of the EIC vacuum ring consist of 96 heat rejection components, with 3 arcs, 6 straight sections, 2 full IR sections, and 2 half IR sections. Both models included half IR sections, located at the 6 and 12 o'clock positions, which connect the left and right side. The two models side by side are shown below in Figure 5.



**Figure 5:** Two halves of the Vacuum system ring

### III. Radio Frequency System

#### 1002 RF System

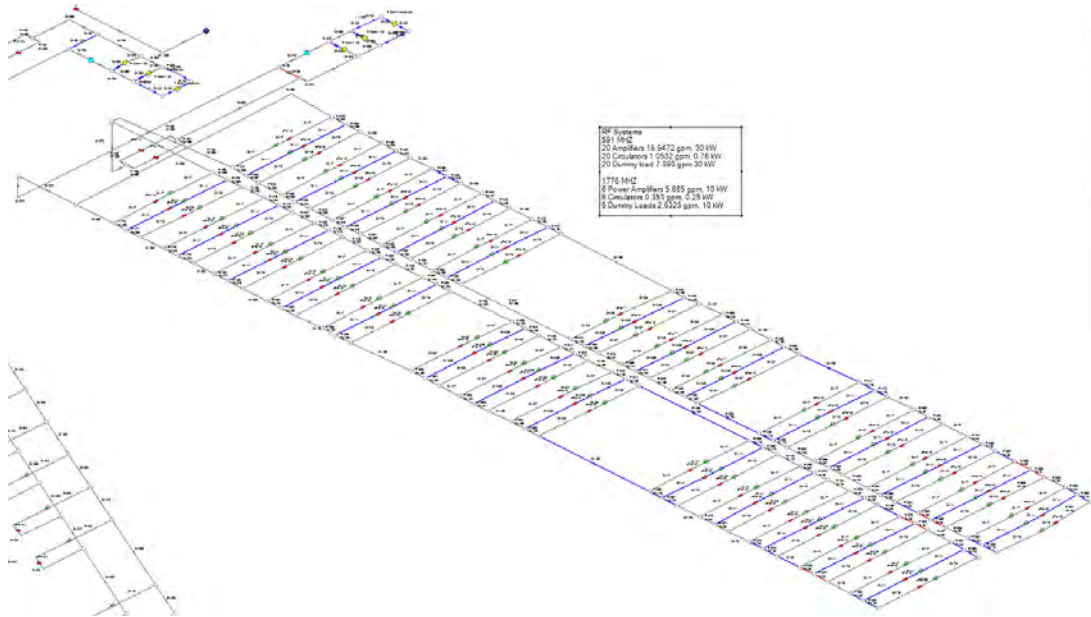
There are 26 sections of heat rejection components in the 1002 building RF system. These heat rejection components are in place for the 13 cavities in the 1002 building. Each section consists of a power amplifier (PA), circulator, and dummy load (absorber). All RF systems consist of these three components per unit. The sections are also broken up into twenty 591 MHz cavities, and six 1773 MHz cavities. The heat load for these cavities was calculated by summing up the heat load values provided by the RF system group, and dividing the total heat load amongst the total number of sections in the RF system. This was done because the original number of cavities was modified, so the heat load and flow rate per cavity had to be adjusted. The data tables below summarize the cooling requirements and Figure 6 shows the 1002 RF layout.

**Table 2: 591 MHZ unit cooling specifications**

<b>Section</b>	<b>Heat Load (kW)</b>	<b>Pressure Drop (psi)</b>	<b>Temperature Rise (°F)</b>	<b>Flow Rate (gpm)</b>	<b>Loss Factor, K</b>
Power Amplifier	30	45	11.9	16.95	200.09
Circulator	0.76	30	4.9	1.05	213.18
Dummy Load	30	30	25.9	7.898	87.42

**Table 3: 1773 MHZ unit cooling specifications**

<b>Section</b>	<b>Heat Load (kW)</b>	<b>Pressure Drop (psi)</b>	<b>Temperature Rise (°F)</b>	<b>Flow Rate (gpm)</b>	<b>Loss Factor, K</b>
Power Amplifier	10	45	11.9	5.66	257.01
Circulator	0.25	30	4.9	0.351	1980.65
Dummy Load	10	30	25.9	2.63	108.8



**Figure 6: 1002 RF Building Overall Layout**

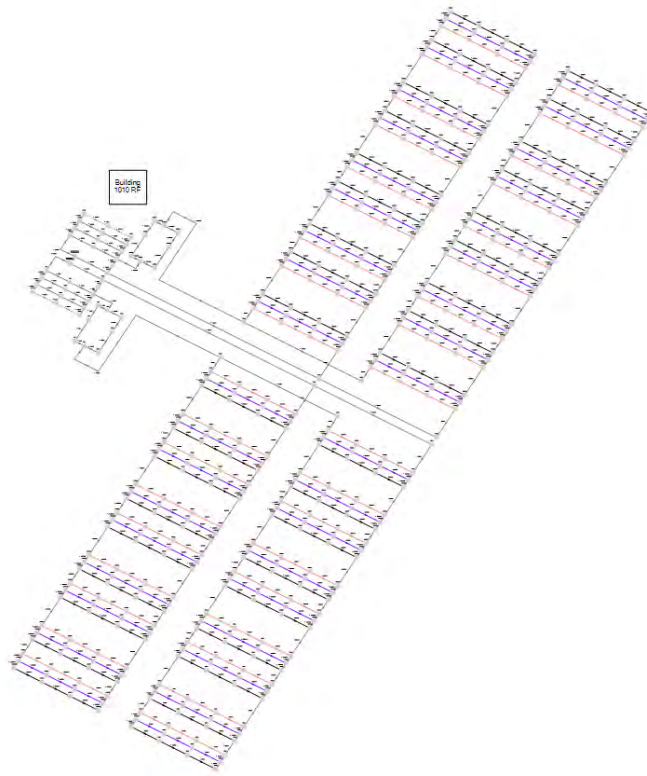
### 1010 RF System

There are 36 sections of heat rejection components in the 1010 RF system. Originally, there were supposed to be 34 units, but the heat load and flow rates were totaled and divided among 36. The resulting calculations are summarized in Table 4 below.

**Table 4: RF 1010 unit cooling specifications**

Section	Heat Load (kW)	Pressure Drop (psi)	Temperature Rise (°F)	Flow Rate (gpm)	Loss Factor, K
Power Amplifier	377.78	45	12	226.34	98.93
Circulator	9.44	30	4.9	14.04	84.00
Dummy Load	53.36	30	5.0	77.09	96.40

Two rows of 18 were created and each unit was grouped in pairs to replicate the Creo layout of the 1010 building. The layout used is shown in Figure 7.



**Figure 7:** 1010 RF Building Overall Layout

A close-up of the pumping system is displayed in Figure 8. There are three pumps functioning in each circuit, with a total of 6 pumps. There is one standby pump each as a backup. Both circuits have two plate and frame heat exchangers to cool down the water.





**Figure 8:** RF 1010 Pumping System

### Optimization

When designing a system, optimizing the model is a large part of analysis. This stems from the realization that two pipe sizes can result in the velocity parameter being between 5 ft/s and 10 ft/s. Part of the optimization requires analyzing the benefits of using a pipe size that provides a cheaper installation cost versus one that provides a cheaper pump power cost. One example was for the power amplifier section of the RF system. Three pipes for each of the 36 sections could either be 3 inches in diameter, or 4. When analyzing the power cost, it was seen that the 3 inch pipe had a higher cost total. However, the total cost for piping installation was higher for the 4 inch pipe. By comparing the two costs, it can be seen that the power cost is much higher than the installation, so the 4 inch pipe should be chosen for the system, as it results in a lower overall cost. Table 5A summarizes the information.

**Table 5A:** Optimization of pipe sizes for power cost and pipe cost

<b>Pipe Size (in)</b>	<b>Velocity in pipe (ft/s)</b>	<b>Power (kW)</b>	<b>Cost for Power</b>	<b>Cost per unit</b>	<b>Total Cost for piping</b>
3	8.7	470.10	\$2,783,462.10	\$145.50	\$147,711.60
4	5.1	391.86	\$2,320,203.06	\$163.00	\$165,477.60
		<b>Difference</b>	\$463,259.04		\$(17,766.00)

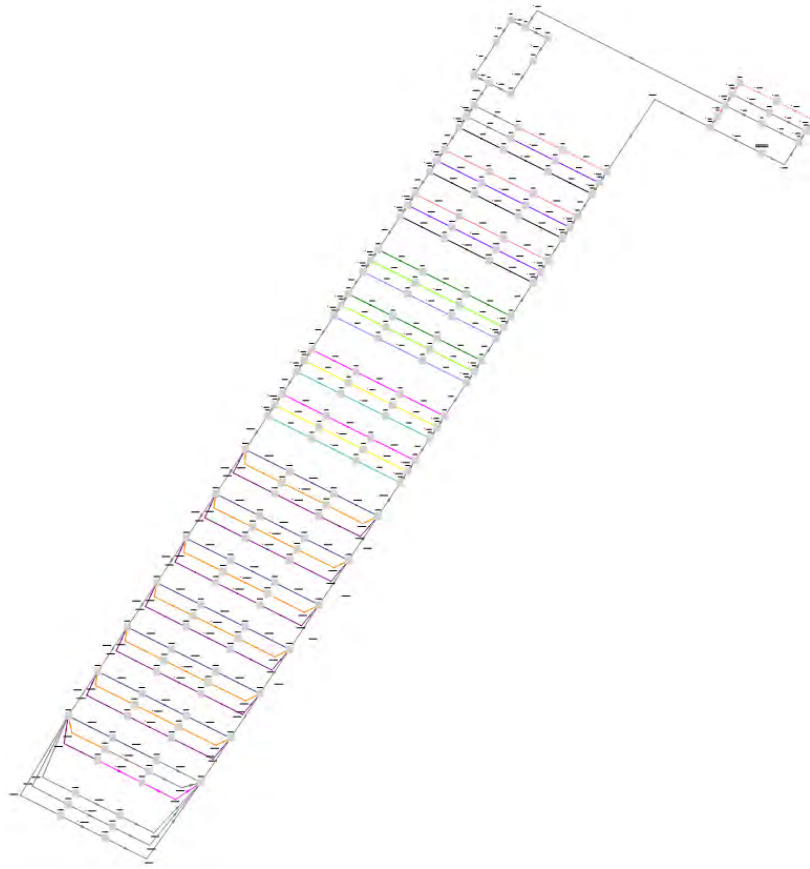
Furthermore, saving on power costs results in savings for carbon emissions as well. Based on information from the U.S. Energy Information Administration, there are 0.92 pounds of CO<sub>2</sub> emissions per kWh [1]. Using this information, a calculation can be done to determine exactly how much CO<sub>2</sub> is being used per year. Referring to Table 5B, multiplying columns A through D together by 7 days a week results in column E. This provides the emissions of CO<sub>2</sub> per year. As we can see, the 3 inch pipe has a much higher emission of CO<sub>2</sub> and assuming the EIC will run for 20 years, the emissions already make 40.6 million pounds. This is compared to 33.9 million pounds of CO<sub>2</sub> from the 4 inch pipe. The difference is a factor of 1.9, so these calculations should be used for determining the best pipe size, not only for cost, but in terms of lower carbon emissions as well.

**Table 5B:** Carbon Emissions comparison for operating 3 inch pipe vs. 4 inch pipe

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
<b>Pipe Size (in)</b>	<b>Power (kW)</b>	<b>Operating Weeks per year</b>	<b>Operating Hours per Day</b>	<b>Pounds of CO2 per kWh</b>	<b>Pounds of CO2 emission per year</b>
3	470.10	28	24	0.92	2,034,442
4	391.86	28	24	0.92	1,695,845
				<b>Savings per year</b>	338,598

### 1004 RF System

The 1004 RF System is angled like the 1010 building, except it is in the 4 O’Clock location of the EIC ring. This system is much smaller than the previous system and includes only one pumping circuit with 15 units. The full configuration is shown below in Figure 9.



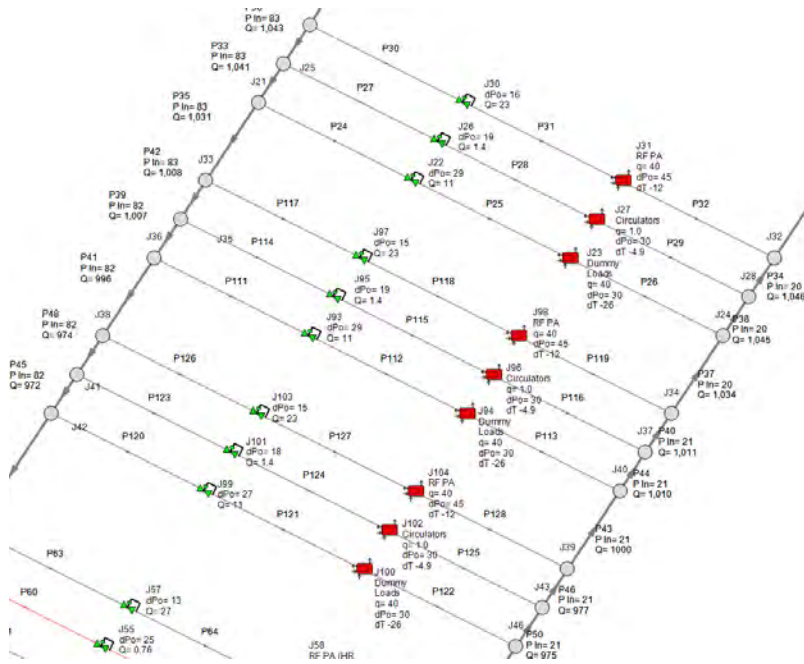
**Figure 9:** RF 1004 Full Configuration

Although the system is smaller, the units are not all uniform. This is much harder to design since all components need to be carefully matched to the respective cooling specification. There are a total of 5 groups and the data is summarized in Table 6 below, which includes the number of units each.

**Table 6:** RF 1004 Cooling Specifications for each group

<b>Number of Units</b>	<b>Heat Rejection Component</b>	<b>Heat Rate (kW)</b>	<b>Flow Rate (gpm)</b>	<b>Pressure Drop (psig)</b>	<b>Temperature Rise (°F)</b>
3	Power Amplifier	40	22.63	45	12.1
	Circulator	1	1.40	30	4.9
	Dummy Load	40	10.53	30	25.9
2	Power Amplifier	70	26.54	44	18
	Circulator	1	0.76	30	9
	Dummy Load	70	17.69	30	27
2	Power Amplifier	80	30.33	44	18
	Circulator	1	0.76	30	9
	Dummy Load	80	20.22	30	25.9
7	Power Amplifier	120	67.9	45	12.1
	Circulator	3	4.21	30	4.9
	Dummy Load	120	31.59	30	25.9
1	Power Amplifier	50	28.30	45	12.1
	Circulator	1.3	1.76	30	4.9
	Dummy Load	50	13.16	30	25.9

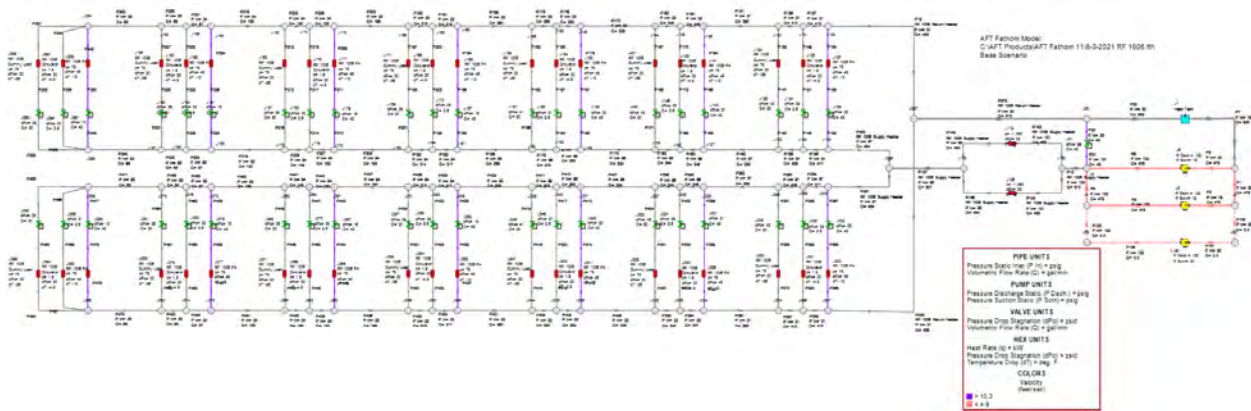
The first group with three units is shown in the image below in Figure 10.



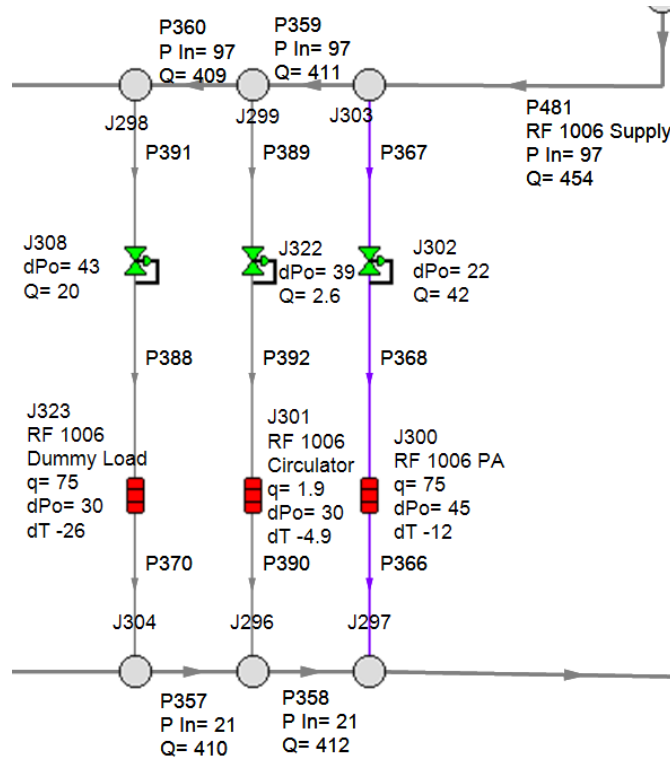
**Figure 10:** RF 1004 close-up of first group of units

### 1006 RF System

The 1006 RF System is laid out vertically in a horizontal building. All 14 units are identical, which makes it easier to model. The units have been laid out in two rows of 7. The full model is in Figure 11, while a close-up of one RF unit is in Figure 12.



**Figure 11:** RF 1006 Building with 14 RF Units and Data

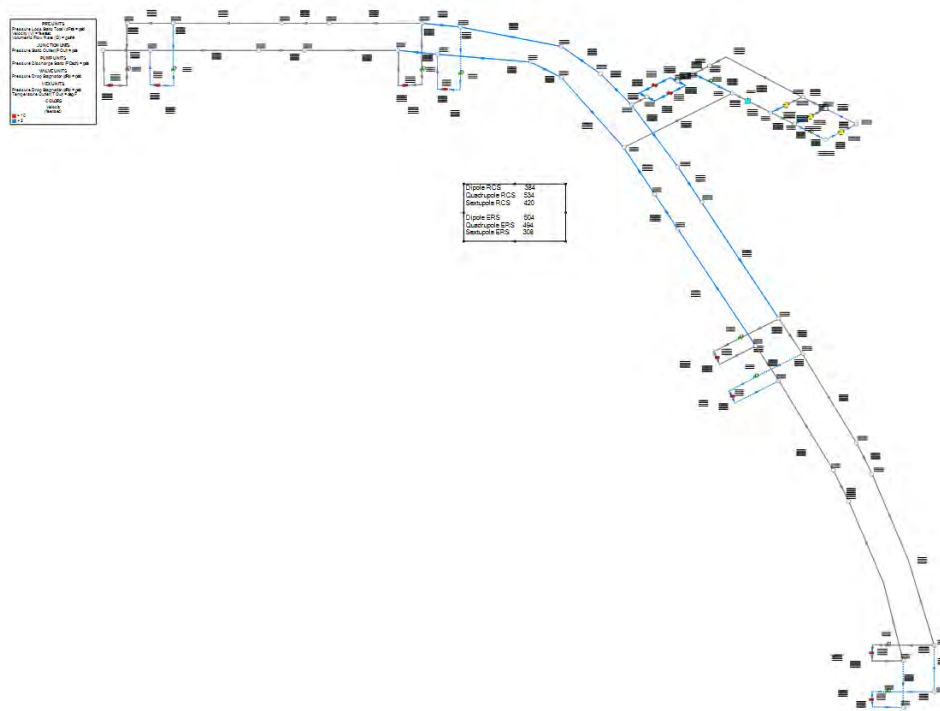


**Figure 12:** One unit from RF 1006 Layout zoomed in

#### IV. Magnets and Power Supplies

The magnet and power supply system was made to follow the circumference of the ring, like the vacuum system. Since the information for the exact number of magnets is currently in flux, the best method of approach was to create 4 magnet and power supply cooling components per quadrant. A master spreadsheet provides the information for each magnet for the ring, so the total flow rate and total heat load was calculated. Then, the values were divided by 16 to provide a single value for each of the control valves and heat rejection components. The total flow rate was found to be 999.1 gpm and the total heat load was 4381.90 kW. Therefore, each heat rejection component would need to have a flow rate of 62.44 gpm and a heat load of 273.86 kW. Values for the power supply system have not been implemented into the simulation, so place

holders have been set for them. Closed control valves were used to keep any flow from going through these areas. The finished magnet layout is shown below.



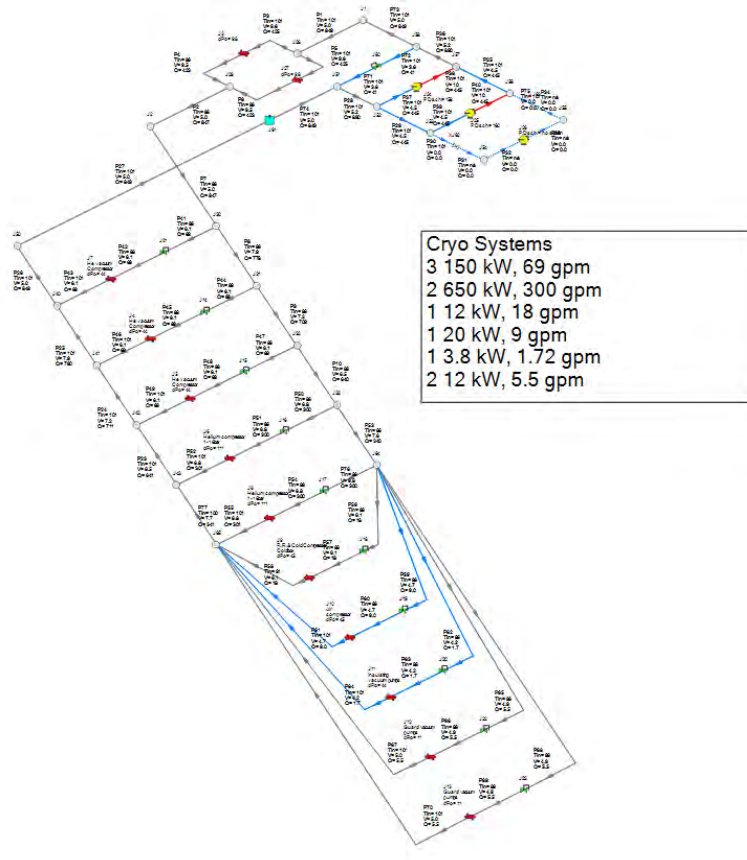
**Figure 13: Magnet System**

## V. Cryogenics

### 1002 Cryo

The 1002 Cryo building was dimensioned to be approximately 140 feet in length and 50 feet in width. The building layout has not been finalized, so the position of the systems are arbitrary. Each component in the system is grouped with components of the same type. The system is made up of 10 heat rejection components: 3 vacuum compressors, 2 helium compressors, 1 R&R cold compressor, 1 air compressor, 1 insulating vacuum pump, and 2 guard vacuum pumps. A manifold or “home run” layout was made at the end because the supply and return headers need to be 2 inches in diameter or more. Once the size goes below, the pressure in the system starts to get very high. To keep it under control, the manifolds are used. The layout is in Figure 14.





**Figure 14:** 1002 Cryogenics System Overall Layout

### 1006 Cryo

The 1006 Cryo building was dimensioned as 30.32 feet in length and 58.95 feet in width. This layout is very similar to the 1002 Cryo building with a manifold portion at the end to prevent higher pressure. The 1006 building has only 8 units, so it is much smaller. Figure 15 provides a representation of the units.

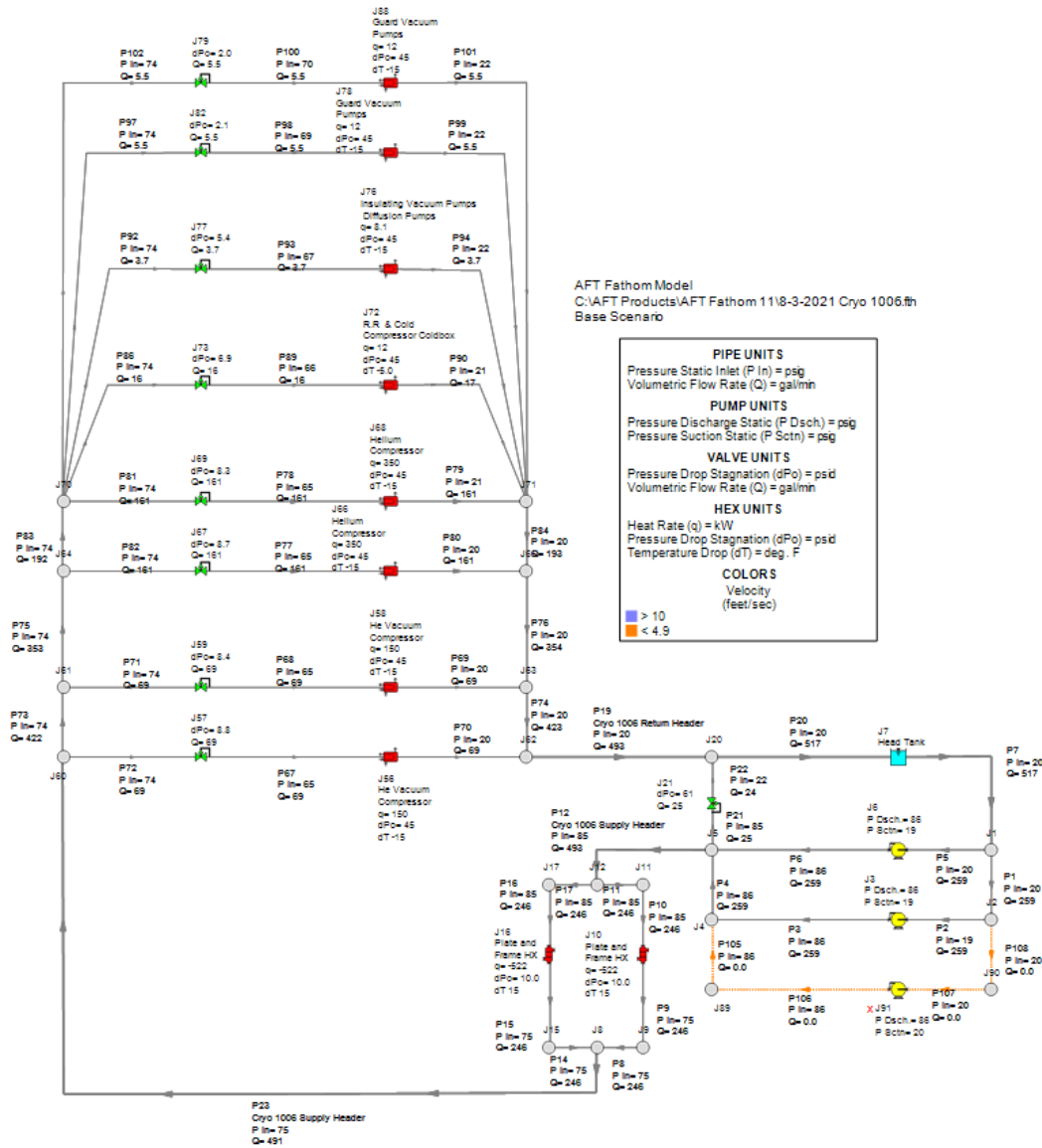
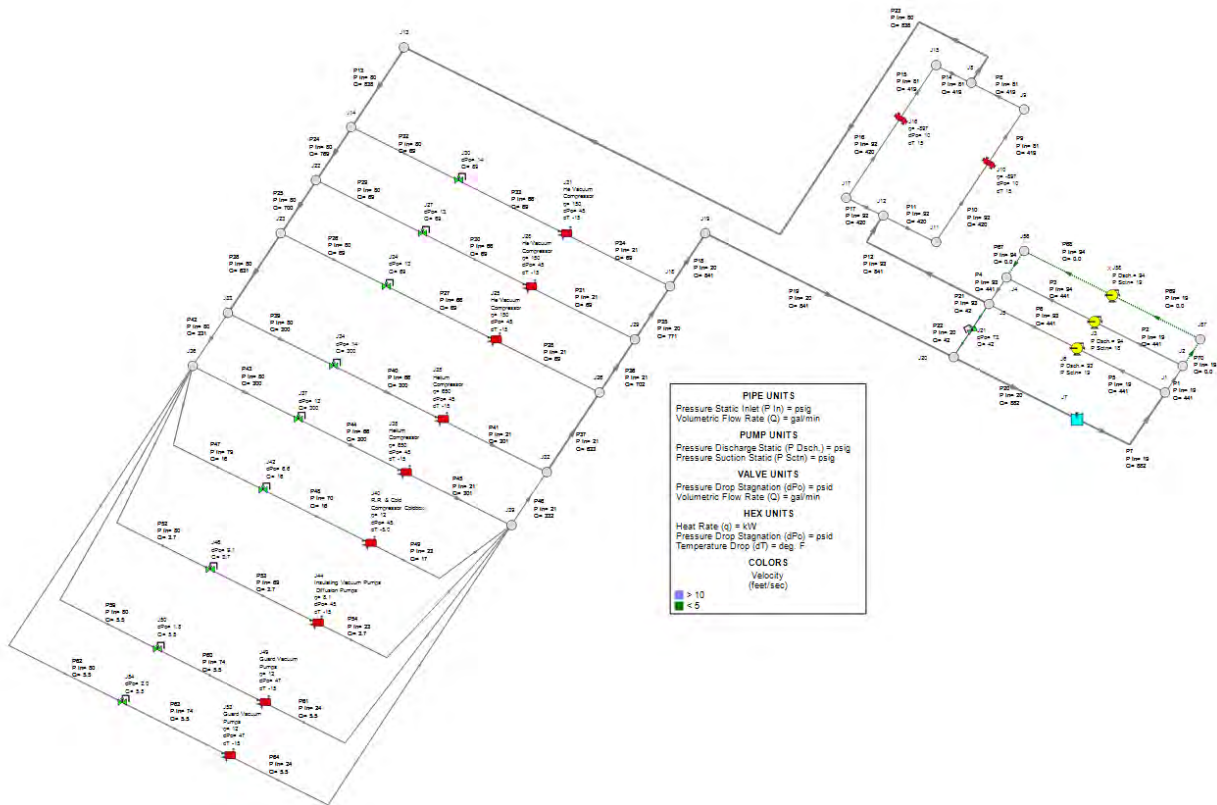


Figure 15: Layout of 1006 building with 8 heat rejection components

### 1010 Cryo

The 1010 Cryo building was dimensioned from the HDR Drawings to be 101.33 feet in length and 48 feet in width. The full layout of the components within the building were not finalized, so the cooling layout was made to follow the building dimensions. This system had 9 units in total and had the same units as the 1006 Cryo system, except it also includes an insulating vacuum pump. This is displayed in Figure 16.



**Figure 16:** 1010 Cryo building with manifolds and pumping system

## VI. Conclusion

By designing large-scale models in a simulation software, we were able to create a modeling tool that can be used to reduce the cost and improve effectiveness of the cooling system, thus lowering the carbon footprint by reducing the power consumption. The team utilized principles of fluid dynamics to design a cooling system model that would handle the heat rejection requirements for the EIC accelerator support systems. Recommendations would be to get real pump and manufacturer data to update the model to replicate real-life cooling systems. The model also should be updated to replicate design changes to the EIC. Additionally, the model should continue to be used for cost savings and carbon emission savings. Ten systems were created during the summer program. We were able to fully complete our project for such complex cooling systems that would contribute meaningful results to the EIC.

## **VII. References**

[1]“Frequently Asked Questions (Faqs) - U.S. Energy Information Administration (EIA).”

*Frequently Asked Questions (FAQs) - U.S. Energy Information Administration (EIA)*, 15 Dec. 2020, [www.eia.gov/tools/faqs/faq.php?id=74&t=11](http://www.eia.gov/tools/faqs/faq.php?id=74&t=11).

## **VIII. Acknowledgements**

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

We would also like to thank Charlie Folz and Ram Srinivasan for guiding us through the project and providing us exposure to several topics on cooling systems and fluid dynamics. They met with us regularly to provide their career advice, experience, and help for designing the project. We want to thank the EIC Infrastructure Team, the CAD Water Group, the RF Systems Group, Cryogenics Group, Magnets Systems Group, the Vacuum Group and the Technical Systems Division for all of their data inputs and suggestions. We also thank them for meeting with us to go over the systems in depth to match the cooling requirements.

# Optimizing neural networks for next-gen AI

Andrew Deutsch<sup>1</sup>, Yihui Ren<sup>2</sup>, Sandeep Mittal<sup>3</sup>

<sup>1</sup>*Department of Physics, Stevens Institute of Technology, Hoboken, NJ 07030*

<sup>2</sup>*Mentor and* <sup>3</sup>*Co-mentor, Computational Science Initiative, Brookhaven National Lab, Upton, NY 11973*

*August 11, 2021*

Most scientific data challenges involve real-time decision making on high-volume data flows. Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs) have been promising solutions for many challenging problems, but are limited to training and evaluation in an offline manner on a Graphics Processing Unit (GPU). In this work, we aim to transform pre-trained DNNs and CNNs into Spiking Neural Networks (SNNs), a better, biologically inspired form that can be implemented and deployed on novel hardware. Unlike discrete matrix multiplication based DNNs, an SNN acts on continuous trains of signals. First, we investigate the sparsity nature of DNNs, pruning techniques, and their connections to SNNs. We then investigate DNN to SNN conversion using NengoDL, an existing framework based on TensorFlow. Specifically, we optimize the accuracy of converted models by properly scaling the firing rates of neurons and applying a lowpass filter over the spike trains. We systematically run these experiments using the popular benchmark image datasets MNIST and CIFAR-10. For further optimization, we limit scaling of firing rates to only one layer and we find evidence that scaling late layers in a converted DNN and early layers in a converted CNN leads to higher accuracy. Further, we investigate the cause of this discrepancy by extracting firing rate data from both networks. Our findings provide insight into the co-design of novel energy-efficient neuromorphic chips for SNNs.

## I. INTRODUCTION

Spiking Neural Networks (Figure 1) are fundamentally time-dependent, as they operate on continuous trains of discrete spikes. Instead of representing information as a vector of real numbers, as is done with conventional DNNs and CNNs, the SNN encodes these values in the frequency domain with higher spiking rates corresponding to higher values and vice versa. The Rectified Linear Unit (ReLU) activation functions at each neuron of conventional networks are replaced with Leaky Integrate and Fire (LIF)

neurons (Figure 2), which compute a membrane potential as a weighted sum of the input spikes followed by exponential decay. When the

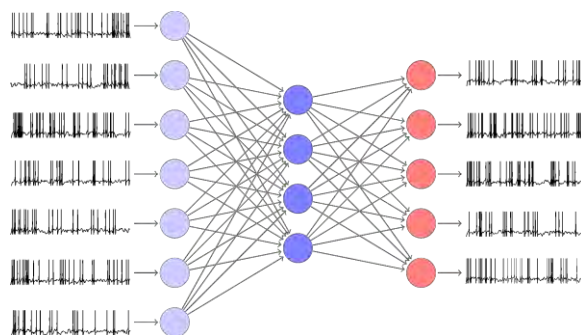


Figure 1: An illustration of a simple 3-layer Spiking Neural Network, complete with input and output spike trains.

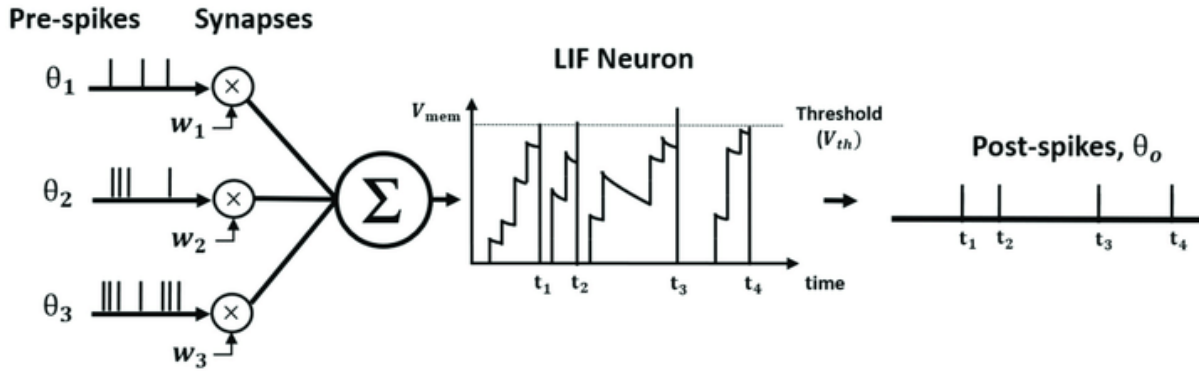


Figure 2: The Leaky Integrate and Fire (LIF) neuron, transforming three presynaptic spike trains into a single postsynaptic spike train.

membrane potential exceeds a threshold, the neuron fires, thus transforming any number of input spike trains to one output spike train.

The sparsity nature of SNNs means they provide approximately the same functionality as DNNs and CNNs, but much less communication is required. While all of the results in this paper are obtained from tests on a digital computer, SNNs are capable of implementation of novel energy-efficient analog neuromorphic chips. It is for this reason that their optimization is of great importance.

We begin with an analysis of network pruning techniques. Network pruning is a common method of reducing the size of a neural network while maintaining a desired level of accuracy. During the training process, neuron weights that are found to have little to no effect on the outcome of the network are sent to zero. The result is that much less information is required to store these networks and, in some cases, accuracy can improve. Often times, the prescribed network is much larger than needed to solve the problem at hand. We can further specify a sparsity schedule to prune the network after a particular epoch. In this work, we test the limits of network pruning on a simple CNN trained to read handwritten digits in the MNIST dataset (Figure 3). While network pruning is a crucial part of optimizing

any network, it is not specific to SNNs and we instead choose to focus on optimizing the conversion process of DNNs and CNNs to SNNs.



Figure 3: A small subset of the MNIST image dataset of handwritten digits used to train these pruned networks. It consists of 70,000 images belonging to 10 classes.

Our analysis of SNNs begins as a two-dimensional optimization problem. We then test the effects of scaling the firing rates in individual layers. Finally, we justify our results by directly measuring the firing rates in layers throughout our SNNs. All SNN optimization tests are systematically performed using NengoDL on the benchmark image dataset CIFAR-10 shown below.

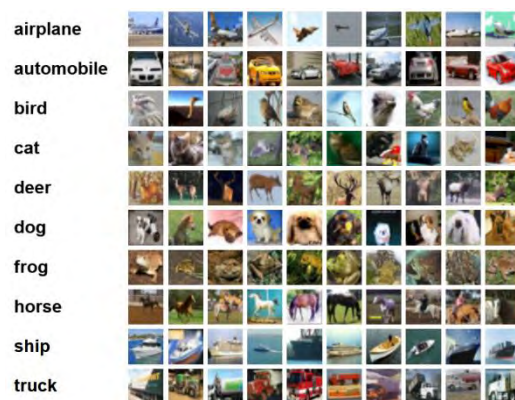


Figure 4: The CIFAR-10 image dataset used to train and test our SNNs consists of 60,000 images belonging to 10 classes.



## II. NETWORK PRUNING

To test the effect of network pruning, we construct a simple CNN in the TensorFlow framework. This network consists of two sets of 3x3 convolutional and 2x2 pooling layers with 28 and 56 neurons, respectively. These layers are followed by a Flatten() layer, a fully connected Dense() layer with 300 neurons and finally a Dense() output layer with softmax activation functions. We utilize the traditional ReLU activation function for all other layers.

The MNIST dataset is then normalized to make all input values between 0 and 1, and one-hot encoding is used on the labels.

Now, we are ready to train out network with varying levels of sparsity. Sparsity values range from 0% to 90% and represent the proportion of weights in the original network sent to zero during training or, equivalently, the reduction in size of the network. The sparsity schedule is specified to be a polynomial pruning which takes effect after epoch 5. The plots below show the training accuracy recorded over 20 epochs for each of the networks. For simplicity, we only present the training accuracy for networks with sparsity above 60%.

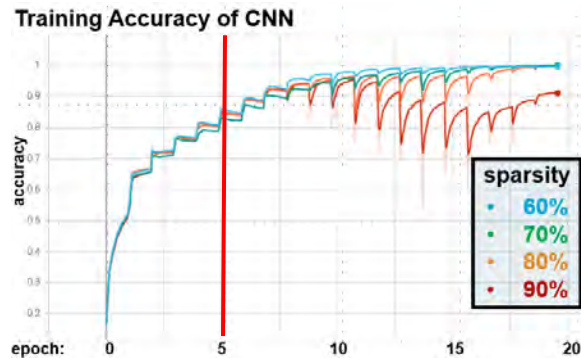


Figure 5: Training accuracy over 20 epochs of the same CNN pruned with varying sparsity after epoch 5.

Our findings indicate that this simple CNN can be pruned by up to 80% (only 20% of the initial weights remain) before the training accuracy

begins to suffer. Only at 90% sparsity do we observe a dip in training accuracy after 20 epochs. This result hints at the ease of training networks for MNIST; very few neurons are required for a simple CNN to correctly interpret these handwritten digits 99% of the time. This result serves mainly as a proof-of-concept that accuracy can be maintained while drastically reducing the size of a network.

## III. SNN OPTIMIZATION

To optimize the conversion process of DNNs and CNNs to SNNs, we make use of the NengoDL framework. This provides us with the tools necessary to swap neuron types and encode all information in the spike trains. However, a brute force translation to SNNs usually results in very poor accuracy. To remedy this, we make use two parameters: `synapse` and `scale_firing_rates`. The `synapse` parameter is the time constant  $\tau$  of a low-pass filter which is applied over all the spike trains. This behaves as a rolling average in an exponential window function.

$$f(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

The `scale_firing_rates` parameter simply scales up the firing rates so as to increase the amount of information propagating through the network while making communication less sparse.

For the sake of comparison, we run all the following optimization tests on both a converted DNN and a converted CNN, with architectures shown to the right.

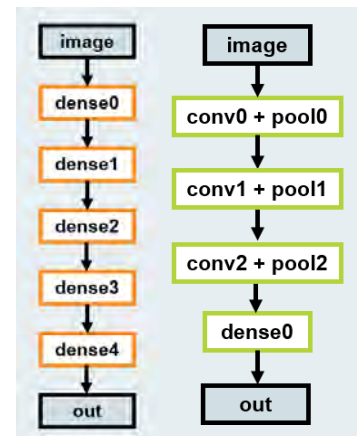


Figure 6: Architectures for DNN (left) and CNN (right)

#### IV. SYNAPSE AND SCALING

We are now presented with a two-dimensional optimization problem: how can we obtain optimal test accuracy for our converted SNN as we vary both synapse and scale\_firing\_rates? To investigate this problem, we generate a heat map of the test accuracy for various combinations of these two parameters. Specifically, we vary synapse from 0 (None) to 0.045 seconds and scale\_firing\_rates from 1 to 91.

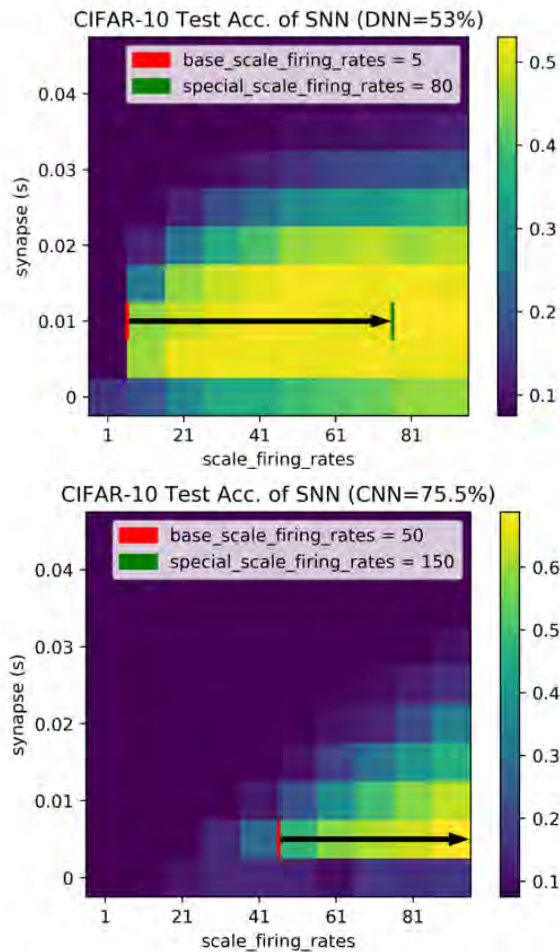


Figure 7: Test accuracy heat maps for the converted DNN (top, with non-spiking accuracy of 53%) and CNN (bottom, with non-spiking accuracy of 75.5%). Arrows illustrate scaling of individual layers (Section V)

Accuracy is observed to increase with scale\_firing\_rates directly, while synapse has an optimal value in both cases. While increasing

the firing rates results in higher accuracy, this comes with a tradeoff: communication is less sparse. As we increase the firing rates to infinity, the accuracy of the SNN approaches that of the non-spiking network, but this defeats the entire purpose of the SNN as it will require an enormous amount of information processing.

Somewhat more interestingly, as the low-pass filter time constant synapse increases, we see an initial jump, followed by a steady decrease in accuracy. This means we can extract these optimal values of synapse for each network (0.01 seconds for DNN and 0.005 seconds for CNN). The behavior of synapse indicates that there exist specific low-pass filters which provide optimal accuracy for each network.

#### V. SCALING BY LAYER

NengoDL also allows us to limit scaling of firing rates to specific layers, posing another optimization problem. Perhaps we can achieve the same levels of accuracy by only scaling the firing rates in one layer. We do exactly this, only we provide a base scaling to all layers so as to maintain decent accuracy. This can be visualized in Figure 7, where we scale all layers by base\_scale\_firing\_rates and only one layer by special\_scale\_firing\_rates. For example, the following is the scaling configuration in the spiking DNN for dense4.

LAYER:	input	dense0	dense1	dense2	dense3	dense4	output
SCALE:	5	5	5	5	5	80	5



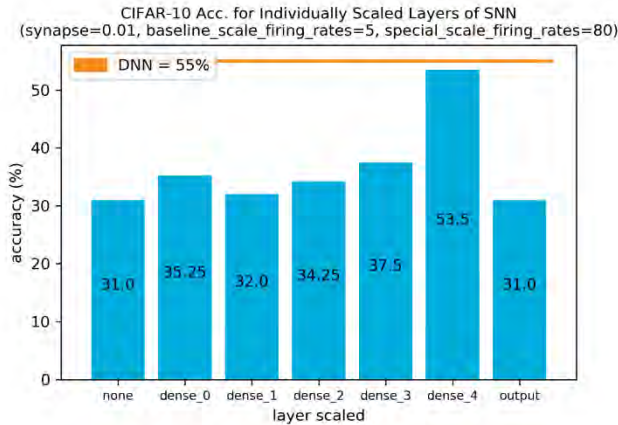


Figure 8: Test accuracy of the spiking DNN by layer scaled.

Here, we notice that for the converted DNN the accuracy is universally poor when only one layer is scaled, with the exception of dense4, the last hidden layer. In this case, the accuracy is very close (53.5%) to the baseline non-spiking accuracy (55%). Scaling the output layer has no effect. These results indicate that scaling later, rather than earlier, in the converted DNN leads to higher accuracy. The final hidden layer appears to be especially important.

Now, we repeat this test for the CNN. Scaling of pooling layers was found to have no effect, so they are omitted from the graph below.

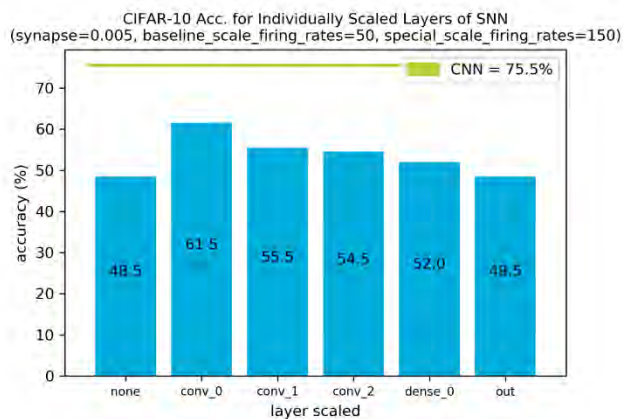


Figure 9: Test accuracy of the spiking CNN by layer scaled.

We find the opposite to be true in the case of the converted CNN: scaling the initial convolutional layer results in the highest

accuracy. We see a steady decrease afterwards. Another discrepancy is the wider gap between the highest SNN accuracy (61.5%) and the baseline non-spiking accuracy (75.5%).

To eliminate unwanted degrees of freedom, both networks were subsequently changed so that the number of neurons in each layer was the same. The conclusion was essentially identical, showing that these results are not due to the relative number of neurons in each layer.

## VI. FIRING RATES

To further investigate the cause of these results, we directly probe the neurons in each network and extract firing rate statistics by layer.

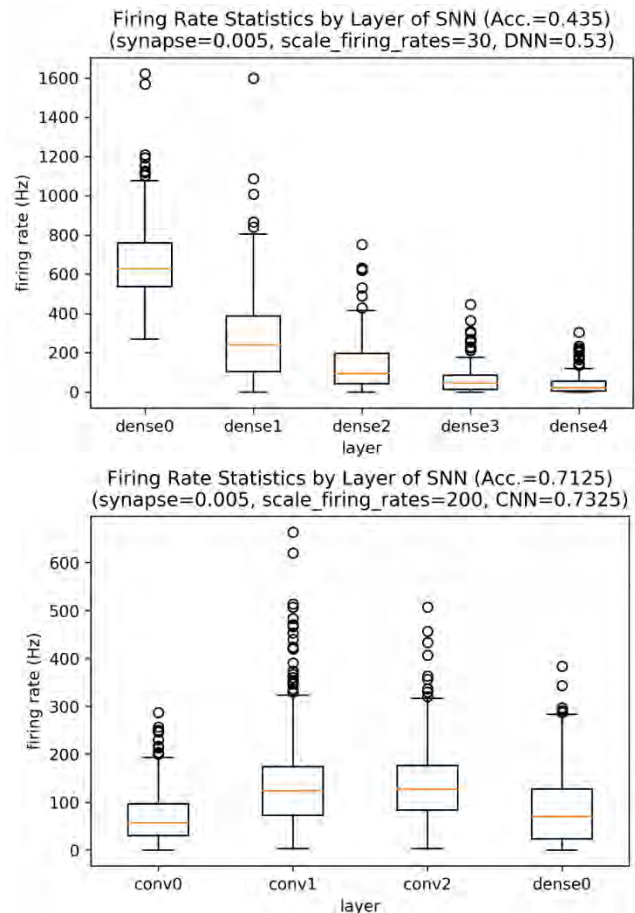


Figure 10: Box-and-whisker plots showing the minimum, median (orange), IQR, maximum, and outliers for the firing rate data by layer of the converted DNN (top) and CNN (bottom)

In the case of the converted DNN, spiking rates die off drastically as we move deeper into the network. The opposite appears to be true in the case of the converted CNN, where firing rates generally increase across the convolutional layers.

Looking at the firing rate statistics, together with the results from *Section V*, we can see a clear relationship that explains our findings. Scaling the firing rates is most beneficial when firing rates are comparatively low. In the DNN, this occurs in the final hidden layer and in the CNN, this occurs in the first convolutional layer. We can now conceptualize `scale_firing_rates` as a parameter which, when applied correctly, revives spike trains that have become too sparse.

To look at this result from another angle, we plot the firing rate data as a histogram with a best-fit curve.

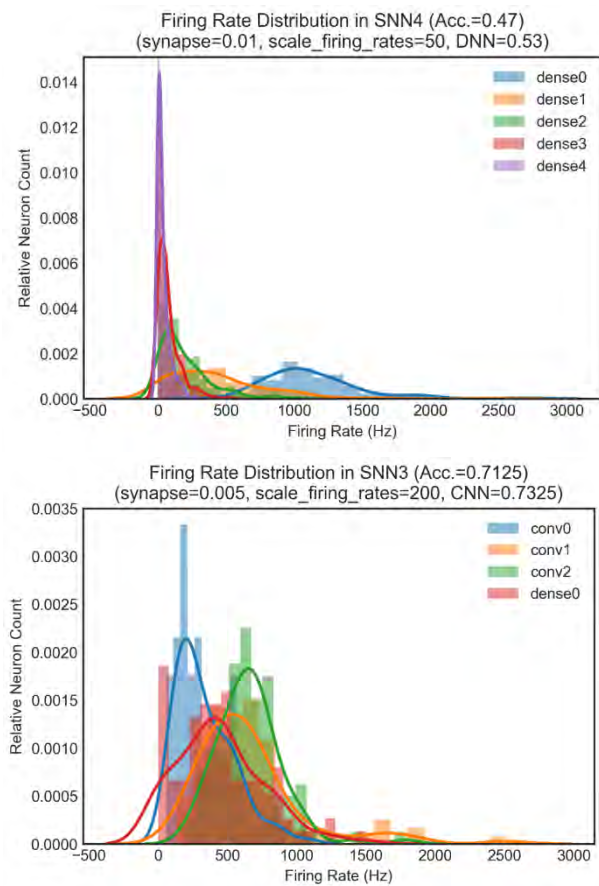


Figure 11: *Firing rate distributions by layer of the converted DNN (top) and CNN (bottom)*

Finally, we notice in the converted DNN an initially wide-spread distribution of firing rates, indicative of healthy activity. As we move deeper into the network, however, more and more neurons refrain from firing. In the converted CNN, the opposite is true: initially, many of the neurons are firing at low rates and, as we move deeper into the network, the distribution of firing rates has a greater deviation and fires at a generally higher rate.

## VII. CONCLUSION

These results indicate that the conversion of traditional neural networks to Spiking Neural Networks can be optimized (1) by pre-conversion pruning techniques, (2) through the specification of a low-pass filter time constant, and (3) by scaling the firing rates in layers with comparatively sparse spike trains.

Future work involves the expansion of these networks into deeper residual spiking networks, which are capable of obtaining higher test accuracy than the networks presented.

## VIII. ACKNOWLEDGEMENTS

I would like to give special thanks to my mentor, Ray, for all his guidance and support through the SULI program. I would also like to thank Sandeep Mittal for his thoughtful help and generosity. In addition, I wish to thank the Office of Educational Programs (OEP) and my OEP team leader Amy Engel for fostering a virtual environment conducive of advanced research and science communication. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## IX. REFERENCES

<sup>1</sup>Bodo Rueckauer, Iulia-Alexandra Lungu, Yuhuang Hu, Michael Pfeiffer and Shih-Chii Liu, “Conversion of Continuous-Values Deep Networks to Efficient Event-Driven Networks for Image Classification.” <https://www.frontiersin.org/articles/10.3389/fnins.2017.00682/full>

<sup>2</sup>Abhronil Sengupta, Yuting Ye, Robert Wang, Chiao Liu, Kaushik Roy, “Going Deeper in Spiking Neural Networks: VGG and Residual Architectures.” <https://www.frontiersin.org/articles/10.3389/fnins.2019.00095/full>

<sup>3</sup>K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

<sup>4</sup>Figure 1: <https://fzenke.net/index.php/2017/02/19/learning-in-multi-layer-spiking-neural-networks/>

<sup>5</sup>Figure 2: Lee, Chankyu & Sarwar, Syed & Panda, Priyadarshini & Srinivasan, Gopalakrishnan & Roy, Kaushik. (2020). “Enabling Spike-Based Backpropagation for Training Deep Neural Network Architectures.” *Frontiers in Neuroscience*. 14.119.10.3389/fnins.2020.00119.

<sup>6</sup>Figure 3: <https://deepai.org/dataset/mnist>

<sup>7</sup>Figure 4: <https://www.cs.toronto.edu/~kriz/cifar.html>

Analysis of the sPHENIX on-carriage data rack cooling water piping system

David Drysdale

Mechanical Engineering, SUNY Polytechnic Institute, Utica, NY 13502

Russell Feder

Physics Department, Brookhaven National Lab, Upton, NY 11973

## **Abstract**

The new sPHENIX detector at Brookhaven National Laboratory is under construction and will focus on studying strongly interacting particles produced in heavy ion collisions in the RHIC particle accelerator. The impact that sPHENIX will bring to Brookhaven National Laboratory is that scientists and physicians will have a better understanding of the Quark Gluon Plasma (QGP). My SULI 2021 internship focused on the engineering of the sPHENIX cooling water system used to remove heat from detector electronics and high performance detector server racks mounted on the sPHENIX carriage superstructure. I used the software package PipeFlow Expert to simulate and optimize the flow of water in the sPHENIX piping system, calculating pressure drop, flow rates and other hydraulic properties. I also used the 3D CAD software package AutoDesk Fusion 360 to view detailed models of sPHENIX and used those models to develop simplified simulation models in PipeFlow Expert. My accomplishments in this internship is that I learned to use the new engineering software packages. This helped me develop and grow a repertoire of engineering tools. To get a working model, we switched to using the pump model and selecting an ideal pump based on the pump curves. We knew the total flow rate needed based on the number of racks and flow rate at each rack. Based on that known flow rate we selected a pump that could provide the pressure-head needed to get water to all the racks.

# **Introduction**

## **A. Background**

The purpose of this paper is to discuss how I contributed to the cooling of the racks on sPHENIX. I used mainly Pipeflow expert to calculate desired flow rates and pressure drops for the racks of sPHENIX and what type of pipe material that needs to be used for sPHENIX. Also, I used Fusion 360 to look at the model in a 3d cad software. What I focused on is the derivation of the cooling system of sPHENIX. By using Pipeflow expert, I can model a schematic of what the piping system would look like and can calculate the desired flow rates and pressure drops in the system.

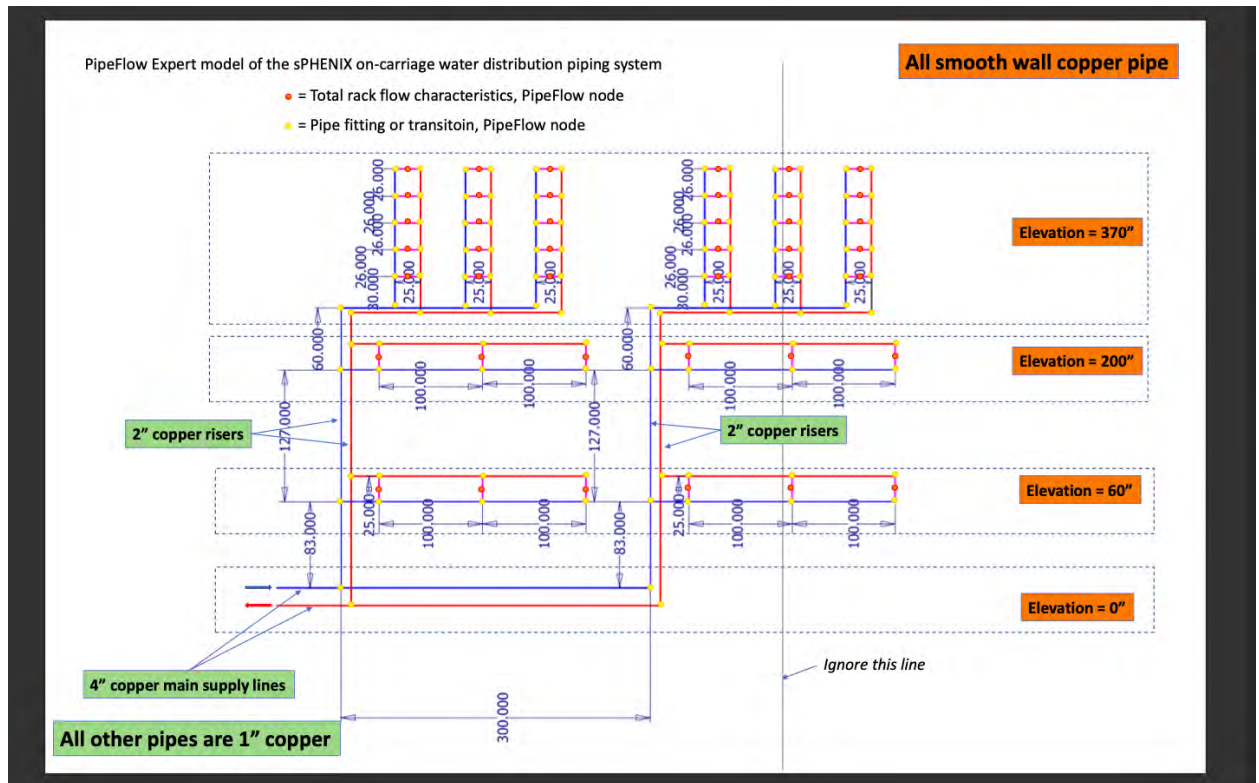
## **B. What is sPHENIX?**

sPHENIX is a new particle detector for the RHIC collider currently under construction that is going to analyze the “perfect fluid” called Quark Gluon Plasma (QGP)<sup>1</sup>. A QGP is generated from the collisions of high energy particles inside the Relativistic Heavy Ion Collider(RHIC).<sup>2</sup>. A group of the physicists collaborated and designed a new detector called sPHENIX based on experience gained on the PHENIX project as well as detectors at CERN on the LHC.. The “s” stands for the strongly interacting particles and PHENIX stands for anticipated use of key detector components and experiments location in the RHIC ring once the existing PHENIX systems complete their data-taking lifetime at the end of this year’s run<sup>3</sup>.

<sup>1</sup> sPHENIX, US Dept of Energy, “Introducing...sPHENIX”, June, 2016

<sup>2</sup> sPHENIX, US Dept of Energy, “Introducing...sPHENIX”, June, 2016

<sup>3</sup> sPHENIX, US Dept of Energy, “Introducing...sPHENIX”, June, 2016

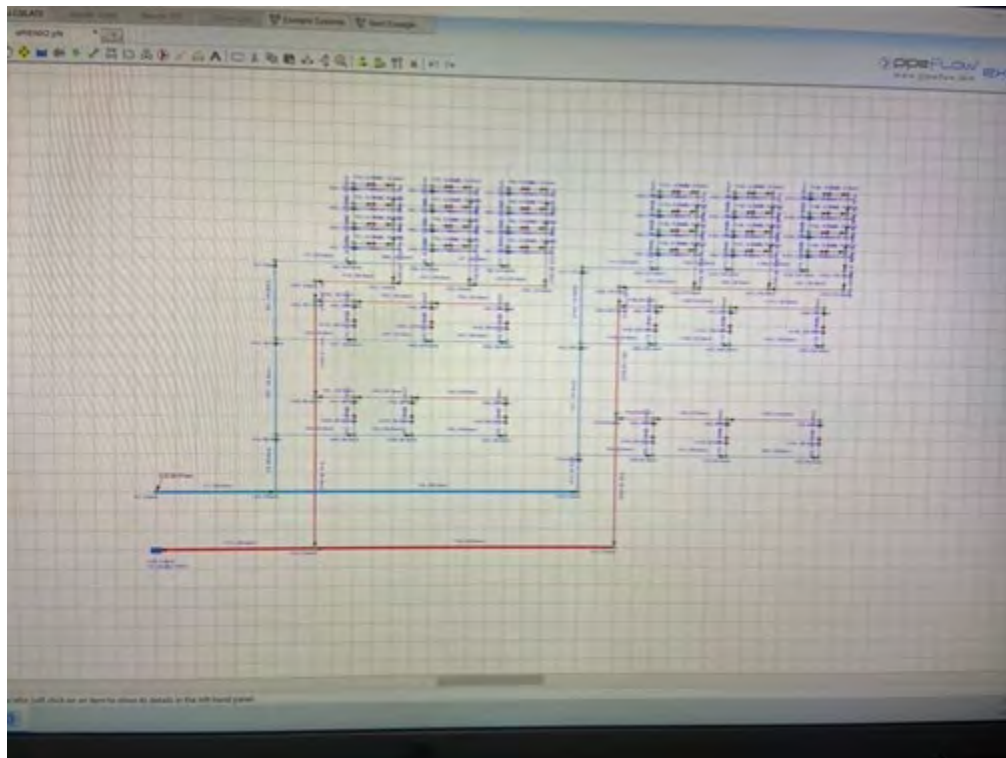


## Main Body

The scope of my project was to help analyze a cooling system for sPHENIX server racks. The figure above shows the first draft of the cooling system model in PipeFlow Expert. The blue lines indicate the supply and the red lines indicate the return. The purpose of this model is as the blue lines enter into the red dots and then turn into red lines. After the blue lines turn into red lines, the red lines will turn into blue lines again and



the process is a closed loop. The red dots indicate the server racks for sPHENIX as well as the yellow dots indicate the elbows and tees . The server racks in sPHENIX hold certain things that are important for sPHENIX. The figure below shows the model implemented in Pipeflow expert.



We were able to get this initial model to work. The only problem that we had was that there were negative pressure drops along the racks and the flow rate into the racks was very close to 0 gallons per minute. So in this model, you can see that on the blue side there is an arrow which indicates a demanded flow into that node. The indicated flow is 200 gallons per minute. So, my mentor and I decided to meet with sPHENIX engineers Dan Cacace and Robert Pisani to discuss this issue. The advice that Robert Pisani and Dan Cacace gave us was to put a flow restriction valve at the racks to set each of the racks to 8 gallons per minute and also change from a demand flow into a demand

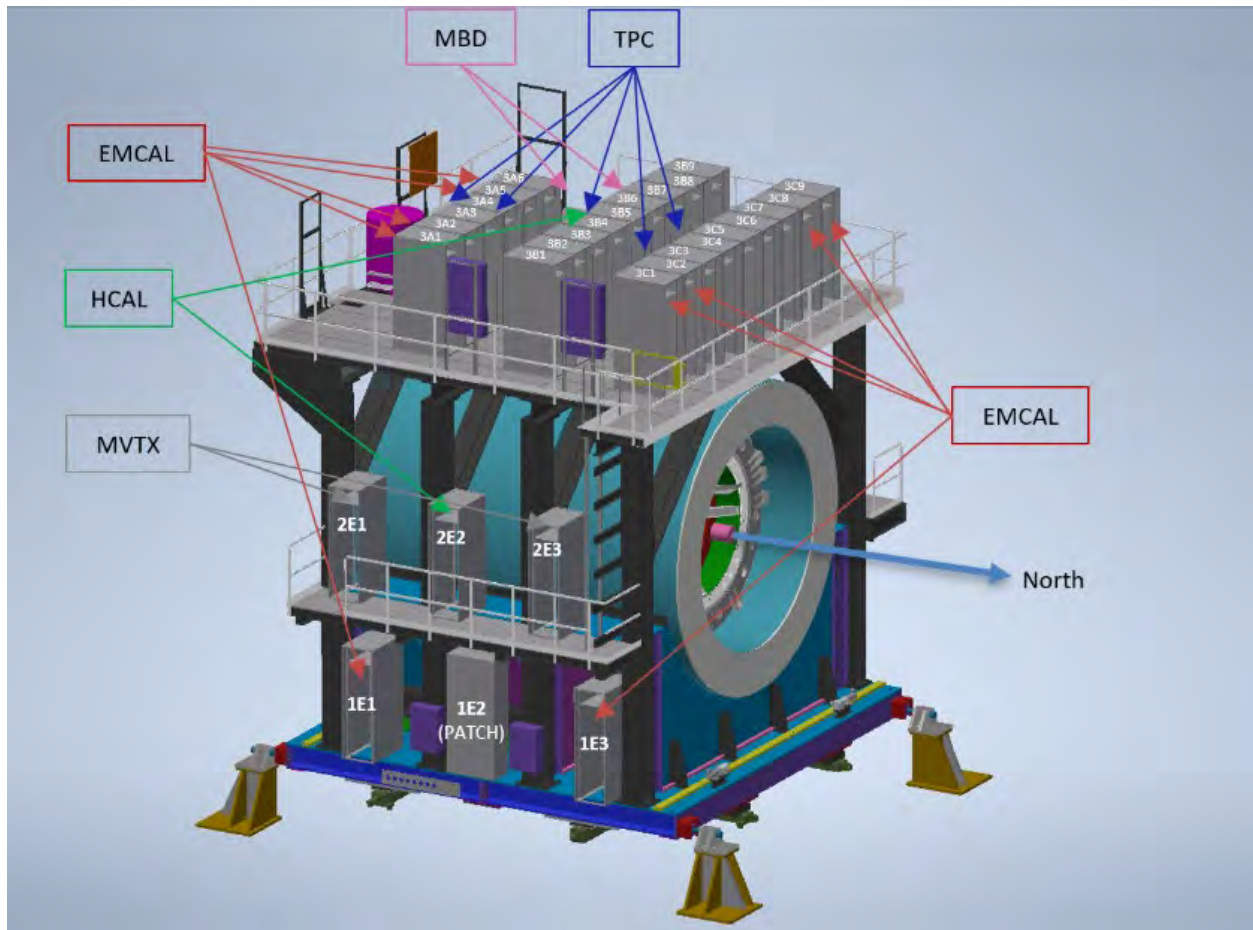


pressure by adding a tank with pressurized water. After making these changes, the flow into the racks were constant and the pressure drop was still negative and we were not sure why it was like that. Robert told my mentor and I to cut and paste a portion of the racks and see if we were able to get a positive pressure drop. I made a discovery that the negative pressure drops were a result of the tanks having 0 pressure. So, I changed the pressures so the difference would be 10 psi. For instance I set the supply pressure to 40 psi and I set the return psi to 30 psi. That fixed the issue. So, I implemented this into the existing model and it fixed that problem. The next modeling step was to change the layout of the sPHENIX model to match to ICD-42. ICD-42 is a Interface Control Document that the sPHENIX engineers use to define the configuration and requirements of the data racks. By doing this the PipeFlow Expert model will match the day-1 operational

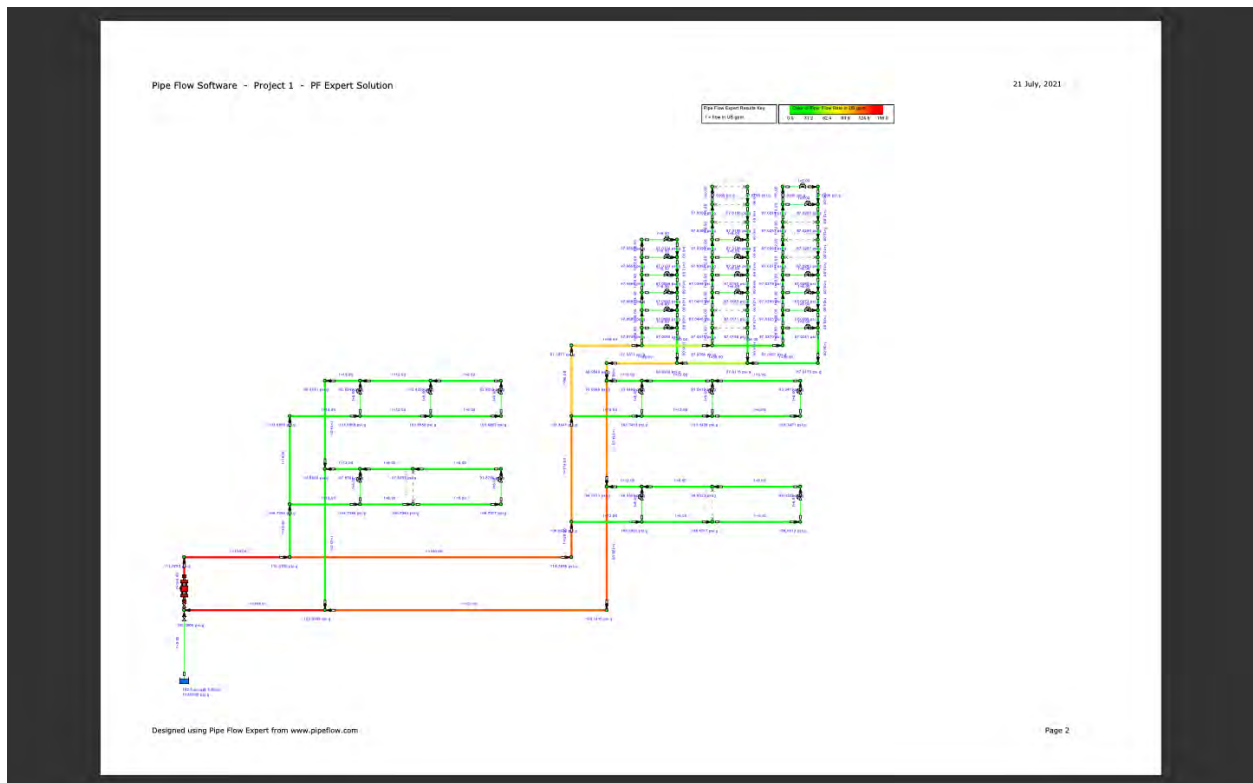
configuration

of

sPHENIX.



As you can see from this model, the top racks are laid out differently so we had to make some changes to our current model.



So, this is the final draft of sPHENIX in Pipeflow expert. The colors indicate the flow rate in the system. From green to red means slow to fast, respectively. We were able to get all the racks to have a constant flow rate as six gallons per minute. We have a maximum flow rate that we anticipated and that value was 276 Gallons per minute, and the maximum flow rate in this model is 156 gallons per minute which is a lot less than the max. There was a lot of ambiguity with the open tanks. By defining a specific pump we were able to determine exactly what pressure head we need for a given pipe diameter. We looked at what pump is needed for 4" pipes and what pump is needed for 3" pipes.

## **Conclusion**

\_\_\_\_\_ It was a great summer researching this. If I were going into my junior year of college, I would definitely apply for this position again. I was able to apply what I learned in class into this internship, and it helped a lot. I learned a new Cad software called Fusion 360 which I am still a beginner at now. I can use that information and can apply it to my projects in school. I thought I was going to be bored with the research that I did this summer. But on the contrary, I was really intrigued. I wish this internship was in person so I can actually see what I was working on. It made me realize that I should definitely try new things in my engineering career. I always thought that I wanted to go straight into aviation, but this internship definitely helped me open my eyes more.

## **References**

\_\_\_\_sPHENIX, US Dept of Energy, "Introducing...sPHENIX", June, 2016

Numerical simulations of small-scale junction field effect transistors

Dominique Duliepre, Department of Electrical Engineering Technology, SUNY at Farmingdale State College, Farmingdale, NY 11735

Gabriele Giacomini, Instrumentation Division, Brookhaven National Laboratory  
Upton, NY 11973

## ABSTRACT

The viability of the metal-oxide semiconductor field effect transistor (MOSFET) in substitution with the junction field effect transistor (JFET) in a low-noise input transistor application pivots on the length of the channels and the difference in length between them. In the past, JFETs were ubiquitous in front-end electronics as a low-noise input transistor, but it has been replaced by MOSFET integrated circuits (IC) for its low output-noise characteristics. With modern fabrication technology, a reduction in channel length between the source and the drain is an attainable method to diminish the interelectrode capacitance but to increase the transconductance ( $g_{m0}$ ) at zero gate-voltage. This research analyzes the viability of the junction field effect device physics and the variation of device geometry using the Silvaco Technology Computer Aided Design (TCAD). Through the utility of the Silvaco DeckBuild program, cylindrical model JFETs of submicron gate-lengths have been designed. The output characteristics, input characteristics and the transconductance of the transistors have been modeled with Tonyplot and external python compilers to compare the efficiency of current MOSFET transistor technology in front-end electronics. Concurrent on the analysis of the various geometries, the efficacious device iteration produced in the simulation should allow for a more efficient device while also reducing the gate-to-drain capacitance ( $C_{gd}$ ) and increasing the gate-to-source off-voltage ( $V_{gs-off}$ ) to compete with its MOSFET counterpart. This program has enriched my repertoire of Python programing, semiconductor device physics, geometry and specifically device doping characteristics. Furthermore, I am now a proficient user of both the Tonyplot, and the DeckBuild programs available in the Silvaco software tool.

## I. INTRODUCTION

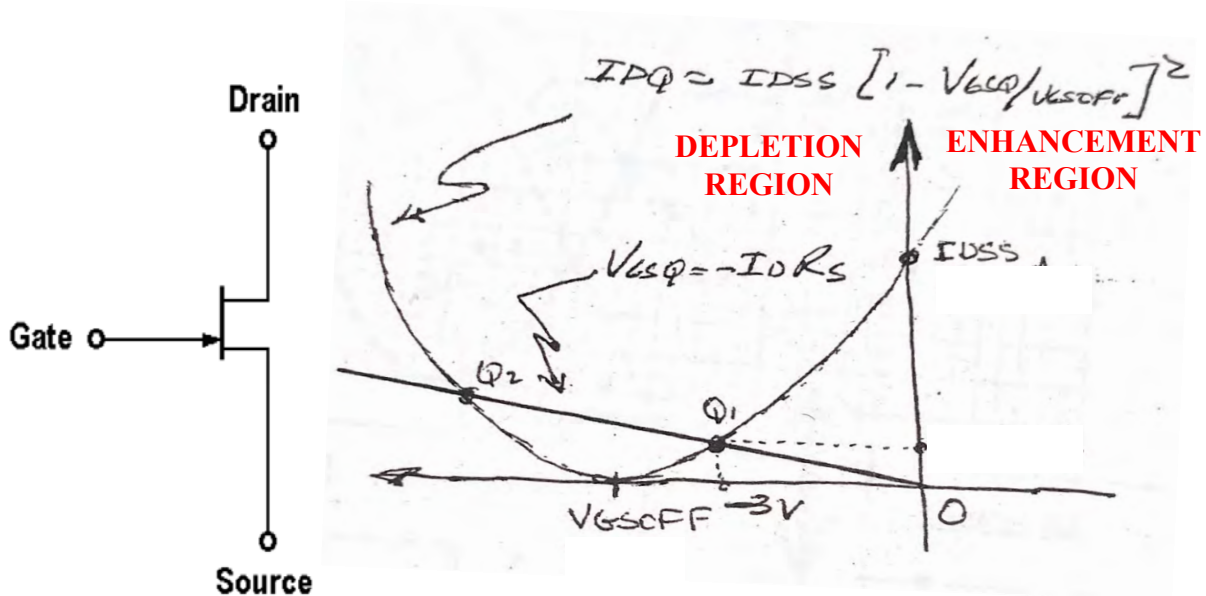
The Junction Field Effect Transistor is a normally-on voltage-controlled current source transistor (see FIG. 1.0) which means that the JFET is controlled by an electric field, operational only in the depletion region of device. The (p-type) gate of the (n-channel) JFET is biased with a voltage at or below zero to turn on the device, but as the device gate-length decreases it is found that  $V_{gs-off}$  increases rapidly, this makes it more inefficient than a MOSFET transistor.<sup>1</sup> This drawback is one of the crucial challenges with the JFET for low-noise applications. Another reason why the MOSFET has been adopted in modern-day industry is due to the metal-oxide layer shielding the gate terminal from the drain and the source terminals. The MOSFET can be manufactured in smaller dimensions compared to the current JFET alternative, this makes it convenient for small scale applications. The JFET has an upper hand on the MOSFET since its gate leakage current and  $1/f$  noise are essentially null.

## II. SCOPE & OBJECTIVE

The project consists of three stages. The first stage is to model the JFET using the Silvaco TCAD: this allows for the analysis of the transconductance characteristics, the interelectrode capacitance between the electrodes, the output impedance, and the doping characteristics. After a device passes the first analysis of the JFET, it is tested for frequency response to record the maximum input frequency before the output waveform becomes distorted. The next step is to be fabricated onto a silicon wafer for further practical simulations and examinations. The third and final steps of the project are the peer review (and possibly the proceeding of a patent application for the transistor). For the project, a JFET of high input sensitivity is of the essence while skewing the gate-to-drain interelectrode capacitance  $C_{gd}$  but also increasing the transconductance. As shown in **FIG. 2**, At this point, the inter-electrode capacitance developed at

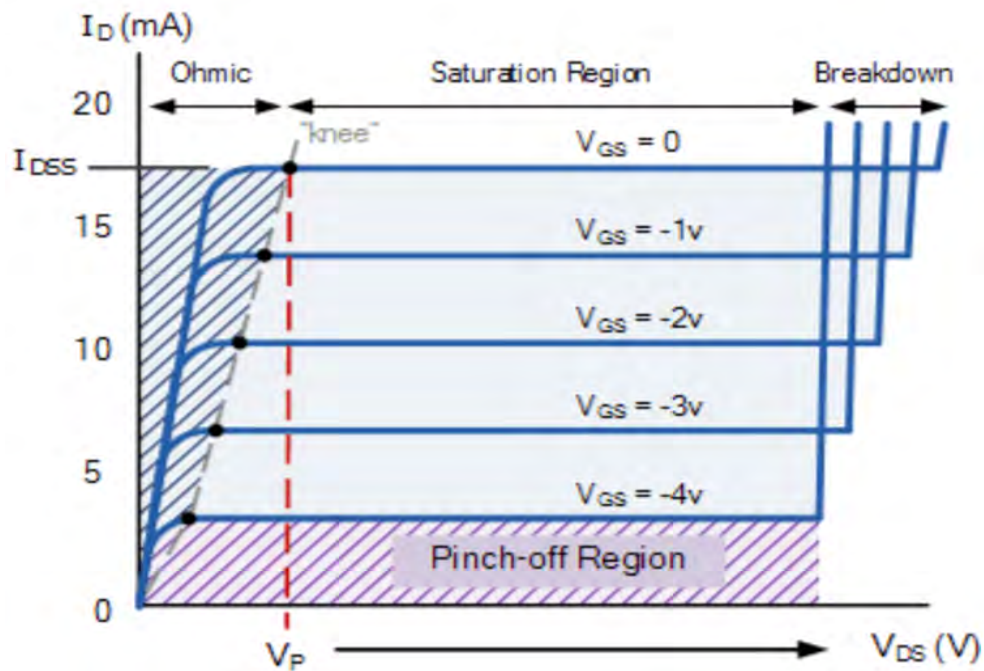


the Gate-to-Drain junction is at its lowest. This has been confirmed when testing the JFET gate-length of 2-microns, the gate-length of 1-micron, the gate-length of 0.5-micron and the gate-length of 0.2-micron. The primary test unveiled that  $V_{gs-off}$  decreases with respect to the narrowing of the boron implant below the top-gate electrode. The second test involved changing the doping concentration of the bottom-gate, and the substrate. The result was a lowered gate-to-drain capacitance at  $V_d = 12V$ . The subsequent test conducted consists of a tetrode mode transistor with the bottom-gate grounded for each of the JFETs. As expected, this series reproduced a transconductance of higher amplitude at  $V_g = 0V$  than all previous tests but also caused an increase in  $C_{gd}$ . After analyzing the results, a new mesh was needed to increase the data resolution. When the mesh was varied,  $V_{gs-off}$  decreased by two volts while maintaining the drain current position. The final test was an iteration of a tetrode once again but with new mesh coordinates and new doping profiles.



**Figure 1:** The  $n$ -channel JFET.

**Figure 1.1:** The input characteristic plot displays the quadratic polynomial for device operation between  $V_{gs-off}$  and  $V_{gs} = 0V$ .<sup>2</sup>



**Figure 1.2:** The output characteristic plot displays the V-I curves for JFET device operation.<sup>3</sup>

### III. METHODS & OUTCOMES

#### A. Deep-implant tetrode mode transistor

The first JFET test series (0.2-, 0.5-, 1.0- & 2.0-micron gate-length) conducted were in the tetrode mode; where the bottom-gate electrode is held at ground-zero (see FIG. 1.3 APPENDIX A). Whereas a transistor in the triode mode effectively synchronizes both the top-gate and the bottom-gate electrodes to the applied gate voltage. The first test product generated input characteristic plots with high  $V_{gs}$ -off at about 30V (see FIG. 1.5 & 1.8 Appendix A). When the 0.5 and 0.2  $\mu\text{m}$  gate-length transistor iterations were analyzed in FIG. 2.1 & 2.4 Appendix A, it is observed that  $V_{gs}$ -off increases at a high rate and that this also causes a negative impact on the transconductance of both transistors as seen in FIG. 2.6 Appendix A. The

high  $V_{gs-off}$  is a disadvantage to the efficiency of the device since it requires high-voltage to turn-off the device. It is well known that JFET transistors with high pinch-off voltage require larger drain-to-source voltage to enter the saturation region in FIG. 2.0 Appendix A; This low  $V_{gs-off}$  characteristic is a disadvantage to compact integrated circuits with insufficient space to regulate and convert high voltages. The low  $V_{gs-off}$  would allow for a higher overage in the applicable peak-to-peak input signal but that is arbitrary for the low-noise application of interest. The 0.2-micron iteration is an exaggerated example proving that a further reduction in  $V_{gs-off}$  will extend the Ohmic-region of the transistor, making it more applicable in large signal applications. The modern MOSFET technology defeats the low  $V_{gs-off}$  via its metal-oxide layer. The metal-oxide layer creates higher input impedance, these are the reasons why MOSFETs are highly used in power electronics.

Equation (1), the transconductance (measured in Siemens) at  $V_g = 0V$  is  $gm_0$ . It is essentially a measure of how much more influence the gate-voltage has on the drain-current compared to the drain-voltage on the drain-current.

$$gm_0 = dx \left( \frac{!''##}{|&'( ) *++|} \right) \quad (1)$$

Equation (2), with the  $gm_0$  computed any mutual transconductance value above or below  $V_g = 0V$  is verified with the mutual transconductance formula.

$$gm = gm_0 \left( 1 - \frac{\&'(}{\&'( ) *++} \right) \quad (2)$$

In FIG. 2.7 Appendix A, the R-out is the measure of the output impedance measured in Ohms; it is the inverse slope of the saturation region linear area on the output curve. The inverse slope is measured using a secant line on near the  $V_d = 20V$  pinch-off point onward. The R-out of the 0.2-micron gate-length was measured at  $V_d = 30V$ . Due to the higher saturation point on the curve, this may explain the change in slope at the 0.2-micron point where the R-out values are plotted using Python's Matplotlib 2D library. The transconductance was reviewed and the consensus was that a greater  $g_m$  and can be achieved while further decreasing  $C_{gd}$  and  $V_{gs-off}$ .

#### B. Deep-implant triode mode transistor with new doping

The second round of JFET tests conducted were iterations of the first, the bottom-gate boron implant doping concentration was reduced by a magnitude of  $2e13 \text{ cm}^2$ . This test was a movement in the right direction since it effectively increased the  $V_{gs-off}$  (see FIG. 3.0, 3.3, 3.6 & 3.9). As a result of the change in doping characteristics the gate-to-drain capacitance was also reduced. The  $V_{gs-off}$  decreased by over 18-volts for the 2.0- & 1.0-micron model which is a 60% improvement from the previous test iteration. For the 0.2- & 0.5-micron gate-length models, the  $V_{gs-off}$  increased substantially from -50v & -100V up to -15V & -30V. As shown in FIG. 4.7 Appendix B, the diffusion time at 1000 degrees centigrade was also reduced by 20-minutes. The phosphorus doping concentration for the channel region lowered by a magnitude of  $1e13 \text{ cm}^2$ . It had been predicted that the device data would make a compromise between the  $C_{gd}$ , and the transconductance. In terms of the interelectrode capacitance in FIG 4.1 Appendix B, the first test of the 2-micron gate-length model produced a  $C_{gd}$  at 22.5 femto-Farads and after the doping change it sits at 172 femto-Farads this indicates a 7.64x increase. For the 1.0-micron

gate-length model, the first test produced a Cgd just under 24 femto-Farads and after doping and changing the mode it sits at 155 femto-Farads this indicates a 6.45x increase. For the 0.5-micron model the Cgd was previously 13 femto-Farads and after the revisions it increased by 12x to 155 femto-Farads. The 0.2-micron model had emerged from 11.5 femto-Farads from the previous test to 139 femto-Farads, indicating a 12x increase. Another important observation has been made: this observation corresponds to the transconductance for the 1.0-micron model. From the previous model to the new one is a 1.1x increase in performance, but the transconductance of the 0.5-micron increased by 1.9x, the other gate-length models lie in within this range. The prediction was correct after compiling the data of each transistor gate-length in Python because it is understood that the deep doping implant method is sound for the 1.0- & 2.0-micron models. This method is not compatible for the development of the 0.2- & 0.5-micron-gate models because the transistor channels would not pinch-off below 5V, a new approach is necessary for this to be achieved.

### C. Shallow-Implant tetrode mode transistor with new doping & mesh

The new method commenced with a completely revised doping and mesh geometry; the shallow-implant device with an implant kinetic energy of 10 kilo-electron volts was modeled with a denser mesh due to the reduced depth. The mesh point-concentration will allow for more accurate results per unit-area of the structure. This sacrificed time for a more accurate computational analysis of the device. As a result of also shallowing the boron doping dept of the top-gate The Vgs-off of this test group increased further. This was done intentionally (see FIG. 4.4, 4.7, 5.0 & 5.3 Appendix C), the Vgs-off now is near an efficacious -3.0V for each of the gate-length variants which leads to a low pinch-off voltage at the output to properly close the

channel without high-voltage. This is very efficient for real-world low-noise applications where high voltage is not easily generated without consuming valuable space and power on a circuit. At the output, the saturation drain-current of the device at  $V_g = 0V$  is just below 1.75 milli-amperes compared to the previous iteration which produced an order of magnitude higher due to the doping variation and depth of the implants. Prior to the remodeling it was predicted that interelectrode capacitance would be significantly reduced across each of the electrode junctions throughout the entire device not just  $C_{gd}$ . As seen in FIG 6.0 Appendix C, the interelectrode capacitance when  $V_g = 0V$  is measured at a the single-digit picofarad range for the 0.2-micron and the 0.5-micron range, this has been improved by 1 order of magnitude from the last deep-implant series tested. The interelectrode capacitance  $C_{gd}$  for 1.0- & 2.0-micron produced similar improvements. Although the low transconductance of the device has yet to be rectified Without diffusion characteristics considered, this model has the potential to operate very efficiently with the improvement of the gate-to-gate voltage and the  $C_{gd}$  in FIG. 6.0 & 6.1. This in turn can lead to an increase in device performance and a reduction in overall device dimensions for small integrated-circuit applications. In prospective, this iteration shall be refined in DeckBuild to maximize the performances.

#### **IV. CONCLUSION**

The shallow-doping model is near completion, further refinery of the device series will lead to an increase in  $g_m$  and to account for the diffusion times. Subsequently, the theoretical frequency analysis of each gate-length model must be done to address the attainable frequency bandwidth with respect to the transistor gain. Once this this stage has passed, it will be reproduced onto a 10cm silicon wafer for further examination. The MOSFET has been in use for

quite some time now but with modern fabrication techniques the JFET remains of great potential for future innovations both in terms of relative efficiency and cost effectiveness.

## V. REFERENCES

### - Literature

<sup>1</sup>J. Sahoo, A. GB, Shijin, S. Maity, Sreedilal, A. Rashid, Hafte, Vrijeshpandey, Vikas, Suresh, and M. Kumar, Electronic Circuits and Diagrams-Electronic Projects and Design (2009).

<https://www.circuitstoday.com/jfet-and-mosfet-comparison>

### - Photographs

<sup>2</sup>J. Fiorillo, (JFET Biasing Class Lecture).

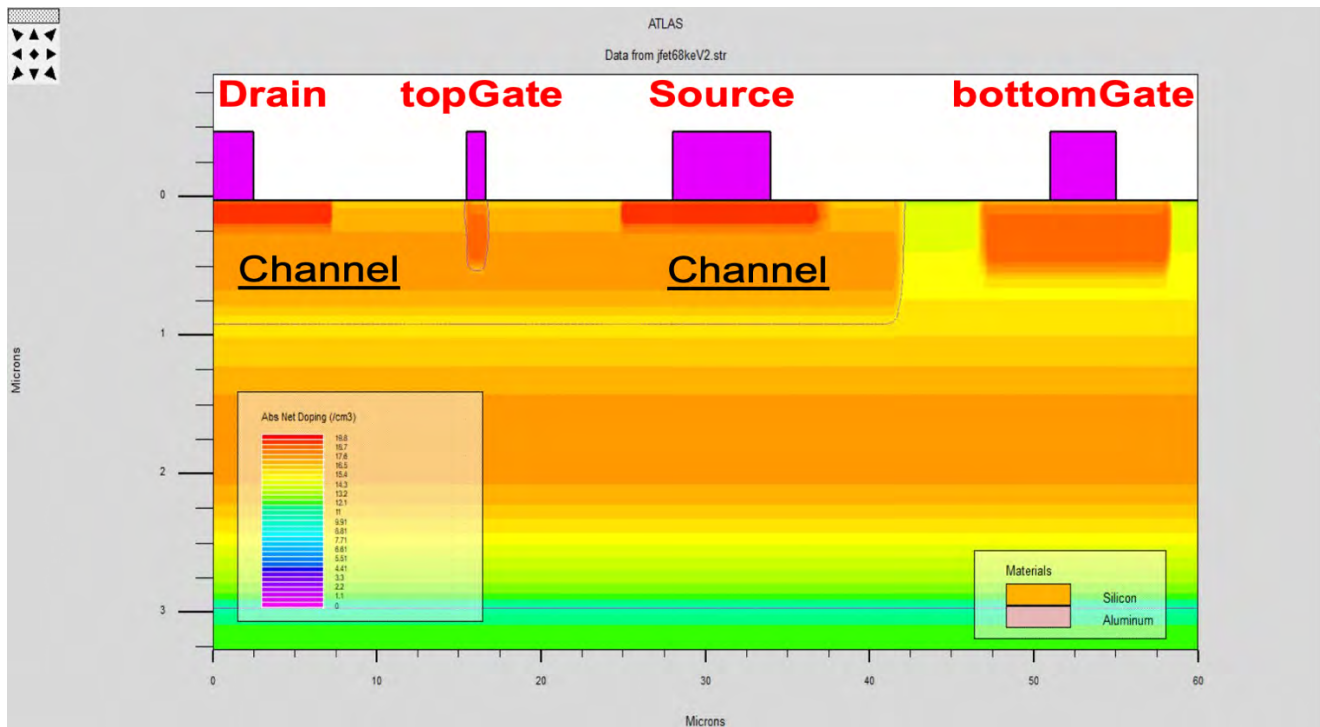
<sup>3</sup>Basic Electronics Tutorials (2018).

[https://www.electronics-tutorials.ws/transistor/tran\\_5.html](https://www.electronics-tutorials.ws/transistor/tran_5.html)

## VI. ACKNOWLEDGEMENTS

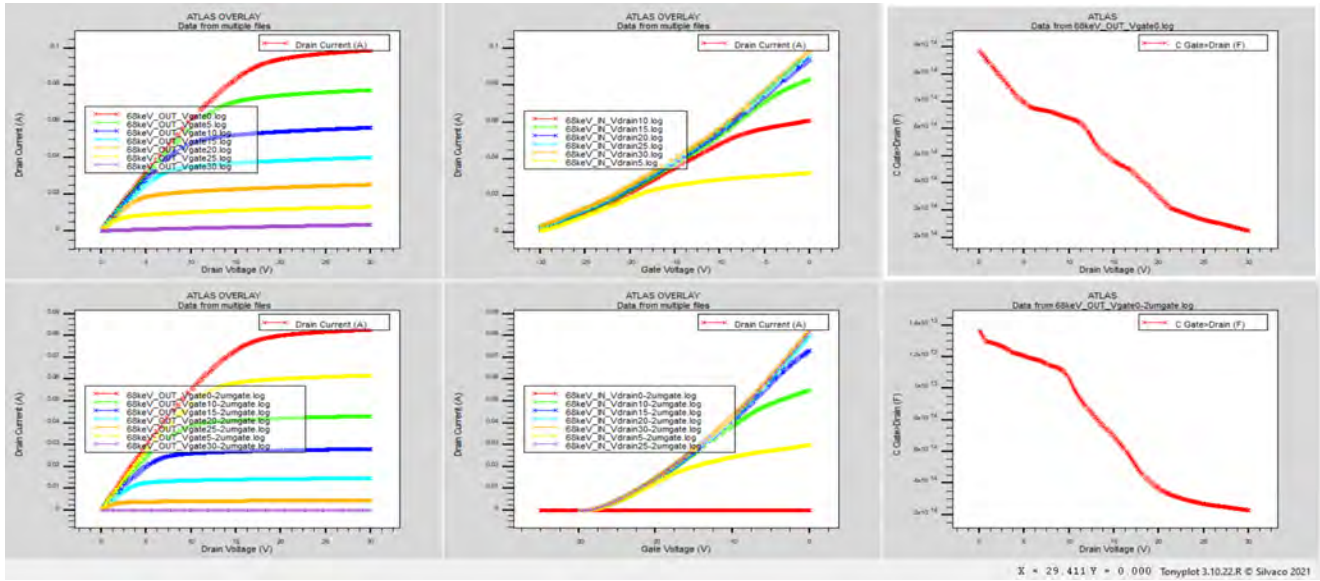
I wish to extend my special thanks to my mentor Dr. Gabriele Giacomini, the advancements of the project couldn't have been reinforced by anyone else. I also wish to thank Dr. Marco Bomben from Paris, France for his contributions. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## VII. APPENDIX A

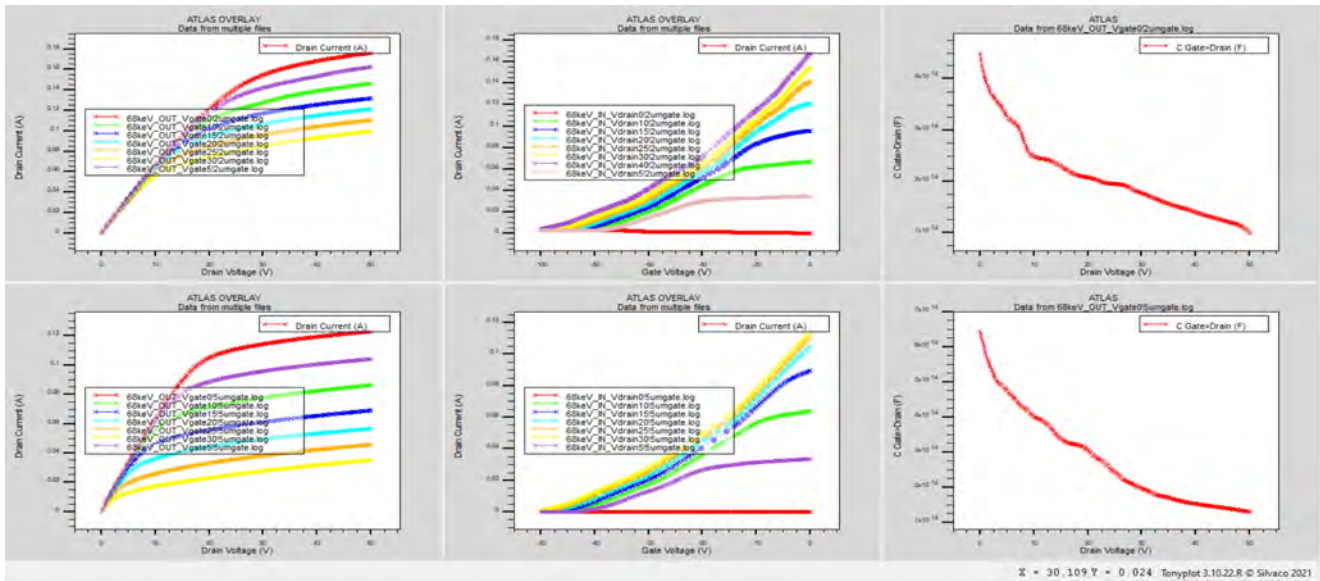


**FIG. 1.3:** The deep-implant method produces a channel dept just under 1-micron and a substrate doping less than 2.5-microns.

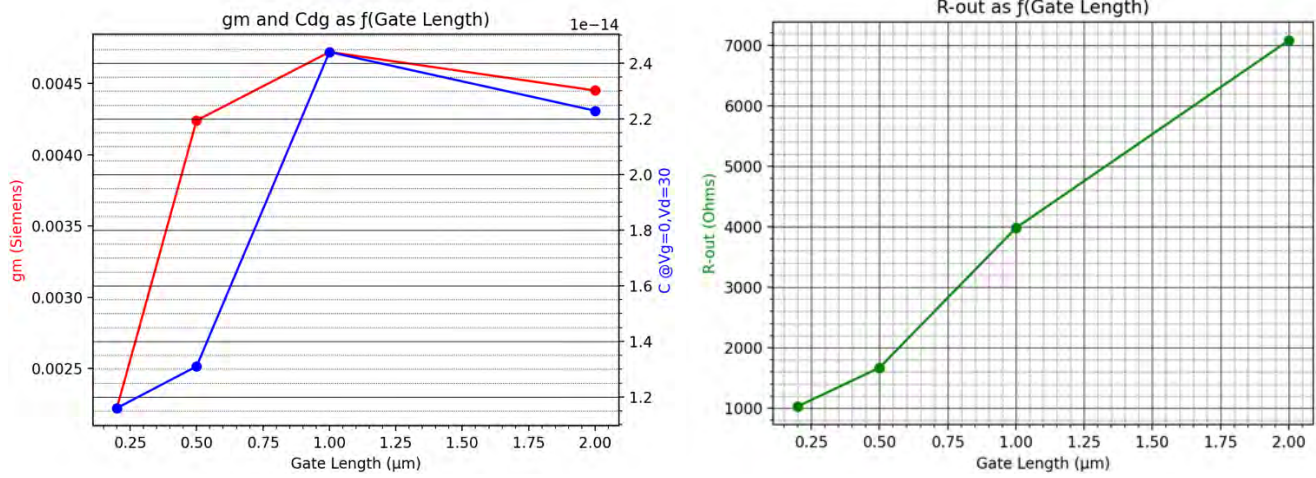




**FIG. 1.4-1.9:** A Tonyplot pictorial of the input/output characteristics and the Cgd for 1  $\mu\text{m}$  & 2  $\mu\text{m}$  gate-length JFET.



**FIG. 2.0-2.5:** A Tonyplot pictorial of the input/output characteristics and the Cgd for 0.5  $\mu\text{m}$  & 0.2  $\mu\text{m}$  gate-length tetrode JFET



**FIG. 2.6-2.7:** Python diagrams compiled for the gm, Cgd & the Rout of each gate-length iteration of the tetrode JFETs.

## VIII. APPENDIX B

***#In geometry (\*in file) replaced the lines:***

```
////////////////////////////////////  
/////
```

```
implant boron dose=2e13 energy=1000 pearson crystal print.mom tilt=7  
implant boron dose=4e13 energy=1000 pearson amorph print.mom tilt=7
```

***#and the lines:***

```
diffus time=30 temp=700 t.final=1000 nitro press=1.00  
diffus time=30 temp=1000 dryo2 press=1.00 hcl.pc=0  
diffus time=50 temp=1000 nitro press=1.00  
diffus time=60 temp=1000 t.final=700 nitro press=1.00
```

***#and the line:***

```
the line: implant phosphor dose=3.1e13 energy=330 pearson tilt=7 amorph
```

```
////////////////////////////////////  
/////
```

***#with the lines:***

```
implant boron dose=1e13 energy=1000 pearson crystal print.mom tilt=7  
implant boron dose=3e13 energy=1000 pearson amorph print.mom tilt=7
```

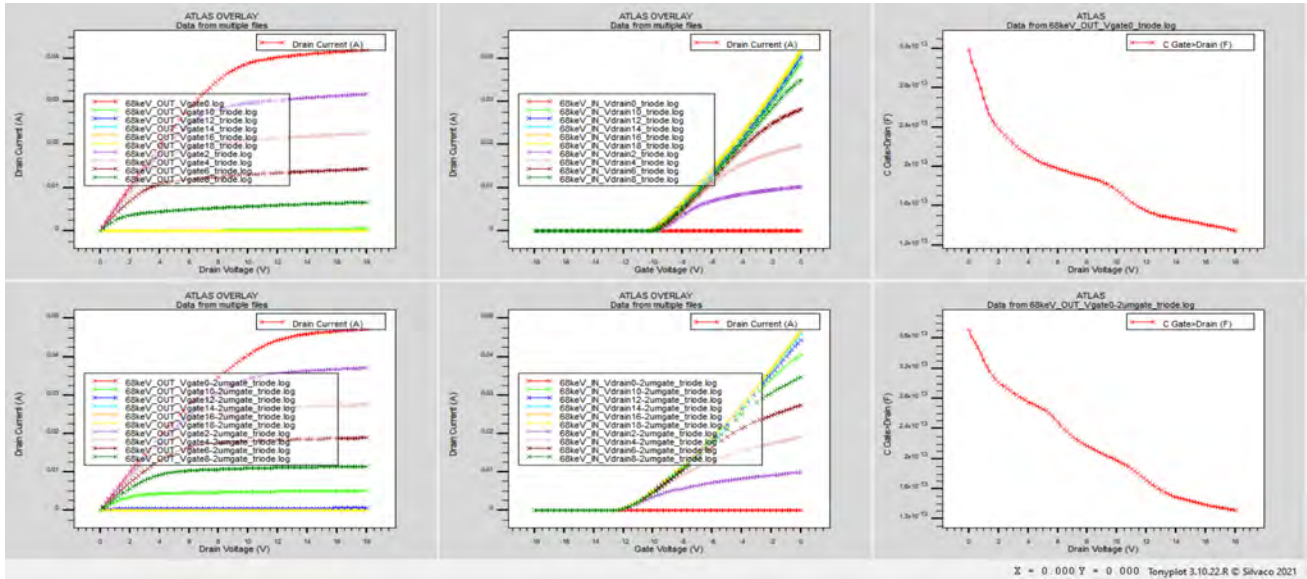
***#With the lines:***

```
diffus time=30 temp=700 t.final=1000 nitro press=1.00  
diffus time=30 temp=1000 dryo2 press=1.00 hcl.pc=0  
diffus time=30 temp=1000 nitro press=1.00  
diffus time=60 temp=1000 t.final=700 nitro press=1.00
```

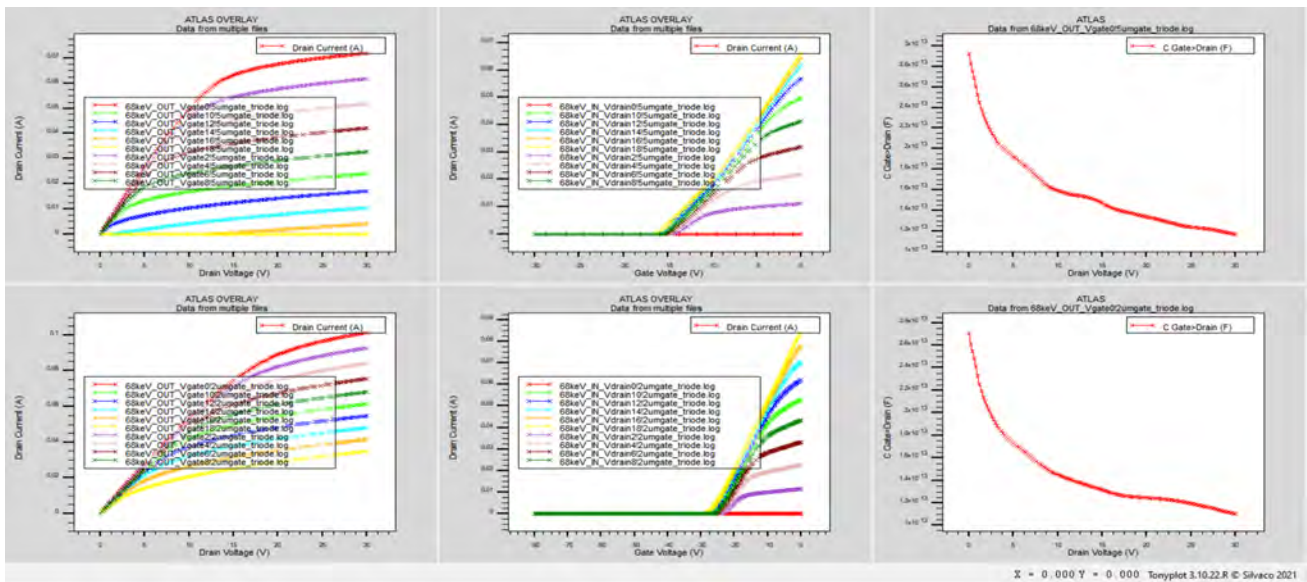
***#with the line***

```
implant phosphor dose=2.1e13 energy=330 pearson tilt=7 amorph
```

**FIG. 2.8:** The bottom-gate doping concentration instructed as implant boron-dose is reduced by a magnitude of  $1e13 \text{ cm}^2$  in both cases. The equivalent diffusion time to 1000-degrees centigrade reduced to 30-minutes from 50-minutes.

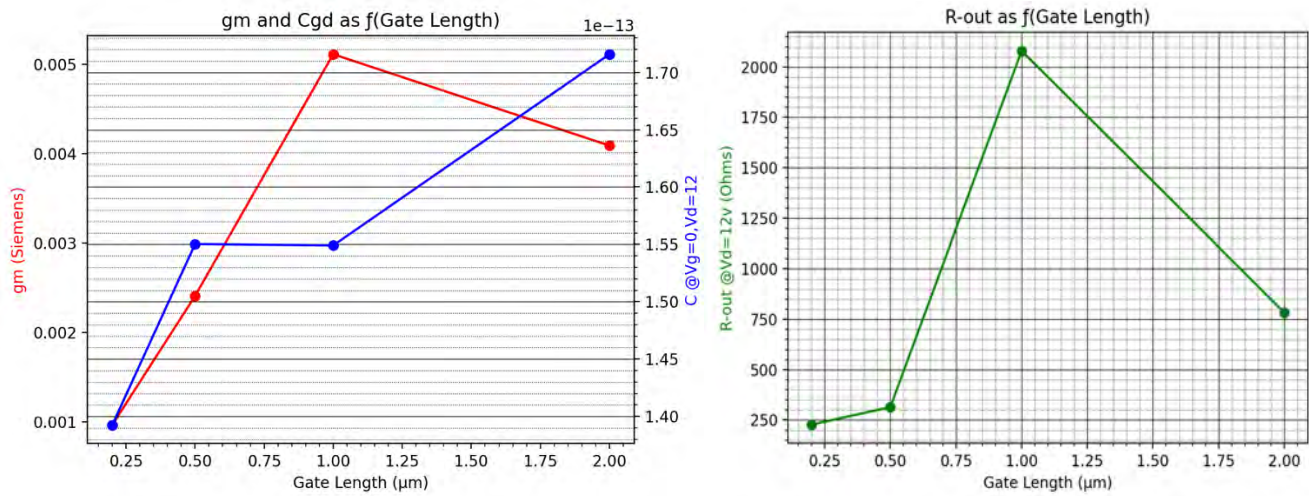


**FIG. 2.9-3.4:** A Tonyplot pictorial of the input/output characteristics and the Cgd for 0.5 μm & 0.2 μm gate-length tetrode JFET – computed from a triode iteration & the reduction of initial boron dose and diffusion time.



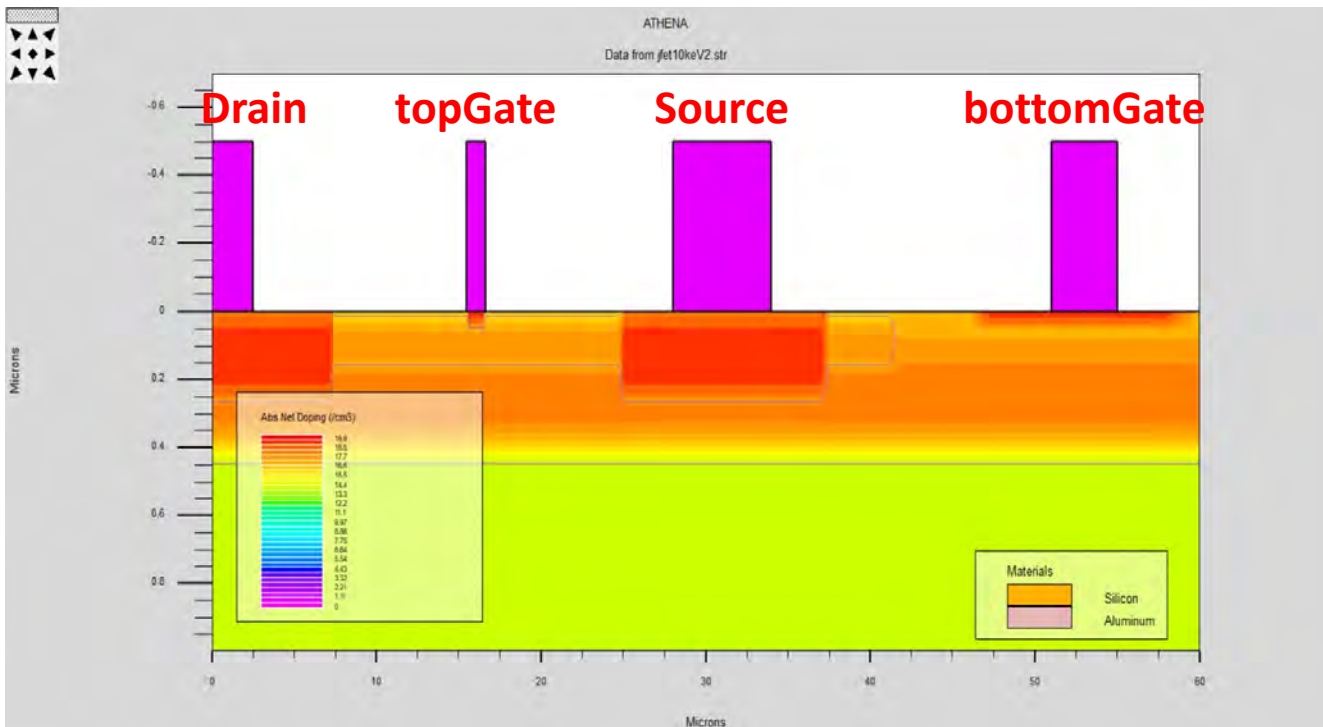
**FIG. 3.5-4.0:** A Tonyplot pictorial of the input/output characteristics and the Cgd for 0.5 μm & 0.2 μm gate-length tetrode JFET – computed from a triode iteration & the reduction of initial boron dose and diffusion time.



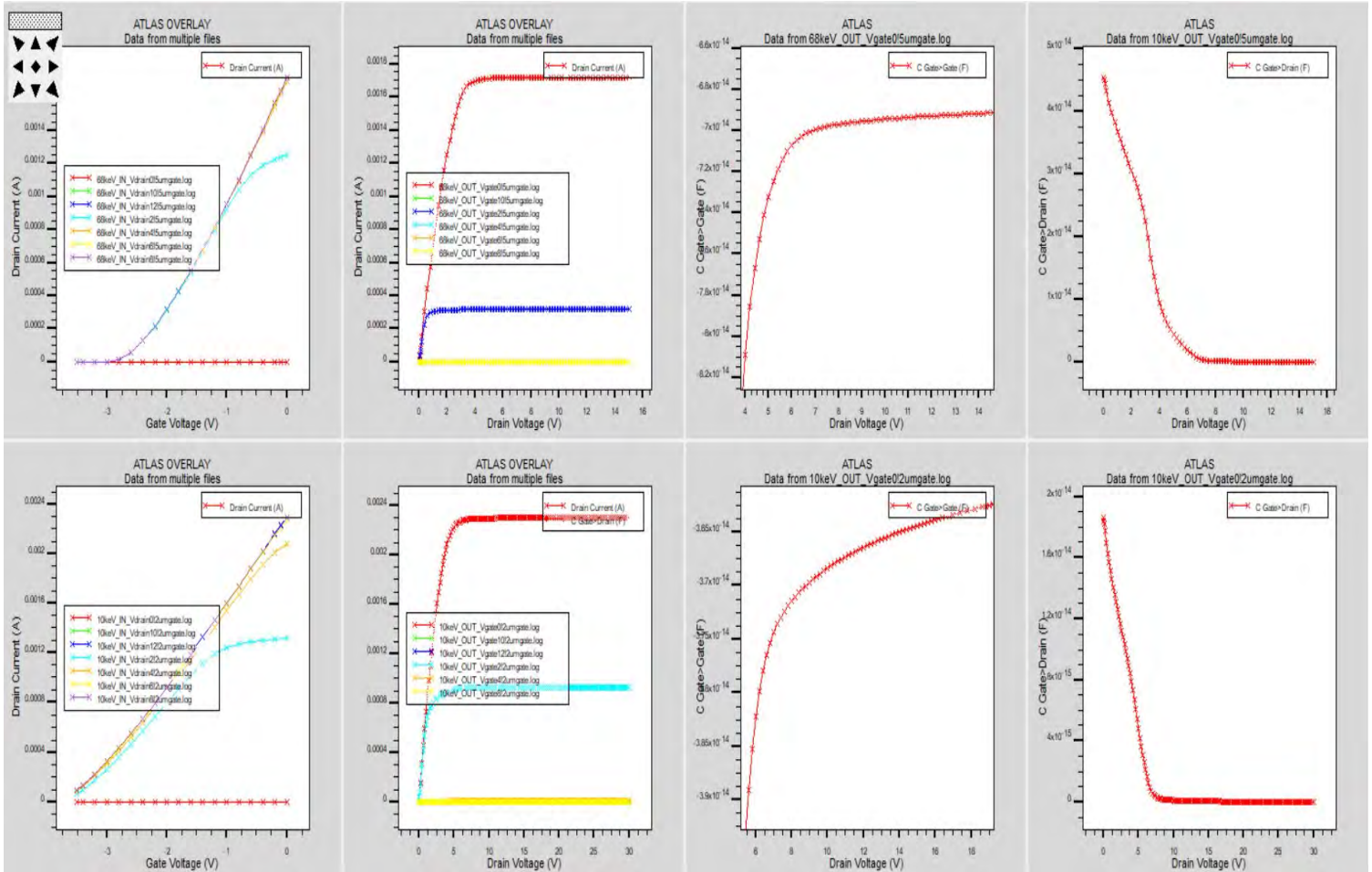


**FIG. 4.1-4.2:** Python diagrams compiled for the gm, Cgd & R-out of each gate-length iteration of the tetrode JFETs – computed from the triode configuration the reduction of initial boron dose and diffusion time.

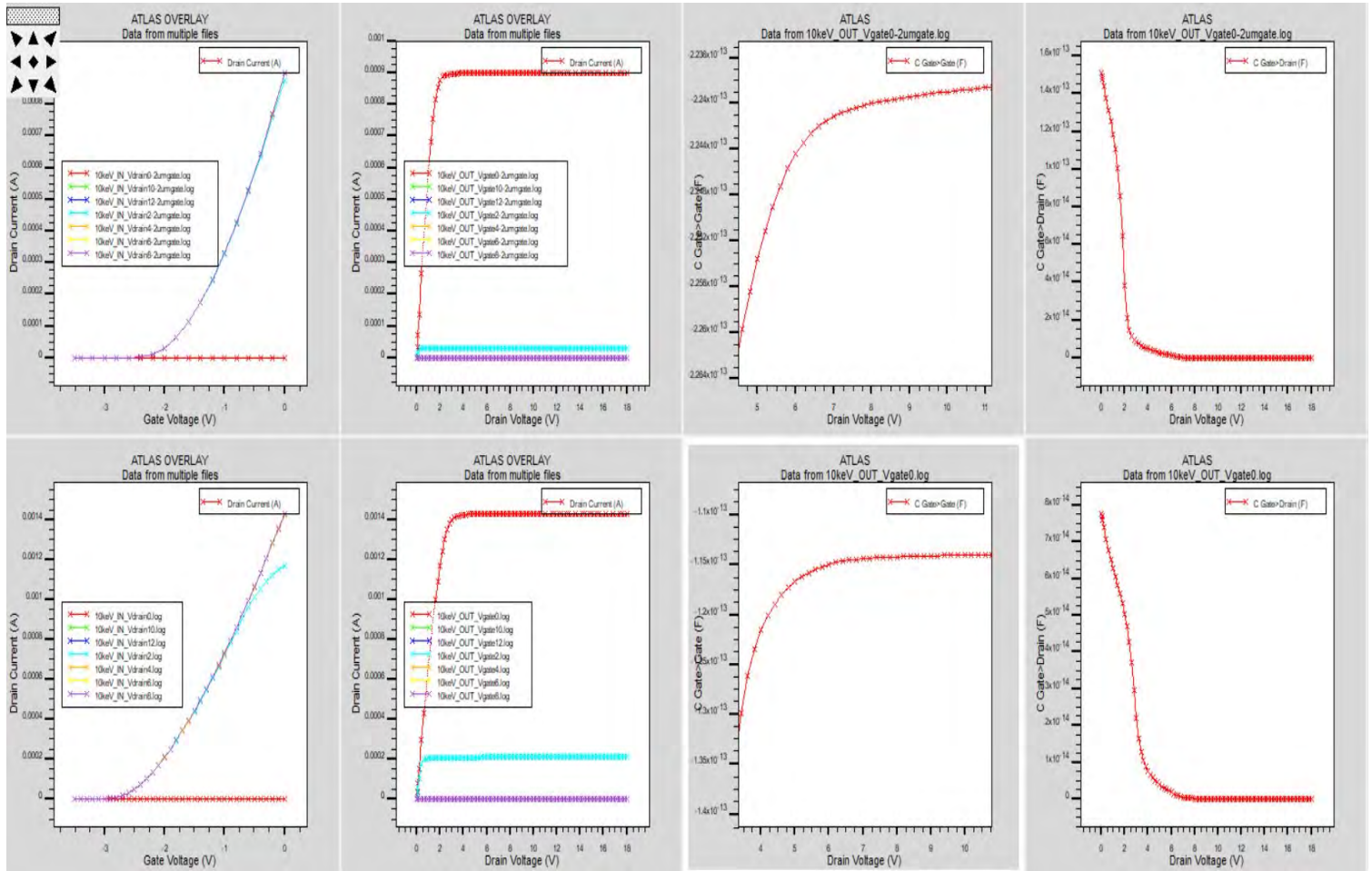
## IX. APPENDIX C



**FIG. 4.3:** The shallow-doping method produces a channel depth less than 200-nanometers and a substrate doping less than 600-nanometers.

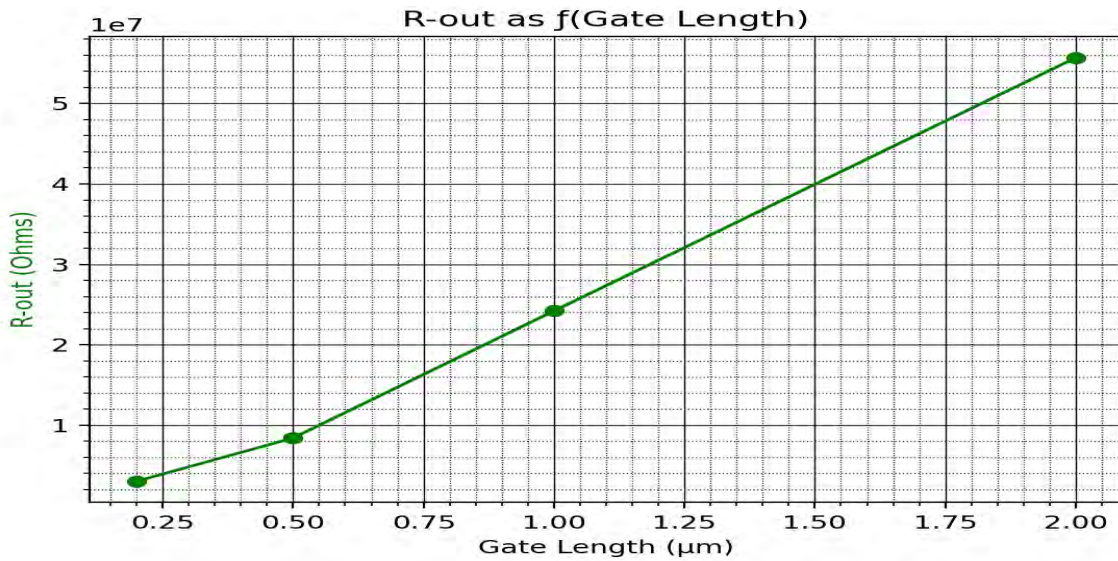
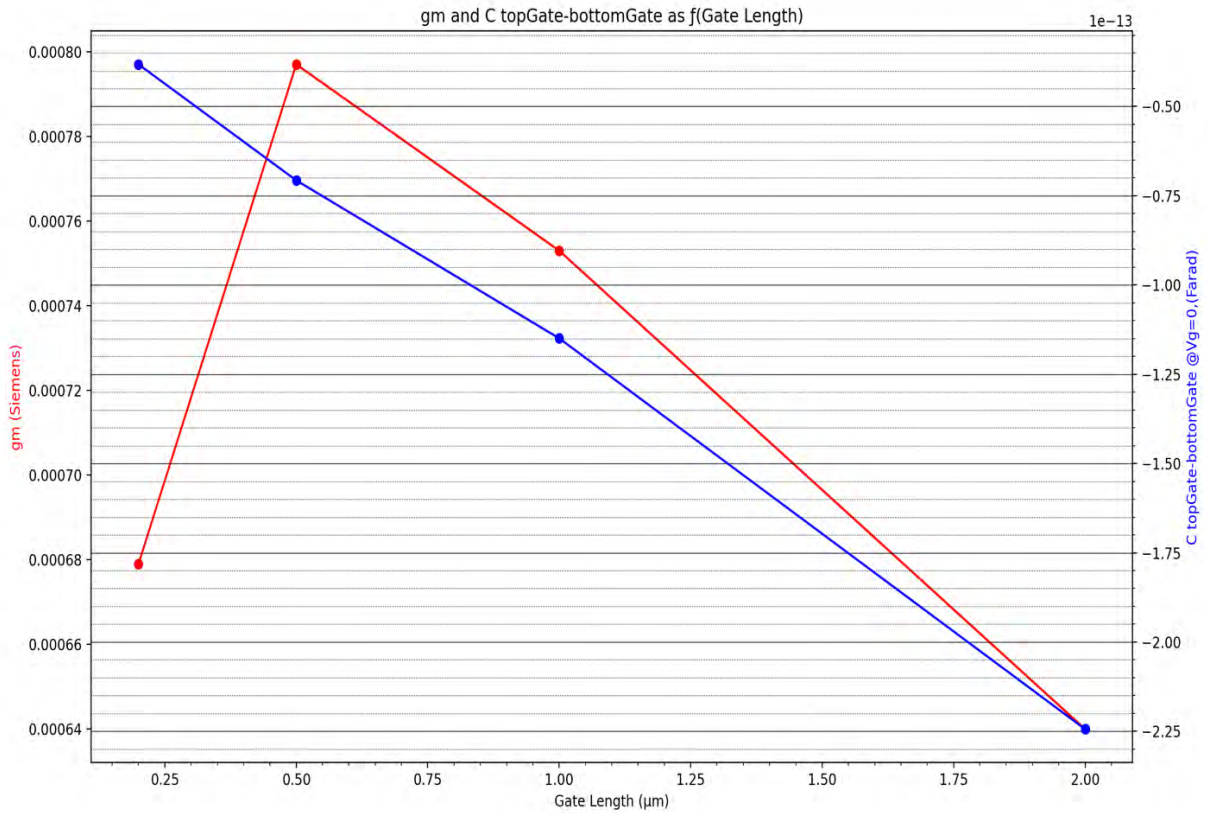


**FIG: 4.4-5.1:** The input & output characteristics of the 0.5-micron shallow implant model on the top-left. The input & output characteristics of the 0.2-micron shallow implant model on the bottom-left.  $V_{gs-off}$  is seen at -3V for the 0.5-micron model and nearing -3.5V for the 0.2-micron model. On the output model, the pinch-off voltage is near 6V for both devices. On the top-left is the gate-to-gate capacitance and the gate-to-drain capacitance of the 0.5-micron model. On the bottom-left is the gate-to-gate capacitance and the gate-to-drain capacitance of the 0.2-micron model.



**FIG: 5.2-5.9:** The input & output characteristics of the 2.0-micron shallow-implant model on the top-left. The input & output characteristics of the 1.0-micron shallow implant model on the bottom-left.  $V_{gs-off}$  is seen at -3V for the 0.5-micron model and nearing -3.5V for the 0.2-micron model. On the output model, the pinch-off voltage is near 6V for both devices. On the top-left is the gate-to-gate capacitance and the gate-to-drain capacitance of the 0.5-micron model. On the bottom-left is the gate-to-gate capacitance and the gate-to-drain capacitance of the 0.2-micron model.





**FIG. 6.0-6.1:** The shallow-implant tetrode JFET Python diagrams compiled for the gm, C top-gate-bottom-gate & R-out of each gate-length iteration – computed from the tetrode & reconfiguration of the initial doping concentrations.



# Machine learning for classifying neutron resonances

Mary Fucci<sup>1</sup> and Gustavo Nobre<sup>2</sup>

<sup>1</sup>*Department of Physics, University at Albany, Albany, NY 12222*

<sup>2</sup>*National Nuclear Data Center, Brookhaven National Laboratory, Upton, NY 11973*

**Abstract.** Understanding the nuclear formation and decay processes of elements is crucial to unravelling the mysteries of our universe. These processes depend on intrinsic properties of nuclei such as nuclear level densities, decay strength functions, and other nuclear data. It is thus critical that a reliable nuclear database is produced. The focus of this project is to use current measurements of resonance states observed in compound nuclei (formed by neutron-induced reactions) and develop a machine learning algorithm to automate and assess this data and correctly classify neutron resonances according to their spin groups. Synthetic data was used to train machine learning algorithms to classify resonances, considering their widths and spacings. Performance comparisons were ran on Scikit-learn classifiers (such as Random Forest, Nearest Neighbors, Neural Network, etc.) to assess accuracies when varying hyper-parameters. [1] Continued optimization allows for application of transfer learning to predict spin assignments in real nuclei, as compiled in evaluated files or in the Atlas of Neutron Resonances. Having an accurate nuclear database has many applications within astrophysics and nuclear energy which promotes future discoveries in physics.

## INTRODUCTION

### Neutron Resonance

Neutron-induced reactions are nuclear reactions where a neutron is fired at a nucleus; these reactions can lead to the formation of a compound nucleus. The combination of the neutron and nucleus excites the nucleus into a higher energy state where it either decays by emitting a neutron or a gamma. Resonance peaks are observed when the incident neutron energy is near the excitation energy of a given level in the compound nucleus. Studying and understanding neutron resonance is important as scientists can learn more about the properties and the stability of nuclei which plays a role in nuclear energy, for example.

### Motivation and Purpose

The *Atlas of Neutron Resonances* is the most comprehensive nuclear database for evaluated neutron resonance. It compiles resonance properties and parameters, such as resonance widths and spacings, for individual elements and isotopes. [2] Currently, even though the *Atlas* is the main reference in resonance data, it is full of issues related to uncertainty quantification and evaluation reproducibility; there are many gaps in the data and lack of documentation

of resonances. The National Nuclear Data Center (NNDC) at Brookhaven National Laboratory is seeking to address these issues.

This paper reports a particular contribution to this effort where the main focus was on optimizing and automating a Python-based machine learning algorithm to identify missing neutron resonances and effectively reclassify those that have misassigned spins. Optimization and automation was centered around testing and fine-tuning the machine's classifiers, including Scikit-learn classifiers: Random Forest, Nearest Neighbors, Neural Network, etc.

## METHODS

Our resonance classification code operates by inputting a jumbled sequence of resonances and organizes the data according to spingroups. Jumbled sequences are those with resonances that are incorrectly labeled on purpose. Training data used for this project was comprised of a synthetic sequence of resonances based on actual data from chromium-52. The sequences are jumbled to different degrees to replicate the non-uniformity in the data. This also allows us to view how much the classifiers can handle before they begin to break down.

Twelve sequences, each with increasing degrees of jumbledness, are cycled through the machine 10 times so that an average of accuracy with respect to jumbledness can be recorded. The classifier's hyper-parameters are adjusted and an average is taken for each adjustment. Hyper-parameters control the classifier's learning values. Accuracies for each hyper-parameter variation are then plotted against jumbledness factors where comparisons can be made and we can see where the classifier begins to stabilize. This procedure is tested with our Gamma-gamma ( $\Gamma_\gamma$ ) feature toggled on and off. This feature controls how the machine interprets the capture width measurement in the data. The capture width is the width of the resonance in a capture cross section. Many instances of this quantity are unreliable, since in some cases it was measured, but the majority are averaged, biasing the classification. Accuracies are inflated when this feature is on since the algorithm can "cheat" and group resonances with similar capture widths to the same class.

In this paper, the classifiers were ran using the angular momentum number ("L") label mode to make class predictions. In the future, tests will be done using other labels modes, such as spin group which compares the sets of spin numbers and  $\frac{0}{1}$  which is a binary variable that acts as a true/false indicator.

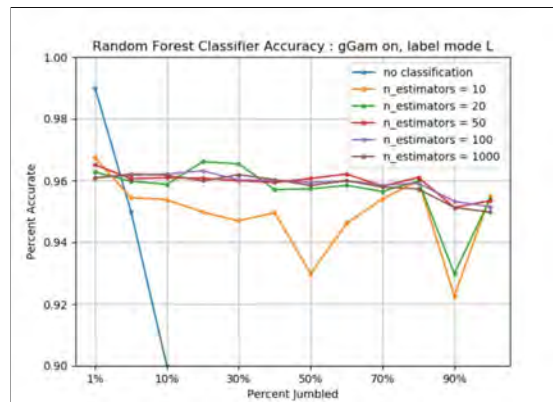
## Automation

Automation is the operation of minimizing human interference in a process. Automating the machine learning code centered around creating nested loops that would quickly sift through all of the training data while switching

through different features and making predictions called labels. In the future, we would like to be able to automatically filter through hyper-parameter variations. The code was also able to save each accuracy output and store it into an array which would be used to plot the accuracy values. We utilized a wrapping script so that arguments could be quickly edited without hard-wiring temporary features into our main code. In addition to these steps, we also found a way to significantly decrease run time by utilizing the NNDC's multi-core computer cluster. We wrote a code that enabled running batch calculations in parallel. These batches were automatically submitted to a multitude of computational nodes which have higher-power processors and lower traffic compared to the mother node. As a result of this automation, the run time decreased from 4 hours to 2 minutes.

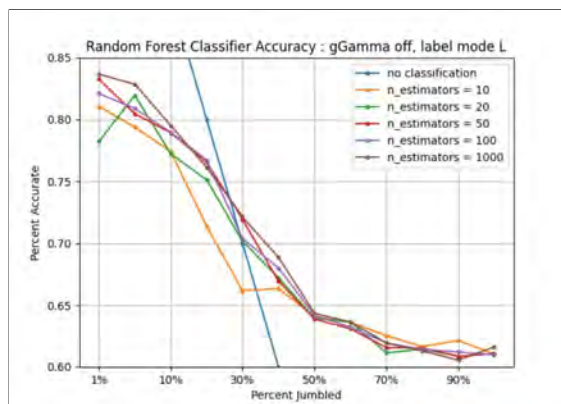
## RESULTS

I ran preliminary performance tests on each of the classifiers. Random Forest was chosen first for optimization based on its run time relative to the other two and its minimal number of hyper-parameters. I chose "n estimators" to adjust first and view how changing it affected accuracy scores. "n estimators" controls the number of decision trees the classifier uses to come to a class prediction. I tested values of 10, 20, 100, and 1,000 and cycled the training data 10 times for each value to calculate an average. Figure 1, below, displays a plot of accuracy scores against jumbledness factors of the sequences with the capture width feature turned on.



**FIGURE 1.** Plot of accuracies against jumbledness factors for Random Forest Classifier testing variations of hyper-parameter n estimators with label mode L and capture width feature on.

It can be seen in Fig. 1 that the classifier stabilized when there were 50 or more trees present and was unsteady when there were 10 and 20 trees used. Knowing this lets us conclude that as long as we use a number of trees greater than 50, then this hyper-parameter will allow the classifier to run well. Testing with a larger number of trees will likely demonstrate a similar convergence. Figure 2 exhibits the accuracies versus percent jumbled for the same set of sequences after I computed them with the capture width feature off.



**FIGURE 2.** Plot of accuracies against jumbledness factors for Random Forest Classifier testing variations of hyper-parameter n estimators with label mode L and capture width feature off.

It is shown that using 10 and 20 trees is still unsteady here and again, a stabilization can be seen using 50 trees or more. It is also notable that the classifier was depending on the capture width quantity for reclassification, since an inverse relationship between accuracy and jumbledness percentage is prevalent. It is worthwhile to remark that the false high accuracy rate given with the capture width on is not a characteristic expected for all data we use. This depends on the available information on the isotope's resonances being used. The discrepancy between these two plots was anticipated based on the data available for chromium-52.

The blue line plotted in Fig. 1 and Fig. 2 serves as a "reality check" to ensure that reclassification attempts are beneficial. It is a "do nothing" line that we use to gauge if our efforts for reclassifying are better than doing nothing to it. Ideally, once the classifiers are fine-tuned, all accuracy points would be above this line.

## CONCLUSION

There are more classifiers and hyper-parameters to be tested and optimized before profitably running real data from other elements with misclassified or missing resonances through the machine. Up to this point, the machine learning algorithm has been better automated which will speed up the process of optimizing the classifiers. The first steps were taken in optimizing the Random Forest classifier which resulted in a better understanding of the impact

that the number of decision trees has on the classifier's ability to make predictions. Continued optimization allows for application of transfer learning to predict spin assignments in real nuclei. Zirconium-90, Iron-56, Lead-208, and Indium-115 have preemptively been selected to test as real data. Neutron resonance physics plays a part in other areas of science, such as astrophysics and nuclear energy. Therefore, having an dependable nuclear resonance database promotes future research in physics; scientists can rely on this database and be able to make decisions and discoveries that have sound footing.

## **ACKNOWLEDGMENTS**

I would like to thank my internship mentor Gustavo Nobre, group leader David Brown, and Amber Coles who are great guides and resources. I would also like to thank my fellow interns Rose Crawford, Sergio Ruiz, and Sergey Scoville. This project was successful through the combined effort of this team and dedication to meet our goals.

This project was supported in part by Brookhaven National Laboratory, the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## APPENDIX

Participant Name	Affiliation	Contribution
Gustavo Nobre	National Nuclear Data Center at Brookhaven National Labs	Mentor
David Brown	National Nuclear Data Center at Brookhaven National Labs	Group Leader
Rose Crawford	Willamette University	SULI Intern, Investigated the impact of missing resonances on the average resonance spacing for a resonance sequence to aid the development of an automated algorithm which can extract reliable resonance spacings and their uncertainties.
Sergio Ruiz	Georgia Institute of Technology	SULI Intern, Validated BRR on real resonance data for the first time by gathering data from different resonance experiments, jumbling it and testing BRR on it.
Sergey Scoville	Rensselaer Polytechnic Institute	SURP Intern, Developed a utility that gathers data on any nucleus from the Atlas of Neutron Resonances to build and populate a resonance table class applicable on a wide range of machine learning classifiers including methods to address untracked resonance parameters.

## REFERENCES

1. scikitlearn, <https://scikit-learn.org/stable/>
2. S. F. Mughabghab, *Atlas of Neutron Resonances, Vol.1 and 2*. Elsevier, Amsterdam (2018).

# Quantum simulation of spin entanglement in quantum chromodynamics strings using a Kondo-type spin model

Wenjie Gong<sup>1</sup>, Zhoudunming Tu<sup>2</sup>, and Raju Venugopalan<sup>2</sup>

<sup>1</sup>*Department of Physics, Harvard University, Cambridge, Massachusetts 02138*

<sup>2</sup>*Department of Physics, Brookhaven National Laboratory, Upton, New York 11973*

## Abstract

Experiments at high-energy colliders have probed the internal structure of nucleons, yielding information regarding the momentum and spin of their quark and gluon constituents. However, much remains unknown regarding the dynamics of the partons within nuclear matter, including their interactions prior to hadronization. Quantum computers, which capitalize upon the evolution of quantum ensembles to perform computational tasks, provide an attractive platform for simulating the quantum many-body interactions of partons before fragmentation. Here, we quantum simulate the dynamical evolution of spin correlations within strings of partons. We model these quantum chromodynamics (QCD) strings by letting strange quarks act as effective spin impurities mediated by lighter (up and down) quark matter. The evolution of entanglement between the spin impurities is examined using a modified version of the Kondo Lattice Model at half-filling, derived with applicability to universal digital quantum computing platforms. Entanglement between a pair of spins is assessed using two-particle correlations and the partial positive transpose (PPT) criterion of entanglement. We find that spin entanglement between strange particles in a QCD string oscillates at antiferromagnetic coupling. Initial quantum simulations on IBM Q's noisy intermediate scale

quantum (NISQ) devices show that the noise introduced by the quantum hardware prohibits the extraction of useful information; this indicates the necessity of targeted error-correction codes in the NISQ era.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>The KLM as a spin Hamiltonian</b>	<b>5</b>
<b>3</b>	<b>Entanglement evolution of a single embedded singlet</b>	<b>11</b>
<b>4</b>	<b>Discussion and conclusions</b>	<b>14</b>
<b>5</b>	<b>Acknowledgements</b>	<b>15</b>
<b>A</b>	<b>Details of the two-flavor spin chain model</b>	<b>15</b>
<b>B</b>	<b>Measuring the PPT criterion of entanglement</b>	<b>22</b>

## 1 Introduction

Deep inelastic scattering (DIS) has enabled the precise examination of the quark and gluon, or parton, composition of nucleons, providing the means to determine the kinematic distribution of partons within a targeted nucleon [1–3]. In particular, data from the DIS Hadron-Electron Ring Accelerator (HERA), fixed target experiments, the Tevatron, and the Large Hadron Collider (LHC) have contributed to a well-established understanding of the parton distribution function of the proton, a necessary input to any predictions made at proton colliders [4]. However, many aspects of partons, including their interactions prior to their combination into colorless particles, a process known as hadronization, remain mysterious.



A useful framework within which to explore parton behavior before particle formation consists of the Lund string model [5]. As the strong force does not fall off with spatial separation, the gluon field lines, or gluon strings, that stretch between color charges only lengthen as the charges move apart. Therefore, it becomes energetically favorable at some point for the gluon strings to split, forming a quark anti-quark pair at the site of the breakage. In the Lund string picture of hadronization, strings of gluons embedded with quark anti-quark pairs stretch between quark fragments of particles; the quarks within these "quantum chromodynamics (QCD) strings" subsequently combine with local neighbors to produce colorless hadrons. These strings are fundamental to our understanding of QCD; their dynamics are used to populate the Monte Carlo (MC) generators that simulate events at high energy colliders. Thus, it is useful to further clarify the behavior of the partons within QCD strings prior to string fragmentation and hadronization— a greater understanding of QCD strings can significantly improve phenomenology.

One outstanding question regarding QCD strings concerns the possible entanglement among quarks within the string. The discovery of azimuthal angular correlations between particles widely separated in rapidity, or "ridge" correlations, in proton-proton collisions by the Compact Muon Solenoid (CMS) experiment at the LHC [6] brings this issue of entanglement to further urgency; similar ridge correlations were previously only observed in heavy-ion collisions. As important features of proton-proton collisions can be explained by the existence of quantum correlations between the partons that form the colliding particles [7, 8], quantum entanglement between partons may very well be responsible for the ridge correlations in both proton-proton and heavy-ion collisions. This entanglement would likely be generated by the QCD strings from which the partons originated. Indeed, many experiments at electron-positron, [9], DIS [10, 11], and hadron [12] colliders address the effects of parton correlations and entanglement, underscoring the mounting interest in possible entanglement within QCD strings.

Spin entanglement between quark anti-quark pairs in a QCD string— and especially

between strange anti-strange quarks— can lend insight into the distribution and evolution of entanglement within the string.  $s\bar{s}$  pairs are of specific interest as their spin polarizations and entanglement can be measured via the self-analyzing weak decay of  $\Lambda\bar{\Lambda}$  hyperons [13, 14]. This makes the reasonable assumption that the spin of the  $\Lambda$ s are carried by the strange quark [15, 16]. Previous work using a static two-flavor model of a QCD string <sup>1</sup> has shown that two-particle spin correlation between the quarks of the flavor of interest, here  $s\bar{s}$ , decreases as approximately  $\sim 1/N^2$  with increasing multiplicity  $N$  of the other quark flavor in the QCD string ensemble; quantum entanglement thus becomes washed with increasing length of the string out due to fermionic indistinguishability [14]. While this study looks at the effect of parton multiplicity upon spin entanglement in the static case, entanglement between  $s\bar{s}$  may also exhibit non-trivial dynamical behavior owing to interactions among partons in the string prior to hadronization.

The dynamical behavior of entangled quarks in a QCD string, a complex quantum many-body problem, is ideally suited for simulation on quantum architecture; as per the original idea of Feynman [17], a quantum computer can replicate the behavior of quantum system whose features are difficult to access via classical computation. Indeed, classical simulation of the real-time dynamics of quantum many-body systems must contend with a dynamical sign problem that exponentially increases the difficulty of simulation at long timescales [18]. In recent years, analog quantum simulators, or quantum ensembles with natural evolution corresponding to that of a specific target Hamiltonian, have been successfully built and implemented to simulate the Hubbard model [19, 20] and many-body localization [21–23]. While analog simulators are effective, they are limited in versatility and scalability. Various problems in physics and chemistry have also been investigated using a universal digital quantum computer, such as finding the ground state energy of a molecule with a variational quantum eigensolver (VQE) [24] and simulating the Ising [25] and Schwinger [26] models.

---

<sup>1</sup>This work was also completed by me this summer as a part of the SULI 2021 internship. However, as the model ties into a previous project, the details are omitted from the main narrative of this report. Specifics regarding the two-flavor model are included in Appendix A.

Even though the noisy intermediate scale quantum (NISQ) devices currently at our disposal are not fully fault tolerant, they hold promise for extracting, at the very least, qualitative information regarding a wide array of quantum many-body systems [27, 28]. It is therefore of interest to further probe the dynamics of entanglement within QCD strings using a quantum computer, thereby developing applications of quantum information to high energy and nuclear physics in the NISQ era.

Here, we consider the dynamical problem of the evolution of spin entanglement between  $s\bar{s}$  in a QCD string prior to hadronization; state-of-the-art quantum architecture is then critically assessed as a tool for computing such time evolution. Using a Kondo-type formalism, we model the  $s$  quarks and valence quarks of the QCD string as localized spin impurities within a sea of "conduction" spins representative of the lighter up ( $u$ ) and down ( $d$ ) quarks in the string. As  $s$  quarks are several orders of magnitude more massive than  $u$  or  $d$ , treating  $s$  quarks as localized impurities that interact with the dynamical  $u$  and  $d$  constitutes a good model of a QCD string. Due to the relativistic speeds of the valence quarks, they are also effectively quite massive and thus static and localized. Starting with the Kondo Lattice Model (KLM) at half-filling as reference, Section 2 will derive a spin Hamiltonian that captures this scenario. In Section 3, we simulate the real-time evolution of the entanglement between a spin singlet  $s\bar{s}$  embedded in a QCD string under the derived Kondo-type Hamiltonian, comparing results obtained from both numerical methods and from *IBMQ*'s quantum hardware. An analysis of our results as well as perspectives on the future of NISQ-era quantum computing in nuclear physics are presented in Section 4.

## 2 The KLM as a spin Hamiltonian

The KLM describes the interaction of electrons in a conduction band with localized spins. Model properties result from an interband exchange between the two subsystems.

The Hamiltonian is given by:

$$H = H_s + H_{sf} + H.c. \quad (2.1)$$

$$H_s = \sum_{\langle i,j \rangle \sigma} T_{ij} c_{i\sigma}^\dagger c_{j\sigma} \quad (2.2)$$

$$H_{sf} = -J \sum_i \mathbf{s}_i \cdot \mathbf{S}_i \quad (2.3)$$

Here,  $\mathbf{s}_i = \frac{1}{2}(c_{i\uparrow}^\dagger c_{i\downarrow}^\dagger) \sigma \begin{pmatrix} c_{i\uparrow} \\ c_{i\downarrow} \end{pmatrix}$  where  $\sigma$  are the Pauli matrices, and  $\mathbf{S}_i$  are the spin matrices of the localized spins. Now,  $H_s$  is the part of the Hamiltonian that describes uncorrelated electrons in a non-degenerate energy band. More specifically, it denotes hopping of conduction electrons between nearest-neighbor sites [29].  $c_{i,\sigma}^\dagger (c_{i,\sigma})$  creates (annihilates) an electron with spin  $\sigma$  at lattice site  $i$ .  $T_{ij}$  are the hopping integrals, indicating the coupling between nearest neighbors  $i, j$  on the lattice; if constant, they can be pulled out of the sum.  $H_{sf}$  represents the interaction between itinerant and localized spins  $\mathbf{s}_i$  and  $\mathbf{S}_i$ . For  $J > 0$  ( $J < 0$ ), a parallel (antiparallel) alignment is favored between itinerant and localized electrons, signifying ferromagnetic (antiferromagnetic) coupling. The coupling of most physical systems is antiferromagnetic. The exchange interaction can also be expanded into second-quantized form

$$H_{sf} = -\frac{1}{2}J \sum_i [S_i^z (n_{i\uparrow} - n_{i\downarrow}) + S_i^+ (c_{i\downarrow}^\dagger c_{i\uparrow}) + S_i^- (c_{i\uparrow}^\dagger c_{i\downarrow})] \quad (2.4)$$

where  $n_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$ . The first term is an Ising-like interaction between the  $z$  components of the localized and itinerant spins, while the second term and third terms encapsulate spin-exchange processes [30].

We must now address the issue of converting Eq. 2.1 into spin operator formalism; only a Hamiltonian of spin operators can be directly implemented on a digital quantum device. When facing this problem, our first impulse may be to utilize the Jordan-Wigner

transformation [31], which maps spinless fermions to spin operators:

$$\begin{aligned} c_j^\dagger c_j &= \frac{\sigma_j^z + 1}{2} \\ c_j &= \prod_{k=1}^{j-1} (-\sigma_k^z) \sigma_j^- \end{aligned} \quad (2.5)$$

However, here we run into two complications. The first consists of the fact that our electrons are spin-full, so we cannot use Jordan-Wigner to directly relate a creation operator such as  $c_{i\uparrow}$ , which has two quantum labels, to a Pauli operator. The second obstacle is that we cannot use the lattice label  $i$  as a unique qubit identifier  $j$ . Currently, both electron spins  $s_i$  and localized spins  $S_i$  are enumerated by lattice site, but when we map this system onto a quantum circuit, electrons and localized spins should be represented by different qubits. Thus, we must translate the lattice configuration appropriately onto an ensemble of qubits.

To address these two issues in tandem, for a 1D lattice of length  $L$ , we can relabel the operators in the Hamiltonian Eq. 2.1 as

$$\begin{aligned} S_i &\rightarrow S_j \\ c_{i\uparrow} &\rightarrow c_{j+L} \\ c_{i\downarrow} &\rightarrow c_{j+2L} \end{aligned} \quad (2.6)$$

where the subscript now refers to qubit index rather than lattice site. Note that as we consider 1D QCD strings, we can specifically target the 1D KLM. Then, using the Jordan-Wigner transformation Eq. 2.5, we get

$$\begin{aligned} H &= \sum_{\langle j,k \rangle} T_{jk} (\sigma_{j+L}^x \sigma_{k+L}^x + \sigma_{j+L}^y \sigma_{k+L}^y + \sigma_{j+2L}^x \sigma_{k+2L}^x + \sigma_{j+2L}^y \sigma_{k+2L}^y) \\ &\quad - J \sum_j \left[ -\frac{\sigma_j^z}{4} (\sigma_{j+L}^z - \sigma_{j+2L}^z) + \sigma_{j+L+1}^z \dots \sigma_{j+2L-1}^z (\sigma_j^- \sigma_{j+L}^+ \sigma_{j+2L}^- + \sigma_j^+ \sigma_{j+2L}^+ \sigma_{j+L}^-) \right] \end{aligned} \quad (2.7)$$

where we have used that  $\sigma^\pm = \frac{\sigma^x \pm i\sigma^y}{2}$  and  $S_j = \frac{1}{2}\sigma_j$ .

While, in principle, Eq. 2.7 can now be simulated on a quantum computer, it has converted the spin-flip interactions in  $H_{sf}$  into a long chain of  $\sigma^z$  interactions. Indeed, the additional chain of  $\sigma^z$ s in the Jordan-Wigner transformation (Eq. 2.5), which keeps fermionic creation operators from directly equating to spin creation operators, is known as a Jordan-Wigner string; it serves to keep track of fermionic parity. The string can turn two-particle interactions into multi-qubit interactions, greatly increasing the complexity of the quantum implementation. This issue with the Jordan-Wigner transformation is well-known— due to the higher computational cost introduced by Jordan-Wigner strings, it may no longer be efficient to simulate fermionic Hamiltonians on quantum rather than classical hardware [32]. Therefore, rather than working with the unwieldy converted Hamiltonian Eq. 2.7, we specifically consider the case of the KLM at half-filling. This is arguably one of the most interesting cases of the KLM, with half-filling giving rise to the Kondo gap and Kondo insulator. [33] As we will see, since the assumption of half-filling allows us to treat the conduction electrons as spins, it is also the most applicable to our model of the QCD string. In our model, we have related quarks to spins, with the up and down quarks comprising the "conduction band."

Half-filling indicates that the number of electrons in the conduction band is equal to the number of lattice sites  $L$ . In this case, if each lattice site is occupied, then there can only be one electron— either spin-up or spin-down— at each site. In this manner, we can treat the electrons as effective spins and modify the original KLM Hamiltonian Eq. 2.1 appropriately. Indeed, the localized spins in the Kondo model itself are a model of half-filled electron orbitals. Changing Eq. 2.1 to only emphasize the conduction electrons' spin degrees of freedom, we have

$$H = \sum_{\langle i,j \rangle} T_{ij} (s_i^x s_j^x + s_i^y s_j^y) - \frac{1}{2} J \sum_i [2S_i^z s_i^z + S_i^+(s_i^-) + S_i^-(s_i^+)]. \quad (2.8)$$

Here, we have analogized  $H_s = \sum_{\langle i,j \rangle \sigma} T_{ij} c_{i\sigma}^\dagger c_{j\sigma}$  to a spin-hopping term  $\sum_{\langle i,j \rangle} T_{ij} (s_i^x s_j^x + s_i^y s_j^y)$

such as that in the Heisenberg model. Moreover, we have noted that the factor of  $(n_{i\uparrow} - n_{i\downarrow})$  in the interaction term  $H_{sf}$  is essentially equal to  $\langle \sigma_i^z \rangle = 2 \langle s_i^z \rangle$ , so we have replaced it with  $2s_i^z$ . Using  $s_i^\pm = s^x \pm is^y$ , relabeling by qubit number instead of lattice site, and taking spin operators to Pauli operators, we obtain the spin Hamiltonian for the 1D KLM at half-filling

$$H = T \sum_{\langle j,k \rangle} (\sigma_{j+L}^x \sigma_{k+L}^x + \sigma_{j+L}^y \sigma_{k+L}^y) - J \sum_j [\sigma_j^z \sigma_{j+L}^z + \sigma_j^x \sigma_{j+L}^x + \sigma_j^y \sigma_{j+L}^y]. \quad (2.9)$$

We have assumed that  $T_{jk}$  is a constant  $T$  for simplicity. This is much simpler to work with than the Hamiltonian we recovered for the 1D KLM from the Jordan-Wigner transformation.

As a sanity check to confirm that Eq. 2.9 reproduces the expected features of the 1D KLM, we can assess one of the most prominent characteristics of the KLM at half-filling—the Kondo gap. In the ground state of the half-filled KLM, the Kondo gap forms due to hybridization between conduction and localized spins, a phenomenon known as the Kondo effect. This requires the commensurability condition, in which the spins of localized and conduction electrons for each polarization sum to one at each lattice site  $i$ : [34]

$$\langle n_{\uparrow,i}^l \rangle + \langle n_{\uparrow,i}^c \rangle = 1 \quad (2.10)$$

$$\langle n_{\downarrow,i}^l \rangle + \langle n_{\downarrow,i}^c \rangle = 1 \quad (2.11)$$

To test this with our Hamiltonian, we will evolve an ensemble into the ground state of Eq. 2.9 and compute commensurability at each lattice site. Note that the localized spin at site  $j$  is qubit  $j$ , while the conduction spin is qubit  $j + L$ . Moreover, for a qubit, we can define

$$\langle n_{\uparrow} \rangle = \frac{n_0}{n_0 + n_1} \quad (2.12)$$

where  $n_i$  is the number of counts of measurements of the qubit in state  $i$ .

To ensure that we attain the ground state of our Hamiltonian Eq. 2.9, we can define a

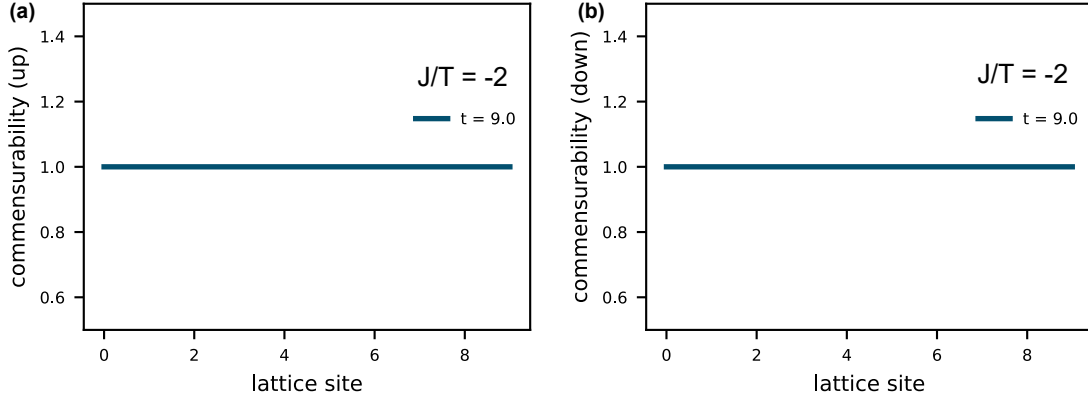


Figure 2.1: (a)  $\langle n_{\uparrow,i}^l \rangle + \langle n_{\uparrow,i}^c \rangle$  for a 1D Kondo lattice with 10 sites and coupling  $J/T = -2$ . (b)  $\langle n_{\downarrow,i}^l \rangle + \langle n_{\downarrow,i}^c \rangle$  for the same lattice.

time-dependent Hamiltonian  $H$ :

$$H(t) = \left(1 - \frac{t}{t_f}\right)H_0 + \frac{t}{t_f}H_1 \quad (2.13)$$

By initializing our system in the ground state of  $H_0$  and sweeping  $t$  from 0 to  $t_f$ , our system should evolve into the ground state of  $H_1$  by the adiabatic theorem. Here  $H_1$  is our Hamiltonian of interest Eq. 2.9, and  $H_0 = -\sum_i^{2L} \sigma_i^x$ . The ground state of  $H_0$  is simply the  $\sigma^x$  eigenstate across each the qubit. At the end of our sweep, we expect the commensurability conditions to be satisfied.

The adiabatic evolution and subsequent commensurability calculations were simulated using Python's QuTip package. The commensurabilities obtained for a 1D KLM with parameters  $J/T = -2$ , a lattice length of 10, and a sweep of length  $t_f = 9$  are shown in Fig. 2.1. As visible in Fig. 2.1, the commensurability along each spin direction is precisely equal to one, just as predicted for a 1D KLM at half filling. This validates our effective spin Hamiltonian for the 1D KLM at half-filling.



### 3 Entanglement evolution of a single embedded singlet

We now have the tools to simulate the real-time evolution of a spin singlet  $s\bar{s}$  embedded in a QCD string using our Hamiltonian for a half-filled 1D KLM, Eq. 2.9. We will start with a quick overview of how apply Hamiltonian evolution to a quantum circuit. The basic motivation is to break down a complicated Hamiltonian into a sum of simpler pieces. For a general Hamiltonian  $\hat{H} = \hat{A} + \hat{B}$ , we can use a process known as symmetric Trotter decomposition to rewrite time evolution by  $\hat{H}$  as: [28]

$$e^{-i\hat{H}t} = (e^{-i\hat{H}\Delta t})^M = (e^{-i\frac{1}{2}\hat{A}\Delta t}e^{-i\hat{B}\Delta t}e^{-i\frac{1}{2}\hat{A}\Delta t})^M + \mathcal{O}(\Delta t^3)^M \quad (3.1)$$

$M$  is known as the number of Trotter steps. For our specific Hamiltonian, we can consider the following operators:

$$\begin{aligned} \hat{A}_j &= e^{iJ(\sigma_j^x\sigma_{j+L}^x + \sigma_j^y\sigma_{j+L}^y + \sigma_j^z\sigma_{j+L}^z)\frac{\Delta t}{2}} \\ \hat{B}_k &= e^{-iT(\sigma_k^x\sigma_{k+1}^x + \sigma_k^y\sigma_{k+1}^y)\frac{\Delta t}{2}} \\ \hat{C}_k &= e^{-iT(\sigma_k^x\sigma_{k+1}^x + \sigma_k^y\sigma_{k+1}^y)\Delta t} \end{aligned} \quad (3.2)$$

Here we make a distinction between indices  $j$  and  $k$ :  $j$  goes from  $1, \dots, L$  and  $k$  goes from  $L + 1, \dots, 2L$ . Then, the symmetric Trotter decomposition is:

$$e^{-i\hat{H}\Delta t} = \left(\prod_j \hat{A}_j\right)\left(\prod_{k \text{ even}} \hat{B}_k\right)\left(\prod_{k \text{ odd}} \hat{C}_k\right)\left(\prod_{k \text{ even}} \hat{B}_k\right)\left(\prod_j \hat{A}_j\right) + \mathcal{O}(\Delta t^3) \quad (3.3)$$

The circuit schematic for this type of decomposition is shown in Fig. 3.1(a). Circuit implementations of the operators necessary are shown in Fig. 3.1(b). We will not go into detail regarding the derivation of these circuits as we want to focus on our simulation results. However, a brief explanation of how to implement operators of the form  $N(\alpha, \beta, \gamma) = \exp[i(\alpha\sigma^x \otimes \sigma^x + \beta\sigma^y \otimes \sigma^y + \gamma\sigma^z \otimes \sigma^z)]$  is given in Ref. [28].

Now, we consider the state of a  $s\bar{s}$  singlet embedded in a QCD string with two non-

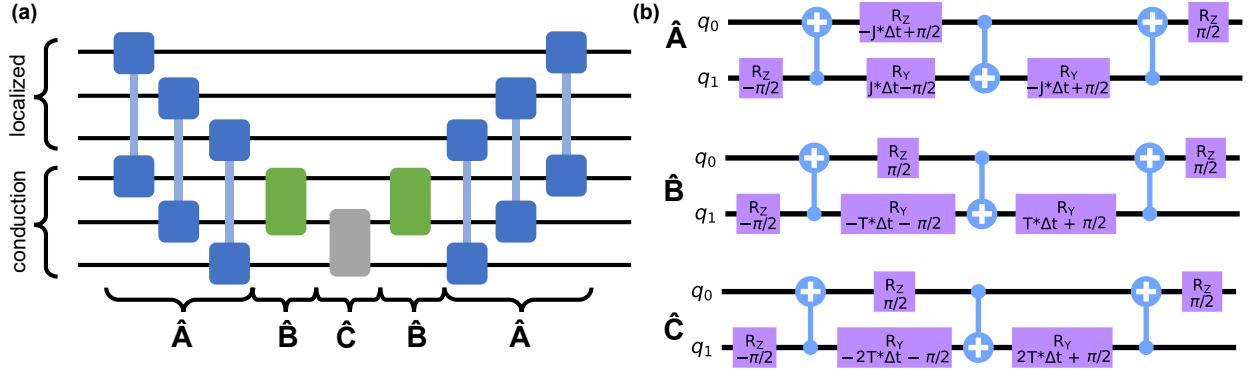


Figure 3.1: (a) A schematic diagram of one Trotter step for the Hamiltonian Eq. 2.9. (b) The implementation of each operator in Eq. 3.2 is also shown. Details of how these circuits can be obtained are given in Ref. [28].

strange valence quarks. This necessitates a lattice of length 4. We initialize the localized spins as

$$|+\rangle \otimes \frac{|01\rangle - |10\rangle}{\sqrt{2}} \otimes |+\rangle \quad (3.4)$$

and the conduction spins at each lattice site as

$$|+\rangle \otimes |+\rangle \otimes |+\rangle \otimes |+\rangle \quad (3.5)$$

where  $|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}$  is the  $\sigma^z$  eigenstate. First, we simulate this state noiselessly using IBM Qiskit's numerical simulator `qasm_simulator`. The evolution of the expectation of the partial-positive transpose (PPT) operator  $(\hat{A}^2)^{T_2} = 4I_1 \otimes I_2 - 4(-\sigma_{1x}\sigma_{2x} - \sigma_{1y}\sigma_{2y} - \sigma_{1z}\sigma_{2z})$  for the singlet is presented in Fig. 3.2(a) for various couplings  $J/T$ . Recall that  $\langle (\hat{A}^2)^{T_2} \rangle < 0$  indicates entanglement; the details are given in Appendix B. We see that entanglement quickly damps out for ferromagnetic coupling. However, for antiferromagnetic coupling, entanglement undergoes a coherent oscillation. This oscillation manifests in an oscillating amplitude of the two-particle correlation function of the singlet, shown in Fig. 3.2(b).

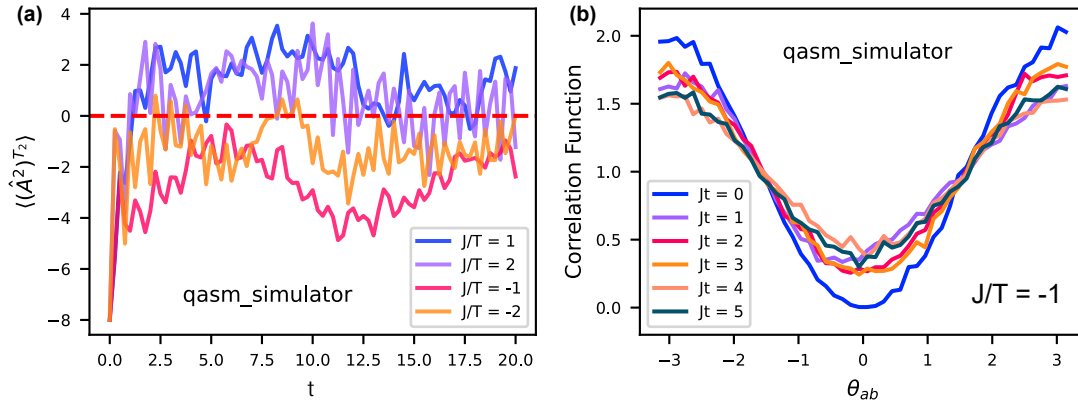


Figure 3.2: (a) Evolution of the PPT measure for various couplings  $J/T$  using  $\Delta t = 0.25$ . Recall that positive (negative)  $J/T$  indicates ferromagnetic (antiferromagnetic) coupling. (b) Two-particle correlation function for the singlet at various timestamps  $Jt$ , evolved using  $\Delta t = 1$ .

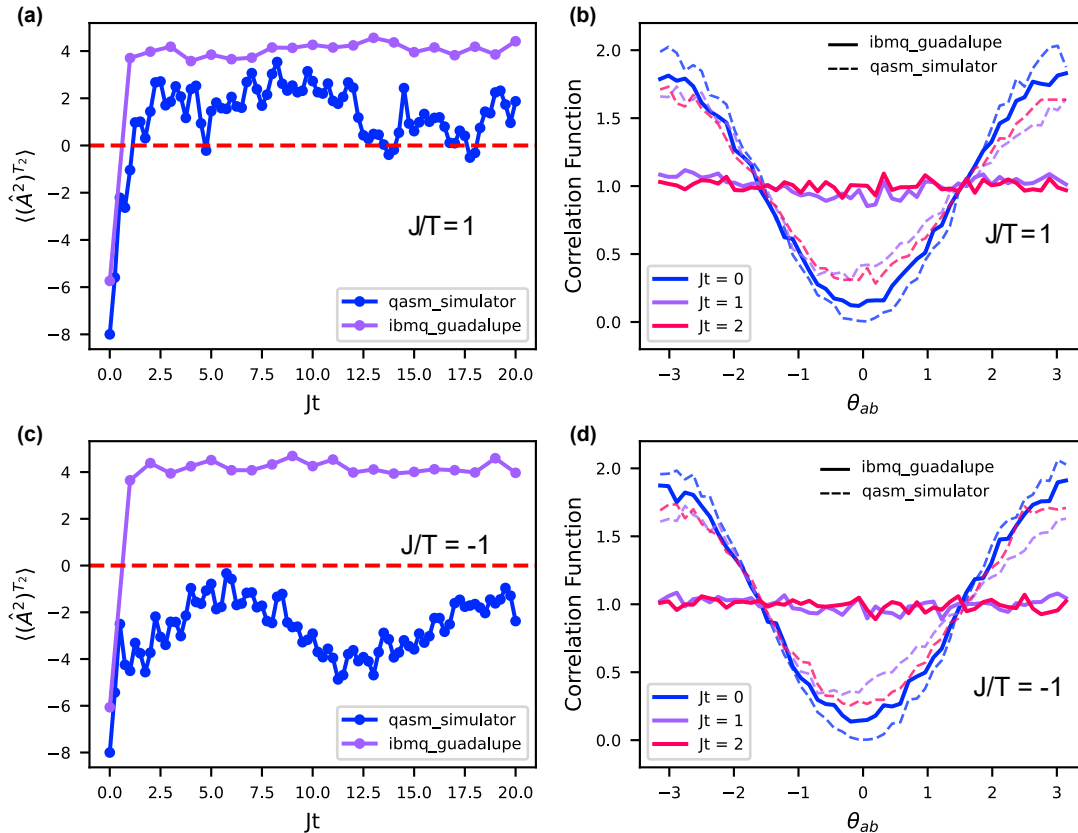


Figure 3.3: (a) PPT measure for  $J/T = 1$ , evolved on `ibmq_guadalupe` using  $\Delta t = 1$ . The numerical results are plotted for comparison. (b) Two-particle correlation function for the singlet for  $J/T = 1$ , evolved on `ibmq_guadalupe` using  $\Delta t = 1$ . (c) PPT measure for  $J/T = -1$ . (d) Two-particle correlation for  $J/T = -1$ .

We also performed the same simulation on IBM Q's 16 qubit computer `ibmq_guadalupe`.

The PPT measure and two-particle correlation function for  $J/T = \pm 1$  obtained using `ibmq_guadalupe` are shown in Fig. 3.3. We see that even after one Trotter step, the noise introduced by the quantum hardware is prohibitive to extracting any physics. Indeed, the correlation functions in 3.3(b) and (d) flatten out immediately after one time step.

## 4 Discussion and conclusions

In this work, we have derived an effective spin Hamiltonian for the 1D KLM at half-filling (Eq. 2.9). We then applied this Hamiltonian to an initial state consisting of one  $s\bar{s}$  singlet embedded in a QCD string with non-strange valence quarks. Numerical simulations of this system's evolution indicate that at weak antiferromagnetic coupling, entanglement undergoes coherent oscillations rather than immediately being washed out by interactions with the itinerant quarks. This suggests that if parton spin entanglement persists in the hadronic products of high energy collisions, then quarks in QCD strings prior to hadronization may experience a weak antiferromagnetic interaction.

Quantum simulations of the QCD string were largely unfruitful. The gate errors of IBM's quantum architecture prevent the determination of any useful information after only one time step of evolution. This highlights the importance of developing targeted quantum error correction schemes in the NISQ era. Until quantum computers are fully fault-tolerant, methods of reducing machine error are crucial to exploiting the quantum devices currently at our disposal.

Future work consists of simulating entanglement evolution in various types of QCD strings over a wider range of parameter space. Existing error correction methods can also be tested on quantum simulations of this system with the goal of reducing noise and amplifying signal.

## 5 Acknowledgements

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## A Details of the two-flavor spin chain model

We provide here details of the computation of the two flavor spin chain model containing four different types of particles:  $s, \bar{s}, u$  and  $\bar{u}$  comprising  $N$  total particles. The particles and antiparticles are coupled into singlets, where there are  $a$  singlets of type  $s\bar{s}$ ,  $b$  singlets of type  $s\bar{u}$  or  $u\bar{s}$ , and  $N/2 - a - b$  singlets of type  $u\bar{u}$ .

This implies that there are  $2a + b$  particles of type  $s$  or  $\bar{s}$  and  $N - 2a - b$  particles of type  $u$  or  $\bar{u}$ . We can therefore number the particles as follows:  $1 \dots a + b/2$  are  $s$ ,  $a + b/2 + 1 \dots 2a + b$  are  $\bar{s}$ ,  $2a + b + 1 \dots a + b/2 + N/2$  are  $u$ , and  $a + b/2 + N/2 + 1 \dots N$  are  $\bar{u}$ .

We can set up the wavefunction by imagining that these particles occupy the lowest  $N/2$  levels of a system, with  $s\bar{s}$  on levels  $1 \dots a$ ,  $s\bar{u}$  on levels  $a + 1 \dots a + b/2$ ,  $u\bar{s}$  on levels  $a + b/2 + 1 \dots a + b$ , and  $u\bar{u}$  on levels  $a + b + 1 \dots N/2$ . The single particle spatial wavefunctions consist of the set:

$$\{\phi^\alpha \mid \alpha \in \{1, 2, \dots, N/2\}\} \quad (\text{A.1})$$

The notation we will use here are superscripts to denote energy level and subscripts to denote particle.

Now, we want to construct a wavefunction that is antisymmetric under exchange of  $s$  with  $s$ ,  $\bar{s}$  with  $\bar{s}$ ,  $u$  with  $u$ , and  $\bar{u}$  with  $\bar{u}$ , in which  $s\bar{s}$ ,  $s\bar{u}$ ,  $u\bar{s}$ , and  $u\bar{u}$  are paired into singlets. Such a wavefunction is given as follows:

$$|\Psi\rangle \sim \sum_{\sigma(s)} \sum_{C(\bar{s})} \sum_{\sigma(u)} \sum_{C(\bar{u})} \text{sgn}(\sigma(s)) \text{sgn}(\sigma(u))$$

$$\begin{aligned}
& \times \underbrace{|S_{\sigma(1), C(a+b/2+1)\dots}\rangle}_{a \ s\bar{s}} \underbrace{|S_{\sigma(a+1), C(a+b/2+N/2+1)\dots}\rangle}_{b/2 \ s\bar{u}} \\
& \times \underbrace{|S_{C(2a+b/2+1), \sigma(2a+b+1)\dots}\rangle}_{b/2 \ \bar{s}u} \underbrace{|S_{\sigma(2a+3b/2+1), C(a+b+N/2+1)\dots}\rangle}_{N/2-a-b \ u\bar{u}} \\
& \times \Phi^+((\sigma(1), \sigma(a + b/2 + 1), \dots, (\sigma(2a + 3b/2 + 1), \sigma(a + b + N/2 + 1))) \quad (A.2)
\end{aligned}$$

The expression above denotes the sum over permutations of the  $s$  and  $u$  particles, as well as the sum over combinations of the  $\bar{s}$  and  $\bar{u}$  particles. For example, we freely permute particles  $1\dots a + b/2$ , or the  $s$  particles. However, for the  $a + b/2$  total  $\bar{s}$  particles, we must choose  $\binom{a + b/2}{a}$   $a$  particles to pair with the permuted  $s$  particles, leaving the rest to pair with the permuted  $u$  particles. In this manner, we count all the possible permutations that account for indistinguishable  $s$ ,  $\bar{s}$ ,  $u$ , and  $\bar{u}$  while keeping the required singlet structures.

The  $\Phi^+$  operator completely symmetrizes the spatial wavefunctions across all the pairs  $(\sigma(1), \sigma(a + \frac{b}{2} + 1))\dots(\sigma(2a + \frac{3b}{2} + 1), \sigma(a + b + \frac{N}{2} + 1))\dots$  as follows:

$$\begin{aligned}
\Phi^+((1, \sigma(a + b/2 + 1))\dots) \sim & \left[ \dots\phi^\alpha(\mathbf{x}_{\sigma(1)})\phi^\alpha(\mathbf{x}_{\sigma(a+b/2+1)})\phi^\beta(\mathbf{x}_{\sigma(2)})\phi^\beta(\mathbf{x}_{\sigma(a+b/2+2)})\dots \right. \\
& \left. + \dots\phi^\alpha(\mathbf{x}_{\sigma(2)})\phi^\alpha(\mathbf{x}_{\sigma(a+b/2+2)})\phi^\beta(\mathbf{x}_{\sigma(1)})\phi^\beta(\mathbf{x}_{\sigma(a+b/2+1)})\dots \right] \quad (A.3)
\end{aligned}$$

Thus the pairs  $(\sigma(1), \sigma(a + \frac{b}{2} + 1))\dots(\sigma(2a + \frac{3b}{2} + 1), \sigma(a + b + \frac{N}{2} + 1))\dots$  all each have an equal probability of occupying any level 1 through  $\frac{N}{2}$ .

The full density matrix of the system is given by  $\rho = |\Psi\rangle\langle\Psi|$ . However, since in our simple model we only care about the spin subsystem, we compute the reduced density matrix  $\rho_{\text{spin}}$  by tracing out the spatial degrees of freedom:  $\rho_{\text{spin}} = \text{Tr}_{\text{spatial}}(|\Psi\rangle\langle\Psi|)$ . If we assume that the spatial wavefunctions are orthogonal,  $\langle\phi^\alpha(\mathbf{x}_i)|\phi^\beta(\mathbf{x}_i)\rangle = 0$  for  $\alpha \neq \beta$ , the symmetrized spatial wavefunctions are then also orthogonal. This allows us to simplify the spin density

matrix to

$$\begin{aligned}
\rho_{\text{spin}} &= \frac{1}{\mathcal{N}} \sum_{\sigma(s)} \sum_{C(\bar{s})} \sum_{\sigma(u)} \sum_{C(\bar{u})} \\
&\times \underbrace{|S_{\sigma(1), C(a+\frac{b}{2}+1)} \cdots \rangle}_{a \ s\bar{s}} \underbrace{\langle S_{\sigma(1), C(a+\frac{b}{2}+1)} \cdots |}_{a \ s\bar{s}} \\
&\times \underbrace{|S_{\sigma(a+1), C(a+\frac{b}{2}+\frac{N}{2}+1)} \cdots \rangle}_{\frac{b}{2} \ s\bar{u}} \underbrace{\langle S_{\sigma(a+1), C(a+\frac{b}{2}+\frac{N}{2}+1)} \cdots |}_{\frac{b}{2} \ s\bar{u}} \\
&\times \underbrace{|S_{C(2a+\frac{b}{2}+1), \sigma(2a+b+1)} \cdots \rangle}_{\frac{b}{2} \ \bar{s}u} \underbrace{\langle S_{C(2a+\frac{b}{2}+1), \sigma(2a+b+1)} \cdots |}_{\frac{b}{2} \ \bar{s}u} \\
&\times \underbrace{|S_{\sigma(2a+\frac{3b}{2}+1), C(a+b+\frac{N}{2}+1)} \cdots \rangle}_{\frac{N}{2}-a-b \ u\bar{u}} \underbrace{\langle S_{\sigma(2a+\frac{3b}{2}+1), C(a+b+\frac{N}{2}+1)} \cdots |}_{\frac{N}{2}-a-b \ u\bar{u}}
\end{aligned} \tag{A.4}$$

where the normalization

$$\mathcal{N} = \left(a + \frac{b}{2}\right)! \left(\frac{N}{2} - a - \frac{b}{2}\right)! \binom{a + \frac{b}{2}}{a} \binom{\frac{N}{2} - a - \frac{b}{2}}{\frac{b}{2}}. \tag{A.5}$$

Now, we want to compute two-particle correlations of the form  $\frac{P(|\hat{n}_1\rangle, |\hat{n}_2\rangle)}{P(|\hat{n}_1\rangle)P(|\hat{n}_2\rangle)}$ , where  $P(|\hat{n}_i\rangle)$  indicates the probability that a spin points in the  $\hat{n}_i$  direction. This form of correlation is of specific interest to us; if the correlation function has the form  $1 - p \cos(\theta_2 - \theta_1)$ , then this indicates that the two-particle state is rotationally invariant, defies a classical explanation by way of local hidden variable theory (LHVT), and, if  $p > \frac{1}{2}$ , is entangled by the entanglement fidelity criterion [14]. To compute  $s\bar{s}$  correlations, we must first compute the reduced density matrix:  $\rho_{1, a+\frac{b}{2}+1} = \text{Tr}_{\text{rest}}(\rho_{\text{spin}})$ . To do so, we consider terms of type  $|S_{1, a+\frac{b}{2}+1}\rangle \langle S_{1, a+\frac{b}{2}+1}|$  and  $|S_{1l}\rangle |S_{a+\frac{b}{2}+1, k}\rangle \langle S_{1l}| \langle S_{a+\frac{b}{2}+1, k}|$  in the full density matrix  $\rho_{\text{spin}}$ .

For  $|S_{1,a+\frac{b}{2}+1}\rangle \langle S_{1,a+\frac{b}{2}+1}|$ , the contribution to the reduced density matrix is

$$\begin{aligned} & \frac{(a+b/2-1)!(N/2-a-b/2)! \binom{a+b/2-1}{a-1} \binom{N/2-a-b/2}{b/2}}{(a+b/2)!(N/2-a-b/2)! \binom{a+b/2}{a} \binom{N/2-a-b/2}{b/2}} \\ & \times |S_{1,a+b/2+1}\rangle \langle S_{1,a+b/2+1}| \\ & = \frac{a}{(a+b/2)^2} |S_{1,a+b/2+1}\rangle \langle S_{1,a+b/2+1}| \end{aligned} \quad (\text{A.6})$$

Likewise, for  $|S_{1l}\rangle |S_{a+\frac{b}{2}+1,k}\rangle \langle S_{1l}| \langle S_{a+\frac{b}{2}+1,k}|$ , the contribution to the reduced density matrix is

$$\frac{(a+\frac{b}{2})^2 - a}{(a+\frac{b}{2})^2} \frac{1}{4} I_1 \otimes I_{a+\frac{b}{2}+1}. \quad (\text{A.7})$$

Thus the net contribution to the reduced density matrix is

$$\rho_{1,a+b/2+1} = \frac{a}{(a+b/2)^2} |S_{1,a+b/2+1}\rangle \langle S_{1,a+b/2+1}| + \frac{(a+b/2)^2 - a}{(a+b/2)^2} \frac{1}{4} I_1 \otimes I_{a+b/2+1} \quad (\text{A.8})$$

With this, we can calculate the two particle probability,

$$\begin{aligned} P(|\hat{n}_1\rangle, |\hat{n}_2\rangle) &= \text{Tr}(|\theta_1\theta_2\rangle \langle\theta_1\theta_2| \rho_{1,a+b/2+1}) \\ &= \frac{a}{2(a+b/2)^2} \sin\left(\frac{\theta_2 - \theta_1}{2}\right)^2 + \frac{(a+b/2)^2 - a}{4(a+b/2)^2} \end{aligned} \quad (\text{A.9})$$

where  $|\hat{n}\rangle = |\theta\rangle = \cos(\theta/2) |0\rangle + \sin(\theta/2) |1\rangle$ . Similarly, the single particle reduced density matrix is

$$\text{Tr}_{a+\frac{b}{2}+1}(\rho_{1,a+\frac{b}{2}+1}) = \frac{a}{2(a+\frac{b}{2})^2} I_1 + \frac{(a+\frac{b}{2})^2 - a}{2(a+\frac{b}{2})^2} I_1$$



$$= \frac{1}{2}I_1, \quad (\text{A.10})$$

giving  $P(|\hat{n}_1\rangle) = \frac{1}{2}$ . Thus the correlation function for  $s\bar{s}$  is

$$\boxed{\frac{P(|\hat{n}_1\rangle, |\hat{n}_2\rangle)}{P(|\hat{n}_1\rangle)P(|\hat{n}_2\rangle)} = 1 - \frac{a}{(a + b/2)^2} \cos(\theta_2 - \theta_1)} \quad (\text{A.11})$$

This result only depends on  $a$  and  $b$ , which makes sense, since  $a$  and  $b$  are the only parameters that determine the number of  $s\bar{s}$ ,  $s\bar{u}$ , and  $u\bar{s}$  singlets.

Using the same procedure, we also calculate the  $s\bar{u}(u\bar{s})$  correlations:

$$\boxed{\frac{P(|\hat{n}_1\rangle, |\hat{n}_2\rangle)}{P(|\hat{n}_1\rangle)P(|\hat{n}_2\rangle)} = 1 - \frac{b/2}{(a + \frac{b}{2})(\frac{N}{2} - a - \frac{b}{2})} \cos(\theta_2 - \theta_1)}, \quad (\text{A.12})$$

and likewise for  $u\bar{u}$ ,

$$\boxed{\frac{P(|\hat{n}_1\rangle, |\hat{n}_2\rangle)}{P(|\hat{n}_1\rangle)P(|\hat{n}_2\rangle)} = 1 - \frac{(N/2 - a - b)}{(\frac{N}{2} - a - \frac{b}{2})^2} \cos(\theta_2 - \theta_1)}. \quad (\text{A.13})$$

As an initial demonstration of simulating many-body fermionic ensembles on quantum hardware, we can also run a correlation function simulation on a quantum circuit in attempt to recreate Eq. A.11. To do so, we must first address how to simulate mixed states  $\rho$  on quantum hardware,

$$\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|. \quad (\text{A.14})$$

This initialization procedure is not obvious as quantum computers are designed to be initialized in pure states. The most straightforward approach– and the one that we adopt– is to initialize and test each constituent pure state  $|\psi_i\rangle \langle\psi_i|$ , subsequently weighting the results by their appropriate probabilities  $p_i$ .

We use this method to try to recover Eq. A.11 on a representative case of the two flavor

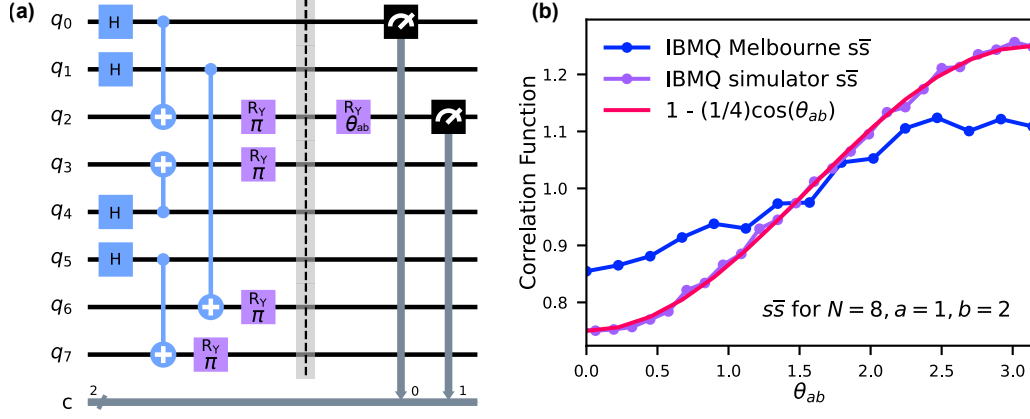


Figure A.1: (a) One of 16 circuits necessary to initialize and simulate the state of  $N = 8$  particles with  $a = 1$   $s\bar{s}$  singlets,  $b = 2$   $s\bar{u}$  and  $u\bar{s}$  singlets and one  $u\bar{u}$  singlet (Eq. A.15). The barrier separates initialization from simulation of the correlation function. (b) Quantum simulation results for the  $s\bar{s}$  correlation function. The analytical prediction (Eq. A.11), is indicated for comparison.

spin chain model: we consider an ensemble of  $N = 8$  particles, with  $a = 1$   $s\bar{s}$  singlets,  $b = 2$  total  $s\bar{u}$  and  $u\bar{s}$  singlets, and one  $u\bar{u}$  singlet. From Eq. A.4, we see that our initial state is then given by the spin density matrix:

$$\begin{aligned}
\rho_{spin} = & \frac{1}{16} ( |S_{13}S_{27}S_{45}S_{68}\rangle \langle S_{13}S_{27}S_{45}S_{68}| + |S_{13}S_{27}S_{46}S_{58}\rangle \langle S_{13}S_{27}S_{46}S_{58}| \\
& + |S_{23}S_{17}S_{45}S_{68}\rangle \langle S_{23}S_{17}S_{45}S_{68}| + |S_{23}S_{17}S_{46}S_{58}\rangle \langle S_{23}S_{17}S_{46}S_{58}| \\
& + |S_{14}S_{27}S_{35}S_{68}\rangle \langle S_{14}S_{27}S_{35}S_{68}| + |S_{14}S_{27}S_{36}S_{58}\rangle \langle S_{14}S_{27}S_{36}S_{58}| \\
& + |S_{24}S_{17}S_{35}S_{68}\rangle \langle S_{24}S_{17}S_{35}S_{68}| + |S_{24}S_{17}S_{36}S_{58}\rangle \langle S_{24}S_{17}S_{36}S_{58}| \\
& + |S_{13}S_{28}S_{45}S_{67}\rangle \langle S_{13}S_{28}S_{45}S_{67}| + |S_{13}S_{28}S_{46}S_{57}\rangle \langle S_{13}S_{28}S_{46}S_{57}| \\
& + |S_{23}S_{18}S_{45}S_{67}\rangle \langle S_{23}S_{18}S_{45}S_{67}| + |S_{23}S_{18}S_{46}S_{57}\rangle \langle S_{23}S_{18}S_{46}S_{57}| \\
& + |S_{14}S_{28}S_{35}S_{67}\rangle \langle S_{14}S_{28}S_{35}S_{67}| + |S_{14}S_{28}S_{36}S_{57}\rangle \langle S_{14}S_{28}S_{36}S_{57}| \\
& + |S_{24}S_{18}S_{35}S_{67}\rangle \langle S_{24}S_{18}S_{35}S_{67}| + |S_{24}S_{18}S_{36}S_{57}\rangle \langle S_{24}S_{18}S_{36}S_{57}| ) \quad (A.15)
\end{aligned}$$

The circuit that initializes the first term in Eq. A.15 on a quantum circuit is shown in Fig. A.1(a).

To simulate a correlation function measurement of Eq. A.11– the second part of the

circuit in Fig. A.1(a)– we first select two representative qubits to probe. For  $s\bar{s}$  correlations, we choose qubits  $q_0$ , representing a  $s$  spin, and  $q_2$ , representing a  $\bar{s}$  spin. Note that as the spins are indistinguishable within the same particle type, our particular selection of qubits does not matter. We then rotate one of the selected qubits,  $q_2$  in our case, by a varying  $\theta_{ab}$  about the  $y$  axis, thus demonstrating varying spin polarizations projected into the  $x - z$  plane. For each  $\theta_{ab}$ , we measure  $q_0$  and  $q_2$  in the  $z$  basis for a total of 1000 shots. The joint probability that both qubits point in the same direction is thus  $n_{11}/1000$ , while the individual probabilities are  $n_{10}/1000$  and  $n_{01}/1000$ , where  $n_{11}$  indicates the counts of the measurement result '11'. For each  $\theta_{ab}$ , we thus see from the general form of the correlation function  $\frac{P(|\hat{n}_1\rangle, |\hat{n}_2\rangle)}{P(|\hat{n}_1\rangle)P(|\hat{n}_2\rangle)}$  that the correlation is

$$1000 \cdot \frac{n_{11}}{n_{10}n_{01}}. \quad (\text{A.16})$$

Averaging this across all 16 circuits, each representing a term in Eq. A.15, for each  $\theta_{ab}$  value produces the plot in Fig. A.1(b). We see from Fig. A.1(b) that though qualitative aspects of the correlation are retained for the computation on IBM's 14 qubit quantum device `ibmq_16_melbourne`, the correlation is much weaker than in the noiseless numerical simulation, which follows our prediction Eq. A.11.

Though this method of simulating mixed states via constructing pure states and averaging is straightforward, it can be cumbersome and redundant as the same simulation must be run multiple times for a single result. Instead, it is useful to note a mixed state Eq. A.14 can also be initialized by constructing the pure state

$$\rho' = \sum_{i=0}^{\mathcal{N}-1} \sqrt{p_i} |\psi_i\rangle |i\rangle. \quad (\text{A.17})$$

Here  $\mathcal{N}$  is the total number of terms in the mixed state, and  $i$  counts over the terms in binary. After initializing  $\rho'$ , if we then only work in the subspace spanned by the  $|\psi_i\rangle$ , we effectively have the mixed state Eq. A.14. As an example, we can construct our ensemble

$N = 8$ ,  $a = 1$ ,  $b = 2$  using this method. The requisite circuit is given in Fig. A.2. For our states, we need a total of  $\log_2(\mathcal{N})$  ancilla initialized in an equal superposition of  $|0\rangle$  to  $|\mathcal{N}\rangle$ . We then perform cSWAP controlled on each of  $i$  on the ancilla,  $i \in [0, \mathcal{N}]$ , targeting the necessary qubits needed to swap from one configuration of permutations to another. Note that though this way of initialization is more concise, it has a much higher cost of multi-qubit gates.

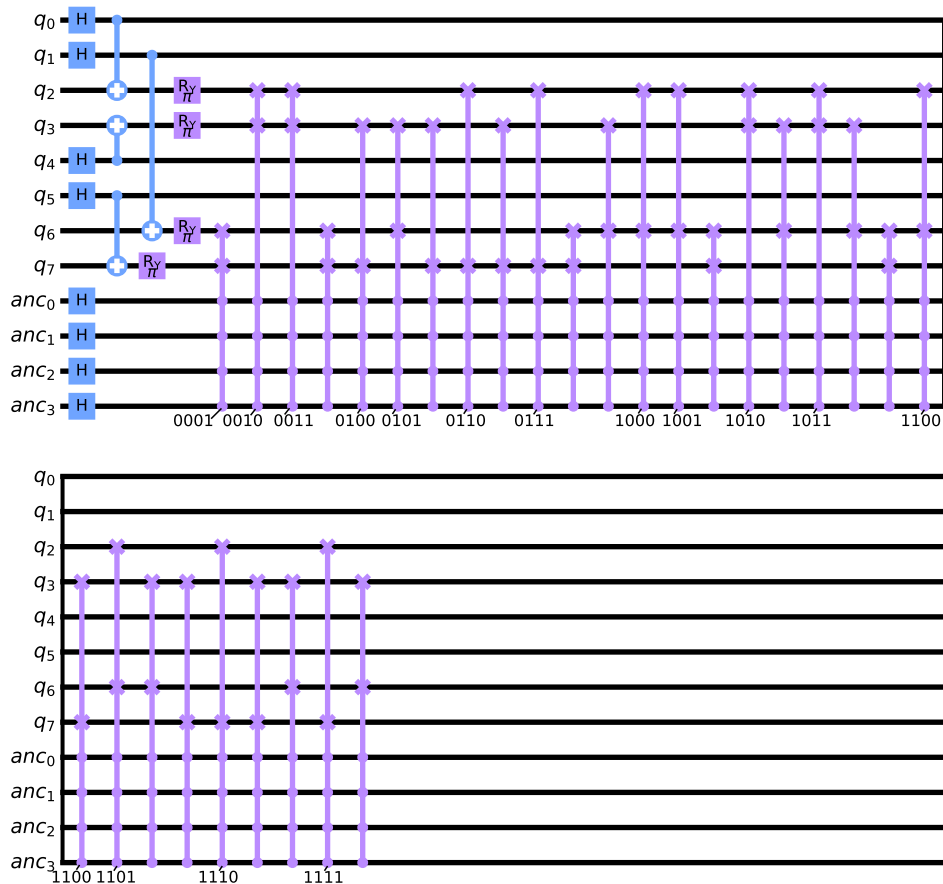


Figure A.2: A circuit that initializes Eq. A.15 in the main quantum register. Swaps controlled on each of  $i \in [0, 16]$ , where  $i$  counts in binary, are applied to initialize a full state of the form Eq. A.17. The binary form of  $i$  for each group of cSWAPs is listed in the circuit.

## B Measuring the PPT criterion of entanglement

For mixed states, the Peres-Horodecki criterion [35, 36], also known as the positive partial transpose (PPT) criterion, can be used to detect separability. This criterion is a

necessary condition<sup>2</sup> for the separability of the joint density matrix  $\rho$  of two systems  $A$  and  $B$ . We first describe this criterion and then briefly discuss observables that can signal its violation, indicating bipartite entanglement between  $A$  and  $B$ .

If we have  $\rho$  on  $\mathcal{H}_A \otimes \mathcal{H}_B$ ,

$$\rho = \sum_{ijkl} p_{kl}^{ij} |i\rangle \langle j| \otimes |k\rangle \langle l|, \quad (\text{B.1})$$

where  $|i\rangle, |j\rangle$  ( $|k\rangle, |l\rangle$ ) label an orthonormal basis for  $A$  ( $B$ ), the partial transpose is defined as follows:

$$\begin{aligned} \rho^{T_B} &= (I \otimes T)\rho \\ &= \sum_{ijkl} p_{kl}^{ij} |i\rangle \langle j| \otimes (|k\rangle \langle l|)^T \\ &= \sum_{ijkl} p_{kl}^{ij} |i\rangle \langle j| \otimes |l\rangle \langle k|. \end{aligned} \quad (\text{B.2})$$

If  $\rho$  is separable, then all the eigenvalues of  $\rho^{T_B}$  are non-negative. If  $\rho^{T_B}$  has one or more negative eigenvalues, then  $\rho$  is entangled.

It is enlightening to first consider this criterion on a Werner state [37], or a two-particle mixed state consisting of maximally entangled and maximally mixed components:

$$\rho_{AB} = p |S_{AB}\rangle \langle S_{AB}| + \frac{(1-p)}{4} I_A \otimes I_B, \quad (\text{B.3})$$

$$\rho_A = \frac{1}{2} I_A. \quad (\text{B.4})$$

We note that our two-particle reduced density matrix in the two-flavor case, Eq. A.8, adopts precisely this form. Then, taking the partial transpose over subspace  $B$ , we find that the

---

<sup>2</sup>It is also sufficient if the dimension of the product space is  $2 \times 2$  or  $2 \times 3$ .

eigenvalues of  $\rho_{AB}^{TB}$  are:

$$\nu_1 = \frac{1+p}{4}; \nu_2 = \frac{1-3p}{4}. \quad (\text{B.5})$$

Here  $\nu_1$  has multiplicity 3. Since a negative eigenvalue indicates entanglement, we deduce that  $\rho$  is entangled for  $1/3 < p \leq 1$ .

As the two-flavor two-particle reduced density matrix is of the specific form of a Werner state, we see that the condition for entanglement according to PPT here is that the coefficient of the cosine in Eq. A.11 be greater than  $\frac{1}{3}$ — is actually less restrictive than that given by entanglement fidelity, which requires the coefficient to be greater than  $\frac{1}{2}$  [14].

We can now question whether accordance with the PPT criterion can be detected via the measurement of observables in a high energy collision. It is important to note that if  $\rho^{TB}$  does indeed have negative eigenvalues, then  $\rho^{TB}$  is not a physical state, and thus the act of partial transposition does not correspond to a physical process. However, for any observable  $\hat{A}$ : [38]

$$\langle \hat{A} \rangle_{\rho^{TB}} = \langle \hat{A}^{TB} \rangle_{\rho}. \quad (\text{B.6})$$

Thus if we can find a positive operator  $\hat{A}^2$ , with corresponding  $(\hat{A}^2)^{TB}$ , such that

$$\langle \hat{A}^2 \rangle_{\rho^{TB}} = \langle (\hat{A}^2)^{TB} \rangle_{\rho} < 0, \quad (\text{B.7})$$

this then indicates that  $\rho^{TB}$  is not positive; hence  $\rho$  is entangled by the PPT criterion. For two-particle density matrices of the form of the Werner state in Eq. B.3, we consider

$$\hat{A} = I_1 \otimes I_2 + \sigma_{1x}\sigma_{2x} - \sigma_{1y}\sigma_{2y} + \sigma_{1z}\sigma_{2z}, \quad (\text{B.8})$$

$$\hat{A}^2 = 4I_1 \otimes I_2 - 4(-\sigma_{1x}\sigma_{2x} + \sigma_{1y}\sigma_{2y} - \sigma_{1z}\sigma_{2z}), \quad (\text{B.9})$$

$$(\hat{A}^2)^{T_2} = 4I_1 \otimes I_2 - 4(-\sigma_{1x}\sigma_{2x} - \sigma_{1y}\sigma_{2y} - \sigma_{1z}\sigma_{2z}). \quad (\text{B.10})$$

For the  $\rho$  given in Eq. B.3,

$$\begin{aligned}\langle \hat{A}^2 \rangle_{\rho^{T_B}} &= \langle (\hat{A}^2)^{T_B} \rangle_{\rho} = \text{Tr}((\hat{A}^2)^{T_2} \rho) \\ &= 4 - 4\text{Tr}(3p |S_{AB}\rangle \langle S_{AB}|) = 4 - 12p.\end{aligned}\tag{B.11}$$

This becomes negative for  $p > \frac{1}{3}$ , as we expect from our previous work. Again, as the reduced two-particle density matrix in the two-flavor case is a Werner state, using  $\hat{A}$  given in Eq. B.8 to assess entanglement via the PPT criterion would also yield that the state is entangled when the coefficient of the cosine in Eq. A.11 is greater than  $\frac{1}{3}$ .

## References

- [1] A. Airapetian, N. Akopov, Z. Akopov, A. Andrus, E. C. Aschenauer, W. Augustyniak, R. Avakian, A. Avetissian, E. Avetissian, S. Belostotski, N. Bianchi, H. P. Blok, H. Böttcher, C. Bonomo, A. Borissov, A. Brüll, V. Bryzgalov, J. Burns, M. Capiluppi, G. P. Capitani, E. Cisbani, G. Ciullo, M. Contalbrigo, P. F. Dalpiaz, W. Deconinck, R. De Leo, M. Demey, L. De Nardo, E. De Sanctis, M. Diefenthaler, P. Di Nezza, J. Dreschler, M. Düren, M. Ehrenfried, A. Elalaoui-Moulay, G. Elbakian, F. Ellinghaus, U. Elschenbroich, R. Fabbri, A. Fantoni, L. Felawka, S. Frullani, A. Funel, D. Gabbert, G. Gapienko, V. Gapienko, F. Garibaldi, G. Gavrillov, V. Gharibyan, F. Giordano, S. Gliske, I. M. Gregor, H. Guler, C. Hadjidakis, D. Hasch, T. Hasegawa, W. H. Hesselink, G. Hill, A. Hillenbrand, M. Hoek, Y. Holler, B. Homme, I. Hristova, G. Iarygin, Y. Imazu, A. Ivanilov, A. Izotov, H. E. Jackson, A. Jgoun, S. Joosten, R. Kaiser, T. Keri, E. Kinney, A. Kisselev, T. Kobayashi, M. Kopytin, V. Korotkov, V. Kozlov, P. Kravchenko, V. G. Krivokhijine, L. Lagamba, R. Lamb, L. Lapikás, I. Lehmann, P. Lenisa, P. Liebing, L. A. Linden-Levy, A. Lopez Ruiz, W. Lorenzon, S. Lu, X. R. Lu, B. Q. Ma, D. Mahon, B. Maiheu, N. C. Makins, L. Manfré, Y. Mao, B. Marianski, H. Marukyan, V. Mexner, C. A. Miller, Y. Miyachi, V. Muccifora, M. Murray, A. Muss-

- giller, A. Nagaitsev, E. Nappi, Y. Naryshkin, A. Nass, M. Negodaev, W. D. Nowak, A. Osborne, L. L. Pappalardo, R. Perez-Benito, N. Pickert, M. Raithel, D. Reggiani, P. E. Reimer, A. Reischl, A. R. Reolon, C. Riedl, K. Rith, S. E. Rock, G. Rosner, A. Rostomyan, L. Rubacek, J. Rubin, D. Ryckbosch, Y. Salomatin, I. Sanjiev, A. Schäfer, G. Schnell, K. P. Schüler, B. Seitz, C. Shearer, T. A. Shibata, V. Shutov, M. Stancari, M. Statera, E. Steffens, J. J. Steijger, H. Stenzel, J. Stewart, F. Stinzing, J. Streit, P. Tait, S. Taroian, B. Tchuiko, A. Terkulov, A. Trzcinski, M. Tytgat, A. Vandenbroucke, P. B. van der Nat, G. van der Steenhoven, Y. van Haarlem, C. van Hulse, M. Varanda, D. Veretennikov, V. Vikhrov, I. Vilardi, C. Vogel, S. Wang, S. Yaschenko, H. Ye, Y. Ye, Z. Ye, S. Yen, W. Yu, D. Zeiler, B. Zihlmann, and P. Zupranski, “Measurement of parton distributions of strange quarks in the nucleon from charged-kaon production in deep-inelastic scattering on the deuteron,” *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, vol. 666, no. 5, pp. 446–450, 2008, ISSN: 03702693. DOI: [10.1016/j.physletb.2008.07.090](https://doi.org/10.1016/j.physletb.2008.07.090). arXiv: [0803.2993](https://arxiv.org/abs/0803.2993).
- [2] D. De Florian, R. Sassot, M. Stratmann, and W. Vogelsang, “Extraction of spin-dependent parton densities and their uncertainties,” *Physical Review D - Particles, Fields, Gravitation and Cosmology*, vol. 80, no. 3, pp. 1–26, 2009, ISSN: 15507998. DOI: [10.1103/PhysRevD.80.034030](https://doi.org/10.1103/PhysRevD.80.034030).
- [3] T. Gehrmann and W. J. Stirling, “Spin-dependent parton distributions from polarized structure function data,” *Zeitschrift für Physik C Particles and Fields*, vol. 65, no. 3, pp. 461–470, 1995, ISSN: 01709739. DOI: [10.1007/BF01556134](https://doi.org/10.1007/BF01556134).
- [4] R. Placakyte, “Parton Distribution Functions,” *Nuclear Physics B - Proceedings Supplements*, vol. 53, no. 1-3, pp. 69–80, Nov. 2011, ISSN: 09205632. DOI: [10.1016/S0920-5632\(96\)00600-7](https://doi.org/10.1016/S0920-5632(96)00600-7). arXiv: [1111.5452](https://arxiv.org/abs/1111.5452). [Online]. Available: <http://arxiv.org/abs/1111.5452>.



- [5] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, “Parton fragmentation and string dynamics,” *Physics Reports*, vol. 97, no. 2-3, pp. 31–145, 1983, ISSN: 03701573. DOI: [10.1016/0370-1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7).
- [6] CMS, “Observation of long-range, near-side angular correlations in proton-proton collisions at the LHC,” *Journal of High Energy Physics*, vol. 2010, no. 91, 2010, ISSN: 10298479. DOI: [10.1007/JHEP09\(2010\)091](https://doi.org/10.1007/JHEP09(2010)091). arXiv: [1009.4122](https://arxiv.org/abs/1009.4122).
- [7] A. Dumitru, K. Dusling, F. Gelis, J. Jalilian-Marian, T. Lappi, and R. Venugopalan, “The ridge in proton-proton collisions at the LHC,” *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, vol. 697, no. 1, pp. 21–25, 2011, ISSN: 03702693. DOI: [10.1016/j.physletb.2011.01.024](https://doi.org/10.1016/j.physletb.2011.01.024). arXiv: [1009.5295](https://arxiv.org/abs/1009.5295). [Online]. Available: <http://dx.doi.org/10.1016/j.physletb.2011.01.024>.
- [8] A. Kovner and V. V. Skokov, “Bose enhancement, the Liouville effective action, and the high multiplicity tail in p-A collisions,” *Physical Review D*, vol. 98, no. 1, pp. 1–15, 2018, ISSN: 24700029. DOI: [10.1103/PhysRevD.98.014004](https://doi.org/10.1103/PhysRevD.98.014004). arXiv: [1805.09296](https://arxiv.org/abs/1805.09296).
- [9] Y.-C. Chen, C.-W. Lin, Y.-J. Lee, and P. Chang, “Search for long range flow-like correlation in hadronic  $e^+e^-$  collisions with Belle,” *PoS*, vol. ICHEP2020, p. 541, 2021. DOI: [10.22323/1.390.0541](https://doi.org/10.22323/1.390.0541).
- [10] V. Andreev *et al.*, “Measurement of charged particle multiplicity distributions in DIS at HERA and its implication to entanglement entropy of partons,” *Eur. Phys. J. C*, vol. 81, no. 3, p. 212, 2021. DOI: [10.1140/epjc/s10052-021-08896-1](https://doi.org/10.1140/epjc/s10052-021-08896-1). arXiv: [2011.01812](https://arxiv.org/abs/2011.01812) [hep-ex].
- [11] Z. Tu, “HERA data on azimuthal decorrelation and charged particle multiplicity spectra probing QCD dynamics and quantum entanglement effects,” *PoS*, vol. ICHEP2020, p. 513, 2021. DOI: [10.22323/1.390.0513](https://doi.org/10.22323/1.390.0513). arXiv: [2011.02875](https://arxiv.org/abs/2011.02875) [hep-ex].

- [12] Z. Tu, D. E. Kharzeev, and T. Ullrich, “Einstein-Podolsky-Rosen Paradox and Quantum Entanglement at Subnucleonic Scales,” *Phys. Rev. Lett.*, vol. 124, no. 6, p. 062 001, 2020. DOI: [10.1103/PhysRevLett.124.062001](https://doi.org/10.1103/PhysRevLett.124.062001). arXiv: [1904.11974](https://arxiv.org/abs/1904.11974) [hep-ph].
- [13] N. A. Tornqvist, “Suggestion for Einstein-podolsky-rosen Experiments Using Reactions Like  $e^+e^- \rightarrow \Lambda\bar{\Lambda} \rightarrow \pi^-p\pi^+\bar{p}$ ,” *Found. Phys.*, vol. 11, pp. 171–177, 1981. DOI: [10.1007/BF00715204](https://doi.org/10.1007/BF00715204).
- [14] W. Gong, G. Parida, Z. Tu, and R. Venugopalan, “Bell-type inequality tests and quantum entanglement from  $\Lambda$ -hyperon spin correlations at high energy colliders,” Jul. 2021. arXiv: [2107.13007](https://arxiv.org/abs/2107.13007). [Online]. Available: <http://arxiv.org/abs/2107.13007>.
- [15] M. Burkardt and R. L. Jaffe, “Polarized  $q \rightarrow \Lambda$  fragmentation functions from  $e^+e^- \rightarrow \Lambda + X$ ,” *Phys. Rev. Lett.*, vol. 70, pp. 2537–2540, 1993. DOI: [10.1103/PhysRevLett.70.2537](https://doi.org/10.1103/PhysRevLett.70.2537). arXiv: [hep-ph/9302232](https://arxiv.org/abs/hep-ph/9302232).
- [16] A. Moretti, “Transversity and  $\Lambda$  polarization at COMPASS,” *PoS*, vol. SPIN2018, P. Lenisa, G. Ciullo, M. Contalbrigo, and L. Pappalardo, Eds., p. 138, 2018. DOI: [10.22323/1.346.0138](https://doi.org/10.22323/1.346.0138). arXiv: [1901.01735](https://arxiv.org/abs/1901.01735) [hep-ex].
- [17] R. P. Feynman, “Simulating physics with computers,” *International Journal of Theoretical Physics*, vol. 21, no. 6-7, pp. 467–488, Jun. 1982, ISSN: 0020-7748. DOI: [10.1007/BF02650179](https://doi.org/10.1007/BF02650179). [Online]. Available: <http://link.springer.com/10.1007/BF02650179>.
- [18] G. Cohen, E. Gull, D. R. Reichman, and A. J. Millis, “Taming the dynamical sign problem in real-time evolution of quantum many-body problems,” Oct. 2015. DOI: [10.1103/PhysRevLett.115.266802](https://doi.org/10.1103/PhysRevLett.115.266802). arXiv: [1510.03534](https://arxiv.org/abs/1510.03534). [Online]. Available: <http://arxiv.org/abs/1510.03534%20http://dx.doi.org/10.1103/PhysRevLett.115.266802>.
- [19] D. Mitra, P. T. Brown, E. Guardado-Sanchez, S. S. Kondov, T. Devakul, D. A. Huse, P. Schauß, and W. S. Bakr, “Quantum gas microscopy of an attractive Fermi–Hubbard system,” *Nature Physics*, vol. 14, no. 2, pp. 173–177, Feb. 2018, ISSN: 1745-2473. DOI:

- [10.1038/nphys4297](https://doi.org/10.1038/nphys4297). [Online]. Available: <http://www.nature.com/articles/nphys4297>.
- [20] W. S. Bakr, J. I. Gillen, A. Peng, S. Fölling, and M. Greiner, “A quantum gas microscope for detecting single atoms in a Hubbard-regime optical lattice,” *Nature*, vol. 462, no. 7269, pp. 74–77, Nov. 2009, ISSN: 0028-0836. DOI: [10.1038/nature08482](https://doi.org/10.1038/nature08482). [Online]. Available: <http://www.nature.com/articles/nature08482>.
- [21] J.-y. Choi, S. Hild, J. Zeiher, P. Schauss, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross, “Exploring the many-body localization transition in two dimensions,” *Science*, vol. 352, no. 6293, pp. 1547–1552, Jun. 2016, ISSN: 0036-8075. DOI: [10.1126/science.aaf8834](https://doi.org/10.1126/science.aaf8834). [Online]. Available: <https://www.sciencemag.org/lookup/doi/10.1126/science.aaf8834>.
- [22] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Lüschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch, “Observation of many-body localization of interacting fermions in a quasirandom optical lattice,” *Science*, vol. 349, no. 6250, pp. 842–845, Aug. 2015, ISSN: 0036-8075. DOI: [10.1126/science.aaa7432](https://doi.org/10.1126/science.aaa7432). [Online]. Available: <https://www.sciencemag.org/lookup/doi/10.1126/science.aaa7432>.
- [23] P. Bordia, H. Lüschen, S. Scherg, S. Gopalakrishnan, M. Knap, U. Schneider, and I. Bloch, “Probing Slow Relaxation and Many-Body Localization in Two-Dimensional Quasiperiodic Systems,” *Physical Review X*, vol. 7, no. 4, p. 041 047, Nov. 2017, ISSN: 2160-3308. DOI: [10.1103/PhysRevX.7.041047](https://doi.org/10.1103/PhysRevX.7.041047). [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevX.7.041047>.
- [24] P. J. J. O’Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Nealey, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, “Scalable Quan-

- tum Simulation of Molecular Energies,” *Physical Review X*, vol. 6, no. 3, p. 031 007, Jul. 2016, ISSN: 2160-3308. DOI: [10.1103/PhysRevX.6.031007](https://doi.org/10.1103/PhysRevX.6.031007). [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevX.6.031007>.
- [25] A. Cervera-Lierta, “Exact Ising model simulation on a quantum computer,” Jul. 2018. DOI: [10.22331/q-2018-12-21-114](https://doi.org/10.22331/q-2018-12-21-114). arXiv: [1807.07112](https://arxiv.org/abs/1807.07112). [Online]. Available: <http://arxiv.org/abs/1807.07112%20http://dx.doi.org/10.22331/q-2018-12-21-114>.
- [26] N. Klco, E. F. Dumitrescu, A. J. McCaskey, T. D. Morris, R. C. Pooser, M. Sanz, E. Solano, P. Lougovski, and M. J. Savage, “Quantum-classical computation of Schwinger model dynamics using quantum computers,” *Physical Review A*, vol. 98, no. 3, p. 032 331, Sep. 2018, ISSN: 2469-9926. DOI: [10.1103/PhysRevA.98.032331](https://doi.org/10.1103/PhysRevA.98.032331). [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.98.032331>.
- [27] J. Preskill, “Quantum Computing in the NISQ era and beyond,” *Quantum*, vol. 2, p. 79, Aug. 2018, ISSN: 2521-327X. DOI: [10.22331/q-2018-08-06-79](https://doi.org/10.22331/q-2018-08-06-79). [Online]. Available: <https://quantum-journal.org/papers/q-2018-08-06-79/>.
- [28] A. Smith, M. S. Kim, F. Pollmann, and J. Knolle, “Simulating quantum many-body dynamics on a current digital quantum computer,” *npj Quantum Information*, vol. 5, no. 1, pp. 1–16, 2019, ISSN: 20566387. DOI: [10.1038/s41534-019-0217-0](https://doi.org/10.1038/s41534-019-0217-0). arXiv: [1906.06343](https://arxiv.org/abs/1906.06343).
- [29] M. Troyer, “Ferromagnetism of the one-dimensional Kondo-lattice model: A quantum Monte Carlo study,” *Physical Review B - Condensed Matter and Materials Physics*, vol. 47, no. 5, pp. 4–7, 1993.
- [30] C. Santos and W. Nolting, “Ferromagnetism in the Kondo-lattice model,” *Physical Review B - Condensed Matter and Materials Physics*, vol. 65, no. 14, pp. 1–11, 2002, ISSN: 1550235X. DOI: [10.1103/PhysRevB.65.144419](https://doi.org/10.1103/PhysRevB.65.144419).

- [31] P. Jordan and E. Wigner, “Über das Paulische Äquivalenzverbot,” *Zeitschrift für Physik*, vol. 47, no. 9-10, pp. 631–651, Sep. 1928, ISSN: 1434-6001. DOI: [10.1007/BF01331938](https://doi.org/10.1007/BF01331938). [Online]. Available: <http://link.springer.com/10.1007/BF01331938>.
- [32] J.-m. Reiner, “Analog and Digital Quantum Simulation of Fermionic Systems : Superconducting Emulator Circuits and Effects of Gate Errors,” Ph.D. dissertation, 2019.
- [33] H. Tsunetsugu, Y. Hatsugai, K. Ueda, and M. Sigrist, “Spin-liquid ground state of the half-filled Kondo lattice in one dimension,” *Physical Review B*, vol. 46, no. 5, pp. 1–4, 1992. DOI: <https://doi.org/10.1103/PhysRevB.46.3175>.
- [34] R. Peters and N. Kawakami, “Ferromagnetic state in the one-dimensional Kondo lattice model,” *Physical Review B - Condensed Matter and Materials Physics*, vol. 165107, no. July, pp. 1–7, 2012. DOI: [10.1103/PhysRevB.86.165107](https://doi.org/10.1103/PhysRevB.86.165107).
- [35] A. Peres, “Separability criterion for density matrices,” *Phys. Rev. Lett.*, vol. 77, pp. 1413–1415, 1996. DOI: [10.1103/PhysRevLett.77.1413](https://doi.org/10.1103/PhysRevLett.77.1413). arXiv: [quant-ph/9604005](https://arxiv.org/abs/quant-ph/9604005).
- [36] P. Horodecki, “Separability criterion and inseparable mixed states with positive partial transposition,” *Phys. Lett. A*, vol. 232, p. 333, 1997. DOI: [10.1016/S0375-9601\(97\)00416-7](https://doi.org/10.1016/S0375-9601(97)00416-7). arXiv: [quant-ph/9703004](https://arxiv.org/abs/quant-ph/9703004).
- [37] R. F. Werner, “Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model,” *Phys. Rev. A*, vol. 40, pp. 4277–4281, 8 Oct. 1989. DOI: [10.1103/PhysRevA.40.4277](https://doi.org/10.1103/PhysRevA.40.4277). [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.40.4277>.
- [38] Y. B. Band and P. A. Mello, “Partial transposition in a finite-dimensional Hilbert space: physical interpretation, measurement of observables, and entanglement,” *Quantum Studies: Mathematics and Foundations*, vol. 5, no. 2, pp. 177–188, 2018, ISSN: 21965617. DOI: [10.1007/s40509-017-0120-3](https://doi.org/10.1007/s40509-017-0120-3). arXiv: [1705.09613](https://arxiv.org/abs/1705.09613).

Optical model for a vacuum transport upgrade

Student Intern

Ben Herr, CCI, Suffolk County Community College

Lab Mentors

Marcus Babzien, Mikhail Polyanskiy, Accelerator Test Facility, Brookhaven National Laboratory

## I. Abstract

When light travels through an open space, it has a chance to be diffracted by small anomalies in the air. Modeling beam propagation and refraction through transmissive elements for a vacuum transport system will allow the laser to traverse the system without outside interference. The objective of this project is to validate the mechanical design through numerical simulations of the beam propagation. Simulating the beam transport through these optics validates the mechanical design. Apertures are set up at exact measurements to the nearest millimeter when we transport the beam over long distances. The mirrors are finite sized, have sharp edges, and reflect on one side. The model will show that the laser will have an undistorted beam profile throughout the apparatus. Using online resources as well as Zemax, we are researching optimizations within 3-dimensional modeling and ray-tracing. This project is being undertaken at the laser test facility at Brookhaven National Labs to optimize the already functioning laser system.

## II. Introduction

“Infrared light is light with a vacuum wavelength longer than  $\approx 700\text{--}800$  nm, the upper limit of the visible wavelength range. That limit is not well-defined, as the responsivity of the eye is reduced very gradually in that spectral region. Although the responsivity e.g. at 700 nm is already very low, even the light from some laser diodes at wavelengths beyond 750 nm can be seen if that light is sufficiently intense. Such light may be harmful for the eye even if it is not perceived as very bright. The upper limit of the infrared spectral region in terms of wavelength is also not precisely defined; it is usually understood to be roughly 1 mm.”<sup>1</sup> The objective of this project is to validate the mechanical design through numerical simulations of the beam propagation. Simulating the beam transport through these optics validates the mechanical design. We placed mirrors and lenses in specific locations to make sure the beam doesn't get clipped inside the transport tube. Ray-tracing is an idealized propagation for a beam which has no diffraction while physical optics propagate waves that diffract as they propagate.



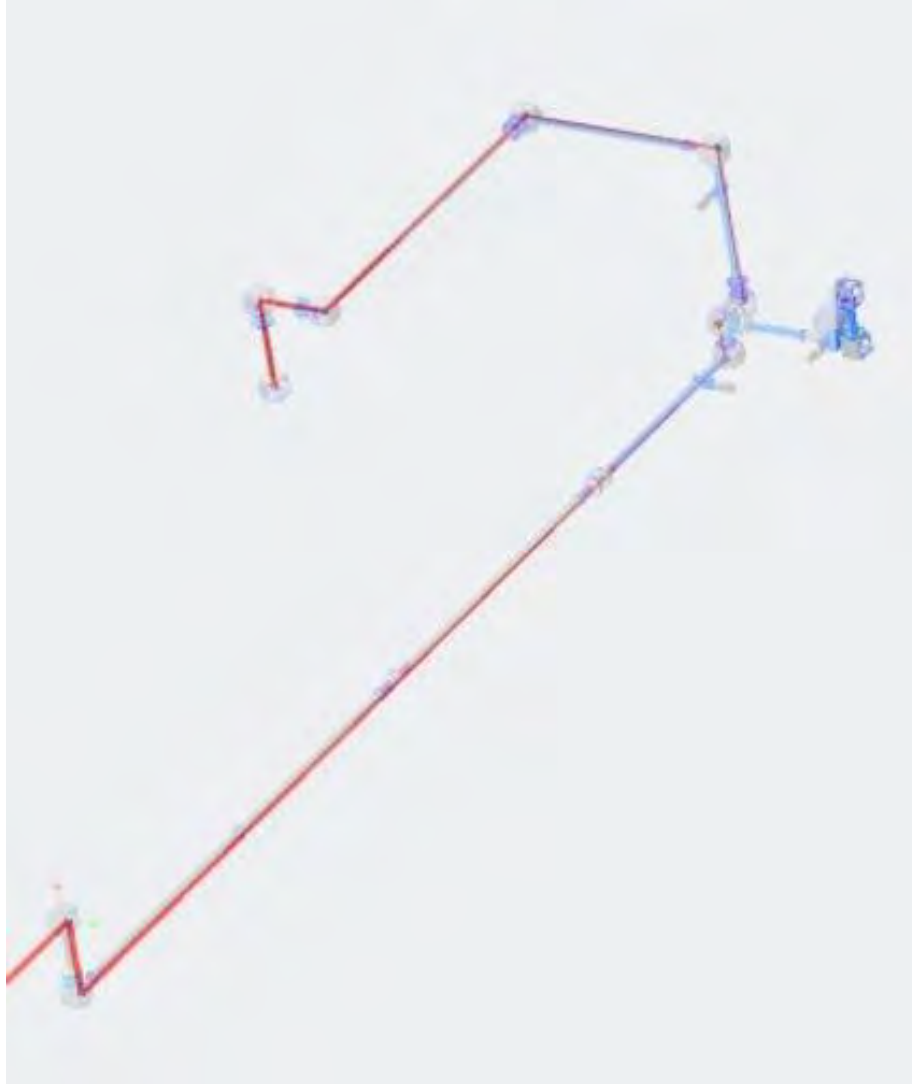
### III. Zemax

The screenshot displays a detailed Zemax table with columns for Surface Type, Comment, Aperture, Thickness, Material, Coating, Curvature, and various ray parameters. The table is color-coded: blue for lenses, red for coordinate breaks, and grey for mirrors. The data includes surface numbers (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100) and their corresponding properties.

#### Example 1: Full System Zemax Table

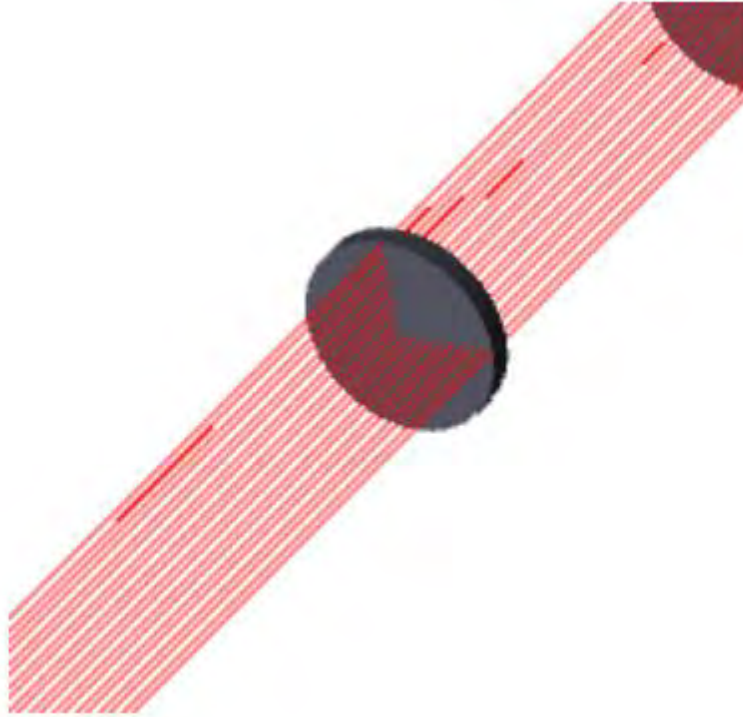
Example 1 shows the Zemax table of the numerical simulations where the laser was computed.

A few lines to point out are the blue (lenses), red (coordinate breaks), and grey (mirrors). Lenses are designed to focus, homogenize, or shape the laser beams. The type of material used for the lenses are fused silica. Laser mirrors are ideal for beam steering applications. If the mirrors were off by a minute fraction, the beam would have clipped with the transport tube and would not make it to the end of the system. Coordinate breaks are dummy surfaces, meaning they have no refractive or reflective power and cannot bend rays. Its sole purpose is to define a new coordinate system in terms of the current one. Tilting about the XYZ axis' will allow the user to change the lasers beam path.



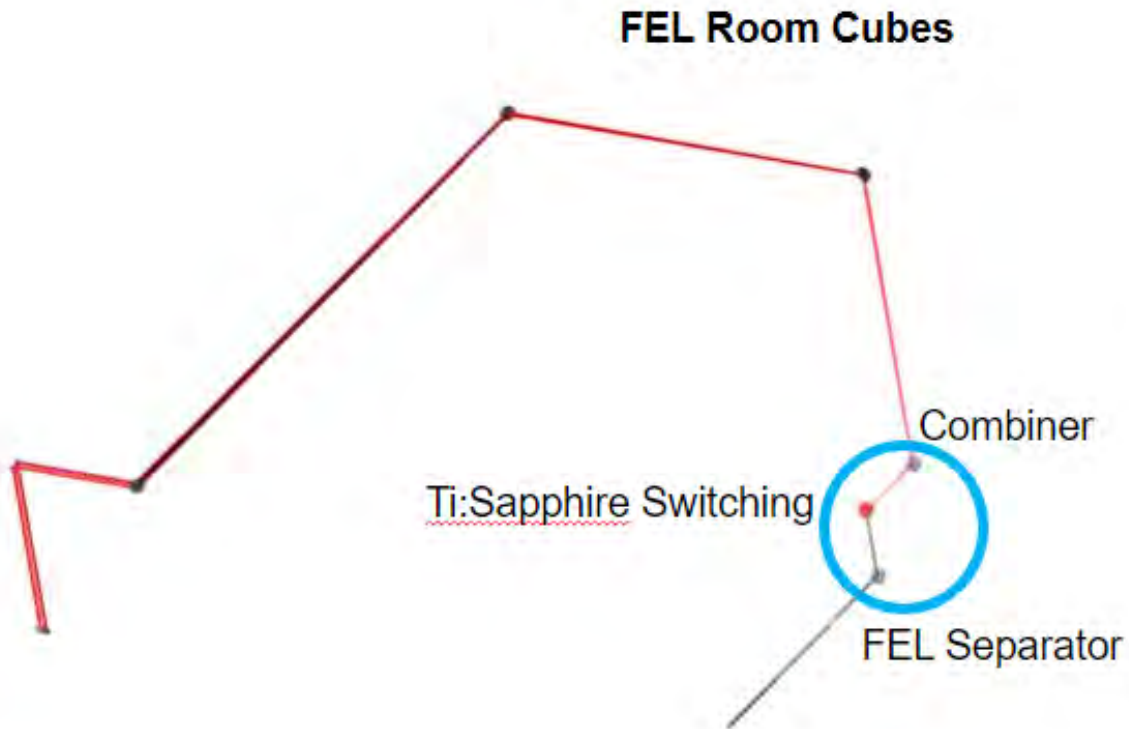
### **Example 2: Laser Optics overlaid with mechanical design**

Example 2 shows the downstream segment of the transport. The laser is overlaid with the 3d rendering of mechanical design. Angles had to be precise and correct in order to present the optical and mechanical simulations together. If numerical values were input incorrectly, the mirrors could be off and sending the beam in the completely opposite direction. Discussions and approach to working with mechanical engineer allowed projection angles to line up their mechanical rendering to the optical simulations.



### **Example 3: Laser Lens**

Two separate lenses were used in the beginning of the laser system. One side of both lenses has a convex curve, making sure the laser does focus at a certain point later down the system. The first lens has the laser hitting the convex side first, making the laser spread in between the first and second lens. The second side of the first lens is flat, having a radius of infinity. The laser reaches the second lens, which the laser hits the flat side first, then traveling out the convex side, continuing on to the rest of the laser system.



#### Example 4: FEL Room Cubes

Ray-tracing is shown in Example 4, with the red beam being the 800 nm wavelength while the black beam is 1.064  $\mu\text{m}$  wavelength. Zoomed in fully you can see that at the FEL Separator the 800 nm wavelength is more focused than the 1.064  $\mu\text{m}$  wavelength. The Ti:Sapphire Switching station is optimal for taking measurements of the beam as both the 800 nm and the 1.064  $\mu\text{m}$  wavelength are identical at this location. There are extra cubes attached splitting off to the right of the configuration known as the table penetration, allowing the lab to take those measurements. When the laser gets to the Combiner mirror, the 1.064  $\mu\text{m}$  wavelength is more focused than the 800 nm wavelength. This is due to the fused silica lenses that were placed at the start of the system.

#### IV. Conclusion

We were able to fully simulate the Near-Wave Infrared (NWIR) laser system to overlay with the 3d model of the mechanical design. Part of the simulation was to ray-trace the validated imaging configuration used. This configuration was shown to minimize pointing instability at the location of the FEL room cubes. The high level of agreement between the mechanical projection and optical simulation is what gives high confidence that the project will meet the design goals. The project is going forward based on the simulations that lead us to believe there will be little to no clipping of the laser. With more time, we would have continued to simulate the Long-Wave Infrared (LWIR) laser.

#### V. Acknowledgements

I'd like to thank the laser upgrade division for welcoming me into their team. A special thank you to Marcus and Mikhail for guiding me through the operations that I didn't understand. I'd also like to thank the intern team leads for helping us interns throughout this process. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program (CCI),

## VI. References

1) <https://www.rp-photonics.com/>

Experimental design for the measurement and analysis  
of vibration-rotation spectra of molecules relevant to  
extra-terrestrial bodies

Joshua Heuvel-Horwitz, Chemistry Department, Stony Brook University,  
Stony Brook, NY 11790

Trevor Sears, Chemistry Department, Stony Brook University, Stony Brook,  
NY 11790

## Abstract

High-precision spectroscopy in the near-infrared (NIR) will be central to improving future measurements in many fields, including meteorology and astronomy. For the accurate spectroscopic analysis of planetary atmospheres, including Earth's, high-precision NIR spectra for simple molecules such as acetylene, methane, water, and carbon dioxide are needed, but current data still contains myriad ambiguous and unassigned features at lower than ideal precision. Consequently, there still exists data from NASA missions such as the Cassini probe that have been unable to be analyzed due to poor laboratory references. As more space missions are initiated, notably NASA's "New Frontiers" program and the Titan Saturn System Mission, new planetary atmospheric data will be collected, and such data will require adequate high-precision laboratory data in order to make confident conclusions about those atmospheres. Recording all this data is a massive undertaking, and it will require many spectroscopists to accomplish, so cultivating interest in the field is a crucial step towards accelerating progress. Therefore, we present an experimental design that records NIR absorption spectra of simple molecules using relatively cheap materials as compared to the highest precision experiments but achieves a much higher precision than traditional off-the-shelf spectrometers traditionally used in teaching laboratories. We record several lines of acetylene in helium as a proof of concept, and we measure Boltzmann's constant as an example for a teaching laboratory and as a benchmark for the quality of the spectrometer.



# 1 Introduction

Spectroscopy is broadly the study of how electromagnetic radiation interacts with matter. It is a tool ubiquitous in modern science across most if not all fields. Spectroscopy in the near-infrared region is particularly useful in remote-sensing measurements, specifically in the detection of simple molecules such as acetylene, methane, water, and carbon-dioxide. The atmosphere of Titan, Saturn's largest moon, has been extensively studied by remote measurements and in-situ measurements alike, and remains a topic of great interest due to the complex organic photochemistry in the haze layer, which may help to understand early Earth chemistry and the conditions that gave rise to life.

Measurements of Titan's atmosphere were collected by the Cassini spacecraft and its Huygens probe and also by the ALMA observatory from Earth [1]. These measurements revealed all of what we know about Titan and its atmosphere today. However, even though the last of these measurements were taken in 2017, much of the data has not yet been analyzed, mostly due to shortcomings in our current spectroscopic models. The high-resolution transmission molecular absorption (HITRAN) database is, by their own description "a compilation of spectroscopic parameters that a variety of computer codes use to predict and simulate the transmission and emission of light in the atmosphere" [2]. Currently HITRAN effectively only supports the Voigt line shape model, a convolution of an exponential function due to the Doppler Effect from the distribution of molecular velocities and a Lorentzian function due to collisional, lifetime, broadening for the vast majority of transitions, due to the lack of higher order data available. For proper analysis of data from Cassini-Huygens and future missions, however, the more accurate Hartmann-Tran Profile (HTP) line shape model is necessary; the Voigt profile does not account for higher order effects [3]. Measuring the parameters that would make modeling with the

HTP viable requires high-precision spectroscopy experiments involving an optical frequency comb. Such experiments are difficult and expensive to assemble, and data measurement with them is painstaking and time costly. With new space research missions and projects planned such as the Dragonfly Probe mission to Titan (under the broader NASA New Frontiers Program), the Transiting Exoplanet Survey Satellite, and the James Webb Space Telescope, demand for high resolution laboratory data will only increase [4].

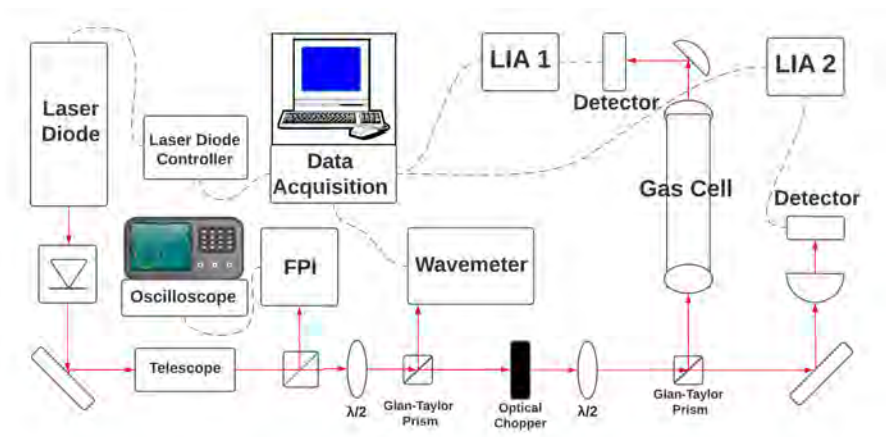
To keep up with this demand, it is imperative that undergraduate chemistry students continue to pursue careers in high resolution spectroscopy research. In an effort to generate interest and teach undergraduates the underlying principles of these experiments, we present an experimental design for a spectrometer for use in an undergraduate chemistry laboratory course. This spectrometer will measure in the near-infrared region, and is intended for use primarily with acetylene or methane in a mostly inert gas mixture. Unlike off-the-shelf Fourier Transform Infrared Spectrometers (FTIRs) typically used in undergraduate laboratories, our spectrometer can measure much smaller frequency steps ( $0.001\text{ cm}^{-1}$  compared to  $0.2\text{ cm}^{-1}$ ), and its optical components are easily seen for better learning. We demonstrate the functionality of our spectrometer by measuring the P(15) line of the  $\nu_1 + \nu_3$  band of acetylene at various pressures. As a sample experiment, we measure the Boltzmann constant from our fitted results.

## 2 Experimental Design

Our experiment follows similar principles as those by McRaven et al., Arteaga et al., and Fasci et al. [5] [6] [7]. We start with a Sacher model Lynx extended cavity diode laser (ECDL). The beam exits the cavity and goes through an optical isolator to eliminate reflections and feedback into the cav-

ity, and then through a telescope to reduce the beam size. Then 4% of the beam is split off to a Fabry Perot Interferometer (FPI) the output of which is monitored on an oscilloscope. We do this to ensure that the ECDL is lasing properly. The beam then passes through a half-wave plate and a Glan-Taylor (GT) prism that deflects a variable amount of light to our Bristol Instruments 621 Series Wavelength Meter (WM). We could have simply deflected 4% of the light again, but we do this because the optical alignment of the WM is in our opinion the most difficult part of the assembly of this experiment due to the need for completely co-linear beams. Allowing for more light to be deflected during alignment can somewhat reduce this difficulty, especially when working with lower power lasers. Next, the beam goes through a Thorlabs MC2000B Optical Chopper, which provides amplitude modulation to the beam and acts as a reference signal to our lock-in amplifiers (LIAs). The beam then passes through another half-wave plate and GT prism, this time to balance the beam power so it is split 50% two ways. One path goes to a lens which focuses the beam onto a New Focus 2053 detector that serves as a reference. The other goes through our gas cell and to an off-axis parabola (OAP) that focuses the beam onto another detector which serves as the signal. Signals from both detectors are sent to two lock-in amplifiers, Optical Finesse  $\mu$ LIA 320s, which use the chopper amplitude modulation reference frequency to eliminate unwanted noise frequencies. Comparing the signal and reference gives us a proportion of light absorbed by the gas, which is plotted over a range of frequencies to obtain line shapes.

For data acquisition and instrument communication with the WM, LIAs, and ECDL controller, we use a custom Python program written for this project. Inconveniently, each of our devices uses a different communication protocol, as shown in table 1. However, this resulted in a data acquisition program (DAQ)



**Figure 1:** Optical schematic of the experiment. LIA = Lock-In Amplifier, FPI = Fabry-Perot Interferometer. The laser is a Sacher model Lynx Extended Cavity Laser Diode. The wavemeter is a Bristol Instruments 621 Series Wavelength Meter. The Lock-In Amplifiers are Optical Finesse uLIA-320s. We use a ThorLabs MC2000B Optical Chopper.

that provides a good example of how to properly communicate using each one in Python.

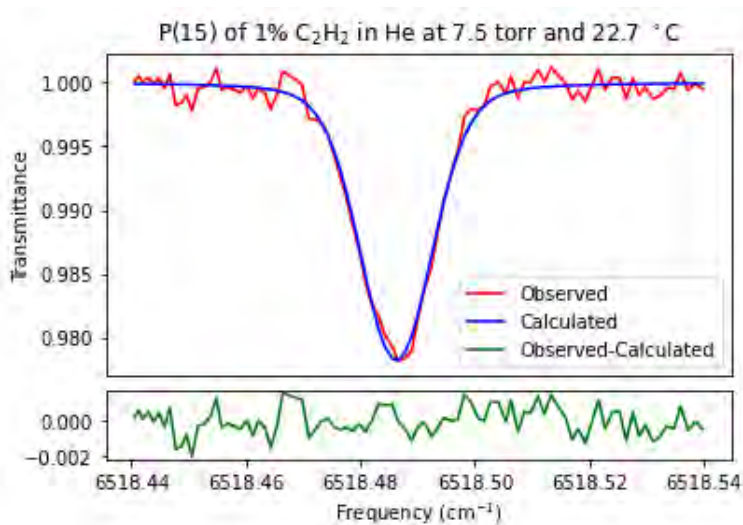
Layer:	Device:	Bristol Series 621 WM	Optical Finesse uLIA-320	Sacher Model Lynx ECLD
Application		Python DAQ	Python DAQ	Python DAQ
Session		Python Bristol Class	C Object (ctypes)	Python Sacher Class
Data Link		RFC 1549	Proprietary Format	GPIB
Physical		UART/USB	USB	USB

**Table 1:** Communication with the devices in our experiment as Open System Interconnection (OSI) model layers. DAQ = Data Acquisition system.

### 3 Results

We recorded the P(15) line of acetylene at 7.5, 11.6, 14.3, 21.6, and 30.1 torr, all at approximately room temperature. We fitted the data using a Voigt line profile. These results are available in figures 2 through 7. In figure 7 we

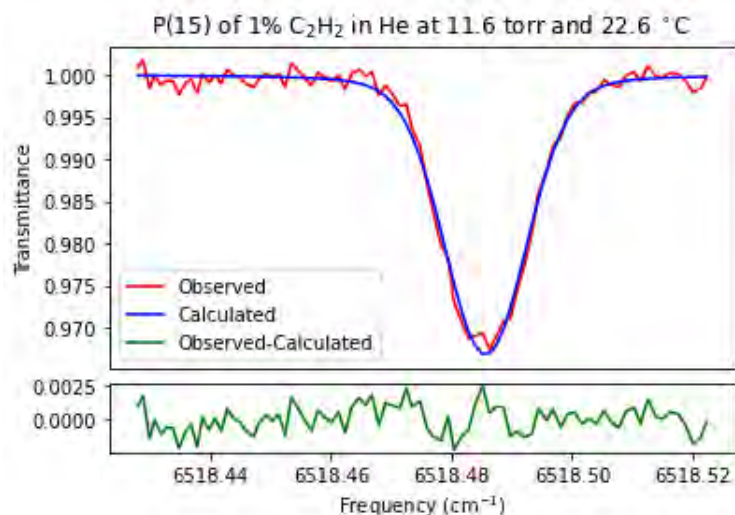
combine all the fitted profiles to show the effects of pressure broadening. From our line fits, we calculate the Boltzmann Constant, in  $10^{-23}$ J/K, increasing in pressure, to be:  $0.95 \pm 0.52$ ,  $1.19 \pm 0.38$ ,  $1.04 \pm 0.31$ ,  $1.21 \pm 0.24$ , and  $0.96 \pm 0.20$ .



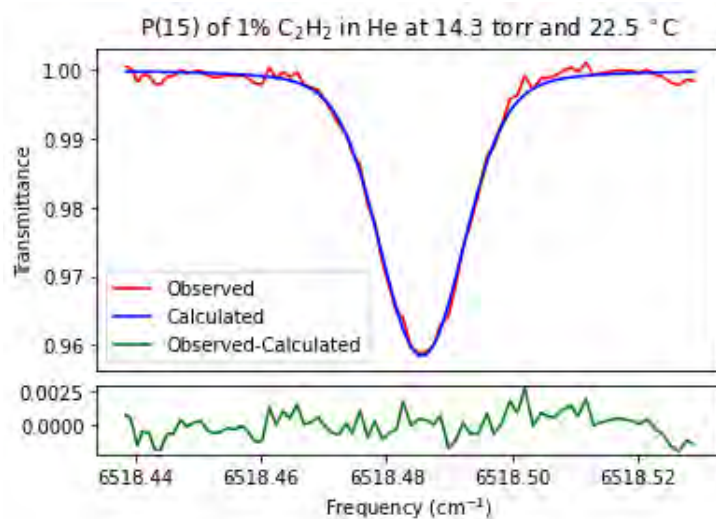
**Figure 2:** Absorption data, fit and residual for the P(15) line of the  $\nu_1 + \nu_3$  band of C<sub>2</sub>H<sub>2</sub> at 7.5 torr and 22.7 °C.

## 4 Discussion

Theoretically, one would expect a spectroscopic transition to be an infinitesimally thin line at a single frequency. We do not see this in experiments, instead we see a curve that we call a line profile. A line profile is a statistical distribution of frequencies that can be described by three characteristics: line center, line shape, and line intensity. Figure 7 shows the pressure broadening effect on line shape and intensity. As pressure increases, the probability of the P(15) transition occurring increases, increasing the intensity. A higher pressure also means more collisions, which increases the likelihood of an interruption during the absorption and emission process, which increases the uncertainty of

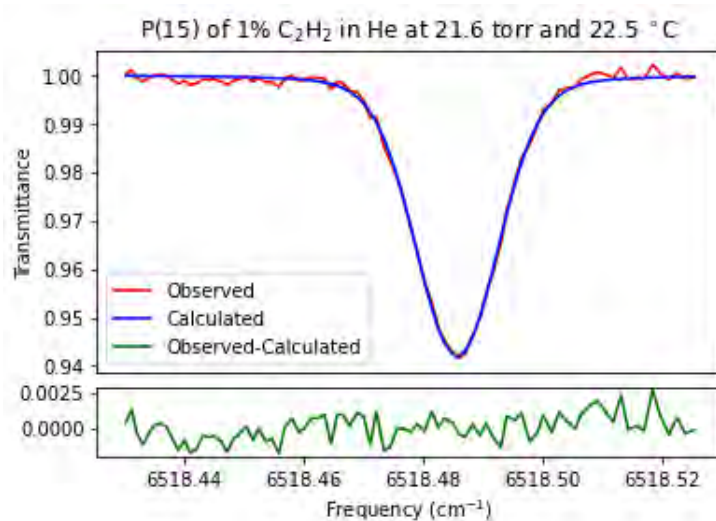


**Figure 3:** Absorption data, fit and residual for the P(15) line of the  $\nu_1 + \nu_3$  band of C<sub>2</sub>H<sub>2</sub> at 11.6 torr and 22.6 °C.

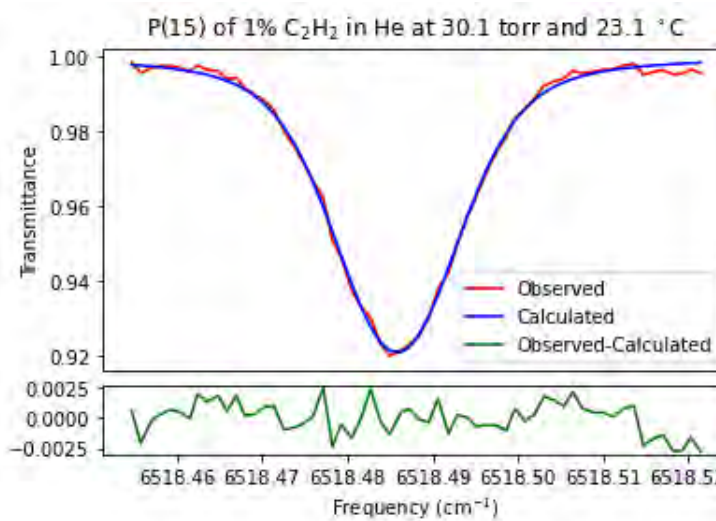


**Figure 4:** Absorption data, fit and residual for the P(15) line of the  $\nu_1 + \nu_3$  band of C<sub>2</sub>H<sub>2</sub> at 14.3 torr and 22.5 °C.

the wavelength absorbed. In our 1% C<sub>2</sub>H<sub>2</sub> in He sample, He provides a broadening effect without increasing intensity. Pressure and inert gas broadening is



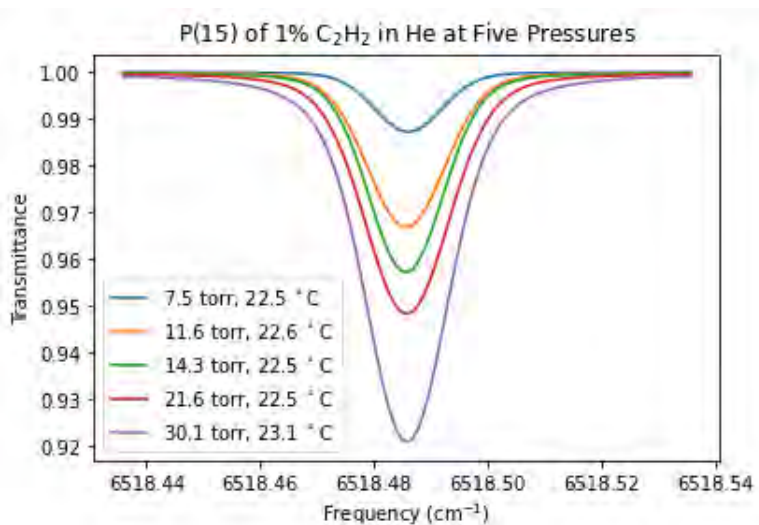
**Figure 5:** Absorption data, fit and residual for the P(15) line of the  $\nu_1 + \nu_3$  band of C<sub>2</sub>H<sub>2</sub> at 21.6 torr and 22.5 °C.



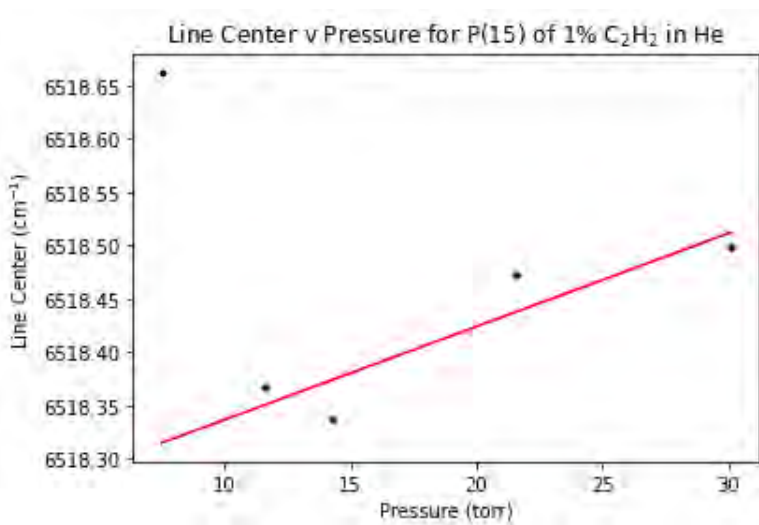
**Figure 6:** Absorption data, fit and residual for the P(15) line of the  $\nu_1 + \nu_3$  band of C<sub>2</sub>H<sub>2</sub> at 30.1 torr and 23.1 °C.

described by the following equation:

$$\gamma(p, T) = \left(\frac{T_{ref}}{T}\right)^{n_{He}} (\gamma_{He}(p_{ref}, T_{ref})(p - p_{self}) + \gamma_{self}(p_{ref}, T_{ref})p_{self}) \quad (1)$$



**Figure 7:** The fitted lines of each of the five pressures recorded overlain.



**Figure 8:** The shift in line centers for each of the five pressures recorded.

where  $\gamma(p, T)$  is the Lorentzian (pressure broadened) Half Width at Half Maximum (HWHM) at pressure  $p$  (atm) and temperature  $T$  (K),  $\gamma_{He}(p_{ref}, T_{ref})$ , and  $\gamma_{self}(p_{ref}, T_{ref})$  are the He-broadened HWHM and the self-broadened HWHM at 1 atm and 296 K, respectively,  $p_{self}$  is the partial pressure of the analyte gas,



and  $n_{He}$  is the coefficient of temperature dependence of  $\gamma_{He}$ .  $\gamma_{He}(p_{ref}, T_{ref})$  was obtained from Arteaga et al. [5]. We see in figure 7 that our data is what we expect from pressure broadening; as we increase pressure, line intensity increases and the line shape becomes broader. This shows that our spectrometer is recording realistic data. The temperature dependence is included in this equation 1, but we neglect this in our fitting as we recorded near 296 K. The effect of a higher temperature is of course to increase the speed of gas molecules. It then becomes more likely that a molecule will be moving at a higher speed towards or away from an incoming photon, which shifts its frequency in accordance with the Doppler Effect and leads to a wider range of observed frequencies. Line center,  $\nu$ , can also be affected by pressure. As pressure increases and more molecules collide with each other at faster speeds, the potential energy of transitions is shifted slightly upwards. This pressure shift, denoted  $\delta$ , affects line center by the following relationship:

$$\nu^* = \nu + \delta(p_{ref})p \quad (2)$$

We see this effect, though very small, in figure 8. We see that, overall, line center increases with pressure in He. Of course, more data and recorded lines in the future would help to confirm this. The fact that we have an outlier at our lowest pressure makes some sense. For one thing, the pressure gauge we use has a percentage error of the total reading, 2%, meaning the lower the pressure, the higher the error. Additionally, the lower line intensity means that there is a higher proportion of both frequency and intensity errors to line width, leading to lower resolution.

The accepted value of the Boltzmann constant to 3 significant figures is  $1.38 \cdot 10^{-23}$  J/K. Obtained values from our fitted lines all agree within uncertainty. Our uncertainties were quite large, though, likely due to inefficiencies in our fitter program which gives only a first approximation of parameter errors.

This may be acceptable for an undergraduate experiment, but these errors probably indicate that there is some fine tuning to be done in both the experiment and the fitting program. That said, these values are at least physically realistic.

With the data and fits we have recorded, we have demonstrated the functionality of our spectrometer. In its present state, it could be used in a teaching experiment. However, there are still several areas in which improvement could be made. One of the most noticeable omissions in this report is a lack of comprehensive numerical error data. This is due to the complexity of the myriad sources of error present, which can only all be accounted for properly by performing Monte-Carlo Simulations, which we were not able to do as of the writing of this report. However, we can identify potential sources of error and consider the extent to which they are relevant.

One of the biggest potential oversights that we realized after beginning to take data was the difference in the path length in air after the beam is split equally at the second GT prism. The path length in air for the reference detector was about 29 cm, while the path length for the signal detector was about 17 cm. This can matter when there are nearby absorption lines for molecules present in the air, namely water. If the path lengths in air after the split were even, then the effects of a water line would be eliminated because the effect would be equally applied to both detectors and cancelled out in the ratio. However, the longer path length to our reference detector in our case would mean a greater absorption of a water line showing in our data. In fact, there is a water line somewhat nearby P(15) of  $C_2H_2$ . The P(15) line is centered at  $6518.4858\text{ cm}^{-1}$ , and there is a nearby water line at  $6518.4360\text{ cm}^{-1}$ , though its intensity is about a factor of  $10^9$  less than P(15). However, the water vapor in the air is at a much higher pressure than the  $C_2H_2$  in the cell, so it could be significant. Luckily, this can usually be resolved by subtraction of a quadratic from the baseline,

which we did for several of our recorded pressures. Ideally, the air path lengths to each detector should be as even as possible.

We previously mentioned that the error in our pressure gauge was a flat percentage, specifically 2%. Also for consideration is the precision of this pressure gauge, which is only 0.1 torr. This precision makes it difficult to detect small leaks in the cell, which could alter data slightly as the spectrometer is scanning. We neglect temperature deviations in our fitting algorithm, but we may want to consider this in the future as the room in which we conduct this experiment can drift up to 0.5 °C over an hour. Temperature drift can change the frequency of the laser, which can alter the polarization. Our GT prisms, responsible for splitting the beam evenly to signal and reference, are sensitive to polarization changes, and any change will cause a change to the baseline we record. In normal operation, the frequency changes regularly, and thus polarization changes are linear and predictable. Temperature drifts could introduce additional baseline errors.

## 5 Impact on DOE Missions

The ultimate practical intention of this experimental design is to encourage the hastening of high-resolution spectroscopic data collection. With strong databases of high order line parameters, we will be able to come to more accurate conclusions about remote measurements of all kinds. This will undoubtedly lead to a vastly improved knowledge of the composition of our own atmosphere, which is changing at a faster rate than at any time in recent history. It may also lead to advances in meteorology. Monitoring changing concentrations of greenhouse gases in our atmosphere will be crucial to manage the environmental catastrophes that climate change will bring in the coming decades. This is directly in line with the DOE's core mission, "to ensure America's security

and prosperity by addressing its energy, environmental and nuclear challenges through transformative science and technology solutions.”

## **6 Acknowledgements**

I would like to thank my P.I. Trevor Sears and his graduate student Eisen Gross for their invaluable and enthusiastic help on this project. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## References

- [1] M. A. Cordiner, N. A. Teanby, C. A. Nixon, V. Vuitton, A. E. Thelen, and S. B. Charnley. Alma spectral imaging of titan contemporaneous with cassini's grand finale. *The Astronomical Journal*, 158(2):76, Jul 2019.
- [2] I.E. Gordon, L.S. Rothman, C. Hill, R.V. Kochanov, Y. Tan, P.F. Bernath, M. Birk, V. Boudon, A. Campargue, K.V. Chance, B.J. Drouin, J.-M. Flaud, R.R. Gamache, J.T. Hodges, D. Jacquemart, V.I. Perevalov, A. Perrin, K.P. Shine, M.-A.H. Smith, J. Tennyson, G.C. Toon, H. Tran, V.G. Tyuterev, A. Barbe, A.G. Császár, V.M. Devi, T. Furtenbacher, J.J. Harrison, J.-M. Hartmann, A. Jolly, T.J. Johnson, T. Karman, I. Kleiner, A.A. Kyuberis, J. Loos, O.M. Lyulin, S.T. Massie, S.N. Mikhailenko, N. Moazzen-Ahmadi, H.S.P. Müller, O.V. Naumenko, A.V. Nikitin, O.L. Polyansky, M. Rey, M. Rotger, S.W. Sharpe, K. Sung, E. Starikova, S.A. Tashkun, J. Vander Auwera, G. Wagner, J. Wilzewski, P. Wcisło, S. Yu, and E.J. Zak. The hitran2016 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 203:3–69, 2017. HITRAN2016 Special Issue.
- [3] N.H. Ngo, D. Lisak, H. Tran, and J.-M. Hartmann. An isolated line-shape model to go beyond the voigt profile in spectroscopic databases and radiative transfer codes. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 129:89–100, 2013.
- [4] J.J. Fortney, T.D. Robinson, S. Domagal-Goldman, D.S. Amundsen, M. Brogi, M. Claire, D. Crisp, E. Hebrard, H. Imanaka, R. de Kok, et al. The need for laboratory work to aid in the understanding of exoplanetary atmospheres. *arXiv preprint arXiv:1602.06305*, 2016.
- [5] S.W. Arteaga, C.M. Bejger, J.L. Gerecke, J.L. Hardwick, Z.T. Martin,

- J. Mayo, E.A. McIlhattan, J.-M.F. Moreau, M.J. Pilkenton, M.J. Polston, B.T. Robertson, and E.N. Wolf. Line broadening and shift coefficients of acetylene at 1550nm. *Journal of Molecular Spectroscopy*, 243(2):253–266, 2007. PRAHA2006, The 19th International Conference on High Resolution Molecular Spectroscopy.
- [6] E. Fasci, T. A. Odintsova, A. Castrillo, M. D. De Vizia, A. Merlone, F. Bertiglia, L. Moretti, and L. Gianfrani. Dual-laser absorption spectroscopy of  $\text{c}_2\text{h}_2$  at  $1.4\mu\text{m}$ . *Phys. Rev. A*, 93:042513, Apr 2016.
- [7] C. McRaven, M. Cich, Gary Lopez, Trevor Sears, Daniel Hurtmans, and Arlan Mantz. Frequency comb-referenced measurements of self- and nitrogen-broadening in the  $\nu_1 + \nu_3$  band of acetylene. *Journal of Molecular Spectroscopy - J MOL SPECTROSC*, 266:43–51, 03 2011.

Analysis of precipitable water vapor retrievals for the TRACER campaign

Marcie Hogan

Applied Physics and Mathematics, State University of New York at Geneseo, Geneseo, NY  
14454

Dié Wang and Michael Jensen

Environmental and Climate Sciences Department, Brookhaven National Laboratory, Upton, NY  
11973

Grace Kowalski

Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette,  
Indiana 47907

## **Abstract**

The Tracking Aerosol Convection Interactions Experiment (TRACER) is an upcoming campaign to study deep convection cloud-aerosol interactions and the variability of atmospheric phenomena over the Houston metropolitan area. To investigate the environmental conditions in the region we analyzed the characteristics of precipitable water vapor (PWV), which represents the depth of water in a column of the atmosphere if all its vapor is condensed. 10 years of daily PWV retrievals from HoustonNet Global Positioning System (GPS) Network sites are used in order to analyze PWV variability at different locations in Houston. Sixteen different weather regimes—periods of similar meteorological conditions over Houston in the summertime months, previously identified through machine learning technique, Self-Organizing Maps—were utilized to organize and better understand trends in PWV. Emphasis has been placed on the difference in PWV statistics for sites at varying distances from the coast and under different meteorological conditions. Upon examination of plots created using Python scripts, it has been noted that PWV follows a diurnal cycle, peaking in the latter half of the day, and is generally higher in regimes representative of anti-cyclonic conditions. Inland sites were found to have more consistent diurnal cycles while the coastal site had greater variability across most regimes. Furthermore, days with radar observations confirmed convection have been found to have a stronger diurnal cycle versus no-precipitation days.



## **I. Background**

### **A. TRACER Campaign**

The TRACER campaign is set to begin on October 1<sup>st</sup>, 2021, and will run until September 30<sup>th</sup>, 2022. During this time, a comprehensive suite of instrumentations will be deployed to Houston and the greater-Houston region at three primary locations: the Atmospheric Radiation Measurement (ARM) Mobile Facility in La Porte, TX on the northwestern edge of the Galveston Bay, the C-Band Scanning ARM Precipitation Radar located approximately 30 km to the south-southwest of LaPorte, and an ancillary site in a rural area southwest of Houston that will be added during the summer of 2022 <sup>1</sup>. La Porte and the rural area were selected due to high air pollution and lower air pollution, respectively. The TRACER campaign will acquire data for the duration of its period and the data will subsequently be analyzed to grasp a better hold on Houston cloud-aerosol interactions. Houston has been selected as the location of the campaign due to its unique status as being a major metropolitan area, with an abundance of human-generated pollution, located right along the Gulf of Mexico, which lends itself to frequent isolated convective systems <sup>1</sup>. Through analysis of atmospheric variables—one such example being precipitable water vapor (PWV)—from previously recorded meteorological data, we will better understand convective patterns in the Houston region in hopes to lessen the strain of post-TRACER analysis.

### **B. PWV from Global Positioning System measurements**

PWV is one of many variables being studied in preparation for the TRACER campaign and is the focus of this paper. PWV represents the depth of water in a column of the atmospheric if all its vapor is condensed and is typically associated with precipitation <sup>4</sup>. In short, it can be

thought of as a measure of total atmospheric humidity. PWV is typically measured in millimeters, and readings can be likened to a hypothetical rain gauge.

PWV has long been measured using atmospheric soundings based on radiosondes carried by weather balloons. Radiosonde-based PWV measurements are costly and launching stations are typically several hundred miles apart—a characteristic which would be problematic for looking at atmospheric data over a small region, such as Houston. Globally, the radiosonde technique for PWV data acquisition is increasingly becoming replaced by ground-based GPS networks<sup>5</sup>. GPS can be used to measure PVW as the presence of atmospheric water vapor affects the amount of refractivity that GPS signals encounter, allowing for measurements based on satellite-receiver path delays. Unlike radiosonde observations, which are labor intensive and only performed a few times per day requires fair weather, GPS sites can obtain continuous PWV data regardless of atmospheric conditions—making the HoustonNet GPS network the optimal source of PWV data for the TRACER campaign<sup>5</sup>.

### C. Purpose

In analyzing Houston PWV measurements, this paper aims to discover how PWV differs under a variety of categories—weather pattern, coastal proximity, and time of day as well as validate previously identified PWV-related phenomena. Through statistical analysis via Python plotting and ArcGIS Pro, data visualization with both plots and maps will be utilized to identify historical patterns in the cyclic nature of PWV which can be useful in future TRACER-related analysis when aerosols are factored into play.

## II. Data & Methods

### A. Scope of data

Ten years (2010-2019) of daily PWV measurements recorded every five minutes, from the summer months of June through September where isolated convection is most common, are being used in this study. While the HoustonNet GPS Network encompasses numerous sites from which PWV measurements have been databased, this study has identified three sites of peak interest—TDAM, a coastal site located in Galveston, NASA, an inland site located near La Porte, and UH01, an urban site in the heart of Houston. The sites are in a perpendicular line from the coast and are relatively evenly spaced. The selection of TDAM, NASA, and UH01 came from the desire to understand how PWV may differ at a coastal location seeing the greatest sea breeze influence, an urban site where industrial pollution is most prevalent, and a site between the two.

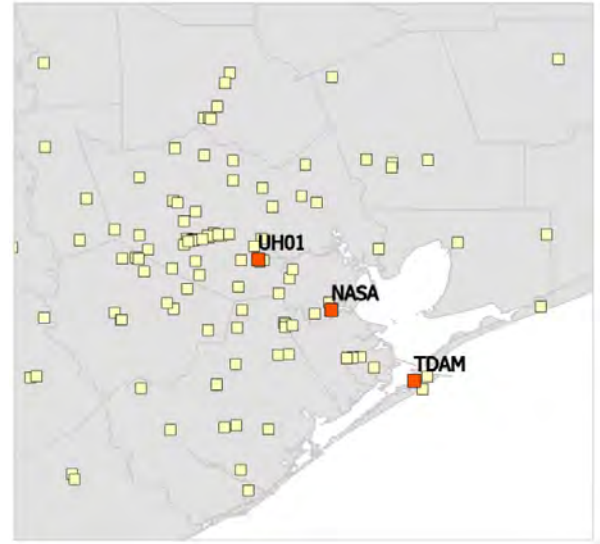


Figure 1: The three focal GPS sites symbolized in red and labeled, unused HoustonNet GPS sites in yellow

#### B. Self-Organizing maps as a means of weather classification

The overarching classification method of PWV data for the duration of this paper is rooted in previously identified groupings of weather patterns, namely regimes. Using the machine-learning technique of self-organizing maps, each day in June through September for each year was identified to be in one of 16 groupings of weather patterns, referred to as SOM nodes. Nodes 0,4,8, and 12 are classified as days with strong anticyclonic features influenced by the Bermuda High. These nodes generally see higher temperatures and more precipitation. Nodes 3,7,11, and 15 are dominated by post-trough characteristics and are cooler and drier.

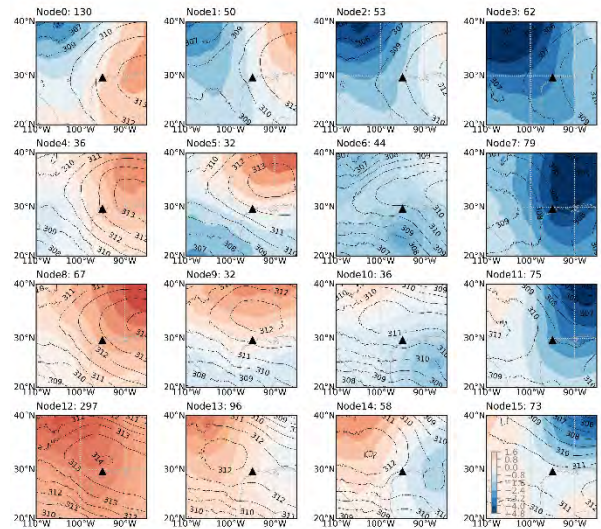


Figure 2: 700hpa geopotential anomalies of the 16 atmospheric nodes

All other nodes have been identified as “transitional regimes” and are influenced by more than one large-scale weather phenomenon <sup>2</sup>.

### C. Python scripting as means for analysis

To create meaningful imagery, a primary Python script was written to read in GPS PWV files for a given regime and specified grouping of GPS sites. The script outputted a CSV file containing data on site name, date, time, and corresponding PWV measurement. CSV files for all 16 regimes were generated. These files were subsequently read into other novel scripts which created plots using the Matplotlib, Pandas, and NumPy libraries. Line plots and box plots made up most of the created imagery as they best allowed for the comparison of time series and subplots were commonly used for means of direct comparison between certain distinguishing features.

## III. Results

A. Different weather patterns affect the cyclic nature of PWV

Perhaps the most important result was that PWV behaved differently under the different large-scale weather conditions found in each regime. When plotting mean PWV versus UTC hour for each regime, it was found that regimes dominated by anticyclonic features (nodes 0,1,4,8,12) showed the strongest and most consistent diurnal cycle of PWV across all three sites.

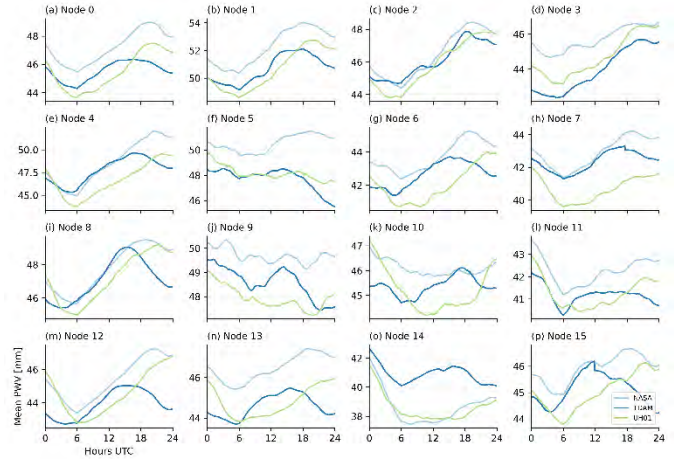


Figure 3: Subplot of mean PWV versus UTC hour for all 16 SOM nodes

Regime 1 had the highest overall PWV, with the NASA site reaching a maximum mean value of over 54 millimeters, a strong outlier in comparison to others. When looking at previously identified

Node	Large-scale Convection	Sea Breeze Convection	Little Precipitation	No Precipitation	Clear Skies
0	36.5	3.1	20.8	36.5	3.1
1	52.6	13.2	18.4	15.8	0.0
2	41.5	4.9	24.3	26.8	2.4
3	44.8	1.7	27.6	22.4	3.4
4	33.3	24.2	9.1	33.3	0.0
5	30.7	3.9	30.8	34.6	0.0
6	31.7	0.0	21.9	46.3	0.0
7	25.7	1.4	15.7	44.3	12.9
8	35.4	25	10.4	22.9	6.3
9	38.7	19.3	19.3	19.35	3.2
10	35.5	3.2	25.8	29	6.4
11	27.3	3.0	9.1	43.9	16.7
12	23.2	20.3	11.6	38.6	6.3
13	32.4	14.9	13.5	35.1	4.0
14	25.0	3.8	15.4	42.3	13.5
15	23.4	14.1	17.2	34.4	10.9

Figure 4: Previously-made table containing data (in percentages) of observed weather type seen in each regime, classified through radar imaging. Colored boxes represent highest values.

weather pattern classifications of radar imagery for each regime, Regime 1 was found to have the highest percentage of large-scale convection days, 52.6%.

Across all the anti-cyclonic regimes, coastal site TDAM had lower PWV and a cycle which peaked earlier, with the exception being regime 8. Regime 8 was identified to have the highest percentage of sea breeze convection days, with 25%, which can explain why TDAM has higher PWV, as convection has a chief influence on PWV.

In contrast, regimes 11 and 14 were outliers with regards to low PWV and an atypical diurnal cycle, peaking much later at 24 UTC. This is likely due to the cooler, dry air in these regimes caused by post-trough conditions. Regimes 11 and 14 recorded high percentages of clear skies days, with 43.9% and 42.3% respectively, which is consistent with low PWV.

### B. Isolating data by convection type

In addition, PWV data was sorted using radar observations to better understand how the daily PWV cycle differs under varying levels of convection.

Lists were created of all days identified through visual inspection of radar imagery

to have large scale, sea breeze, or no convection, and their data was subsequently plotted. Only days from regime 12 were considered, as regime 12 had the greatest number of days, allowing for a more robust average.

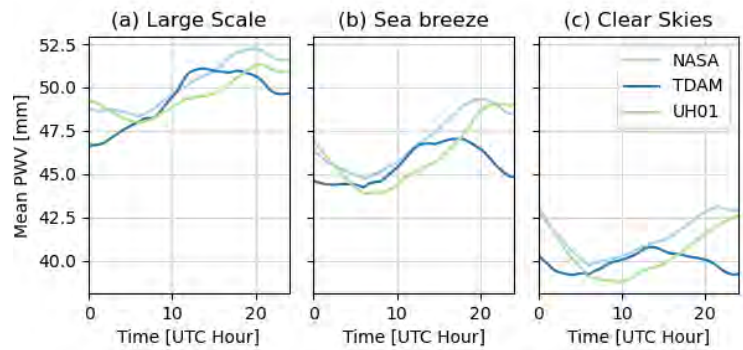


Figure 5: Subplot of mean PWV versus UTC hour for identified large scale, sea breeze, and clear sky days

Figure 4 shows that on days when large scale precipitation is present, PWV tends to be higher. Sea breeze days also lend themselves to higher PWV than clear skies days, however we see a significantly lower maximum PWV value across all sites than large scale days. Both large scale and sea breeze conditions cause a strong diurnal cycle which sees inland sites peaking at roughly 20UTC and the coastal site peaking earlier around 12UTC. Lower temperature and early onset of sea breeze are likely factors for earlier peaking. The same phenomenon is seen on clear sky days as well, although inland sites see a deviation from the typical diurnal cycle and instead peak at 24UTC—there is no convection to drive the PWV peak earlier and higher.



### C. Understanding variability across regimes

Although PWV does converge to a given value across each regime, there are many instances where values have been recorded far from their average. A grouped boxplot was created to understand the level of variability seen in total precipitable water vapor variability seen at each regime for all three sites.

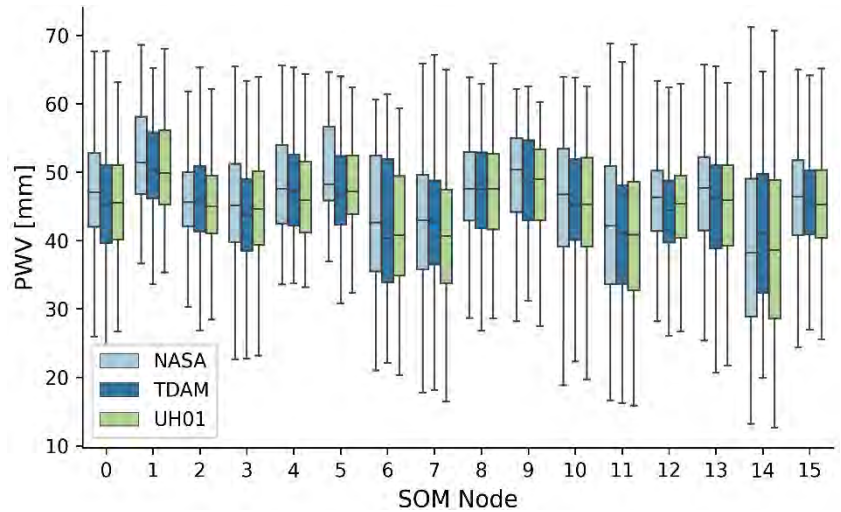


Figure 6: Grouped boxplot of PWV versus SOM node

Post-trough regimes 7, 11, and 14 saw both the lowest median PWV and the highest variability, with a greater than fifty-millimeter difference between the high and low ends of their whiskers. These regimes were earlier noted to have strayed the furthest from a typical diurnal cycle and saw a greater number of clear skies days than the other regimes on average; they can be looked at as regimes in which weather patterns are the least predictable. Furthermore, it should be noted that proximity to the coast had little effect on the overall variability in PWV values across all regimes.

### D. Precipitation onset and PWV spikes

A previous study by Sapucci et. al. found that PWV spikes immediately before intense rainfall and will subsequently drop

rapidly. We sought to confirm this with Houston data. Radar imagery was used to identify days

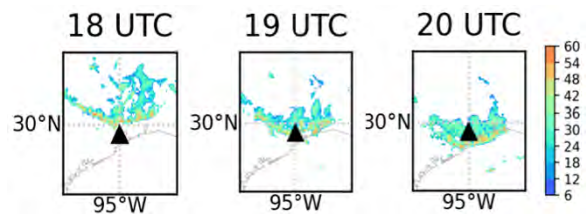


Figure 7: Subplot of radar reflectivity over the Houston area from 18-20 UTC for 8/25/2015; the front hits Houston at roughly 19UTC

with large scale convection and plots were generated to show the radar reflectivity for the given period where a front begins to pass over the Houston epicenter. The radar imaging was then compared to the corresponding PWV time series for that day to assess whether the phenomenon described by Sapucci was present in our data <sup>3</sup>.

Many days which showed strong, large-scale convection passing over the Houston region were observed to have a sharp decline in PWV occurring at the same time which convection began. which was consistent with what Sapucci describes. On days with less precipitation, such as sea breeze, a small peak and subsequent decline

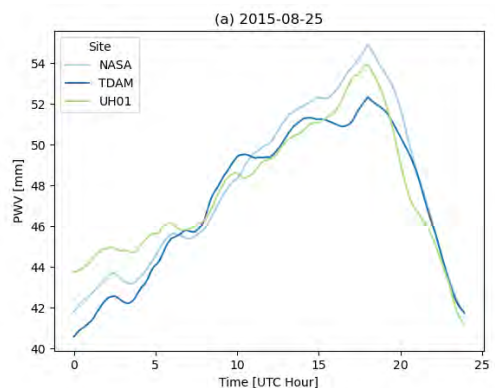


Figure 8: Time series of PWV for 8/25/2015, showing a peak at 18UTC

were noted, although they were not as prominent as those seen on large scale convection days. The ability to use PWV data as a means for precipitation prediction for the TRACER campaign seems viable, as a rapid increase in PWV has been noted as fundamental for intense rainfall <sup>3</sup>.

#### IV. Discussion

PWV has been found to peak earlier in the day in anti-cyclonic regimes than post-trough regimes. This can be attributed to the lower amount of convection seen in post-trough regimes, as increased PWV is correlated largely with precipitation. Furthermore, coastal Houston, regardless of weather pattern, sees PWV peaking lower and earlier in the day, likely due to proximity to cooler air over the water as well as being the first land affected by sea breeze. With regards to variability, greater PWV variability is seen in post-trough conditions, potentially meaning weather patterns may not be as predictable in identified post-trough regimes. As for differences in convection, it should be noted that large scale and sea breeze convection follow a



similar PWV cycle across all three sites, as the presence of precipitation is a likely driving factor in their daily PWV time series.

The findings from this research will serve to inform scientists of prior atmospheric conditions in the Houston area with regards to PWV. The trends identified in this historical PWV analysis can be compared to future data collected in the forthcoming TRACER campaign, where aerosols may begin to shed some light on new PWV patterns.

## V. References

<sup>1</sup>Jensen et al., 2021 <https://www.arm.gov/research/campaigns/amf2021tracer>

<sup>2</sup>Wang et al., 2021 <https://asr.science.energy.gov/meetings/stm/posters/abstract/2689>

<sup>3</sup>Sapucci, LF, Machado, LAT, de Souza, EM, Campos, TB. Global Positioning System precipitable water vapour (GPS-PWV) jumps before intense rain events: A potential application to nowcasting. *Meteorol Appl.* 2019; 26: 49– 63. <https://doi.org/10.1002/met.1735>

<sup>4</sup>Precipitable water. (2020, October 16). In *Wikipedia*.  
[https://en.wikipedia.org/wiki/Precipitable\\_water](https://en.wikipedia.org/wiki/Precipitable_water)

<sup>5</sup>Campos-Arias, P. (2019). GPS Precipitable Water Vapor Estimations over Costa Rica: A Comparison against Atmospheric Sounding and Moderate Resolution Imaging Spectrometer (MODIS).



# Charge sharing in pixelated sensors

Andrea Houck, Department of Natural Sciences, Lawrence Technological University, Southfield,  
MI 48075

Ivan Kotov, Instrumentation Division, Brookhaven National Laboratory, Upton NY 11973

## Abstract

The National Synchrotron Light Source II (NSLS II) Soft Inelastic X-ray Scattering (SIX) beam line at Brookhaven National Laboratory studies electronic excitations in materials using resonant inelastic x-ray scattering. Scattered x-rays are detected by an Electron Multiplying Charged Coupled Device (EMCCD). The x-ray interacts with the silicon lattice of the EMCCD and creates a number of electron hole pairs that is proportional to the x-ray energy. The charges diffuse in the field free region that is  $8\mu m$  thick and collect on EMCCD pixels, which are  $16\mu m$  squares. The number of electrons in each pixel provides information about the distribution of charges, which helps us measure x-ray coordinates with accuracy better than the pixel pitch. Our project investigates radial distribution of charges taking into consideration pixels square shape. We focus on the diffusion of charges in the field-free region and the radial distribution in the EMCCD. We also investigate the effect of initial conditions on the charge distribution by finding the solution to the diffusion equation for both a Delta function and a Gaussian function initial condition. We write an equation for the radial spreading of charge across the EMCCD pixels. Using this equation and numerically integrating over each pixel, we create a Lego plot of the charge distribution over the pixels. Our study aims to provide a predictive function for determining what distribution of charges should be observed in the EMCCD pixels.

## I. Introduction

A charged coupled device (CCD) is composed of arrays of metal-oxide-silicon structures (MOS) and measures energy by detecting incoming photons from the x-rays. A photon enters the CCD and is converted into electron-hole pairs. The electron travels first through the field-free region, then the depletion layer. CCDs have pixels, which are essentially potential wells that collect electrons. By looking at the number of electrons in each pixel, images can be created. An electron multiplying CCD (EMCCD) can be used in place of a CCD to reduce the readout noise by amplifying the signal prior to the signal being output from the device.<sup>1</sup> This is especially useful for low signal levels because the signal to noise ratio is smaller, so noise reduction has more impact.

In the SIX beam line, the CCDs have a field-free region  $8\mu m$  thick and a depletion layer  $6-8\mu m$  thick. The pixels are  $16\mu m$  squares.<sup>2</sup> Charge moves through the CCD by drift or diffusion, depending on if an electric field is present. In the field-free region, the charge moves by diffusion whereas there is also drift to consider in the depletion layer. We investigate the diffusion in the field-free region.

The charge distribution is largely dominated by the boundary conditions and the growth of the electron cloud from diffusion. The initial electron cloud size also contributes, but is not a key factor. Initial conditions of the diffusion equation give us the electron cloud size and shape. We incorporate the specifications required at the boundaries using boundary conditions when solving the diffusion equation.

## II. Diffusion equation

The diffusion equation solution depends on the initial conditions of the charge distribution and the conditions at the boundaries. The diffusion equation is

$$D\nabla^2\rho(\mathbf{r}, z, t) = \frac{\partial\rho(r, z, t)}{\partial t} \quad (1)$$

where  $D$  is the diffusion coefficient,  $\nabla^2$  is the Laplacian,  $\rho(\mathbf{r}, z, t)$  is the charge density,  $\mathbf{r}$  is the radial vector in cylindrical coordinates, and  $z$  is the depth perpendicular to the EMCCD surface.<sup>3</sup> We equate  $\rho(\mathbf{r}, z, t) = \rho(\mathbf{x}, t)$  because  $\mathbf{r}$ , the radius in spherical coordinates, has an  $x$ ,  $y$ , and  $z$  position associated with it so  $\mathbf{r}$  and  $z$  can be written in terms of  $\mathbf{x}$ .

## A. General solution

The diffusion equation is separable, so we have the form

$$\rho(\mathbf{x}, t) = X(x)Y(y)Z(z)T(t) \quad (2)$$

Following Ref. [3] and solving the equation, we get

$$D \left( \frac{\partial^2 \rho}{\partial x^2} + \frac{\partial^2 \rho}{\partial y^2} + \frac{\partial^2 \rho}{\partial z^2} \right) = \frac{\partial \rho}{\partial t} \quad (3)$$

$$\frac{1}{DT} \frac{\partial T}{\partial t} = \frac{1}{X} \frac{\partial^2 X}{\partial x^2} + \frac{1}{Y} \frac{\partial^2 Y}{\partial y^2} + \frac{1}{Z} \frac{\partial^2 Z}{\partial z^2} \quad (4)$$

which we split into independent equations

$$\begin{aligned} \frac{1}{DT} \frac{\partial T}{\partial t} &= - (m_1^2 + m_2^2 + m_3^2) \\ \frac{d^2 X}{dx^2} + m_1^2 X &= 0 \\ \frac{d^2 Y}{dy^2} + m_2^2 Y &= 0 \\ \frac{d^2 Z}{dz^2} + m_3^2 Z &= 0 \end{aligned} \quad (5)$$

where  $m_1, m_2$  and  $m_3$  are constants. All of these can be referenced by using the notation  $m_i$  where  $i = 1, 2, 3$ . Eq. (5) can be solved as an eigenvalue problem to find the general solutions given in Eqs. (6), (7), and (8).  $Q_j$  in Eqs. (7) and (8) is the solution for the spatial components, but it represents three different solutions since there are the three coordinate components equations in Eq.

(5). Euler's formula allows us to write the spatial solution as either exponential or trigonometric.

$$T(t) \propto e^{-\sum m_i^2 D t} \quad (6)$$

$$Q_j(x_j) \propto a_j e^{i m_j x_j} + b_j e^{-i m_j x_j} \quad (7)$$

$$Q_j(x_j) \propto a_j \cos(m_j x_j) + b_j \sin(m_j x_j) \quad (8)$$

Ref. [3] then uses the Fourier transform, which is defined as

$$\rho(\mathbf{x}, t) = \int_{-\infty}^{\infty} \rho(\mathbf{k}, t) e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \quad (9)$$

We then take the derivative to write the right-hand side of Eq. (1) as

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} = \int_{-\infty}^{\infty} \frac{\partial \rho(\mathbf{k}, t)}{\partial t} e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \quad (10)$$

Multiplying the transform by  $D$  and taking the Laplacian, we get the left-hand side of Eq. (1) to be

$$D \nabla^2 \rho(\mathbf{x}, t) = \int_{-\infty}^{\infty} -D \rho(\mathbf{k}, t) (\mathbf{k} \cdot \mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \quad (11)$$

where  $\rho(\mathbf{k}, t)$  can be written as

$$\rho(\mathbf{k}, t) = \rho(\mathbf{k}, 0) e^{-D k^2 t} \quad (12)$$

and

$$\rho(\mathbf{k}, 0) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \rho(\mathbf{x}, 0) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x} \quad (13)$$

and  $\rho(\mathbf{x}, 0)$  is the initial condition.

## B. Initial conditions

The initial condition is often taken to be a delta function.<sup>3,4</sup> This is not a perfect representation because it assumes all the charge is within one single point. This is not true because when the

x-ray interacts with the silicon in the lattice structure, photoelectrons are knocked out. These photoelectrons have range, and they ionize other silicon atoms in the lattice, creating a cloud of initial charges. Since the initial cloud is very small, the delta function is a good simplification for the initial condition, but the next step is to give the charge cloud a size. Therefore, we explore other options for the initial condition to solve the diffusion equation, noting that the chosen function must satisfy certain conditions. The function must describe a bell-shaped curve and converge sufficiently fast in an infinite sum. First, we look at the solution using the delta function as an initial condition, then we use the Gaussian function as an initial condition.

## 1. Delta function

Using the initial condition  $\rho(\mathbf{x}, 0) = \delta(\mathbf{x} - \mathbf{x}_0)$  in Eq. (12), we get

$$\rho(\mathbf{k}, 0) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \delta(\mathbf{x} - \mathbf{x}_0) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x} \quad (14)$$

$$= \frac{1}{(2\pi)^n} e^{i\mathbf{k}\cdot\mathbf{x}_0} \quad (15)$$

After doing substitutions to return to an equation for  $\rho(\mathbf{x}, t)$ , writing the equation in terms of vector components, and simplifying, we get Eq. (16).<sup>3</sup>

$$\rho(\mathbf{x}, t) = \frac{1}{(4\pi Dt)^{\frac{n}{2}}} e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{4Dt}} \quad (16)$$

## 2. Gaussian function

The Gaussian function follows a bell-shape curve. The normalized Gaussian function where  $x_0$  is the position of the peak's center and  $\sigma_0$  is the standard deviation is represented by Eq. (17).

$$\rho(x, 0) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-x_0)^2}{2\sigma_0^2}} \quad (17)$$



The one-dimensional Gaussian function is sufficient even though the diffusion equation is three-dimensional because separation of variables allows us to split the diffusion equation into four different equations that are summed together. This sum allows us to still get a three-dimensional solution because we use the one-dimensional equation for each one-dimensional component.

Plugging in the Gaussian function as the initial condition in Eq. (12) but in one dimension, we get

$$\rho(k, 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-x_0)^2}{2\sigma_0^2}} e^{ik \cdot x} dx \quad (18)$$

Following the process outlined in Section II.A., we plug Eq. (18) into Eq. (12).

$$\rho(k, t) = \frac{1}{2\pi} e^{-Dk^2 t} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-x_0)^2}{2\sigma_0^2}} e^{ik \cdot x} dx \quad (19)$$

$$= \frac{1}{2\pi} e^{-Dk^2 t} \frac{1}{\sqrt{2\pi}\sigma_0} \int_{-\infty}^{\infty} e^{-\frac{(x-x_0)^2}{2\sigma_0^2} + ik \cdot x} dx \quad (20)$$

For simplicity, we solve the integral independently, then plug the solution back into Eq. (20). We start by completing the square for the terms in the exponent.

$$-\frac{(x-x_0)^2}{2\sigma_0^2} + ik \cdot x = -\frac{1}{2\sigma_0^2} (x^2 - 2x_0x + x_0^2 - 2\sigma_0^2 ik \cdot x) \quad (21)$$

$$= -\frac{1}{2\sigma_0^2} (x^2 - 2x(x_0 + ik\sigma_0^2) + x_0^2) \quad (22)$$

$$= -\frac{1}{2\sigma_0^2} (x^2 - 2x(x_0 + ik\sigma_0^2) + x_0^2 + (x_0 + ik\sigma_0^2)^2 - (x_0 + ik\sigma_0^2)^2) \quad (23)$$

$$= -\frac{1}{2\sigma_0^2} (x - (x_0 + ik\sigma_0^2))^2 - \frac{1}{2\sigma_0^2} (x_0^2 - x_0^2 - 2ikx_0\sigma_0^2 + k^2\sigma_0^4) \quad (24)$$

$$= -u^2 - \frac{1}{2\sigma_0^2} (-2ikx_0\sigma_0^2 + k^2\sigma_0^4) \quad (25)$$

$$= -u^2 + ikx_0 - \frac{k^2\sigma_0^2}{2} \quad (26)$$

We use the substitution for  $u$  given in Eq. (27). When plugging the exponent back into the integral, we use the substitution for  $dx$  given in Eq. (29).

$$u = \frac{x + x_0 - ik\sigma_0^2}{\sqrt{2}\sigma_0} \quad (27)$$

$$du = \frac{1}{\sqrt{2}\sigma_0} dx \quad (28)$$

$$dx = \sqrt{2}\sigma_0 du \quad (29)$$

This gives us the integral

$$\int_{-\infty}^{\infty} e^{-\frac{(x-x_0)^2}{2\sigma_0^2} + ik \cdot x} dx = \sqrt{2}\sigma_0 e^{ix_0k - \frac{k^2\sigma_0^2}{2}} \int_{-\infty}^{\infty} e^{-u^2} du \quad (30)$$

Note that  $\int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi}$ . This comes from the error function definition and simplifies our calculation. Using this in Eq. (30), we get

$$\sqrt{2}\sigma_0 e^{ix_0k - \frac{k^2\sigma_0^2}{2}} \int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{2\pi} \sigma_0 e^{ix_0k - \frac{k^2\sigma_0^2}{2}} \quad (31)$$

Now, we plug Eq. (31) back into Eq. (20) in place of the integral and simplify.

$$\rho(k, t) = \frac{1}{2\pi} e^{-Dk^2t} \frac{1}{\sqrt{2\pi}\sigma_0} \sqrt{2\pi} \sigma_0 e^{ix_0k - \frac{k^2\sigma_0^2}{2}} \quad (32)$$

$$= \frac{1}{2\pi} e^{-Dk^2t + ix_0k - \frac{k^2\sigma_0^2}{2}} \quad (33)$$

Completing the Fourier Transform, we plug Eq. (33) into Eq. (9).

$$\rho(x, t) = \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-Dk^2t + ix_0k - \frac{k^2\sigma_0^2}{2}} e^{-ik \cdot x} dk \quad (34)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-Dk^2t - \frac{k^2\sigma_0^2}{2} + ix_0k - ik \cdot x} dk \quad (35)$$

$$\rho(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\left[\left(Dt + \frac{\sigma_0^2}{2}\right)k^2 + (-ix_0 + ix)k\right]} dk \quad (36)$$

Let's define

$$\sigma(t) = \sigma = \sqrt{2Dt + \sigma_0^2} \quad (37)$$

Note that in normalized variables with  $D' = \frac{D}{\hbar^2}$  and  $\sigma'_0 = \frac{\sigma_0}{\hbar}$ ,

$$\sigma'(t) = \sigma' = \sqrt{2D't + \sigma_0'^2} \quad (38)$$

We substitute Eq. (37) into Eq. (36) and complete the square in the exponent.

$$\left(Dt + \frac{\sigma^2}{2}\right)k^2 + (-ix_0 + ix)k = \left(\frac{1}{2}\sigma^2\right)\left(k^2 + \frac{-ix_0 + ix}{\frac{1}{2}\sigma^2}k\right) \quad (39)$$

$$= \frac{1}{2}\sigma^2\left(k^2 + \frac{-ix_0 + ix}{\frac{\sigma^2}{2}}k + \left(\frac{-ix_0 + ix}{\sigma^2}\right)^2\right) - \frac{1}{2}\sigma^2\left(\frac{-ix_0 + ix}{\sigma^2}\right)^2 \quad (40)$$

$$= \frac{\sigma^2}{2}\left(k + \frac{-ix_0 + ix}{\sigma^2}\right)^2 - \frac{(-ix_0 + ix)^2}{2\sigma^2} \quad (41)$$

$$= \left(\frac{\sigma}{\sqrt{2}}k + \frac{-ix_0 + ix}{\sqrt{2}\sigma}\right)^2 - \frac{(-ix_0 + ix)^2}{2\sigma^2} \quad (42)$$

Using the substitution

$$u = \frac{\sigma^2 k + ix - ix_0}{\sqrt{2} \sigma} \quad (43)$$

$$du = \frac{\sigma}{\sqrt{2}} dk \quad (44)$$

$$dk = \frac{\sqrt{2}}{\sigma} du \quad (45)$$

we plug Eq. (42) back into the integral in Eq. (36) and solve.

$$\int_{-\infty}^{\infty} e^{-\left[\left(Dt + \frac{\sigma^2}{2}\right)k^2 + (-ix_0 + ix)k\right]} dk = \int_{-\infty}^{\infty} e^{-\left(\left(\frac{\sigma}{\sqrt{2}}k + \frac{ix_0 + ix}{\sqrt{2}\sigma}\right)^2 - \frac{(-ix_0 + ix)^2}{2\sigma^2}\right)} dk \quad (46)$$

$$= \int_{-\infty}^{\infty} \frac{\sqrt{2}e^{-u^2 + \frac{(ix - ix_0)^2}{2\sigma^2}}}{\sigma^2} du \quad (47)$$

$$= \frac{\sqrt{\pi} e^{\frac{(ix-ix_0)^2}{2\sigma^2}}}{\sqrt{2} \sigma} \int_{-\infty}^{\infty} \frac{2e^{-u^2}}{\sqrt{\pi}} du \quad (48)$$

$$= \frac{\sqrt{2\pi} e^{\frac{(ix-ix_0)^2}{2\sigma^2}}}{\sigma} \quad (49)$$

$$= \frac{\sqrt{2\pi} e^{\frac{i^2(x-x_0)^2}{2\sigma^2}}}{\sigma} \quad (50)$$

$$= \frac{\sqrt{2\pi} e^{-\frac{(x-x_0)^2}{2\sigma^2}}}{\sigma} \quad (51)$$

Plugging Eq. (51) back into Eq. (36), we get the solution in Eq. (52) for the one-dimensional case, or the solution in Eq. (53) for n-dimensions.

$$\rho(x, t) = \frac{e^{-\frac{(x-x_0)^2}{2\sigma^2}}}{\sqrt{2\pi} \sigma} \quad (52)$$

$$\rho(\mathbf{x}, t) = \left( \frac{e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\sigma^2}}}{\sqrt{2\pi} \sigma} \right)^n \quad (53)$$

Writing Eq. (52) and Eq. (53) in terms of the normalized variables, we get

$$\rho(x', t) = \frac{e^{-\frac{(x'-x'_0)^2}{2\sigma'^2}}}{\sqrt{2\pi} \sigma'} \quad (54)$$

$$\rho(\mathbf{x}', t) = \left( \frac{e^{-\frac{(\mathbf{x}'-\mathbf{x}'_0)^2}{2\sigma'^2}}}{\sqrt{2\pi} \sigma'} \right)^n \quad (55)$$

### C. Boundary conditions

Only the z-direction has boundary conditions. Our boundary conditions require that the charge density at the interface of the field-free region and the depletion zone ( $z = 0$ ) is zero as indicated in Eq. (56) because once charge gets to the depletion zone, we assume the electric field immediately picks up the electrons.

$$\rho(\mathbf{r}, z, t)|_{z=0} = 0 \quad (56)$$

This corresponds to the Dirichlet boundary conditions as solved in the heat equation.<sup>5</sup> The flux at a reflective boundary is zero. The partial derivative of the charge density with respect to  $z$  represents the flux. When  $z = h$  where  $h$  is the depth of the field-free region, the flux is zero. Therefore, the partial derivative must be zero, as indicated in Eq. (57).

$$\left. \frac{\partial \rho(\mathbf{r}, z, t)}{\partial z} \right|_{z=h} = 0 \quad (57)$$

This corresponds to the Neumann boundary conditions.<sup>5</sup> Since we have both a Dirichlet boundary condition and a Neumann boundary condition, the mixed boundary condition solution must be used to find the solution.<sup>3</sup>

In Section A., we present the general solutions to the diffusion equation. Using Eq. (8), we apply the boundary condition given in Eq. (56).

$$Z(z) = a \cos(mz) + b \sin(mz) \Big|_{z=0} = 0 \quad (58)$$

When  $z = 0$ ,  $b$  can take any value because  $b \sin(mz) = 0$ . The only way to ensure  $a \cos(mz) = 0$  for any  $m$  is to set  $a = 0$ . This gives us

$$Z(z) = b \sin(mz) \quad (59)$$

To find the value of  $b$ , we use the other boundary condition, given in Eq. (57).

$$\left. \frac{dZ}{dz} \right|_{z=h} = b m \cos(mh) = 0 \quad (60)$$

If  $b = 0$ , the solution is trivial, so we need to find where  $\cos(mh) = 0$ . This happens when

$$mh = \frac{\pi}{2} + n\pi \quad (61)$$

so

$$Z(z) = b_n \sin\left(\left(2n + 1\right)\frac{\pi}{2} \frac{z}{h}\right) \quad (62)$$

To simplify this equation, we set  $\alpha_n = n\pi + \frac{\pi}{2}$  and  $z' = \frac{z}{h}$ , which makes  $m = \frac{\alpha_n}{h}$ . Writing in terms of an infinite sum over  $n$ , we get

$$Z(z') = \sum_{n=0}^{\infty} b_n \sin(\alpha_n z') \quad (63)$$

We now refer back to the initial conditions to determine  $b_n$  because

$$Z(z')|_{t=0} = \sum_{n=0}^{\infty} b_n \sin(\alpha_n z') = \phi(z') \quad (64)$$

where  $\phi(z')$  is the initial condition. Even before applying specific initial conditions, we use the properties of orthogonality to get an expression for  $b_n$  in terms of  $\phi(z')$ .

$$\int_0^1 \sin^2(\alpha_n z') dz' = \frac{1}{2} \quad (65)$$

$$b_n = 2 \int_0^1 \phi(z') \sin(\alpha_n z') dz' \quad (66)$$

## D. Delta function applied to the boundary conditions

Ref. [3] takes us through the solution with the delta function. Using  $\phi(z') = \delta(z' - z'_0)$  in Eq. (66), we simplify the expression and plug it into Eq. (64). Multiplying Eq. (64) by the general solution for  $T(t)$  given in Eq. (6), we get the solution

$$Z(z')T(t) = 2 \sum_{n=0}^{\infty} e^{-\alpha_n^2 D' t} \sin(\alpha_n z'_0) \sin(\alpha_n z') \quad (67)$$

## E. Gaussian function applied to the boundary conditions

When applying the Gaussian function to the boundary conditions, we use the normalized variables  $z' = \frac{z}{h}$ ,  $z'_0 = \frac{z_0}{h}$ , and  $\sigma_0 = \frac{\sigma_0}{h}$  for simplicity. Using  $\phi(z') = \frac{1}{\sqrt{2\pi}\sigma'_0} e^{-\frac{(z'-z'_0)^2}{2\sigma'^2_0}}$  in Eq. (66), we get

$$b_n = 2 \int_0^1 \frac{1}{\sqrt{2\pi}\sigma'_0} e^{-\frac{(z'-z'_0)^2}{2\sigma'^2_0}} \sin(\alpha_n z') dz' \quad (68)$$

Using Euler's formula, we get the identity

$$\sin(x) = \frac{e^{ix} - e^{-ix}}{2i} \quad (69)$$

and rewrite the sine term in Eq. (68).

$$b_n = 2 \int_0^1 \frac{1}{\sqrt{2\pi}\sigma'_0} e^{-\frac{(z'-z'_0)^2}{2\sigma'^2_0}} \frac{e^{i\alpha_n z'} - e^{-i\alpha_n z'}}{2i} dz' \quad (70)$$

$$= \int_0^1 -\frac{i}{\sqrt{2\pi}\sigma'_0} e^{-\frac{(z'-z'_0)^2}{2\sigma'^2_0}} (e^{i\alpha_n z'} - e^{-i\alpha_n z'}) dz' \quad (71)$$

$$= -\frac{i}{\sqrt{2\pi}\sigma'_0} \int_0^1 e^{-\frac{(z'-z'_0)^2}{2\sigma'^2_0} + i\alpha_n z'} - e^{-\frac{(z'-z'_0)^2}{2\sigma'^2_0} - i\alpha_n z'} dz' \quad (72)$$

$$= -\frac{i}{\sqrt{2\pi}\sigma'_0} \int_0^1 e^{-\frac{(z'-z'_0)^2}{2\sigma'^2_0} + i\alpha_n z'} dz' + \frac{i}{\sqrt{2\pi}\sigma'_0} \int_0^1 e^{-\frac{(z'-z'_0)^2}{2\sigma'^2_0} - i\alpha_n z'} dz' \quad (73)$$

Now, we can solve each integral in Eq. (73) separately by completing the square for both exponents, simplifying both expressions, and recombining the two integral solutions into one solution for Eq. (73).

Completing the square for the exponent in the first integral in Eq. (73) where  $u = \frac{z'+z'_0-i\alpha_n\sigma_0'^2}{\sqrt{2}\sigma'_0}$  and  $dz' = \sqrt{2}\sigma'_0 du$ , we get

$$-\frac{(z' - z'_0)^2}{2\sigma_0'^2} + i\alpha_n z' = -\frac{1}{2\sigma_0'^2} (z'^2 - 2z'_0 z' + z_0'^2 - 2\sigma_0'^2 i\alpha_n z') \quad (74)$$

$$= -\frac{1}{2\sigma_0'^2} (z'^2 - 2z'(z'_0 + i\alpha_n\sigma_0'^2) + z_0'^2) \quad (75)$$

$$= -\frac{1}{2\sigma_0'^2} (z'^2 - 2z'(z'_0 + i\alpha_n\sigma_0'^2) + z_0'^2 + (z'_0 + i\alpha_n\sigma_0'^2)^2 - (z'_0 + i\alpha_n\sigma_0'^2)^2) \quad (76)$$

$$= -\frac{1}{2\sigma_0'^2} (z' - (z_0' - (i\alpha_n\sigma_0'^2)))^2 - \frac{1}{2\sigma_0'^2} (z_0'^2 - z_0'^2 - 2i\alpha_n z_0'\sigma_0'^2 + \alpha_n^2\sigma_0'^4) \quad (77)$$

$$= -u^2 - \frac{1}{2\sigma_0'^2} (-2i\alpha_n z_0'\sigma_0'^2 + \alpha_n^2\sigma_0'^4) \quad (78)$$

$$= -u^2 + i\alpha_n z_0' - \frac{\alpha_n\sigma_0'^2}{2} \quad (79)$$

We now plug Eq. (79) into the exponent in the first integral in Eq. (73) to get

$$\frac{-i}{\sqrt{2\pi\sigma_0}} \int_0^1 e^{-\frac{(z'-z_0')^2}{2\sigma_0'^2} + i\alpha_n z'} dz' = \frac{-i}{\sqrt{2\pi\sigma_0}} \int_0^1 e^{-u^2 + i\alpha_n z_0' - \frac{\alpha_n\sigma_0'^2}{2}} dz' \quad (80)$$

$$= -\frac{\sqrt{2}\sigma_0 i}{2\sqrt{2}\sigma_0} e^{-\frac{\alpha_n\sigma_0'^2}{2} + i\alpha_n z_0'} \int_0^1 \frac{2}{\sqrt{\pi}} e^{-u^2} du \quad (81)$$

$$= -\frac{i}{2} e^{-\frac{\alpha_n\sigma_0'^2}{2} + i\alpha_n z_0'} \operatorname{erf}(1) \quad (82)$$

Completing the square for the exponent in the second integral in Eq. (73) where  $u = \frac{z'-z_0'+i\alpha_n\sigma_0'^2}{\sqrt{2}\sigma_0'}$  and  $dz' = \sqrt{2}\sigma_0' du$ , we get

$$-\frac{(z' - z_0')^2}{2\sigma_0'^2} + i\alpha_n z' = -\frac{1}{2\sigma_0'^2} (z'^2 - 2z_0'z' + z_0'^2 - 2\sigma_0'^2 i\alpha_n z') \quad (83)$$

$$= -\frac{1}{2\sigma_0'^2} (z'^2 - 2z'(z_0' - i\alpha_n\sigma_0'^2) + z_0'^2) \quad (84)$$

$$= -\frac{1}{2\sigma_0'^2} (z'^2 - 2z'(z_0' - i\alpha_n\sigma_0'^2) + z_0'^2 + (z_0' - i\alpha_n\sigma_0'^2)^2 - (z_0' - i\alpha_n\sigma_0'^2)^2) \quad (85)$$

$$= -\frac{1}{2\sigma_0'^2} (z' - (z_0' - (i\alpha_n\sigma_0'^2)))^2 - \frac{1}{2\sigma_0'^2} (z_0'^2 - z_0'^2 + 2i\alpha_n z_0'\sigma_0'^2 + \alpha_n^2\sigma_0'^4) \quad (86)$$

$$= -u^2 - \frac{1}{2\sigma_0'^2} (2i\alpha_n z_0'\sigma_0'^2 + \alpha_n^2\sigma_0'^4) \quad (87)$$

$$= -u^2 - i\alpha_n z_0' - \frac{\alpha_n\sigma_0'^2}{2} \quad (88)$$

Plugging Eq. (88) into the exponent in the second integral in Eq. (73), we get

$$\frac{i}{\sqrt{2\pi\sigma_0}} \int_0^1 e^{-\frac{(z'-z_0')^2}{2\sigma_0'^2} + i\alpha_n z'} dz' = \frac{i}{\sqrt{2\pi\sigma_0}} \int_0^1 e^{-u^2 - i\alpha_n z_0' - \frac{\alpha_n\sigma_0'^2}{2}} dz' \quad (89)$$

$$= \frac{\sqrt{2}\sigma_0 i}{2\sqrt{2}\sigma_0} e^{-\frac{\alpha_n\sigma_0'^2}{2} - i\alpha_n z_0'} \int_0^1 \frac{2}{\sqrt{\pi}} e^{-u^2} du \quad (90)$$



$$= \frac{i}{2} e^{-\frac{\alpha_n \sigma_0'^2}{2} - i\alpha_n z_0'} \operatorname{erf}(1) \quad (91)$$

Now plugging Eq. (82) and Eq. (91) back into Eq. (73), we get

$$b_n = -\frac{i}{2} e^{-\frac{\alpha_n^2 \sigma_0'^2}{2} + i\alpha_n z_0'} \operatorname{erf}(1) + \frac{i}{2} e^{-\frac{\alpha_n^2 \sigma_0'^2}{2} - i\alpha_n z_0'} \operatorname{erf}(1) \quad (92)$$

$$= -\frac{i}{2} \operatorname{erf}(1) e^{-\frac{\alpha_n^2 \sigma_0'^2}{2}} (e^{i\alpha_n z_0'} - e^{-i\alpha_n z_0'}) \quad (93)$$

$$= \operatorname{erf}(1) e^{-\frac{\alpha_n^2 \sigma_0'^2}{2}} \left( \frac{e^{i\alpha_n z_0'} - e^{-i\alpha_n z_0'}}{2i} \right) \quad (94)$$

$$= \operatorname{erf}(1) e^{-\frac{\alpha_n^2 \sigma_0'^2}{2}} \sin(\alpha_n z_0') \quad (95)$$

Plugging this back into Eq. (64), we get

$$Z(z')|_{t=0} = \operatorname{erf}(1) \sum_{n=0}^{\infty} e^{-\frac{\alpha_n^2 \sigma_0'^2}{2}} \sin(\alpha_n z_0') \sin(\alpha_n z') \quad (96)$$

This gives us the solution

$$Z(z')T(t) = \operatorname{erf}(1) \sum_{n=0}^{\infty} e^{-\frac{\alpha_n^2 \sigma_0'^2}{2}} \sin(\alpha_n z_0') \sin(\alpha_n z') e^{-\frac{\alpha_n^2 D t}{h^2}} \quad (97)$$

Since we know that the normalization of the diffusion coefficient is  $D' = \frac{D}{h^2}$ , we can simplify our solution to be

$$Z(z')T(t) = \operatorname{erf}(1) \sum_{n=0}^{\infty} e^{-\frac{\alpha_n^2 \sigma_0'^2 - 2\alpha_n^2 D' t}{2}} \sin(\alpha_n z_0') \sin(\alpha_n z') \quad (98)$$

$$Z(z')T(t) = \operatorname{erf}(1) \sum_{n=0}^{\infty} e^{-\frac{\alpha_n^2}{2} (2D' t + \sigma_0'^2)} \sin(\alpha_n z_0') \sin(\alpha_n z') \quad (99)$$

Using the definition of  $\sigma'(t)$  in Eq. (38), our solution is

$$Z(z')T(t) = \operatorname{erf}(1) \sum_{n=0}^{\infty} e^{-\frac{\alpha_n^2}{2} \sigma'^2} \sin(\alpha_n z_0') \sin(\alpha_n z') \quad (100)$$

## F. Final solutions

Combining the solution in the  $x$  and  $y$  directions with the solution in the  $z$  direction for both initial conditions, we obtain the final solutions. For the Delta function initial condition, we get

$$\rho(\mathbf{x}, t) = \frac{e^{-\frac{(x-x_0)^2+(y-y_0)^2}{4Dt}}}{\sqrt{4\pi Dt}} 2 \sum_{n=0}^{\infty} e^{-\alpha_n^2 D't} \sin(\alpha_n z'_0) \sin(\alpha_n z') \quad (101)$$

For the Gaussian function initial condition, we get

$$\rho(\mathbf{x}, t) = \frac{e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma}} \text{erf}(1) \sum_{n=0}^{\infty} e^{-\frac{\alpha_n^2}{2}\sigma'^2} \sin(\alpha_n z'_0) \sin(\alpha_n z') \quad (102)$$

The solution with the Gaussian function as the initial condition is very similar to that of the delta function initial condition.

Now, we find the flux through the  $z = 0$  boundary. This is the charge that leaves the field-free region, therefore will give us the charge distribution.

## III. Charge spreading

A reflecting boundary ensures all charge is collected. This causes an increase in radial spreading at reflecting boundaries.<sup>6</sup> The charge flux at the interface of the field-free zone and the depletion zone integrated over time gives the radial distribution, as given in Eq. (103).

$$q(\mathbf{r}) = D \int_0^{\infty} \frac{\partial \rho}{\partial z} \Big|_{z=0} dt \quad (103)$$

Ref. [3] gives us

$$\frac{\partial \rho(\mathbf{r}, z, t)}{\partial z} \Big|_{z=0} = \frac{Q_0}{4\pi Dth^2} \cdot 2 \sum_{n=0}^{\infty} \alpha_n \sin(\alpha_n z'_0) e^{-\frac{\alpha_n^2 Dt}{h^2} - \frac{r^2}{4Dt}} \quad (104)$$

where  $Q_0$  is the initial charge.

Integrating first with respect to time then space, we get

$$q(\mathbf{r}) = \frac{Q_0}{\pi h^2} \sum_{n=0}^{\infty} \alpha_n \sin(\alpha_n z'_0) K_0\left(\alpha_n \frac{r}{h}\right) \quad (105)$$

where  $q(\mathbf{r})$  is the flux through the surface element and  $K_0$  is a modified Bessel function of the second kind. We must switch from cylindrical to Cartesian coordinates because the pixels are square. To get the charge distribution, we integrate Eq. (105) over the pixel area. This becomes problematic because the Bessel function creates a singularity in the function. Switching the order of integration eliminates this problem.

### A. Integrating with respect to space before time

We now start with the spatial part of the integral. A double integral is necessary because the pixels have a dimension in both the  $x$  and  $y$  directions. The integral we calculate comes from Ref. [3]

$$q = \left( \frac{Q_0}{h^2} \sum_{n=0}^{\infty} \alpha_n \sin(\alpha_n z'_0) e^{-\frac{\alpha_n^2 D t}{h^2}} \right) \frac{1}{2\pi D t} \int_B^C \int_E^F e^{-\frac{r^2}{4Dt}} dx dy \quad (106)$$

where  $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$ . Temporarily ignoring constants and substituting for  $r$  to convert from cylindrical to Cartesian coordinates, we get

$$q_{pixel} = \frac{1}{2\pi D t} \int_B^C \int_E^F e^{-\frac{(x-x_0)^2 + (y-y_0)^2}{4Dt}} dx dy \quad (107)$$

We separate the integral in Eq. (107) into separate integrals for  $x$  and  $y$ .

$$q_{pixel} = \frac{1}{2\pi D t} \int_B^C e^{-\frac{(x-x_0)^2}{4Dt}} dx \int_E^F e^{-\frac{(y-y_0)^2}{4Dt}} dy \quad (108)$$

We write Eq. (108) in the form of the error function in Eq. (109).

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2} du \quad (109)$$

Because of the limits of integration, the result is a product of differences of the error function.

$$q_{pixel} = \frac{1}{2\pi Dt} \int_B^C e^{-\left(\frac{x-x_0}{\sqrt{4Dt}}\right)^2} dx \int_E^F e^{-\left(\frac{y-y_0}{\sqrt{4Dt}}\right)^2} dy \quad (110)$$

$$= \frac{1}{2\pi Dt} \left( \int_0^C e^{-\left(\frac{x-x_0}{\sqrt{4Dt}}\right)^2} dx - \int_0^B e^{-\left(\frac{x-x_0}{\sqrt{4Dt}}\right)^2} dx \right) \times \left( \int_0^F e^{-\left(\frac{y-y_0}{\sqrt{4Dt}}\right)^2} dy - \int_0^E e^{-\left(\frac{y-y_0}{\sqrt{4Dt}}\right)^2} dy \right) \quad (111)$$

We now look at each integral in Eq. (111) independently. All integrals in Eq. (111) are of the same form, therefore follow the same solution process as the first integral in the following equations.

$$\int_0^C e^{-\left(\frac{x-x_0}{\sqrt{4Dt}}\right)^2} dx = \frac{\sqrt{\pi}}{2} \int_0^C \frac{2}{\sqrt{\pi}} e^{-\left(\frac{x-x_0}{\sqrt{4Dt}}\right)^2} dx \quad (112)$$

$$= -\frac{\sqrt{\pi}}{2} 2\sqrt{Dt} \operatorname{erf}\left(\frac{x_0 - x}{2\sqrt{Dt}}\right) \Big|_0^C \quad (113)$$

$$= -\sqrt{\pi Dt} \left( \operatorname{erf}\left(\frac{x_0 - C}{2\sqrt{Dt}}\right) - \operatorname{erf}\left(\frac{x_0}{2\sqrt{Dt}}\right) \right) \quad (114)$$

$$= \sqrt{\pi Dt} \left( \operatorname{erf}\left(\frac{x_0}{2\sqrt{Dt}}\right) - \operatorname{erf}\left(\frac{x_0 - C}{2\sqrt{Dt}}\right) \right) \quad (115)$$

$$\int_0^B e^{-\left(\frac{x-x_0}{\sqrt{4Dt}}\right)^2} dx = \sqrt{\pi Dt} \left( \operatorname{erf}\left(\frac{x_0}{2\sqrt{Dt}}\right) - \operatorname{erf}\left(\frac{x_0 - B}{2\sqrt{Dt}}\right) \right) \quad (116)$$

$$\int_0^F e^{-\left(\frac{y-y_0}{\sqrt{4Dt}}\right)^2} dy = \sqrt{\pi Dt} \left( \operatorname{erf}\left(\frac{y_0}{2\sqrt{Dt}}\right) - \operatorname{erf}\left(\frac{y_0 - F}{2\sqrt{Dt}}\right) \right) \quad (117)$$

$$\int_0^E e^{-\left(\frac{y-y_0}{\sqrt{4Dt}}\right)^2} dy = \sqrt{\pi Dt} \left( \operatorname{erf}\left(\frac{y_0}{2\sqrt{Dt}}\right) - \operatorname{erf}\left(\frac{y_0 - E}{2\sqrt{Dt}}\right) \right) \quad (118)$$

Now, we find the difference of the integral from 0 to C and 0 to B to simplify the first term in the product and the difference of the integral from 0 to F and 0 to E to simplify the second term in the

product.

$$\int_0^C e^{-\left(\frac{x-x_0}{\sqrt{4Dt}}\right)^2} dx - \int_0^B e^{-\left(\frac{x-x_0}{\sqrt{4Dt}}\right)^2} dx = \sqrt{\pi Dt} \left( \operatorname{erf} \left( \frac{x_0}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{x_0 - C}{2\sqrt{Dt}} \right) \right) \quad (119)$$

$$- \sqrt{\pi Dt} \left( \operatorname{erf} \left( \frac{x_0}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{x_0 - B}{2\sqrt{Dt}} \right) \right) \\ = \sqrt{\pi Dt} \left( \operatorname{erf} \left( \frac{x_0 - B}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{x_0 - C}{2\sqrt{Dt}} \right) \right) \quad (120)$$

$$\int_0^F e^{-\left(\frac{y-y_0}{\sqrt{4Dt}}\right)^2} dx - \int_0^E e^{-\left(\frac{y-y_0}{\sqrt{4Dt}}\right)^2} dx = \sqrt{\pi Dt} \left( \operatorname{erf} \left( \frac{y_0}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{y_0 - F}{2\sqrt{Dt}} \right) \right) \quad (121)$$

$$- \sqrt{\pi Dt} \left( \operatorname{erf} \left( \frac{y_0}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{y_0 - E}{2\sqrt{Dt}} \right) \right) \\ = \sqrt{\pi Dt} \left( \operatorname{erf} \left( \frac{y_0 - E}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{y_0 - F}{2\sqrt{Dt}} \right) \right) \quad (122)$$

Plugging Eqs. (120) and (122) into Eq. (111) and simplifying, we get

$$q_{pixel} = \frac{1}{2\pi Dt} \left[ \sqrt{\pi Dt} \left( \operatorname{erf} \left( \frac{x_0 - B}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{x_0 - C}{2\sqrt{Dt}} \right) \right) \right] \quad (123)$$

$$\left[ \sqrt{\pi Dt} \left( \operatorname{erf} \left( \frac{y_0 - E}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{y_0 - F}{2\sqrt{Dt}} \right) \right) \right] \\ = \frac{1}{2} \left[ \operatorname{erf} \left( \frac{x_0 - B}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{x_0 - C}{2\sqrt{Dt}} \right) \right] \left[ \operatorname{erf} \left( \frac{y_0 - E}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{y_0 - F}{2\sqrt{Dt}} \right) \right] \quad (124)$$

Going back to Eq. (106), we also integrate over time. Using Eq. (124) in place of the portion of the equation outside the parenthesis in Eq. (106) and again ignoring the extra coefficients, we get

$$q_{time} = \frac{1}{2h^2} \int_0^\infty \sum_{n=0}^\infty e^{-\frac{\alpha_n^2 Dt}{h^2}} \frac{1}{2} \left[ \operatorname{erf} \left( \frac{x_0 - B}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{x_0 - C}{2\sqrt{Dt}} \right) \right] \\ \left[ \operatorname{erf} \left( \frac{y_0 - E}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{y_0 - F}{2\sqrt{Dt}} \right) \right] dt \quad (125)$$

$$= \frac{1}{4h^2} \sum_{n=0}^\infty \int_0^\infty e^{-\frac{\alpha_n^2 Dt}{h^2}} \left[ \operatorname{erf} \left( \frac{x_0 - B}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{x_0 - C}{2\sqrt{Dt}} \right) \right] \\ \left[ \operatorname{erf} \left( \frac{y_0 - E}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{y_0 - F}{2\sqrt{Dt}} \right) \right] dt \quad (126)$$

We plug this back into Eq. (106) to solve for  $q$ .

$$q = \frac{Q_0}{4h^2} \sum_{n=0}^{\infty} \alpha_n \sin(\alpha_n z'_0) \int_0^{\infty} e^{-\frac{\alpha_n^2 Dt}{h^2}} \left[ \operatorname{erf} \left( \frac{x_0 - B}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{x_0 - C}{2\sqrt{Dt}} \right) \right] \times \left[ \operatorname{erf} \left( \frac{y_0 - E}{2\sqrt{Dt}} \right) - \operatorname{erf} \left( \frac{y_0 - F}{2\sqrt{Dt}} \right) \right] dt \quad (127)$$

Writing the equation in terms of normalized variables simplifies the equation. We normalize variables by using the depth of the field-free region  $h$  as a scaling factor. Let  $B' = \frac{B-x_0}{\sigma} = \frac{b}{\sigma}$ ,  $C' = \frac{C-x_0}{\sigma} = \frac{c}{\sigma}$ ,  $E' = \frac{E-y_0}{\sigma} = \frac{e}{\sigma}$ ,  $F' = \frac{F-y_0}{\sigma} = \frac{f}{\sigma}$ , and  $\sigma^2 = \frac{4Dt}{h^2}$ . Eq. (127) becomes

$$q = \frac{Q_0}{2} \sum_{n=0}^{\infty} \alpha_n \sin(\alpha_n z'_0) \int_0^{\infty} e^{-\frac{\alpha_n^2 \sigma^2}{4}} \frac{1}{4} \left( \operatorname{erf} \left( \frac{c}{\sigma} \right) - \operatorname{erf} \left( \frac{b}{\sigma} \right) \right) \left( \operatorname{erf} \left( \frac{f}{\sigma} \right) - \operatorname{erf} \left( \frac{e}{\sigma} \right) \right) d\sigma^2 \quad (128)$$

where in the equations above,

$$\alpha_n = \frac{(2n+1)\pi}{2} \quad (129)$$

## B. Numeric integration: the midpoint method

The midpoint method for numerically integrating a function uses the formula given in Eq. (130) where  $b$  is the upper limit of the interval,  $a$  is the lower limit of the interval,  $c$  is the number of rectangles into which the interval is split, and  $x_i$  are the endpoints of each of these rectangles. The  $f$  indicates that these values are plugged back into the function being integrated to get the value at those points.

$$\frac{b-a}{c} \left[ f \left( \frac{x_0+x_1}{2} \right) + f \left( \frac{x_1+x_2}{2} \right) + \dots + f \left( \frac{x_{n-2}+x_{n-1}}{2} \right) + f \left( \frac{x_{n-1}+x_n}{2} \right) \right] \quad (130)$$

### 1. Numerically integrating the Gaussian function

We start by testing the usefulness of the midpoint method using the integral of the Gaussian function  $\int_{-\infty}^{\infty} e^{-x^2} dx$ . Plotting this function in Fig. (1), we estimate limits from -4 to 4 would be

appropriate for approximating a numerical solution. From the shape of the function, we see that

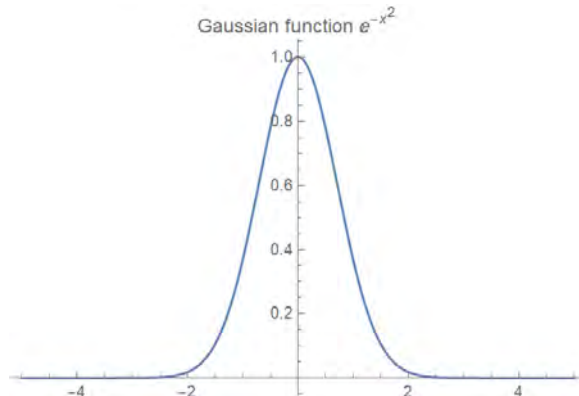


Figure 1: Plot of the Gaussian function. The limits of integration can be approximated by looking where the function has a value very close to zero.

using more rectangles, or a higher  $c$  value, will provide a better approximation. For simplicity in calculation, we show the case where  $c = 8$ .

$$\int_{-4}^4 e^{-x^2} dx \simeq \frac{4 - (-4)}{8} \left[ f\left(\frac{-4 + (-3)}{2}\right) + f\left(\frac{-3 + (-2)}{2}\right) + f\left(\frac{-2 + (-1)}{2}\right) \right] \quad (131)$$

$$+ f\left(\frac{-1 + 0}{2}\right) + f\left(\frac{0 + 1}{2}\right) + f\left(\frac{1 + 2}{2}\right) + f\left(\frac{2 + 3}{2}\right) + f\left(\frac{3 + 4}{2}\right) \Big] \\ \simeq e^{-\frac{49}{4}} + e^{-\frac{25}{4}} + e^{-\frac{9}{4}} + e^{-\frac{1}{4}} + e^{-\frac{1}{4}} + e^{-\frac{9}{4}} + e^{-\frac{25}{4}} + e^{-\frac{49}{4}} \quad (132)$$

$$\simeq 1.77227049 \quad (133)$$

We know from our analytic solution of the integral of the Gaussian function from  $-\infty$  to  $\infty$  that the value we expect is  $\sqrt{\pi} \simeq 1.77245385$ . Our approximation is close. To get a more accurate approximation, we write a C++ code to perform the numeric integration but with a larger  $c$ . Our value from the program agrees with our solution above with the input  $c = 8$ . With  $c = 10$ , the integral is equal to 1.77245, which matches more closely to the expected value. The results from running our code are in Figures 2 and 3. Therefore, the midpoint method is sufficient for evaluating the integral given reasonable choices for the interval limits and the value of  $c$ .

```

Start of interval: -4
End of interval: 4
Number of rectangles to split interval into: 8
Integral is equal to: 1.77227
Absolute integration error is: 0.000188141
Relative integration error is: 0.0106147%

```

Figure 2: Output of C++ code for integrating  $e^{-x^2}$  from -4 to 4 with 8 rectangles.

```

Start of interval: -4
End of interval: 4
Number of rectangles to split interval into: 10
Integral is equal to: 1.77245
Absolute integration error is: 2.59679e-06
Relative integration error is: 0.000146508%

```

Figure 3: Output of C++ code for integrating  $e^{-x^2}$  from -4 to 4 with 10 rectangles.

## 2. Numeric integration of charge distribution equation

We numerically integrate our more complex charge distribution equation given in Eq. (128). First, we plot the function since we cannot use infinity as a limit when integrating numerically. Because  $h = 8\mu m$  and the pixels are  $16\mu m$ , we get the numerators of the arguments of the error functions to be 1 for the center pixel. We also use the substitution that  $\sigma^2 = u$  so that the variable we are integrating over is not squared. For the plot in Figure 4 and numeric integration in Figure 5, we use  $n = 0$  to obtain the solution to the integral for a single term in the series. This gives us

$$\int_0^\infty e^{-\frac{((2n+1)\pi)^2}{4}u} \left( \operatorname{erf}\left(\frac{1}{\sqrt{u}}\right) - \operatorname{erf}\left(\frac{-1}{\sqrt{u}}\right) \right) \left( \operatorname{erf}\left(\frac{1}{\sqrt{u}}\right) - \operatorname{erf}\left(\frac{-1}{\sqrt{u}}\right) \right) du \quad (134)$$

We evaluate the integral using our C++ code. From Figure 4 we see that the function value quickly decreases, so contributions to the solution from outside our chosen limits of 0 to 20 are minimal. When using the midpoint rule, we use a large number of rectangles to calculate the integral value to obtain a more accurate result. Evaluating the integral, we get 4.26841, as shown in Figure 5. We compare this value to the result from the built-in numeric integration function in Mathematica to ensure our code is working properly. The difference between the integral results is less than 0.002%, therefore assuring our code is working properly. Note that this result is only the result



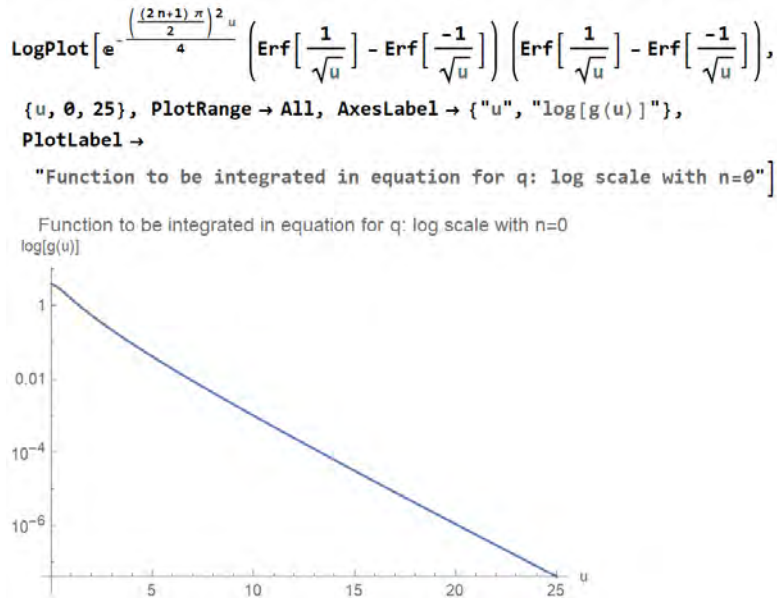


Figure 4: Plot of function inside the integral,  $g(u)$ , in the equation for  $q$ . The value of the function evaluated with  $n = 0$  is displayed for values of  $u$  up to 25 to visualize how fast the function approaches 0.

```

Start of interval: 0
End of interval: 20
Number of rectangles to split interval into: 999999
Integral is equal to: 4.26841
Absolute integration error is: 7.99996e-05
Relative integration error is: 0.00187419%

```

Figure 5: Result of integration from 0 to 20.

of the integral, not the full solution for the amount of charge in the center pixel. The full solution requires us to multiply by the constants we ignored and perform the sum over  $n$  instead of taking only the  $n = 0$  term. For our analysis, we take  $Q_0 = 1$  and  $z'_0 = 1$ .

We calculate the charge distribution throughout the pixels by finding the results of these numeric integrations using the edges of each pixel for the limits of integration. We still use  $x_0 = 0$  and  $y_0 = 0$  because we assume the incoming charges are still focused at the center of the center pixel. Figure 6 shows the layout we use to find these values.

The integral for all the pixels can be computed at the same time using our C++ code. The values are stored in an array, then plotted using ROOT. Due to the difference in amount of charge between the center pixel and the outer pixels, we choose to use a log scale on our z-axis to make

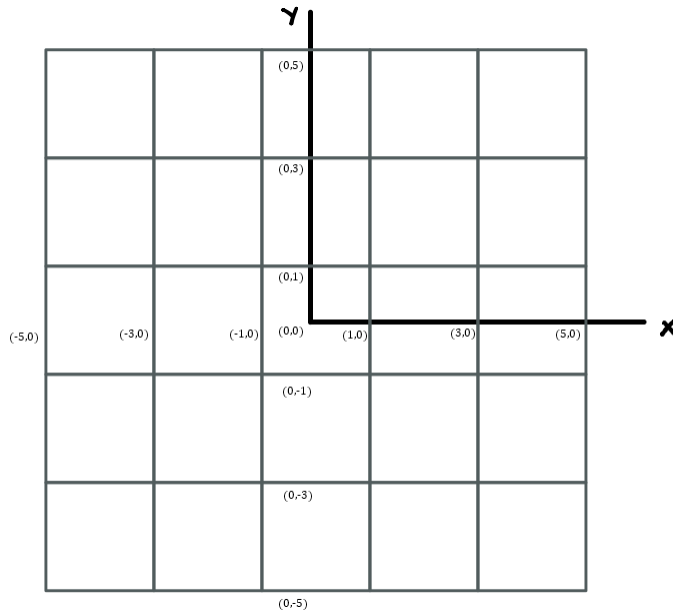


Figure 6: Pixels on coordinate system.

viewing and analysis easier. Note that in our original formula,  $n$  was being summed over from 0 to infinity, but at this point we are only using single values of  $n$  to determine the shape of the charge distribution for each term in the sum. To compare the graphs at each  $n$  value, we normalize the amount of charge in the pixels by dividing all pixels by the height of the center pixel for that graph. Figure 7 shows the difference in the charge distribution shape for different values of  $n$ .

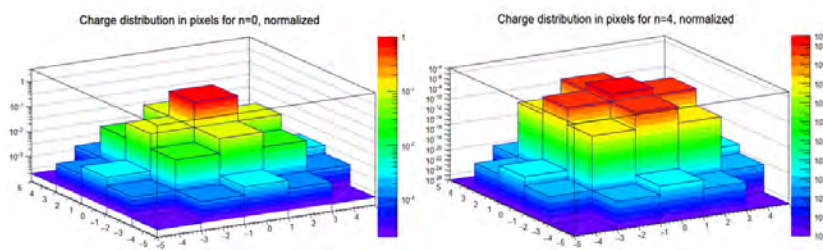


Figure 7: Charge distribution for different values of  $n$

We further investigate this by creating a plot of the dependence of the integral in the charge distribution equation on the value of  $n$ . As we see in Figure 8, the value of the integral depends on  $n$  more when  $n$  is small, but very little as  $n$  increases. The graph on the right in Figure 8 shows the dependence of the entire equation for  $q$  on  $n$ . This graph shows us that even with the oscillations

of the sine function, as  $n$  increases the value of the  $n^{th}$  term approaches 0. From this, we conclude

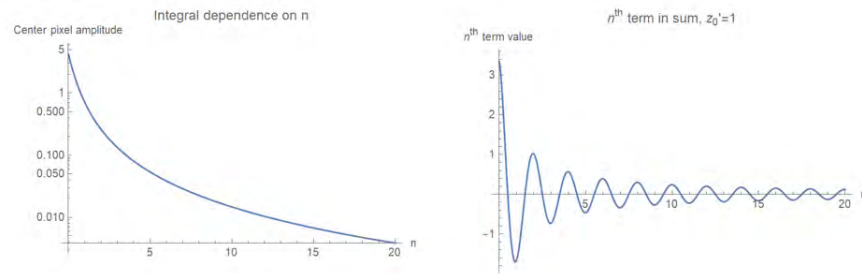


Figure 8: Left: Center pixel dependence on  $n$ . The value of the integral depends on  $n$  less as  $n$  increases. Right: Value of  $n^{th}$  term in sum converges. The sine function causes oscillations in the solution.

that the distribution shape depends on  $n$ .

Note that changing the value of  $z'_0$  affects the behavior of the function, as seen in Figure 9.

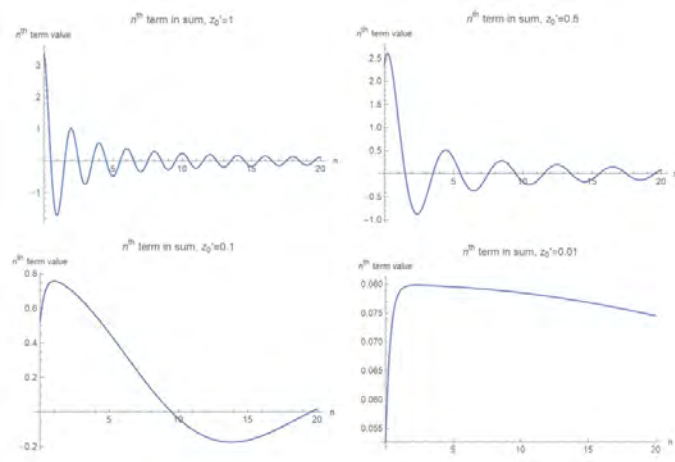


Figure 9: Value of  $n^{th}$  term in sum behavior at different  $z'_0$  values. The value of  $z'_0$  will affect the charge distribution, but more analysis is necessary.

We create plots for individual values of  $n$  to determine how fast the series converges. Using  $n = 0, n = 4$ , and  $n = 10$ , we see that as  $n$  increases, the contribution of the term to the solution decreases. The center pixel contribution for each term is the greatest of all the pixels, so we focus our analysis on the center pixel. In the  $n = 0$  plot in Figure 10, the center pixel contribution is 0.83812, whereas in the  $n = 4$  plot, the center pixel contribution decreases to 0.14147 and by  $n = 10$ , the contribution is only 0.06062. By  $n = 100$ , the contribution from the center pixel is down to 0.00263.

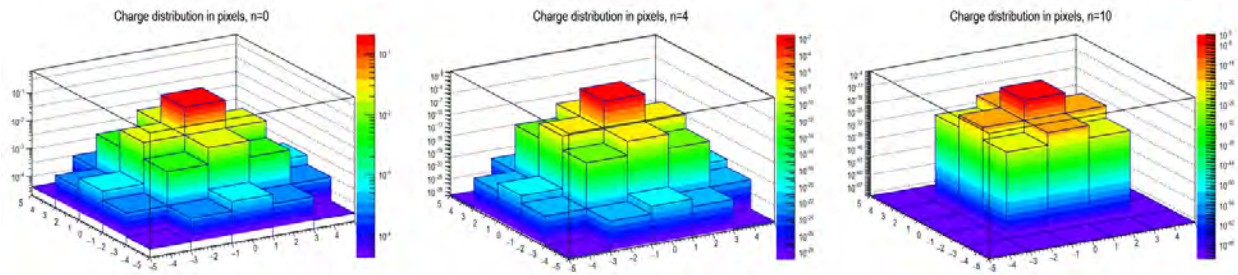


Figure 10: Charge distribution for different terms in the series.

The next step is to include the sum over  $n$ . Figure 11 shows the charge distribution when  $n$  is summed over from 0 to 100, which approximates the charge distribution over all  $n$ .

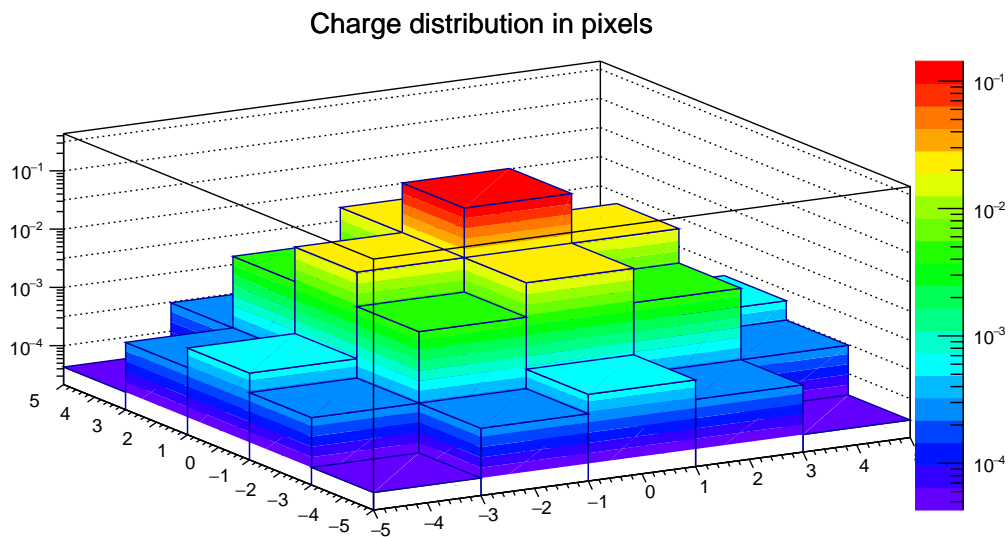


Figure 11: Charge distribution plot using terms from  $n = 0$  through  $n = 100$ . The center pixel has the most charge, which is expected since the x-ray beam is centered at  $(0,0)$ . The distribution is symmetric.

## IV. Conclusion

We solve the diffusion equation for the delta function initial condition, which assumes all initial charge is in a single point, and for the Gaussian function initial condition that gives the initial charge shape. We also solve for the charge distribution over all the pixels in the EMCCD. Our equation gives fast convergence and takes square pixel shape into consideration. When the

incoming x-rays are focused in the center of the center pixel, we see that the charges are focused in the center pixel and there is less charge in the surrounding pixels. The charge distribution is symmetric across the pixels. A more accurate predictive function for charge distribution will be useful at the SIX beamline as well as in other applications with EMCCDs.

## V. Acknowledgements

I would like to thank my mentor Dr. Ivan Kotov for his guidance and for giving me the opportunity to work on this project. I would also like to thank Mikhail Smirnov for his collaboration on different aspects of this project. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## VI. References

<sup>1</sup> T.J. Fellers and M.W. Davidson, *Electron multiplying charge-coupled devices (EMCCDs)*, <<https://hamamatsu.magnet.fsu.edu/articles/emccds.html>>

<sup>2</sup> I. Kotov, S. Hall, D. Gopinath, A. Barbour, J. Li, Y. Gu, K. Holland, A. Holland, I. Jarriage, J. Pellicciari, M. Soman, S. Wilkins, and V. Bisogni, "Analysis of the EMCCD point-source response using x-rays", *Nuclear Inst. and Methods in Physics Research, A*, <<https://doi.org/10.1016/j.nima.2020.164706>> (2020).

<sup>3</sup> I. V. Kotov, "Diffusion in the field-free zone", Personal communications (2021).

<sup>4</sup> T. Toyabe, T. Shinoda, M. Aoki, H. Kawamoto, K. Mitsusada, T. Masuhara, and S. Asai, *IEEE transactions on electron devices* (1982).

<sup>5</sup> *Heat equation*, <<https://web.stanford.edu/class/math220b/handouts/heatqn.pdf>>.

<sup>6</sup> G. R. Hopkinson, "Charge diffusion effects in CCD x-ray detectors", *Nuclear Instruments and Methods* **216**, 423-429 (1983).

Text mining from the scientific literature with natural language processing methods

Darrien Hunt

Computer Programming & Information Systems, Farmingdale State College, Farmingdale, NY 11735

Dr. Line Pouchard

Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973

Dr. Gilchan Park

Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973

## Abstract

Scientists often manually review scientific literature for pertinent information, which proves a time-consuming task. Many useful published articles are formatted as PDF files which make parsing through this literature using automated machine learning algorithms arduous because the information is presented in a layout-based format. In this paper, I evaluate the effectiveness of two PDF parsing tools, PyMuPDF [2] and PDFPlumber [3], in comparison to previously evaluated tools through their ability to distinguish paragraph boundaries, determine the reading order, determine semantic roles, translate ligatures, and merge hyphenated words. The success in the evaluation of PDF parsing tools will identify a useful method for parsing and evaluating scientific literature. In comparison to previously evaluated tools, both PyMuPDF [2] and PDFPlumber [3] produced low error rates when processing PDF files (less than ~0.03%), which is good in comparison to the most effective tools presented in the evaluation. However, in some categories, e.g., reproducing the correct number of newlines, these tools performed worse than other evaluated tools. PyMuPDF [2] and PDFPlumber [3] produced results that varied 112.4% and 189.1% respectively, relative to the number of newlines in the accurately produced comparison files. Following the evaluation, we leverage PyMuPDF [2] to help develop, extract, and process a scientific corpus. Applying these tools to extract annotations on relevant articles is essential for training a model that leverages Bidirectional Encoder Representations from Transformers, a natural language processing technique, to semantically analyze relevant information. Information extracted on the purification of MoFe and Fe proteins to develop a referable X-Ray Crystal Structure will be useful for biologists to study recently discovered Nitrogenase-like enzymes and rationalize why they produce methane as opposed to Nitrogenase

enzymes which produce Nitrogen. Conducting this research has exposed me to valuable Natural Language Processing concepts and techniques, to a variety of useful Python packages, and to research methodology I can apply to many projects.

## **Introduction**

PDFs prove to be a popular and consistent format for scientific literature, especially in the domain of biology. Databases such as *bioRxiv.org* and *rcsb.org* contain various articles and papers related to the Crystallization process that scientists can use to garner a better understanding of various proteins and enzymes. This methodology can be retrieved from manually scanning literature, however, there are a vast number of articles that may or may not contain the desired information: this manual process can prove tedious and time-consuming. Various methods exist today to automatically process text, but parsing PDFs presents a unique set of challenges because of their layout-based format. This layout-based format makes it difficult to process the text semantically as PDF articles specify the fonts and positions of each character rather than recognizing their meanings and roles.

PDF parsing tools have been created to mitigate issues surrounding extracting text from PDF formatted literature but vary in effectiveness. *A Benchmark and Evaluation for Text Extraction from PDF* [4] presents evidence of this as the evaluation conducted on 14 different tools showcases the functionality of each tool. These tools were evaluated across several metrics including their ability to accurately identify newlines, paragraphs, and words. By using a generated benchmark, comparisons could be made to articles processed by each of these tools to determine whether the tool made a substantial number of errors or not. In searching for the most effective PDF parsing tools, we identified two that were not previously evaluated in *A*



*Benchmark and Evaluation for Text Extraction from PDF* [4]. This project serves as an evaluation of these two tools, PyMuPDF [2] and PDFPlumber [3], using the same methodology provided in [4]. It also provides an opportunity to experiment with leveraging these tools to build a scientific corpus and extract information on the crystallization process, with the intent of developing an algorithm to automatically complete this task.

## **Background**

Extracting text from PDF literature can be a daunting task, but many current tools have shown to be efficient in this process. A paper formatted in PDF presents information in a layout-based format, which means the characters are recognized by either their fonts or positions and not their semantics. A paper published in 2017 titled *A Benchmark and Evaluation for Text Extraction from PDF* [4] provides solid evidence of this. This study identifies various tools for text extraction from PDFs and evaluates them on a number of metrics. The tools are as follows: *pdftotext*, *pdftohtml*, *pdftoxml*, *PdfBox*, *pdf2xml*, *ParsCit*, *LA-PdfText*, *PdfMiner*, and *IceCite*. The final tool listed, *IceCite*, is a method created by the authors.

## **Metrics**

In order to assess the effectiveness of the tools listed above, a number of metrics have to be established. Tools are evaluated based on how well they can extract semantic information from a PDF. The paper identifies the various kinds of semantic information as *word identification*, *word order*, *paragraph boundaries*, and *semantic roles* [4]. Determining whether words can be accurately detected in the right order despite spacing between the characters and lines is pertinent. The ability to detect where a paragraph begins and ends and what roles the text plays

in the literature (e.g. distinguishing the title, body text, figures, etc.) is also critical to gauging whether or not a tool has extracted the text as it is positioned.

The tools possess a variety of features that allowed for the creation of evaluation metrics such as: *identification of paragraph boundaries, identification of the reading order, identification of semantic roles, translation of ligatures, extraction of characters with diacritics as single characters, and the merging of hyphenated words* [4]. The metrics used in the evaluation process are defined as: *the number of spurious newlines, the number of missing newlines, the number of spurious paragraphs, the number of missing paragraphs, the number of reordered paragraphs, the number of spurious words, the number of missing words, and the number of misspelled words* [4].

## **Evaluation Process**

The method of evaluation of PDF text extraction tools was provided under <https://github.com/ckorzen/arxiv-benchmark>, the GitHub repository identified in the paper. This process includes generating a functional PDF extraction benchmark in order to utilize the tools. The benchmark consists of a diverse set of 12,099 scientific articles ranging in topics and publication dates. All of the articles included in the benchmark were extracted from *arXiv.org*, but are regenerated PDF files using the corresponding TeX files that *arXiv.org* provides. A benchmark generator was also created in the evaluation process, which was responsible for the generation of TeX files that distinguish logical text blocks and split them into *ground truth files*. The regenerated PDF files and *ground truth files* are compared to determine how accurately each PDF tool extracts the text. The GitHub repository includes instructions on how to access the benchmark generator and benchmark articles. All the files are available for download given the instructions.

Provided the repository has all of the tools necessary to complete the evaluation, copying the repository to my own environment proved an efficient process. Each tool and its respective files for operation are included within the repository. Custom files that identify methods of translating text files, the command needed to run the text extraction mechanism, and other helpful information and tips also represented the makeup of the evaluation process. The primary scripts, *extractor.py* and *evaluator.py*, is code used to extract and evaluate the text accordingly given a number of optional parameters.

## **Evaluation**

In order to conduct this evaluation on two newly identified PDF text extraction tools, PyMuPDF and PDFPlumber, I first sought to reproduce the evaluation. I followed the steps provided by the GitHub Repository to automatically generate text files and achieve the same evaluation results using all the tools. Only a handful of tools produced text files that corresponded to the desired output, while others produced blank files. I believe that this may be the result of configuration variances given that the repository is identical to the one provided by the authors. As a result of this, the next steps of modifying the existing tool configurations for the extraction process to custom fit the two new tools wasn't as anticipated.

Rather than use the *extractor.py* files to extract the text from the benchmark and create output text files, I simply designed a script for each tool that extracted the text from each article and directed the created text files to their respective folders with the desired formatting and file naming convention.

From this point, I was able to return to the instructions for the evaluation process using the *evaluator.py* file to compare with the ground truth files. After modifying the configuration

files for each new tool and running the *evaluator.py* script, I was able to retrieve evaluation results. The script prints the results in a format similar to the one presented in the table, providing absolute number and percentage of *affected words relative to the number of words in the ground truth files*, *the number of PDF files which could not be processed by a tool*, and *the average time needed to process a single PDF*. In the case of the spurious newlines and missing newlines which are included in the evaluation results, the percentage gives *the absolute number relative to the number of newlines in the ground truth files*.

The PDF text extraction tools were evaluated through ten distinct criteria. These criteria are the detection of the number of false newlines, missing newlines, false paragraphs, missing paragraphs, rearranged paragraphs, false words, missing words, misspelled words, unprocessed files, and the time needed to process a single PDF. They are defined by both an absolute number and percentage.

## **Results**

This study was able to replicate the results presented in [4] for the evaluation of pdf tools with a new algorithm that we designed following the inability to use “extractor.py” proposed in [4]. In addition, we evaluated 2 new algorithms, PyMuPDF and PDFPlumber using the same criteria. Results are presented in Table 1.

The evaluation results on the identified 14 PDF text extraction tools that this study performed were provided in [4] in a concise table. Table 1 contains the results from the original evaluation.

I reproduced the table and appended the results from the evaluation of PyMuPDF and PDFPlumber. The results are shown in tables 2. Both PyMuPDF and PDF Plumber produced some of the lowest error rates (1 and 3 unprocessed files respectively), but some of the highest

processing times for each PDF file (49 seconds and 57 seconds respectively). The category both tools performed best in, *the number of rearranged paragraphs*, produced low rates of error (0.1% and 2.7%). The category yielding the highest results for the two tools is the number of missing newlines, indicating that they are not as effective as other tools in properly identifying when a newline is supposed to be present within the text.

Tool	NL <sup>+</sup>	NL <sup>-</sup>	P <sup>+</sup>	P <sup>-</sup>	P <sup>↑↓</sup>	W <sup>+</sup>	W <sup>-</sup>
pdftotext	14 (16%)	44 (53%)	60 (29%)	2.3 (0.6%)	1.4 (1.9%)	24 (0.7%)	2.4 (0.1%)
pdftohtml	3.6 (4.3%)	70 (84%)	9.2 (31%)	4.2 (3.2%)	0.1 (0.1%)	16 (0.5%)	1.6 (0.0%)
pdftoxml	33 (40%)	20 (25%)	80 (31%)	1.8 (0.5%)	0.1 (0.1%)	21 (0.6%)	1.5 (0.0%)
PdfBox	3.0 (3.6%)	70 (85%)	7.6 (27%)	0.9 (0.2%)	0.0 (0.1%)	17 (0.5%)	1.5 (0.0%)
pdf2xml	33 (40%)	39 (48%)	44 (21%)	40 (30%)	7.8 (9.5%)	8.6 (0.3%)	3.6 (0.1%)
ParsCit	15 (18%)	39 (47%)	10 (10%)	14 (6.4%)	1.3 (1.8%)	16 (0.5%)	2.3 (0.1%)
LA-PdfText	5.5 (6.4%)	23 (28%)	4.8 (3.1%)	52 (73%)	2.9 (5.9%)	5.7 (0.1%)	6.1 (0.1%)
PdfMiner	32 (38%)	18 (21%)	84 (30%)	3.6 (1.0%)	1.4 (2.1%)	34 (1.0%)	2.6 (0.1%)
pdfXtk	7.9 (9.7%)	68 (84%)	12 (29%)	4.5 (3.5%)	0.1 (0.1%)	59 (1.8%)	6.1 (0.2%)
pdf-extract	95 (114%)	53 (64%)	99 (32%)	8.4 (3.1%)	4.1 (7.7%)	74 (2.1%)	41 (1.2%)
PDFExtract	9.5 (11%)	33 (40%)	28 (21%)	22 (25%)	0.8 (0.9%)	12 (0.4%)	2.8 (0.1%)
Grobid	9.5 (11%)	30 (36%)	7.5 (6.7%)	11 (15%)	0.0 (0.0%)	14 (0.4%)	1.6 (0.0%)
Icecite	3.4 (4.0%)	10 (13%)	6.2 (4.2%)	7.7 (5.5%)	0.1 (0.1%)	10 (0.3%)	1.7 (0.1%)
PyMuPDF	6.1 (7.3%)	93.4 (112.4%)	27.7 (38.6%)	28.3 (2.1%)	0.0 (0.1%)	16 (0.5%)	3.9 (0.1%)
PDF Plumber	45.5 (53.6%)	158.9 (189.1%)	61.4 (46.6%)	65.3 (25.9%)	4.4 (2.7%)	61.0 (1.7%)	5.7 (0.2%)

**Table 1:** The table presents the evaluation results from *A Benchmark and Evaluation for Text Extraction from PDF*. The evaluation results of PyMuPDF and PDFPlumber have been appended to the results of the original evaluation. The caption provided for the original table provides insights on the table characteristics: *”Summary of the evaluation results of 14 PDF extraction tools. The second column gives a summary of Table 1, for convenience. The evaluation results are given in columns 3-10, broken down into the criteria NL+: the*

*number of spurious newlines; NL-: the number of missing newlines; P +: the number of spurious paragraphs; P -: the number of missing paragraphs; P ↑↓ : the number of reordered paragraphs; W +: the number of spurious words; W -: the number of missing words; W ~: the number of misspelled words. For each criterion, its absolute value and a percentage is given, which is computed as follows: for NL+ and NL-, it is the absolute value divided by the number of newlines in the ground truth; for the other criteria, it is the number of affected words relative to the number of words in the ground truth files. The best values in each criteria are printed in blue and bold, the two worst values in red. The column ERR gives the aggregated number of PDF files where (a) the extraction process resulted in an error or (b) the runtime of the extraction process exceeded the timeout of five minutes. The column Tθ gives the average time needed to process a single PDF file, in seconds.” [4]*

On a smaller scale, experimenting with PyMuPDF and PDFPlumber has shown that these tools convert PDF files to text files. PyMuPDF was effective in my processing and manipulating of PDF files to develop a scientific corpus relative to the crystallization process.

PyMuPDF has proven to possess a reliable annotation detection tools that allow for the extraction of specific text by copying and pasting relevant lines of text into the comments of the PDF file. This tool will be used in the pre-processing of scientific literature to train a machine-learning model for extracting experimental conditions .

## **Conclusion**

The benchmark provided by *A Benchmark and Evaluation for Text Extraction from PDF* serves as a reliable source for articles to evaluate emerging PDF parsing tools. The evaluation I conducted showed that PyMuPDF and PDFPlumber are effective in extracting text, as they produced some of the lowest error rates in processing articles of all the tools, but of course possess faults. PyMuPDF and PDFPlumber are among the least effective at processing newlines, as they missed the most newlines of all the tools.

This process has also shown that, IceCite, the most effective tool evaluated, may not be scalable for all PDF articles, as it proved to only function with PDF literature regenerated from TeX files. Text extraction of standard PDF files may be best suited for the other identified tools.

This research has proven essential because scientific literature presents an abundance of information in the form of text, figures, and data that scientists and students often reference and digest during their researching efforts. In the particular discipline of biology, this research has been thin. The development of a model to automate text extraction from PDFs will be a great effort in the near future, while it begins with identifying the best available PDF parsing tools.

### **Acknowledgements**

Mentors - Dr. Line Pouchard & Dr. Gilchan Park (Computational Science Initiative of Brookhaven National Laboratory)

Contributor – Dr. Dale Kreidler (National Synchrotron Light Source II)

Science Undergraduate Laboratory Internships (SULI)

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## References

McKie, J. X. (2015). *PyMuPDF documentation*. PyMuPDF Documentation - PyMuPDF 1.18.16 documentation. <https://pymupdf.readthedocs.io/en/latest/>.

McKie, J. X. (2015). *pymupdf/PyMuPDF: Python bindings For Mupdf's rendering library*. GitHub. <https://github.com/pymupdf/PyMuPDF>.

Singer-Vine, J. (2015). *Jsvine/Pdfplumber: Plumb a pdf for detailed information about each char, rectangle, line, et cetera - and easily extract text and tables*. GitHub. <https://github.com/jsvine/pdfplumber>.

H. Bast and C. Korzen, "A Benchmark and Evaluation for Text Extraction from PDF," 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL), 2017, pp. 1-10, doi: 10.1109/JCDL.2017.7991564.

Cold Spring Harbor Laboratory. (2013). *biorxiv*. bioRxiv. <https://www.biorxiv.org/>.

Wikipedia contributors. (2021, July 12). Crystallization. Wikipedia. <https://en.wikipedia.org/wiki/Crystallization#:~:text=Crystallization%20or%20crystallisation%20is%20the,deposition%20directly%20from%20a%20gas>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

θ



## **Neutronics Modeling For Intentional Forensics**

Shia Jones

Department of Civil Mechanical Engineering Tech. & Nuclear Engineering, South Carolina State University, Orangeburg, SC 29115, USA

Caroline Sears

Picker Engineering Program, Smith College, Northampton, MA 01063, USA

Eric Schess

A. James Clark School of Engineering, University of Maryland, College Park, MD 20740, USA

Arantzazu Cuadra-Gascon

Nuclear Science and Technology Department, Brookhaven National Laboratory, Upton, NY 11973-5000, USA

## I. ABSTRACT

The security of nuclear fuel is a top concern in the nuclear sector of the United States. One security concern of note is the possibility of fuel being diverted, stolen, or misplaced. As part of the international effort to deter misuse and theft of nuclear fuel, Brookhaven National Laboratory has devoted resources to the research of intentional forensics. This effort includes screening materials that would enhance the traceability of Uranium Dioxide fuel rods. Our research explores the possibility of using taggants in nuclear fuel rods via coatings, or by adding dopants in the fuel or cladding. The candidates chromium, chromia, tungsten, molybdenum, nickel, zirconium, hafnium, and vanadium were screened using neutron transport (neutronics) codes for their viability to determine the fuel's place of origin in the event that fuel were to be lost and then recovered. Key metrics for this research include the effect of candidate doping or coating on the fuel's burn cycle length, reactivity coefficients, sensitivity to cross-section libraries and uncertainties, and the survivability of the taggant during burnup.

## II. INTRODUCTION

As of 2021, close to 20% of the United States' electrical power supply is generated by the 93 U.S. Nuclear Regulatory Commission (NRC) licensed nuclear power plants in current operation. 63 of those plants employ the use of Pressurized Water Reactors (PWR), which makes PWRs the most common type of industrial nuclear reactor<sup>1</sup>. The core of a standard PWR contains 157 or 193 fuel assemblies, which are each composed of a 17-pin array. Each pin cell in the array is filled with either a fuel rod, control rod, or instrumentation rod. The fuel rod, in which the fissionable material resides, is a cylindrical fuel pellet surrounded by a layer of cladding, with a gap of helium between the two materials allowing for the expansion and contraction of the pellet and providing a space for the storage of fission gasses produced during burnup. The rods are cylindrical, but fit into a square assembly, leaving space for the flow of borated water (the moderator), which cools and controls fission in the fuel rods.

PWRs typically use enriched Uranium Dioxide ( $\text{UO}_2$ ) fuel, as it is a well studied and predictable fissionable material that produces significant heat. The heat produced during the fission reaction is then processed via a turbine generator and converted to the desired form of energy, electricity. The fuel is replenished by using a process referred to as a "three-batch burn cycle". This is a fuel replacement system designed to keep the reactor running continuously with consistent neutron flux across the core. A burn cycle is the time frame in which the reactor is allowed to run before. The first batch in the cycle is the "fresh" batch of fuel, never burnt before. The second and third batches have each been in the reactor for one and two burn cycles, respectively. Once the fresh batch has completed its first cycle of depletion in the reactor, that prompts the reactor management to remove the oldest batch, now three-times burned, and make room for a new set of fresh fuel rods, and so repeats this process. Each of these refueling cycles is ideally 18 months (or 520 days) long. The timeline of a typical PWR three-batch burn cycle allows for efficient depletion of the fuel as well as predictability of the reactor maintenance, as refuelings are best performed during Fall or Spring when energy usage is not at peak demand in the U.S.

As with any type of fuel source, the large amount of energy stored in enriched  $\text{UO}_2$  makes it a useful but potentially dangerous material. Managing  $\text{UO}_2$  fuel requires responsibility and care throughout its lifecycle, from its design and production stage until it is spent and put into waste storage. One aspect of nuclear fuel management is knowing the location

of a fuel rod at all times to ensure security. The concept of intentional forensics, a design tactic where products are created with built-in tracking measures, has been suggested to aid in the traceability of nuclear fuel<sup>2</sup>. Intentional forensics suggests if a chemical tagging agent or “taggant” were integrated into the design of the fuel rod, one could distinguish fuel originating from different facilities and preemptively deter misuse and theft.

Potential taggant candidates are desired to have little to no impact on reactor performance. Thus, taggants are chosen for viability screening depending on their ability to be neutron-transparent (small neutron absorption cross section), the amount and quality of nuclear data available, and their material properties such as water insolubility and the ability to withstand high temperatures. The cross section size and material properties directly affect where the taggant can reasonably be placed in the fuel rod to minimize its impact, with the main two areas of interest being the fuel pellet and the cladding. Taggants chosen to act as a fuel pellet dopant must be as neutron transparent as possible to minimize the probability of the taggant absorbing more neutrons than the fuel itself. Clad dopants have somewhat more leeway in this regard, and could potentially be functional even with taggants of slightly lower neutron transparency, but must be able to withstand the presence of the moderator, as it will be in direct contact with the borated water. In order to make these decisions about our model, it is necessary to consider the amount of nuclear cross section data available for each taggant because it is central to performing neutronics calculations and ensures the conclusions being made about the element during research are accurate. These attributes are important to take into account to ensure the viability of the material and to prevent introducing system failures.

The suggestion of intentional forensics poses two fundamental questions, the first being what type of elements are appropriate for use as taggants, and subsequently, how do those taggant candidates affect the functioning of a PWR? Our study focuses on the exploration of these questions. The guiding design constraint of our taggant screening study was to find candidates which minimize the effects of both the taggant on the reactor as well as the reactor on the taggant. We will be screening for potential effects through three metrics: the burn cycle length, the coefficients of reactivity (safety coefficients), and the survivability of taggants’ isotopics.

### III. METHODS

Seven elements and one compound were identified as taggant screening candidates based on the desired characteristics and metrics described (availability of and size of nuclear cross sections, material properties). Each candidate was evaluated as to whether it would be best suited for use as a fuel dopant, clad dopant, coating, or multiple areas. The candidates tungsten, nickel, and zirconium were selected to be clad dopants, as each is water insoluble and could thus remain viable while in contact with the moderator. Zirconium was chosen in particular because it is already a main component in the Zircaloy cladding. Molybdenum, hafnium, chromium, and vanadium were selected to be fuel dopants. Molybdenum was of particular interest as it is already a common nuclear fuel component, and is a fission product of uranium. Chromia was tested in two modes, as a coating and a fuel dopant. Different areas of interest were explored for each type of taggant. For the candidates with the largest average cross sections (tungsten, hafnium, and vanadium), a concentration study was conducted to determine the threshold amount of taggant which could be added to the fuel/clad without compromising efficiency. Concentration is defined here as the mole fraction relative to the amount of moles of  $^{235}\text{U}$  in the fuel, and is given in parts per million (ppm). The concentrations of taggant used as clad dopants were as low as 50 ppm and as high as 10,000 ppm. All candidates tested as fuel dopants were screened with a concentration of 1000 ppm. Additionally, a survivability study was conducted for all fuel dopants to track the isotopic signatures of the dopants in the fuel throughout the burn time.

Two neutronics codes, SCALE 6.2 (Standardized Computer Analyses Licensing Evaluation)<sup>3</sup> and MCNP6 (Monte Carlo N-Particle Transport Code)<sup>4</sup>, were used for our screening process. These codes model the fuel inside a reactor at given conditions. SCALE is a deterministic code package from which we primarily used the Transport Rigor Implemented with Time-dependent Operation for Neutronic Depletion (TRITON) code. This code calculates K-effective values, allowing us to simulate the burn cycle of a 2D, quarter-array fuel assembly. K-effective is the number of neutrons present after a given generation over the number of neutrons present before that generation. We use K-effective to calculate our primary reactor-based screening metrics, burn cycle length and reactivity coefficients. The burn cycle length is found as the intercept over the slope of the linear relationship between burnup and reactivity, as stated by the Linear Reactivity Model<sup>5</sup> (see EQN. 1). Due to the

inevitable leakage, some present neutrons will not induce fission . This slightly hindered reactivity is accounted for in the equation with an added leakage factor of 0.03. (see EQN. 2)

$$\rho = \rho_0 + AB \quad (1)$$

$$\rho = \frac{k_{\text{eff}1} - k_{\text{eff}2}}{k_{\text{eff}1} * k_{\text{eff}2} + 0.03} \quad (2)$$

Also utilizing K-effective, three different reactivity coefficients were calculated: the Moderator Temperature Coefficient, The Doppler Coefficient (also known as the Fuel Temperature Coefficient), and the Boron Coefficient. These coefficients show the change in reactivity per degree temperature changed (in the case of the temperature coefficients) or per change in concentration (in the case of the Boron coefficient). We plotted all three coefficients against burnup, and then against temperature or concentration, for the two types of coefficients, respectively.

Additionally, we added the Origen-S Post-Processing Utility and Plotting Program for SCALE (OPUS) tool to our TRITON inputs to calculate a list of the 500 most abundant isotopes and their concentrations at each time step during the burn cycle. We conducted our survivability studies by graphing the concentrations of select isotopes over time. An ideal candidate would retain its isotopic concentrations throughout the burn cycle to ensure it is identifiable, and a significant change in the isotopics would ruin that functionality. Separate OPUS files were produced for the fuel pellet and the cladding, although the scope of this project only encompassed survivability studies in which the dopant was present in the fuel pellet. Clad dopants were identified as a lesser concern regarding survivability due to the lower temperatures experienced in comparison to the fuel pellet.

Lastly, the Tools for Sensitivity and Uncertainty Analysis Methodology Implementation (TSUNAMI) package of SCALE was used for finding the sensitivity of the K-effective to the covariances of the neutron absorption cross section library data. This was used to determine which nuclear cross section covariances had the largest effect on our K-effective values.

Conversely, MCNP is a Monte-Carlo code package, in which we calculated the K-effective values for single fuel-pins. While we used the K-effective values from TRITON to calculate our primary screening metrics, we used the K-effective values from MCNP to compare different Evaluated Nuclear Data File (ENDF) libraries. MCNP is more versatile than

SCALE in regards to its ability to use many different data libraries. While SCALE 6.2 can only use ENDF/B-VII.1<sup>7</sup>, MCNP can use any version of ENDF, including the current version ENDF/B-VIII.0<sup>6</sup>. Two versions of the same MCNP input were run using these respective libraries in order to see if the results were consistent so that we could conclude whether the effect on K-effective is negligible.

#### IV. RESULTS

Burn cycle length was the key component used in evaluating the taggants' effect on the reactor. Any taggants which caused the burn cycle length to dip significantly lower than 18 months (520 days) were not considered viable because of their effect on the neutronic operation of the reactor. Based on our data, the candidates eliminated by this criteria are tungsten and hafnium in concentrations greater than 100 PPM and 500 PPM, respectively (see FIG. 1). All other taggant candidates had viable burn cycle lengths and thus neutronic effects on the reactor are not a concern.

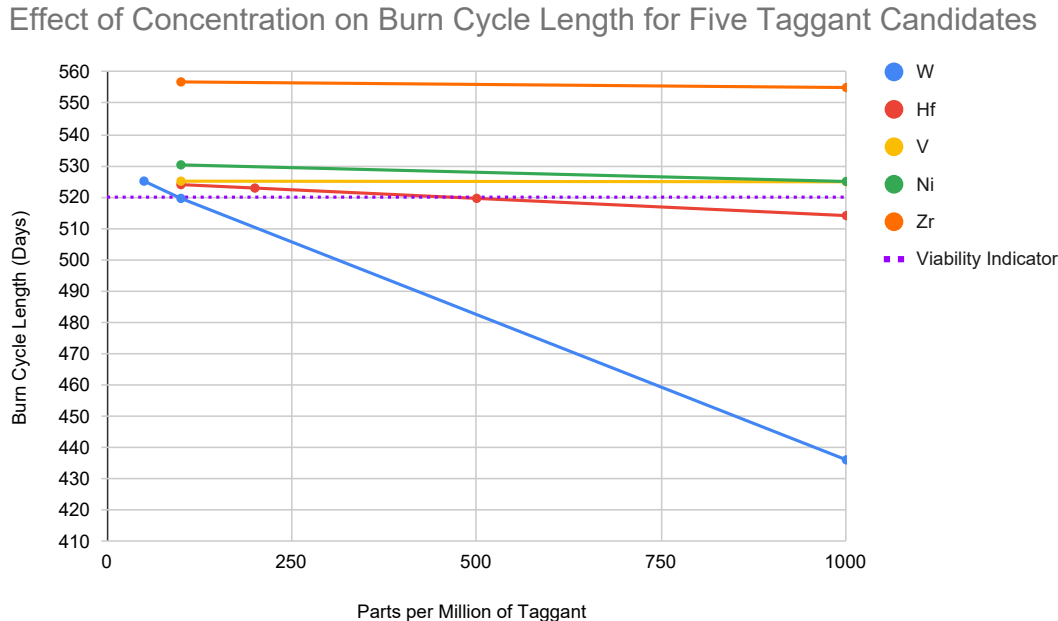


FIG. 1. Graph of burn cycle lengths resulting from SCALE burnup models with respective taggant candidates, correlated with ppm concentration of taggant used in model.

For the candidates with suitable burn cycle lengths, safety becomes the next most important metric for screening. We expected all reactivity coefficient plots for viable candidates

to show us negative values, due to the fact that increasing the moderator and fuel temperatures decreases their respective densities, resulting in diminished reaction rates. We expect a similar correlation in the case of the boron coefficient. As a neutron poison, boron will absorb neutrons from the present population, limiting the amount of neutrons available for fission. For all taggants at 1000 ppm concentration, our expectation of negative values held true, and shows us the reactivity coefficients are within safety parameters for all candidates (see FIG. 2, 3, and 4)

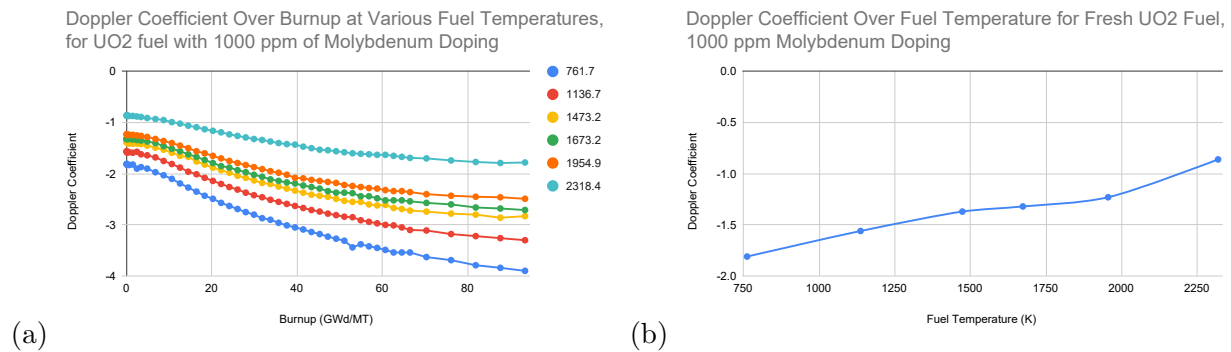


FIG. 2. Doppler coefficient (fuel temperature coefficient) plots, sample shown for molybdenum fuel dopant case, calculated with SCALE. (a) Doppler coefficient plotted against burnup for respective temperature branches of the fuel assembly, (b) Doppler coefficient plotted against initial temperature of each branch for initial fuel conditions.

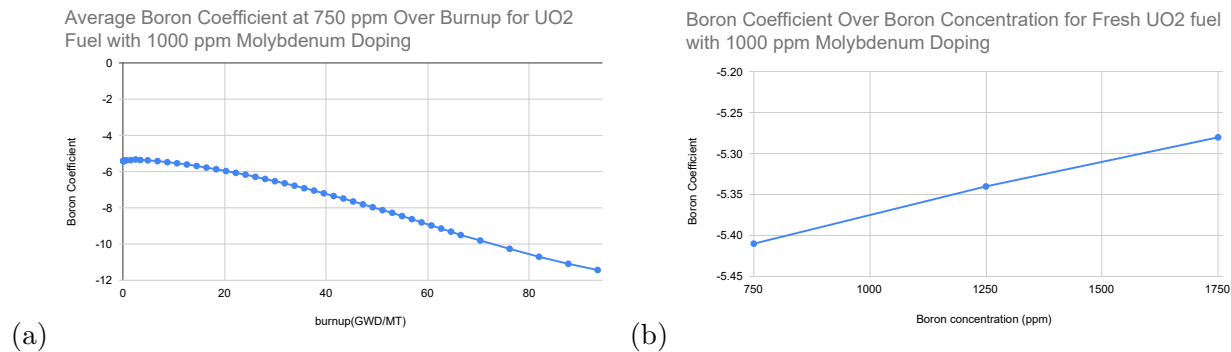


FIG. 3. Boron coefficient plots, sample shown for molybdenum fuel dopant case. (a) Boron coefficient plotted against burnup for 750 ppm concentration of boron in moderator, (b) Boron coefficient plotted against boron concentration for initial fuel conditions

The metric of isotopic survivability indicates whether the isotopic signature of a taggant will be preserved throughout the fuel rod's lifecycle, including after its depletion. Most taggants retained consistent isotopic concentrations with the exception of molybdenum and hafnium, which underwent a significant change in isotopics during the burn cycle. In the case of molybdenum (see FIG. 5), this can be attributed to the fact that many of its isotopes are fission products of uranium. The fission process increased the number of molybdenum



Moderator Temperature Coefficient Related to Burnup for 1000 ppm Molybdenum Fuel Dopant Model

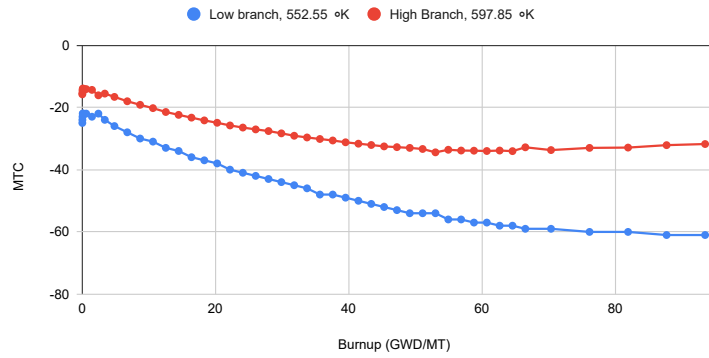


FIG. 4. Moderator Temperature coefficient plot, sample shown for molybdenum fuel dopant case. Moderator Temperature coefficient plotted against burnup for two moderator temperatures, the "high branch", 597.85°K, and the "low branch", 552.55°K, relative to the reference branch, 576.55°K.

Natural Molybdenum Isotopic Vectors Over Burn Cycle (Triton), For UO<sub>2</sub> fuel with 1000 ppm Molybdenum Doping

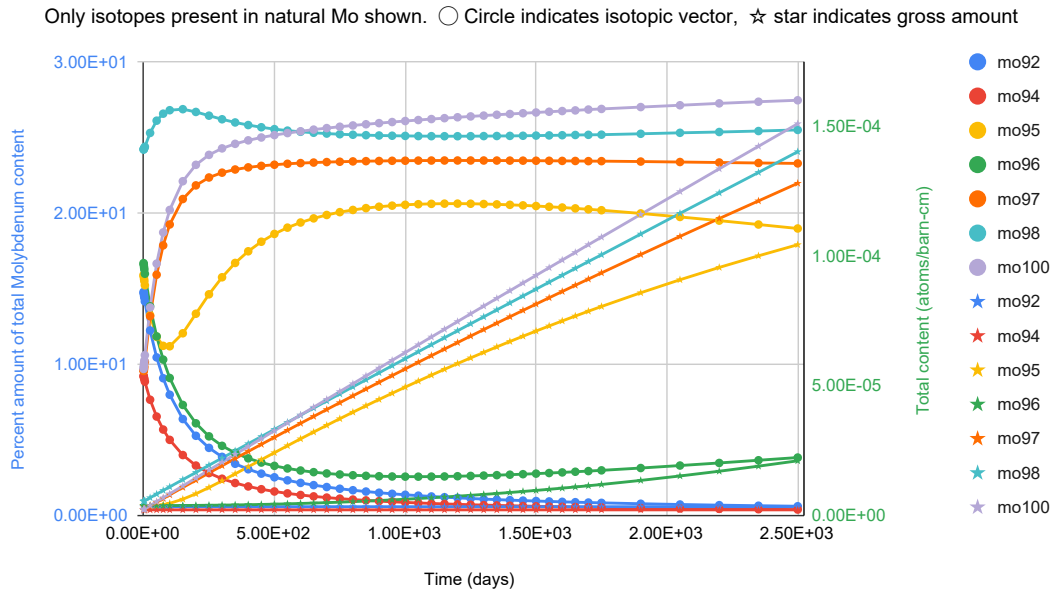


FIG. 5. Survivability results of 1000 ppm natural molybdenum during UO<sub>2</sub> fuel burnup. Isotopic vector (percent amount over total molybdenum content) shown on left vertical axis, total amount of each isotope (in ppm) shown on right vertical axis.

atoms present in the fuel pellet over time, including the addition of isotopes not found in the natural molybdenum taggant used in our model. We have interpreted this observation of molybdenum to suggest that fission products are not ideal taggant candidates, because the fission reaction hinders their ability to pass the survivability test.

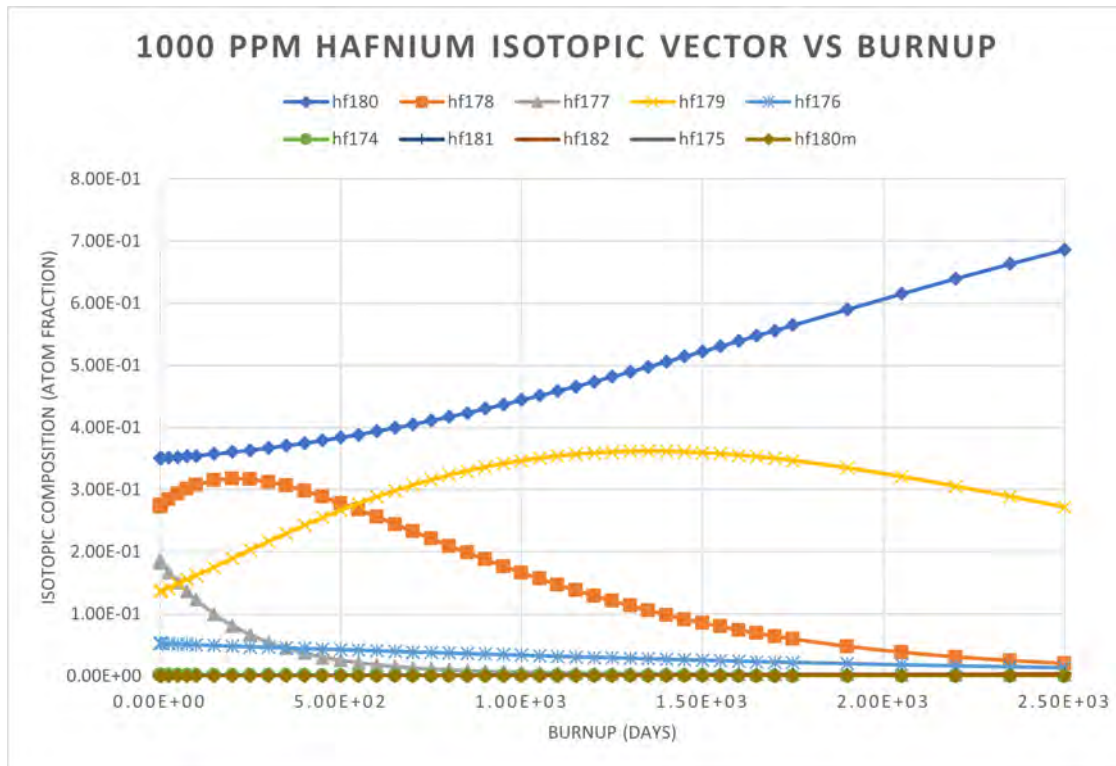


FIG. 6. Survivability results of 1000 ppm natural hafnium during UO<sub>2</sub> fuel burnup. Isotopic vector (percent amount over total hafnium content) shown.

While hafnium also did not pass the survivability test, it is not a fission product, which suggests there are other factors besides fission product status that affects the isotopic composition of a potential taggant over a burn cycle. Notably, the hafnium isotope concentrations did not universally increase, as they did in the case of molybdenum (see FIG. 6). This indicates the change in isotopes was caused by the transformation of hafnium isotopes, with isotopes such as <sup>176</sup>Hf, <sup>177</sup>Hf, and <sup>178</sup>Hf, absorbing neutrons and transforming into larger isotopes (such as <sup>180</sup>Hf) which we observed to have increasing concentrations. This suggests absorption cross section can be used as an indicator for a taggants' survivability.

In addition to our taggant viability metrics, we observed two other peripheral quantities related to the neutron absorption cross section data, the effect of the cross-section covariances and the ENDF library version on K-effective. We were interested in quantifying if and how the nuclear data, particularly in regards to the differences between ENDF/B-VIII and ENDF/B-VII.1, had an impact on our K-effective calculations. Our results show noticeably but consistently lower values of K-effective for the MCNP data calculated with ENDF/B-VIII (see FIG. 7). Due to the consistency of results, it appears the difference in K-effective is

not a direct result from the added taggant, but rather the uranium cross section data. This is corroborated by the TSUNAMI file outputs which consistently show the largest contributor to the covariance calculation for all taggants were  $^{238}\text{U}$  and  $^{235}\text{U}$ .

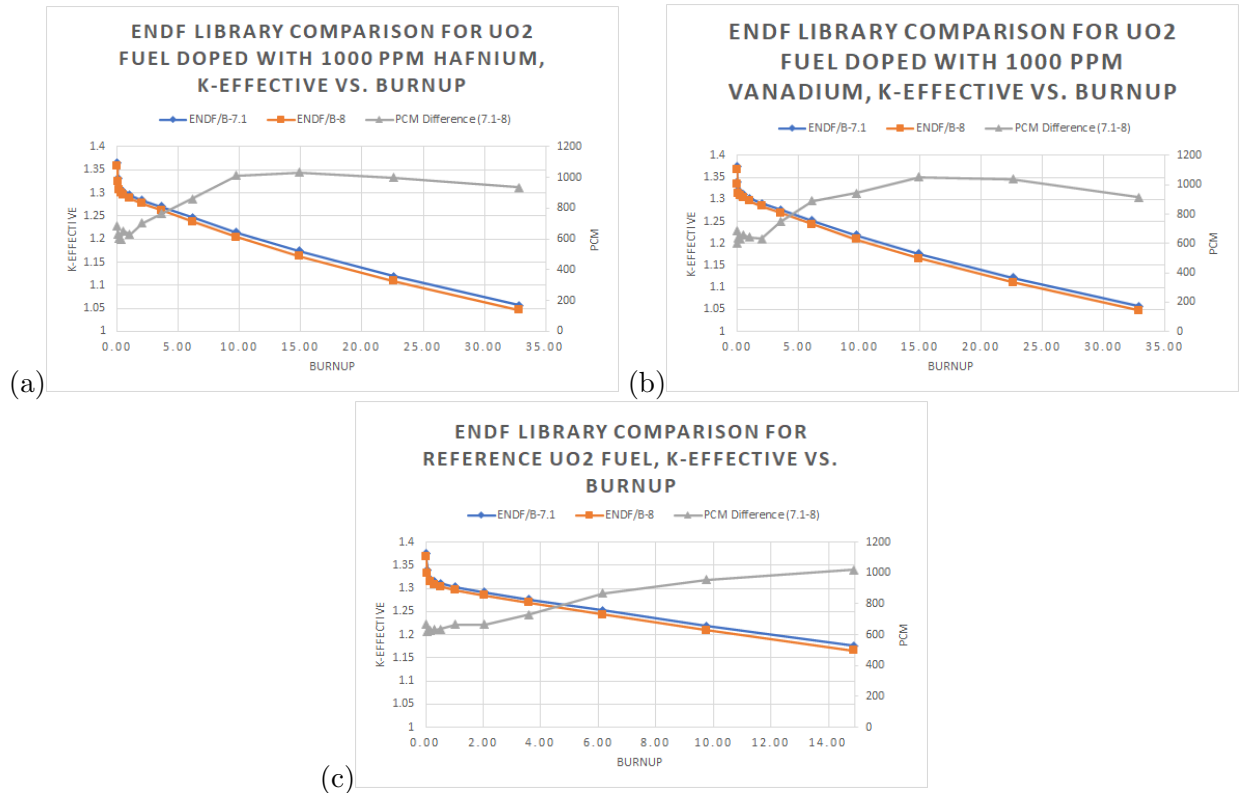


FIG. 7. ENDF/B-VII.1 vs. ENDF/B-VIII.0 comparisons for two taggants, (a) Hafnium, and (b) Vanadium, with (c) un-tagged UO<sub>2</sub> fuel as a reference case

## V. CONCLUSIONS AND DISCUSSION

The primary conclusion of our study is that the largest indicator of taggant candidate success is its neutron absorption cross section. This property directly impacts a taggant's ability to be neutron-transparent, have consistent isotopics throughout a burn cycle, and sets the upper limit on its potential concentration. For the candidates with the largest cross sections, Tungsten and Hafnium, the viable concentration was capped at a much smaller amount than other elements with smaller average cross sections, which had the potential to be viable in far higher concentrations with little impact on the fuel's burn cycle length. Overall, our data supports the viability of taggant candidates with smaller cross sections, but we do believe candidates with larger average cross sections can be suitable in sufficiently

low amounts and do not need to be discounted entirely from further study.

An additional takeaway from this work is the recommendation that fission products ought not to be considered as viable fuel dopant candidates, as shown through the case of molybdenum. However, we must note that our screening process solely considered molybdenum as a fuel dopant and this recommendation may not be relevant for taggants used as a clad dopant, where the creation of fission products would have a significantly lesser impact on the isotopics. We believe further screening of molybdenum and other candidates that are fission products could be conducted under the stipulation that they must be screened as a clad dopant.

We believe further research is required and may be guided by the results of our screening study. For all candidates which survived the screening issue-free, they can move on to further testing, in the form of more advanced models or physical tests in reactors. These candidates are tungsten, zirconium, and nickel in the clad, and vanadium and chromium in the fuel, with chromia being viable in all areas of the fuel rod, including as a coating. As hafnium and molybdenum both failed their survivability tests, further research of these taggants would instead consist of further screening of their survivability in the cladding, which operates at a lower temperature and may help maintain the taggant's isotopic signature.

Due to the reliance of this screening process on the ENDF cross section libraries, it cannot be overstated the importance of having sufficient, high-quality nuclear data with which to perform studies such as these. As shown through the sensitivity of the K-effective values on the cross section libraries of  $^{238}\text{U}$  and  $^{235}\text{U}$ , the accuracy and scope of these calculations are both highly influenced by and potentially limited by the current landscape of nuclear data. Our conclusions are based on the data currently available, and further screening studies could be recommended in tandem with new data evaluations.

## ACKNOWLEDGMENTS

We would like to thank our mentor Arantzazu Cuadra-Gascon, and also to thank Dave Brown from the NNDC and Arnie Aronson, Michael Todosow and Lap-Yan Cheng from the NSTD for their help on this project. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program

(SULI). This project was also supported in part by the Brookhaven National Laboratory (BNL), National Nuclear Data Center, under the BNL Supplemental Undergraduate Research Program (SURP).

## REFERENCES

- <sup>1</sup>“Power reactors,” <https://www.nrc.gov/reactors/power.html>, accessed: 08-04-2021.
- <sup>2</sup>N. R. Brown, M. Todosow, and A. Cuadra, “Screening of advanced cladding materials and un-u3si5 fuel,” *Journal of Nuclear Materials* **462**, 26–42 (2015).
- <sup>3</sup>B. T. Rearden and E. M. A. Jessee, “Scale code system, ornl/tm-2005/39, version 6.2,” (2016).
- <sup>4</sup>C. J. Werner *et al.*, “Mcnp users manual-code version 6.2,” Los Alamos National Laboratory (2017).
- <sup>5</sup>M. J. Driscoll, T. J. Downar, and E. E. Pilat, *The linear reactivity model for nuclear fuel management* (Amer Nuclear Society, 1990).
- <sup>6</sup>D. A. Brown, M. Chadwick, R. Capote, A. Kahler, A. Trkov, M. Herman, A. Sonzogni, Y. Danon, A. Carlson, M. Dunn, *et al.*, “Endf/b-viii. 0: The 8th major release of the nuclear reaction data library with cielo-project cross sections, new standards and thermal scattering data,” *Nuclear Data Sheets* **148**, 1–142 (2018).
- <sup>7</sup>M. Chadwick, M. Herman, P. Obložinský, *et al.*, “ENDF/B-VII.1 nuclear data for science and technology: Cross sections, covariances, fission product yields and decay data,” *Nuclear Data Sheets* **112**, 2887 – 2996 (2011), special Issue on ENDF/B-VII.1 Library.

## VI. APPENDIX

TABLE I. Burn Cycle Length for All Taggant Candidate Models

Candidate	Concentration	Location	Burn Cycle Length	Average Neutron Absorption Cross Section
-	ppm	-	Days	bn
Ref.	N/A	N/A	526.8	N/A
Mo	1000	fuel	530.32	0.2
Cr	1000	fuel	530.7	0.01
W	50	clad	525.12	1
W	100	clad	519.6	1
W	1000	clad	436.1	1
W	10000	clad	-155	1
Hf	100	fuel	524	1
Hf	200	fuel	522.9	1
Hf	500	fuel	519.6	1
Hf	1000	fuel	514.1	1
Hf	10000	fuel	417	1
V	100	fuel	525.1	0.1
V	1000	fuel	524.9	0.1
V	10000	fuel	522.8	0.1
Ni	100	clad	530.3	0.09
Ni	1000	clad	525	0.09
Zr	100	clad	556.8	0.08
Zr	1000	clad	555	0.08
Chromia	1000	Fuel	525.1	N/A
Chromia	1000	Clad (full)	659.8	N/A
Chromia	1000	Clad (half)	364.3	N/A

# Simulation and analysis of neutrino fluxes from the Large Hadron Collider

Karan Kumar<sup>1</sup> and Milind Diwan<sup>2</sup>

<sup>1</sup>Physics Department, Stony Brook University, Stony Brook, New York, 11794

<sup>2</sup>Mentor, Physics Department, Brookhaven National Laboratory, Upton, New York, 11973

August 11, 2021

## Abstract

A large flux of neutrinos is expected in the forward direction of the beam axis for pp collisions [1] at the Large Hadron Collider (LHC) at CERN. Several experiments have recently been proposed at CERN to detect these neutrinos and discussion has started on the possibility of building a Forward Physics Facility grouping many of them. We have included next-to-leading order (NLO) QCD radiative correction terms in our calculation of this flux. We have also studied the effect of a non-perturbative Gaussian intrinsic  $\langle k_T \rangle$ . This  $\langle k_T \rangle$  effect mimics the contribution from missing higher-order terms in QCD calculations. We present the study of uncertainties due to scale variations and Parton Distribution Function (PDF) variations in the production rate of  $D_s^\pm$  and tau neutrinos in the far-forward production region at the Large Hadron Collider.

## Introduction

Neutrinos are one of the fundamental particles that make up the universe. They are one of the most abundant particle in the universe and are also the least understood. The neutrino is a subatomic particle that resembles an electron but is electrically neutral and has a very small mass compared to other elementary particles. Neutrinos belong to the lepton family in the standard model. Lepton is a family of particles that doesn't experience strong interactions. Out of the four fundamental forces, neutrinos interact only through the weak force and gravity. As the weak nuclear force is very short-range and gravitational interaction of the neutrino is extremely weak, they rarely interact with normal matter. There are trillions of neutrinos passing through us every second.

Neutrinos come in three flavors: electron, muon, and tau neutrinos. They are labeled after their charged partners within the Standard Model. In charged current neutrino interactions in matter, neutrinos of a given type result in the emission of their charged partner. Out of the three flavors, tau neutrinos are the least studied as there is not sufficient data available to study them. The collision of proton beams at 14 TeV at the Large Hadron Collider (LHC) produces a large flux of hadrons. Some of the hadrons produced in the proton-proton collisions decay further into neutrinos. In particular, the decays of  $D_s^\pm$  produce a large flux of tau neutrinos in the forward direction of the collision; however, there have been no experiments to observe these neutrinos. A new set of experiments have been proposed at CERN[2]. This paper is about the calculation of the flux of neutrinos in the forward direction at the LHC. Along with the flux we study the uncertainties in our predictions due to Parton Distributions Functions. The forward region can be understood with the kinematic variable called pseudo-rapidity. Pseudo-rapidity ( $\eta$ ) is a geometric quantity and is a function of the angle  $\theta$  with respect to the collision axis as seen in Fig.1a. In this project, we are exploring the region with  $\eta > 6.9$ ,  $7.2 < \eta < 8.6$ ,  $8.0 < \eta < 9.2$ , and  $\eta > 8.9$ , which approximately covers the forward detector coverage for the proposed experiments at CERN.

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \quad (1)$$

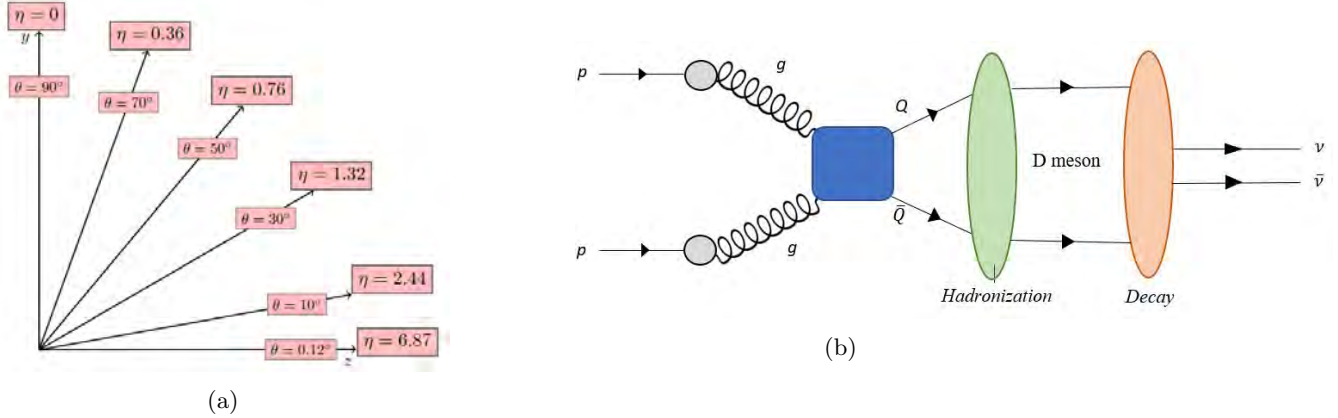


Figure 1: (a). The relation between  $\eta$  and  $\theta$ : An angle of zero is along the beam axis. Generally, particles in the high pseudorapidity regime escape through space in the detector along with the beam axis (forward direction), and this project includes the study of tau neutrino in the forward direction. (b). Proton-proton collision: Gluons from colliding proton interact during the collision. Gluon-Gluon interaction produce a quark and anti-quark which further produces hadrons like  $D_s^\pm$  meson.  $D_s^\pm$  meson decay into tau neutrinos.

There are two problems that arise in the Quantum Chromodynamics (QCD) perturbation theory that is used to calculate meson production in pp collisions: ultra-violet (UV) and infrared (IR) divergences. To include the higher-order term in perturbation theory, we come across Feynman graphs with closed loops, that are associated with energy unbounded from above. Because of unconstrained energy, the integral associated with such Feynman diagrams tend to diverge. Such divergences are called UV divergences. These divergences are not physical. The UV divergence are cured by introducing the renormalization factor ( $\mu_R$ ). We also encounter Feynman diagrams which include massless particles of energy approaching zero. The integral of Feynman diagrams including massless particles with zero energy also diverge. The divergence due to massless particles is known as Infrared(IR) divergences. The problem of IR divergence is solved by introducing the factorization factor ( $\mu_F$ ). The renormalization and factorization scales ( $\mu_R, \mu_F$ ) are defined to be the factors ( $N_R, N_F$ ) multiplied by the transverse mass  $m_{T,2}$ .

$$m_{T,2} = \sqrt{4m_c^2 + p_T^2},$$

where  $p_T$  is the magnitude of the transverse momentum of the charm quark and  $m_c$  is the charm quark mass.

The other variable in our calculation is the Parton Distribution Function (PDF). Parton, name given by Richard Feynman, refers to particle (quarks and gluons) constituents within the protons, neutrons and other hadrons. In the proton-proton collision, what actually happen is that these partons collide with each other as seen in Fig.1b. The colliding parton carries a fraction of the momentum of the proton. Parton distributions functions are momentum distribution functions of the partons within the proton. At leading order in pQCD, they are probability densities to find a parton with momentum with fraction  $x$  at an energy scale of  $\mu^2$ . We can calculate the hadronic cross section using these PDF.

## Theory

The most important concept in collisions of subatomic particles is their cross section. The word cross section is first introduced in mathematics as a intersection between a plane and a three dimensional



object. In physics, the word cross section also has units of area, but comes from a different consideration. The cross section with units of area governs the probability that two particles will collide or interact to produce a certain outcome and is denoted by  $\sigma$ . For example, the total cross section for production of tau neutrinos could be written as:

$$\sigma = \frac{\# \text{ of tau neutrinos produced per unit time}}{\text{Luminosity of protons per unit area per unit time}} \quad (2)$$

The other important term is Luminosity. In the above formula, it can be seen that Luminosity is the ratio of the number of events detected ( $N$ ) in a certain time ( $t$ ) to the cross-section. Using cross section and luminosity, we can calculate the number of events ( $dN/dt$  events/sec) by the following:

$$dN/dt = L \times \sigma \quad (3)$$

As mentioned earlier, we include NLO QCD corrections to the heavy-quark (HQ) production cross section. The HQ production cross section under perturbative QCD is as follow [4]:

$$E \frac{d^3\sigma}{dp^3} = \sum_{i,j} \int dx_1 dx_2 f_i^{H_1}(x_1, \mu_F^2) f_j^{H_2}(x_2, \mu_F^2) \left[ E \frac{d^3\hat{\sigma}_{ij}(x_1 P_{H_1}, x_2 P_{H_2}, p, m^2, \mu_F^2, \mu_R^2)}{dp^3} \right] \quad (4)$$

where  $f_i^{H_1}(x_1, \mu_F^2)$  and  $f_j^{H_2}(x_2, \mu_F^2)$  are parton distribution functions (PDFs),  $\mu_F^2$  and  $\mu_R^2$  are factorization and renormalization scales, respectively. As we discussed before, we need to account for mean transverse momentum in our calculation. We use a Gaussian approximation for transverse momentum in 2 dimensions.

$$f(\vec{k}_T) = \frac{1}{\pi \langle k_T^2 \rangle} e^{-\frac{k_T^2}{\langle k_T^2 \rangle}} \quad (5)$$

After including the  $k_T$  effect and integrating over it, the heavy quark production cross section becomes:

$$E \frac{d^2\sigma}{dp_z d^2p_T} = \int d^2k_T \int d^2p'_T f(\vec{k}_T) E \frac{d^2\sigma}{dp_z d^2p'_T} \delta^2(\vec{p}_T - \vec{p}'_T - \vec{k}_T) \quad (6)$$

The theoretical evaluation of the production of heavy quarks like charm has been studied for quite a long time, and is already implemented in a computer program called HVQ[5, 6] using the FORTRAN language. We used this program with some modifications to run simulations in our study. The HVQ code uses the Vegas algorithm [7] to calculate the integrals for the heavy quark production cross section. As there is no data available in the forward direction for production of particles, we use LHCb data for the production of  $D_s^\pm$  to compare with our theoretical predictions. We provide predictions by varying three parameters:  $k_T$ ,  $\mu_F^2$  and  $\mu_R^2$ . By varying these parameters, we tried to fit the data with our predictions using different transverse masses as discussed above. From the previous studies, we found that  $k_T = 0.7$  GeV fit well with the data [8, 9]. The scales variations ( $N_R, N_F$ ) in  $(\mu_R, \mu_F) = (N_R, N_F)m_{T,2}$ , where  $m_{T,2} = \sqrt{4m_c^2 + p_T^2}$  are [(1, 1), (0.5, 0.5), (2, 2), (0.5, 1), (1, 0.5), (2, 1), (1, 2)] with (1,1) as central scale choice. By examining the scale variations, we find the best central scale assumptions of transverse mass with  $k_T$ . In our studies, we are using PROSA [10] PDF. The PDF are created using data from Deep Inelastic Scattering (DIS) of leptons and fit with different models and parameters. PROSA contains 40 variations that account for PDF fit uncertainty, PDF parameter uncertainty, PDF model uncertainty. The fit uncertainties come from the data that is used to create the PDF, parameter uncertainties arise from the parameter used to fit the data, and model uncertainties originate from the assumed model for the fit for the PDF. The total uncertainties are obtained by adding fit, model, and parameter uncertainties in quadrature.

## Results

In Fig. 2, we compare our prediction of  $D_s^\pm$  double differential cross section with the LHCb data. By comparing figures 2a and 2b, we conclude that transverse momentum  $m_{T,2} = \sqrt{4m_c^2 + p_T^2}$  with

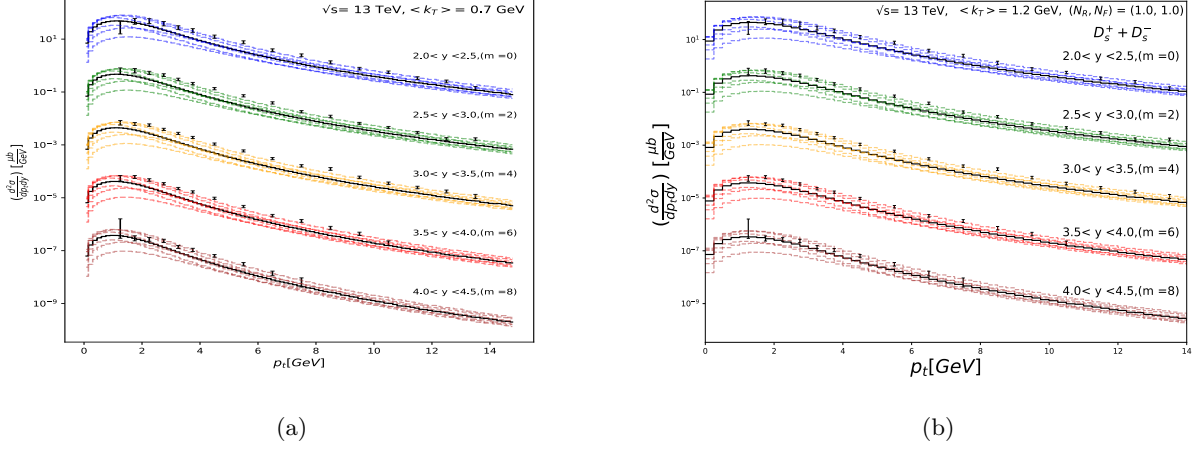


Figure 2: Comparison between our predictions and LHCb experimental data on double-differential cross section for  $D_s^\pm$  production. Data and predictions for different  $y$  bins are shifted by  $10^{-m}$  where values of  $m = 0, 2, 4, 6$  and  $8$ . The Fig. (a) refer to the central scale  $N_R = 1.0$ ,  $N_F = 1.0$  with  $m_{T,2} = \sqrt{4m_c^2 + p_T^2}$  with  $\langle k_T \rangle = 0.7$ . The Fig. (b) shows to the central scale  $N_R = 1.0$ ,  $N_F = 1.0$  with  $m_{T,2} = \sqrt{4m_c^2 + p_T^2}$  with  $\langle k_T \rangle = 1.2$ . The colored portion shows the uncertainty band of seven scale variations.

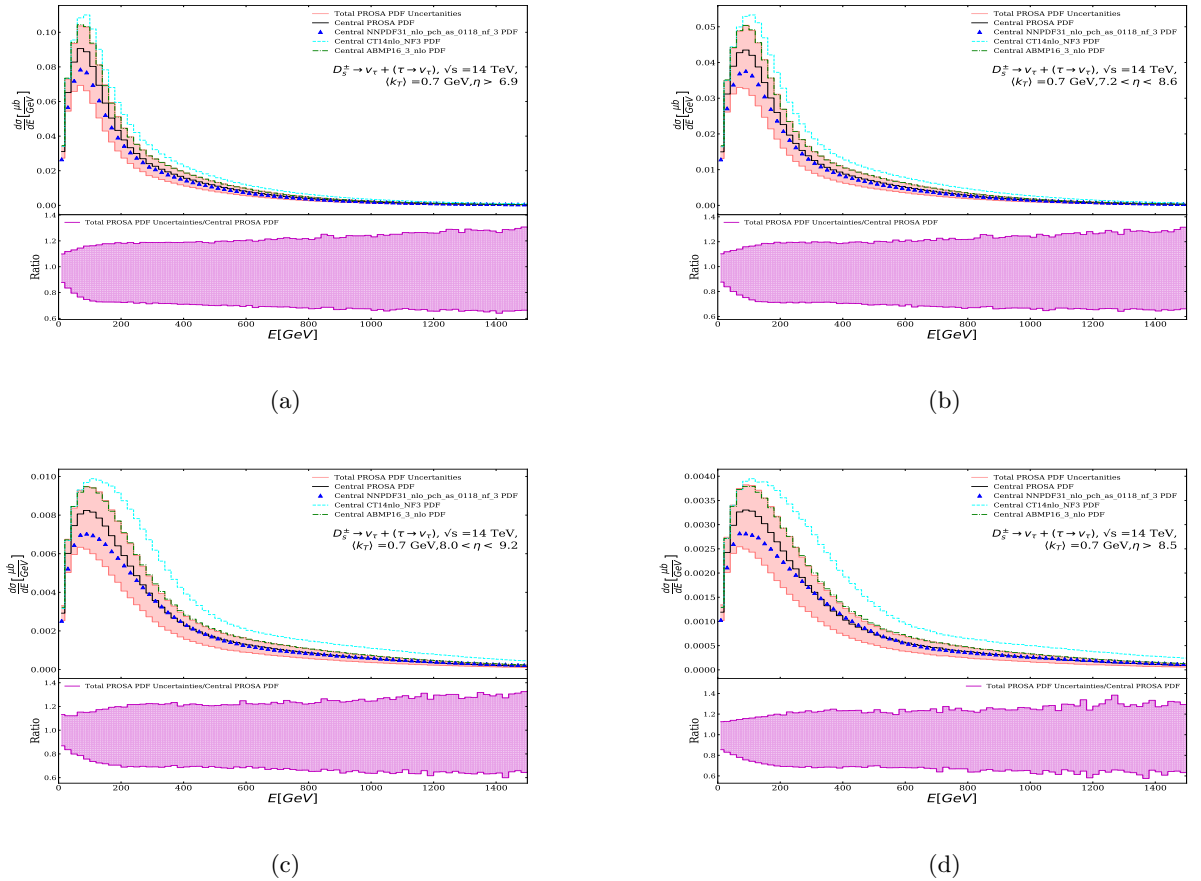


Figure 3: These figures show the uncertainties in the calculation of cross section of tau neutrinos in the rapidity range from  $\eta > 6.9$ ,  $7.2 < \eta < 8.6$ ,  $8.0 < \eta < 9.2$ , and  $\eta > 8.5$ . The selected rapidity ranges represent the forward region of the pp collision.

$k_T = 0.7$  GeV produces better prediction. We select the parameter,  $\langle k_T \rangle = 0.7$  GeV, and  $m_{T,2} =$

$\sqrt{4m_c^2 + p_T^2}$  with central scale of  $(N_R, N_F) = (1.0, 1.0)$ , to further study the uncertainties in PDFs for tau neutrino production in the forward region. The Fig.3 shows the uncertainties as bands around the central PDF assumption due to variations in PDFs in the cross section of tau neutrinos in the forward region of the pp collision. From fig. 3, we can see that the uncertainties remain approximately the same as  $etaincreases$  and are approximately 20%. We also compare the central value of other PDF and noticed that *NNPDF31\_nlo\_pch\_as\_0118\_nf\_3* and *ABMP16.3\_nlo* lies within the uncertainty band while *CT14nlo\_NF3* is slightly deviates from the uncertainties band. These model assumptions vary from different collaborations who have produced these PDFs. The root of the problem arises from the lack of knowledge of the structure of the nucleon. To get better estimates of the production rate in the forward direction, we need better understanding of model assumptions for PDFs.

## Conclusion

Neutrinos are mysterious and everywhere around us. Understanding neutrinos is a challenging task as they rarely interact with matter. FASER-nu [2] experiment has been proposed to measure the flux of neutrinos in the forward direction at the LHC. This paper focused on flux calculation with PDF uncertainties in the production of  $D_s^\pm$  and tau neutrinos from decays of these mesons. According to our results, the calculation of  $D_s^\pm$  double differential cross section with transverse momentum  $m_{T,2} = \sqrt{4m_c^2 + p_T^2}$  with  $\langle k_t \rangle = 0.7$  is the best fit for the LHCb data. The analysis of PDF uncertainties is still under way. According to our initial studies, approximately 20% of uncertainties are due to PDF. The studies will continue to examine the various elements of uncertainties due to the PDFs and find ways to reduce those uncertainties in our calculation.

## Acknowledgments

I would like to thank my mentor Dr. Milind Diwan at Brookhaven National Lab for his support and knowledge on this project. I would also like to thank Prof. Mary Hall Reno (University of Iowa), Prof. Maria Vittoria Garzelli (Hamburg University), Dr. Weidong Bai (University of Iowa), and Dr. Yu Seon Jeong (CERN) for their time and valuable suggestions on this project. This project was supported in part by the Brookhaven National Laboratory (BNL), Physics Department under the BNL Supplemental Undergraduate Research Program (SURP).

## References

- [1] A. De Rujula and R. Ruckl, "Neutrino and muon physics in the collider mode of future accelerators," doi:10.5170/CERN-1984-010-V-2.571
- [2] FASER collaboration, A. Ariga et al., Technical Proposal for FASER: ForwArd Search Experiment at the LHC, 1812.09139.
- [3] K. Winter, Detection of the tau-neutrino at the LHC, in ECFA Large Hadron Collider Workshop, Aachen, Germany, 4-9 Oct 1990: Proceedings.2., pp. 37–49, 1990.
- [4] P. Nason, S. Dawson and R. K. Ellis, The Total Cross-Section for the Production of Heavy Quarks in Hadronic Collisions, Nucl. Phys. B303 (1988) 607–633.
- [5] M. L. Mangano, P. Nason and G. Ridolfi, Nucl. Phys. B **373**, 295-345 (1992) doi:10.1016/0550-3213(92)90435-E
- [6] P. Nason, S. Dawson and R. K. Ellis, Nucl. Phys. B **327**, 49-92 (1989) [erratum: Nucl. Phys. B **335**, 260-260 (1990)] doi:10.1016/0550-3213(89)90286-1
- [7] G. P. Lepage, J. Comput. Phys. **27**, 192 (1978) doi:10.1016/0021-9991(78)90004-9

- [8] Bai, W., and Reno, M. H. (2019). Prompt neutrinos and intrinsic charm at SHiP. *Journal of High Energy Physics*, 2019(2). doi:10.1007/JHEP02(2019)077
- [9] W. Bai, M. Diwan, M. V. Garzelli, Y. S. Jeong and M. H. Reno, *JHEP* **06**, 032 (2020) doi:10.1007/JHEP06(2020)032 [arXiv:2002.03012 [hep-ph]].
- [10] PROSA collaboration, O. Zenaiev, M. Garzelli, K. Lipka, S. Moch, A. Cooper-Sarkar, F. Olness et al., Improved constraints on parton distributions using LHCb, ALICE and HERA heavy-flavour measurements and implications for the predictions for prompt atmospheric-neutrino fluxes, *JHEP*04(2020) 118, [1911.13164].

Converting National Synchrotron Light Source II accelerator status pages from HTML to  
Phoebus

Stephanie Keyes, College of Science, Oregon State University, Corvallis, OR 97331

Raymond Fliiller\*, Reid Smith, Emil Zitvogel, National Synchrotron Light Source II,  
Brookhaven National Laboratory, Upton, NY 11973

## ***Abstract***

The National Synchrotron Light Source II is a synchrotron radiation facility at Brookhaven National Laboratory in New York. Users, technicians, management, and engineers require access to detailed statistics of accelerator operation for various reasons. Currently, the only publicly viewable source of this information is a site called FullStat. This site is occasionally unstable; it has a tendency to crash on a nearly daily basis. While access to the controls system is the main alternative, this requires multiple logins and security measures, making it impractical for regular use. Solving this problem involves taking data from the controls system and feeding it into a public display that is more stable while maintaining security. In addition, the new site should be easily modified and expanded. The Phoebus architecture provides this advantage over the HTML format currently in use. However, the transition to this software architecture presents compatibility issues. These may include visual glitches or failure to display data due to an issue with a file path or process variable. These problems have been addressed, and additional pages created to aid in understanding the status pages. GitLab is used to facilitate changes made for the transition. My personal benefit is new experience gained with software version control and the concept of Git. This combines with my understanding of accelerator physics to provide vital skills to particle accelerator facilities such as the NSLS-II. Creating easily accessible public status pages is in line with the Department of Energy's commitment to the Open Government Movement.

## ***Introduction***

There is a demand outside of the NSLS-II control room to see various aspects of accelerator operations. Technicians, operators, and engineers clearly have a need to monitor

accelerator systems, but people outside the controls system firewall (such as students, beamline staff, and users) also watch the systems. This is currently facilitated by a website called FullStat<sup>1</sup>. There are two different kinds of pages on the site. The Python-based pages run reliably, but require manipulation of HTML code to modify. Pages on the site labeled as “X2GO” use screenshots taken directly from the controls system. These have proven unreliable on occasion, crashing on a regular basis without any immediately noticeable issue. A typical crash involves a failure to update the page, so there is a risk of reporting inaccurate data.

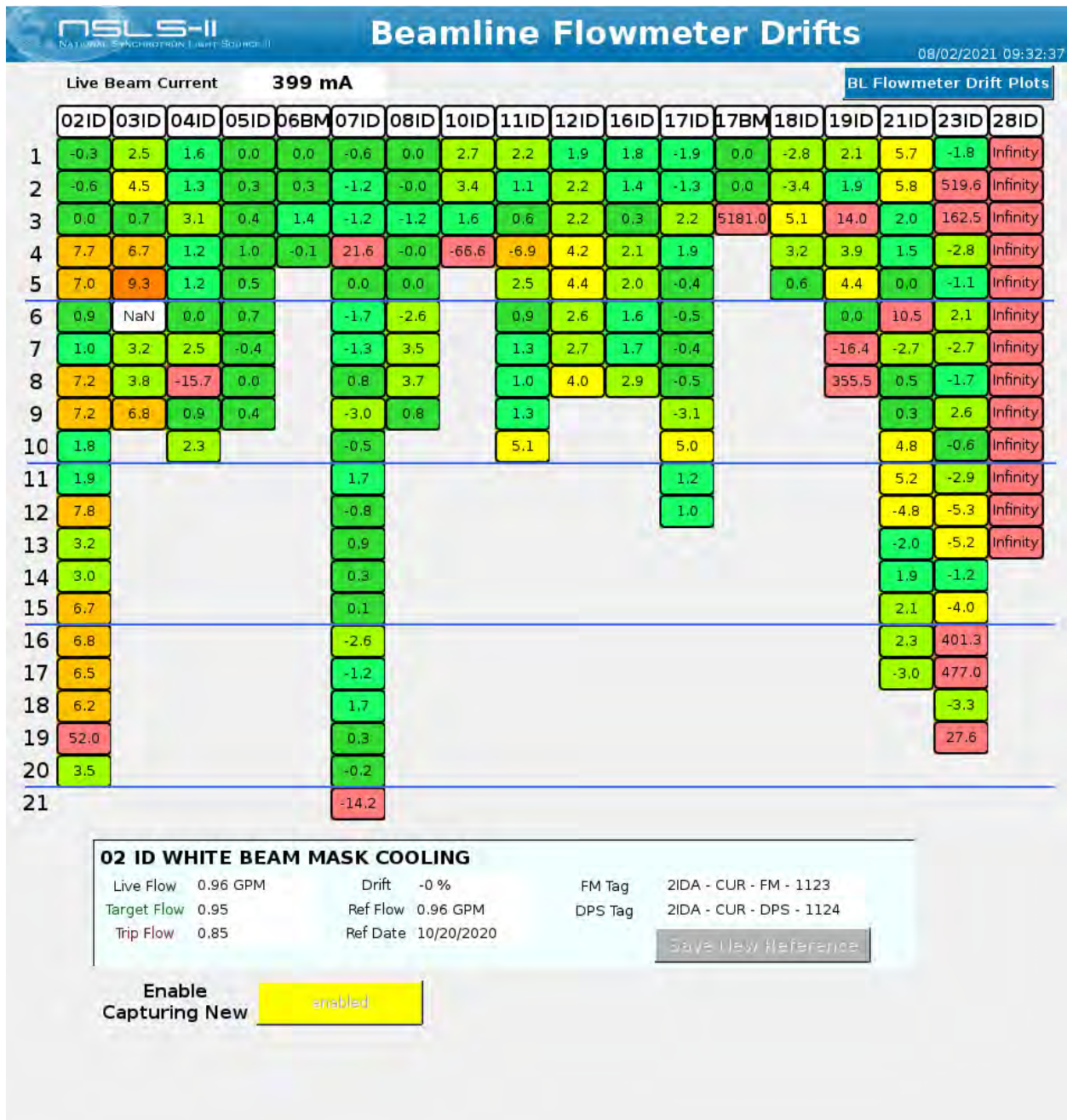


Figure 1: A frozen FullStat page; this screenshot was taken on August 3, 2021. It had been frozen for 24 hours.

The scope of this project involves moving existing pages to a new software environment, and modifying them so that they will properly display in the software architecture they are being moved to. Doing so involves use of the Controls System Studio (CSS) program used by the NSLS-II operations group. CSS has been the program of choice since the beginning of



operations in 2014. It is used to create pages for the operations group to monitor and operate the accelerator systems. The program itself has two different versions each running in their own software environment. The newer version is being used to generate the pages for this project. Completed pages are pushed into a repository, then Brookhaven National Laboratory's Information and Technology Department (ITD) takes over, merges the changes with its own repository, and puts the pages into a publicly viewable format.

### *Narrative*

FullStat is the website NSLS-II currently uses to publicly display accelerator statuses. This is done for the X2GO pages by displaying screenshots from the controls system, a process that is prone to failures. The Python-based pages operate on HTML code provided by the controls system without allowing outside access to accelerator controls. The version of CSS that facilitates these real-time updates uses a back-end architecture called Eclipse. Use of CSS in an Eclipse environment has been described as “slow” and “clunky”. Being a programming environment that has its basis in Java, Eclipse is very Java dependent. Eclipse puts a high demand on processing power and the program occasionally freezes. The computer it's running on must have specific operating system settings. Subsequently, there is a higher risk for errors. Additionally, working with HTML pages is labor intensive. Positions of objects and text on the page must be manipulated through raw HTML code.

While this setup provides the necessary security for this application, the operations group sought out an alternative due to these issues. Most alternatives did not have the necessary security level, and so could not be implemented. Then, the whole controls system was migrated from Eclipse to a new back-end software architecture called Phoebus. It offered a secure

webview feature that Eclipse did not. Pages could now be created using the menus and tools of CSS rather than raw HTML code.

## *Process*

Beginning the project required access to Controls System Studio, a program used by NSLS-II operators. There were already pages created, but these pages did not function properly with the Phoebus software architecture. Figure 2 is a screenshot of CSS as it appears in Phoebus. Everything that requires laborious coding in an HTML based website is easily done with menus instead. Any component of the page, or “widget” can be selected using the list on the left. The page itself can easily be viewed as it would appear in a webview, or in the editor as seen here. New widgets can be added using the menu to the right of the page being edited, and their properties can be edited using the menu right of that.

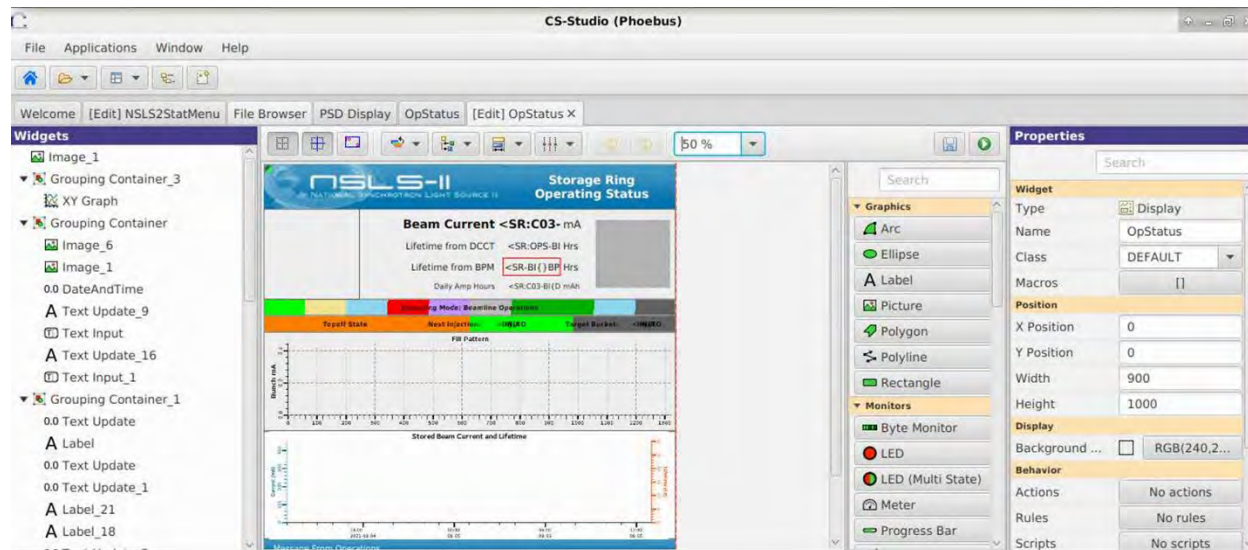


Figure 2: Screenshot of Controls System Studio (CSS) display editor

The process began with changing filename extensions to ensure compatibility with Phoebus. While the pages were given the “.opi” extension by their original Eclipse environment,

Phoebus has its own filename extension, “.bob”. The .opi files could be opened in Phoebus, but needed to be saved with a .bob extension for the project. As for plots and image files, no change in filename extension was necessary.

The transition to new back-end architecture caused some glitches to occur in the pages. Some objects were misaligned and incorrectly sized. Other times, error images would show instead of the intended image. Correcting these glitches was trivial. Then there were issues with data display. Due to an error in process variables (such as incorrect names or gateway issues), data fields would be highlighted in purple and fail to display the desired information.

All work done on the project was pushed into a repository on GitLab. This allows for file sharing and software version control.

	ID Gap	BMFS	GV1	Flag	Shutters	IDPS	FV	SSA	SSB	GV2&BL	Gap 1	Gap 2	Beamline Control
2 SIX	Closed	Open	Open	Out	Enabled	Open	Open	Open	Open	Open	21.9		Remote
3 HXN	Closed	Open	Open	Out	Enabled	Open	Open	Open	Open	Open	<SR:C3		Remote
4 ISR	Closed	Open	Open	Out	Enabled	Open	Open	Open	Open	Open	13.00		Remote
5 SRX	Closed	Open	Open	Out	Enabled	Open	Open	Open	Open	Open	<SR:C5		Remote
8 ISS	Closed	Open	Open	Out	Enabled	Open	Open	Open	Open	Open	15.0	15.0	Remote
0BBM TES	None	Open	Open	N/A	Enabled	Open	N/A	Open	Open	Open			Remote
10 IXS	Closed	Open	Open	Out	Enabled	Open	Open	Open	Open	Open	NaN		Remote
11 CHX	Closed	Open	Open	Out	Enabled	Open	Open	Open	Open	Open	<SR:C1		Remote
11BM CMS	Remote	Open	Open	N/A	Enabled	Open	N/A	Open	Open	Open	0.0		Remote
12 SMI	Closed	Open	Open	Out	Enabled	Open	Open	Open	Open	Open	7.30		Remote
16 LIX	Closed	Open	Open	Out	Enabled	Open	Open	Open	Open	Open	6.05		Remote
17 [A/F]MX	Closed	Open	Open	Out	Enabled	Open	Open	Open	Open	Open	<SR:C1	<SR:C1	Remote
17BM XFP	Remote	Open	Open	N/A	Enabled	Closed	N/A	Closed	Closed	Open	0.0		Remote
18 FXI	Closed	Open	Open	None	Enabled	Open	Open	Open	Open	Open	15.0	15.0	Remote
19 NYX	Closed	Open	Open	Out	Enabled	Closed	Open	Closed	Closed	Closed	6.41		Remote
21 ESM	Closed	Open	Open	Out	Enabled	Open	Open	Open	Open	Open	220.0		Remote

Figure 3: A status page as it appeared before corrections were made

This figure shows a status page before filename extension change and corrections were done. Note that the header shows error images. This is because the filepath is invalid- the image that is intended to appear does not exist in the folder where the program is looking for it. Fixing the issue involved changing where the program tried to get the image- in this case, the NSLS-II logo and page header. Some data fields are highlighted in purple, indicating a gateway issue or incorrect process variable name.

### ***Current status***

A feature was added that did not exist in the FullStat website: reference pages. The reference pages provide context for site users. Not all users will be familiar with every system, and this will aid in understanding displays that are typically made with experts in mind. They are screenshots of status pages, highlighted in colored rectangles and annotated either with a short description or a number corresponding to a list of descriptions. Rectangle color choice is largely arbitrary; colors were chosen to draw attention to certain parts of the screenshot and to contrast neighboring rectangles for easier visual differentiation. The following figures are examples of these reference pages. In the case of a more complex reference page such as Figure 5, text annotations needed to be put into a numbered list. It was not feasible to arrange annotations as they appeared in Figure 4.



### Insertion Device & Beamline Front End status reference

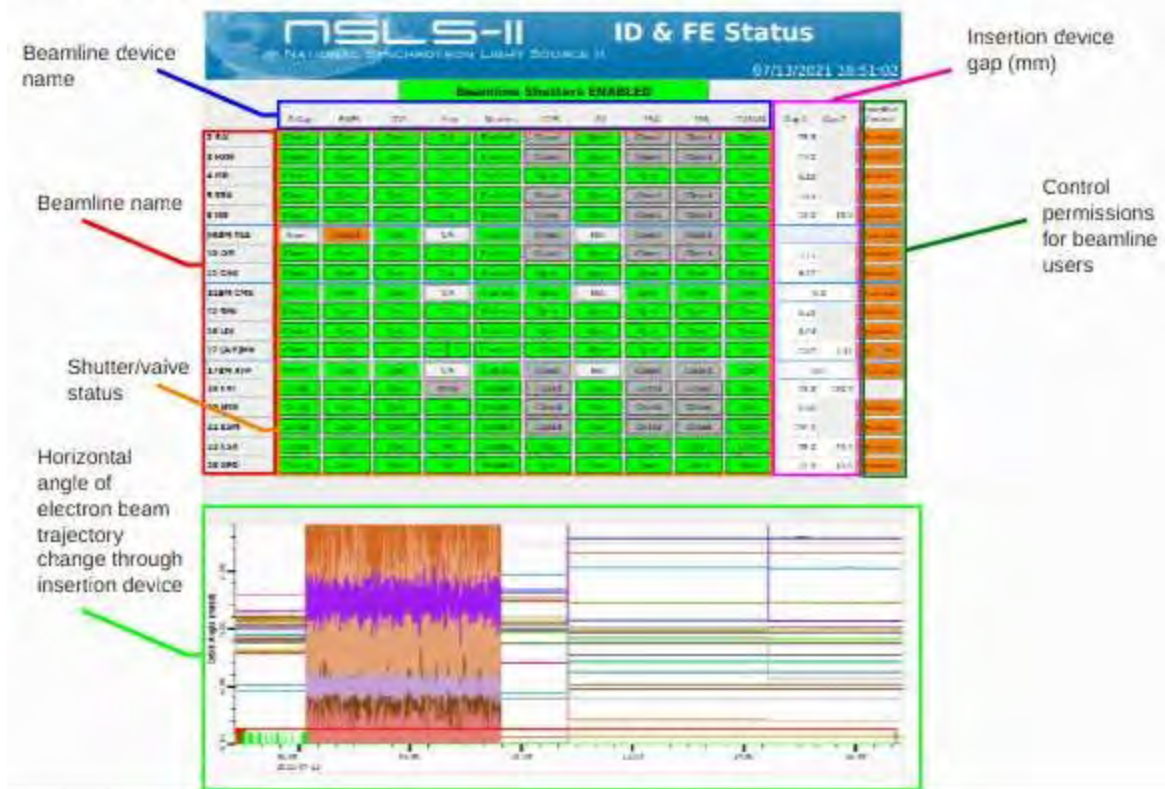


Figure 4: Status page for insertion devices and beamline front-ends; this page is easily annotated.

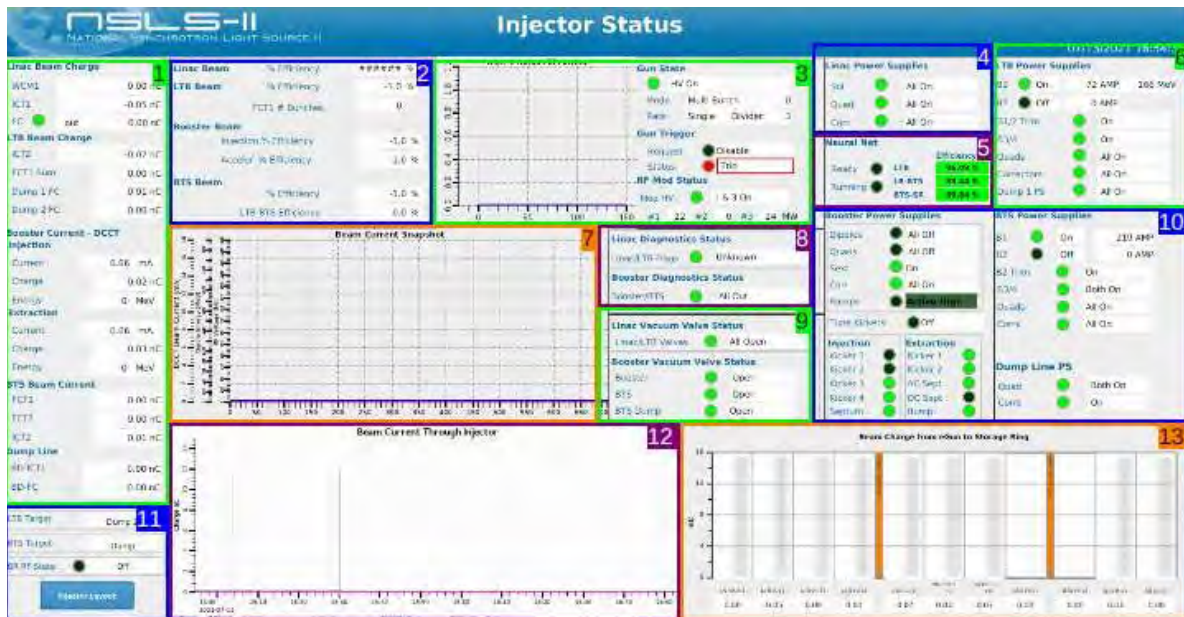


Figure 5: More detailed status page for the injectors and transfer lines

1. Beam charge and energy at various points in the injector/transfer line system
2. Beam transfer efficiency; ie. how much of beam makes it through the linac, linac-to-booster transfer line, booster, and booster-to-storage-ring transfer line
3. Status of electron gun; bunch pattern generated by its pulses, linac RF modulator status and power
4. Status of the power supplies to the linac's solenoid, quadrupole, and corrector magnets
5. At one point in time, a neural net program was used to operate the injectors. While this program is no longer in use, this status window serves as an additional reference for beam transfer efficiency.
6. Status of the power supplies for the bending, corrector, and quadrupole magnets in the linac-to-booster transfer line. Also shown is current in the bending magnets and the energy of any beam passing through.
7. Graphs showing beam current, RF amplitude, and the energy of any beam circulating in the booster.
8. Position of beam diagnostic devices in the linac and booster

9. Status of vacuum valves in the injectors and transfer lines
10. Status of booster magnet power supplies
11. Intended destination (or "target") for a beam that's in a given section of the accelerator complex; also status of the booster RF.
12. Beam current in various locations in the injectors
13. Beam charge in various locations in the injectors; also shows status of beam shutters in transfer lines

Figure 6: The numbered annotation list corresponding to the injector status reference page

### ***Next steps***

Currently, 5 reference pages have been completed, and the new site will have 8 status pages. BNL ITD has begun work on converting these pages into a publicly viewable format. Their work will involve setting up a secure server that can correctly display pages created and modified in this project. Issues they face in this transition involve failures to update color codes according to rules set in the Phoebus pages, and charts building up a 24 hour operations history instead of reading back data from the past 24 hours.

This project demonstrates that migration of CSS pages into Phoebus is feasible, and that secure webview pages are possible. It also shows the difficulties that are encountered in the process. More pages will be made available, beyond what has been created and modified here. At this moment, the new website is only available internally as testing is still in progress. Once the issues have been resolved and a secure server made, the site will be made publicly available.

Through this project, I have learned about the concept of Git and software version control. These things are vital in programming work of any kind. Complex machines such as particle accelerators require computer programming to make operations possible. With the

knowledge I have gained, I can begin similar projects at other particle accelerator facilities. Many have some form of publicly viewable status site, but not all are as detailed as the existing and new status sites for NSLS-II.

### ***Acknowledgements***

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

### ***References***

1. NSLS II Full Status, <https://status.nsls2.bnl.gov/fullstat.html>

### ***Footnotes***

\*Raymond Filler transferred to the BNL Collider-Accelerator Department during this project.



Variability in the areal characteristics of precipitation in the Houston, TX region

Grace Kowalski

Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette,  
IN 47907

Michael Jensen

Department of Environmental and Climate Sciences, Brookhaven National Laboratory, Upton,  
NY 11973

Dié Wang

Department of Environmental and Climate Sciences, Brookhaven National Laboratory, Upton,  
NY 11973

Brian Witteman

Department of Earth and Atmospheric Sciences, State University of New York at Oneonta, NY  
13820

## Abstract

Convective clouds drive the transport and circulation of heat, moisture, and aerosols throughout the troposphere. Currently, there is limited understanding of how aerosols impact convective clouds. Together with the BNL Climate Sciences Department, through the Tracking Aerosol Convection Interactions Experiment (TRACER) campaign, we are working to gain a better understanding of the interactions between aerosols and convective cloud processes. We are using data from the Houston, Texas region collected by National Weather Service's Next Generation Radar to analyze the variability in the areal coverage of precipitation within weather regimes with common atmospheric behaviors. Our findings from this analysis have coincided with expected precipitation behavior driven by a combination of the large-scale circulation and diurnal forcing from land-sea breezes. We have analyzed each synoptic regime's radar reflectivity along with visible satellite imagery to determine which kinds of weather events (e.g., sea breeze, large-scale precipitation, clear skies) occur most per regime. We also find that there are clear differences in the precipitation coverage at different locations. These results help to quantify the variability of precipitation and convective cloud properties among different large-scale meteorological conditions. This is an important step towards isolating the effects of aerosols on deep convective systems, a major target of TRACER research. Overall, findings show that regimes dominated by the westward expansion of the Bermuda High are generally moister and with a greater amount of large-scale convection. The regimes driven by a continental anticyclone have weaker winds and more sea breeze days, but both types of anticyclones show a clear diurnal cycle. The trough regimes range from more large-scale convection to many clear sky days, but do not tend to follow the same diurnal cycle patterns as the anticyclone regimes.

Throughout this project, I have developed my python coding abilities, and I have gained more meteorology knowledge.

## Introduction

The impacts of aerosols on deep convective clouds is a topic that remains uncertain and is the focus of active research in the atmospheric sciences. Convective clouds are an important driving of global, synoptic and mesoscale circulations, and transport moisture, heat, aerosols, and momentum vertically through the depth of the troposphere (Jensen et al., 2019). Aerosols, serving as cloud condensation nuclei, will impact the cloud droplet number concentration, and subsequently the formation of precipitation, impacting the lifecycle of the convective cloud system (Heikenfeld et al., 2019). On an even broader scale, the aerosol-convective interactions will further impact the global radiation balance with potential feedbacks that we do not yet comprehend. The United States National Academy of Science asserts that “the largest of all the uncertainties about global climate forcing—is probably the indirect effect of aerosols on clouds” (National Research Council, 2005).

The Tracking Aerosol Convection Interactions ExpeRiment (TRACER) has the goal of deepening scientific understanding of aerosol-convective interactions by researching the Houston, Texas region. Houston is ideal for this due to its unique conditions including having numerous isolated convective systems and a spectrum of aerosol conditions ranging from urban aerosols to cleaner, rural areas. TRACER will have a particular focus on sea breeze convection during the summertime. A sea breeze is caused by differential heating in the daytime where the land is warmer than the sea, causing convection to occur as the warm air over the land raises and circulates to the cold sea air (Miller et al., 2003).

When the TRACER campaign starts in October 2021, the aerosol-convective interactions and the impact of sea breeze will be measured with a comprehensive network of state-of-the-art instrumentation. In preparation, we analyzed large-scale meteorological and precipitation

characteristics, specifically areal coverage of precipitation. I focused on the Houston, TX region for the years 2010 to 2017 in the summer months of June to September. Overall, we found that there are distinct differences in the temporal and spatial variability of precipitation under different weather regimes.

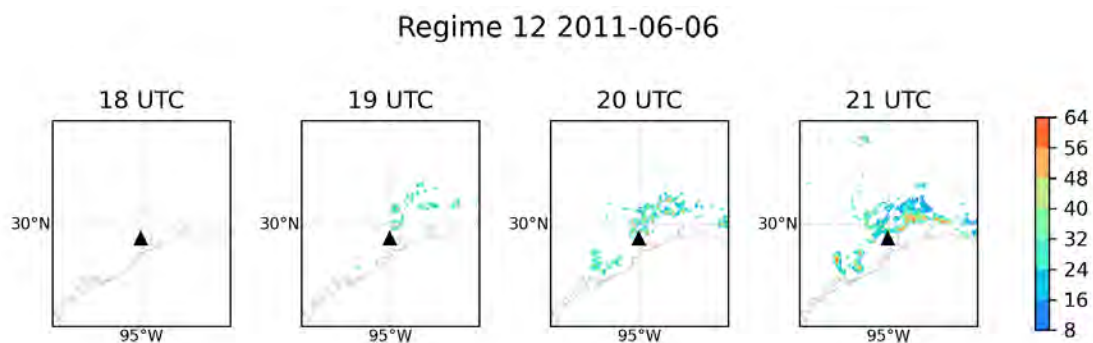
### **Scopes and Objectives**

The objectives of this research include understanding the deep convection properties (diurnal cycle, frequency, area coverage) over the region during summer months, understanding the variability in the quantities for different synoptic weather regimes, and generally preparing this information for the field phase of the TRACER campaign. This work will provide insights into the overarching question of how aerosol-convection interactions impact climate change and the radiation balance. The relevant scope is limited to the Houston, Texas region during the warmer summer months when sea breeze is more prevalent, specifically in the afternoon hours from 13 UTC to 23 UTC.

### **Methods**

Prior to any analysis of precipitation coverage, self-organizing maps (Kohonen, 2001), a machine learning technique, were used to classify the weather states into 16 synoptic nodes using large-scale weather regime analysis (Wang et al., 2021). These nodes can be split in half as 8 trough-related nodes and 8 anticyclone-dominated nodes. There are several transitional nodes that include characteristics of both anticyclone and trough regimes. Some of the anticyclone regimes are dominated by the Bermuda High and others by an anticyclone over the middle of Texas. For some of the trough regimes, the trough has already passed the region meaning they are post-trough patterns. All further analysis was done per node to evaluate the differences among them.

The radar reflectivity data used for each step of this analysis was collected from the National Weather Service's (NWS) Next Generation Radar (NEXRAD). Specifically, NEXRAD Level 2 data, GridRad which is an hourly, three-dimensional radar reflectivity product that merged 125 high-resolution S-band Doppler weather radars operated by the NWS (Homeyer & Bowman, 2017; Cooney et al., 2018). We plotted the radar reflectivity at a constant altitude of 2-km per day for each node in order to categorize each day with a specific precipitation weather condition. Figure 1 shows an example of the radar reflectivity at a height of 2 km for a sea breeze day in node 12.

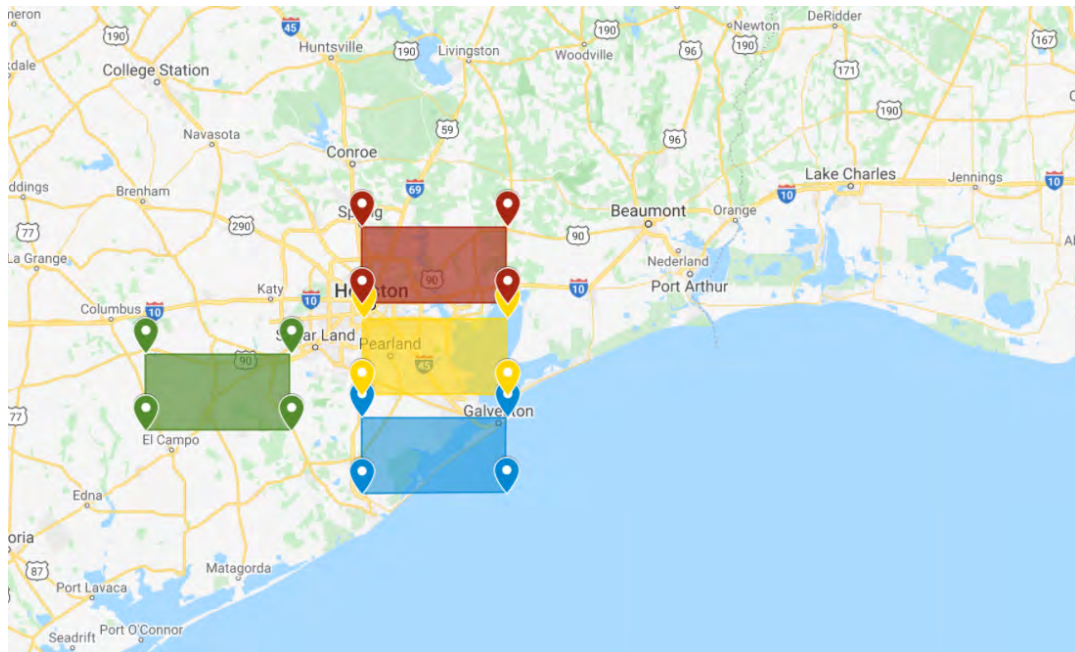


**Figure 1:** Radar reflectivity scan of sea breeze day (June 6, 2011) in node 12

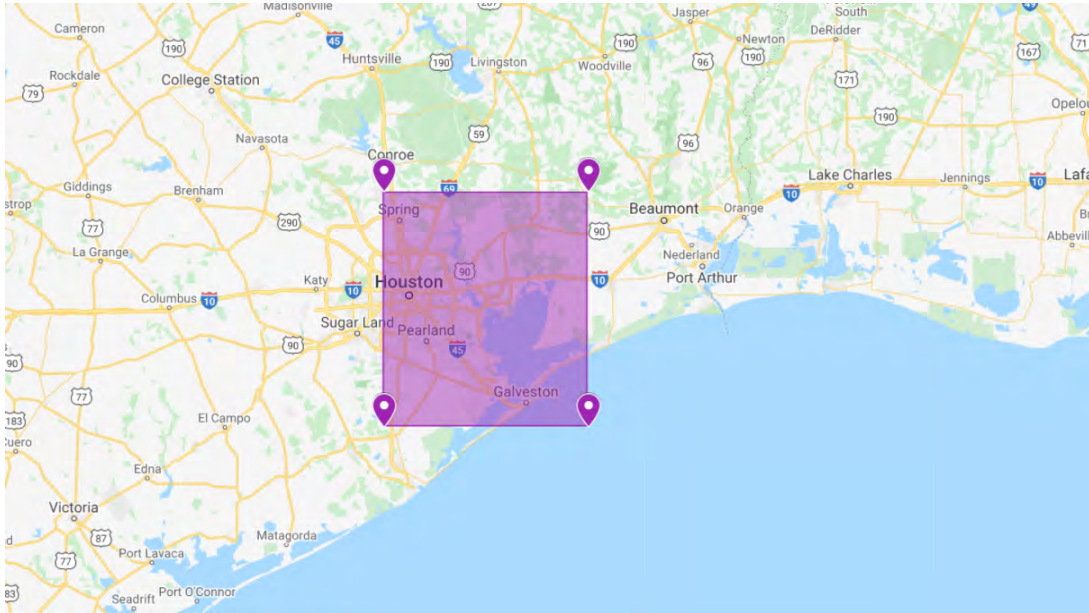
In order to classify each day per node as having a specific weather condition, I visually inspected radar imagery like Figure 1 in conjunction with NASA's visible satellite imagery, AQUA Modis, to classify the weather conditions and convection type of that day (Gumley et al., 2010).

Python was used to analyze and visualize all data for these findings. The data being analyzed includes several different domains within the Houston, Texas region. We analyzed the differences in precipitation characteristics moving from coastal to further inland areas. Figure 2 shows these domains of analysis. The blue box is closest to the ocean at latitudes between

29.024003 and 29.324003 degrees, and longitudes between -94.763353 and -95.413353 degrees. This will be called the coastal location. The yellow box is in between the farthest inland and most coastal box and is described as the inland location. This location's latitudes are between 29.411434 and 29.711434 degrees, and longitudes are between -94.763353 and -95.413353 degrees. The red box is the farthest inland and is called the urban location. The latitudes are between 29.767430 and 30.067430 degrees, with longitudes between -94.763353 and -95.413353 degrees. Lastly in Figure 2 is the ancillary location in the green box which is near TRACER's ancillary site. Its latitudes are between 29.27272 and 29.57272 degrees, and longitudes are between -96.385832 and -95.735832 degrees. Each of these locations are 0.30 latitude by 0.65 longitude degrees. Figure 3 shows the larger area that encompasses the blue, yellow, and red box which makes up the larger TRACER area being researched. This will be called the TRACER region and lies between latitudes of 29.2 and 30.2 degrees and longitudes between -94.5 and -95.5 degrees. It is a 1 degree by 1 degree box.



**Figure 2:** *The coastal, inland, urban, and ancillary locations*



*Figure 3: The TRACER region*

## **Results**

### **I. Weather Conditions per Node**

As stated previously, radar reflectivity plots and visible satellite imagery were used to classify each day in each node as having a specific weather conditions. These states break down into five overarching categories seen throughout the nodes, all of which were analyzed over the La Porte coast. One is the “large-scale convection day” which is when there is a large amount of precipitation caused by synoptic weather, such as frontal systems. There are “sea breeze convection days”, days with precipitation parallel to the coast caused by the sea breeze. Next are “little precipitation days” which had slight amounts of visible precipitation on the radar, but may have had cloudiness or sea breeze circulation without convection on the satellite images. Then there are “no precipitation days” which had no visible convection or precipitation, but had some form of cloudiness or sea breeze circulation. Lastly are the “clear sky days” with no precipitation or cloudiness.



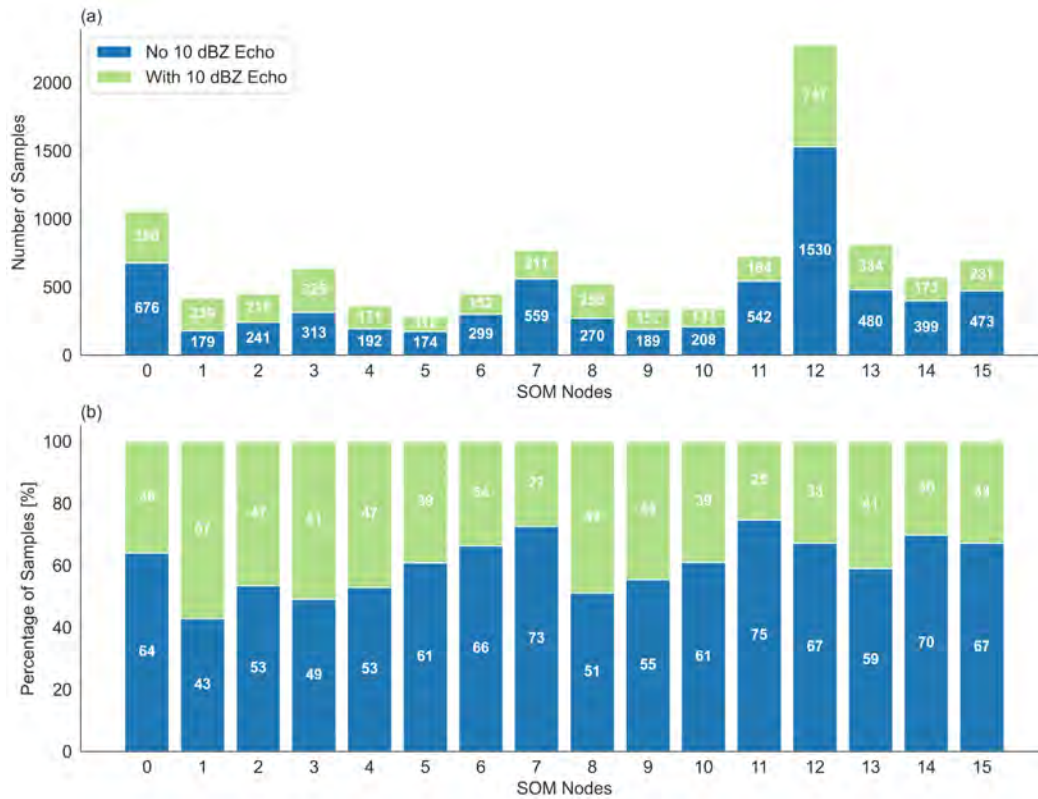
In Table 1, nodes highlighted in peach on the left column have the highest percentages of large-scale convection days. All of these regimes are associated with the westward expansion of the Bermuda High, with trough influence in nodes 2 and 3, causing strong moisture transport and unstable conditions allowing for large-scale convection. The blue highlighted nodes had the most sea breeze days, and are all anticyclone regimes with one being a transitional regime (node 9). These regimes have weaker winds than the ones influenced by large-scale systems, and mostly southerly wind flow which helps to propagate the sea breeze. The yellow highlighted regimes had the most days with little precipitation. Node 3, as stated previously, is an unstable trough regime and often experiences showers. Node 5 is a transitional regime. The green regimes have the highest no precipitation days are composed of trough regimes (nodes 7 and 11) or transitional regimes (nodes 5 and 6) with stabler conditions. The lavender regimes had the most clear sky days. They are post-trough regimes that have the most stable conditions out of all the regimes.

Node	Large-scale Convection	Sea Breeze Convection	Little Precipitation	No Precipitation	Clear Skies
0	36.5	3.1	20.8	36.5	3.1
1	52.6	13.2	18.4	15.8	0.0
2	41.5	4.9	24.3	26.8	2.4
3	44.8	1.7	27.6	22.4	3.4
4	33.3	24.2	9.1	33.3	0.0
5	30.7	3.9	30.8	34.6	0.0
6	31.7	0.0	21.9	46.3	0.0
7	25.7	1.4	15.7	44.3	12.9
8	35.4	25	10.4	22.9	6.3
9	38.7	19.3	19.3	19.35	3.2
10	35.5	3.2	25.8	29	6.4
11	27.3	3.0	9.1	43.9	16.7
12	23.2	20.3	11.6	38.6	6.3
13	32.4	14.9	13.5	35.1	4.0
14	25.0	3.8	15.4	42.3	13.5
15	23.4	14.1	17.2	34.4	10.9

**Table 1:** Percentages of each organizational state per regime [%]

## II. Amount of Radar Samples per Node

Each node has a different number of days within it, therefore each one also has a different number of radar samples. The nodes with more days have more samples. This is displayed in panel (a) of Figure 4. Node 12 has the greatest number of observations with precipitation (radar reflectivity > 10 dBZ) during the afternoon hours, and the most days within this node, while node 5 has the least of both.. Panel (b) of Figure 4 shows the percentage breakdown of how many of those samples are equal to zero, meaning there is no radar echo > 10dBZ and therefore no precipitation, in dark blue. The cases with all radar echo < 10dBZ are shown in light green. Nodes 1 and 3 have the highest percentage of time with radar reflectivity > 10dBZ reaching 57% and 51%, respectively. This is due to enhanced cloud and precipitation formation resulting from the high moisture content these regimes have from the westward expansion of the Bermuda High. Nodes 7 and 11 have the lowest percent of precipitation cases (27% and 25%). These are calmer post-trough regimes resulting in drier conditions. The rest of the nodes lay somewhere between 30% to 49% days with precipitation echo.

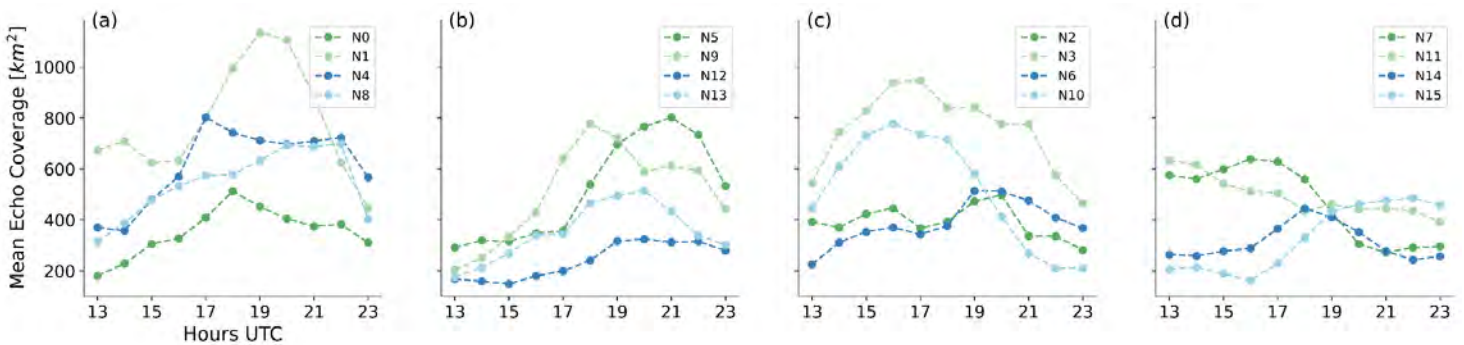


**Figure 4:** Amount and percentage of samples of precipitation (radar reflectivity > 10 dBZ) area per node

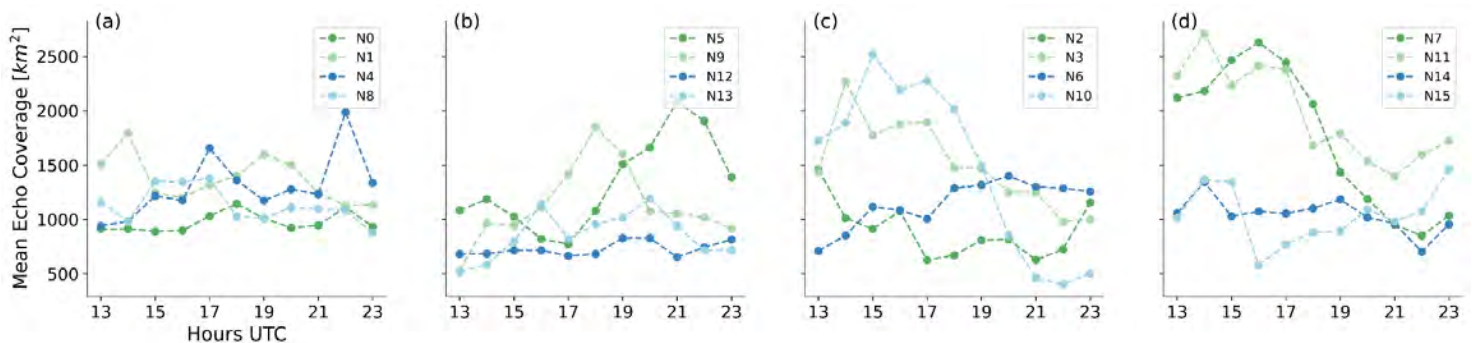
### III. Average Mean Precipitation Coverage

The averages for Figures 5-8 were taken over the TRACER region. Figure 5 demonstrates again the precipitation amount in different nodes. Nodes 1 and 3 have the highest average echo coverage on Figure 5. These are two of the moistest and most precipitating regimes, with a high level of moisture transport from the Gulf of Mexico. Node 11 is one of the lowest, due to it being a drier regime as stated previously. Node 12 is the overall lowest on this Figure 5. It is a continental anticyclone which, from Figure 4, is only moderately precipitating with 33% of radar samples recording precipitation. Figure 6 combined with Figures 4 and 5 show more of the whole story. Figure 6 is the same information as Figure 5, but with the zero samples removed, so this is only the days with precipitation being averaged. In this case, nodes 7 and 11

are the highest overall. Therefore, while they are drier regimes with less days with precipitation, when it does precipitate, it tends to be large-scale convection with larger precipitation coverage. Node 12 remains the smallest in this case still. It is only moderately precipitating and favors sea breeze induced convection. Sea breeze is a meso-scale circulation, so convection induced by this is smaller compared to large-scale induced convection.



**Figure 5:** Average echo precipitation coverage per node per hour at 10dBZ, with zero samples

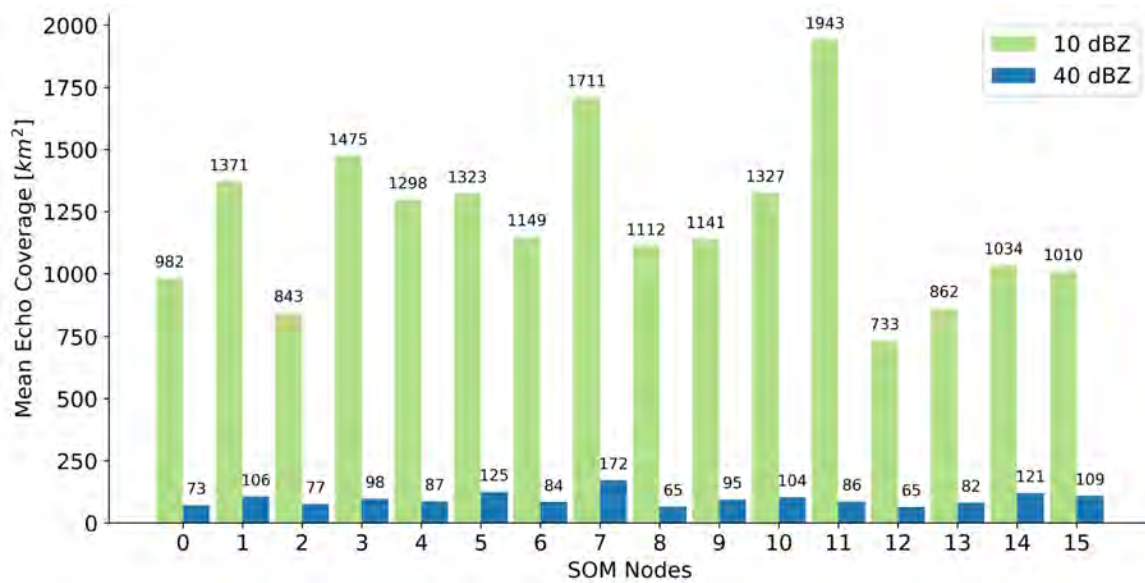


**Figure 6:** Average echo precipitation coverage per node per hour at 10dBZ, without zero samples

Figure 5 also demonstrates a clear diurnal cycle for anticyclone regimes. Panels (a) and (b) show nearly each anticyclone regime peaking in the afternoon hours, at 19-21UTC. As the sun rises, the land takes a while to heat and warm the lower layers of the air, causing convection to start as it heats. Trough regimes peak earlier in the day, with transitional regimes being less

predictable and with less of a diurnal influence as shown in panel (c). Panel (d) shows post-trough regimes tend to stay flatter, with more stable conditions.

Figure 7 shows similar analysis, but at different radar reflectivity thresholds. 40dBZ is a commonly used threshold for deep convective precipitation. At the 10dBZ level, node 12 has the smallest precipitation coverage, while at the 40dBZ level, nodes 8 and 12 have the smallest precipitation coverage. Node 8 also favors sea breeze induced convection, so it is smaller for similar reasons to node 12. Node 11 has the largest echo surface areas at the 10dBZ level, while node 7 does at the 40dBZ level. Node 7 is also the second largest at 10dBZ, both being drier regimes with large-scale precipitation when it occurs.

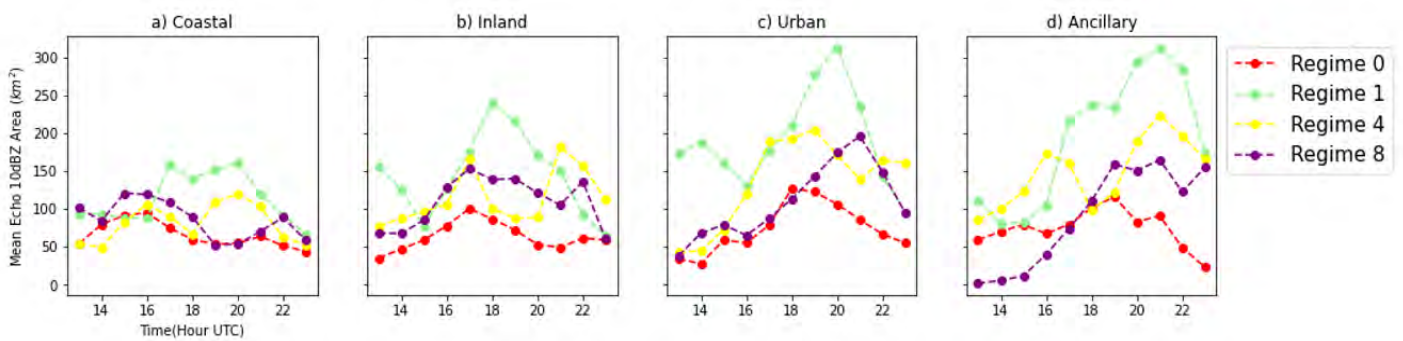


**Figure 7:** Overall average echo coverage > 10dBZ vs > 40dBZ

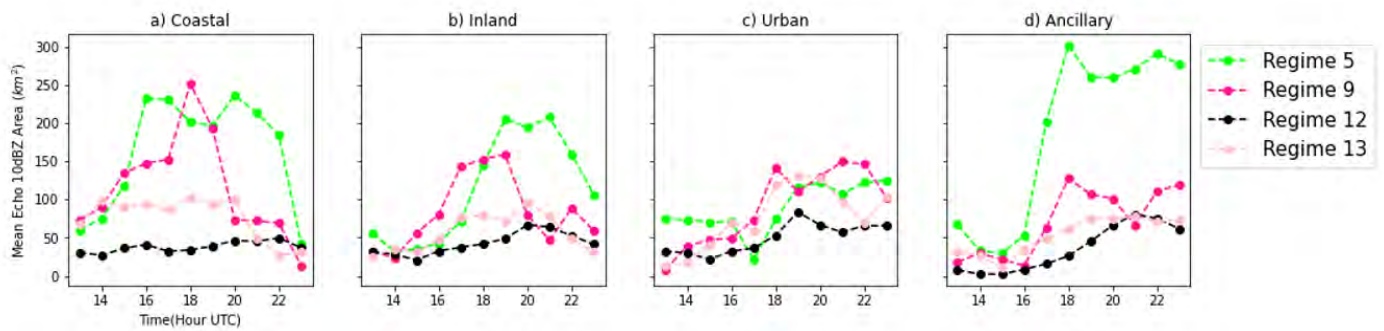
#### IV. Variability of Time Peak Precipitation Coverage with Location

The anticyclone regimes' time of maximum precipitation coverage, or peak time, delays moving from the coast to inland. These locations are the ones defined in the Methods section. Figures 8 and 9 display this increase in peak time for anticyclone regimes. For example, node 9 peaks at 18 UTC, then 19 UTC, then 21 UTC when moving from the coastal location to the

furthest urban location. For every case, the ancillary location is most closely resemblant of the urban location. Some regimes, like node 1, have two peak times and this pattern is true for only one of them. For the anticyclone regimes dominated by sea breeze, like nodes 8 and 12, the increase in peak time could be due to the impact of sea breeze starting along the coast in the earlier afternoon, then propagating inland as the winds carry it. For other regimes, it could be caused by the south-easterly winds propagating other convection inland still.



**Figure 8:** First group of anticyclone regimes moving inland

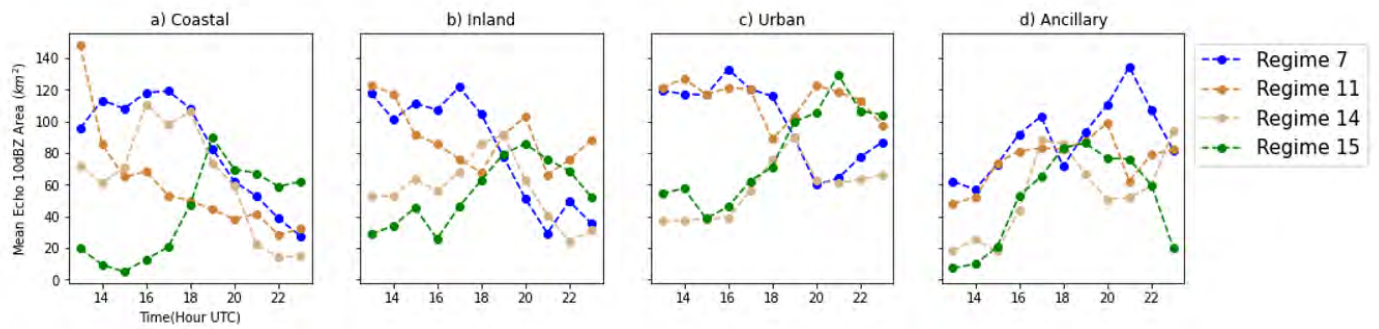


**Figure 9:** Second group of anticyclone regimes moving inland

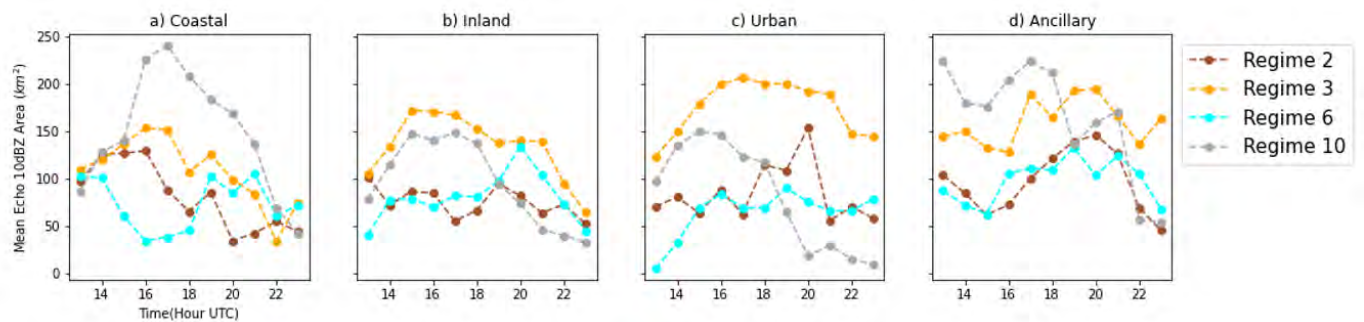
Trough regimes do not show this clear increasing pattern, and do not show any clear pattern in general. Figure 10 shows post-trough regimes. Here, node 15 happens to follow the same increasing pattern moving inland, but none of the rest do. Figure 11 shows the other trough regimes also do not follow any specific pattern. For example, node 3 has an earlier peak time at



the inland location than the coastal and node 6 has its earliest peak at the urban location. This could be due to them having a different direction of windflow blowing it away from the coast, such as winds from the north instead of the south.



*Figure 10: Post-trough regimes moving inland*



*Figure 11: Trough regimes moving inland*

## Conclusion

Overall, our hypothesis was proven true: there are distinct differences in the temporal and spatial variability of precipitation under different weather regimes. Regimes associated with the westward expansion of the Bermuda High are the most precipitating while post-trough regimes are the least. The continental anticyclone and other trough regimes are moderately precipitative. The post-trough regimes have greater areas of precipitation while regimes with large sea breeze influence have lesser areas of precipitation. Sea breeze days are most common in the anticyclone

regimes with weaker, southwesterly winds. When the post-trough regimes do have rainfall, it is often deep convection. The deep convective rainfall is, on the whole, smaller in precipitation coverage than the overall precipitation and moisture transport for each regime. Lastly, the anticyclone regimes all have clear diurnal cycle influence and increasing peak time moving inland while trough regimes do not. All of this analysis was based on large-scale, synoptic regimes, but did not capture the nuances of local Houston conditions, such as pollution, aerosols, and urban island convection. To better understand those components, the TRACER campaign will be looking at these, specifically aerosol influence. To conclude, these results help to quantify the variability of precipitation and convective cloud properties among different large-scale meteorological conditions, an important step towards isolating the effects of aerosols on deep convective systems.



## Bibliography

- Cooney, J. W., Bowman, K. P., Homeyer, C. R., & Fenske, T. M. (2018). Ten year analysis of TROPOPAUSE-OVERSHOOTING Convection Using Gridrad data. *Journal of Geophysical Research: Atmospheres*, 123(1), 329–343.  
<https://doi.org/10.1002/2017jd027718>
- Gumley, L., Descloitres, J., & Schmaltz, J. (2010). Creating Reprojected True Color MODIS Images: A Tutorial. *NASA*
- Heikenfeld, M., White, B., Labbouz, L., & Stier, P. (2019). Aerosol effects on deep convection: The propagation of aerosol perturbations through convective cloud microphysics. *Atmospheric Chemistry and Physics*, 19(4), 2601–2627.  
<https://doi.org/10.5194/acp-19-2601-2019>
- Homeyer, C. R., & Bowman, K. P. (2017). *3-D gridded nexrad WSR-88D radar data*. GridRad. <http://gridrad.org/index.html>.
- Jensen, M. P., Xu, Y., van Lier Walqui, M., Usenko, S., Snyder, J., Oue, M., Nowotarski, C., Matsui, T., Kumijian, M., Hu, J., Giangrande, S., Defer, E., Brooks, S., Ryzhkov, A., Kuang, C., Fridlind, A., Bruning, E., Zhang, G., Wang, Y., ... Collins, D. (2019). Tracking aerosol Convection Interactions Experiment (TRACER) Science Plan. *ARM*. <https://doi.org/10.2172/1561242>
- Kohonen, T. (2001). *Self-organizing maps* (Third, Extended, Vol. 30, Ser. Series in Information Sciences). Springer .

Miller, S. T., Keim, B. D., Talbot, R. W., & Mao, H. (2003). Sea breeze: Structure, forecasting, and impacts. *Reviews of Geophysics*, 41(3).

<https://doi.org/10.1029/2003rg000124>

National Research Council. (2005). *Radiative forcing of climate change: Expanding the concept and addressing uncertainties*. National Academies Press, Washington, D.C.

Wang, D., Giangrande, S., Jensen, M., & Park, J. (2021). *Linking synoptic patterns to cloud properties and local circulations over the Houston, TX region* [Poster Abstract].

2021 JOINT ARM USER FACILITY AND ASR PI MEETING.

## Design of High Current Electron Transport System

Danielius Krivickas

Department of Engineering/Industrial Technology, Suffolk County Community College, NY,  
11946

## **ABSTRACT**

The Coherent Electron Cooling Project group is in the process of upgrading their current Cathode Deposition System to allow for more diverse yields and incorporate a way to layer the cathode in cesium, potassium, and antimonide molecular beams. My primary role in this project was to help make a 3D model of a proposed beamline that would be fitted into the updated system and make changes to it as the project went along. During the project, I was able to successfully make a model that meets the teams' requirements and with it, we can start planning as to where we can attach sensors and pumps to it. As a result of this summer, I have improved my proficiency with Autodesk Inventor and have learned how to use a piece of open-source software called gnuplot to inspect cross-sections of the beam.

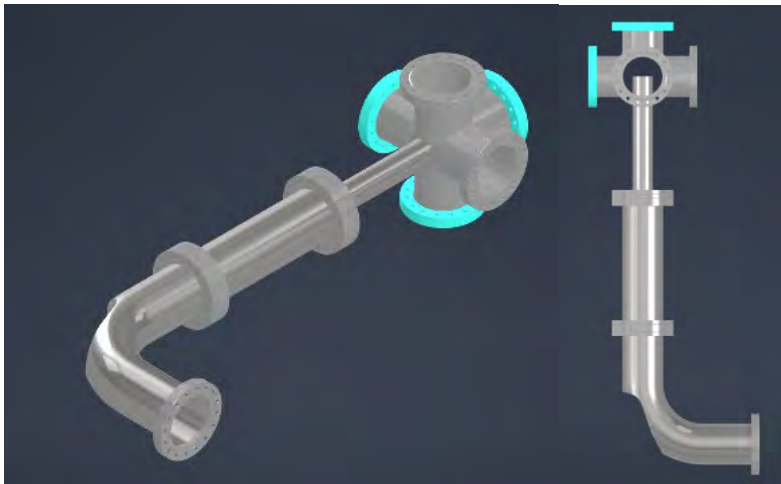
## **Introduction**

During my time at the SURP program at Brookhaven National Laboratory, the assigned task was to make a 3D Model of a beamline for the Coherent Electron Cooling (CeC) Project and learn about the system in place and why it is being modified. From the general information that was given the current installation was meant to be a temporary proof of concept that would be built upon to make it a more permanent installation. Though the system proved to be quite effective at producing the required cathodes and was kept operating for quite a few years. The new beamline should allow for more diverse yields and incorporate a way to layer the cathode in cesium, potassium, and antimonide molecular beams. My primary objective is to make a 3D model of the beamline as when it comes time for assembly, the group would like to be able to have a model on file that can then be incorporated into a larger assembly file to see if any other accommodations would need to be made for the new beamline fit. This will allow for when it comes time to assemble the whole system no unforeseen issues will come up.

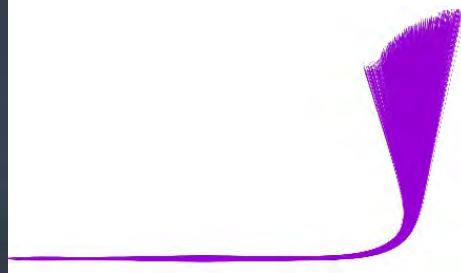
## Methods

The main methodology of this project was to take all existing material about the beamline in question and create a model from it. This required me to collect a basic schematic of the requested beamline, all data pertaining to the cross-section of the beam to get the scale correct, and find out what additional features would be required to be added to make the model more complete. I would then use these collected resources to create the model in Autodesk Inventor, but before I could fully dive in, I needed to input a text file into gnuplot that had thousands of plots in order to plot out the beam in to collect the necessary dimensions from it.

## Results

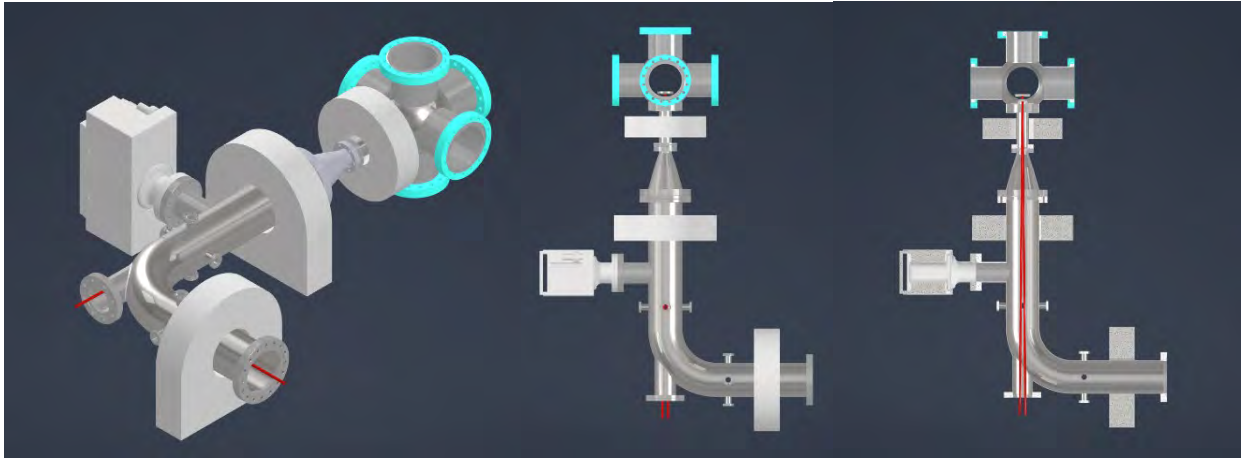


**Figure 1.** Initial beamline model

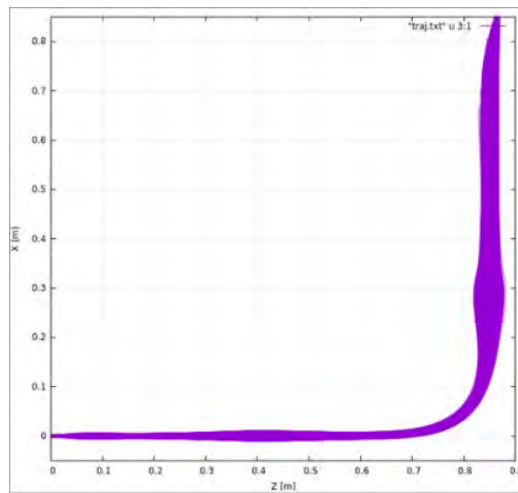


**Figure 2.** Initial beam cross-section

The is the initial model of the beamline shown in Figure 1 was made with this cross-section of the beam, Figure 2, provided by Jyoti Biswas as a reference. The original plan was to have a 4-inch tube at the bottom of the beamline that would carry the beam to the 5-way cross, but it was later revised to a 2 ½-inch tube. I had also made the initial mistake of using a tubular cross instead of a spherical one, so that would have needed a modification down the line.



**Figure 3.** Final beamline model



**Figure 4.** Most recent beam cross-section

In the final version of the beamline that I modeled as seen in Figure 3 the 5-way tubular cross was changed to a 6-way spherical cross that had one of the flanges cut off to 1 inch. This model includes dummy solenoids, an ion vacuum pump, and 8 ports for the 2 sets of quad detectors for beam alignment to show possible interference when the beamline will be connected to the general assembly. Two conical reducers were also used to take up less space and use less material. The conical reducers were also modified to remove a shared flange between them to mitigate a possible leak in the system. The revised 2.5-inch outer diameter bottom tube was also implemented by this stage as well. The 8 portholes will be used for quad detectors to measure the beam alignment and an additional one for an ion pump to maintain ultra-high vacuum pressure.

Figure 4 shows the final cross-section of the beam that I had received from Jyoti Biswas and I needed to make sure that the beam still fit within the beamline which it did. The final item that I added to the model was the anode and cathode plate which can be seen within the 6-way cross.

## **Conclusion**

The SURP program at Brookhaven National Lab had given me a unique opportunity of furthering my skills in 3D modeling and to pick the mind of my mentor occasionally. The project that I had been tasked with in the CeC project had taught me more about Autodesk Inventor's feature set and I'm grateful for walking away with such a useful skill.

## **References**

Figures 2 and 4 were provided and made by Jyoti Biswas.

## **Acknowledgments**

I would like to thank the DOE and BNL for the phenomenal opportunity to intern here. I would also like to thank my mentor Nathaniel Speece-Moyer, Luca Cultrera, Jyoti Biswas, and John Walsh.

This project was supported in part by the Brookhaven National Laboratory (BNL), Instrumentation Division, under the BNL Supplemental Undergraduate Research Program (SURP).

Integrated Facility Management: Remote support for proactive building maintenance

Angela Lagnese, Watson College of Engineering, Binghamton University, Binghamton, NY  
13902

Daniel Forino, Integrated Facility Management, Brookhaven National Laboratory, Upton, NY  
11973



## **I. Abstract**

The Integrated Facility Management Division (IFM) manages and maintains the hundreds of buildings on the Brookhaven National Laboratory campus. IFM helps to continue the mission of the laboratory, “a passion for discovery,” by maintaining adequate indoor environmental conditions so that science and support operations can be conducted in comfortable and well operated buildings. I utilized controls logic and active and historical trending data to report on deficiencies and anomalies in several buildings to identify the root causes for various issues. To provide remote monitoring of the building’s environmental conditions, overall condition, and operation of the Heating, Ventilating and Air Conditioning (HVAC) equipment in the buildings the Building Management System (BMS) utilizing an Automated Logic (ALC) WebCtrl platform was used. The BMS platform is critical to the laboratory’s success, as remote work environments are becoming the new normal due to the recent pandemic. Factors such as reduced space occupancy, increased ventilation rates, and modified operating hours all come into play impacting the way monitoring and maintenance can be conducted. After the issues are reported on, a meeting with the Facility Project Managers (FPMs) occurs and they can turn the issues into work orders. Communication with the FPMs is essential to get any operational critical issues or future concerns taken care of for proactive and preventative maintenance. More reports mean better control over the buildings and greater feasibility for discoveries and innovations to take place. The goal of a well-operated building is that you do not realize it is being operated, with the goal of eliminating downtime and proactively identifying smaller issues before they can become substantial.

## II. Introduction to WebCtrl

WebCtrl is utilized to manage and operate the 900+ buildings on the BNL campus. This system allows the user to see each building and every HVAC unit that has logic and fault detection and diagnostics associated with it. The controls logic used allows for a continuously updated look at all buildings and units at any given time. The units are programmed and linked to the system so, for example, if the chilled water system in an air handling unit opens to a higher percentage to help in lowering the zone temperature, the system will be updated to reflect that in a matter of seconds. The logic also allows for the system to set boundaries and setpoints for alarms using a series of if-then statements. Using the high zone temperature alarm as an example, if the lower and upper setpoints for a particular unit are 65°F and 70°F and the zone temperature is 72°F, the logic is constructed so the alarm will become active. From that point, the logic will change something in the unit like increasing the amount of chilled water going through the system, or if there is nothing the system can do, the alarm will stay active until someone is notified to service it.

Additionally, since the HVAC units are regulating temperature in multiple sections of a building, the interface depicts the range of comfort in the form of colors for each section (See Appendix A for a screenshot of the interface). For example, a green colored section represents a comfortable temperature that is within the setpoints of the respective section and a red colored section represents an uncomfortable temperature that is outside the setpoints of the section. These colors represent the temperature; however, they do not represent the units themselves and how they are operating. A green colored unit, for example, can have a dirty air filter or a fan that is not working. This stresses the importance of reporting every unit in a building, regardless of what temperature the section is maintaining.

In addition to the controls logic and color key, WebCtrl uses Fault Detection and Diagnostics (FDD) to enable automatic detection and diagnosis of equipment<sup>1</sup>. This is used when alarms are activated. It will detect the issue and give some suggestions as to what the problem might be. The objective of this is to improve operating efficiency of HVAC systems by 10-30%. This aids in reporting on the active alarms and completing the reports for review.

### **III. My Tasks**

Throughout this summer, I have been utilizing the ALC WebCtrl system to identify and report issues in various buildings. Reporting is a tool to create a database of past operations and issues to eventually recognize trends and recurring issues in the system. My task was to review and report on buildings at BNL for recordkeeping and pinpointing issues in the building. The flow of work to identify and resolve issues are as follows: Enter the WebCtrl system and select a building, complete a full report on all units in the building, consolidate issues into a Red Report, meet with the FPM to review the reports on their building, create work orders for issues, get the work order completed, and then continuously monitor the system to keep these issues at a minimum. This flow is explained further in the following paragraphs.

To effectively compile data from the system, I made a report template (See Appendix B for the remote monitoring template). This allows for a uniform recording document to be completed for every building. The important topics to be included were determined and those became the main boxes to fill in for each report (i.e., Building No., unit name, FPM associated to the building, date, observations, active alarms, comments section, graphics). Identifying the building number and unit name on the report allow for the FPM to easily address any problems

<sup>1</sup> AutomatedLogic. (2021, April 27). *Fault detection & diagnostics*. YouTube. <https://www.youtube.com/watch?v=V2jYcRW8TZc>.

with a specific unit. If there is a listed active alarm on a particular unit, and they will know what building the unit is in and when the alarm was active. The goal for these reports is to review every unit in a building, including all observations and alarms, to see the trends and identify issues that are outstanding from normal operation.

Once the reporting is complete, a more consolidated report is completed. I decided to name it the “Red Report” due to the content of the report, which lists all the issues highlighted in red from the building report (See Appendix C for the Red Report template). The purpose of the red report is to list everything that needs a follow-up, to reduce the time spent reviewing reports and effectively solve the issues listed. When creating work orders for the building, this report will be useful. For example, one of the reports I completed this summer was for National Synchrotron Light Source II (NSLSII). This building is 2 floors in the shape of a ring, with other satellite buildings associated with it. Being a large building, it made the building report extremely lengthy, which can make it difficult to recognize the issues quickly amongst the rest of the report. This is where the red report becomes useful.

The Facility Project Managers are essential to this process of reporting deficiencies and issues in buildings. After the building report and the red report are completed, a meeting with the FPM is scheduled. During this meeting, the reports are reviewed, addressing any identified issues, and answering questions listed in the questions/comments section of the report. This is where the FPM’s knowledge of their building is utilized to answer these questions. From this point, it is the job of the FPM to turn any identified issues that need maintenance into work orders. These types of work orders are mostly corrective, meaning they are not scheduled maintenance. For example, an exhaust fan gets maintained on a fixed schedule in the

preventative maintenance system, but if the exhaust fan failure alarm becomes active, a corrective work order will need to be used to address the alarm.

The flow of identifying and resolving issues with the help of the WebCtrl system and the FPMs is straightforward and effective. It is important for IFM to stay updated on any issues happening in the buildings as well as the trends that keep them well-operated.

#### **IV. Importance**

This process of using the BMS to report on buildings and address deficiencies is crucial for continuing operations and experiments at the laboratory by keeping downtime at a minimum, maintaining units proactively, and adapting to present-day work environments. Equipment downtime in a lab setting is not ideal for any scientist. Some experiments are time-sensitive, meaning they need full operation of their lab to produce accurate results. By utilizing the WebCtrl system to identify any deficiencies in the units, the maintenance can be completed quicker, hopefully before more issues or damage to any units occur. Continuous monitoring is essential in this sense. Recognizing trends in units, such as temperature during operation, and schedules in the building help to pinpoint deficiencies and anomalies in the system. This is formed by keeping record of the reports and the dates they were completed. Lastly, due to the recent Covid-19 pandemic, remote working has increasing dramatically. At the lab, this caused many employees to be virtually completing their work. Being a computer system with wireless capabilities and controls logic, people who have access to WebCtrl can view the buildings from anywhere at any time. Working virtually with a remote monitoring system can help significantly in identifying issues and deficiencies in the equipment at the laboratory. The BMS offers many benefits to BNL and plays an important role in achieving the goal of a well-operated laboratory campus.

## **V. Conclusion**

The reporting process I completed this summer has the potential to be utilized to proactively identify and resolve issues in equipment at BNL. This process of reporting deficiencies and anomalies in buildings through the WebCtrl system and notifying the Facility Project Managers to correct them is essential for a well-operated campus. Continuously reviewing, reporting, and maintaining buildings will keep active alarms and issues to a minimum, which is important in building management. The data collected from these reports will build up a historical database where trends can be determined and improved upon.

## VI. Acknowledgements

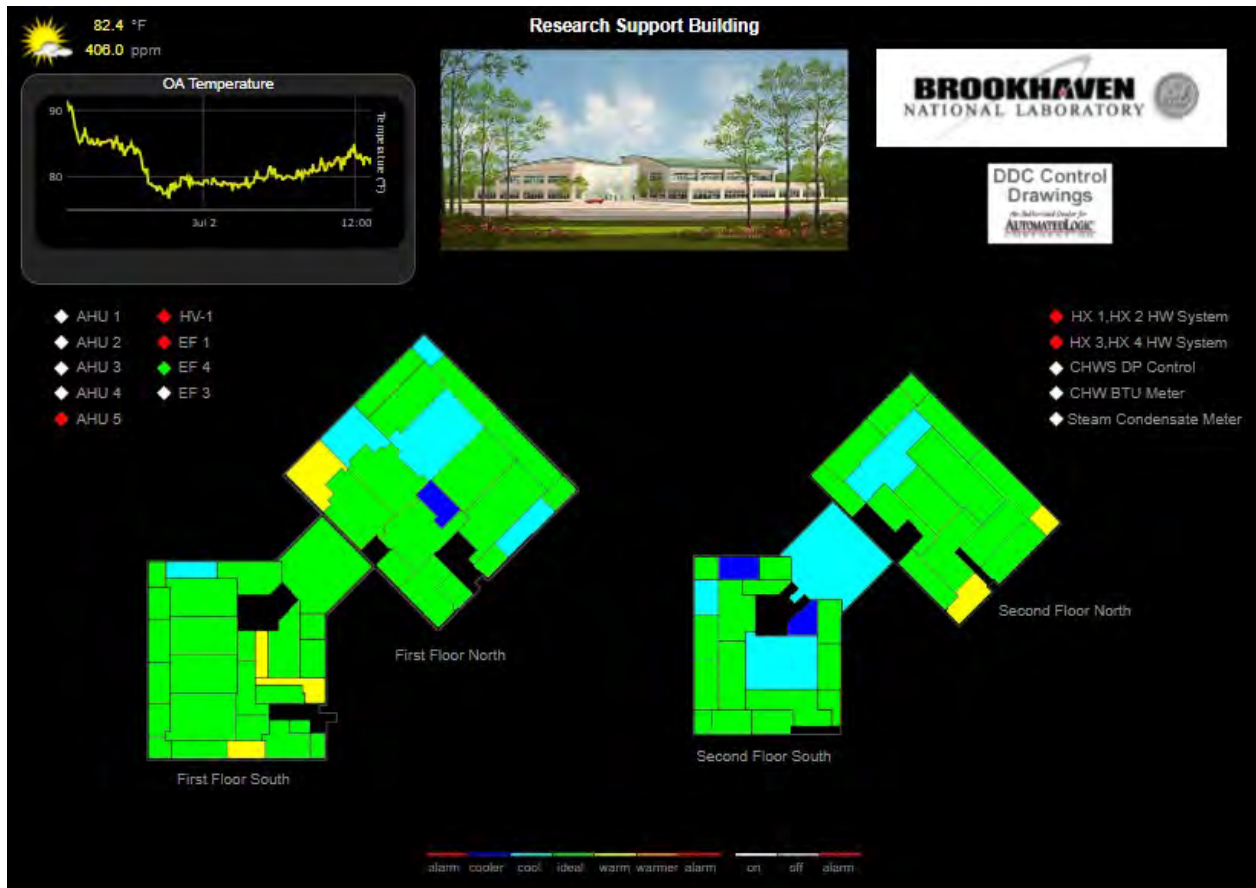
I would like to thank my mentor, Dan Forino, and the rest of the Integrated Facility Management Division for helping me to complete my project and internship at Brookhaven National Laboratory. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## VII. References

AutomatedLogic. (2021, April 27). *Fault detection & diagnostics*. YouTube. <https://www.youtube.com/watch?v=V2jYcRW8TZc>.

## VIII. Appendices

### Appendix A: Screenshot of WebCtrl Interface for Building 400 on BNL campus



**Appendix B: WebCtrl Remote Monitoring Report Template**

**Web CTRL Remote Monitoring Report**

Completed By: Angela Lagnese

**Red Font = Follow-up needed**

<b>Building:</b>	<b>Unit:</b>	<b>FPM:</b>	<b>Date:</b>
<b>Observations</b>	•		
<b>Active Alarms</b>	•		
<b>Questions/Comments</b>	•		

**Graphics:**

**Appendix C: WebCtrl Remote Monitoring Red Report Template**

**Web CTRL Remote Monitoring Red Report**

Completed By: Angela Lagnese		<b>Includes all alarms/issues in red from all reports in this building</b>
Building:	FPM:	

<b>From Report:</b>		
<b>Unit</b>	<b>Date</b>	<b>Active Alarms/Issues</b>

<b>From Report:</b>		
<b>Unit</b>	<b>Date</b>	<b>Active Alarms/Issues</b>



Development of data analysis and automation routines for the Inner Shell Spectroscopy beamline

John Lictro

Computer Science, Manhattan College, Bronx, New York, 10471

Denis Leshchev

National Synchrotron Light Source-II, Brookhaven National Laboratory, Upton, NY, 11973

## Abstract

The Inner Shell Spectroscopy (ISS) beamline at National Synchrotron Light Source-II (NSLS-II) investigates the properties of a wide variety of materials through x-ray absorption spectroscopy. Each experiment generates large datasets on a multitude of samples with the goal to capture their element-specific dynamics and transformations as a function of external parameters, such as temperature, voltage, etc. Presently, there is no streamlined way to detect and characterize the spectra associated with species transiently formed in the sample during *in situ* experiments. This project focuses on building data analysis software that can be used to automate the detection and extraction of the components from the recorded datasets. Our approach consists of utilizing the singular value decomposition of the data matrix with proceeding testing of the significance of each component using a variety of strategies, including scree plot, residual calculation, and autocorrelation calculation, and then integrating this with a Graphical User Interface. Our goal is to wrap this analysis into a python package that could be installed and used at ISS or elsewhere to perform such analysis. We have come very far in this effort, and fully developed the SVD analysis as well as the GUI. As a result of this summer internship, I have been able to enhance my linear algebra abilities and add data analysis to my repertoire of skills. I have also become much more familiar with the process of software development and feel more comfortable using the Ubuntu OS and the Conda package management system for Python.

## **I. Background**

### **A. The beamline**

The Inner Shell Spectroscopy beamline at National Synchrotron Light Source-II(NSLS-II) is used to perform experiments in a wide range of disciplines spanning chemistry, biology, physics, and materials science. The ISS beamline uses x-ray absorption spectroscopy – a unique tool to probe electronic and geometric properties of materials with elemental specificity. The goal of these experiments is to capture the dynamics and transformations of these materials as a function of external parameters, such as temperature, voltage, chemical treatments, etc. These experiments generate hundreds of spectra at a time that capture the dynamics of the processes under investigation. During such an experiment researchers are interested in learning about the number of steps that the reaction/process takes and what are their characteristic time scales. These questions can be, in general, answered by a combination of tools based on matrix factorization of the data. However, the existing most wide-spread tools used to analyze spectroscopy data only allow analysis of only a few spectra at a time making the process highly time consuming and inefficient. Currently existing packages that do have capabilities to handle such datasets are non-intuitive, require a high level of expertise, and are difficult to integrate within the beamline infrastructure, which limits their application during actual experiments. The goal of this project is to create a data analysis tool that is efficient and easy to use, and could be installed and used at the ISS beamline or elsewhere in order to automate the detection and extraction of significant components from various datasets. In the following we describe this process

## **II. Body**

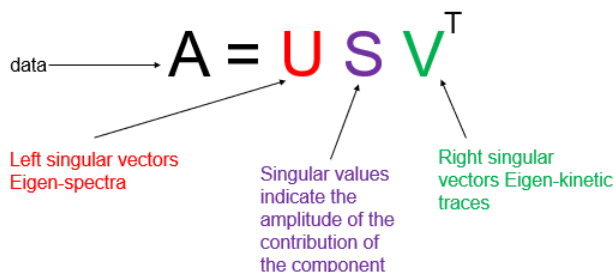
### **A. Scope and Objectives**

The objective of this project is to create an easy to use tool to perform data analysis on data acquired by the ISS beamline, as well as other beamlines as needed. The initial goal was to build a tool for determination of the number of significant components, corresponding to the species present in the sample, in the dataset by means of computing the singular value decomposition of the data matrix. Determination of significance of each component is then tested using a variety of methods. These methods include scree plots (i.e. plot of singular values as a function of the component index), residual calculation, and autocorrelation calculation. Eventually, once the data analysis portion was completed, the goal was to integrate this with a graphical user interface. This interface would allow users to easily import and display data. It would also allow them to easily filter through parts of the data based on a variety of criteria.

### **B. Methods**

This project was completed in the Python programming language. The Python libraries NumPy and Matplotlib were used in the program. NumPy was used to perform the singular value decomposition on an input data matrix. It was also used to compute things like residuals and autocorrelation. Matplotlib was used in order to take these pieces of data and plot them on figures. Qt Designer was used to create the graphical user interface that would allow users to actually interact with the methods that were programmed to perform and display the data analysis.

The first step in the process was to compute the singular value decomposition of a data matrix. Data matrix  $A$  – a rectangular matrix with dimensions  $N_E \times N_t$  with  $N_E$  being the number of energy points in each spectrum and  $N_t$  being the total number of spectra in the series – is passed in and has the singular value decomposition run on it. The result of this is three matrices,  $U$ ,  $S$ , and  $V$ :



**Figure 1:** Visualization of singular value decomposition

where matrices  $U$  (with size  $N_E \times N_E$ ) and  $V$  (with size  $N_t \times N_t$ ) are left and right singular vectors;  $U$  and  $V$  are unitary and orthogonal; matrix  $S$  (with size  $N_E \times N_t$ ) is diagonal and composed of singular values sorted in a descending order and indicating the relative significance of each component.. Once the SVD was run on the data matrix  $A$ , the next step was to plot these matrices so that the user can view each of them and their components. A scree plot was used to display the singular values of matrix  $S$  as a function of component index, allowing the user to look at all of the components and compare them to see which are the most outstanding. As an add on to the scree plot, users can also inspect the importance of the subset of components by computing low rank approximation of the dataset and calculating the total residual amplitude between that and the original dataset. Here, the low rank approximation of a dataset corresponds to the data matrix  $A$  reconstructed only using  $n$  first component and can be denoted as  $A_n$ ; the total residuals  $\chi^2$  can be computed as follows:

$$\chi^2 = \sum_{i,j} ((A)_{i,j} - (A_n)_{i,j})^2$$

Here, the subset of significant components is determined qualitatively based on the loss in the total residual value as a function of the number of components included in the low rank approximation calculation; if all components are included,  $A_n \equiv A$  and  $\chi^2 = 0$ .

In addition to that, users can judge the relative importance of components qualitatively by inspecting the general shape of vectors of U and V matrices. Here, the smooth components are generally related to the useful signal carrying chemical information, while noisy components correspond to the noise found in the dataset. The next step in our process was to compute the autocorrelation of matrices U and V to quantify the smoothness of the data. Autocorrelation was computed for each vector using standard approach with the lag of 1:

$$a_{U_j} = \sum_{i=1}^{N_E-1} U_{i-1,j} U_{i,j}$$

$$a_{V_j} = \sum_{i=1}^{N_t-1} V_{i-1,j} V_{i,j}$$

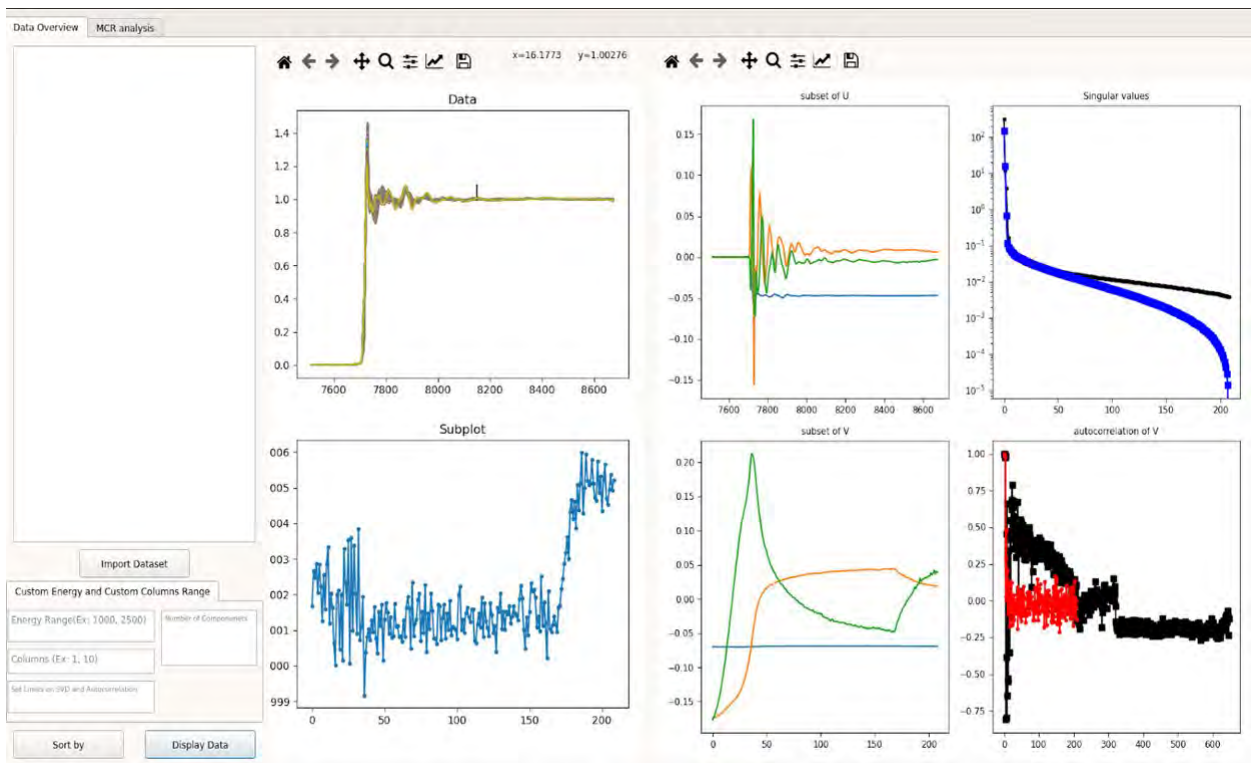
Determined as such, the autocorrelation will take values between 1 and -1, since U and V matrices are unitary. Components in the data that have a high autocorrelation appear to be smooth, whereas those with a lower autocorrelation appear noisier. Components that have meaningful signals in your data will be smooth, and therefore have a higher autocorrelation. The threshold that we used for determining whether a component was meaningful was an autocorrelation of 0.8 or higher as is recommended.<sup>1</sup>

The next major step of the project was to create a graphical user interface that would allow the user to interact with the program that performs the data analysis presented above. The GUI has simple buttons that allow the user to upload their data set file, and another button to

confirm that they want to display it. The SVD analysis will then be performed in the background, and all of the relevant graphs will be displayed to the user. This includes the graphs of matrices U and V, the graph of singular values, and the graph of the autocorrelation of U and V. It also includes a graph of the overall dataset, and one graph for an optional display of a subset of the overall dataset. The GUI also contains several empty and optional fields. These fields allow the user to filter what kinds of data they want to display at any given time. Filling in these fields and then pressing the “Display Data” button will refresh the data and display the updated graphs with the users filters applied.

### III. Results

The singular value decomposition analysis is fully completed and functional. The graphical user interface for displaying and interfacing with the SVD, is also complete. The GUI has several methods for allowing the user to filter through and display their data.



**Figure 2:** Image of the working GUI

There are many fields that allow the user to control the data that is displayed, as can be seen in the bottom left of figure 2. Users are able to input an energy range or set of columns to display a certain range of data. They can also choose to display a certain number of components in the data, or set a number of points to display on the graphs of singular values and autocorrelation. Filling in these fields and then pressing “Display Data” will then redisplay the data with the users specific preferences. The user also has the option to right click on the graph of the overall dataset, and “Add a line.” Adding a line will add a line that goes through the dataset at the location that the user clicked. A subplot containing the data that was sliced through will be displayed. The user can add as many slices as they want in order to compare specific parts of the data, they also have the option to normalize these subplots. Figure 2 shows the final GUI that was developed during this project. The dataset for testing the GUI/analysis capabilities was taken from previously published work by Porosoff et al<sup>2</sup>.

There is still room for improvement in the future on this project. For example, the next step of the factor analysis workflow includes the so-called multivariate curve resolution method. Here, the algorithm allows the user to retrieve spectra corresponding to each component in the reaction/process. This is achieved by supplying an educated guess for the starting solutions as well as imposing physical constraints on the shapes of the retrieved spectra.. This would also require expansion of the GUI so that it can be displayed and understood by users. The I/O capabilities of the GUI could also be expanded to be able to take in a larger variety of file types from users, or even multiple files at the same time that users can sort through.



#### IV. Conclusion

The tool that we have developed allows users to easily analyze and view their data. The analysis based on the singular value decomposition is fully complete and the GUI is fully functional. It allows users an easy to use and understand way to understand their data. The project is fully functional in its current state, but there is still room for improvement. Another analysis method could be added and integrated with the GUI. This has been a very exciting project to work on and I am grateful for having had this opportunity.

#### V. References

<sup>1</sup>A. Martini, S.A. Guda, A.A. Guda, G. Smolentsev, A. Algasov, O. Usoltsev, M.A. Soldatov, A. Bugaev, Yu. Rusalev, C. Lamberti, A.V. Soldatov, PyFitit: The software for quantitative analysis of XANES spectra using machine-learning algorithms, *Computer Physics Communications*, Volume 250, 2020, 107064, ISSN 0010-4655, <https://doi.org/10.1016/j.cpc.2019.107064>.

<sup>2</sup>Liu, Renjie, Leshchev, Denis, Stavitski, Eli, Juneau, Mitchell, Agwara, Jane N., and Porosoff, Marc D. Selective hydrogenation of CO<sub>2</sub> and CO over potassium promoted Co/ZSM-5. Netherlands: N. p., 2021. Web. <https://doi.org/10.1016/j.apcatb.2020.119787>.

## **VI. Acknowledgements**

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

# The Study of 2D Ion Crystals For Quantum Information Processing

Jian Jun Liu, Department of Physics, Stony Brook University, NY, 11794

Andrei Nomerotski, Mentor, Cosmology and Astrophysics, Brookhaven National  
Laboratory, Upton, NY 11973

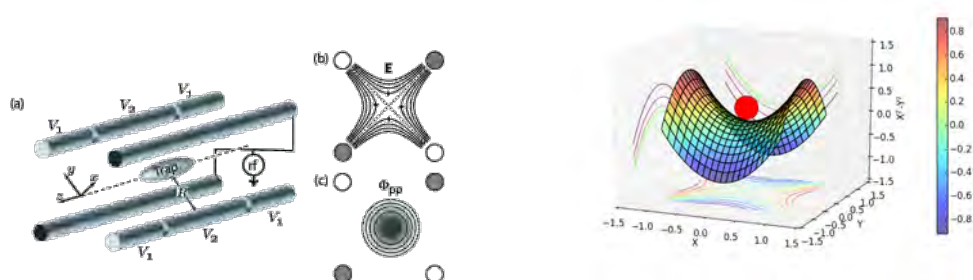
## **I. Abstract**

We study the two-dimensional ion crystals in an RF trap using a novel camera to capture and produce images of barium ions. The data was taken at the university of Washington and was sent here at the bnl to analyze. The goal is to understand the structure of the 2D ion crystals as well as manipulating the 2D coulomb crystals for quantum information processing. Due to the extremely high fidelity state intialization, control and readout, ion crystals makes an idea source for quantum computing. Unlike the one-dimensional ion traps where spacing is hard to maintain for large numbers ions and the scalability due to the constraint of being 1-dimensional, the two-dimensional ion crystals may be well suited to solve these problems. The 2D crystal is prepared by extracting a sample of barium ions and cool it down to temperature close to absolute zero using conventional laser cooling techniques, the fluorene then allows us to collect data which would be analyzed. Our experiment has successfully created the ion crystals using the conventional laser cooling and pumping techniques. We were able to extract the information using the time3pix camera and study its motions. This has also led to the developments of software that are able to track key information such as the spatial coordinate that can be used for mny applications

## **II. Introduction**

To have any kind of ion trapping at the atomic scale, potential with a local minimum in spatial coordinate is required. However, due to the nature of electromagnetism, or specifically by Earnshaw's theorem, that any collections of point charges as is the case of the barium ion that we have cannot be maintained in a

stable equilibrium configuration solely by electrostatic interaction of the charges. That means, given an electrostatic configuration of charges, we cannot have a local minimum of the potential. Thus, to create an ion trapping system, we must go on to utilize electrostatics.

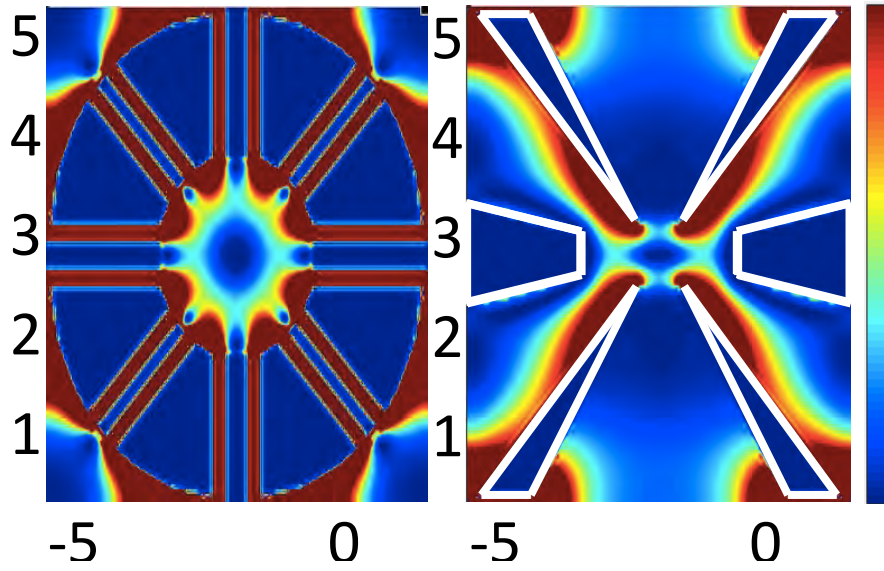


(Figure 1 left. Design of the Paul trap, with 4 electrodes two pairs of opposite polarization. Figure 1 right. Saddle potential diagram of the linear Paul trap)

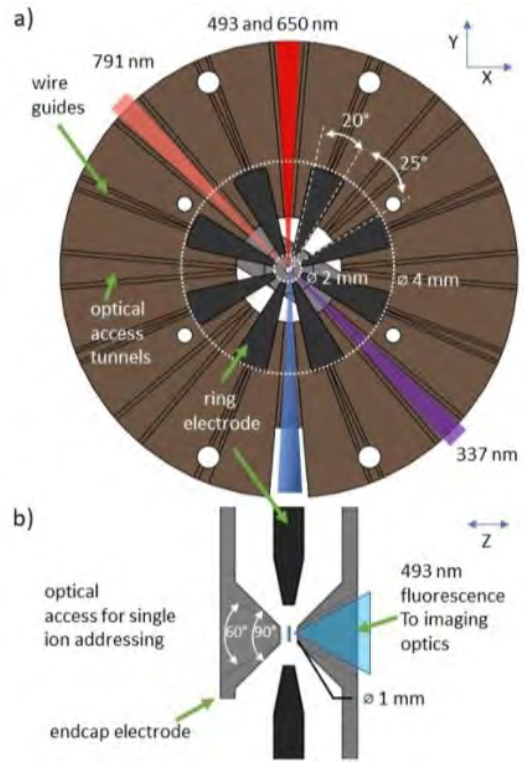
We cannot have a local minimum based on Earnshaw's law for ion trapping, but we could create an effective local minimum so that ion gets confined. This is done by creating a saddle potential and rotating it by some frequency as shown in the figure below. Right before the ion (red ball) is about to fall in the y-direction (due to instability of potential in the y axis), the ion will be dragged up again since after some time the y axis potential will be what the x-axis potential was and vice versa. As a result, in spite of being able to trap the ion, we are faced with this micromotion that is directly related to the frequency of the potential rotating system.

The 2D ion crystals follow the Mathieu differential equation, which we can utilize to create simulations that can be used for a better understanding of the experimental system. As the Mathieu functions, which is the solution to the Mathieu equation implies, there exists not only the micromotion in the trapped ion but also secular motions which are associated with the normals among the ions. Nevertheless,

due to the nature of the convenience of a differential equations, we were able to extract some useful information out of the simulation.



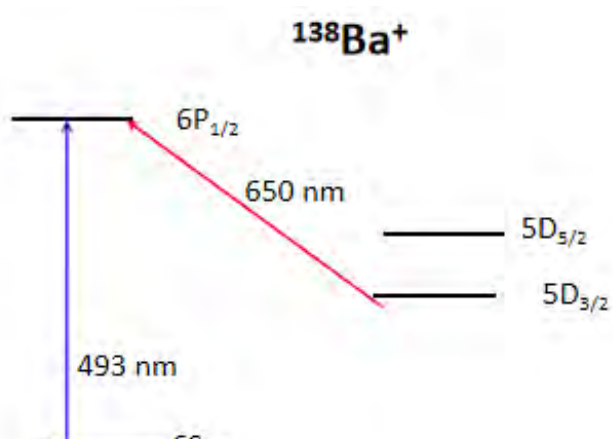
(Figure 2. Design and schematic layout of the ring trap)



(Figure 3. Potential diagram of the sectored-Ring trap)

**III. Methods**

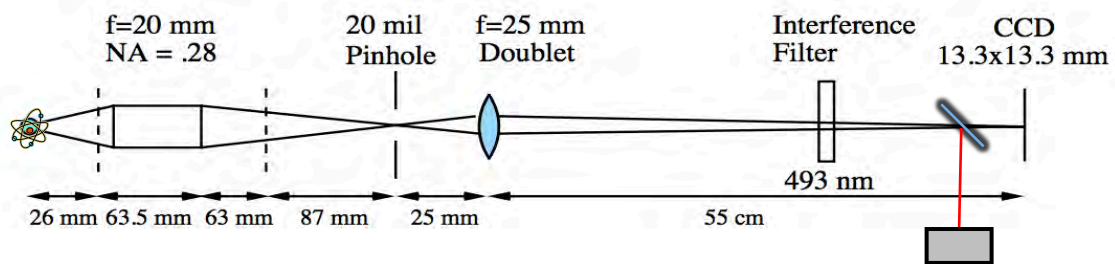
First, Ba atoms are ionized via a two-photon process, provided by a 791 nm ECDL and 337 nm pulsed Nitrogen laser beam. Another ECDL at 493nm and commercial 650 nm diode laser are for the cooling and repump transitions of Ba<sup>+</sup> ions. The prepared barium ion is then cooled to a very low temperature using Doppler laser cooling technique where we have the doppler cooling limit is the minimum temperature achievable with doppler cooling. Then the trapping design used is the Paul trap but with sectored ring electrodes to create two-dimensional ion crystals. The laser cooling of barium ions initiates a state transition cycle in the energy level of



barium ions, causing it to constantly fluorescence which can be capture by the

(Figure 4. Energy diagram of Barium ion)

Timepix3 camera that was used in this experiment. The camera data consisted of “events” which are the records of individual photons incident upon the camera. Integrating over millions of events then allows us to create an image of the trapped ions. Furthermore, for better centroding accuracy, we also post process the raw data to obtain the cluster size(number of pixels it hits) of the incident photon as well as the standard deviation of each centroid position of the cluser. The imaging system is shown in the following.

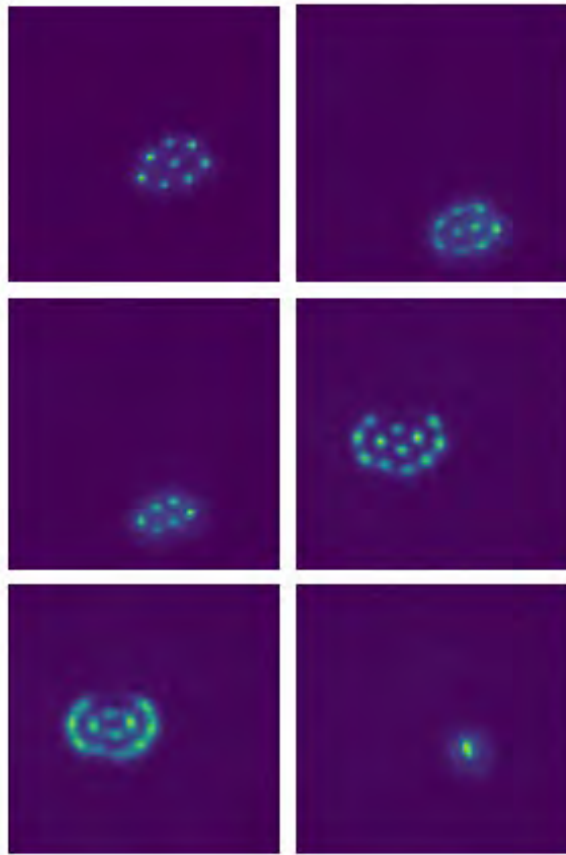


(Figure 4. detect fluorescence 493nm photons from our ions)

We also developed numerical modeling to determine the pseudopotential for arbitrary electrode geometries. We can model the equilibrium positions of the ions in the crystal to inform our electrode design. The pseudopotential equation is a function of the charge and mass of the ions, as well as the radio frequency that we used in the experiment.

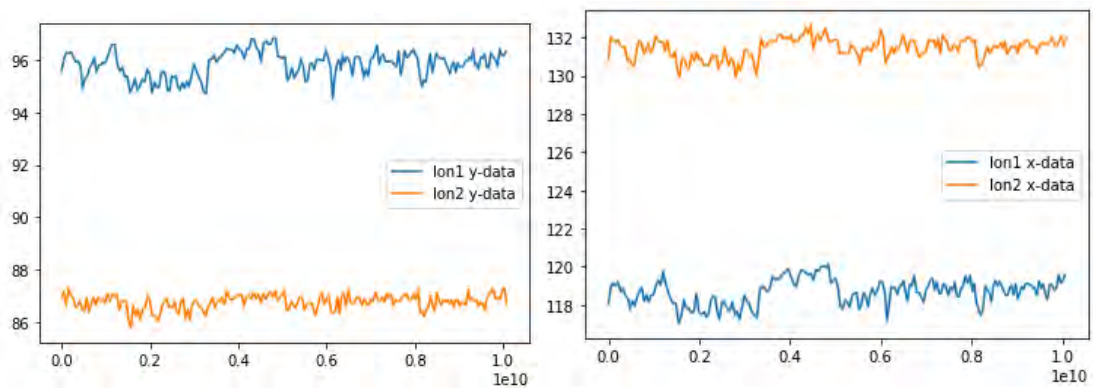


## IV. Results

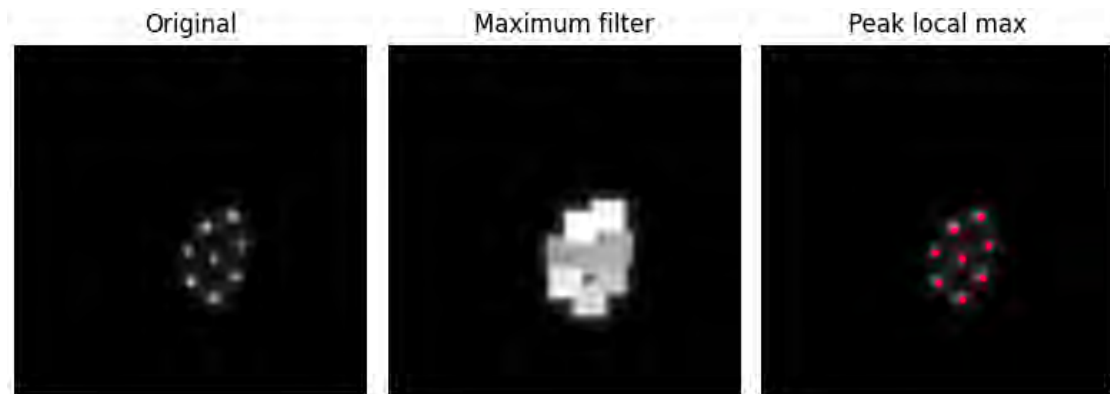


(Figure 6. Sample images of integrated data)

When taking a sample of data that have rapid movements and compare the two brightest ions, the two brightest ions are clearly moving in synchronization, as shown:

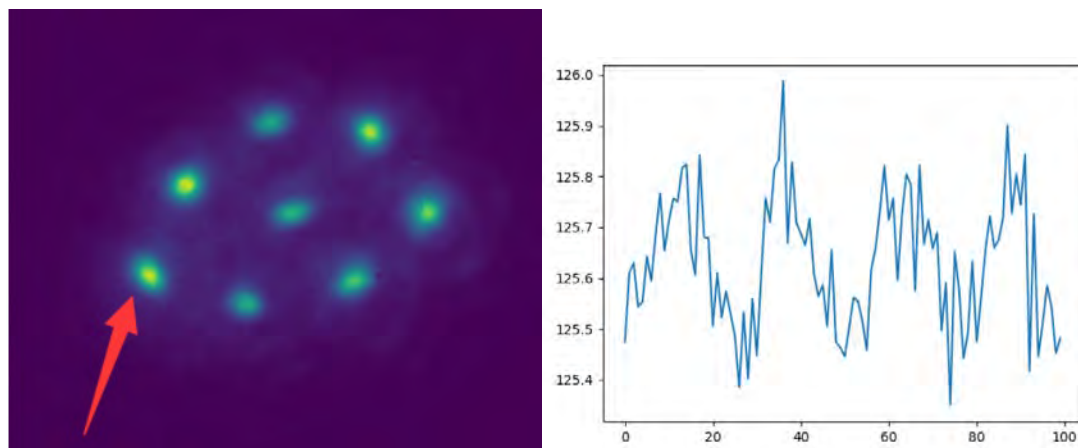


The movement was later on resolved by moving one of the resonators used in the trap. We have also developed a software that are used to calculate the centroid position of the integrated image. The centroid position was the result of peak finding in the intensity flux distribution over the camera, show in figure 6 below.

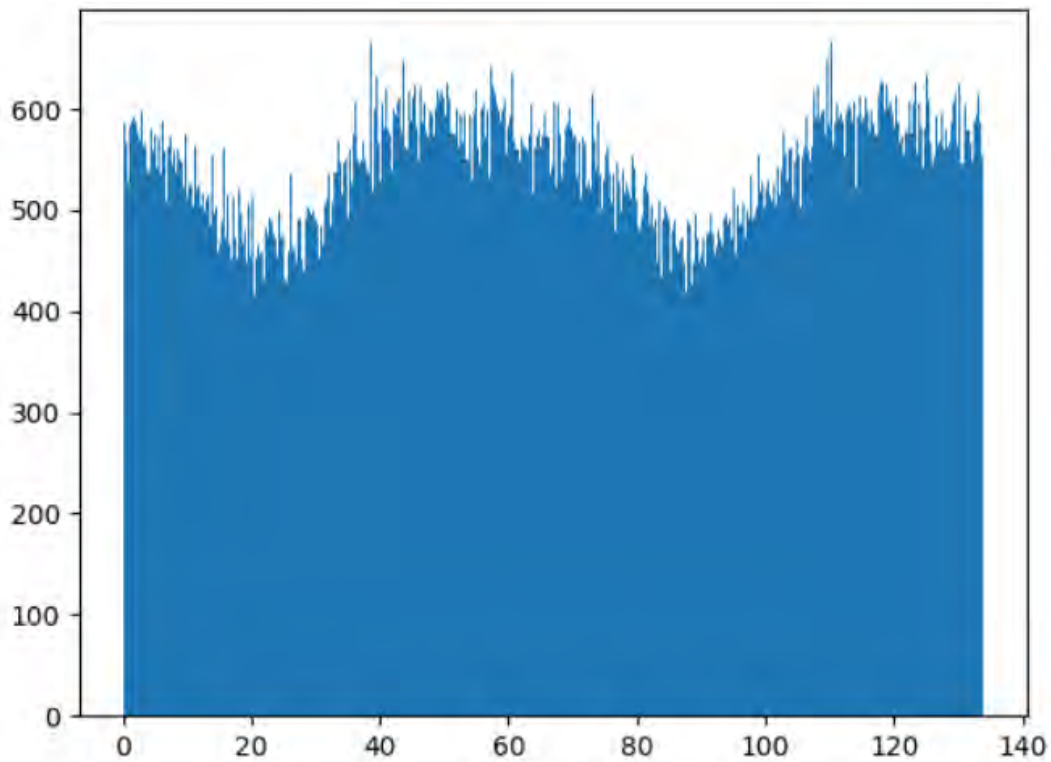


(Figure 7, Peak finding algorithm for centroding results)

We also measure the micromotion amplitude of the ion crystal. The following gives the microtion over four radio frequency period  $\sim 66$  times four ns.



(Figure 8. Left. The ion we are interested in; Right. Ion x-position as a function of frame (100 frames) which translates to 4 radio frequency period.)



(Figure 9. Photon flux over two radio frequency period)

## V. Conclusion

Ion trap is a very complicated process that still requires a lot of theoretical work, and so as a result, even with an extremely sensitive camera, we were not be able to some information that we intended for. In this research project, we worked on studying the motion of trapped ions, including the micromotions caused by the radio frequency that we use to drive the potential. We also study the phase transition of these ion crystals, by increasing the temperature until the crystal melts, and see if we can observe any patterns occurring. We observed that, as we increase the temperature before the crystal melts, that the inner shells and the outer shells form by large crystals begin to loose; In the sense that the inner shells and the outer shells rotate asynchronously. In the early stage of the project, the ion is moving unexpectedly violence, turns out this was caused by an experimental setup issue and was later fixed.

However, this motion persisted but at a much smaller scale, which could indicate the problem is not completely resolved. It is also our great interest to study the micromotion that is the result of the radio frequency, which we have measured the amplitude of the micromotions in each ion, this allows us to map the micromotions to the different crystal structures, which in turn can give us a sense on how well we can scale up this 2D ion crystal for quantum information processing.

## **VI. Acknowledgements**

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

I would also like to thank my mentor Andrei Nomerstki for his support and mentorship of this project, as well as Alexander Kato, my collaborator; Boris Blinov, my ion trap experimental advisor . Without them, I would not have learned and contributed much in just a summer.

## **VII. References**

1. M. Ivory, A. Kato, A. Hasanzadeh and B. Blinov, Review Of Scientific Instruments 91, (2020).
2. M. D’Onofrio, Y. Xie, A. Rasmusson, E. Wolanski, J. Cui and P. Richerme, Physical Review Letters 127, (2021).
3. H. Thomsen and M. Bonitz, Physical Review E 91, (2015).
4. L. D. Landau and E. M. Lifshitz, in Mechanics (Pergamon, New York, 1976), pp. 93–95. A Bautista-Salvador et al 2019 New J. Phys. 21 043011

# Quantum-Assisted telescopes

Ryan Mahon, Department of Physics and Astronomy, Stony Brook University,  
Stony Brook, NY 11794

Andrei Nomerotski, Physics Department, Brookhaven National Laboratory,  
Upton, NY 11973

August 11, 2021

## Abstract

The principles of quantum mechanics allow for better and more accurate measurements of photons and their respective wave functions. These quantum methods are applied to two photon interferometry where a light source is sent into a beam splitter where two photons are then split off and quantumly entangled. Through the Henry Brown & Twiss effect (HBT) we can check the correlation of these two photons entering different channels and their respective time differences. If everything works out, we should see a sharp Gaussian peak around the zero point of a histogram of data taken in the lab. We should also see strong coincidence correlations between the various channel pairs. These methods have many different applications to various fields such as cosmology, astrophysics and can specifically help with problems such as the question of dark matter. If these methods are fully realized, the more precise imaging could allow for better imaging of black holes which could help lead towards a crack in the mystery of dark matter. We have seen promising results where HBT peaks have been found along with other promising results from the analysis. At the same time however we are also seeing some unexpected results from the data which will require further study. These quantum methods allow for much more accurate and precise measuring than typical classical models and if fully realized will forever change the way we look at the universe.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Methods</b>	<b>5</b>
2.1	Lab Setup . . . . .	5
2.2	Methods for data analysis . . . . .	5
<b>3</b>	<b>Results</b>	<b>8</b>
<b>4</b>	<b>Conclusion</b>	<b>11</b>

# 1 Introduction

Since the very beginnings of civilizations throughout our history we have looked upon the stars with wonder. From Ptolemy to Aristotle, to Galileo and to the present day, millions of people everyday are curious about what lies and what exists beyond our relatively small planet. In order to be able to unravel the mysteries of the stars, we have had to develop methods, instruments and techniques to image the sky. The earliest telescopes constructed used principles of magnification and quickly gave excellent results. More recently we have used methods of classical interferometry to image the sky and they have given very good results. The best example of this is the first image of a black hole which was unveiled in April, 2019 which used these classical methods of interferometry with their technique VLBI (Very Long Baseline Interferometry) [1]. An example of a setup involving classical inteferometry can be seen in Fig. 1

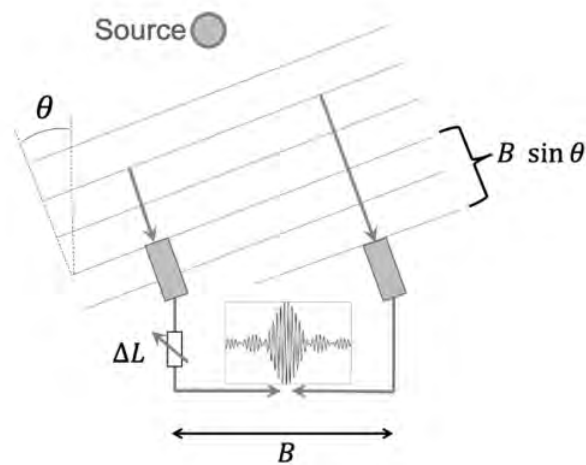


Figure 1: Setup using classical inteferometry methods [2]

While this classical interferometry has indeed done much to further our knowledge of the universe, and help us be able to see such incredible objects as the pre mentioned black hole, it currently limits the resolution of the received images. The angular resolution when using classical interferometry is defined as below.

$$\Delta\theta = \frac{\lambda}{B} \quad (1)$$

Where  $\Delta\theta$  is the optical resolution,  $\lambda$  is the wavelength of the received source and  $B$  is the separation between detectors. Current resolutions typically found using the visible wavelength range tend to be in the order of milli-arcseconds; however much greater resolution in the micro-arcsecond range would provide much better, clearer, and more precise imaging of astronomical bodies [2]. The current methods used with classical interferometry are not very efficient as they require a connection between detection sites. This required connection would need to run over a very large distance using classical methods in order to achieve the resolution wished for [3]. This is very costly in terms of both time and money and very few telescopes have ever been constructed using classical methods meeting these requirements.

The research worked on throughout this internship seeks to increase the efficiency of interferometry by looking to quantum methods instead of solely relying on classical methods. Classical interferometry currently is based upon a single photon source which is interfered with itself [2], this can be seen in Fig. 1. The proposed method of our research group is to eliminate the issue of the connection between the detectors altogether. This is possible if two different entangled photons experience interference from one another. These two photons are connected through the phase difference between them, and upon hitting multiple detectors, if they both hit each detector within a small enough time interval, we consider this to be a coincidence [2]. The Henry Brown & Twiss (HBT) effect predicts we should see peaks of coincidences at some specific time difference value, where the time difference is the difference in time between the first photon and the second photon hitting their respective detectors. We also expect to see correlations between the pair rates according to equation (2). [2]

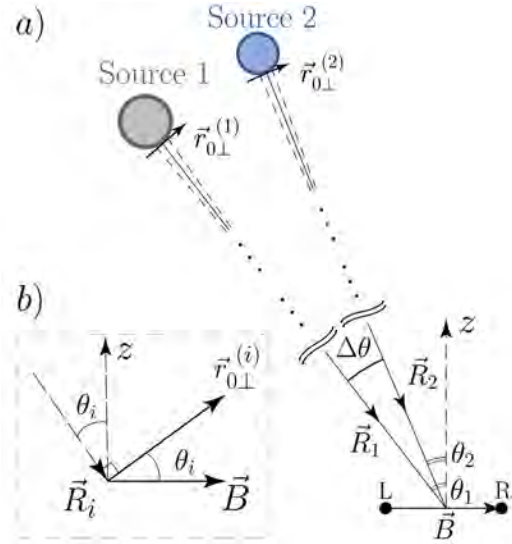


Figure 2: New 2 photon interferometry proposed setup [2]

$$\langle N(xy) \rangle = \frac{k((S_1 + S_2)^2)}{8} [1 \pm V_{2PS} \cos(\delta_1 - \delta_2)] \quad (2)$$

Where S stands for the flux densities of the 2 photon sources,  $\Delta\nu$  is the band-width of the detector, k is a constant, V is the fringe visibility of the pair and N is the number of pair events. This equation is then extended to various different channel pairs through the following 4 equations in equation (3).

$$\begin{aligned} \langle N(13) \rangle (t) &= \overline{N(13)} [1 + V_{13} \cos(\Delta\delta(t))] \\ \langle N(14) \rangle (t) &= \overline{N(14)} [1 - V_{14} \cos(\Delta\delta(t))] \\ \langle N(23) \rangle (t) &= \overline{N(23)} [1 + V_{23} \cos(\Delta\delta(t))] \\ \langle N(24) \rangle (t) &= \overline{N(24)} [1 - V_{24} \cos(\Delta\delta(t))] \end{aligned} \quad (3)$$

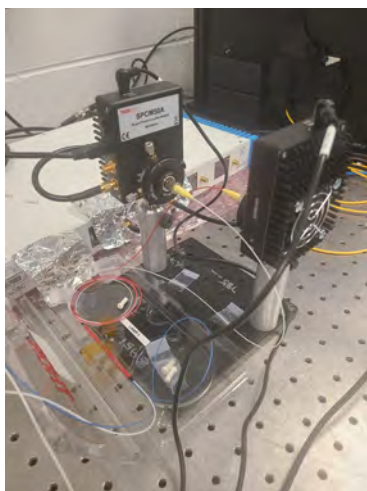
From these equations we can see which pair rates should be correlated or anti correlated with other pair rates. It is clear to see that channels 1 and 3 should have pair rates anti-correlated with channels 1 and 4 and channels 2 and 3 should be correlated with channels 1 and 3. These theories were tested throughout the time of my research and the hope of the research going in was to confirm these theories for the new SPAD detector setup.



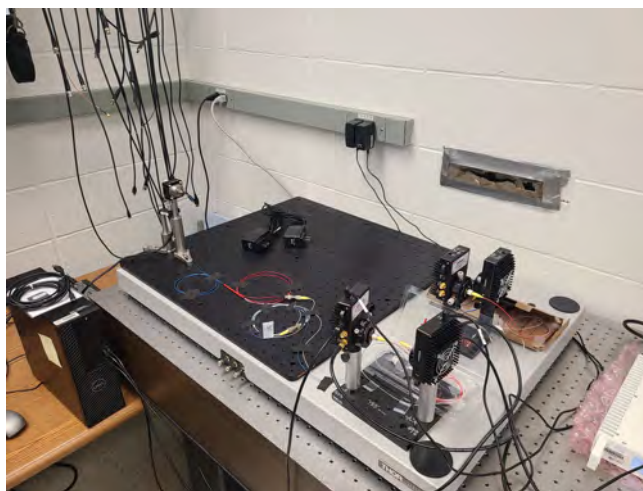
## 2 Methods

### 2.1 Lab Setup

Soon before starting this internship, the old nanowire setup my research group used had broken down; as a result the group decided to switch from the nanowire setup to a SPAD detector setup. This meant that the lion's share of my research was focused on testing the SPAD detectors and making sure the results of the data analysis aligned with the theories stated in the introduction. Figures 3 shows photos from the lab of the new SPAD detector setup.



(a) First Photo



(b) Second Photo

Figure 3: Laboratory setup with SPAD detectors

You can see from Fig. 3b that there are 4 detector channels in the setup. As photons are sent into the configuration from one of the lamps, it is possible for the photons to end up hitting in any one of the 4 channels. The SPAD detectors log the time of arrival of each photon and which channel the photon entered. Using the software quTAG [4], the data can be recorded as data files with separate columns for the time and location of hits.

### 2.2 Methods for data analysis

Using the data files created by quTAG, the next step is to analyze the data. The data analysis I performed was done primarily using Python [5] with the software Jupyter

Notebook [6]. Analyzing the data using python allowed for computation of the large data gathered in reasonable time periods and for construction of plots, and graphs to illustrate the results of the analysis. Specific libraries used often in our analysis which played an instrumental role in retrieving the data includes Numpy [7], Scipy [8] and Matplotlib [9]. Numpy allowed for easy computation of the data arrays and loops throughout the code; Scipy was primarily used to construct the curve fits and Matplotlib was used to produce every plot generated.

A significant amount of the data analysis performed throughout the internship was performed using python code or modifying code written by previous SULI interns Jonathon Schiff and Alex Parsells for the research group [10]. The first step in conducting the data analysis for large data files was to load in the time and channel arrays using Numpy. After assigning variables, to find coincidences, I would check two time values logged in the time array to see if both corresponded to specified separate channels. If this was true and the other condition of both hits being within a specific small interval were met, the difference between the times of the recorded hits would be inputted into a new array logging the time differences. A histogram of the time difference would then be compiled by adding the time differences arrays for all files in the data set.

The next important area to look at to test the theory was the peak pair rates. This was done by looking at values within a small specified range away from the peak of an HBT peak. Typically a sigma range of 1.5 nanoseconds was used. We once again loaded in the time and channel arrays, and this time also loaded in sigma arrays which housed the sigma values from the curve fits of all channel combination time difference plots. The next step is to check if the time difference of two hits in the time array fall within this specific sigma range both to the left and to the right of the HBT peak. If so, the time of these hits, the time difference of the hits and their respective channels are loaded into a brand new array. From the results of this new array, it would then be possible to construct plots to look at the peak pair rates for all channels, the normalization rates for

all channels and scatter plots checking for correlations or anti correlations between the peak pair rates according to equation (3).

Early on in the research, I was asked to develop a function for the convolution of a Lorentzian, Gaussian function and an exponential function. This was accomplished by using the following definition of a convolution.

$$\phi(t) = (f(t) * g(t)) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau \quad (4)$$

Where in this case f(t) is defined as

$$f(t) = e^{-\frac{|t|}{\tau_c}} \quad (5)$$

and g(t) is defined as

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} \quad (6)$$

therefore, the convolution function for these two equations can be defined as

$$\phi(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{|\tau|}{\tau_c}} e^{-\frac{(t - \tau)^2}{2\sigma^2}} d\tau \quad (7)$$

When evaluated, this brings about the following convolution function.

$$\phi(t) = \frac{1}{2} \left[ e^{-\frac{t^2}{2\sigma^2} + \frac{(-\frac{2\sigma^2}{\tau} - 2t)^2}{8\sigma^2}} \left( 1 + \operatorname{erf} \left( \frac{(-\frac{2\sigma^2}{\tau} - 2t)}{2^{3/2}\sigma} \right) \right) + e^{-\frac{t^2}{2\sigma^2} + \frac{(\frac{2\sigma^2}{\tau} - 2t)^2}{8\sigma^2}} \left( 1 - \operatorname{erf} \left( \frac{(\frac{2\sigma^2}{\tau} - 2t)}{2^{3/2}\sigma} \right) \right) \right] \quad (8)$$

A full derivation of this convolution function typed up by myself can be found using the following link:

<https://1drv.ms/b/s!AsxKlvnwU712jZJclaokTQaee-Lc8w>

(Note: the pdf in onedrive occasionally doesn't format equal signs or other areas correctly; for the best experience viewing the above link, download the pdf file to view with your own software).

### 3 Results

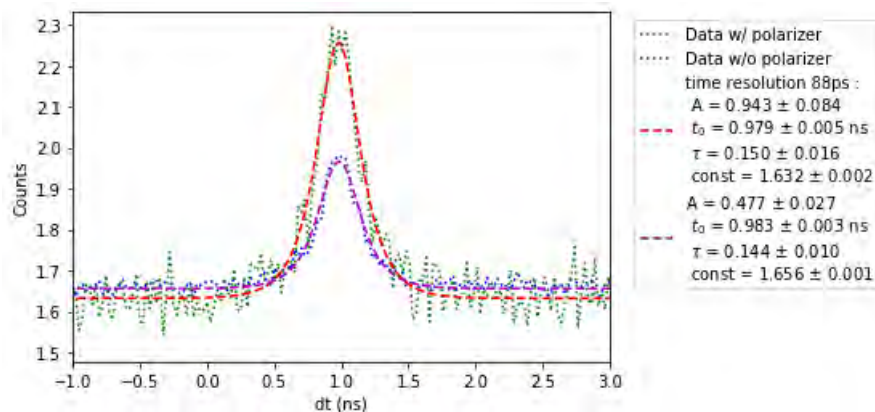


Figure 4: HBT peaks for both polarized and non-polarized data fitted by convolution function

The figure above shows an HBT peak between two channels. As expected from the discussion in the introduction, we are seeing a peak at a time difference value of 1ns. This means that most coincidences for this channel pair occur with 1ns separation. Two separate peaks can be seen: one is a peak of polarized data and one is a peak of non-polarized data. It can be seen that the polarized data gives a higher amplitude than the non-polarized data by about a factor of 2, which is to be expected. The visibility of this data which is defined by the amplitude divided by the constant is therefore also twice as big. The visibility is able to tell us how visible the peak is against the background.

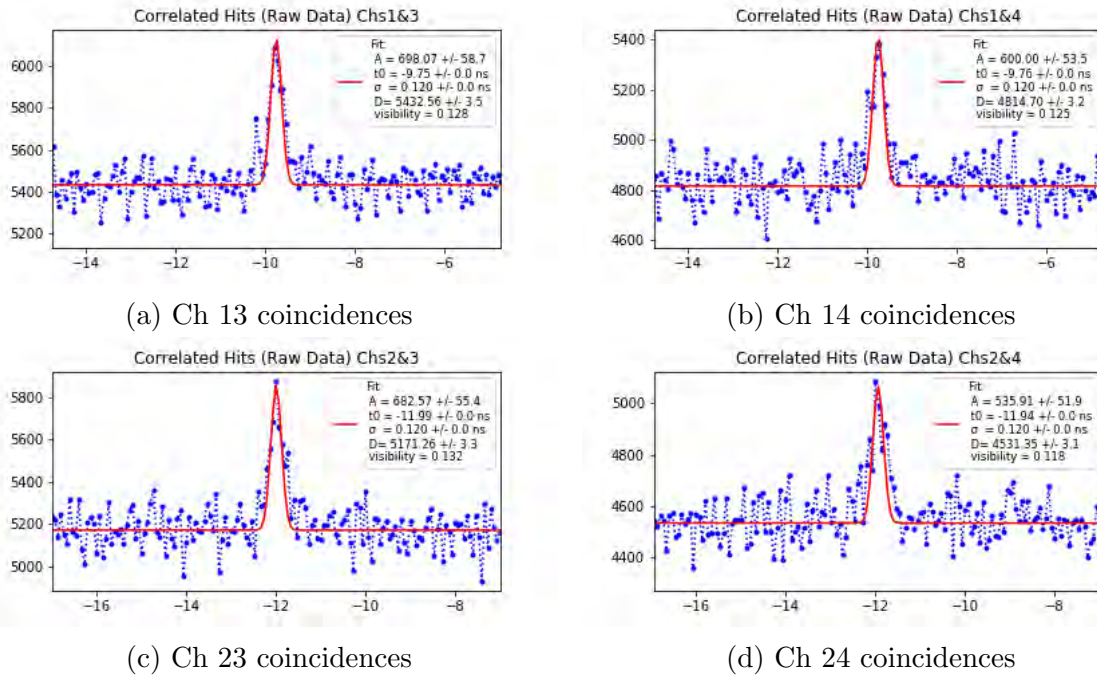


Figure 5: Time difference plots showing HBT peaks of coincidences for 4 different channel combinations from 60 min non-polarized data set

Fig. 5 shows HBT peaks for every channel combination as described in equation (3). These peaks were fitted using Gaussian curves. The offset of about -10ns is due to different cable lengths used in the experiment. Seeing HBT peaks for these channel combinations is crucial in confirming the proposed theories. Uncertainties in the fits were calculated by taking the square root of the proper output from the Scipy curve fitter. The data set these peaks were generated from was as described in the caption and included 2 lamps.

In addition to this data set, I also analyzed a 60 minute polarized 2 lamp data set, 2 separate single lamp 30 minute data sets where each had 1 of the lamps turned on and the other turned off, and 2 more of the same data sets but with polarized light sources. I also looked at 5 more data sets of 30 minutes with specific conditions such as a polarized light source with one lamp and a factor of 2 lower rate from the lamp. All of these data sets showed HBT peaks throughout just like shown in Fig. 5. An interesting aspect of these data sets to look at and compare was the visibility of every channel combination of each data set. Fig. 6 below shows these results.

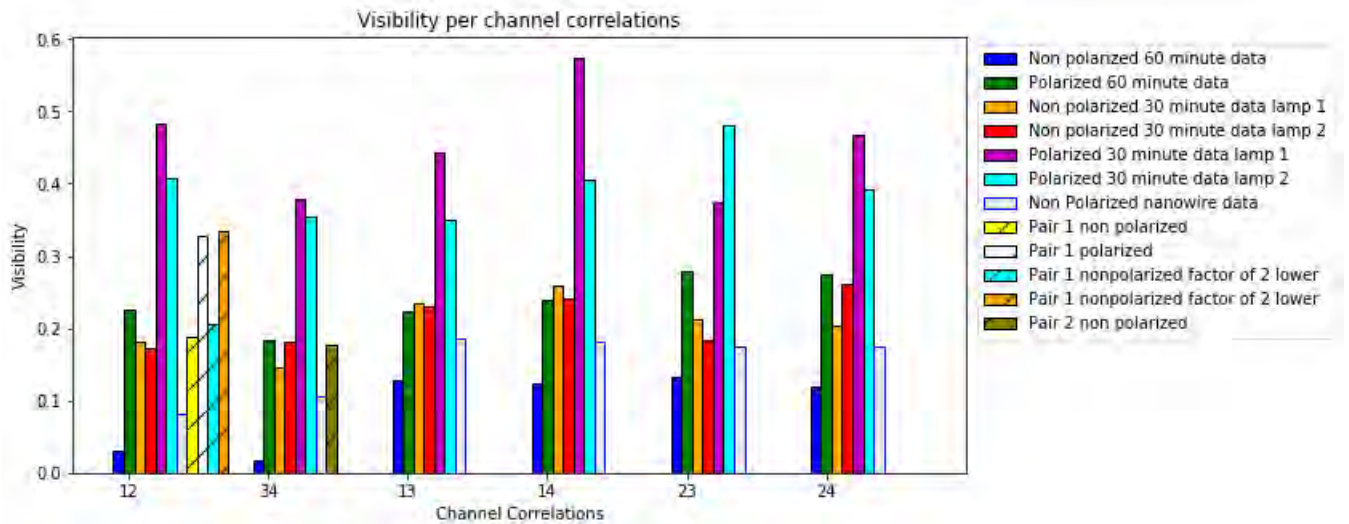
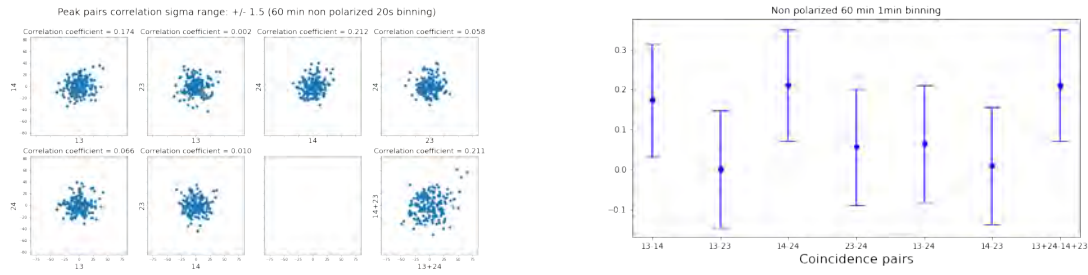


Figure 6: Visibilities of each channel pair for every data set analyzed throughout my internship as well as some previous nanowire data provided from before I joined

This graph shows some very interesting and peculiar results. The first aspect to consider is how does the visibility change between non-polarized and polarized data sets. We are seeing here the results we would expect as we are seeing visibilities twice as high for polarized data sets as their respective non-polarized data sets. Many of the other results from this graph are a bit puzzling though. First we look at the fact that for the non polarized 60 min SPAD detector data and also the nanowire data, we are seeing Ch12 and Ch34 showing smaller visibilities than in all the other channel pairs. The next puzzling result is that we are seeing single lamp data sets with much higher visibilities than the 2 lamp data sets. These results will have to be looked into and researched going forward past the time of my internship.



(a) Scatter plot to determine correlations

(b) Summary of scatter plots showing Pearson coefficients for each channel pair.

Figure 7: Scatter plots and Pearson coefficients plot to check for correlations in peak pair events.

Fig. 7 shows plots to check for correlations or anti correlations in the peak pair events. From equation (3) we know that we should expect to see anti correlations between channel pairs 13 and 14, and correlations between pairs 13 and 23, and similarly for other combinations as described in equation (3). The graphs above were once again from the 60 minute non-polarized data set. We are not seeing the expected behaviors from the graphs above. If 2 pairs are anti correlated, you should see values of the first pair get lower as the other gets higher, and the opposite should be true for correlated pairs. This is not seen here though as even pairs 13 and 14 which should show definite anti correlation have a positive Pearson [11] coefficient which indicates correlation. Oddly in fact, none of these channel pair combinations show any negative Pearson coefficient values which means no anti correlation was seen anywhere for the peak pair events in this data set. Understanding the cause of this will need to continue to be looked into and researched.

## 4 Conclusion

Understanding our world and our universe have been intrinsic characteristics of humankind from the very beginnings of our civilizations. One of the most triumphant ways to exploit this innate curiosity has been to look out into the cosmos and take a glimpse at the wonder that is out there. Classical methods of interferometry have done wonders in this regard and the achievements from this method should not be discounted. However, the cost and inefficiency of building telescopes with these classical methods prevent us



from being able to regularly receive imaging of very precise resolutions. The new 2 photon interferometric method proposed in our research has the potential to revolutionize how we image stellar objects and could massively reduce the costs of building telescopes with precise resolutions. The research I performed this summer revealed mixed results for the new SPAD setup as we indeed did see the HBT peaks everywhere predicted by the theory behind the research, but we did not see the peak pair event correlations the way we expected to. We also were seeing peculiar results with the visibilities of various different data sets. Research and study will need to continue to unravel the mysteries behind these results and bring us closer to achieving the full potential of this research.

## Acknowledgments

I would like to thank my mentor Andrei Nomerotski for all the help he gave me with this research and for giving the opportunity to gain insight into this area of physics which I had never known before. I would also like to thank other members of the research group Paul Stankus, Stephen Vintskevich and Michael Keach for assistance they have given me as well. I also want thank fellow intern Zhi Chen and former interns Jonathon Schiff and Alex Parsells for the same reasons. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## References

- [1] O. Lutz, *How scientists captured the first image of a black hole*, <https://www.jpl.nasa.gov/edu/news/2019/4/19/how-scientists-captured-the-first-image-of-a-black-hole/>, Accessed 2021-08-09, Apr. 2019.
- [2] P. Stankus, A. Nomerotski, A. Slosar, and S. Vintskevich, “Two-photon amplitude interferometry for precision astrometry,” *arXiv preprint arXiv:2010.09100*, 2020.



- [3] A. Nomerotski, P. Stankus, A. Slosar, S. Vintskevich, S. Andrews, G. Carini, D. Dolzhenko, D. England, E. Figueroa, S. Gera, *et al.*, “Quantum-assisted optical interferometers: Instrument requirements,” in *Optical and Infrared Interferometry and Imaging VII*, International Society for Optics and Photonics, vol. 11446, 2020, p. 1 144 617.
- [4] *Qutag time-to-digital converter manual*, English, version 1.4, Apr. 2019.
- [5] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009, ISBN: 1441412697.
- [6] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, and C. Willing, “Jupyter notebooks – a publishing format for reproducible computational workflows,” in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, F. Loizides and B. Schmidt, Eds., IOS Press, 2016, pp. 87–90.
- [7] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. DOI: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2). [Online]. Available: <https://doi.org/10.1038/s41586-020-2649-2>.
- [8] E. Jones, T. Oliphant, P. Peterson, *et al.*, *SciPy: Open source scientific tools for Python*, 2001–. [Online]. Available: <http://www.scipy.org/>.
- [9] J. D. Hunter, “Matplotlib: A 2d graphics environment,” *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007. DOI: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- [10] J. Schiff and A. Parsells, *Quantum-astrometry*, <https://github.com/jschiff11/Quantum-Astrometry>, 2021.

- [11] “Pearson’s correlation coefficient,” in *Encyclopedia of Public Health*, W. Kirch, Ed. Dordrecht: Springer Netherlands, 2008, pp. 1090–1091, ISBN: 978-1-4020-5614-7. DOI: [10.1007/978-1-4020-5614-7\\_2569](https://doi.org/10.1007/978-1-4020-5614-7_2569). [Online]. Available: [https://doi.org/10.1007/978-1-4020-5614-7\\_2569](https://doi.org/10.1007/978-1-4020-5614-7_2569).

**Cosmic rays in astronomical charge-coupled devices**

Jocelyn McMahon

Physics, Susquehanna University, Selinsgrove PA 17870

Andrei Nomerotski

Cosmology and Astrophysics Group, Brookhaven National Laboratory, Upton, NY, 11973

## **ABSTRACT**

The cosmic rays in astronomical charge coupled devices (CCDs) project at Brookhaven National Lab (BNL) examines the energy deposition tracks of various particles to improve the functionality and capabilities of the Large Synoptic Survey Telescope's (LSST) camera. Using data from the Hyper Suprime-Cam Subaru telescope in Hawaii, we mapped the energy deposition of various particles. The particle type was identified by the shape and intensity of the track it left. We examined the distributions of energy deposition, fit the data, and calculated the deposition per unit length to determine values of beta and the subsequent velocity of the particle. The code developed in this project will increase the quality of data analysis at the LSST, which will use the same kind of CCD sensors. Improved algorithms would benefit the classification and identification of various particles, and the code can continue to be refined at BNL for the LSST and current uses.

## I. INTRODUCTION

Cosmic rays are high-energy protons or atomic nuclei that travel near the speed of light, originating from our sun, outside the solar system, and other galaxies. When cosmic rays collide with atomic nuclei in the upper atmosphere of Earth, pions and kaons are created. These particles then quickly decay to muons. A muon has an average lifetime of 2.2 micro seconds and travels near the speed of light.<sup>1</sup> Muons can reach the surface of Earth due to special relativity. Protons from cosmic rays also travel through the atmosphere and reach Earth's surface. When particles from the upper atmosphere are observed in sensors like charge-coupled devices, or CCDs, much of the flux captured in these devices at sea level are from muons.<sup>2</sup> In this project, the tracks of energy deposition from muons and protons in CCDs were examined.

CCDs are sensors that convert light into electrons. Light and particles pass through these sensitive semi-conductor elements whose energy can be converted into electrons.<sup>3</sup> Muons and protons are some of these particles. Their energy that is captured in the CCDs can then be converted into images represented by thousands of square pixels. The energy can be seen in these images as bright tracks in the otherwise dark image. Tracks created by particles from cosmic rays rely on sharp edges in the imaging to be identified.

Data used in this project was collected at the Prime Focus Spectrograph (PFS) of the Subaru telescope and was provided to BNL by Ropert Lupton.<sup>4</sup> The CCDs at the PFS have a

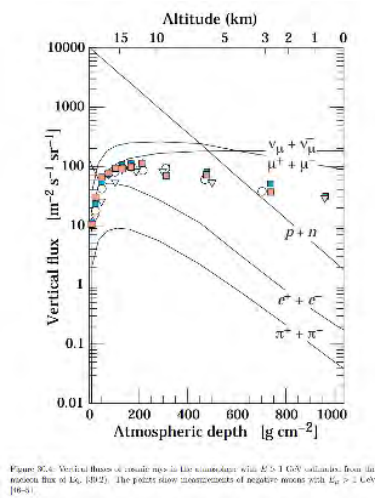
<sup>1</sup> D. B. Chitwood, T. I. Banks, M. J. Barnes, S. Battu, R. M. Carey, S. Cheekatmalla, ... & MuLan Collaboration, Improved measurement of the positive-muon lifetime and determination of the Fermi constant, *Physical review letters*, 99(3), 032001 (2007).

<sup>2</sup> M. Fisher-Levine and A. Nomerotski, "Characterising ccds with cosmic rays", *Journal of Instrumentation*10, C08006 (2015).

<sup>3</sup> M. Fisher-Levine & A. Nomerotski, Characterising CCDs with cosmic rays. *Journal of Instrumentation*, 10(08), C08006 (2015).

<sup>4</sup> R. Lupton (2021). PFS cosmics tracks.

pitch of 15 microns and a thickness of 200 microns. The PSF is at approximately 4 km above sea level.<sup>5</sup> At that altitude, the amount of flux from muons and protons can be predicted.



**Figure 1.** Predicted flux of various particles, including muons,  $\mu^+ + \mu^-$  and protons, p.

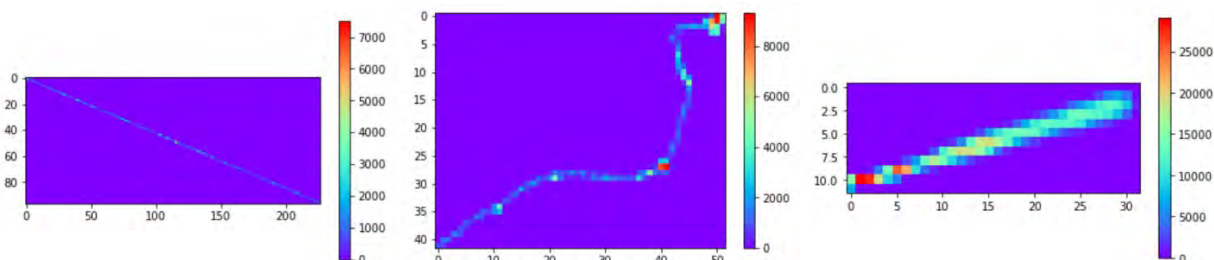
Figure 1 provides information on cosmic ray spectra and composition as they interact with the atmosphere.<sup>6</sup> It shows the predicted flux from various particles, including the muons and protons that are the focus of this project, at various altitudes. The solid lines represent the predicted amount of flux from those particles, and the points represent measure values. At sea level, muons are the dominant particle creating flux, but at 4 km above sea level, flux from protons is approximately 10 times greater than at sea level.<sup>7</sup> Proton flux is still expected to be significantly smaller than flux from muons, but the increase explains why protons could be more prominent in the data set.

<sup>5</sup> H. Sugai, H. Karoji, N. Takato, N. Tamura, A. Shimono, Y. Ohshima, ... & C. Yan. H, Prime focus spectrograph: Subaru's future. Ground-based and Airborne Instrumentation for Astronomy IV, 8446(84460Y) (2012).

<sup>6</sup> J. Beringer, Particle data group. Phys. Rev. D, 86(010001) (2012).

<sup>7</sup> J. Beringer, Particle data group. Phys. Rev. D, 86(010001) (2012).

There are three types of tracks that can be found in CCD sensors: straight tracks created by high-energy muons, curved tracks from electrons, and “fat” tracks, that are hypothesized to be created by protons.



**Figure 2, 3, and 4.** From left to right is “fat”, straight, and curved tracks.<sup>8</sup>

The focus of this project is on the “fat” tracks to test if they are created by protons. Particles that produce “fat” energy deposition tracks are hypothesized to be protons. “Fat” tracks have higher energy deposition over an observably wider track. Python Jupyter code was used to filter out extraneous straight and curved tracks, isolate “fat” tracks for analysis, and mathematically examine the properties of these tracks.

Protons are the primary component of cosmic rays, with a mass of 0.938 GeV. They have an average energy of 1 GeV at sea level once they have traveled through and lost some energy to the atmosphere.

<sup>8</sup> M. Fisher-Levine & A. Nomerotski, Characterising CCDs with cosmic rays. *Journal of Instrumentation*, 10(08), C08006 (2015).

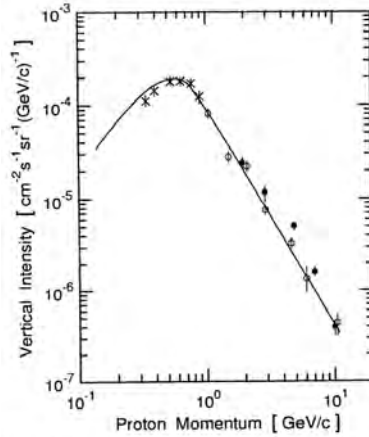


Figure 3.2: Differential momentum spectrum of protons at sea level (Barber et al., 1980).

**Figure 5.** Differential momentum spectrum of protons at sea level.<sup>9</sup>

Figure 5 shows the intensity vs momentum of protons at sea level. A similar graph for muons could not be found, but are expected to have a mean energy of 4 GeV at sea level. Muons have a mass of 0.106 GeV, and they lose approximately 2 GeV in the atmosphere.<sup>10</sup> This provides the basis for understanding the range of energy these two particles can have as they pass through the PFS's CCD sensors.

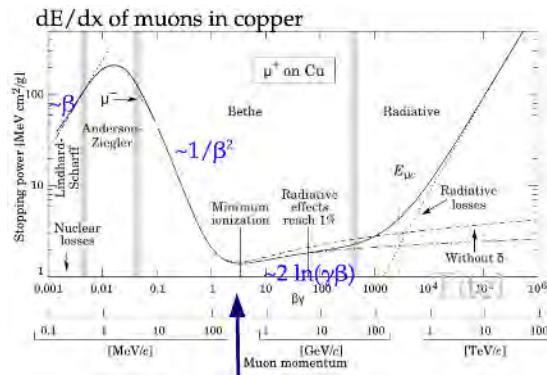
This energy can be expressed in terms of stopping power, which shows how much energy a particle loses while moving through matter.<sup>11</sup>

<sup>9</sup> J. Beringer, Particle data group. Phys. Rev. D, 86(010001) (2012).

<sup>10</sup> M. Fisher-Levine & A. Nomerotski, Characterising CCDs with cosmic rays. Journal of Instrumentation, 10(08), C08006 (2015).

<sup>11</sup> A. A. Correa, Calculating electronic stopping power in materials from first principles. Computational Materials Science, 150, 291-303 (2018).





**Figure 6.** Stopping power vs. muon momentum, or energy deposition per unit length, in copper.<sup>12</sup>

Stopping power strongly depends on  $\beta$ , which is defined as  $\frac{v}{c}$ , where  $v$  is the velocity of the particle, and  $c$  is the speed of light, as well as  $\gamma$ , defined as  $\frac{1}{\sqrt{1-\beta^2}}$ .<sup>13</sup> Based on the energies of muons and protons at approximately 4 km above sea level, the range 0.1 to 10 GeV is of interest to this project. For that range, the stopping power depends on the equation  $1/\beta^2$ . For low energies,  $\gamma = 1$ , so the equation depends on  $\beta$ . At high energies,  $\beta = 1$ , so the equation is dependent on  $\gamma$ . Protons are expected to have a small  $\beta$  to deposit more energy than muons with  $\beta = 1$ . Combined with knowledge from Figure 5, muons with low  $\beta$  are expected to be very rare, and protons with  $\beta = 1$  are very rare.

## II. METHODS

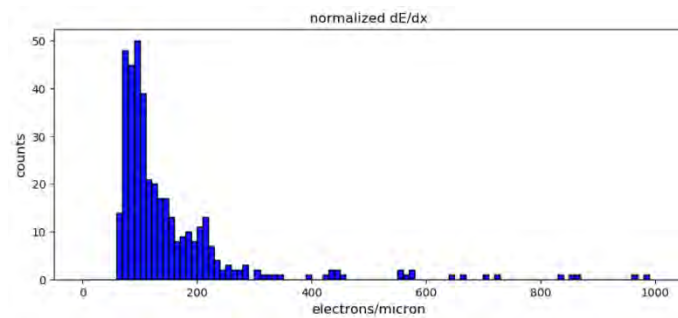
Beginning with 406 calibrated track cutouts provided by Robert Lupton, each track was analyzed using Jupyter code to see if it fit the expected properties of a track created by a proton.

<sup>12</sup> J. Beringer, Particle data group. Phys. Rev. D, 86(010001) (2012).

<sup>13</sup> M. Fisher-Levine & A. Nomerotski, Characterising CCDs with cosmic rays. Journal of Instrumentation, 10(08), C08006 (2015).

The length of the track was estimated assuming the track was a straight cutout diagonal. The expected number of hit pixels was compared to the measured number of hit pixels to reject curved tracks. The tracks that met the selection requirements were corrected for entrance angle and thickness of the sensor. Then the total charge deposition was calculated via integration of each pixel charge. Energy deposition was assumed to be constant to determine the normalized energy deposition per unit length, represented by  $\frac{dE}{dx}$  in units of  $e^-/\text{micron}$ .

The “fat” tracks were selected for by checking if  $\frac{dE}{dx}$  was greater than the threshold value of  $400 e^-/\text{micron}$ .<sup>14</sup>

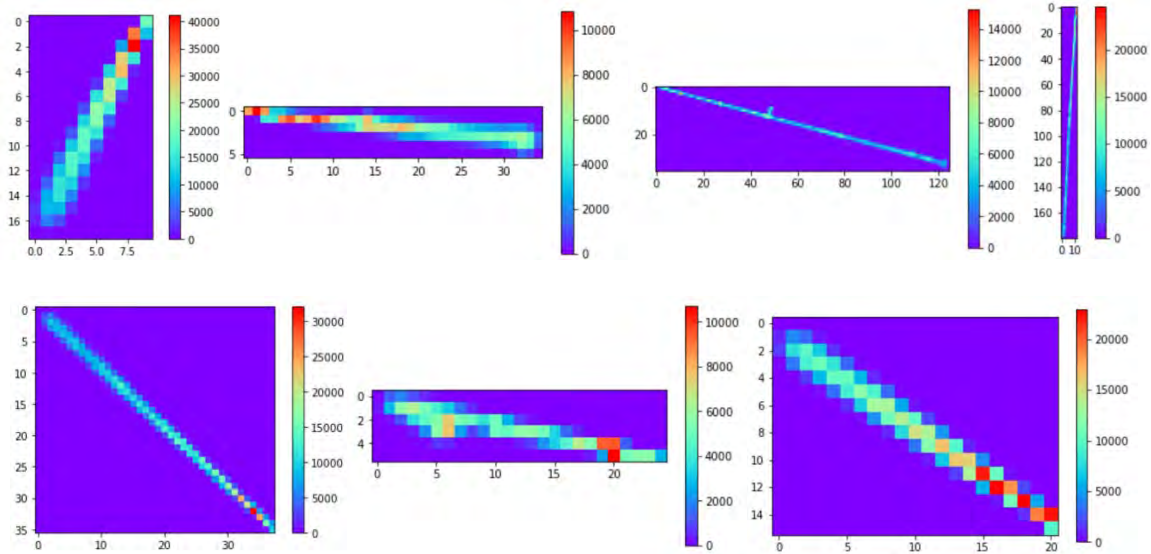


**Figure 7.** The normalized values of  $\frac{dE}{dx}$  in  $e^-/\text{micron}$ .

This selection gives 26 possible “fat” tracks from the 406 original tracks, or 6.4% of the tracks.

Below are examples of the “fat” tracks that were selected.

<sup>14</sup> M. Fisher-Levine & A. Nomerotski, Characterising CCDs with cosmic rays. *Journal of Instrumentation*, 10(08), C08006 (2015).



**Figure 8.** Seven of the “fat” tracks separated from the data set.

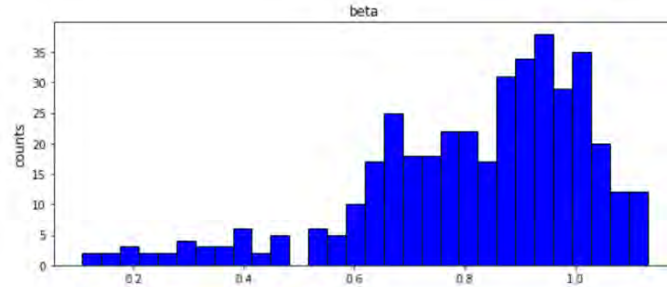
After calculating the values of  $\frac{dE}{dx}$ , values of  $\beta$  for each selected track can then be estimated with the equations:

$$\frac{dE}{dx} = \frac{\alpha}{\beta^2} \quad \text{or, solving for } \beta: \quad \beta = \sqrt{\frac{\alpha}{dE/dx}}$$

where  $\alpha$  is a constant and  $\beta = v/c$ .<sup>15</sup> The normalized energy deposition for a certain material can be calculated directly for the sensor. The CCDs at the PFS are made of silicon, which gives an expected value of  $\alpha$  to be  $80 e^-/\text{micron}$  for the sensors used in this project.<sup>16</sup> Using the peak value from Figure 5, the value of  $\alpha$  was chosen as  $100 e^-/\text{micron}$ . This would give  $\beta$  values close to 1. Calculating these values for each track, it was expected that they would be less than one as nothing can travel faster than the speed of light.

<sup>15</sup> D. B. Chitwood, T. I. Banks, M. J. Barnes, S. Battu, R. M. Carey, S. Cheekatmalla, ... & MuLan Collaboration, Improved measurement of the positive-muon lifetime and determination of the Fermi constant, Physical review letters, 99(3), 032001 (2007).

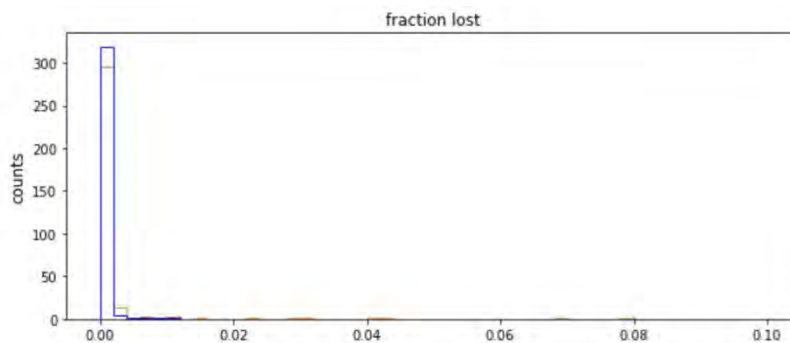
<sup>16</sup> M. Fisher-Levine & A. Nomerotski, Characterising CCDs with cosmic rays. Journal of Instrumentation, 10(08), C08006 (2015).



**Figure 9.** Distribution of  $\beta$  for all selected tracks.

Values greater than one can be attributed to uncertainty in the calculations. The peak near 0.9 is significant, representing that the particles are mostly traveling at approximately 90% the speed of light. This is slightly less than the average velocity of muons from cosmic rays,<sup>17</sup> and could point to protons being present in the data set.

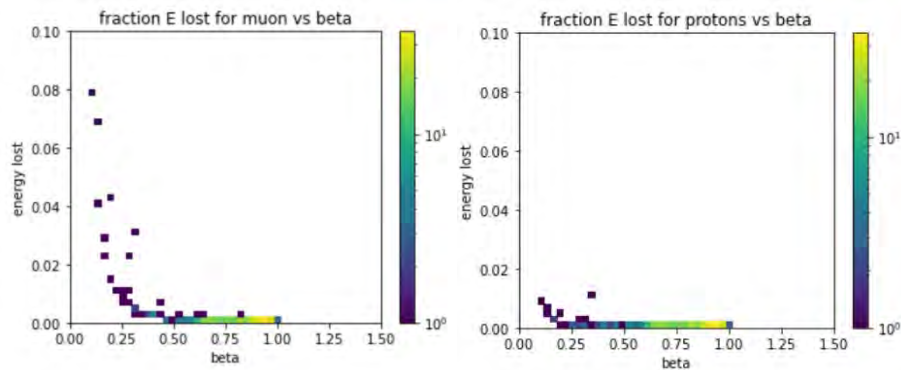
It is then possible to visualize the relationship between  $\beta$  and the fractional energy lost by the particle.



**Figure 10.** Fractional energy lost for muons (blue) and protons (orange). Energy lost for most tracks, both for muon and hypothesized protons, is small.

<sup>17</sup> D. B. Chitwood, T. I. Banks, M. J. Barnes, S. Battu, R. M. Carey, S. Cheekatmalla, ... & MuLan Collaboration, Improved measurement of the positive-muon lifetime and determination of the Fermi constant, Physical review letters, 99(3), 032001 (2007).

This histogram shows the fractional energy lost for each predicted type of particle. From the tracks selected, each had its momentum calculated as if it was a muon or a proton, using the masses of each respectively. The fractional energy lost was then determined by dividing the total signal in MeV by the calculated momentum.



**Figure 11 and 12.** For muons, as  $\beta$  increases towards one, the energy loss decreases as  $\frac{1}{\beta^2}$

towards zero. For protons, as  $\beta$  increases to one, the energy loss decreases as  $\frac{1}{\beta}$  towards zero.

For both masses, the tracks showed only a small amount of energy lost. For muons, since the energy lost for a particle with high  $\beta$  is approximately 0,  $\beta$  can be assumed to be constant.

Protons have a low energy loss for most values of  $\beta$ , so  $\beta$  can also be assumed to be constant for protons. These two observations mean that the particle slowing does not have to be taken into account in any calculations.

### III. RESULTS

Selecting for “fat” tracks in the data set gave valuable information on which variables were efficient for this process. Using the total signal, number of pixels hit on the diagonal, and a

large normalized energy deposition per unit length all provided useful bounds to sort out the possible “fat” tracks. This allowed for easier testing of the possibility that these tracks were created by protons. The selection resulted in 26 of the original 406 tracks, which is 6% of the data collected from the PFS at 4 km. If these tracks were created by protons, this provides data for the number of protons at sea level that had not been previously observed.

The “fat” tracks that are left in the CCDs correlate to low momentum particles, which supports the hypothesis that they are created by protons. This agrees with the expected values of  $\beta$  for protons and muons from their momentum distributions. However, there were no selections to prove definitively that the particles were protons.

#### **IV. DISCUSSION**

The track fitting and selection observables can be improved to find a way to clearly select for protons or muons. While it was not possible to find the best way to select specifically for protons during this ten-week project, more time would give an opportunity to improve the selection process.

In addition to improving the selection, the presence of delta rays in the CCD sensors could provide errors in the data. Delta rays are slow moving electrons with low penetrative power that are ejected from atoms by the particles passing through the sensor<sup>18</sup>. However, they can create extra signal in the CCDs that can cause the code to mistake lower energy tracks with delta rays for higher energy “fat” tracks.

<sup>18</sup> S. P. Swordy, D. Muller, & A. T. Have, On the generation of delta-rays in detectors for high-energy cosmic-ray nuclei. International Cosmic Ray Conference, 8(55) (1983).

With a moderately successful sort, the normalized energy per unit length was expected to be  $80 e^-/\text{micron}$ . The calculations in this project resulted in a peak  $\frac{dE}{dx}$  of  $100 e^-/\text{micron}$ .

This error might be a result of a factor in calculations during the correction for angle. There is also the possibility that “fat” tracks are not left in the CCD by protons, but with muons with lower velocities.

The code used in this project has many future uses both at BNL and with future projects. It can be used to increase the quality of data analysis at the Large Synoptic Survey Telescope (LSST), or Rubin Telescope, which will use similar CCD sensors. Improved algorithms will allow for easier classification and identification of various particles. The code can also continue to be refined at Brookhaven in future projects.

## V. CONCLUSIONS

The cosmic rays in astronomical CCDs project was able to determine that protons are a candidate for the particle that creates “fat” tracks. These tracks deposit more charge with a larger width than would be expected for other particles, such as muons. Using Jupyter allowed for the sorting of the tracks to find the most interesting. This method of analysis was unable to definitively find ways to select for “fat” tracks in data set and could not determine the best method of determining if these tracks are created by slow-moving muons or protons. In the future, this research could be continued by examining more ways to select for protons in the data sets. Improving the code and optimizing functions would also be good next steps in order to improve functionality for the LSST. This research will help to guide discoveries and characterize particles that come from cosmic rays.

## VI. REFERENCES

<sup>1</sup>D. B. Chitwood, T. I. Banks, M. J. Barnes, S. Battu, R. M. Carey, S. Cheekatmalla, ... & MuLan Collaboration, Improved measurement of the positive-muon lifetime and determination of the Fermi constant, *Physical review letters*, 99(3), 032001 (2007).

<sup>2</sup>M. Fisher-Levine & A. Nomerotski, Characterising CCDs with cosmic rays. *Journal of Instrumentation*, 10(08), C08006 (2015).

<sup>4</sup>R. Lupton (2021). PFS cosmics tracks.

<sup>5</sup>H. Sugai, H. Karoji, N. Takato, N. Tamura, A. Shimono, Y. Ohyama, ... & C. Yan. H, Prime focus spectrograph: Subaru's future. *Ground-based and Airborne Instrumentation for Astronomy IV*, 8446(84460Y) (2012).

<sup>6</sup>J. Beringer, Particle data group. *Phys. Rev. D*, 86(010001) (2012).

<sup>11</sup>A. A. Correa, Calculating electronic stopping power in materials from first principles. *Computational Materials Science*, 150(291) (2018).

<sup>17</sup>S. P. Swordy, D. Muller, & A. T. Have, On the generation of delta-rays in detectors for high-energy cosmic-ray nuclei. *International Cosmic Ray Conference*, 8(55) (1983).



## VII. ACKNOWLEDGEMENTS

I would like to thank the Science Undergraduate Laboratory Internships (SULI). This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

I would also like to thank Andrei Nomerotski, for mentoring me through this project, Robert Lupton, for identifying the tracks and creating the cutouts used, and HyeYun Park, for giving advice and help with code.

## VIII. APPENDIX

### *Participants:*

<b>Name</b>	<b>Institution</b>	<b>Role</b>
Dr. Andrei Nomerotski	Brookhaven National Laboratory	Mentor and lead researcher
Robert Lupton	Princeton University and LSST	Data collection and processing
HyeYun Park	Brookhaven National Laboratory	Post-doc secondary mentor
Jocelyn McMahan	Susquehanna University	SULI intern

### *Scientific Facilities:*

- Prime Focus Spectrograph, used by Robert Lupton to provide the project with track data

### *Notable Outcomes:*

- Presentation at SAWG meeting. Virtual. Jocelyn McMahan and Andrei Nomerotski.
- Internal presentation at the Cosmology and Astrophysics Group department meeting. Virtual. Jocelyn McMahan and Andrei Nomerotski.
- Report “Cosmic rays in astronomical charge-coupled devices”. Jocelyn McMahan and Andrei Nomerotski.
- Poster “Cosmic rays in astronomical charge-coupled devices”. Jocelyn McMahan and Andrei Nomerotski.

## **Analysis and simulation of polarization control in quantum networks**

Liam McSweeney<sup>1</sup>, Denis Dolzhenko<sup>2</sup>, and Paul Stankus<sup>3</sup>

<sup>1</sup>The University of Edinburgh, 33 Buccleuch St, Edinburgh, Scotland, EH8 9JS

<sup>2</sup>Purdue University, 610 Purdue Mall, West Lafayette, United States, IN 47907

<sup>3</sup>Quantum Information Sciences, Brookhaven National Laboratory, Upton, NY, United States 11973

## **Abstract**

Polarization control is a fundamental step in the development of quantum cryptography based on polarization entanglement. Without it, information is lost in the transmission of photons and the process of quantum key distribution is impossible to carry out. We investigate two approaches to polarization control. The first is a hardware approach wherein photons are passed through a 3-paddle fiber-twister which is actively tuned to reverse any changes to photon polarizations incurred as they pass through quantum network components. Preliminary results are promising, showing that polarization can be recovered to approximately  $5^\circ$  on the Poincaré sphere. This corresponds to a 98% measurement certainty in practice. The remaining 2% error rate is easily corrected by established methods in Information Theory. The second method explores the possibility of an adaptive measurement scheme, in which members at either end of the quantum network switch between measurement bases to increase the correlation between their measured polarization states regardless of transformations due to the quantum channel. The method suggests a worst-case scenario error rate of approximately 15%. A combination of both methods may also be viable. Further tests are in progress and the full extent of the success of these methods is not yet established.

## Introduction

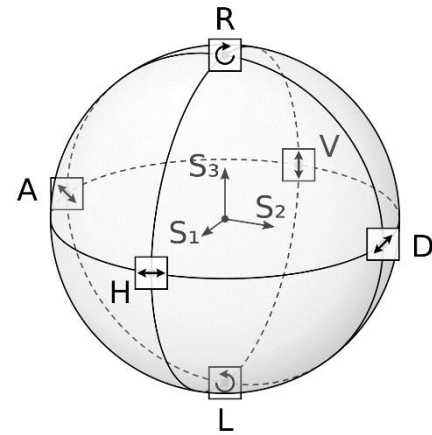
The publication of Shor's algorithm in 1994 <sup>[1]</sup> meant that the advent of large quantum computers would break the world's current cryptography system. The development of quantum cryptography offers an uncrackable alternative to the modern approach, however, a few hurdles stand in the way. Quantum Cryptography encodes data in the polarization of photons and makes use of quantum mechanics to share this data securely. However, quantum mechanical systems are inherently sensitive, and the polarization of transmitted photons will change as the photons travel through the quantum network, leading to the loss of information. Polarization control is a fundamental step in the proper implementation of quantum cryptography on large scales at large distances, as it aims to revert photons back to their initial polarization states, thus retrieving the lost information. In this report we introduce and explore two possible approaches to polarization control. The first is a hardware correction which retrieves the original polarization by sending incoming photons through a calibrated 3-paddle fiber twister, a standard device for inducing polarization transformations in optical fiber. The second is the simulation of an alternative measurement procedure, which makes use of statistical features of the system to create a high degree of correlation between the measured and transmitted polarization states even in the presence of uncontrolled polarization changes. In this way, the input state can be deduced from the measured state, and the intended information retrieved. The hardware implementation is showing promising results when used in real-time in the laboratory. Results for the simulation are still pending at this time.

## Methods

Two methods of polarization control are explored; the first, involves the use of a 3-paddle fiber-twister to revert incoming photon polarizations back to the initial states; the second, involves the simulation of a different measurement scheme to maintain a high degree of correlation between observed and states at either end of the quantum channel.

### The hardware correction

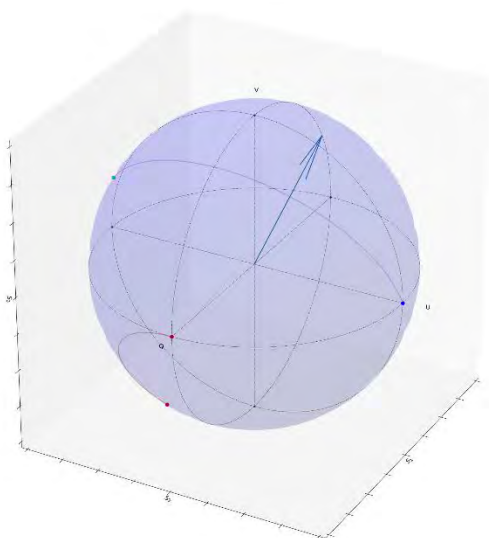
The hardware correction requires the regular mapping of the three paddle settings to a particular rotation of two points on the surface of the Poincare sphere (Fig.1) which describes all the possible polarization states of a photon. The 3-paddle fiber-twister works by twisting fiber-optic cables by some preset angle determined by the paddle settings. The resulting database, linking paddle settings to a rotation, is used to calibrate the 3-paddle fiber-twister so that incoming photon polarizations can be transformed back to their initial states, which are predetermined by regulating the input photon polarizations, so that data can be shared reliably.



*Figure 1 : Poincare Sphere- Polarization states are depicted as points on the surface of the sphere with coordinates determined by the Stokes' vectors:  $s_1$ ,  $s_2$ ,  $s_3$ . Points on opposite sides of the sphere are orthogonal and can be used to describe the full state space.*

To build up a database of transformations, a laser is shone into a circular polarizer and then into a linear polarizer set at an angle of  $0^\circ$  for the Horizontal polarization state ( $|H\rangle$ ) and an angle of  $45^\circ$  for the Diagonal polarization state ( $|D\rangle$ ) at the transmission source. These define the initial polarization states. Both these states are required for the paddle settings to be mapped to a unique rotation on the Poincare sphere. The photons are then passed through the 3-paddle fiber-twister and into a polarimeter which will measure the three Stokes' vectors (Fig. 1) that describe the point on the

surface of the Poincare sphere which defines the photon's polarization. Each measurement is repeated four times for each angle of the linear polarizer and each paddle setting from  $0^\circ$  to  $170^\circ$  with  $10^\circ$  increments. The initial and final positions of the  $|H\rangle$  and  $|D\rangle$  states can be used to describe the rotation carried out by the fiber-twister at each paddle setting (Fig.2). The transformation on the Poincare sphere is a solid-body rotation, which can be written as a clockwise turn through some angle about a particular axis. This is shown in Figure 2 where the basis states on the equator are mapped onto their counterparts somewhere else on the Poincare sphere.



*Figure 2: The effect of the 3-paddle fiber-twister on polarization states. The fiber-twister can be described as rotating a point clockwise about some axis by some angle.*

Once established this database is used in the regular calibration of the measurement apparatus so that the fiber-twister can reverse any changes to the polarization incurred as the photons travel through the quantum network. To test the effectiveness of our database in describing any possible transformation on the photon polarizations, it was regularly searched for a close fit to 10,000 random transformations of the basis states. The results from this showed that the database does cover the full state space

and real-time testing could move forward.

Applying this method in real-time required that a new transformation database be created. Once this was done, the method was tested by scanning the database for a transformation that maps the incoming photon polarizations, to some arbitrary points on the Poincare sphere. With this test being successfully completed, the full calibration and correction scheme was implemented in real-

time. The transformation database was collected once more and the fiber-optic cables leading to the fiber-twister were either replaced or twisted such that the effect of the channel on the polarization of the incoming photons was changed. A series of  $|H\rangle$  and  $|D\rangle$  polarized photons were input into the channel and their final polarization states were recorded. The rotation needed to revert these final states back to their respective initial states was calculated and the database was searched for the three paddle settings that best approximated this rotation. The three-paddle fiber-twister was reset using these paddle settings and more  $|H\rangle$  and  $|D\rangle$  polarized photons were sent through the channel. These states were recovered to within  $5^\circ$  of their intended states on the Poincare sphere, suggesting that the hardware approach can be considered a successful method of polarization control. These results are shown in Figures 5 and 6.

### **The simulation of an alternative measurement scheme.**

The potential of a different measurement scheme in maintaining a high degree of correlation between measured states was explored as a possible alternative to recovering lost information without the use of the hardware correction. The principle is illustrated in Figure 3 through the process of quantum key distribution. An entangled pair source (EPS) produces two photons in the  $|\Psi^-\rangle$  Bell state which are distributed to people who want to communicate securely, Alice and Bob, both of whom make a measurement on the polarization of their photons. In a perfect world, if Alice were to measure the  $|H\rangle$  state, she would immediately know that Bob measured the  $|V\rangle$  state. This is the principle of quantum entanglement. If they don't agree on this measurement, it means that someone has intercepted the photon somewhere along the channel and is eavesdropping on their conversation. Alice and Bob would know that their communication channel is compromised, and can take appropriate measures. This scheme requires that Alice and Bob's measurements are

highly, if not completely, correlated. However, as we know, the photon polarizations change as they travel through the quantum channel and the measured states will differ.

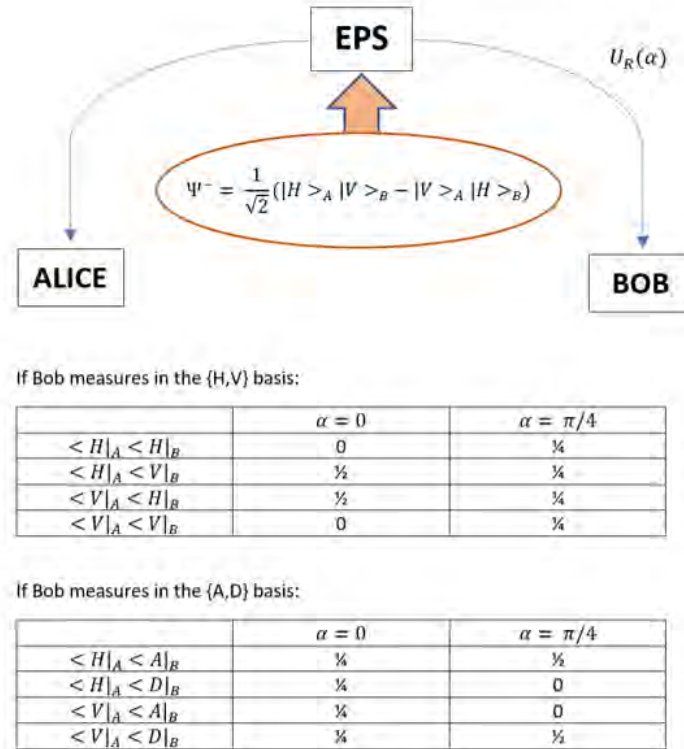


Figure 3: The use of a different measurement scheme may maintain a high degree of correlation between measured states. The polarization is only changed along Bob’s channel by some angle,  $\alpha$ , around the equator of the Poincare sphere.

Figure 3 considers the case when the polarization is changed only along Bob’s channel and the transformation is a rotation by some angle,  $\alpha$ , around the equator of the Poincare sphere. The mathematics are not shown here, but the table shows that a perfect correlation between Alice and Bob’s results can be recovered in the  $\alpha = \pi/4$  case if Bob changes his measurement basis from the {H,V} to the {D,A} basis. Extending this result to any kind of rotation of the polarization (not just around the equator) applied to both Alice and Bob’s channels is the next step in exploring how well this method performs at regaining a perfect correlation between states. Our approach is to study this



method by randomly rotating the photon polarizations along Alice and Bob's channel and testing how effective each measurement basis is in recovering a correlation between states. Based on this information, we can develop an optimal measurement procedure and determine whether this has the potential of limiting error rates sufficiently to be implementable in a real system.

## Results

The hardware correction to the polarization of incoming photons is successful when tested in the laboratory. Testing the database on 10,000 random transformations finds a corresponding transformation in the database that produces an average of over 98% measurement certainty. These data are shown in the histograms in Figure 4 where the final state due to the original rotation of the basis states is compared with that from the rotation identified in the database by measuring the overlap of the two states. The square of this overlap gives the degree of certainty in the measurement.

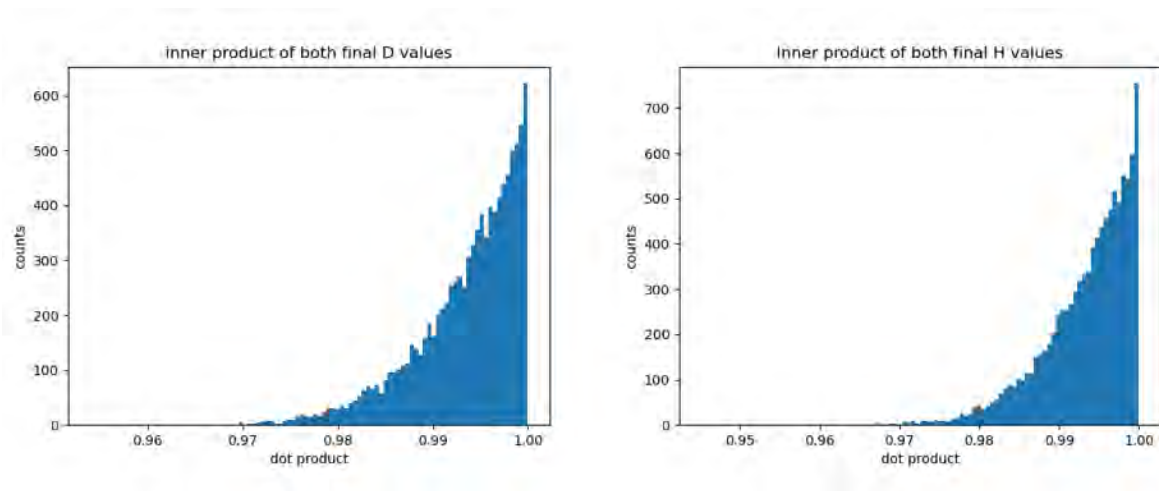


Figure 4: The Inner Product between final states: D (left), H (right) taken from the database with  $10^\circ$  increments of the paddle settings

These results suggest that the database of transformations can successfully map the basis states to any point on the Poincare sphere and vice versa. We considered if sampling the paddle settings over  $10^\circ$  increments creates a database that is too detailed for our purposes, and that we may be able to save time by creating a database over  $20^\circ$  increments, which would be accurate enough for our purposes while requiring one eighth the number of settings. The same test over 10,000 random inputs was repeated using the same database filtered by only using paddle settings at  $20^\circ$  increments and the resulting overlaps are shown below.

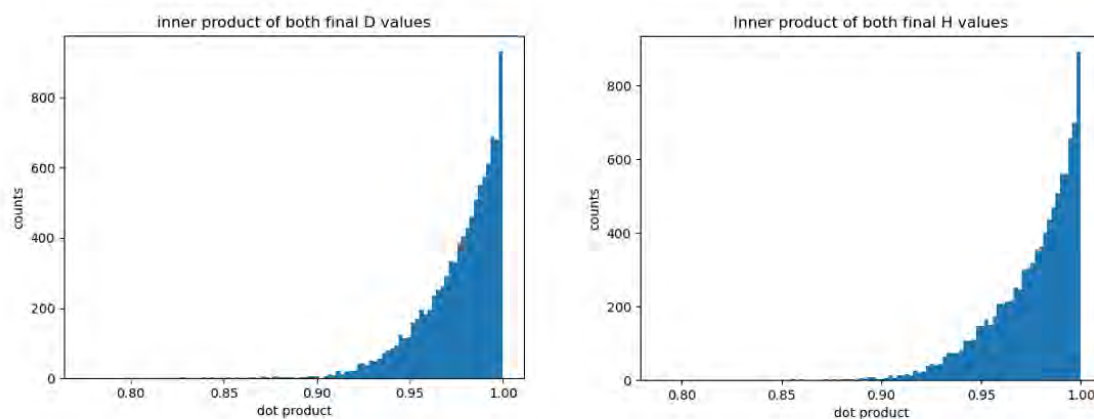


Figure 5: The Inner Product between final states: D (left), H (right) taken from the database with  $20^\circ$  increments of the paddle settings

Even with a significantly smaller database, we can average an overlap of over 95% corresponding to a 90% measurement certainty, making the correction scheme successful even with the smaller database.

The real-time testing of this scheme serves to check that no errors were made in the process of representing the needed rotations both to create the database and to identify the target rotation in the calibration process. Two tests were carried out. The first test involved finding the rotation in the database that would map arbitrary starting points, as determined by the test setup, to some select points on the Poincare sphere. This showed some success when applied to real photons passing

through the setup. With this test done, we tested the real-time calibration procedure on the test setup. We were successful in showing that the hardware correction to polarization control can be used to negate any transformations incurred on the polarization basis states. Images of the benchtop setup, including the setup used to build up the database of transformations are shown in Appendix A. The effect of the channel on the incoming polarization was changed by twisting the fiber optic cable between the motorized rotator and the 3-paddle fiber-twister. The first test setup is shown in Image 2 of Appendix A. The polarization state of the incoming photons was measured, and the fiber-twister was calibrated. Passing photons into the calibrated setup corrected the polarization states to the desired H and D states with reasonable accuracy. The transmission fiber-optic cable was then twisted again, and the process was repeated. Figure 6 shows the results for the second setup. The left image shows the final polarization state as the photons pass through the uncompensated

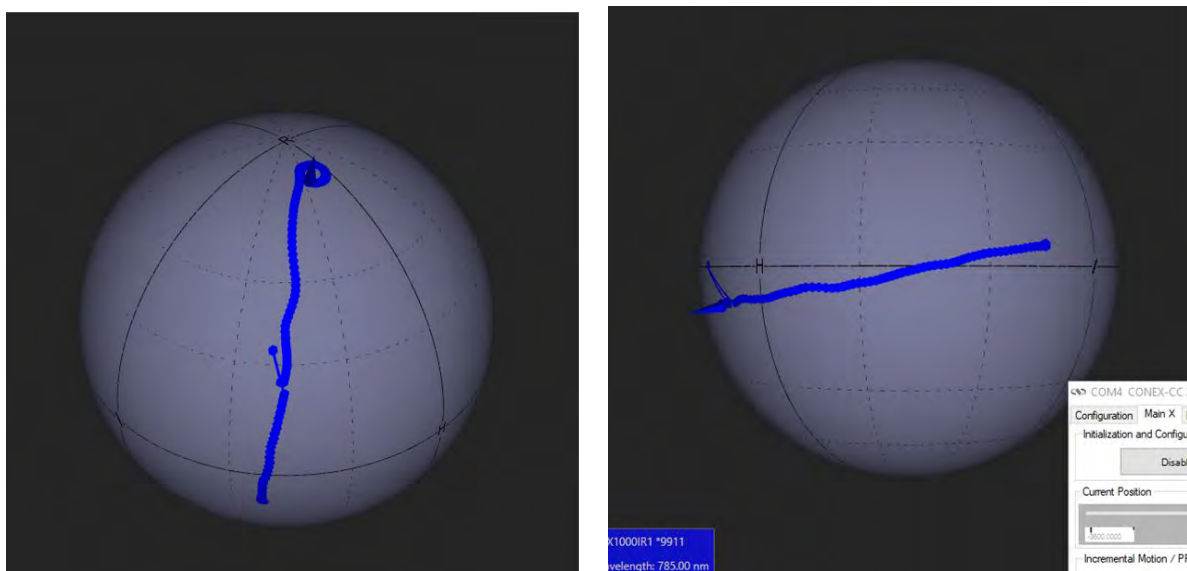


Figure 6: Benchtop results for the hardware correction. (Left) The final polarization states before the calibration of the 3-paddle fiber-twister. (Right) The final polarization states after the calibration of the 3-paddle fiber-twister.

setup. The right image shows the final polarization states after the calibration of the 3-paddle fiber twister. The H and D states aren't recovered completely, however, this degree of accuracy is sufficient to recover the proper H and D states most of the time, allowing for data sharing.

Analysis of the alternative measurement procedure is still in its early stages. However, defining a measure of the correlation between results gives us some ideas as to how well this method works in the simplified case. The correlation was defined as,

$$C = \left| \frac{P(X_A Y_B) - P(X_A X_B)}{P(X_A Y_B) + P(X_A X_B)} \right|$$

where  $P(X_A Y_B)$  is the probability that Alice measures the X state and Bob measures the Y state. In the simplified case described above, Alice only measures in the  $\{H,V\}$  basis and Bob will switch between the  $\{H,V\}$  and the  $\{A,D\}$  bases. The correlation plot is shown below over all possible rotation angles around the equator of the Poincare sphere. Using this scheme, the minimum possible correlation between results is at  $\alpha = (2n+1)*\pi/8$  where  $C = 0.707$ . This result is interpreted as the correct result being  $\sim 5.8$  times more likely to be measured than the incorrect one, or that Bob and Alice's results are correlated 85% of the time. It remains to be quantified whether this correlation is sufficient for data to be efficiently shared across the network as we consider the three-dimensional rotations along both Alice and Bob's channels.

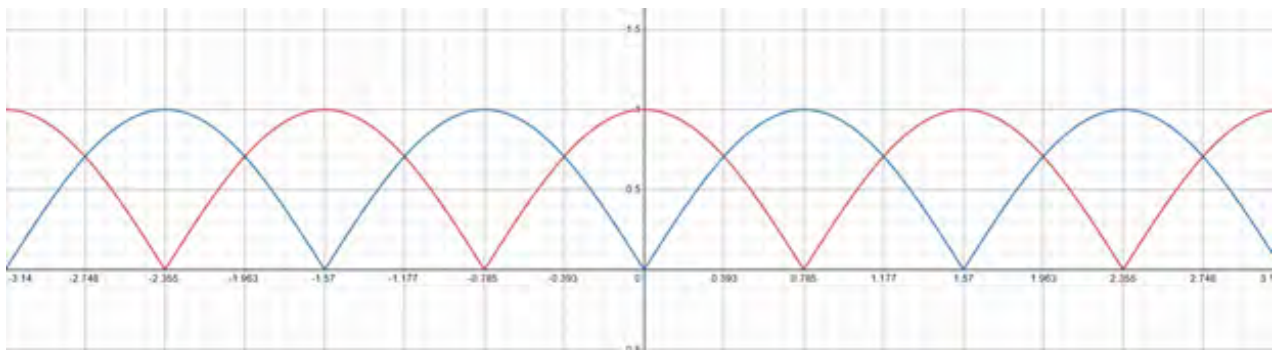


Figure 7: The plot of the correlation between results depending on which basis Bob measures in. The red curve represents the correlation when Bob measures only in the  $\{H,V\}$  basis, the blue curve is for measurements in the  $\{D,A\}$  basis.

## Discussion

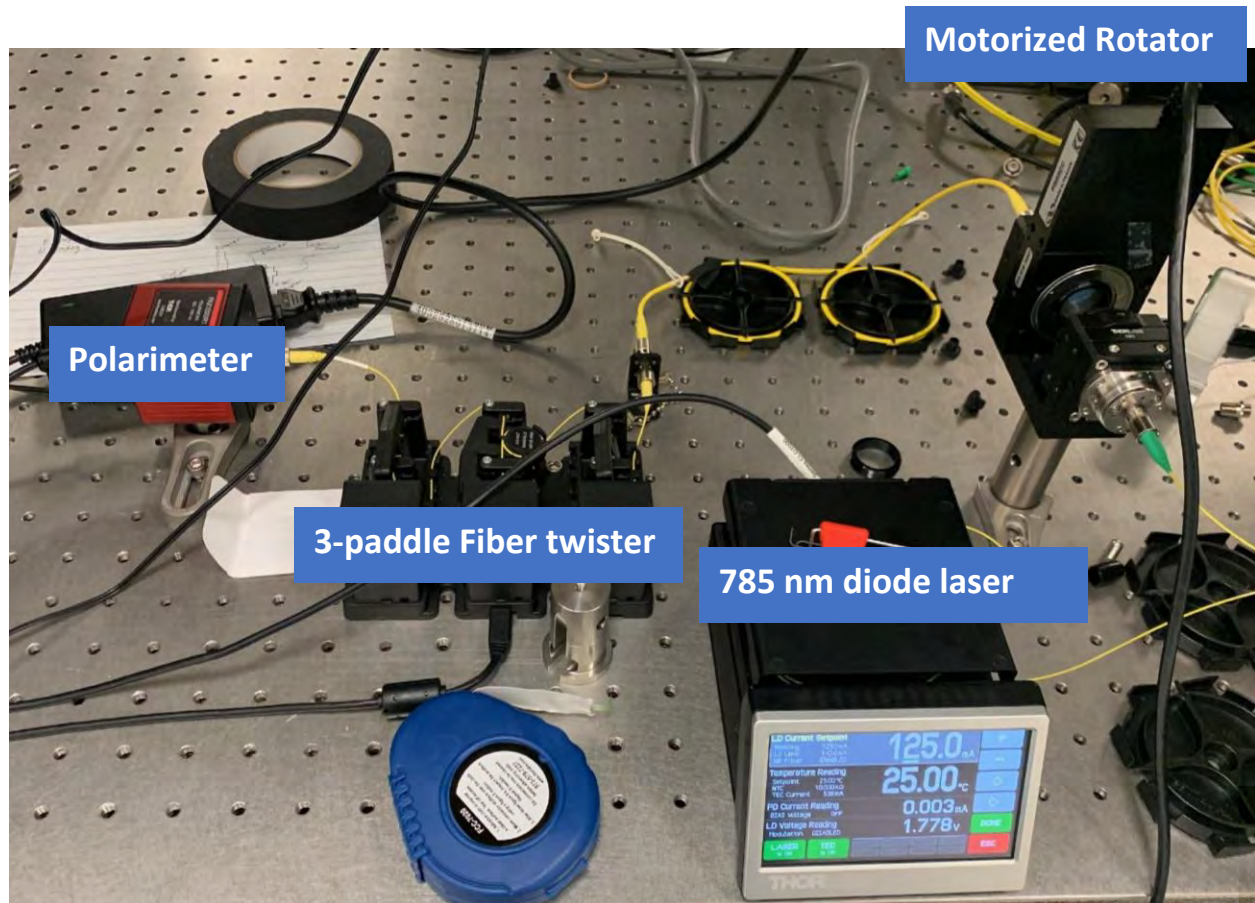
We have demonstrated a practical method for polarization control which works by passing photons through a pre-calibrated fiber-twister. With the benchtop tests of the hardware correction successful, the method will now be implemented into the quantum network linking Brookhaven National Laboratory and Stony Brook University. For the time being, this network is made up of fiber-optic cables but the transmission of photons through free-space is another project currently underway. The hardware correction may then be tested in this system too. The success of this procedure will lead to this research being published in the coming months.

The simulation of the new measurement scheme is still underway and research into this method will extend beyond the end of the internship.

## References

<sup>[1]</sup> P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," *Proceedings 35th Annual Symposium on Foundations of Computer Science*, 1994, pp. 124-134, doi: 10.1109/SFCS.1994.365700.

## Appendix A

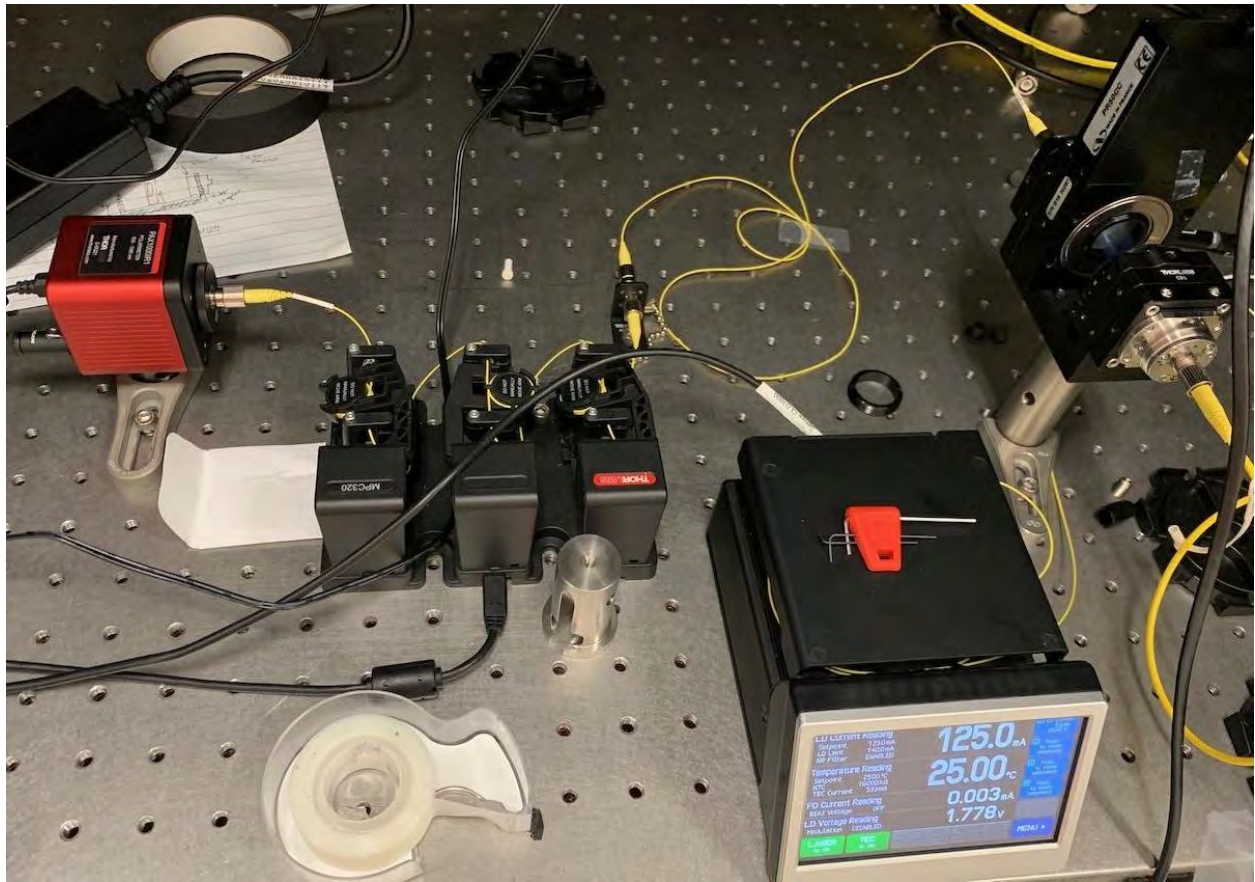


*Image 1: The setup used to build up the database of transformations. Photons are created at the 785 nm diode laser and passed through the fiber-optic cable to the motorized rotator where the  $|H\rangle$  and  $|D\rangle$  states are selected. The photons then travel through the fiber-optic cables to the 3-paddle fiber-twister where their polarization is changed, and then on to the polarimeter where the final polarization is measured.*





*Image 2: The first test setup used to test the correction algorithm. The yellow fiber optic cable was twisted to change the effect of the channel on the incoming polarizations.*



*Image 3: The second setup used to test the correction algorithm. The yellow fiber-optic cable was twisted in a different way.*



# Structural analysis of materials used for energy storage

Marie Françoise C. Millares,<sup>a</sup> Amy C. Marschilok,<sup>\*b, c</sup> Kenneth J. Takeuchi,<sup>\*b</sup> Esther S. Takeuchi<sup>\*b, c</sup>

<sup>a</sup> Department of Nanoscale Science and Engineering, Polytechnic Institute (SUNY), Albany, NY, 12206

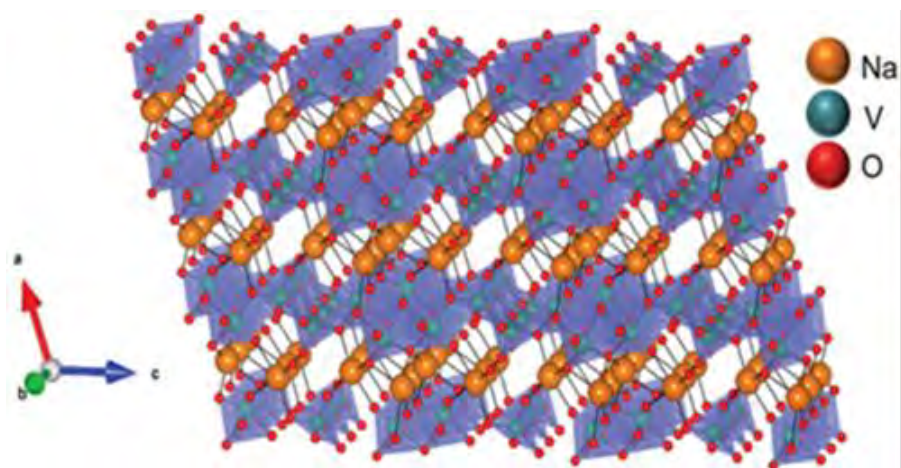
<sup>b</sup> Department of Materials Science and Chemical Engineering and Department of Chemistry, Stony Brook University (SUNY), Stony Brook, NY, 11794

<sup>c</sup> Energy Science Directorate, Brookhaven National Laboratory, Upton, NY, 11973

Microgrids using large-scale energy storage systems are attractive to rural areas in developing nations since they allow electricity to be more accessible at a lower cost. However, when it comes to using intermittent energy sources to power these microgrids, advancements in energy storage systems, such as batteries, need to be made. Even with the rise in market for Li-ion batteries, aqueous Zn-ion batteries are beginning to capture more interest due to their high chemical safety, natural abundance, and low cost. When fabricated with sodium vanadate (NVO) as a cathode material, a unique pillared framework is provided to assist in facilitating Zn-ion migration. In this study, we will be analyzing this crystal framework as a function of temperature and see how it is affected. There were two samples prepared for study: one annealed at 300°C (NVO300) and another at 500°C (NVO500). X-ray powder diffraction (XPD) patterns were collected at Stony Brook University and x-ray absorption spectroscopy (XAS) spectra collected at the National Synchrotron Light Source II (NSLS-II) in Brookhaven National Laboratory (BNL). After the investigation of XPD patterns, it was found that NVO500 contained no water, which was realized to assist in ion transport throughout the electrode. This lack of water depicted a decrease in performance concerning the electrochemistry. Operando XAS spectra was collected for NVO300, which showed a significant jump in energy at the pre-edge, from 5465eV to 5470eV. From the decrease in edge energy, a reduction in vanadate was depicted. These findings are significant since they show how the electrochemistry of the battery can be controlled by changing a material's structural properties through heat addition. As a result of this summer, I can analyze data from XPD and XAS techniques. I am now familiar with several software programs including PeakFit, VESTA, GSAS-II, and Athena.

*Keywords:* energy storage systems, characterization techniques, electrochemistry

## Introduction



**Fig. 1** Crystal structure of sodium vanadium oxide ( $\text{NaV}_3\text{O}_8$ , NVO).

Tang, Christopher & Singh, Gurpreet & Housel, Lisa & Kim, Sung Joo & Quilty, Calvin & Zhu, Yimei & Wang, Lei & Takeuchi, Kenneth & Takeuchi, Esther & Marschilok, Amy. (2021). Impact of Sodium Vanadium Oxide ( $\text{NaV}_3\text{O}_8$ , NVO) Material Synthesis Conditions on Charge Storage Mechanism in Zn-ion Aqueous Batteries. *Physical Chemistry Chemical Physics*. 23. 10.1039/D1CP00516B.

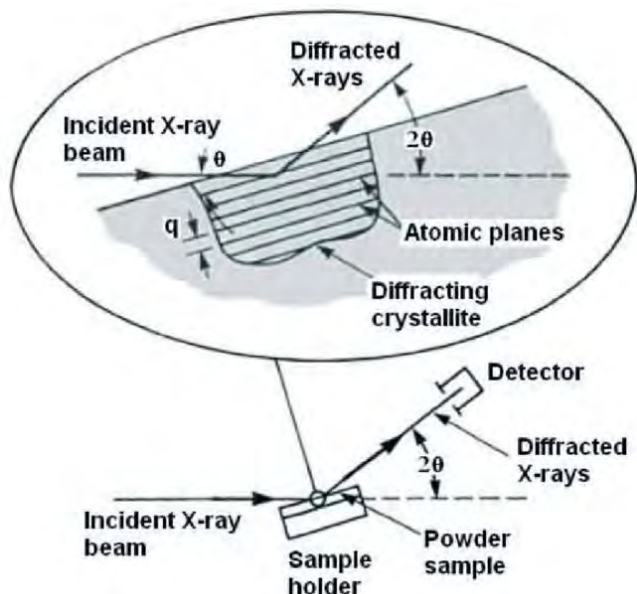
13% of the world does not have access to electricity and many more with access find it unreliable or too expensive.[1] This can pose a barrier to economic development in emerging economies. A possible solution to this is the construction of microgrids in these locations, which are self-contained electric grids that can operate independent of the central power grid. Additionally, microgrids allow for the implementation of sustainable energy production, such as wind and solar power, making this a cheap and efficient source of electricity.[2] To account for the instability in power generation from these intermittent power sources, the use of energy storage systems play an important role.

Zn-ion aqueous based batteries are an attractive alternative for Li-ion batteries for these applications because of its high theoretical capacity (820 mAh/g), low redox potential (-0.76 V vs. SHE), and low toxicity.[3,4] Layered metal vanadate systems ( $\text{M}_x\text{V}_y\text{O}_z$  or MVO, M = metal ion) have been explored as a cathode material because of their architecture, which helps promote ion transfer, and their nanostructured morphology, which is considered to enhance rate capability due to the shortened diffusion length for the Zn ion.[5,6] From these systems, sodium vanadium oxides (NVO) show the most promise with their sheet-like structure (Fig. 1). [7]

To better understand the reaction mechanisms occurring within this material, characterization techniques such as X-ray powder diffraction (XPD) and X-ray absorption spectroscopy (XAS), will be focused on for this study.

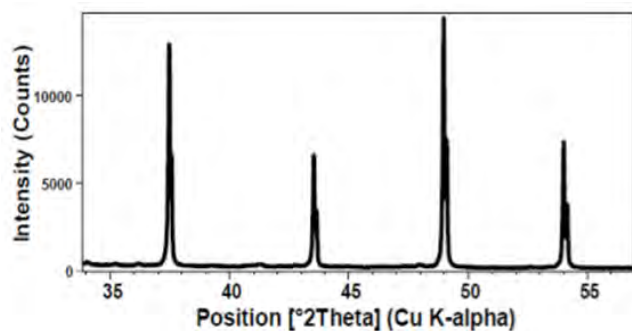
### 1.1 X-ray powder diffraction (XPD)

XPD is a powerful nondestructive characterization technique commonly used to provide information on crystal structure, phases, preferred crystal orientation, grain size, and other structural parameters. This technique is based on the production of constructive interference from diffracted x-rays between the incident x-ray beam and crystalline



**Fig. 2** Schematic of X-ray interacting with lattice planes.

Billah, Areef. (2016). Investigation of multiferroic and photocatalytic properties of Li doped BiFeO<sub>3</sub> nanoparticles prepared by ultrasonication. 10.13140/RG.2.2.23988.76166.



**Fig. 3** Example of XRD pattern.

Bunaciu, Andrei A. & Udriștioiu, Elena & Aboul-Enein, Hassan. (2015). X-Ray Diffraction: Instrumentation and Applications. Critical reviews in analytical chemistry / CRC. 45. 10.1080/10408347.2014.949616.

sample (Fig. 2).[8] Peaks are produced when the detector identifies constructive interference (Fig. 3), and the conditions satisfy Bragg's Law:

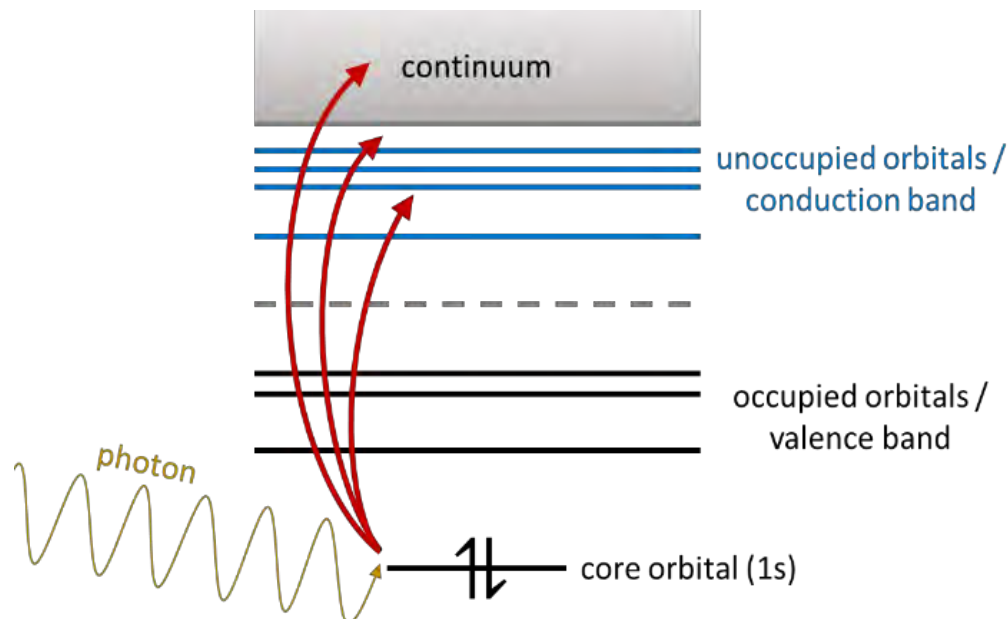
$$n\lambda = 2d\sin\theta \quad (1)$$

where  $n$  is an integer,  $\lambda$  is the wavelength of the X-rays,  $d$  is the interplanar spacing, and  $\theta$  is the diffraction angle.[9] Bragg's Law is important in XPD because it connects the wavelength of electromagnetic radiation to the diffraction and the lattice spacing in the crystalline sample. By converting the diffraction peaks to  $d$ -spacing, compounds can be identified, since each compound has a set of unique  $d$ -spacings. When scanning a sample, a large range of  $2\theta$  values is used to receive all possible diffraction directions of the lattice due to the random orientation of the powdered material. Rietveld refinement can be done on collected pattern. Refinement considers structural parameters, such as micro-strain, unit cell size, and atomic positions, which can generate a 3-dimensional model of the crystal structure. Additionally, sample composition can be determined via phase IDs.

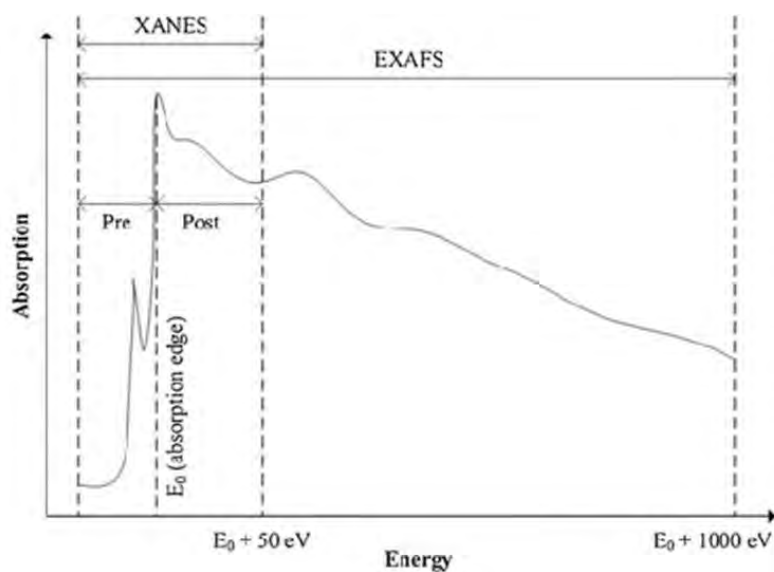
## 1.2 Operando X-ray absorption spectroscopy (XAS)

XAS is a high energy, element specific characterization technique that can give information oxidation states and the geometry of atoms. Essentially, a photon is accelerated towards a core electron and excites it to higher orbitals or to the continuum (Fig. 4).

There are two main regions on a XAS spectra: x-ray absorption near-edge structure (XANES) and extended x-ray absorption fine structure (EXAFS) (Fig. 5). Each region depicts energy levels that the electron can excite to. For example, if an electron were excited to the level of an unoccupied orbital, that only requires a low amount of energy, which would be depicted in the pre-edge XANES region. However, if the electron were



**Fig. 4** Schematic of X-ray interacting with core electrons.



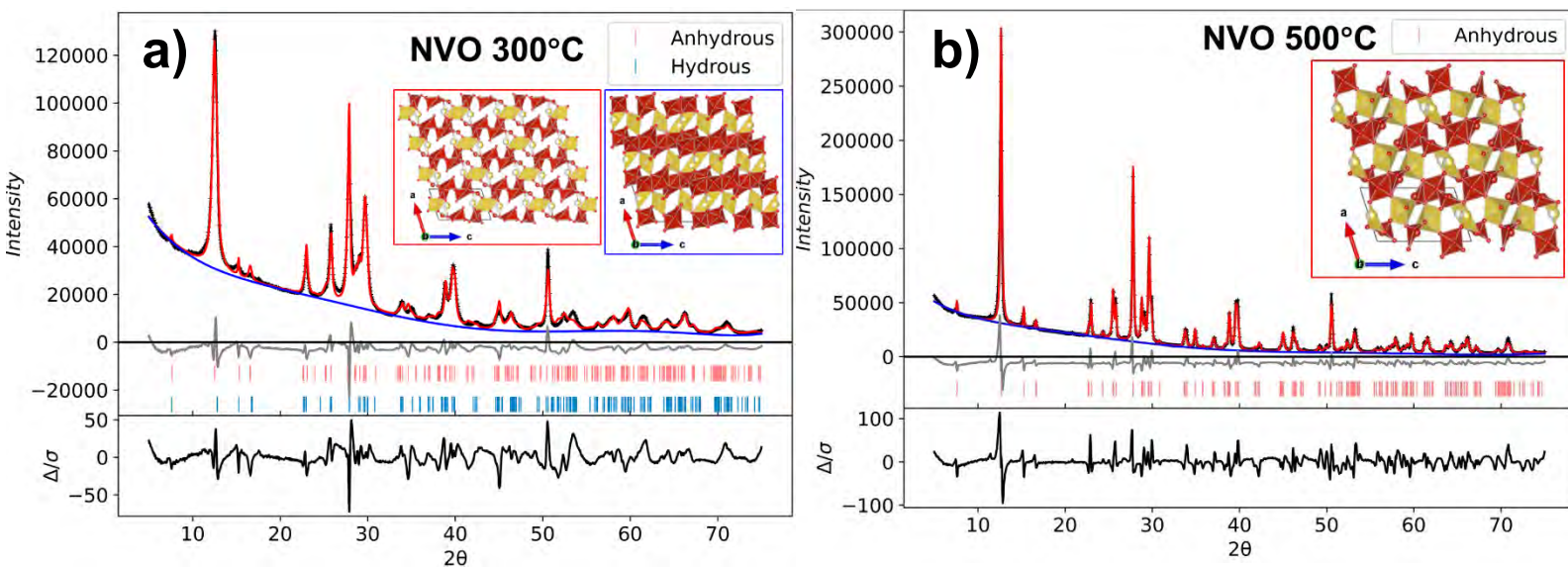
**Fig. 5** Example of XAS spectra showing the two main regions: XANES and EXAFS.

Chae, Sejung & Moon, Juhyuk & Yoon, Seyoon & Levitz, Pierre & Winarski, Robert & Monteiro, P.. (2013). Advanced Nanoscale Characterization of Cement Based Materials Using X-Ray Synchrotron Radiation: A Review. International Journal of Concrete Structures and Materials. 7. 10.1007/s40069-013-0036-1.

excited to the continuum, that requires a lot of energy, which could be seen in the EXAFS region.[11] In this paper, we will mostly be referring to the XANES region, focusing on shifts within the pre-edge and edge energy. Pre-edge shifts correlate to structural changes of the atoms, while edge energy shifts correlate to changes in oxidation states. Operando is a technique where the analysis being done, in this case XAS, is happening while the battery is cycling. This allows for the collection of data in real time during operation. It is a powerful method since it can reveal reaction mechanisms that would not have been apparent otherwise.



## Experimental



The samples used in this study were pre-prepared and then were annealed in air for 3 hours at 300°C for NVO(300) and 500°C for NVO(500). XPD patterns on pristine

**Fig. 6** Rietveld refinement of XPD patterns for (a) NVO300 and (b) NVO500 with accompanying modelled crystal structures.

samples were acquired on a Rigaku Smartlab X-ray diffractometer using a  $K\alpha$  X-ray source. Rietveld refinement was done using the GSAS-II software package. Samples were refined using  $\text{Na}_x\text{V}_3\text{O}_8$  anhydrous and hydrous phases. Operando XAS measurements at the V K edge (5465eV) were taken at the quick x-ray absorption and scattering (QAS) 7-BM beamline in NSLS-II at BNL. All samples were measured with a V metal foil reference simultaneously for correct energy alignment. Spectra for the following standards were also collected:  $\text{V}_2\text{O}_5$ ,  $\text{VO}_2$ , and  $\text{V}_2\text{O}_3$ . XAS spectra was then calibrated using Athena from the Demeter software package.

## Results and Discussion

### 1.1 Rietveld Refinement

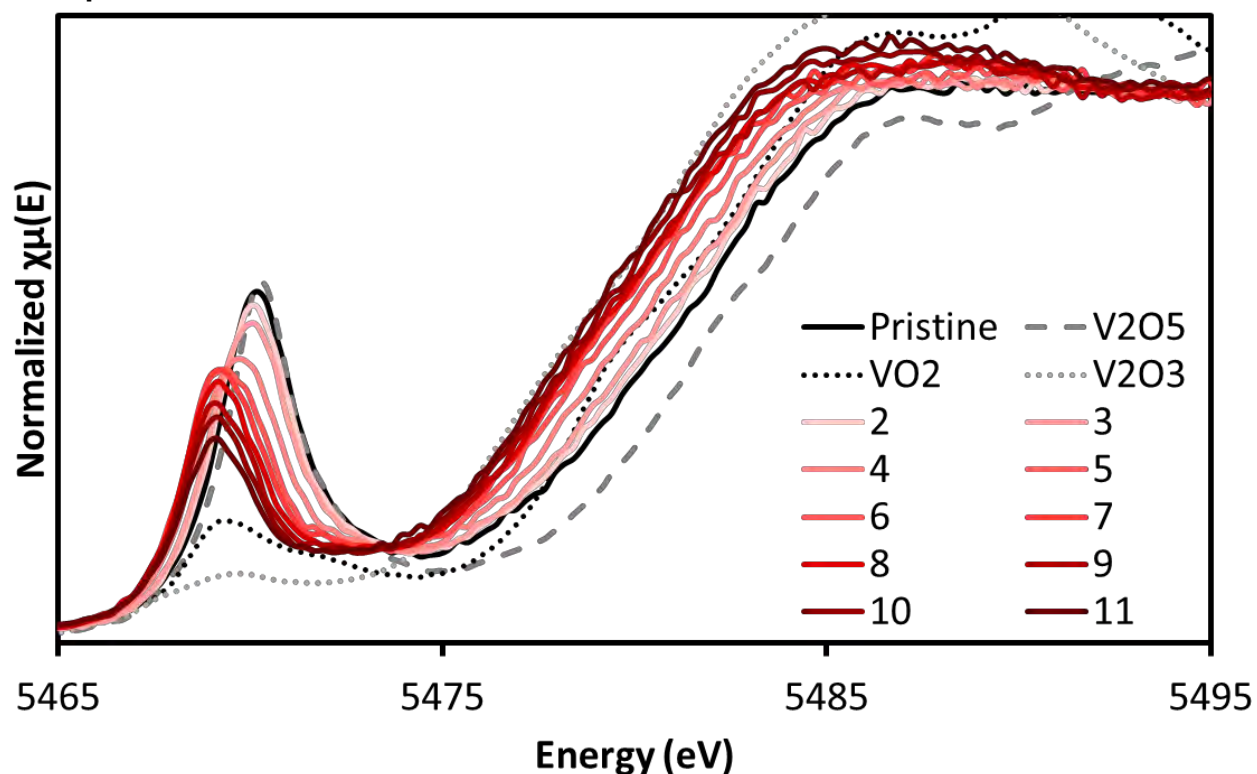
X-ray powder diffraction patterns of NVO300 (Fig. 6a) and NVO500 (Fig. 6b) in the pristine state were collected and fitted using Rietveld refinement. NVO300 was fitted using both anhydrous and hydrous references, while NVO500 only required the anhydrous phase.[12] The refined lattice parameters, crystallite size, micro strain, and unit cell are summarized in Table 1. The NVO300 powder consisted of 60% hydrous phase and 40% anhydrous phase. This shows that NVO300 has higher water content in the interlayer spacing. The smaller crystallite size reported in the NVO300 is

**Table 1: Summary of refined crystal parameters.**

Parameter	NVO300	NVO500
Phase	hydrous/anhydrous	anhydrous
Phase Comparison	60(1)% / 40(1)%	100%
a (Å)	7.236(9) / 7.416(9)	7.312(2)
b (Å)	3.60041(4) / 3.614(1)	3.6093(2)
c (Å)	12.17(1) / 12.12(2)	12.169(5)
$\alpha$ (°)	90.0° / 90.0°	90.0°
$\beta$ (°)	107.65(3)° / 107.54(4)°	107.749(9)°
$\gamma$ (°)	90.0° / 90.0°	90.0°
V (Å <sup>3</sup> )	302.5(1) / 309.7(2)	305.88(3)
Crystallite Size (axial, nm)	0.11(7) $\mu$ m / 0.04(1) $\mu$ m	0.14(2)
Crystallite Size (eqitorial, nm)	0.132(4) $\mu$ m / 0.059(1) $\mu$ m	0.8(7) $\mu$ m
%Rwp	5.40%	7.01%

analogous to the decreased intensity and broadening of peaks when compared to NVO500. Similar unit cell parameters and modelled crystal structures show that annealing alters crystallite size without significant changes in NVO's crystal structure. Additionally, a study reported how smaller crystallite size corresponds to a larger effective diffusion coefficient and improved delivered capacity at all rates.[7] This could also be attributed with the presence of water assisting in the facilitation of ions. According to the same study, after cycling these electrodes in a cell, the presence of two new phases were identified in both samples by using Rietveld refinement. One phase was determined to be  $Zn_4SO_4(OH)_6 \cdot 5H_2O$  (ZHS). ZHS typically forms due to the pH change of mild acidic electrolyte during discharge and can indicate  $H^+$  insertion mechanism. However, on charge, the ZHS phase is not detected, suggesting revisability of the  $H^+$  insertion. The other phase is zinc vanadium oxide,  $Zn_3(OH)_2V_2O_7 \cdot 2H_2O$  (ZVO), which exhibited partial reversibility. This can be related to interactions between the intercalated Zn ions and  $OH^-$  from water dissolution.[13]

## 1.2 Operando XAS



**Fig. 7** Operando XAS scans collected during the first discharge cycle for NVO300.

Vanadium K-edge operando XAS measurements were collected for the NVO300 sample. The spectra received during the charging cycle was analyzed and compared to  $V_2O_5$ ,  $VO_2$ , and  $V_2O_3$  reference materials to estimate the changes in oxidation state (Fig. 7). In its pristine state, NVO300 features a sharp pre-edge peak that is representative of the asymmetric  $VO_5$  square pyramids. Comparison of the sample's pristine state to post discharge depicts a dramatic decrease in intensity and positioning of the pre-edge. The energy shifted from 5470.22 eV to 5468.92 eV during lithiation, which is associated with a loss of crystalline character and the formation of more regular octahedra. The edge energy for the pristine sample is 5481.87 eV, which is most consistent with the  $V_2O_5$  reference, indicating that this material originates as  $V^{5+}$ . As the electrode discharges, the edge position steadily decreases and reaches 5478.44 eV. The spectra is now closer to the  $V_2O_3$  reference, illustrating a shift in oxidation state to  $V^{3+}$ . This is close to the average vanadium oxidation state of 3.2. The decrease in edge energy occurs because as the core orbital is being reduced, the distance between the core orbital and the nucleus increases, requiring less energy to excite these electrons. According to literature, charging the electrode showed that the edge and pre-edge positions return to their pristine state suggesting a reversible oxidation process.[13]

## Conclusions

In this work, samples were pre-prepared and annealed in air for 3 hours at 300°C for NVO(300) and 500°C for NVO(500). XPD characterization depicted smaller crystallite size and higher water content in the NVO300 sample. This proved to be significant since it was found that water is the main driving force for ion insertion throughout the electrode. Through literature search, it was found that other phases started forming as the electrode was cycled. Operando XAS demonstrated that this material has reversible properties, which is important for batteries. Also, it highlighted structural changes in the sample during cycling. Overall, this work proved that the performance of a material can be controlled by making changes in its structural properties.

## Acknowledgements

The XAS measurements used the NIST Beamline for Material Measurement (BMM) 6-BM beamline at the National Synchrotron Light Source II, at U.S. Department of Energy (DOE), Office of Science, at Brookhaven National Laboratory (BNL). This project was supported in part by the DOE, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). This work was completed in collaboration with other interns, Melanie Brito and Olivia Bennett, and the guidance of graduate students from the Marschilok-Takeuchi Research Group at Stonybrook University. It could not have been done without the mentorship of Dr. Kenneth Takeuchi, Dr. Amy Marschilok, and Dr. Esther Takeuchi.

## References

- [1] Ritchie, H. (2020, November 28). Access to Energy. Our World in Data. [https://ourworldindata.org/energy\\_access#low-income-households-lack-access-to-electricity-and-clean-fuels](https://ourworldindata.org/energy_access#low-income-households-lack-access-to-electricity-and-clean-fuels)
- [2] Microgrids and energy storage (n.d.). Environment America. <https://environmentamerica.org/energy-101/microgrids-energy-storage>
- [3] J. W. Choi and D. Aurbach, *Nat. Rev. Mater.*, 2016, 1, 1–16.
- [4] L. M. Housel, W. Li, C. D. Quilty, M. N. Vila, L. Wang, C. R. Tang, D. C. Bock, Q. Wu, X. Tong, A. R. Head, K. J. Takeuchi, A. C. Marschilok and E. S. Takeuchi, *ACS Appl. Mater. Interfaces*, 2019, 11, 37567–37577.
- [5] Xia, C.; Guo, J.; Li, P.; Zhang, X.; Alshareef, H. N. Highly Stable Aqueous Zinc-Ion Storage Using a Layered Calcium Vanadium Oxide Bronze Cathode. *Angew. Chem., Int. Ed.* 2018, 57, 3943–3948.
- [6] Kundu, D.; Adams, B. D.; Duffort, V.; Vajargah, S. H.; Nazar, L. F. A high-capacity and long-life aqueous rechargeable zinc battery using a metal oxide intercalation cathode. *Nat. Energy* 2016, 1, 16119.
- [7] Tang, Christopher & Singh, Gurpreet & Housel, Lisa & Kim, Sung Joo & Quilty, Calvin & Zhu, Yimei & Wang, Lei & Takeuchi, Kenneth & Takeuchi, Esther & Marschilok, Amy. (2021). Impact of Sodium Vanadium Oxide (NaV<sub>3</sub>O<sub>8</sub>, NVO) Material Synthesis Conditions on Charge Storage Mechanism in Zn-Ion Aqueous Batteries. *Physical Chemistry Chemical Physics*. 23. 10.1039/D1CP00516B.
- [8] Billah, Areef. (2016). Investigation of multiferroic and photocatalytic properties of Li doped BiFeO<sub>3</sub> nanoparticles prepared by ultrasonication. 10.13140/RG.2.2.23988.76166.



- [9]** Bunaciu, Andrei A. & Udriștioiu, Elena & Aboul-Enein, Hassan. (2015). X-Ray Diffraction: Instrumentation and Applications. Critical reviews in analytical chemistry / CRC. 45. 10.1080/10408347.2014.949616.
- [10]** Chae, Sejung & Moon, Juhyuk & Yoon, Seyoon & Levitz, Pierre & Winarski, Robert & Monteiro, P.. (2013). Advanced Nanoscale Characterization of Cement Based Materials Using X-Ray Synchrotron Radiation: A Review. International Journal of Concrete Structures and Materials. 7. 10.1007/s40069-013-0036-1.
- [11]** PREM-CIE2M.(2020, May 10) Fundamentals of X-ray Absorption Spectroscopy [Video file]. YouTube. <https://www.youtube.com/watch?v=hsJpPULcbXU&t=3322s>
- [12]** M. Onoda, J. Phys.: Condens. Matter, 2004, 16, 8957
- [13]** Sung Joo Kim, Christopher R. Tang, Gurpreet Singh, Lisa M. Housel, Shize Yang, Kenneth J. Takeuchi, Amy C. Marschilok, Esther S. Takeuchi, and Yimei Zhu. (2020). New Insights into the Reaction Mechanism of Sodium Vanadate for an Aqueous Zn Ion Battery. Chemistry of Materials. 32 (5), 2053-2060. DOI: 10.1021/acs.chemmater.9b05103

# Study of Pad Geometries in AC-coupled Low Gain Avalanche Diodes Using Numerical Simulation

Patricia Moore<sup>1</sup> and Gabriele Giacomini<sup>2</sup>

<sup>1</sup> *Engineering and Computer Science, Suffolk County Community College, Selden, NY, 11784*

<sup>2</sup> *Instrumentation Division, Brookhaven National Laboratory, Upton, NY 11973*

**Abstract.** High Energy Physics (HEP) and Nuclear Physics (NP) experiments - including the proposed Electron-Ion Collider at Brookhaven National Laboratory (BNL) - require the use of high-precision detector technologies. One such variety, the Low Gain Avalanche Diode (LGAD), is a silicon-based sensor which features internal amplification and very good timing. LGADs have been optimized for the detection of Minimum Ionizing Particles (MIPs). When MIPs traverse the LGAD's thin silicon bulk they generate electron/hole pairs along their track. Electrons drift towards a region characterized by a high electric field where they experience impact ionization. However, this is not true for the edge of the pixel, as electric fields are below the threshold for impact ionization rendering it functionally 'dead'. This issue becomes more prevalent with smaller dimensioned LGADs. As a result, the current LGADs used in HEP experiments such as CMS and ATLAS at CERN suffer from a poor spatial resolution, having a relatively large area of  $1.3 \times 1.3 \text{ mm}^2$ . For this reason, a new generation of LGADs has been developed to have higher spatial resolution: AC-coupled LGADs (AC-LGADs). At this time, little is known about how a signal develops within these devices. For this reason, dozens of simulations were conducted using Silvaco's TCAD software that modeled a MIP traversing an AC-LGAD. The resulting current pulses are modeled as if they were read-out by infinite-bandwidth electronics, requiring them to be processed through a low-pass filter to resemble data that would be obtained through non-simulated means. The data were then plotted for analysis. This process was repeated for multiple pad pitches as well as increments of the thickness of the device's oxide layer as a continuation of work done during the Spring 2021 CCI-Tech internship appointment.

## INTRODUCTION

Low Gain Avalanche Diodes are a type of silicon-based sensors which exhibit very good time resolution. They are built on a very thin silicon bulk material, in the order of 50 microns or below. These qualities make them highly suitable for the fast detection of Minimum Ionizing Particles (MIPs) in high energy physics experiments (HEPs).

As the MIPs traverse the bulk, they generate electron/hole pairs along their path, the quantity of which are approximately 80 pairs per micron of silicon. Electrons drift through high electric field regions and undergo impact ionization. This high electric field region is created by the presence of a p enrichment implantation just below the n implant at the surface of the diode (Fig. 1). The short drift of the multiplied holes that eventually reach the back of the diode creates a signal large enough to be detected by fast front- end read-out electronics.

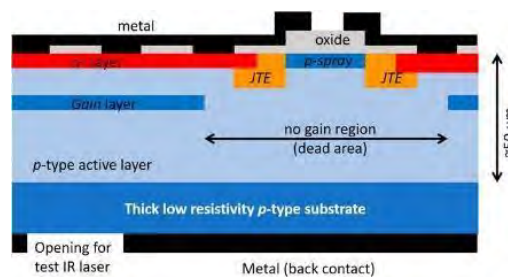
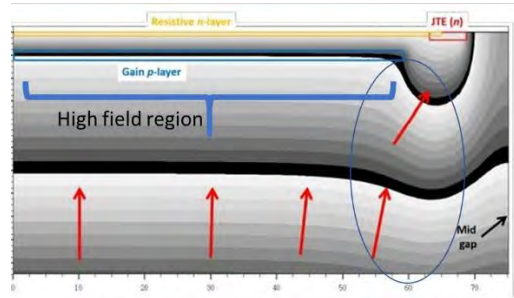


FIGURE 1. Cross-section of an LGAD.

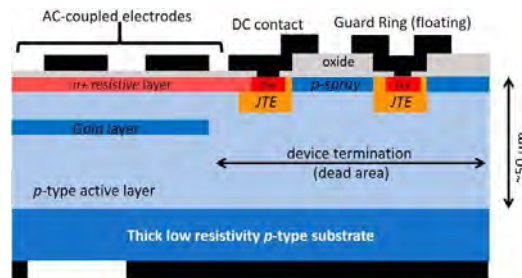
The diode also features a region called the Junction Termination Edge (JTE) at the border of the n implant. The area from JTE to JTE of nearby pixels is considered 'dead'. Electrons created by MIPS that pass through this region

will not experience any impact ionization and as a result the signal in this region will be below the threshold for detection by read-out electronics (Fig. 2). This issue becomes more prevalent with smaller dimensioned LGADs. As a result, the current LGADs used in HEP experiments such as CMS and ATLAS at CERN suffer from a poor spatial resolution, having a relatively large surface area of  $1.3 \times 1.3 \text{ mm}^2$ .



**FIGURE 2.** Electrostatic potential in example simulated LGAD. Red arrows indicate trajectories of electrons. Circled electrons cross low-electric field region and are not amplified.

For this reason, a new generation of LGADs have been developed to have good spatial resolution: AC-LGADs. A main difference is that, in the AC-LGAD, the large uniform n-implant present is about 10 or 100 times less doped than the  $n^+$  implant of a standard LGAD. The n-implant is covered by a thin dielectric layer which is typically silicon oxide, silicon nitride, or both, while remaining grounded through a contact at its border (Fig. 3). A 100% fill factor is achieved, with uniform multiplication and no dead areas. Metal electrodes are patterned over the thin insulator and the signal is AC-coupled to these electrodes, which are connected to the read-out electronics. Since the metal electrodes are separated from the n-layer by an insulator, they do not collect charge and the signal pulse is bipolar (integral of the waveform is zero).



**FIGURE 3.** Signals are capacitively coupled to finely patterned metal –electrodes placed on an insulator above the diode.

At this time, little is known about how the way the signal develops within AC-LGAD devices. To investigate this, multiple numerical simulations were performed using Silvaco’s TCAD suite to simulate a MIP traversing an AC-LGAD. During the Spring 2021 internship appointment this included only one set of simulations – a value for pitch of 100 microns. To expand on this, simulations were performed which doubled that value to 200 microns, and the thickness of the oxide layer was doubled for an additional set of simulations as well. A much smaller set of simulations were also performed without the use of a designated impact ionization model for the purposes of comparison.

## METHODS

### Performing Numerical Simulations with Silvaco's TCAD Suite

Silvaco's TCAD suite of software provided the ability to perform the necessary process and device simulations, allowing the physical and electrical characteristics of AC-LGAD devices to be modeled. This tool was used to create dozens of data sets that modeled a MIP traversing an AC-LGAD with changes made to the geometry of the metal electrodes on the surface of the diode. Previously, a pitch of 100 microns was selected- a relatively standard value. Values for entry point that ranged from 200 to 250 microns were used and were incremented by 5 microns. The chosen values for entry point correspond with the mid-gap and mid-metal locations. For every entry point different values for the gap between two metals were also simulated. Their values ranged from 20 to 80 microns and were incremented by 10microns (Fig. 4).

```
##### Mesh #####  
  
set Pdose = 1e14  
set Bdose = 2.15e12  
set gap = 70 #decrease gap by 10, from 20-80  
set pitch = 100  
set oxthick = .1  
set entrypoint = 210 #increase entrypoint by 5, from 200-250
```

FIGURE 4. Extract of a TCAD simulation.

For this appointment these values were expanded to include a pitch of 200 microns, increasing the overall width of the device to 1000 microns. The MIP entry points used in these simulations ranged from 400 – 500 microns, increasing in increments of 10 microns. These values also correspond with mid-gap and mid-metal locations. Gap values between 20 and 160 microns, incremented by 20, were also simulated for every value of entry point. Additionally, a set of simulations were performed where the thickness of the oxide layer was doubled to .2 microns. This was done for every value of entry point used for the pitch 200 simulations, but only every other value of gap. One final set of simulations were performed which removed a parameter from the deck which defined the impact ionization model used during the simulation process (Fig. 5). This was done for one value of gap and using entry point values which corresponded to a mid-gap location, a mid-metal location, and a point in between. For the pitch 200 data this meant values of 400 microns, 450 microns and 500 microns were selected.

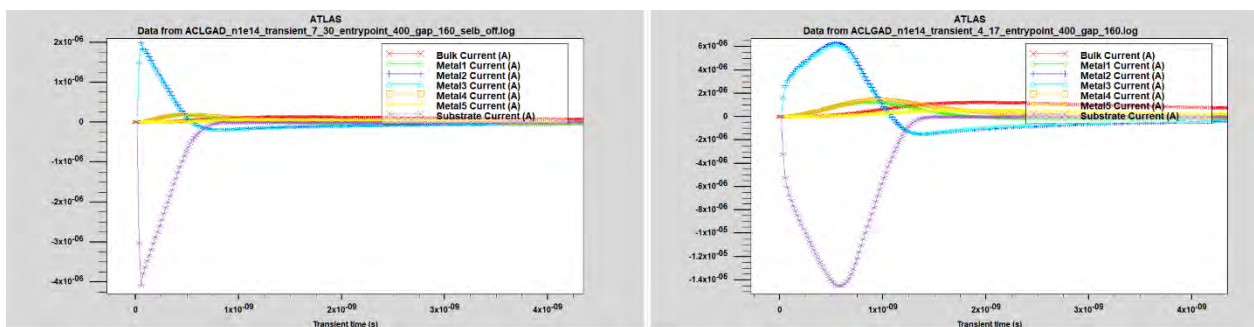
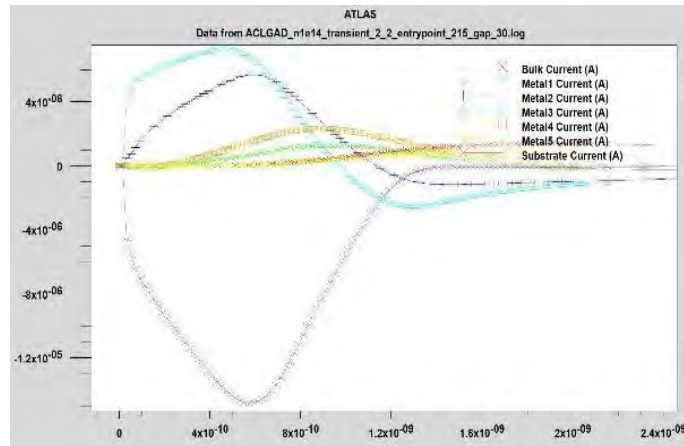


FIGURE 5. Comparison of current pulses with impact self statement off vs. on

Every individual simulation, once completed, generated sampled data that illustrated a current pulse at all the electrodes of the AC-LGAD, as the MIP traversed through it (Fig. 6). This was simulated as if they were read-out by infinite-bandwidth electronics, a case that would never exist in a typical lab setting. This meant that it would be necessary to perform digital signal processing (DSP) to allow the simulated data to reflect more accurately what would be collected through non- simulated means.



**FIGURE 6.** Current pulses at the electrodes.

## Design and Implementation of MATLAB Filtering Script

A custom MATLAB script was designed and written to perform the necessary DSP per the following design considerations.

### *Data Handling and Structure*

The files generated by the Silvaco simulations were text documents which contained columns of data for every parameter simulated. This was truncated to only include the current readings in the metals, bulk and substrate using Silvaco's visualization tool, TonyPlot, and exported as a .csv file for every simulation performed. The script would have to read in the data from every file as it ran, making it necessary to figure out a way to automate the process if the script was going to filter all of the data within one execution. A naming scheme was then employed for the exported files which followed the structure "entrypoint\_\_gap\_\_csv", allowing the values of gap and entry point to be represented in the script as variables. It was desirable to be able to use the same script to import and process additional sets of simulated data without rewriting the script entirely.

The values of gap and entry point used will always be known at runtime. This made it possible to design a function that would take advantage of the naming scheme. A nested loop structure was chosen which would allow the script to iterate through every value of gap simulated for a given value of entry point, before iterating to the next value of entry point. The numerical values for the gap and entry point were concatenated into a string and used to import the files as the script iterated through the inner and outer loops, making the process entirely automatic. It was within the body of these nested loops that the filtering and additional functionality took place. The filtered data is exported as a second .csv at the conclusion of the execution of the inner loop. The same naming scheme is employed, with the addition of "filtered" being appended onto the string representation of the filename.

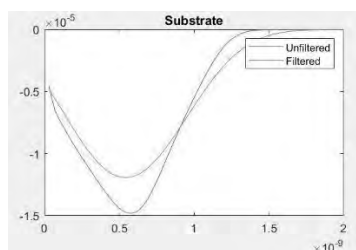
As the loops iterates there is additional data that is vertically appended onto a table. This table is exported after the complete execution of the nested loop structure.

### *Filter Design and Implementation*

A finite impulse response (FIR) filter was chosen for this script, as it is likely to be the same type of filtering performed on data read out with physical scopes. The native FIR function within MATLAB allows for the specification of a variable cutoff frequency, which then must be normalized to the value of the Nyquist frequency for use as an input parameter for the filter function. The order of the filter can also be passed into the function as a variable. For this purpose, an order of 64 was chosen. Of the total samples obtained with the simulation only the first 99 were processed in the filter, as they shared a uniform sample rate. This was included in the script as a variable.

### *Additional Functionality*

It was also desirable to keep track of two additional parameters as the script was being executed- the maximum amplitude was recorded for every metal present in every simulation, and the sum of the currents present at every metal within a single pulse. The values were recorded onto a table as the script iterated through its loop structure and were exported at the end of runtime in a separate .csv file to be analyzed. The script also gave the user the option to plot the filtered data during runtime for faster visualization (Fig. 7).

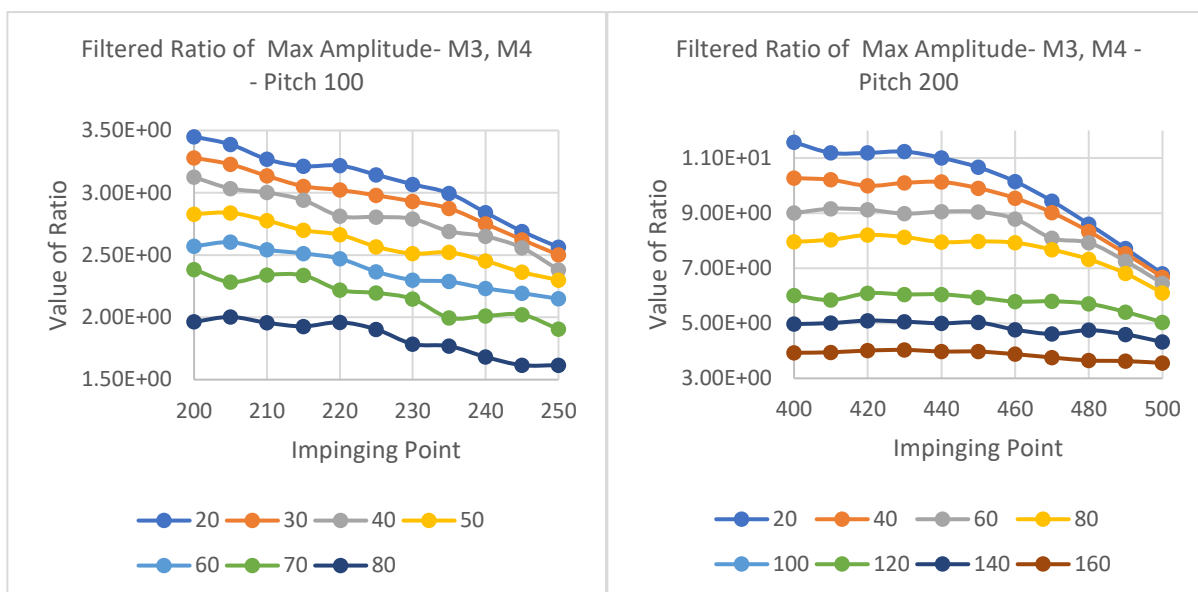


**FIGURE 7.** Filtered waveform with a bandwidth of 1GHz.

### **Organization and Visualization of Filtered Data for Analysis Using Excel Workbooks**

Once the three major sets of simulations were completed and processed through the MATLAB script it became necessary to arrange the resulting data using more concise means than separate, individual .csv files. The maximum amplitudes recorded for each metal were plotted as a function of the entry point for both the filtered and non-filtered MATLAB script results. This allowed us to investigate the values themselves in an easier way, but also identify any commonalities in the shapes of the graphs, and any outliers that may have been present (Fig. 8).

Several other graphs were produced, including the maximum amplitudes plotted as a function of entry point for every value of gap, the sum of the maximum amplitudes recorded at metals 2 & 3, the sum of the maximum amplitudes recorded at metals 2, 3 & 4, and the ratios of the maximum amplitudes recorded at metals 3 & 4 and metals 2 & 3.



**FIGURE 8.** Comparison of ratio of max. amplitudes recorded at metal 3 and metal 4 for pitch 100 & 200 data

### Transient Analysis Using DeckBuild & TINA-TI

During this appointment a significant amount of time was spent learning about the basics of circuit analysis with the purpose of ultimately being able to perform transient analysis in two ways. TI's TINA-TI circuit simulation and analysis software was used to create a crude schematic of the electrical components within an AC-LGAD device. Transient analysis tools within this software were used to attempt to replicate some of the pulse shapes obtained through the DeckBuild simulations that had already been performed. This was found to be difficult with the information obtained through self-study of circuit analysis basics.

Another technique was also started, which would allow for transient analysis to be performed using command within DeckBuild. This deck would take in structure files generated with other decks and perform the necessary steps given values for capacitance and resistance for the important parts of the structure.

### ADDITIONAL WORK & DISCUSSION

The completion of the transient analysis would represent the next major step in this process and would allow us to develop a better understanding of the way the signal develops within these devices. Once the relationship between the aspects of the pad geometries and the resulting signal are understood researchers can begin fabricating physical devices and generate data similar to what had previously been simulated to validate the simulated findings. Eventually, researchers will be able to begin to determine if these parameters are compatible with any of the possible interconnection techniques known by the lab and attempt to use them outside of purely theoretical applications.

### ACKNOWLEDGMENTS

Special thanks are in order to the following individuals for their extensive and ongoing support throughout this endeavor- Gabriele Giacomini, for his knowledge and generosity during the entirety of the internship program, and the entirety of the Office of Educational Programs, Brookhaven National Lab and the Department of Energy at large

for conducting such an enriching and beneficial internship program. Additional thanks to Professors Peter Maritato and Robert Lambiase for their unwavering support and encouragement every step of the way.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program (CCI).

## REFERENCES

Giacomini, G., Chen, W., D'Amen, G., & Tricoli, A. (2019). Fabrication and performance of ac-coupled Igads. *Journal of Instrumentation*, 14(09). <https://doi.org/10.1088/1748-0221/14/09/p09004>

Lutz, G. (1999). *Semiconductor radiation detectors: Device physics*. Springer.

Alexander, C. K., & O., S. M. N. (2021). *Fundamentals of electric circuits*. McGraw-Hill Education.



Interaction energy analysis between diiron enzymes and their natural substrates.

Ambika Natarajan, Department of Chemistry, Emory University, Atlanta, GA 30322.

Atiya Banerjee, Chemistry Division, Brookhaven National Laboratory, Upton, NY  
11973.

## Abstract

Homologous and functional diiron proteins are capable of performing biotransformations of alkanes and alkane-derived substrates, relevant to the cost-effective production of biofuels as well as oil spill recovery. However, the electronic level interactions of  $\Delta^9$  desaturase and alkane hydroxylase (AlkB) and their natural substrates are currently not well understood, warranting a detailed theoretical examination employing both quantum mechanics (QM) and hybrid quantum mechanics/molecular mechanics (QM/MM) methods. To generate putative models, molecular docking was performed of the fatty acids into a crystallographic x-ray structure of  $\Delta^9$  desaturase and a homology model of AlkB using an empirical and knowledge-based scoring function. In the second stage, the docked models containing the highest-scoring conformations were treated using various theoretical methods to gain insight into electronic and geometrical factors influencing their mutual reactivity. Our docking results point towards the accuracy of this method in generating initial structures for detailed study. A parallel approach was employed for the homologous protein, AlkB which has been reported to be responsible for targeted hydroxylation of alkanes. Through *ab initio* computations, we aim to find correlations between the alkyl chain length of substrate with the interaction energies as well as introduce sequence-informed mutations to further probe activity of the latter in encompassing the hydroxylation reaction. Inferences drawn from the electronic level analysis of the orbital exchange, energy decomposition and charge transfer between the enzyme and the substrate are expected to provide insight into their mutual affinities and reaction mechanism. Thermodynamic conclusions drawn from the present work would serve as an initial template for further mechanistic investigations. Through this internship, I have gained experience with molecular docking software and the application of various computational chemistry techniques.

## Introduction

Regio and stereo-selective C-H functionalization, epoxidation, oxidation, hydroxylation, and desaturation of alkanes and alkane-derived products are reactions that are highly useful in the areas of biofuels and oil spill recovery. Non-heme diiron (NHF<sub>2</sub>) enzymes have been well-characterized for their ability to catalyze such reactions and have inspired the development of artificial catalysts that mimic their reactivity.<sup>1</sup>

NHF<sub>2</sub> enzymes contain a core with two tetra, penta or hexa-coordinate irons that are bridged by 1,1 or 1,3-carboxylate,  $\mu$ -oxo, or  $\mu$ -hydroxo ligands. The irons are additionally coordinated to glutamic acid, aspartic acid, histidine, or water ligands.<sup>2</sup> While the overall sequence homology between these proteins is low, it rises to 16-31% when analyzing the four-helix bundle about the active site.<sup>3</sup> The oxygenases are carboxylate rich, a family consisting of enzymes such as methane monooxygenase and toluene-4 monooxygenase. The desaturases and hydroxylases, while differing in reactivity, are similar in that they retain the sequence HX(3 or 4)HX(20-50)HX(2 or 3)HHX(100-200)HX(2 or 3)HH, which corresponds to an imidazole-rich active site.<sup>4</sup> Wirstam, et. al. demonstrated the role of increased hydrogen bonding between one of the histidines and a nearby carboxylate, as well as the presence of Van der Waals forces as stabilizing dioxygen at the active site in hemerythrin.<sup>5</sup> Catalytic activity for all of the enzymes of this class requires activation by dioxygen, although the proposed catalytic intermediates differ substantially in configuration.<sup>6</sup> Such similarities in active site also result in facile “switching” of reactivity, with  $\Delta^9$ -desaturase having been shown to display oxygenase activity with a single mutation at position 199 of a threonine to aspartic acid. The change in reactivity with the aspartic

acid substitution is proposed to arise from the carboxylate's ability to rapidly transfer a proton to the activated dioxygen.<sup>7</sup>

Improving computational models of non-heme diiron enzymes is required to understand the factors contributing to divergence in reactivity. Experimental data reveals insight into the structural, spectroscopic and kinetic aspects of these enzymes, but less is known about the potential energy landscape between the receptors and their substrates.<sup>8</sup> Such information would be invaluable for modelling enzyme-substrate interactions. Common approaches to studying enzyme-substrate interactions include molecular dynamics (MD) or Monte Carlo (MC) methods, which measure the free energy of binding between the enzyme and the substrate as a function of time or distance.<sup>9,10</sup> By contrast, molecular docking is stationary, and is used to sample possible substrate conformations in an active site using a thermodynamic approach that sums the energetic contributions of the intermolecular and intramolecular interactions and searches for a global minimum of the total.<sup>11</sup> Empirical, knowledge-based, force field, or hybrid scoring functions determine how the minimum free energy of binding is reached. Empirical functions sum weighted components of bonding and non-bonding interactions, knowledge-based functions use statistical analysis to determine which conformations are the most probable, and force field functions sum the electrostatic and Van der Waals interactions to generate a potential.<sup>12</sup> A review comparing AutoDock and AutoDock Vina found that AutoDock Vina was better at docking substrates to metalloenzymes, in this case containing Mn(II), Mg(II), and Zn(II) cofactors, in comparison to AutoDock. This assessment was based on enrichment factor (EF1%) values, which measure the number of active ligands when 1% of the ligand and decoy database is searched randomly divided by the number of active ligands that would be expected within 1% of the same database.<sup>13</sup> Additionally, few studies have examined molecular docking on NHFe<sub>2</sub>'s

using AutoDock Vina, one study addressing the docking of iron ions into an apoprotein rather than enzyme-substrate interactions.<sup>14</sup> Since MD, MC, and molecular docking methods address the thermodynamics of the system, these approaches are limited without further consideration of their electronic behavior, and thus reactive nature. This can be performed by computational methods that aim to account for the quantum mechanics of the system, such as Density Functional Theory (DFT) computations.<sup>15</sup>

Computational models are also being used to propose structures for proteins that have not been characterized by X-ray diffraction. Homology modelling starts with the protein amino acid sequence before screening evolutionarily similar proteins that may influence the secondary, tertiary and quaternary structures of the protein. This method has been successful in modelling metalloenzyme systems, including diiron proteins such as ferredoxin.<sup>16</sup> A homology model of the  $\omega$ -alkane hydroxylase, AlkB, has been devised primarily from the starting structure of castor  $\Delta^9$ -desaturase. While AlkB is most closely related to xylene monooxygenase, with a sequence homology of 25%, its secondary structure has a stronger resemblance to the castor  $\Delta^9$ -desaturase.<sup>17</sup> This paper presents a study on molecular docking of natural substrates to diiron enzymes to determine if a) molecular docking is an appropriate method of positioning a substrate onto a  $\text{NHFe}_2$  enzyme and b) docking a natural substrate to the AlkB homology model produces scores within a reasonable range.

## Methods

All X-ray structures were obtained from the Protein Data Bank.<sup>18</sup> Coordinates for the substrates were either obtained from the PubChem Data Bank or built in Avogadro and optimized at the B3LYP/6-31Gp level of theory.<sup>19, 20</sup> Molecular docking was performed using AutoDock Vina, and UCSF Chimera was used as the visualization interface.<sup>21, 22</sup> The latter

software was also used to prepare the receptor X-ray structures by removing precipitation cations and solvent water molecules that were not coordinated to the iron centers, as well as to add appropriate charges. Images were made using UCSF Chimera. Any modifications to the active site to model its dioxygen-activated state were performed using GaussView 6, with reference to reported intermediates.<sup>6</sup> Such alterations were conducted primarily to account for oxidation states within the system. The docking volume was centered about the active site, and the volume was set between 10-25 Å.

Initially, several substrates were docked to dioxygen-activated diiron active sites to obtain a range of scores and approximate enzyme-substrate distances. This required the use of PDB files for toluene-4 monooxygenase (PDB 5TDV), CmlI (PDB 5HYG), Benzoyl Co-A epoxidase (PDB 3Q1G), and castor- $\Delta^9$ -desaturase (PDB 2J2F, 1AFR).<sup>23,24,25,26,27</sup> The respective substrates were toluene, chloramphenicol, and stearic acid, the first two obtained from PubChem.<sup>28,29</sup> In order to test the accuracy of binding pathway-independent docking as a method for positioning substrates, a comparison was made for a single NHFe<sub>2</sub> enzyme with a pathway-dependent molecular dynamics study by Baer et. al.<sup>10</sup> This paper examined the impact of substrate length on the tendency for a substrate to adopt one of two binding motifs, determining the free energy of binding with respect to the distance between the center of masses of the binding pocket and the substrate. The receptor used in the literature study and for comparison was castor  $\Delta^9$ -desaturase (PDB 2XZ1, 1AFR).<sup>30,27</sup> The substrates of capric acid, myristic acid, palmitic acid, and stearic acid were built in Avogadro and optimized at the B3LYP/6-31G level of theory. A different substrate, octane, was built in the same manner for docking on the AlkB homology model. Additional receptor pairings, including xylene/o-toluene monooxygenase (PDB 1T0Q) with xylene and octane, were used as controls.<sup>31,32</sup> Key residues in the substrate

channel at the entrance of the active site, which for human integral membrane- $\Delta^9$ -desaturase and xylene monooxygenase were reported in literature, were examined to evaluate the structural similarities between the binding pockets.

## Results

The initial docking experiments between toluene-4 monooxygenase and toluene, CmII and chloramphenicol, Benzoyl Co-A epoxidase and toluene, and  $\Delta^9$ -desaturase (PDB 1AFR) and stearic acid all produced enzyme-substrate complexes with negative scores, indicating that substrate binding for each pair was favorable. Additional enzyme-substrate complexes were also tested. **Figure 1** shows the shortest distance between the docked substrate and an iron center against the score. The distance between the docked substrate and the closest iron was used to benchmark the general proximity of the substrate to the active site, as the search volumes differed slightly depending on the size of the substrate. The average distance between the substrates and the closest iron was 3.76Å with a standard deviation of 1.55Å. The average score was -5.11 kcal/mol with a standard deviation of 1.52 kcal/mol. The CmII and chloramphenicol complex deviated the most in terms of enzyme-substrate distance, measuring 7.51 Å and falling more than two standard deviations outside of the mean. Toluene-4 monooxygenase deviated the most in score, measuring -2.6 kcal/mol and falling just outside of two standard deviations of the average -5.1 kcal/mol. Although the low  $R^2$  value of 0.3552 indicates poor correlation, the scores and distances are localized between set ranges of -3 and -7 kcal/mol and 2 and 5Å, respectively.

After quantifying the approximate range of binding energies as well as the level of accuracy of molecular docking with diiron systems, the homology model of AlkB was tested. Initial attempts at docking of the natural substrate, octane, to the active site produced negative scores when the substrate was far from the active site, with a search volume as low as 15\*15\*15

Å<sup>3</sup>, and positive scores when the substrate was close to the active site (**Figure 2**). For a score of 1.0 kcal/mol, the distance was 6.55 Å. The distance was over one standard deviation away from the mean, and the score was over two standard deviations from the mean. The positive score represents an unfavorable binding affinity.

## Discussion

While no clear correlation can be drawn between docking score and substrate-active site distance given variations in substrates and docking volume, the closely related o-xylene/toluene monooxygenase and  $\Delta^9$ -desaturase consistently report negative scores between -4 and -7 kcal/mol and have minimum distances between 2 and 5 Å. As the AlkB homology model has substantial sequence similarity with xylene monooxygenase and was modelled with respect to  $\Delta^9$ -desaturase, the score and distance for AlkB and octane appear to be outliers. This is further supported by the docking of octane to xylene monooxygenase and  $\Delta^9$ -desaturase, producing scores of -5.0 and -4.8 kcal/mol and distances of 3.9 and 3.1 Å, respectively (**Figure 3**).

The comparison of the docking results to the free energies of binding reported in the study by Baer et. al. was also in good agreement. This paper performed a steered molecular dynamics study with GROMACS 4.6.2 with a OPLS\_AA force field and a TIP3P water potential on the saturated fatty acid chains of capric acid, myristic acid, palmitic acid, and stearic acid within the  $\Delta^9$ -desaturase enzyme. The results indicated that capric acid adopts motif A, myristic acid is equally likely to adopt motifs A and B, and palmitic and stearic acid will adopt motif B.<sup>10</sup> The molecular docking results on their own had better overlap with the free energies of binding for myristic, palmitic and stearic acid adopting motif A, the less probable pathway (**Figure 4**). Given the maximum difference of about 2.5 kcal/mol between the docking score and molecular



dynamics free energy of binding at motif B for stearic acid, molecular docking appears to be a valid approach to approximating, rather than identifying, possible enzyme-substrate configurations.

A qualitative comparison of the active sites of the enzymes indicated that the homology model of AlkB had similar residues in the substrate channel defining the entrance to the active site. For example, xylene monooxygenase has an entry channel defined by the residues Ile-100, Phe-176, and Phe-196, each separated by 4-6 Å and about 8 Å from the diiron center.<sup>31</sup> In the AlkB model, similar residues of Leu-134, Phe-164, and Leu-265, each separated by 3-4 Å are located about 10 Å from the active site. Additionally, the histidine-rich active site is similar to that of human integral membrane- $\Delta^9$ -desaturase (PDB 4ZYO) and differs from the carboxylate-rich castor- $\Delta^9$ -desaturase (PDB 1AFR) on which it was modelled.<sup>33</sup> Additionally, a search with UCSF Chimera revealed that the castor- $\Delta^9$ -desaturase does not have Phe, Ile, or Trp residues facing its active site, suggesting that its active site entrance is defined by the Leu-108, Thr-140, and Gln-112 residues that lie perpendicular to the position of the substrate entrance with respect to human integral membrane- $\Delta^9$ -desaturase (PDB 1AFR), xylene monooxygenase (PDB 1T0Q), and the AlkB model (**Figure 5**). Variations in the listed receptor X-ray structures, however, have minimal impact on the score and minimum iron-substrate distance, indicating that docking at the AlkB homology model produces unusual scores and distances (**Figure 3**). This may be the result of a narrow binding pocket in the rigid receptor of AlkB, which has a longer depth between the diiron center and the active site entrance residues of Leu-265, Phe-164, and Leu-134, as well as a shorter distance between each of these residues. There may also be steric hindrance due to the orientation of the histidines about the diiron center.

## Conclusion

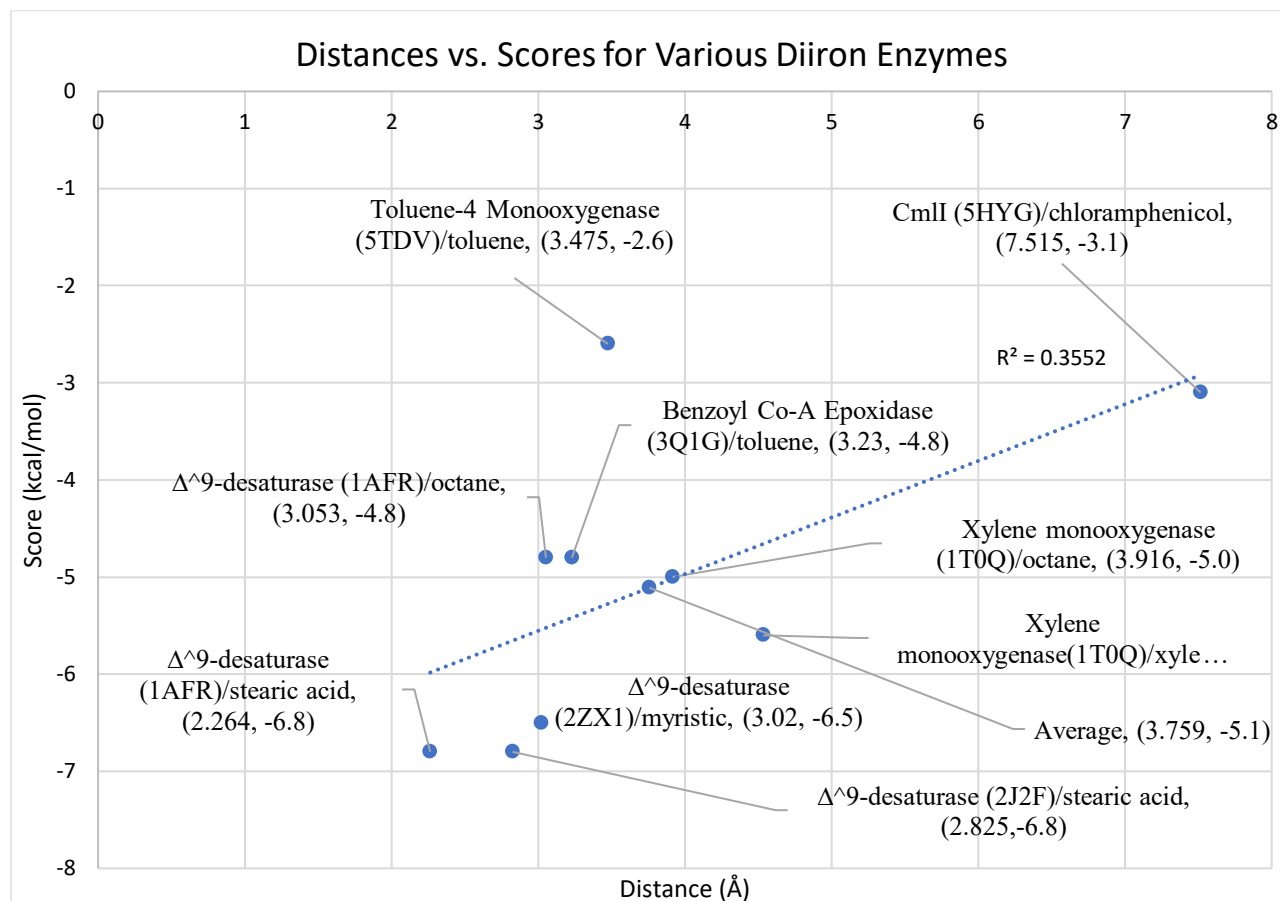
This study showed that molecular docking of substrates to N<sub>H</sub>Fe<sub>2</sub> enzymes by AutoDock Vina is comparable to other methods of substrate positioning, namely steered molecular dynamics. The negative binding affinities resulting from the interactions between enzymes and their substrates, with the substrates close to the active site centers, reveals that this method generates appropriate enzyme-substrate configurations. The AlkB homology model shares distinct similarities with other receptors, such as xylene monooxygenase and human integral membrane- $\Delta^9$ -desaturase. Thus, it is unexpected that docking to this active site should produce positive scores close to the active site. The model may be further tested by docking a range of alkanes of different chain lengths to the AlkB model to see if there is a correlation between chain length and score.

While this study suggests that molecular docking is an adequate method for screening substrates as well as evaluating homology models of diiron enzymes, more sensitive structural and energetic information about the enzyme-substrate complex can be revealed from quantum mechanics/molecular mechanics (QM/MM) data. By optimizing the electronic and therefore geometric structure of these complexes, it would be possible to directly measure the total energy of the system. This energy can be broken into components by type of interaction, providing information into how specific protein residues may facilitate a tendency towards reactivity. This information would be particularly useful in generating data that can improve computational docking methods specific to transition metal systems.

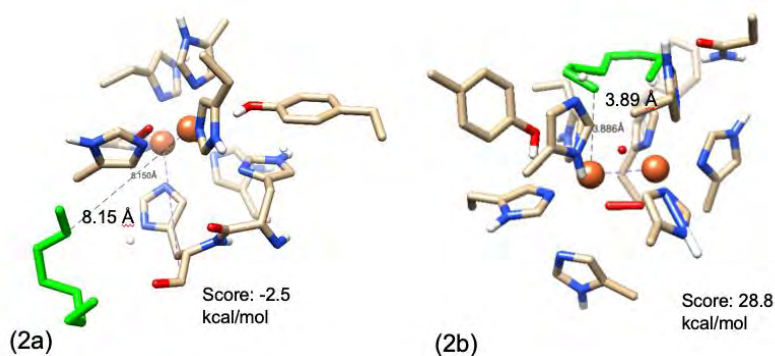
## References:

- <sup>1</sup>Chem. Rev. **104**, 987-1012(2004); doi 10.1021/cr020622y.
- <sup>2</sup>Chem. Rev. **118**, 2554-2592(2018); doi 10.1021/acs.chemrev.7b00457.
- <sup>3</sup>ChemBioChem. **7/8**, 1439-4227(2001); doi 10.1002/1439-7633(20010803)2:7/8<583::AID-CBIC583>3.0.CO;2-5.
- <sup>4</sup>Biochemistry. **33**, 12787-12794(1994); doi 10.1021/bi00209a009.
- <sup>5</sup>J. Am. Chem. Soc. **125**, 3980-3987(2003); doi 10.1021/ja017692r.
- <sup>6</sup>J. Am. Chem. Soc. **142**, 10412-10423(2020); doi 10.1021/jacs.0c01786.
- <sup>7</sup>PNAS. **103(46)**, 17220-17224(2006); doi 10.1073/pnas.0607165103.
- <sup>8</sup>Chem. Rev. **100**, 235-349(2000); doi 10.1021/cr9900275.
- <sup>9</sup>BBA – Bioenergetics. **1862**, 148356(2021); doi 10.1016/j.bbabi.2020.148356.
- <sup>10</sup>Biophys. J. **120**, 1-13(2021); 10.1016/j.bpj.2020.12.036.
- <sup>11</sup>J. Comput. Chem. **31(2)**, 455-461(2010); doi 10.1002/jcc.21334.
- <sup>12</sup>Curr. Comput. Aided. Drug. Des. **7(2)**, 146-157(2011); 10.2174/157340911795677602.
- <sup>13</sup>Appl. Sci. **9**, 4538(2019); doi 10.3390/app9214538.
- <sup>14</sup>JZUS-C. **14(2)**, 118-124(2013).
- <sup>15</sup>J. Biol. Inorg. Chem. **21**, 619-644(2016). Doi 10.1007/s00775-016-1357-8.
- <sup>16</sup>Nucleic Acids Research. **46**, W296-W303(2018); doi 10.1093/nar/gky427.
- <sup>17</sup>J. Bacteriol. **173(5)**, 1690-1695(1991); doi 10.1128/jb.173.5.1690-1695.1991.
- <sup>18</sup>Nucleic Acids Research. **28**, 235-242(2000); doi 10.1093/nar/28.1.235.
- <sup>19</sup>Nucleic Acids Research. **49(D1)**, D1388-D1395(2021); doi 10.1093/nar/gkaa971.
- <sup>20</sup>J. Cheminformatics. **4**, 17(2012).
- <sup>21</sup>J. Comput. Chem. **31**, 455-461(2010).

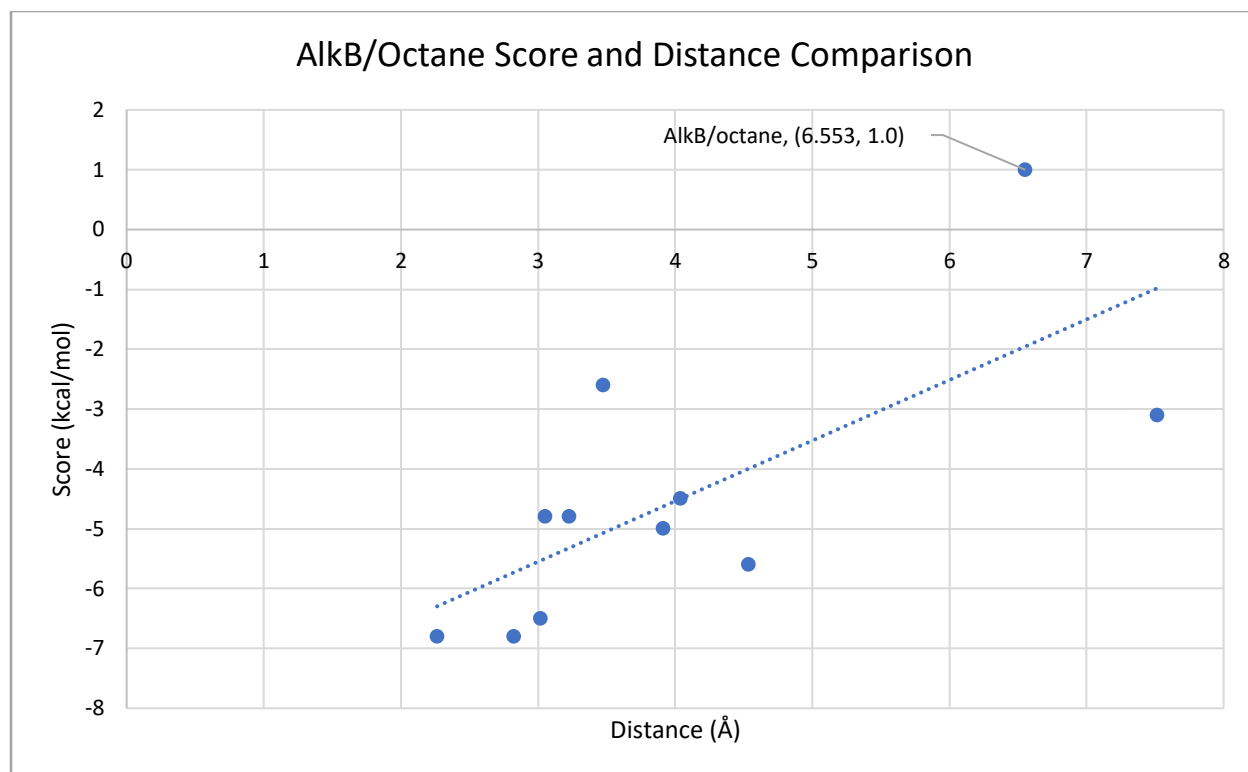
- <sup>22</sup>J. Comput. Chem. **25**(13), 1605-1612(2004); doi 10.1002/jcc.20084.
- <sup>23</sup>Nature. **544**, 191-195(2017); doi 10.1038/nature21681.
- <sup>24</sup>J. Biol. Inorg. Chem. **21**, 589-603(2016); doi 10.1007/s00775-016-1363-x.
- <sup>25</sup>J. Biol. Chem. **286**, 29241-29248(2011); doi 10.1074/jbc.M111.236893.
- <sup>26</sup>Proc. Natl. Acad. Sci. USA. **103**, 17220(2006); doi 10.1073/pnas.0607165103.
- <sup>27</sup>EMBO. J. **15**, 4081-4092(1996).
- <sup>28</sup>National Center for Biotechnology Information (2021). PubChem Compound Summary for CID 1140, Toluene.
- <sup>29</sup>National Center for Biotechnology Information (2021). PubChem Compound Summary for CID 5959, Chloramphenicol.
- <sup>30</sup>Proc. Natl. Acad. Sci. USA. **108**, 16594(2011); doi 10.1073/pnas.1110221108.
- <sup>31</sup>J. Biol. Chem. **279**, 30600-30610(2004); doi 10.1074/jbc.M400710200.
- <sup>32</sup>National Center for Biotechnology Information (2021). PubChem Compound Summary for CID 356, Octane.
- <sup>33</sup>BioRxiv. (2020); doi 10.1101/2020.11.09.373951.



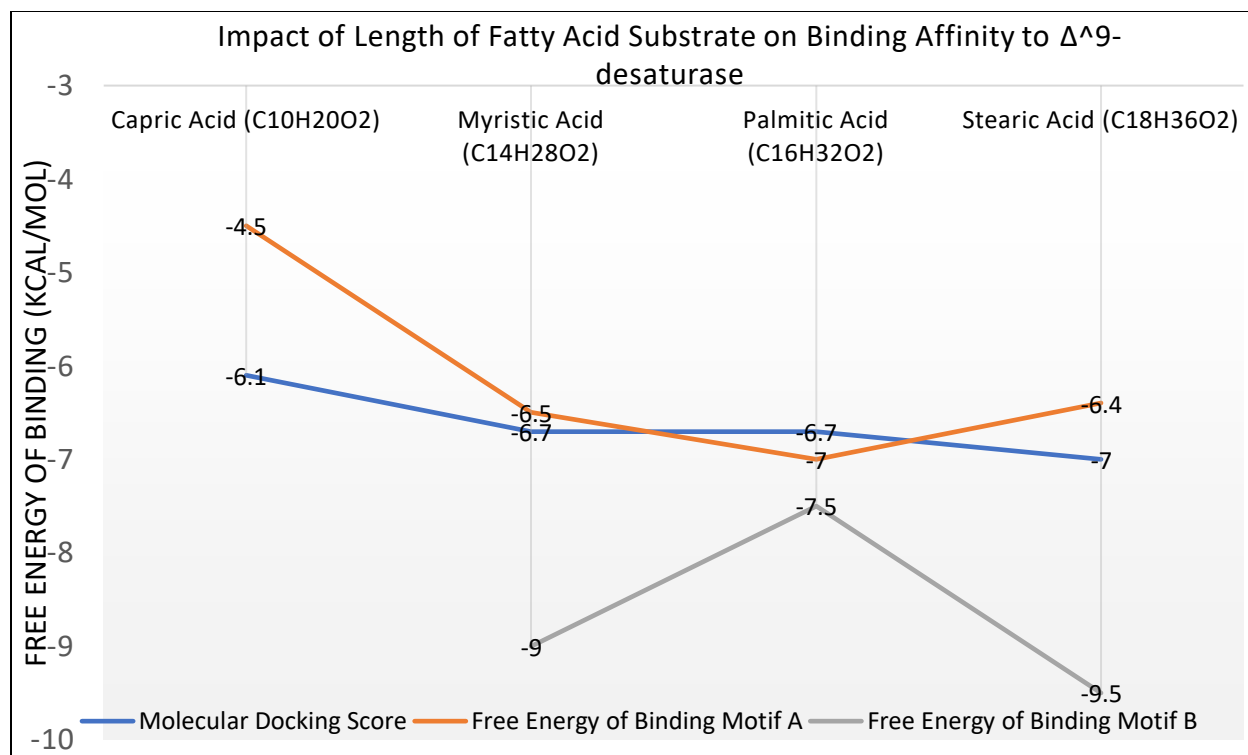
**Figure 1.** Shortest Iron-Substrate Distance vs. Score. The average distance was 3.759 Å and the average score was -5.1 kcal/mol.



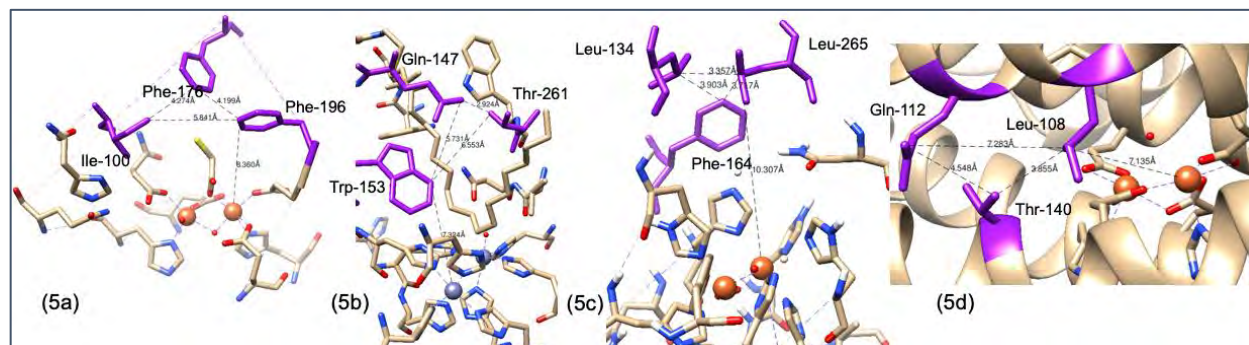
**Figure 2.** Docking of Octane to AlkB Homology Model. Octane is highlighted in green. 2a) Score of -2.5 kcal/mol, docking volume 15\*15\*15 Å<sup>3</sup> centered on active site. Shortest distance Octane/Fe(III): 8.15 Å. 2b) Score of 28.8 kcal/mol, docking volume 8\*8\*8 Å<sup>3</sup> centered on active site. Shortest distance Octane/Fe(III): 3.89 Å.



**Figure 3.** Shortest Iron-Substrate Distance vs. Score, incorporating the AlkB/octane pair. The distance of 6.5 Å and positive score are both uncharacteristic of the docking results produced for similar enzyme/substrate pairs.



**Figure 4.** Comparative assessment of molecular docking scores of fatty acids to castor  $\Delta^9$  desaturase (PDB 2XZ1) with free energies of binding from literature molecular dynamics study.<sup>10</sup>



**Figure 5.** Identification of residues in substrate channel at entrance to active site, highlighted in purple. Typically, three residues define entrance about 8-10Å away from diiron center, and about 3-9 Å away from one another. (5a) Xylene monooxygenase (PDB: 1T0Q). (5b) Human Integral Membrane Stearoyl-ACP desaturase/  $\Delta^9$  desaturase (PDB: 4ZY0). (5c) W-Alkane hydroxylase (AlkB) homology model. (5d) Castor Stearoyl-ACP desaturase/  $\Delta^9$  desaturase (PDB: 1AFR).

# Design and implementation of novel applications for KBase

Boinzemwende Jarmila Roxane Ouango,  
Engineering program, Hostos Community College, Bronx, NY, 10451  
David Dakota Blair,  
Computational Science Initiative, Brookhaven National Laboratory, Upton, NY, 11793

**Abstract**—The Department of Energy Systems Biology Knowledgebase (KBase) is an open-source project enabling collaborative bioinformatics data processing and supporting reproducible science. KBase is designed to be an extensible community resource whose growth is supported by the KBase Software Development Kit (SDK), which enables any developer to build, test, register, and deploy new or existing software as KBase apps. The SDK enables developers to extend the platform’s scientific capabilities. KBase supports data provenance and analysis reproducibility and has a flexible system for sharing data and workflows. Our project at Brookhaven National Laboratory aims to contribute novel applications, modules, and methods to KBase by wrapping an existing tool called JGI-MiniScrub in a KBase app. MiniScrub is a novel Convolutional Neural Network (CNN) based method for de novo identification and subsequent scrubbing of low-quality Nanopore read segments. The technologies required for this include Docker, POSIX shell, python as well as internal KBase APIs. Our approach consists of learning biological topics such as genomes and model organisms, and software engineering concepts such as HTTP, the OSI model, Object-Oriented programming, design patterns, Model-View-Controller (MVC), and public-key cryptography. Additionally, our approach requires effort to achieve familiarity with development tools such as Git, secure shell, and the KBase SDK tool `kb_sdk`. By learning about KBase and its SDK, and implementing MiniScrub in a KBase app, we can promote user-friendly interfaces, and contribute new features to KBase to enable researchers to pursue new avenues of research made possible by these new features.

## I. INTRODUCTION

Long read sequencing has become increasingly important in recent years, with sequencing technologies from companies such as Pacific Biosciences and Oxford Nanopore seeing wide use in a variety of applications including genome assembly, detection of antimicrobial resistance genes, sequencing personal transcriptomes, and improving draft genomes. Sequence assembly is one of the most promising and explored of these applications. Long repeat sections have been shown to be among the most important factors that affect assembly quality, and long sequencing reads are much more capable of resolving these long repeats [1]. Current single molecule, long sequencing reads also have very high error rates, ranging from 5% to 40% per read and often average about 10% to 20% up to as high as 30-40% depending on variables such as the type and version of the sequencing technology and the experiment being performed. These high error rates make assembly and other applications inefficient or error prone. It is thus critical that methods be developed towards addressing this issue so that the potential of

long read sequencing can be fully realized. a novel Convolutional Neural Network (CNN) based method, called MiniScrub, for de novo identification and subsequent scrubbing of low-quality Nanopore read segments.

Our project at Brookhaven National Laboratory aims to contribute novel applications, modules, and methods to the Department of Energy Systems Biology Knowledgebase (KBase) by wrapping MiniScrub in a KBase app.

## II. METHOD

First, In preparation of wrapping MiniScrub in a KBase app, we needed familiarity with biological concepts including genomes and model organism and software engineering concepts such as Object-Oriented programming, and public-key cryptography. we learned about applications development kits such as Git and secure shell. Then, we learned about the KBase software development Kit (SDK) by designing an app called ContigFilters through the `kb-sdk` tutorial available on KBase for internal and external contributions to the platform. In this process, we have installed and gained acquaintance with open-source containerization platform such as Docker. we assimilate how the SDK works and the general concepts involved in making and running apps. In the process of wrapping MiniScrub, we investigate the literature review on the method, its features to understand its aim and how it could potentially support the KBase on its mission of meet the grand challenge of systems biology.

In implementing MiniScrub in the KBase app along with my mentor, we investigate the dependencies required to install the method. These included Docker, and git. Then, we initialized the MiniScrub module containing all the component for the KBase app. This was performed using the KBase SDK commands. Furthermore, we build the Narrative app user interface by creating a directory for the new app through the `spec.json` file and setting the app’s parameters in the `display.yaml` file. Finally, we registered the MiniScrub module in AppDev, the KBase narrative server that is used by app developers to publish in-progress versions of their apps and test and share them using real data.





### III. RESULTS

MiniScrub was wrapped in a KBase app called kb-miniscrub. A demonstration narrative explaining the app's features was created. From the KBase's primary user interface, the Narrative Interface built on the Jupyter10,11 platform, users can (Figure 2) upload their private data, search and retrieve extensive public reference data, select and run the kb\_miniscrub app on their data, view and analyze the results from the app, and record their interpretations along with the analysis steps in the markdown cell. Kb\_miniscrub can be used for de novo genome assembly and large structural variation identification.



Figure 2: kb\_miniscrub user narrative's interface

Kb\_miniscrub takes file a FASTQ as input. To use the app in KBase, users import FASTQ reads objects. Then, MiniScrub performs the following steps by generating reads to reads alignments by MiniMap, encode the alignment into images, and build Convolutional Neutral Network (CNN) models to predict low quality segment that could be scrubbed based on a customized quality cutoff. Finally, the app outputs a new FASTQ file with the scrubbed reads. MiniScrub accurately scrubs out low-quality segments within Nanopore raw reads to improve overall read quality, and the scrubbed reads lead to fewer assembly errors. In addition, MiniScrub outperforms alternative preprocessing tools in terms of leading to fewer mis-assemblies and large indels in de novo assembly.

The implementation of the MiniScrub method on kb-miniscrub for KBase can support new avenues of research and promote novel discoveries at Brookhaven National Laboratory as well as anywhere across KBase. MiniScrub robustly predicts low-quality segments within Nanopore reads. It predicts the percent identity of each segment of a read and scrubs out segments below a user-set threshold, splitting the reads at the low-quality regions [1]. Scrubbing enriches the high-quality read population. scrubbing out a small percentage of low-quality regions nevertheless raises average read percent identity by over 3% (from 83.1% to 86.2%). MiniScrub does not seem to perform much false scrubbing, as high-quality raw reads (particularly those at 90% or higher accuracy) remain almost entirely intact (figure 3) [1].

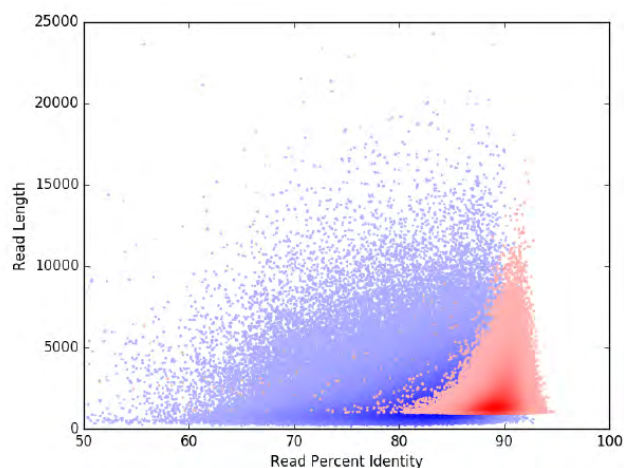


Figure 3: Density scatter plot showing average read quality improvement by MiniScrub versus raw reads [1]

Users are no longer require any sysadmin experience to install the MiniScrub command as it is available on KBase. Data provenance is easily manageable. This aligns with the platform strong commitment to data provenance and its mission of enabling collaborative bioinformatics data processing.

### IV. CONCLUSIONS

The KBase app kb-miniscrub's features can support the grand challenges in long reads sequencing using the MiniScrub method. The implementation of the app narrative's user interface is still in progress as we are updating the app's module to enhance users' experience. In terms of the output, when a sequencing reads is imported in the app, the MiniScrub method returns an empty FASTQ file causing the kb-miniscrub app to output the same. We expected the method and thus the app to return a FASTQ file including scrubbed reads with removed low-quality segments. We have already made changes in the MiniScrub's code locally. Further analysis is needed regarding the MiniScrub's output so we can apply these changes to the app.

### REFERENCES

- [1] MiniScrub: de novo long read scrubbing using approximate alignment and deep learning Nathan LaPierre, Rob Egan, Wei Wang, Zhong Wang bioRxiv 433573; doi: <https://doi.org/10.1101/433573>.

### APPENDIX

I would like to thank my mentor Dr. David Dakota Blair for his generous support and dedication to my learning experience. Also, I want to thank Ziming Yang for her thoughtful contributions to our discussions, and project. Additionally, this project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program (CCI).

***Home-based Citizen Science Program for middle school students to build skills necessary to  
conduct authentic research***

Alyssa Probst, SULi intern, St. John's University, Queens, NY, 11439

Amanda Horn, Department of Educational Programs, Brookhaven National Laboratory

## **I. Abstract**

The Office of Educational Programs at Brookhaven National Lab offers a summer program called the Citizen Science Program that promotes science, builds core observation and research abilities, as well as increasing youth understanding of STEM careers. Remote instruction will be used in this program to engage middle school students in local research and data collection relating to environmental and atmospheric science, plants, insects, and birds. The purpose of this program is to encourage students to participate in larger citizen science projects by collecting and analyzing long-term data. Our study used innovative assessment practices to obtain and document broad learning outcomes based on the curriculum of Introduction to Citizen Science. I have developed assessments that frameworked the curriculum and design-based assessment methods were used to connect and measure growth. Data was collected in the form of pre-and post-assessments of the students to track their progress in scientific skills and knowledge to aid in the creation of future programs. An approach to collecting data on assessing students and checking progress, is through using 5-point entry/exit assessments. These questions range from 1point, 5 questions assessment to 3point and 2 point short/long answer questions. By using short 5-minute exit exams after each weekly session, we were able to access not only progress but background knowledge and engagement. In doing so, we can analyze the data and find connections between the style of assessment, in addition, enrichment between each lesson. Using this format, we can observe and contain quantitative data on students' progress. As a

result of this summer, I have added assessment building to my repertoire of techniques. Additionally, I am now familiar with creating accurate science benchmark assessments, as well as assessing designs in virtual education environments.

## **II. Introduction**

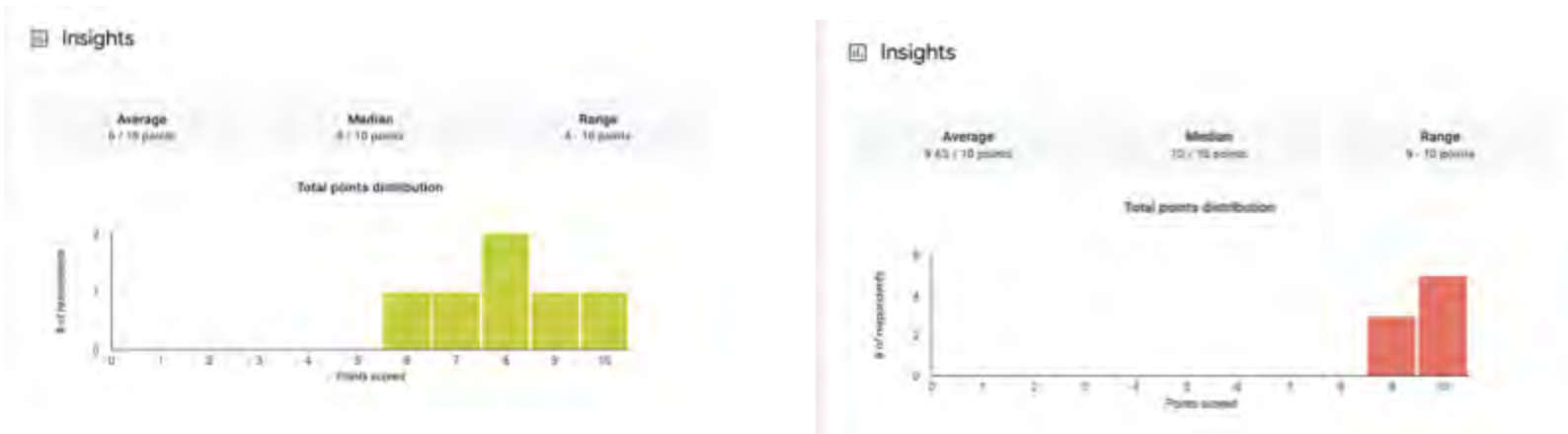
Informal education is a method of developing learning that is smart, fair, and engaging. Conversation, as well as the investigation and expansion of experience, are key components. This project concentrates on the framework of quality assessments. The benchmark assessments that were created addressed key topics in the Citizen Science curriculum and intended to provide information that could be used to monitor students engagement/performance. In addition, the assessments were used in reflection for improving teaching and learning. The alignment of assessments were as the following:

- Consistent with current scientific knowledge
- Fair in expectations for students
- Within ordinary time restrictions
- Sufficient in reliability

This project explored the effectiveness of the assessments, including the following questions:

1. Was exit/entry assessment adequate for this informal program?
2. Do the data and analytical results obtained from these evaluations aid in the development of alternative assessments?

Fig. 1&amp;2



Shown in figure 1&2 above the range is out of ten points. This is because the entry and exit were taken in the same form after class. Explaining the 5 points for the entry and 5 point exit ticket.

Shown in figure 3&4&5 is one example of the pre- assessment questions and answers that was used for week 1. Creating the questions and coming up with those answers, a lot of collaboration goes into building a question. I ran into multiple obstacles while building assessments. Not only do you have to come up with 5 questions and answers based on the curriculum but how to arrange those exact questions and in what order.

- Keep in mind students accessed the assessment through the google site and the assessment was built in google forms.

Fig. 3 &amp;4 &amp; 5

The figure shows three screenshots of Google Forms questions and answers:  
 1. Question: 'What is Citizen Science?'  
 Options:  a participatory approach to environmental research where you can contribute by collecting scientific data;  a broad area of biology;  the study of living organisms;  the study of the structure, function, and evolution of biological systems.  
 2. Question: 'What are good observation skills?'  
 Options:  Focusing on things you see, hear, smell, taste and think;  Critical thinking;  Staying alert and focused;  All of the above.  
 3. Question: 'What is data collection?'  
 Options:  a report;  statistical significance;  study notes;  The gathering of information that can be observed, measured, and reported to address a research problem.  
 A second question is partially visible: 'Which of the following can be a consequence of improperly collecting data?'  
 Options:  It can be misleading to your research;  It can help solve an experiment;  Improperly collected data can help you answer research questions correctly;  Improperly collected data will NOT have an effect on future work.

Which of the following are subjects of biodiversity? \*

1 point

species

genes

ecosystem

all of the above

In 1-2 sentences, why is biodiversity important among plants? \*

3 points

Your answer:

Fig. 1

← Shown here represents what questions from an assessment looked like. I used google forms to collect the data.

Fig. 6&amp;7

### III. Methods of Assessment



Figure 1. The CAESL Quality Assessment Triangle.

Fig. 8

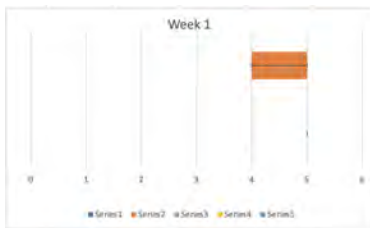
As noted earlier, the assessments consisted of a majority of multiple-choice format items, as they are relatively fast for students to take and relatively easy to score. There are

multiple approach methods to assess students in an informal class, but I chose to use that 5-point entry/exit assessment. Having these 5 minute exams, we were able to acknowledge the background knowledge students are coming into class with and compare and contrast that from what they learned from the lesson. Informal education can be learning projects that are done on our own or education that comes from being involved in a summer program like Introduction to Citizen Science. Informal learning can be more enjoyable to learn and frequently leads to increased creativity and innovation. Citizen Science satisfies curiosity in research while also assisting us in expanding our knowledge base, acquiring good research skills, or honing an existing talent.

**IV. Results**

Week 1	Exit 06/29	Entry Ticket 06	Exit Slip 07/06	Entry Ticket 13	Exit Slip 07/13	Entry Ticket 20th	Exit Slip 07/20	Entry Ticket 07/20	Exit Slip 07/27
5.0	4.9	4.9	4.2	3.8	4.5	3.4	3.7	3.8	4.7
5	5	5	4	4	5	2	4	5	5
5	5	5	4	4	5	2	4	4	4
5	5	5	3	4	5	4	4	4	5
5	5	5	4	2	5	2	3	2	5
5	5	5	5	4	5	3	4	5	5
5	5	5			2	3			4
5	5	5			5	4			5
					4	4			5
									5

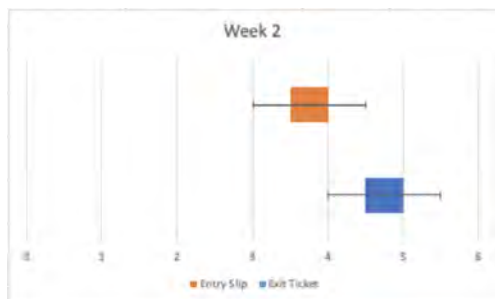
- Cumulative data of test scores for each week



	Variable 1	Variable 2
Mean	5	4.85714286
Variance	0	0.14285714
Observations	7	7
Hypothesized	0	
df	6	
t Stat	1	
P(T<=t) one-	0.17795884	
t Critical one	1.94318028	
P(T<=t) two-	0.35591768	
t Critical two	2.44691185	

- Data plot contains a box and whisker chart of week 1 pre/post assessments. The

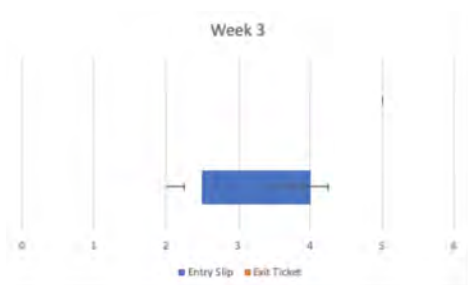
average for these test scores were 5.0-4.9. The p-value for this plot was 0.17795, concluding there was no statistical significance .



t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	Variable 2
Mean	4.8	4.2
Variance	0.2	0.7
Observations	5	5
Hypothesized	0	
df	6	
t Stat	1.41421356	
P(T<=t) one-	0.10351563	
t Critical one	1.94318028	
P(T<=t) two-	0.20703125	
t Critical two	2.44691185	

- Data plot contains a box and whisker chart of week 2 pre/post assessments. The average for those test scores were 4.8-4.8. The p-value

for this plot was 0.103515625, concluding there was no statistical significance.



t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	Variable 2
Mean	3.75	5
Variance	1.58333333	0
Observations	4	4
Hypothesized	0	
df	3	
t Stat	-1.9867985	
P(T<=t) one-	0.07056097	
t Critical one	2.35336343	
P(T<=t) two-	0.14112194	
t Critical two	3.18244631	

- Data plot contains a box and whisker chart of week 3 pre/post assessments. The average for those test scores were 3.7-5. The p-value for this plot was 0.070561, concluding

there was no statistical significance. However, shown in the red circle represents how all students received a 5/5. We can conclude growth and progress in students' understanding of that lesson.

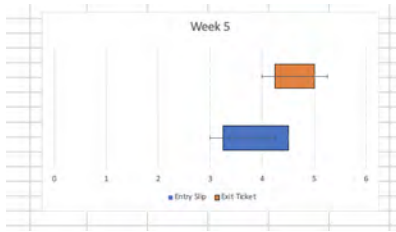


t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	Variable 2
Mean	3.5	3.5
Variance	1.6666667	0.3333333
Observations	4	4
Hypothesized	0	
df	4	
t Stat	0	
P(T<=t) one-t	0.5	
t Critical one	2.1318468	
P(T<=t) two-t	1	
t Critical two	2.7764451	

- Data plot contains a box and whisker chart of week 4 pre/post assessments. The average for those test



scores were 3.5-3.5. The p-value for this plot was 0.5, concluding there was no statistical significance.



t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	Variable 2
Mean	4.25	4.75
Variance	0.916667	0.25
Observation	4	4
Hypothesize	0	
df	5	
t Stat	-0.92582	
P(T<=t) one-tail	0.198513	
t Critical one-tail	2.015048	
P(T<=t) two-tail	0.397027	
t Critical two-tail	2.570582	

- Data plot contains a box and whisker chart of week 5 pre/post assessments. The average of those test scores were 3.8-4.7. The p-value for this

plot was 0.198513, concluding there was no statistical significance.

We highly encourage students to answer the exit tickets but most weeks we would only have an average of about 6 students answer. This made it difficult to analyze the final results due to the compressed data being very minimal in showing results. With future programs, more data and time out of lessons to ensure finalized assessments should be considered. We can tell students have learned something throughout the classes due to the averages per class increase in intervals.

## V. Lesson

Lesson: iNaturalist

I was also given the opportunity to design and implement my own lesson in week 5 of the program. I designed a lesson that assisted students with understanding and the importance of biodiversity. Students will investigate biodiversity as it includes the diversity of plant species, and ecosystems. The students will also touch on the importance of maintaining diversity and the health of the planet with biodiversity. The activity was based off of <https://www.inaturalist.org/>. To reflect on presenting my lesson and implementing my activity, in the future, I will give clear instructions and take more time presenting the lesson. I will also try to get students more engaged by asking more questions throughout.

A. Pictures of slides from the lesson iNaturalist and Biodiversity

## VI. Detailed Lesson Plan Below:

Lesson: Inaturalist

**Title:**iNaturalist

**Time frame:** 40-45 mins

**Overview:** This module is designed to assist students with understanding the importance of biodiversity. Students will investigate biodiversity as it includes the diversity of plant species, and ecosystems. Because all living things are connected, the students will also touch on the importance of maintaining diversity as critical to the health of the planet. This lesson is based off of <https://www.inaturalist.org/>. Since students are coming from all different parts of Long Island it's important to recognize the differences and similarities between where we live.

**Objective:** After completion of this module, the student will be able to

- investigate the characteristics of plants, and classify similarities/differences across various plants.
- demonstrate knowledge of biodiversity, its contributions, and its benefits

### **Materials:**

1. Laptop
2. Camera

Vocabulary:

- Biodiversity
- Species biodiversity
- Species
- Plant diversity

**Launch:** We will discuss “ Compare and Contrast”. Here we discuss the difference between two pictures of different hydrangeas. They use their keen observation skills here.

- Watch a short video on connecting what is biodiversity and why we need biodiversity. This is to review or give insight to students' background knowledge on the subject. This should only take around 5 mins.
- [https://www.youtube.com/watch?v=b6Ua\\_zWDH6U](https://www.youtube.com/watch?v=b6Ua_zWDH6U) (Stop at 1:35)

Additionally, I will present a lesson powerpoint on what biodiversity is, why and how it is relevant to plants/soil and review an example of a species with great diversity. I went in detail on the three subjects of biodiversity and how we will only be focusing on one this lesson. In succeeding slides we will discuss the benefits of biodiversity relevant to plants and species biodiversity.

- The examples used for species and plant diversity was a Fern. Today, ferns are the second-most diverse group of plants on Earth
- The launch should gain their interest into the subject, but their background knowledge should peak their interest into learning more. I will also explain how diversity within soil is important to us as humans and can lead to more crops and food production. Using an example of soil was important for students to be able to connect to something they already know to something new.
- In this lesson I will be using the zone of proximal development to get my students where they are now and understand how biodiversity can relate to previous learned subjects.

**Preparation:** To introduce this activity I will review the lesson slides. In citizen science, scientists observe and collect data, so in this activity students will be asked to compare and contrast two different plants from their yard/house. Students will also learn about resources on iNaturalist and where they can also contribute their observations on the app. Students should think of reasons and critically think on how these plants at home are contributing to biodiversity and why their differences are important.

### Activity (Explore): Field Guide

1. Explain the type of species they discovered.
2. Have students sketch or upload images of their findings
3. Have each student complete the Nature Field Guide exercise.

**Nature Field Guide**

Name: \_\_\_\_\_

The type of specimens: \_\_\_\_\_ and \_\_\_\_\_

Ex: Is it a flower, leaf, etc)?

Sketch your observations here or upload two images

\_\_\_\_\_

\_\_\_\_\_

Date: \_\_\_\_\_

Location: \_\_\_\_\_

Explorer's Notes: (Tell me about your species) Ex: color, size, texture, shape

\_\_\_\_\_

\_\_\_\_\_

One similarity and one difference:

4. Have students compare and contrast one similarity and difference.



**Safety Tips:** Be careful of the plants around you, especially of poison ivy, and make sure to be a good observer!

**Going Further (Extensions):**

Give students the resource link if they want to identify the plants they have discovered. Then students will be asked to map it out. Students will pin where they are located on Long Island to show how diverse as a class we are and how all these different types of plants are native and live across the Island. I want students to walk away with an interest/engagement in iNaturalist and an understanding of the importance of biodiversity. This will help

students realize how every home has one or more different plants. If a student goes from Nassau to Suffolk and looks for differences in plants, they will experience the great variety and quantity of differences.

**Assessment (Evaluation):** This activity will be assessed through an exit ticket form.

**Resources:**

[Exit Ticket Form](#)

[How to use iNaturalist](#)

[Quick intro to iNaturalist](#)

[Plant Identification](#) - (Use this website to identify plant/ shows percentages etc)

## **VII. Problem Solving**

The overall lesson went overall smoothly but I did run into a few issues that I would change in the future. Students had issues taking pictures and uploading these images to the google forms activity. In the future, I would plan to have more resources, like a video of how to upload images to google forms, or just have students email me the images. One student ran into this issue, but we resolved it by receiving an email of the pictures taken. In the future, I would also want students to see me uploading an image into iNaturalist. I offered resources for students outside the program to access that but I believe it would have been engaging for them to see me doing that. In the overall review of the images in week 6, I showed students examples of the images submitted. Students explorers notes and explanation of images were specific and I could tell students were engaged by their answers. Some of the students even got creative and submitted different mushrooms and insects. Here is an example of one of the submitted responses: “The first one is a bug and it looks like an ant with wings. It is really small and it looks like it has orange legs. But the other part of him is black. The second picture is a leaf. The leaf feels firm and the leaf looks straight. The leaf is about the size of my hand.”

## **VIII. Conclusion**

As a result of this summer, I have added assessment building to my repertoire of techniques. Additionally, I am now familiar with creating accurate science benchmark assessments, as well as assessment designs in virtual education environments. I would like to

express my sincere thanks to my mentor Amanda Horn for her guidance and knowledge on this project. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

# **A study of fail-slow fault tolerance of databases**

**Andrei Racila**

**Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas, Richardson, TX, 75080**

**Dr. Shuai Mu**

**Department of Computer Science, Department of Computer Science, Stony Brook University, Stony Brook, NY, 11794**



## **Abstract**

Modern day database systems are susceptible to what are known as fail-slow faults where a database instance may suffer from either a software or hardware failure and continue to function but at a huge loss in performance. Though theoretically these databases implement consensus algorithms that should mitigate the impact of these events, there is still much to be desired. The following report focuses on the performance of the Redis data structure store in a distributed context where a fail-slow fault can occur. Redis instances were profiled on virtual machines provisioned on Microsoft Azure in a configuration where 1 instance served as a leader node and the others as followers keeping in sync with whatever changes were made to the leader. Then, one virtual machine was selected (either a leader or follower node) and was subject to an artificially induced fail-slow event while simultaneously being benchmarked for throughput and latency in answering to client requests. Through these benchmarks, we found that when Redis is injected with certain fail-slow faults, overall system performance in answering certain query types drops significantly. Though overall experimental results are mixed, they indicate that, in certain but plausible contexts, Redis is not tolerant against fail-slow faults. These results reinforce the findings of benchmarks ran against other database systems injected with fail-slow faults and indicates a need to rethink and retool the way programmers and computer scientists design these distributed database systems, especially ones used in environments requiring high-availability such as those encountered by cloud-service providers. As a result of this internship, I have learned more about the science of benchmarking computing systems and have gained a deeper understanding of how modern database systems operate.

## I. Introduction

Databases, have long been the subject of study in terms of optimizing, scaling, etc. To make such systems more fault-tolerant, several consensus protocols have been developed such as Paxos and Raft [1] that have been implemented in many databases. These protocols have mainly been modeled around dealing with fail-hard faults and Byzantine faults. While these protocols may handle fail-hard faults where some component in a cluster may completely fail, a new question arises: How well do these systems deal with fail-**slow** faults? **Fail-slow faults** are faults that occur in some piece of software or hardware where the component still operates but in a degraded mode with worse than expected performance [2]. These faults can and have been documented to occur in many major hardware components. For example, fault-masking in processors allows manufacturers to sell slightly flawed chips under a common name. This causes unexpected differences in cache sizes and performance from what is advertised and misleads developers into developing code for systems that assumes larger cache-sizes than what is available, in effect over-estimating a system's possible performance. [3]. Another example is seen in networks where just a few slow switches can cause messages to accumulate and cause excessive network contention [3]. Fail-slow faults are sinister in nature as, without proper monitoring and handling, they can decrease overall system performance and can realistically happen yet are considered as unexpected events by most database systems and are improperly handled. Several popular database systems have already been tested for fail-slow fault tolerance and have been shown to not be fail-slow fault-tolerant [4], but the focus of this paper will be the results obtained from benchmarking Redis: an in-memory data structure store that can be used as a database, cache, and message broker developed by Redis Labs [5]. Redis was chosen in part because it is open source which allows us to test the community version of it at no additional expense and it enjoys popular usage, being used by many organizations among which are large companies such as Instagram [6]. Due to budgetary constraints and the fact that the team is concurrently profiling several database systems on Microsoft

Azure’s cloud platform, the entire benchmarking pipeline was constrained to run on only 4 virtual machines. Experimental configurations may not be accurate to real-life deployments, but nonetheless should highlight some potential issues with Redis and how it deals with fail-slow faults.

## II. Methods

Before any experiments were ran, 4 virtual machines (VM’s) were provisioned on Microsoft Azure, each with the following specifications:

Component	Specification
CPU	4 vCPU’s
Memory	16 Gib
Operating System	Debian 10 “buster” (Linux)
Disk	16 Gib

**Table 1:** Specification of each virtual machine used in every experiment.

Originally experiments were ran using larger disk sizes, up to 64 Gib through Azure P6 disks, but since we had a limited number of Azure credits, the decision was made to decrease the disk sizes. It was reasoned that these results are meant as a “pilot” study on what faults seem most detrimental performance to inform the development teams behind these systems. The virtual machines were provisioned and setup using a Bash script that would 1. Tell Azure through the az command line utility to provision the 4 VM’s, 2. Write the VM configs (names, public IPs, and master IP’s) to a JSON formatted file, 3. Setup the 3 VM clients by ssh ‘ing into the VM’s and installing the dependencies needed to run redis-server 4. Setup the “client” VM by ssh ‘ing into the VM and installing dependencies for redis-benchmark and memtier\_benchmark. The full code is available here:

[https://github.com/xlabuiuc/slooo\\_mongo/blob/redis/old\\_stuff/scripts/redis/create\\_setup\\_servers\\_ssd.sh](https://github.com/xlabuiuc/slooo_mongo/blob/redis/old_stuff/scripts/redis/create_setup_servers_ssd.sh)

Redis is what is called an **in-memory database** meaning that instead of storing its data and records persistent disks such as hard-disk drives (HDD's) or solid-state drives (SSD's), it instead stores it in main memory such as random-access memory (RAM). Having the records and data only in-memory presents the problem that whenever an instance is completely shut down, all the data in-memory is wiped. Redis does provide measures to counter this including several persistence options, including the option to simply turn off any form of persistence [7]. By default, Redis saves point-in-time database snapshots to disk, thus preventing total data-loss if the instance catastrophically fails and shuts down [7]. These persistence options are not too important as only benchmarking/"dummy" data was being written to the instances, we wanted to keep as many options as default as possible, and the virtual machines were shut down after every experiment ran thus wiping the data from the VM's memory. Having persistence on could have impacted results, but only in a way that exacerbated, rather than mitigate, the effect of an injected fail-slow fault (such as slowing down the disk).

3 VM's served as host to a single Redis instance each. This forced the Redis instances to communicate with each other and the "client" VM over the network rather than through inter-process communication and to better control the resource usage of each instance.

In every experiment, **replication** was enabled on each VM, allowing 1 VM to server as a **leader** instance and the other 2 VM's to serve as **replica/follower** instances (though Redis replication has more flexible options). With replication, follower instances try as best as possible to be exact copies of the leader instance with the leader sending a copy of the stream of commands it receives to its replicas (this was another reason why Redis was selected to be tested: it enabled replication by default). This occurs through **synchronization**: whenever the link is broken or impaired between a replica and leader, a **partial resynchronization** (only attempt to recreate part of the command stream of the leader that was

missed) is initiated and when partial resync fails attempt **full resynchronization** (the leader will snapshot the data and send it to the replica along with its command stream) [8]. With Redis in particular, replication is by default asynchronous to have “low latency and high performance” [8]. However, replicas asynchronously acknowledge the amount of data they received periodically [8]. The master doesn’t wait every time for a command to be processed by the replicas, but knows what replica already processed what command. In addition, replication is largely non-blocking on both the replica and master side, so normally when a replica is doing initial synchronization or partial resynchronization a replica will block incoming requests but still answer queries on the old data (while also deleting the old data) and the master will handle queries as normal (though there will be slight delays due to the synchronizations as data is being deleted and new data is being written. This should have negligible impact on the results obtained) [8]. Thus, though Redis is an in-memory database, slowing down the disk and network was theorized (and can still possibly have) an effect on overall system performance. Again, only basic replication was enabled with all settings set to the defaults save the IP addresses and the port numbers of each Redis instance in both the leader and replica configuration files found here [https://github.com/xlab-uiuc/slooo\\_mongo/tree/redis/old\\_stuff/scripts/redis/redis\\_master\\_confs](https://github.com/xlab-uiuc/slooo_mongo/tree/redis/old_stuff/scripts/redis/redis_master_confs) and here [https://github.com/xlab-uiuc/slooo\\_mongo/tree/redis/old\\_stuff/scripts/redis/redis\\_replica\\_confs](https://github.com/xlab-uiuc/slooo_mongo/tree/redis/old_stuff/scripts/redis/redis_replica_confs) respectively.

Some experiments also tested Redis’ **Sentinel** feature [9]. Sentinels run on top and augment regular Redis replication and provides monitoring and notification features alongside automatic **failover**: when a master is not working, another Sentinel can start a process to become promoted to master and all other Sentinels recognize it as the new master [9]. A **quorum** specifies the number of Sentinels that must agree that the master is unreachable to trigger a failover. However, after a failover is triggered a **majority of all the Sentinels** must agree that the master is unreachable to complete the failover and elect a new master [9]. The quorum number was left to its default value of 2 in the configuration files

for the leader and replica Sentinels here

[https://github.com/xlabuiuc/slooo\\_mongo/blob/redis/old\\_stuff/scripts/redis/sentinel\\_master\\_template.c](https://github.com/xlabuiuc/slooo_mongo/blob/redis/old_stuff/scripts/redis/sentinel_master_template.conf)

[onf](https://github.com/xlab-) and here <https://github.com/xlab->

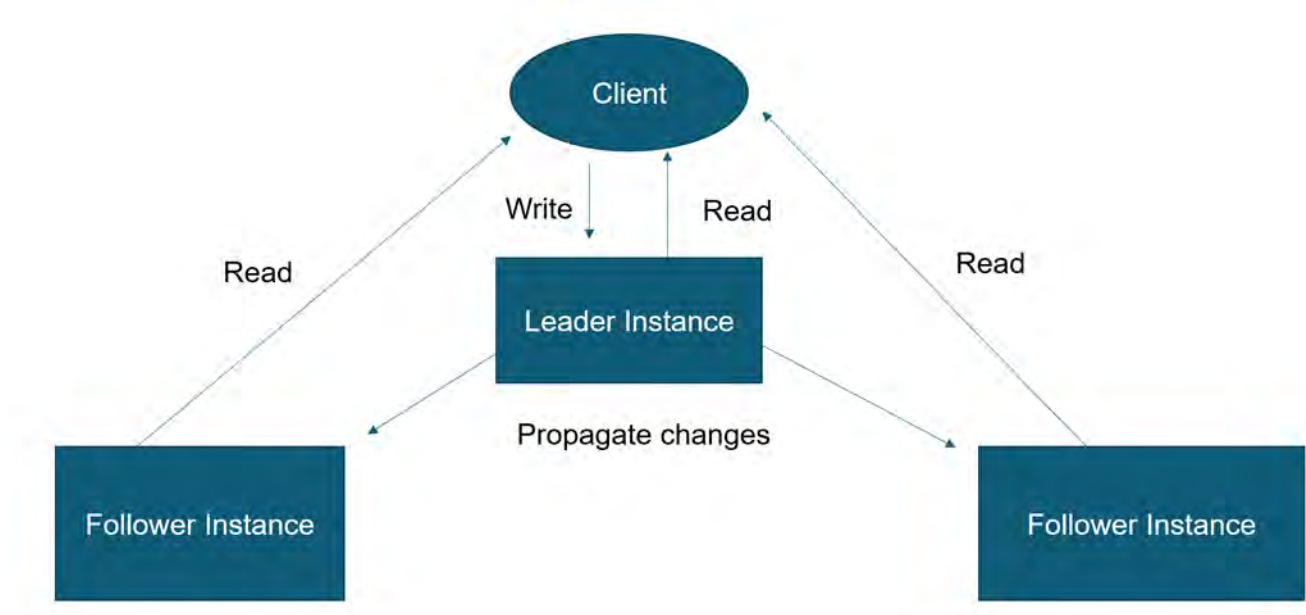
[uiuc/slooo\\_mongo/blob/redis/old\\_stuff/scripts/redis/sentinel\\_follower\\_template.conf](https://github.com/xlabuiuc/slooo_mongo/blob/redis/old_stuff/scripts/redis/sentinel_follower_template.conf) respectively.

This meant that in each experiment, 2 Sentinels (both replica instances) had to trigger and complete the failover to elect a new leader. The experiments didn't track the status of which instance was the leader at any given time, so no information was collected on whether a failover occurred or not; only data concerning overall performance in throughput and latency was collected.

During each experiment, Redis instances were assigned to the same CPU on their respective VM's using `numactl [10, 11]` to reduce the variation in results since there wouldn't be a case where during one trial an instance would be ran off one CPU but during the next trial the VM decided to assign Redis to a different CPU. On the "client VM", benchmarking tools were similarly constrained to run on one specific CPU. More specifically, `numactl --physcpubind=0 redis-server <configuration options for redis-server>` was ran on each VM that would run a Redis instance **only** on the physical CPU with ID 0.

The last VM served as a "client". It would run the benchmarks and in order 1. Start the other virtual machines on Azure, 2. Start a Redis instance on each VM using configuration files that were transferred onto the VM's using `scp` such that 1 VM was a leader, 2 VM's were replicas, 3. Inject a fail-slow fault or inject no fault into a replica instance. 4. Benchmark the leader instance using either `redis-`

benchmark or memtier\_benchmark with the benchmark tool being ran on the “client” VM.



**Figure 1:** Configuration of all 4 VM’s used in each experiment. The “client” VM ran the benchmarks on the leader instance. The leader instance had 2 replica instances. The “client” VM can theoretically read data from any of the instances but can only apply changes to the leader instance.

Two benchmarking tools (both developed by Redis Labs) were utilized: redis-benchmark and memtier\_benchmark. One difference in the tools is that redis-benchmark outputs statistics on a wide range of operations i.e it will output overall latencies and throughputs **per operation type** while memtier\_benchmark will output overall latency and throughput. redis-benchmark by default simulates 50 clients sending 100000 total requests (these defaults were kept) [12]. In addition, by default redis-benchmark runs against a single key but this shouldn’t impact the results too much as Redis is an in-memory store [12]. **Pipelining** allows a Redis instance to send multiple commands which can drastically improve performance. The default pipeline depth that redis-benchmark uses is 1 (i.e no pipelining) [12] which may not reflect real world deployments, but we are aiming to test fail-slow faults, not the maximum performance a Redis instance can achieve on the hardware available to us.

memtier\_benchmark almost exclusively uses to Redis operations to simulate workloads against a Redis instance: SET and GET, which is more accurate to real life workloads [13, 14]. It also allows for more options to configure compared to redis-benchmark, but again all the default options are kept save for the port number and IP address of the leader Redis instance.

Results from redis-benchmark were obtained over 1 trial of all the below fail-slow faults plus a run where no fault was injected, and results obtained from memtier\_benchmark were obtained from taking the median results (latencies, throughput) from 5 trials of all the below faults and a run with no fault injected.

Fault	Description
CPU (slow)	Use cgroup to limit each database process to utilize only a fraction of the CPU.
CPU (contention)	Run a contending program on the same CPU as the RSM.
Disk (slow)	Use cgroup to limit disk I/O bandwidth available for the database process.
Disk (contention)	Run a contending program that writes heavily on the shared disk.
Memory (contention)	Use cgroup to set the maximum amount of user memory for the RSM process.
Network (slow)	Add a delay of 400 ms to the network interface using tc.

**Table 2.** List of fail-slow faults tested and their descriptions which were taken from the DepFast paper previously published by the team [4]. cgroup (control groups) is a Linux kernel feature and allows operating system processes to grouped and organized into hierarchies whose resource usage can then be monitored and controlled [15, 16]. tc is a Linux utility to show and manipulate traffic control settings [17].

All of the code used to run the experiments and graph the results can be found here: [https://github.com/xlab-uiuc/slooo\\_mongo/tree/redis/old\\_stuff/scripts/redis](https://github.com/xlab-uiuc/slooo_mongo/tree/redis/old_stuff/scripts/redis).

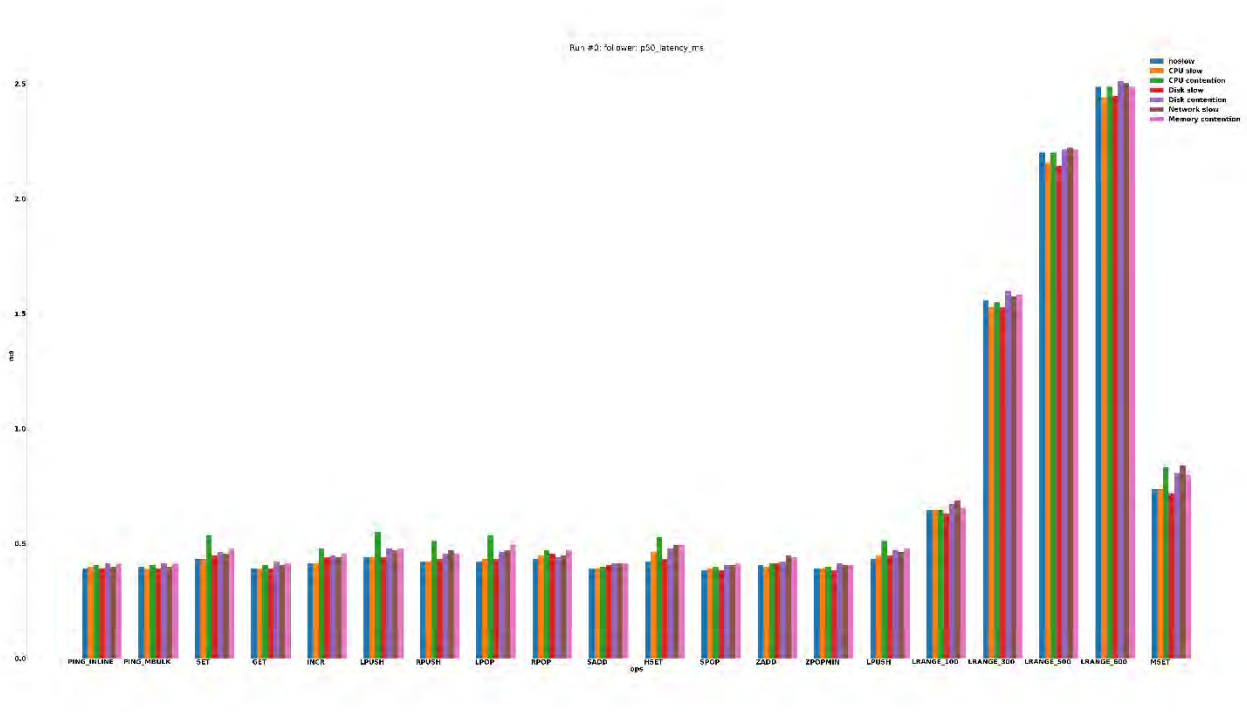


### III. Results

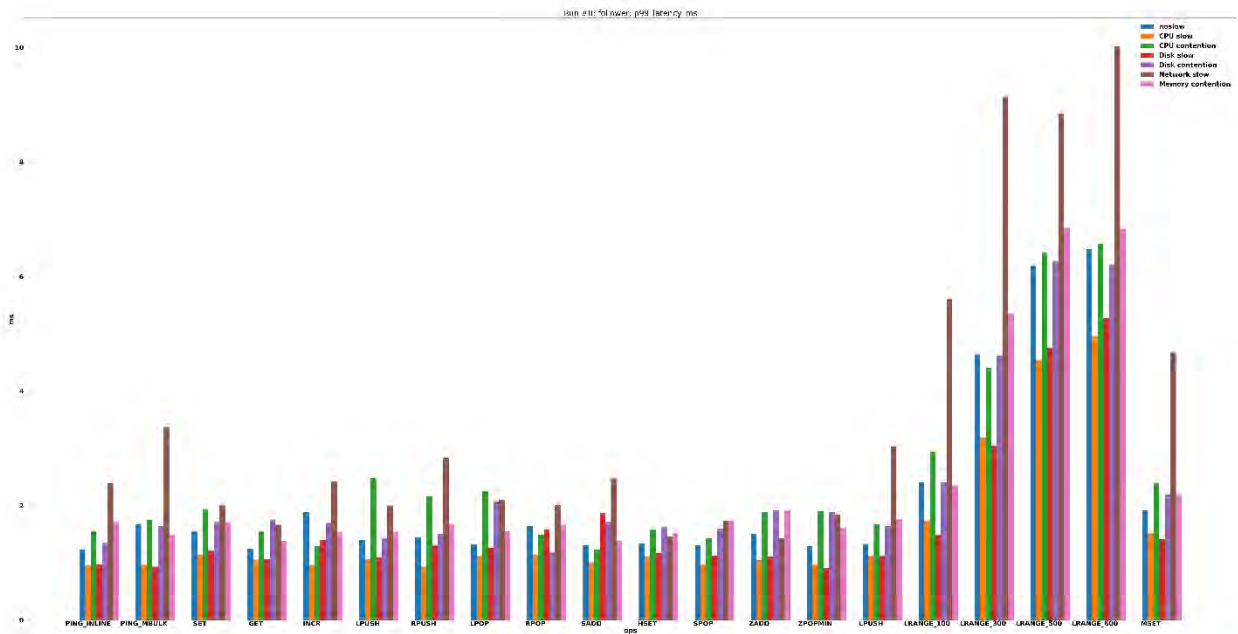
Results were given in terms of either average or percentile latencies (i.e for some latency  $L$  and percentage  $x$ ,  $x\%$  of all queries had latencies less than or equal to  $L$ ).

Overall, results are mixed as redis-benchmark and memtier\_benchmark gave contradictory results. redis-benchmark showed that certain fail-slow faults led to a severe decrease in system performance in answering certain query types. The most noticeable increases in system latency were due to 3 fail-slow faults: **CPU (contention), memory (contention), network (slowdown)**. 50<sup>th</sup> percentile latencies (i.e., median latencies) were uniform as the latencies measured from experiments where fail-slow faults were injected only slightly deviated from the latency the system experienced without any induced fail-slow faults. However, 99<sup>th</sup> percentile latencies showed that certain fail-slow faults (network slowdown, memory contention in particular) caused a considerable increase in system latency, up to roughly a 100% increase in latency for certain types of Redis operations as compared to injecting no faults.

This isn't too surprising: Redis is an in-memory database and is also single-threaded meaning it cannot achieve true parallelization, so running a contending program on the CPU may slow it down if the program is assigned to the same CPU core that redis-server is running on. This is not too surprising, as the Redis documentation recognizes that network bandwidth and latency are naturally very important as instances must communicate with each other over the network to do synchronizations and to answer client queries [12]. The documentation recognizes that Redis favors CPUs with fast single-threaded performance with large caches, though it downplays the importance of RAM and memory bandwidth [12]. It says that Redis runs slower on a VM, but the experiments conducted should still show faults that aren't dependent on the hardware that Redis runs on but with how Redis is designed.



**Figure 2:** 50th percentile/median latencies in milliseconds of each operation type profiled by redis-benchmark after 1 trial of each fail-slow fault and no-fault. The latencies are uniform, though for certain operations, CPU (contention) (green bars) causes slightly higher latencies compared to the other faults injected and no-fault injected.



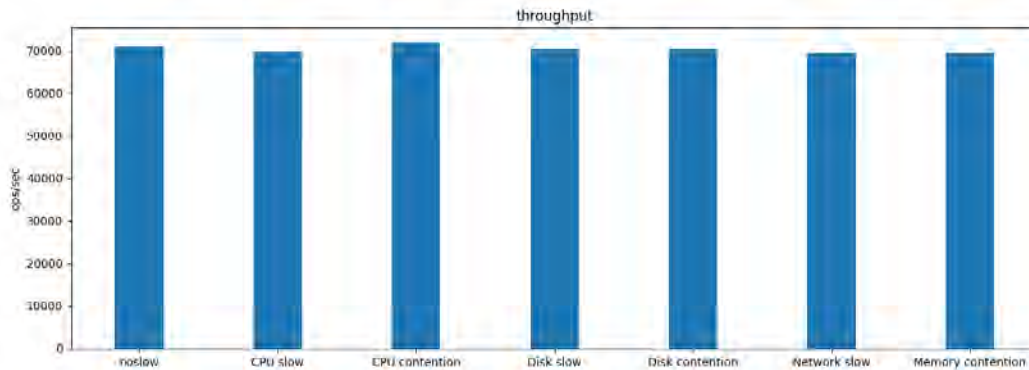
**Figure 3:** 99th percentile latencies in milliseconds of each operation type profiled by redis-benchmark after 1 trial of each fail-slow fault and no-fault. Injecting the **Network (slow)** (brown bars), **CPU (contention)** (green bars), and **Memory (contention)** (pink bars) caused noticeably high latencies as compared to other faults.

<b>Redis Operation</b>	<b>Description</b>
SET key value	Time Complexity: O(1) Set key to hold string value.
GET key	Time Complexity: O(1) Get value of key, returns nil if key doesn't exist.
INCR key	Time Complexity: O(1) Increments number stored key, returns error if key of wrong type. If key doesn't exist, set to 0.
LPUSH key element [element ...]	Time Complexity: O(N) for N elements Inserts all the elements at the head of the list at key or makes a new list if key doesn't exist.
RPUSH key element [element ...]	Time Complexity: O(N) for N elements Inserts all the elements at the tail of the list at key or makes a new list if key doesn't exist.
LPOP key [count]	Time Complexity: O(N) Removes and returns first count elements in key's list.
RPOP key [count]	Time Complexity: O(N) Removes and returns last count elements in key's list.
SADD key member [member ...]	Time Complexity: O(N) Add specified members to the set at key.
HSET key field value [field value ...]	Time Complexity: O(N) for N fields. Sets field in the hash stored at key to value.
SPOP key [count]	Time Complexity: O(N) Removes and returns one or more random members from set at key.
ZADD score member [score member ...]	Time Complexity: O(log(N))

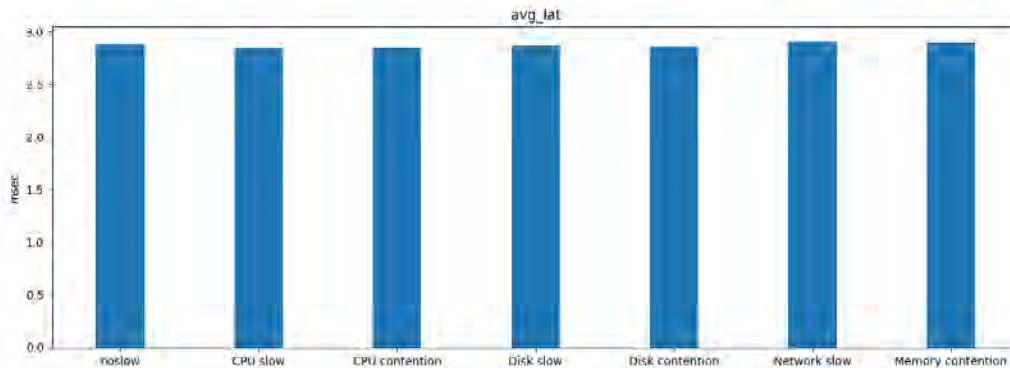
	Adds all specified members with specified scores to the sorted set at key.
ZPOPMIN key [count]	Time Complexity: $O(\log(N)*M)$ where N is the set size and M is the number of popped elements. Removes and returns up to count members with the lowest scores from the sorted set at key.
LRANGE key start stop	Time Complexity: $O(S+N)$ where S is the offset distance and N is the number of elements in the specified range. Returns specified elements of the list at key. In the figures above some bars are labelled with LRANGE_n: n is the number of elements removed by redis-benchmark at some key.
MSET key value [key value ...]	Time Complexity: $O(N)$ Sets given keys to their respective values.

**Table 3:** List of Redis operations tested by redis-benchmark and a brief synopsis of each one. LPUSH is listed twice on the above figures; the LPUSH bars on the right are results from the LPUSH operations used to benchmark LRANGE [18].

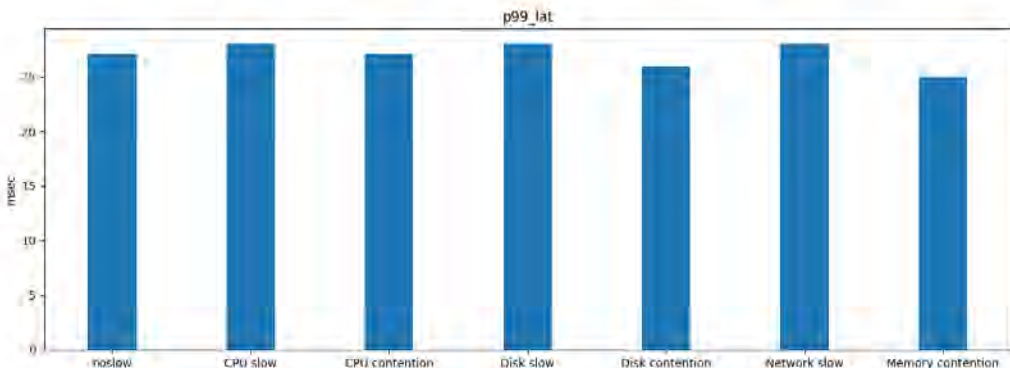
mentier\_benchmark tells a different story. The results show that injecting fail-slow faults into the system caused almost negligible differences in performance; there were no significant increases in latencies, no significant decreases in throughput, even when Sentinels were not used.



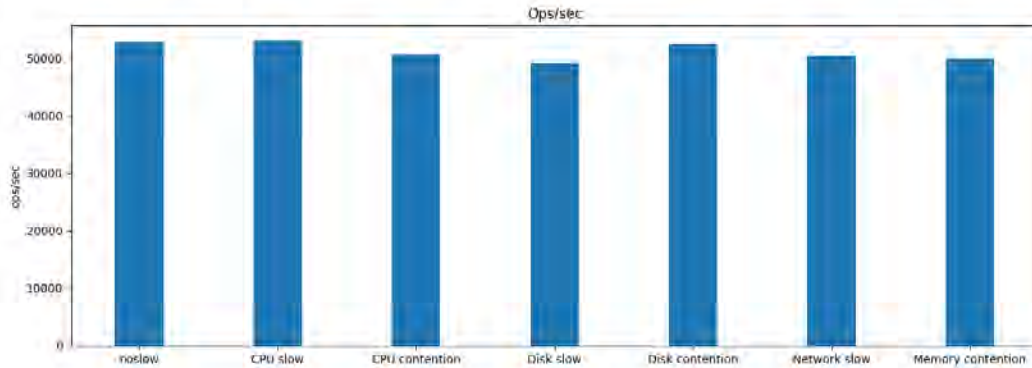
**Figure 4:** memtier\_benchmark throughput results from injecting fail-slow faults without using Redis Sentinels. 5 trials of each type of experiment (noslow, CPU slow, etc.) were ran and the median throughputs from those trials were selected as the results displayed.



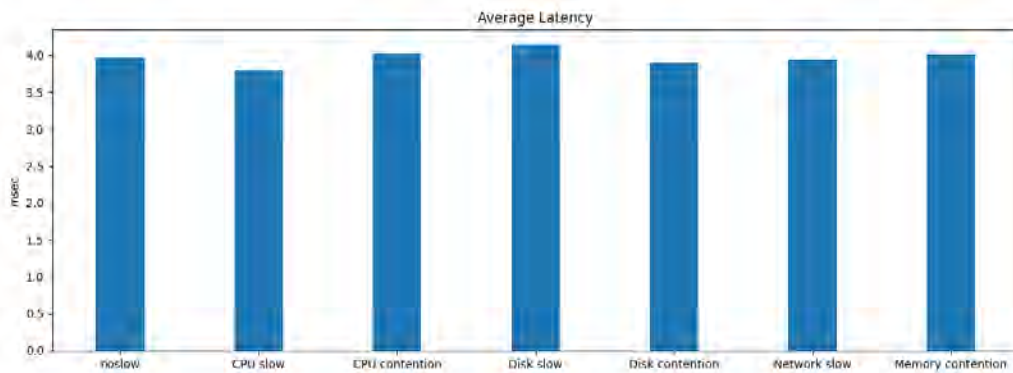
**Figure 5:** memtier\_benchmark average latency results from injecting fail-slow faults without using Redis Sentinels. 5 trials of each type of experiment (noslow, CPU slow, etc.) were ran and the median throughputs from those trials were selected as the results displayed.



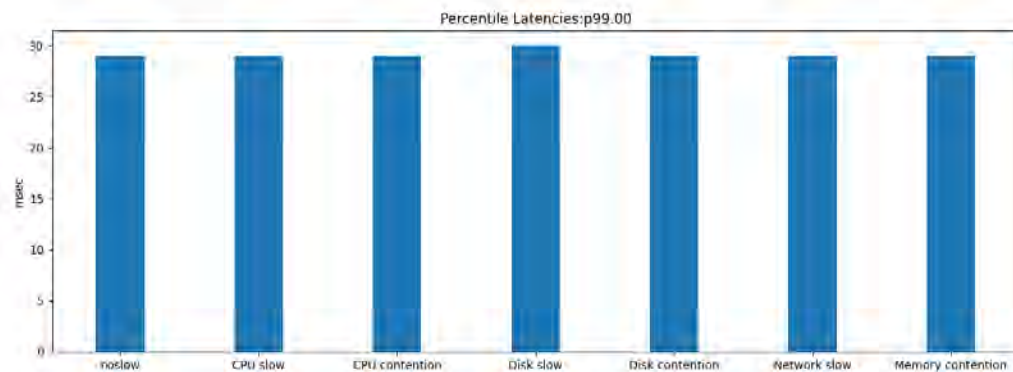
**Figure 6:** memtier\_benchmark 99th percentile latency results from injecting fail-slow faults without using Redis Sentinels. 5 trials of each type of experiment (noslow, CPU slow, etc.) were ran and the median throughputs from those trials were selected as the results displayed.



**Figure 7:** memtier\_benchmark throughput results from injecting fail-slow faults while using Redis Sentinels. 5 trials of each type of experiment (noslow, CPU slow, etc.) were ran and the median throughputs from those trials were selected as the results displayed.



**Figure 8:** memtier\_benchmark average latency results from injecting fail-slow faults while using Redis Sentinels. 5 trials of each type of experiment (noslow, CPU slow, etc.) were ran and the median throughputs from those trials were selected as the results displayed.



**Figure 9:** memtier\_benchmark 99th percentile latency results from injecting fail-slow faults with Redis Sentinels. 5 trials of each type of experiment (noslow, CPU slow, etc.) were ran and the median throughputs from those trials were selected as the results displayed.

The results of memtier\_benchmark using Redis Sentinels (Figures 7-9) may be incorrect as during the experiments ran, some important scripts that injected faults (specifically, scripts that simulated Disk contention CPU contention) and may not have been copied to the VM's. However, even if these results are wrong, the more important results are those obtained without using Sentinels which shows that injecting a fail-slow fault has no effect on the leader instance.

Even just looking at the performance of the SET and GET operations that redis-benchmark ran, the results still show discrepancies in performance when certain fail-slow faults are injected (Network slow, Memory contention) while memtier\_benchmark displays uniform performance no matter what fault (if any) was injected. Thus, further investigation is needed to reconcile these differences.

#### IV. Discussion

There still needs to be more investigation to reconcile the differences between the results obtained from redis-benchmark and memtier\_benchmark. However, these results confirm that Redis has potential issues with dealing with certain classes of fail-slow faults in certain, yet common, scenarios.

The results and methods of this project were derived from previous benchmarks and experiments ran against other popular RSM systems such as MongoDB and TiDB and reinforce a common theme with these systems: a lack of fail-slow fault tolerance. **DepFast**, a framework currently being developed by the team, is aimed at making it easier to develop new systems that are fail-slow fault tolerant. In the future, it is likely that more databases will be tested by my team to see which systems are deficient, which fare better in dealing with faults, etc., and to see which lessons can be applied to the **DepFast** project. During the duration of this project, it became clear that our work was a bit novel so there weren't any open-source tools out there that had, by default, the ability to spin up instances, inject faults, and run benchmark all in one suite.

During this project, it became clear that a new, comprehensive tool (rather than a collection of ad-hoc scripts) was needed to make it easier to run fail-slow experiments against database systems and thus the **slooo** framework was born and is currently under development. Hopefully in the future our tools can be open-sourced so that developers everywhere can improve and benchmark their systems easily for fail-slow faults.



## **V. Acknowledgments**

I'd like to give special thanks the following people for supporting me during my research project:

- Dr. Shuai Mu (Stony Brook) for generously volunteering to be my mentor.
- Dr. Tianyin Xu (UIUC) for helping run the project.
- Varshith B. (UIUC) for assisting in benchmarking.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## VI. References

1. Ongaro, D., & Ousterhout, J. (2014). In Search of an Understandable Consensus Algorithm. In *2014 USENIX Annual Technical Conference (USENIX ATC 14)* (pp. 305-319). Philadelphia: USENIX Association.
2. Haryadi S. Gunawi, Riza O. Suminto, Russell Sears, Casey Golliver, Swaminathan Sundararaman, Xing Lin, Tim Emami, Weiguang Sheng, Nematollah Bidokhti, Caitie McCaffrey, Deepthi Srinivasan, Biswaranjan Panda, Andrew Baptist, Gary Grider, Parks M. Fields, Kevin Harms, Robert B. Ross, Andree Jacobson, Robert Ricci, Kirk Webb, Peter Alvaro, H. B. Runesha, Mingzhe Hao, and Huaicheng Li. 2018. Fail-Slow at Scale: Evidence of Hardware Performance Faults in Large Production Systems. *ACM Trans. Storage* 14, 3, Article 23 (October 2018), 26 pages.
3. Arpaci-Dusseau, R. H., and Arpaci-Dusseau, A. C. Fail-Stutter Fault Tolerance. In *Proceedings of the 8th Workshop on Hot Topics in Operating Systems (HotOS'01)* (May 2001)
4. Yoo, A., Wang, Y., Sinha, R., Mu, S., & Xu, T. (2021). Fail-slow fault tolerance needs programming support. In *Proceedings of the 18th Workshop on Hot Topics in Operating Systems (HotOS-XVIII)*.
5. Redis Labs. (n.d.). Redis. <https://redis.io/>.
6. Instagram Engineering. (2016, December 11). Storing hundreds of millions of simple key-value pairs in Redis. Medium. <https://instagram-engineering.com/storing-hundreds-of-millions-of-simple-key-value-pairs-in-redis-1091ae80f74c>.
7. Redis Labs. (n.d.). Redis persistence. Redis. <https://redis.io/topics/persistence>.
8. Redis Labs. (n.d.). Replication. Redis. <https://redis.io/topics/replication>.
9. Redis Labs. (n.d.). Redis Sentinel Documentation. Redis. <https://redis.io/topics/sentinel>.

10. numactl. numactl(8) - Linux man page. (n.d.). <https://linux.die.net/man/8/numactl>.
11. Numactl. (n.d.). Numactl/Numactl: NUMA support for Linux. GitHub. <https://github.com/numactl/numactl>.
12. Redis Labs. (n.d.). How fast is redis? Redis. <https://redis.io/topics/benchmarks>.
13. Redis Labs. (2020, June 9). Memtier\_Benchmark: A High-Throughput benchmarking tool for Redis & memcached. Redis Labs. [https://redislabs.com/blog/memtier\\_benchmark-a-high-throughput-benchmarking-tool-for-redis-memcached/](https://redislabs.com/blog/memtier_benchmark-a-high-throughput-benchmarking-tool-for-redis-memcached/).
14. RedisLabs. (2020, January 9). RedisLabs/memtier\_benchmark: NoSQL Redis and MEMCACHE traffic generation and benchmarking tool. GitHub. [https://github.com/RedisLabs/memtier\\_benchmark](https://github.com/RedisLabs/memtier_benchmark).
15. cgroups(7) - Linux manual page. (n.d.). <https://man7.org/linux/man-pages/man7/cgroups.7.html>.
16. Linux Kernel Organization, Inc. (n.d.). Control groups. Control Groups - The Linux Kernel documentation. <https://www.kernel.org/doc/html/latest/admin-guide/cgroup-v1/cgroups.html>.
17. tc(8) - Linux manual page. (n.d.). <https://man7.org/linux/man-pages/man8/tc.8.html>.
18. Redis Labs. (n.d.). Command reference. Redis. <https://redis.io/commands>.

# Validating BRR with Polarized-beam data

Sergio Ruiz

School of Physics, Georgia Institute of Technology, Atlanta, GA 30332

Dr. Gustavo Nobre, Dr. David Brown

National Nuclear Data Center, Brookhaven National Laboratory, Upton, NY 11973

August 11, 2021

## I Abstract

A neutron resonance is an event where a slow-moving neutron collides with a target nucleus, gets absorbed by the target nucleus and peaks occur in the cross-sections of the interaction as a function of the energy. The Bayesian Resonance Reclassifier (BRR) developed at Brookhaven National Laboratory reclassifies mis-assigned spins for nucleons after a collision with a neutron. Validating BRR is an important task to ensure its accuracy and applicability to these experiments. To validate BRR, data on a series of various different isotopes had to be gathered from the EXFOR and ATLAS databases. Later, a close look needed to be taken at the data to make sure it came from polarized-beam experiments. This data was then imported into a Python script to change the values of the spins in the data. This data served as test data for BRR to make sure it can identify the spins that were changed. In the process, a big index of all of this data was created tracking key aspects of their respective experiments. It was also discovered that the EXFOR database was missing data on key experiments. When this new data was input into BRR, it was able to correctly detect that we were feeding it increasingly jumbled data. The ATLAS and EXFOR databases should be updated with the indexed data as the current data is either outdated or it is not in those databases for better results. This summer was a great opportunity to develop my coding skills as I learned to use Linux, Git, and interact with computer terminals a lot better, put a lot of my Python and Object-Oriented Programming knowledge to the test, and I am now familiar with a lot of nuclear physics and nuclear resonance terminology.

## II Introduction

When a neutron collides with a target nucleus leading to the formation of a new excited compound nucleus many nuclear processes can occur such as nuclear fission, neutron or gamma decay, etc. During the formation of this new compound nucleus however, because of the neutron's new energy, the new nucleus will now have a new different energy levels scheme. In the case of where a slow-moving neutron collides with a target nucleus a new intermediate state may arise called a neutron resonance (pictured in Figure 1). A neutron resonance is a short-lived state where the energy of the incoming neutron plus the neutron separation energy is close to one of the energies of the new energy level scheme of the excited compound nucleus and the matching of the energies cause peaks in the probability of certain nuclear processes to occur. A particular characteristic of interest of these resonance states is their spin-group. Spin-groups in Quantum Mechanics are described by three quantum numbers: S (sum of ground state spin and intrinsic angular momentum), L (orbital angular momentum) and J (total angular momentum). The L values of the resonance states can be found by observing the peaks in the cross-section as a function of the energy as shown in the Figure 2. The rest of the quantum numbers are dependent on L, the ground-state spin of the target nucleus, and the spin of the projectile which in this case is a neutron whose magnitude of its spin is always  $\frac{1}{2}$ . Any possible combination of these spins L,S and J is called a spin-group. However, measuring the spin-group of these resonance states is not always possible because they are very short-lived states so a lot of simplifying assumptions are made that are prone to misclassifying these spin-groups. The Bayesian Resonance Reclassifier (BRR) developed at Brookhaven National Laboratory (BNL) was developed to address this issue.

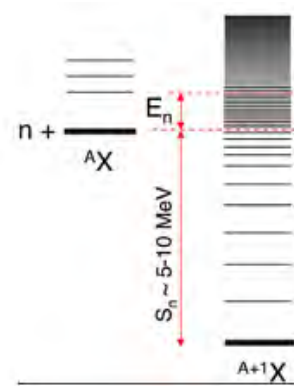


Figure 1: An example of a new energy scheme of a compound nucleus in a resonance state.

BRR is a machine learning algorithm that goes over resonance data to find misclassified spin-groups in resonances and tries to find the correct spin-group for that resonance. To show that BRR can be applied for these purposes, a lot of data is needed to feed it into the algorithm. Polarize-beam data in particular

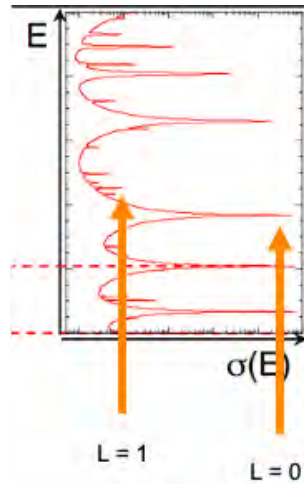


Figure 2: Plot of cross-section as a function of energy. The small peaks are resonances corresponding to  $L = 1$  whereas the tall spikes correspond to  $L = 0$ , as labeled.

was chosen for this purpose due to the fact that data from experiments with this kind of setup has less uncertainty and reduces a lot of simplifying assumptions that would otherwise go into finding the spin-groups of these resonance states. The aim of this project is to validate the BRR algorithm as one that can find any misclassified resonances by training and testing it with polarized-beam data.

### III Methods

To validate an algorithm like BRR, whose goals are to be applicable to all nuclei, a high volume of data needs to be gathered to be separated into training and testing data. The gathering of data started by looking up all the papers referenced in *<sup>1</sup>Parity Violation in compound nuclei: experimental methods and recent results* [Bowman, Penttilä and Sharapov] that contain resonance data on a series of different nuclei and isotopes. All of these references were then kept track of and organized in an Excel spreadsheet that contains different information on the papers such as what atom they studied and the papers' URLs and DOIs. After having created the spreadsheet we needed to verify if the data contained in the papers existed in the EXFOR and/or ATLAS databases and record their entry information in these databases in the spreadsheet. The spreadsheet was then later organized by creating sections for each individual nucleus and creating rows with information on their entries in the databases based on what data was contained in the papers such as resonance energy, average level spacing,  $L$  and  $J$  spin values, etc. After having finalized this stage of the spreadsheet it was time to pick an nucleus to start the project off. We did this simply by choosing the nuclide that contained the most data because it would have been the best to test the algorithm on. Initially we had chosen Rh-103 but upon further inspection of the experimental setup of this paper we found out that the Rh-103 did not

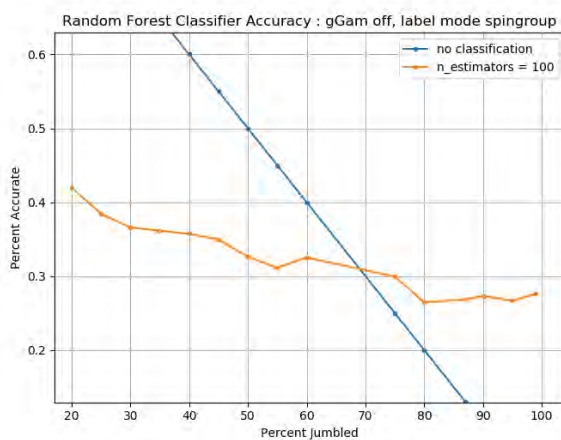
come from a polarized-beam experiment. Also, we found that the evaluated file for 103Rh, which is a source of average resonance properties needed to create a training sequence, has serious problems introduced in its latest evaluation. On the bright side, upon finding this we communicated the issues with the evaluation authors so the resonance file could be fixed. Because of this setback we decided it would be best if we go through all the papers and pay close attention to the experimental setup to make sure the data we were going to use came from polarized-beam experiments. Upon completion of this last step we decided on In-115 because it had the most polarized-beam data and thus was the best candidate for the validation process. The data used for In-115 can be found in <sup>2</sup>*Parity violation in neutron resonances in <sup>115</sup>In*[Stephenson, S].

Target	Reference	DOI	URL	Polarized	EXFOR Entry
Br-81	[67]	<a href="https://doi.org/10.1080/18811248.1981">https://doi.org/10.1080/18811248.1981</a>	<a href="https://www.tandfonline.com/doi/pdf/10.1080/18811248.1981">https://www.tandfonline.com/doi/pdf/10.1080/18811248.1981</a>	NO	21294
Br-81	[67]	<a href="https://doi.org/10.1080/18811248.1981">https://doi.org/10.1080/18811248.1981</a>	<a href="https://www.tandfonline.com/doi/pdf/10.1080/18811248.1981">https://www.tandfonline.com/doi/pdf/10.1080/18811248.1981</a>	NO	21294
Br-81	[67]	<a href="https://doi.org/10.1080/18811248.1981">https://doi.org/10.1080/18811248.1981</a>	<a href="https://www.tandfonline.com/doi/pdf/10.1080/18811248.1981">https://www.tandfonline.com/doi/pdf/10.1080/18811248.1981</a>	NO	21294
Br-81	[67]	<a href="https://doi.org/10.1080/18811248.1981">https://doi.org/10.1080/18811248.1981</a>	<a href="https://www.tandfonline.com/doi/pdf/10.1080/18811248.1981">https://www.tandfonline.com/doi/pdf/10.1080/18811248.1981</a>	NO	21294
Nb-93	[125]	10.1103/PhysRevC.59.1131	<a href="https://link.aps.org/doi/10.1103/PhysRevC.59.1131">https://link.aps.org/doi/10.1103/PhysRevC.59.1131</a>	YES	13780
Nb-93	[125]	10.1103/PhysRevC.59.1131	<a href="https://link.aps.org/doi/10.1103/PhysRevC.59.1131">https://link.aps.org/doi/10.1103/PhysRevC.59.1131</a>	YES	13780
Nb-93	[125]	10.1103/PhysRevC.59.1131	<a href="https://link.aps.org/doi/10.1103/PhysRevC.59.1131">https://link.aps.org/doi/10.1103/PhysRevC.59.1131</a>	YES	13780
Nb-93	[125]	10.1103/PhysRevC.59.1131	<a href="https://link.aps.org/doi/10.1103/PhysRevC.59.1131">https://link.aps.org/doi/10.1103/PhysRevC.59.1131</a>	YES	13780
Nb-93	[125]	10.1103/PhysRevC.59.1131	<a href="https://link.aps.org/doi/10.1103/PhysRevC.59.1131">https://link.aps.org/doi/10.1103/PhysRevC.59.1131</a>	YES	13780
Nb-93	[125]	10.1103/PhysRevC.59.1131	<a href="https://link.aps.org/doi/10.1103/PhysRevC.59.1131">https://link.aps.org/doi/10.1103/PhysRevC.59.1131</a>	YES	13780
Rh-103	[132]	10.1103/PhysRevC.60.045502	<a href="https://link.aps.org/doi/10.1103/PhysRevC.60.045502">https://link.aps.org/doi/10.1103/PhysRevC.60.045502</a>	NO	14285
Rh-103	[132]	10.1103/PhysRevC.60.045502	<a href="https://link.aps.org/doi/10.1103/PhysRevC.60.045502">https://link.aps.org/doi/10.1103/PhysRevC.60.045502</a>	NO	14285
Rh-103	[132]	10.1103/PhysRevC.60.045502	<a href="https://link.aps.org/doi/10.1103/PhysRevC.60.045502">https://link.aps.org/doi/10.1103/PhysRevC.60.045502</a>	NO	14285
Rh-103	[132]	10.1103/PhysRevC.60.045502	<a href="https://link.aps.org/doi/10.1103/PhysRevC.60.045502">https://link.aps.org/doi/10.1103/PhysRevC.60.045502</a>	NO	14285
Rh-103	[132]	10.1103/PhysRevC.60.045502	<a href="https://link.aps.org/doi/10.1103/PhysRevC.60.045502">https://link.aps.org/doi/10.1103/PhysRevC.60.045502</a>	NO	14285

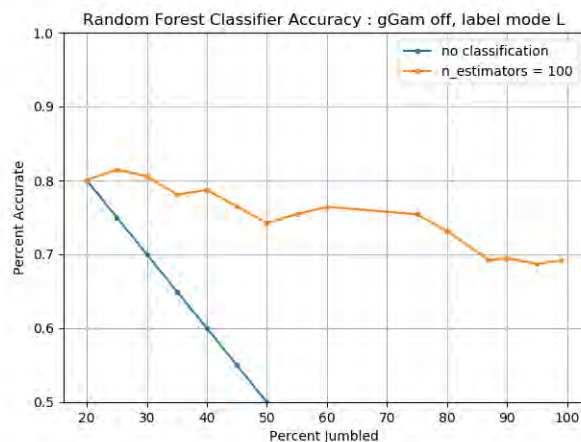
Figure 3: A small look at the spreadsheet where all of the atoms, their papers, and more information were kept track of.

The following step was to extract the In-115 data from either EXFOR or the ATLAS database using coding tools already developed by the team. Once the data was extracted from the ATLAS, the file was fed into a function that would randomly select n resonances to jumble.<sup>1</sup> We ran this code varying the amount of resonances n to mix up, creating several files with test data that have different amounts of mixed up resonances to see if BRR can find them. However, jumbling the data in the test files is not as easy as assigning a new random number to an L or a J value; there are certain quantum mechanical rules that these values follow that had to be taken into account when mixing up the data. Before testing BRR on these files we needed training data. To create a training data file, we needed to find all possible spin-groups for Indium-115 and from there we were able to execute code that will generate a sequence of synthetic data. Once this new file was created, we needed to create several files of training data by jumbling the synthetic data to different degrees. The final step was then to train BRR on these new jumbled files and test it on some of the mixed up files of the real experimental data.

<sup>1</sup>The term "jumble" means to intentionally mess around and mix up some values in a dataset.



(4a) Accuracy when BRR classifies by spin-groups.



(4b) Accuracy when BRR classifies by L values.

Figure 4: BRR's Accuracy at training time with different levels of jumbledness in the training data.

## IV Results

There is now a detailed compilation of all the papers with resonance data with links to the papers of different atoms, their data in EXFOR and ATLAS sorted by the type of measurements that were taken, the number of data entries in each paper to see what atoms would be best to test BRR on next and a column indicating whether the data came from polarized-beam experiments. This was the first time that the BRR method was applied and tested from starting at generating synthetic data to spin-group prediction on real resonance data. Furthermore, the method for artificially changing resonance spins for subsequent detection and validation was fully implemented. In the process of looking for data from the papers in the EXFOR and ATLAS databases I found some issues. Firstly, I discovered that the EXFOR database does not contain necessary data for analysis of several nuclei. Secondly, I also discovered that some of the data in the ATLAS database is either incomplete or it was transcribed to the ATLAS incorrectly by entering incorrect values for certain data entries.

When it comes to running BRR, the classifications can be done using either of two separate sets of labels: either reclassify the resonances by reclassifying the L values or by spin-group. We trained BRR on a series of training data files that were increasingly jumbled and plotted the accuracy as seen in Figure 4. In this figure it can be seen that while training and classifying by spin-groups BRR did not perform particularly well as its best accuracy was at the minimal jumbledness and was only around 41% whereas training and classifying by L values BRR did about as expected with a maximum accuracy roughly at 81%<sup>2</sup>. The accuracy when labeling by L values was about what we expected whereas the accuracy of labeling by spin-groups was interesting

<sup>2</sup>The blue lines represent the accuracy of randomly choosing a resonance and saying it was misclassified.



because randomly choosing misassigned spins did better than BRR on a big interval. This could be due to the fact that there are more spin-groups than L values and thus more choices and more confusion for the algorithm. It is curious however that this peak does not occur at the point of minimal jumbleness. Finally, we made a plot (Figure 5) where we train BRR on files with different levels of jumbleness and validate it on two files, the original In-115 data without jumbling and a file with 15 jumbled resonances in the test data, to see what fraction of resonances it changed. Very interestingly, we found out and as can be seen in Figure 5 by the upward sloping lines and direction of the dots, that BRR recognized that we kept feeding it increasingly jumbled data and thus kept changing more and more resonances. It can also be observed that BRR reclassified more resonances on the file containing actual misclassified spin-groups than in the real non-jumbled file. This is a good thing because BRR reclassified more resonances where there was actual misclassified resonances than in the other file containing no jumbled resonances.

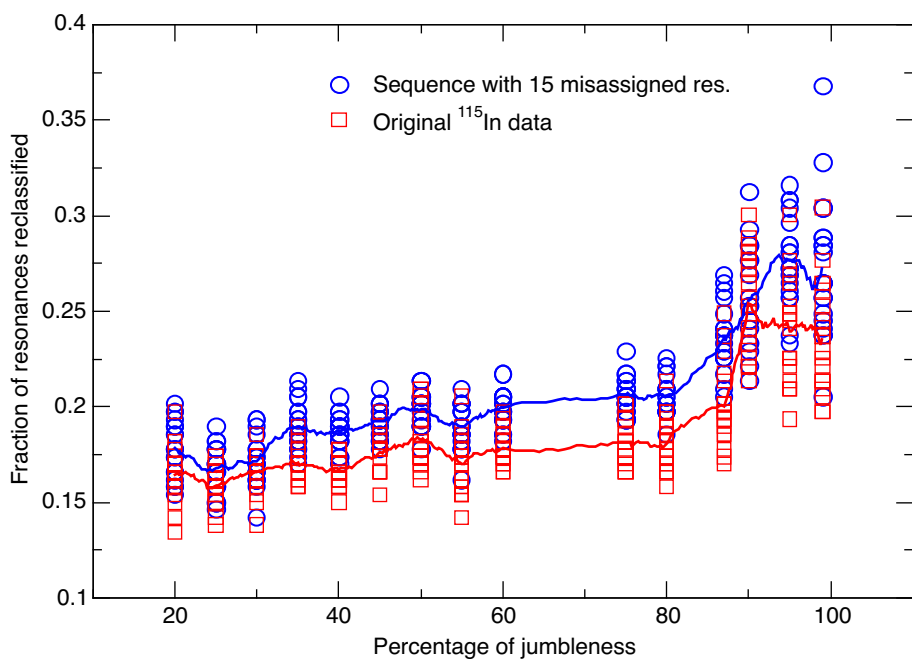


Figure 5: Plots of the fraction of resonances changed as a function of the jumbleness of the training data.

## V Conclusion

Even though BRR was not able to fully and uniquely identify the resonances with modified spin in these preliminary calculations, we were able, for the first time, to apply the BRR method fully from generation of synthetic data to spin group prediction. BRR was also able to identify the increasing levels of jumbledness in the training data, which is a very good sign. Furthermore, the method for artificially changing resonance spins for subsequent detection and validation was fully implemented. This now allows for a detailed research to investigate other possible improvements, optimizations, and validations of BRR. This work opens up many paths for improving a Machine-Learning algorithm to provide reliable and reproducible evaluations of neutron resonances.

## VI Acknowledgements

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). I would also like to thank my mentors Gustavo Nobre and David Brown for helping through my project, always answering my questions and teaching me many useful skills. Finally, I would also like to thank my fellow interns Mary Fucci, Rosie Crawford, and Sergey Scoville for being great coworkers and friends.

## References

- [1] Mitchell, G., Bowman, J., Penttilä, S., Sharapov, É. (2001). Parity violation in compound nuclei: experimental methods and recent results. *Physics Reports*, 354, 157-241.
- [2] Stephenson, S. L. and Bowman, J. D. and Corvi, F. and Crawford, B. E. and Delheij, P. P. J. and Frankle, C. M. and Iinuma, M. and Knudson, J. N. and Lowie, L. Y. and Masaike, A. and Masuda, Y. and Matsuda, Y. and Mitchell, G. E. and Penttilä, S. I. and Postma, H. and Roberson, N. R. and Seestrom, S. J. and Sharapov, E. I. and Shimizu, H. M. and Yen, Y.-F. and Yuan, V. W. and Zanini, L. Parity violation in neutron resonances in In-115. *Physical Review C* 03/06/2000

# Spin reassignment in the Atlas of Neutron Resonances

Sergey Scoville

*Department of Physics, Applied Physics, and Astronomy,  
Rensselaer Polytechnic Institute, Troy, NY, 12180*

Dr. Gustavo Nobre and Dr. David Brown

*Nuclear Science and Technology,  
Brookhaven National Laboratory, Upton, NY 11973*

August 11, 2021

## Contents

<b>1</b>	<b>Introduction (Theory and background)</b>	<b>2</b>
<b>2</b>	<b>Methodology</b>	<b>3</b>
2.1	Training data and Classifiers . . . . .	3
2.2	Atlas of Neutron Resonances Retrieval . . . . .	3
	<b>Table 3: Reference for the headng.cod file</b>	<b>4</b>
<b>3</b>	<b>Results and Discussion</b>	<b>6</b>
<b>4</b>	<b>Conclusion</b>	<b>6</b>
<b>5</b>	<b>Acknowledgments</b>	<b>8</b>

### Abstract

Neutron resonances arise when nuclear states in a compound nuclear system, formed by a neutron and a nucleus, are excited. These resonance sequences are well-documented in such sources as the Atlas of Neutron Resonances and the Reference Input Parameter Library-3. Inconsistencies stemming from difficulties in sorting the resonances into their respective spin groups plague many sources of resonance data. This project aims to build off the prior work on classifying neutron resonance with machine learning methods and feature selection methods. Resonance sequences are dictated by their n-wave group; however, missing labels for total spin and angular momentum cause difficulties with correct sequence assignment. Using data directly from the Atlas of Neutron Resonances, we built a method to determine spin assignments for resonances missing those parameters. Through a careful determination of resonance parameters from available data including neutron, gamma and fission widths, we were able to construct a broad set of features for our algorithms. Data discrepancies encountered with many isotopes were eliminated whenever possible through informed random selection and resonance attribute assignment. While the reassignment of spin groups in well-documented isotopes and elements proved fruitful, difficulties arose with higher atomic numbered and less researched elements and isotopes. We were able to provide a full computational infrastructure that enables all resonance data in the Atlas to be formatted and processed so it can be used by our machine learning algorithm for classification of spin groups, allowing now for systematic and comprehensive investigation. Further work upon this method to collect and use experimental data may prove key in further developing and optimizing the machine learning reassignment of resonances.

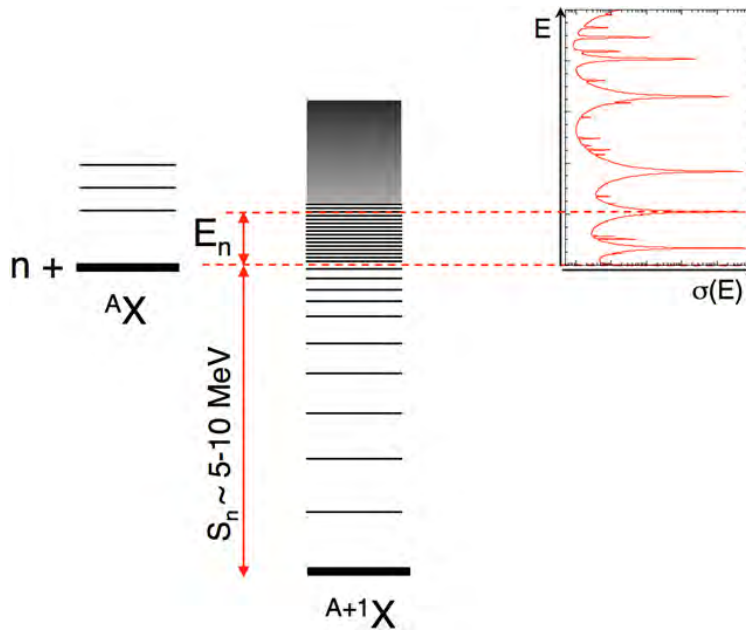


Figure 1: An example of a nuclear cross section. Each line represents an energy resonance. At higher energies, it makes sense to describe the sequence as a continuum rather than the discrete levels that are clearer at lower energies. Fig. taken from Ref. [1].

## 1 Introduction (Theory and background)

Neutron reaction data and scattering in the  $10^{-5}$  eV to 20 MeV neutron energy range are required for the simulation of nuclear systems for applications such as nuclear energy production, stockpile stewardship, non-proliferation, etc. In the energy realm of  $\sim 1$  MeV, only the elastic and capture channels, (as well as fission channels for actinides) are open. In all nuclei, barring the lightest, these channels exhibit strong resonant structure which are identified with the energy levels of the compound nucleus formed by the capture of the neutron into the target state, as shown in Fig. 1.

There are many experimental techniques that help determine the incident neutron orbital angular momentum  $L$  and the total angular momentum  $J$  of each resonance. A few of these techniques include measurements with polarized neutrons,  $\gamma$ -ray multiplicity methods and primary and low-energy  $\gamma$ -rays detected by Ge-Li detectors. These techniques are expensive and rarely used. In optimal cases, determinations of the  $L$  and  $J$  of the resonances is possible with the knowledge of the angular distribution data for scattered neutrons. However for many nuclides, high fidelity angular distribution data are unavailable. The common practice for the evaluator is to determine the  $L$  and  $J$  assignment of the resonances based upon the shape of the observed resonance in cross-section experimental data.

This part of the process is:

1. Very time consuming for the evaluator (who must make assignments for sometimes thousands of resonances)
2. Not reproducible, i.e., dependent on the subjective judgement of the evaluator
3. Does not result in uncertainty estimate on the correct resonance spin assignment
4. Has significant impact on the angular distributions and therefore on the modeling of neutron transport in nuclear systems

The goal of the project is to determine the  $L$  and  $J$  assignments of neutron resonances to correct mis-assigned and missing resonances. The prior approach to this utilized lightweight classifiers from scikit-learn.

These classifiers use a set of features developed from converting resonance data utilizing the elastic width, spacing and prior estimates of the quantum numbers to classify the spin groups of observed resonances, identifying missassignments and missing resonances. Earlier works focused primarily upon the setup and initialization of resonance features.

Currently, there are many different publications of neutron resonance sequences in scholarly articles, papers, and other directories. Some of the most well known examples include the Resonance Input Parameter Library (RIPL-3) [5] and publication of Dr. Said Mughabghab titled *The Atlas of Neutron Resonances* [6]. The Atlas of Neutron Resonances is a world renowned leading source of neutron resonance data, containing resonance data on nearly every element and isotope used in nuclear experiments. This work looks to interpret data from the Atlas and apply this data to our trained classifiers to reassign quantum spin numbers.

## 2 Methodology

### 2.1 Training data and Classifiers

When testing the different classifiers and SHAP metrics, it was crucial to have our data sources mimic what we might expect to see in real world data. To achieve this, we employed the use of generated synthetic resonances that simulated a real neutron resonance sequence [4]. Those generated synthetic resonance sequences were then altered by "jumbling" them. This process reassigned a subset of the resonances randomly to other spingroups.

These jumbled synthetic resonances were used as the input for transfer learning to establish a baseline understanding in our choice of algorithm. Incremental levels of jumbling from 10% to 90% with jumps of 10% as well as 1% and 99% jumbled were created and tested with general trends of higher percentage of jumbled resonances resulting in lower classification accuracy. For the machine learning algorithms, lightweight classifiers from `scikit-learn` [2] were employed which combine ease of usage with powerful algorithms. All of the classifiers operate upon the same base commands which allows for use in a loop to run over all classifiers. The ten different classifiers that we used are listed below:

1. Nearest Neighbors Classifier
2. Linear Support Vector Machine Classifier
3. Radial Basis Function Support Vector Machine Classifier
4. Gaussian Process Classifier
5. Decision Tree Classifier
6. Adaboost Classifier
7. Random Forest Classifier
8. Neural Net Classifier
9. Naive Bayes Classifier
10. Quadratic Discriminant Analysis Classifier

### 2.2 Atlas of Neutron Resonances Retrieval

The Atlas of Neutron Resonances stores resonance data in many different forms. When developing this model we focused upon a set of specific parameters: resonance energy, total compound spin, orbital angular momentum, neutron width, capture width and fission width. The Atlas publication is derived from two different files, namely `txsrp` and `cfmta` files. The former of the two stores the main information of a given element or isotope such as the thermal cross section, bibliographic information and the general resonance

isotopic properties. It also contains the headings for the information stores in the `cmfta` files. The `cmfta` files contain a listings of all the resonances known for a given element as well as observed resonances for all isotopes of said element. `headng.cod` file is simply a file containing all column headers possible in `txsrp` files as well as their descriptors. The ways that these quantities are stored are described below in Table 1

Table 1: Headings in resonance tables in the `txsrp` files for the corresponding `cmfta` files (From the `headng.cod` file)

Dict. ID	Atlas code	# cols.	Longform code	LaTeX	Code explanation
1	A	2	<code>E{-0}</code>	$E_0$	Resonance energy
1	B	1	<code> G</code>	$\Gamma$	Total width of resonance at $E_0$
1	C	4	<code>2g G{-n}</code>	$2g\Gamma_n$	Twice the total neutron width of the resonance at $E_0$ times the statistical factor $g$
1	D	1	<code>J</code>	$J$	$\vec{J} = \vec{I} + \vec{l} + 1/2$ . Total spin of the compound state formed by the resonance at $E_0$
1	E	1	<code>{I1}</code>	$L$	Orbital angular momentum brought in by the projectile to the resonance at $E_0$
1	F	2	<code> G{- g}</code>	$\Gamma_\gamma$	Radiative capture width of resonance at $E_0$
1	G	4	<code>2g G{+0}\{-n}</code>	$2g\Gamma_n^0$	Twice the s-wave neutron width of the resonance at $E_0$ times the statistical factor $g$
1	H	4	<code>2g G{+1}\{-n}</code>	$2g\Gamma_n^1$	Twice the p-wave neutron width of the resonance at $E_0$ times the statistical factor $g$
1	CA	3	<code>g G{-n}</code>	$g\Gamma_n$	Neutron width of resonance at $E_0$ times the statistical factor $g$
1	CB	2	<code> G{-n}</code>	$\Gamma_n$	Neutron width of resonance at $E_0$
1	FA	3	<code>g G{- g}</code>	$g\Gamma_\gamma$	Gamma width of resonance at $E_0$ times the statistical factor $g$
1	FC	4	<code>2g G{- g}</code>	$2g\Gamma_\gamma$	Twice the gamma width of resonance at $E_0$ times the statistical factor $g$
1	IA	7	<code>g G{-n} G{- g}/ G</code>	$g\Gamma_n\Gamma_\gamma/\Gamma$	Total radiative capture area (See Atlas, Eq. (3.4): $A_\gamma/(2\pi^2\lambda^2) = g\Gamma_n\Gamma_\gamma/\Gamma$ )
1	IJ	8	<code>2g G{-n} G{- g}/ G</code>	$2g\Gamma_n\Gamma_\gamma/\Gamma$	Two times the total radiative capture area (See Atlas, Eq. (3.4): $A_\gamma/(2\pi^2\lambda^2) = g\Gamma_n\Gamma_\gamma/\Gamma$ )

Some of the data stored in the Atlas has additional modifiers attached to it including a statistical factor  $g$ . This factor is defined as  $g = (2J + 1)/[2(2I + 1)]$  where  $I$  is the ground state spin of the target nucleus and  $J$  is the total spin of the compound state.

To store all of the data derived from the Atlas files, a resonance table class was created which, among other things, would contain the resonance parameters, ground state spin of the target nucleus, parity of the target nucleus and the projectile. During the setup of the resonance table class the published data of Chromium-52 available in the Atlas of Neutron Resonances was used primarily as a test subject with focus also placed upon Indium-115. The machine learning classification we look to employ on the resonance data from the Atlas requires a complete set of resonance parameters to train and make predictions. When extracting data from the Atlas, many holes in the resonance data was observed which would cause problems with our methods. In order to overcome this we needed a way to fill these gaps somehow. The methods we employed to fill in these holes is discussed further on. When determining the available data we realized that while  $\Gamma_n g$  was available for a large part of the data at low energy, in the higher energy realm this source faltered requiring secondary source for this quantity. The wide availability of the total radiative capture area enabled us to fill in much of the missing data for neutron width as well as some for the capture width.

Setup of the resonance parameters was based primarily upon the quantum numbers  $L$  and  $J$  and the availability of either. Many cases when both were provided, solving for the neutron width and the capture width remained a trivial manner of removing the statistical factor  $g$  or simply just importing the values.

Algebraic manipulation came into play when given the total radiative capture area in conjunction with the L and J quantum numbers and either the capture or neutron width where we solved for the other. The total radiative capture area is defined in Table 1 and we can rearrange to find the neutron width as:

$$\Gamma_n = \frac{\bar{A}_\gamma \Gamma_n}{g \Gamma_n \bar{A}_\gamma} \quad (1)$$

or the capture width as:

$$\Gamma_\gamma = \frac{\bar{A}_\gamma \Gamma_\gamma}{g \Gamma_\gamma \bar{A}_\gamma} \quad (2)$$

where  $\bar{A}$  is defined as  $\frac{A_\gamma}{2\pi^2 \lambda^2}$ .  $\bar{A}$  is dependent upon  $\lambda^2$  where  $\lambda^2$  is defined as:

$$\lambda = \frac{\hbar c}{k} = \frac{197}{k} fm \quad (3)$$

and [3]:

$$k = \frac{\sqrt{2m_n}}{\hbar} \frac{AWRI}{AWRI + 1.0} \sqrt{|E|} \quad (4)$$

where AWRI= ratio of the mass of a particular isotope to that of the neutron and E= laboratory energy in eV.

A final part of the puzzle that we could discern from available data relied upon knowledge of  $\Gamma_n$ ,  $\Gamma_\gamma$ , and  $\bar{A}$  ( $\frac{g\Gamma_n\Gamma_\gamma}{\Gamma}$ ). With this we are able to find the total spin of the compound nucleus. Finding this value wherever we could was paramount as not having to guess parts of the spin group assignment reduces the bias that is fed into our machine learning algorithm. This eliminates the need to guess the total spin in cases where we have orbital momentum but not total spin and eliminates the need to guess both total and orbital in favor of just orbital for cases with both missing spin group assignment.

Many times after building as complete a resonance set as possible, the need for assigning values for missing parameters is necessary. For any of the spin group assignments, this is done with a random guess spin group based on the available values for L and J. For the case where we cannot determine either assignment, one is made based on a random choice from spin-groups already defined by the Atlas. This does pose a problem for sequences that have no fully defined spin groups. However, those sequences are not in the intended application of this method nor is it common for resonance data to be populated as such.

For cases when we need either the orbital or the total spin, this is done through the use of the formula for J range:  $J = \frac{1}{2} + I \pm L$  where I is the ground state spin. For the L values they lie in the range:  $L = [\frac{1}{2} + I - J, J - I - \frac{1}{2}]$ , again with I being the ground state spin.

While this solves the spin group problem there are still times when missing neutron and capture widths plague the resonance table. The neutron widths are generally well documented throughout our methods and have few missing at the lower energy levels. For the times where we are missing them in smaller quantities, we first turned to resonance properties, chiefly the mean s, p and d-wave level spacing correlating to the respective L=0,1,2 orbital angular momenta. In the final cases where this approach does not fill in the method, a final, albeit imperfect, method of imputing values is used with either the mean or median neutron width value used to replace missing values. This method is non-ideal. However, in conjunction with other features, we hope the impact is partially nullified.

The capture width is a larger problem that called for a method to train the algorithm with or without the capture width as one of the resonance features. Even in well documented resonances such as  $^{52}Cr$ , there is an overwhelming lack of capture widths with this sequence only having 22% of capture widths. Because of this, any method that we might employ to fill in the missing blanks can result in highly erroneous training due to the imperfection of assignment. This prompted us to build in an identifier for low percentages of capture width that prompts the user to consider eliminating the capture width as a training parameter. For times when we may employ it however, the same idea as for neutron widths was used with the average resonance capture widths taking precedence over imputing values.

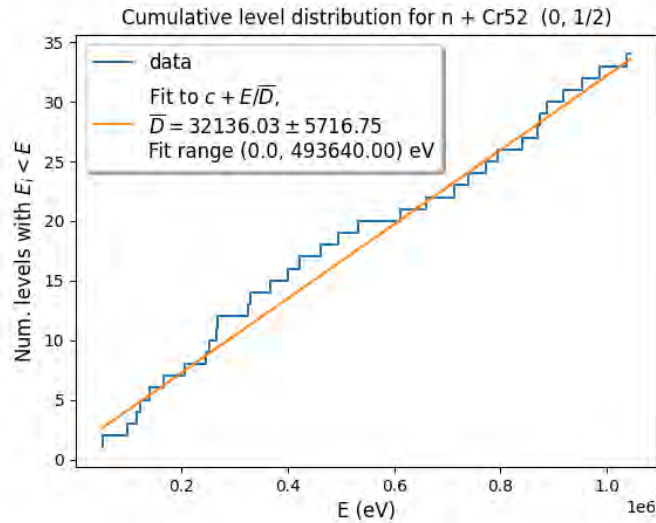


Figure 2: This is the initial Cumulative Level Distribution (CLD) for the  $^{52}\text{Cr}$  isotope using the spin group  $L=0; J=\frac{1}{2}$ . The staircase look with the blue line shows each step as a resonance in the sequence. The orange line shows the computed linear fit of the average resonance spacings, which is the fit is the theoretical path the CLD should follow for the resonance sequence.

### 3 Results and Discussion

The first sequence that we tested was with  $^{52}\text{Cr}$  using the Nearest Neighbors classifier as defined earlier. We first constructed the Cumulative Level Distribution, the resonance level spacings, for the isotope using the  $L=0; J=\frac{1}{2}$  spin group as shown in Figure 2. After application to the machine learning algorithm utilizing the Nearest Neighbors classifier we obtained a resonance set with suggested changes to the orbital angular momentum. Upon applying these changes we then fit the CLD to this reassigned set of resonance data as shown in Figure 3

Our testing yielded varying results, with average reassignment of around 24% of resonances reclassified. This applies to all available spin groups which explains the slight difference in the CLD plots for the  $L=0; J=\frac{1}{2}$  spin groups. Corrections to  $^{52}\text{Cr}$  spin groups were 14% which points to the machine learning algorithm having enough data from the Atlas to make classifications.

Testing with Indium-115 was also done however with much different results than with  $^{52}\text{Cr}$ . When setting up the resonance data, large sources of error were found. These included 78% missing total spin assignments which in turn affected the neutron width assignment. The neutron width was discerned from its storage as  $2g\Gamma_n$ . The statistical factor  $g$  as defined earlier relies upon the knowledge of the  $J$  assignment as for times when we replaced the unknown  $J$  value with a randomly assigned value, this will bias the neutron width due to its reliance upon the unknown value  $J$ . Shown in Figure 4 is the CLD before reclassification.

After classification with the Nearest Neighbors classifier which returned only 9% changed resonances, the CLD for the same spin group is shown in Figure 5.

While the percentage of changed resonances for Chromium is more desirable than Indium, this marks an important step in the project goal. The Atlas of Neutron Resonances remains one of the premier sources of compiled nuclear reaction resonance data. This initial utility gives us the ability to access stored resonance data for any nucleus available in the Atlas of neutron resonances to use for our machine learning algorithm.

### 4 Conclusion

The work by Mary Fucci on optimizing the machine learning algorithms and hyperparametrization is part of the large picture to better our machine learning algorithm. With the completion of this work on translating



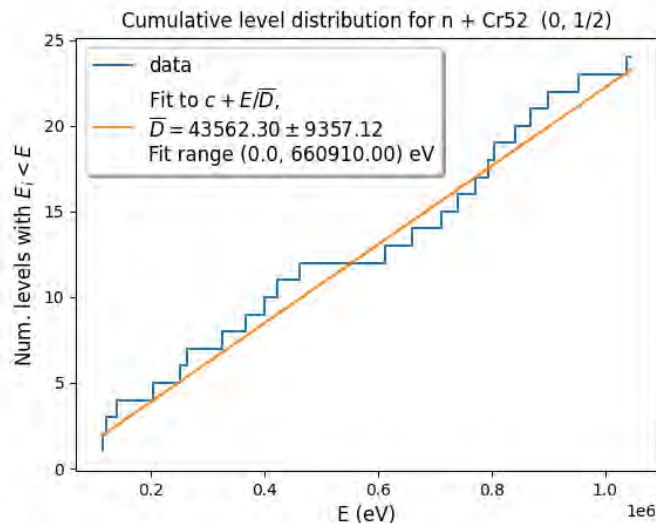


Figure 3: This is the Cumulative Level Distribution (CLD) for the  $^{52}\text{Cr}$  isotope using the spin group  $L=0$ ;  $J=\frac{1}{2}$  after inclusion of reassigned orbital angular momenta values from the Nearest Neighbors classifier. The staircase look with the blue line shows each step as a resonance in the sequence. The orange line shows the computed linear fit of the average resonance spacings which is the theoretical path the CLD should follow for the resonance sequence.

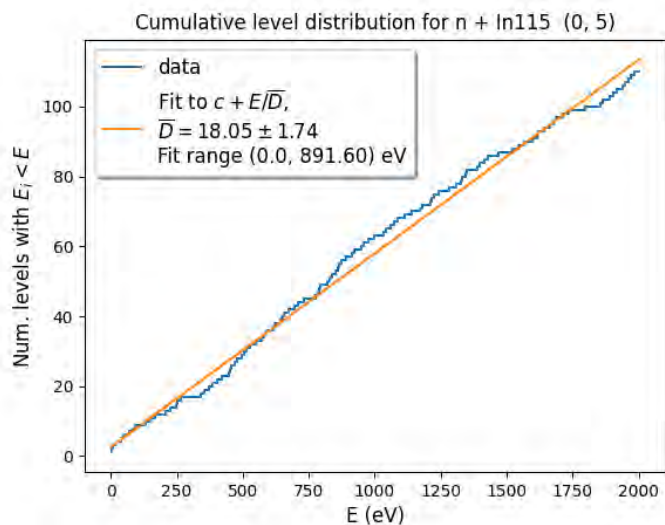


Figure 4: This is the initial Cumulative Level Distribution (CLD) for the  $^{115}\text{In}$  isotope using the spin group  $L=0$ ;  $J=5$ . The staircase look with the blue line shows each step as a resonance in the sequence. The orange line shows the computed linear fit of the average resonance spacings which is the theoretical path the CLD should follow for the resonance sequence.

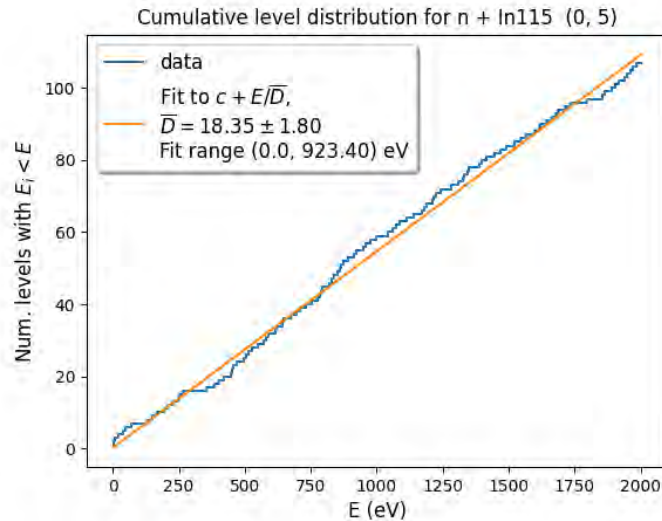


Figure 5: This is the Cumulative Level Distribution (CLD) for the  $^{115}\text{In}$  isotope using the spin group  $L=0$ ;  $J=5$  after inclusion of reassigned orbital angular momenta values from the Nearest Neighbors classifier. The staircase look with the blue line shows each step as a resonance in the sequence. The orange line shows the computed linear fit of the average resonance spacings which is the theoretical path the CLD should follow for the resonance sequence.

Atlas resonance data, we now have access to resonance data gathered nearly every nucleus observed in nuclear experiments. Future exhaustive testing may reveal additional ways to setup data from the Atlas as we encounter more obscure scenarios however addressing those is a relatively simple manner now that the framework is well in place. Additional work by Rose Marie Crawford on the effect of missing resonances on average width calculations maybe used to help improve the Atlas resonance table class for training. Finally, work by Sergio Ruiz on polarized resonance data validation is much easier now that there is an automated way to retrieve Atlas data. Future work on refining spin assignment, width assignment, and machine learning optimization will help improve the accuracy of our classifications and enable us to start correcting resonance sequences.

## 5 Acknowledgments

I'd like to thank my mentor Dr. Gustavo Nobre as well as Dr. David Brown, Dr. Matteo Vorabbi and Dr. Amber Coles for their guidance and intuition throughout this appointment as well as my incredible fellow interns. This project was supported in part by the Brookhaven National Laboratory (BNL), Department of Nuclear Science and Technology under the BNL Supplemental Undergraduate Research Program (SURP).

## References

- [1] From the “Institute of research into the fundamental laws of the Universe – Fundamental Research Division” at the CEA, [http://irfu.cea.fr/en/Phoce/Vie\\_des\\_labos/Ast/ast\\_visu.php?id\\_ast=4237](http://irfu.cea.fr/en/Phoce/Vie_des_labos/Ast/ast_visu.php?id_ast=4237), downloaded Sep. 2020.
- [2] scikitlearn, <https://scikit-learn.org/stable/>
- [3] Trkov, Andre, & Brown, David A. ENDF-6 Formats Manual: Data Formats and Procedures for the Evaluated Nuclear Data Files. BNL-203218-2018-INRE, United States. <https://doi.org/10.2172/1425114>

- [4] D.A. Brown, D. Mulhall, R. Wadgoankar, “A tale of two tools: mres.py, a stochastic resonance generator, and grokres.py, a resonance quality assurance tool,” Technical Report BNL-209313-2018-INRE, Brookhaven National Laboratory, (2018), doi:10.2172/1478482.
- [5] R. Capote, M. Herman, Et. al., RIPL-Reference Input Parameter Library for Calculation of Nuclear Reactions and Nuclear Data Evaluations, 2009
- [6] Atlas of Neutron Resonances (Sixth Edition) (S.F. Mughabghab, ed.), Elsevier, Amsterdam, sixth edition ed., 2018

# User friendly interfaces for Jupyter Notebook data analysis scripts

Ramon Sepulveda<sup>1</sup> and Kenneth Evans-Lutterodt<sup>2</sup>

<sup>1</sup> *Computer Science Department, Inter American University of Puerto Rico, San German, PR, 00683*

<sup>2</sup> *National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY, 11973*

Experimental data produced by the Integrated In situ and Resonant Hard X-ray Studies (ISR) beamline at the National Synchrotron Light Source II (NSLS-II) is primarily conducted by employing user-written software scripts. Normally, these scripts are adapted for new analyses by opening the source code file with a text editor and rewriting program statements. This approach is relatively inefficient and error-prone since a user is often editing a program without recalling its design constraints. Adding to the challenge of editing scientific analysis scripts is the fact that these programs almost always follow a structured programming model which simplifies initial development speed but makes it significantly more difficult to extend its functionality. To solve this problem, we redesigned a selection of existing Jupyter Notebook scripts so that the underlying logic was object-oriented. The new logic was implemented as a single desktop GUI application built entirely in Python using the Tkinter toolkit. The modularity of the application combined with the definition of standard interfaces will allow users to extend functionality by opening a single module and adding one or more function definitions. This new “add it, save it, and forget it” workflow is a major improvement over the previous workflow which required editing multiple files or a single file at multiple points which had to be identified by reading the entirety of the source code. Consequently, we predict that, by utilizing the application and its underlying programming paradigm, software development productivity at the beamline will improve by at least a factor of two.

Computers have become an essential tool in all branches of science<sup>1</sup> with organizations shifting their focus towards exa-scale computing as a way to tackle their most pressing research challenges.<sup>2</sup> However, despite the recognized necessity of computers in modern research, most researchers are never taught the equivalent of basic lab skills for research computing<sup>1</sup>. This skill gap often extends to computer

programming with numerous studies<sup>3,4,5</sup> reporting results that substantiate the claim that the software development skills attained by students over the course of their major are disjoint from the skills desired by employers. Given this background, it would be reasonable to infer that the software being developed for a significant portion of scientific computing applications does not follow generally accepted good programming practices. Our research aims to assess the extent of “non-compliance” with good programming practices for a set of Python scripts currently being employed at the Integrated *In situ* and Resonant Hard X-ray Studies (ISR) beamline at the National Synchrotron Light Source II (NSLS-II) facility within Brookhaven National Laboratory (BNL).

The software scripts in active use at the ISR beamline that were selected for analysis totaled three and were stored/run as Jupyter Notebook files (.ipynb). Each notebook was “self-contained” with all executed code stored within the single file being run. Two of these scripts incorporated graphical user interface (GUI) elements produced by the matplotlib and ipywidgets libraries. In our analysis, the notebook kernel executing these Jupyter Notebooks used Python 3.7.6

The original programs analyzed were written to process data that corresponded to measurements taken by an Eiger R 1M x-ray detector manufactured by Dectris Ltd. At the ISR beamline, a user works on at least one experiment where each experiment is composed of one or more scans and where each scan is composed of one or more “snapshots” or points that each correspond to the x-ray detector measurement for a particular input of x-ray diffractometer variables; these relationships are depicted by the entity-relationship diagram in Figure. Each point is stored as a 32-bit integer array with dimensions (514, 1030) within a Hierarchical Data Format (.h5) file whose path string representation contains, at the very least, a substring that identifies the “parent” scan and a substring that identifies the particular point.

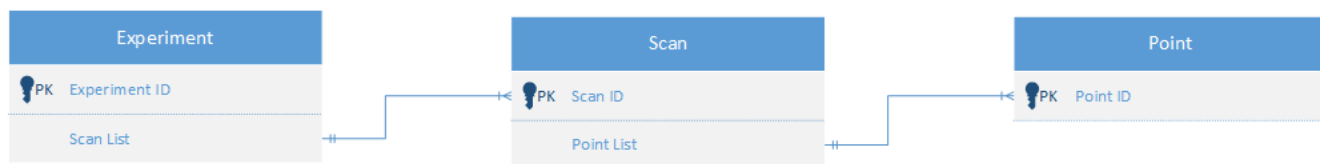


Fig.1 – Entity-relationship diagram for file organizational structure of ISR beamline data

The selected ISR scripts plotted each scan point as a greyscale raster image where the luminosity of each pixel was directly proportional to its corresponding array value as can be seen in Figure. The purpose of this rendering was to allow users to visually inspect a series of scan points to identify regions of interest, determine what diffractometer variables to modify next, start inferring the general structure of the scanned sample. etc. Together with the plots was, in some cases, a simple graphical user interface (GUI) with widgets bound to plotting parameters that offered dynamic “re-rendering”. This GUI was embedded inline within the running Jupyter Notebook with plots displayed in a separate window (see Figure).

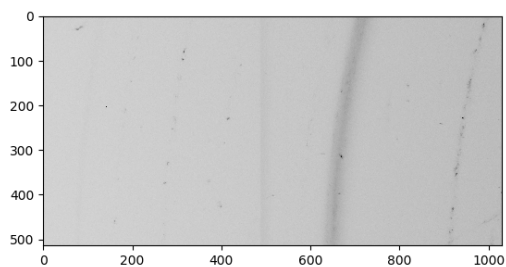


Fig. 2 – Image produced by pyplot package when plotting data captured by Eiger R 1M detector

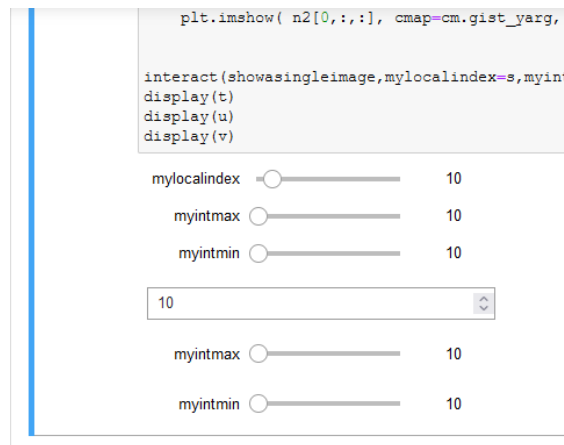


Fig. 3 – Original GUI produced by one of the ISR beamline scripts analyzed

After reading the source code of the selected scripts multiple times, we found that they did not follow many of the generally accepted good programming practices outlined by Wilson et al in their 2017 paper, e.g. decomposing programs into functions, avoiding duplication, giving functions/variables meaningful names, making dependencies and requirements explicit, among other practices. Since we determined that there was room for improvement with regards to the internally developed software being used at the ISR beamline, our focus shifted towards developing a proof-of-concept application that would help substantiate our assertion that good programming practices can be learned “from the ground up” by programmers with the most basic software development experience, i.e. comparable to that of a first-year computer science undergraduate student. In addition, the project application will feature a desktop GUI that is generally considered “user friendly.”<sup>6</sup>

The functionality of all three original scripts was consolidated into a stand-alone GUI application intended for use in desktop environments. As with the original scripts, the application was programmed entirely in Python with graphical elements generated by the tkinter package. The application code was written with an object-oriented structure and stored within a directory structure that was automatically generated by the cookiecutter package. Development was initially done in Jupyter Notebooks to produce proof-of-concepts and then done in Visual Studio Code.

The result of our project is a classic desktop GUI application with that consists of three windows accessed in sequential order. The "root" or start-up window has a simple form layout with four labels, three entries, and two buttons as it is displayed in Figure 4. In the first section, the user must specify the file system path of the parent directory of the selected scan; this path can be specified either by typing in its text representation in the text entry or by pressing the "Select Directory" button and navigating its associated file dialog window. In the second section, the user must type in a Python regular expression or "regex" that uniquely identifies all files belonging to the scan of interest. Finally, the user types a Python regex that uniquely identifies the points of interest within that particular scan. Once all text entries have been filled, the user presses the "Submit" button and the application validates the three text inputs. If these inputs are valid, the display changes to the second application window, also known as the Image Previewer.

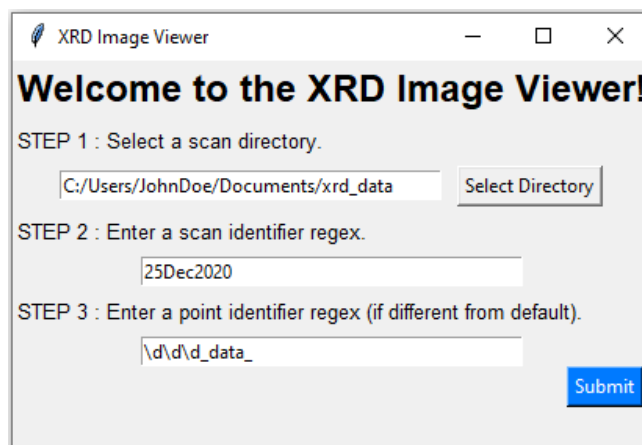


Fig. 4 – Root or start-up window produced by XRD Image Viewer application with example entries

As its name suggests, the Image Previewer's purpose is to allow users to view graphical representations of the selected points before these points are analyzed by one or more functions. Users can quickly skip through the point images, zoom in on a region of interest, and change image parameters to alter contrast by using the widgets on the "Controls" panel on the left side of the plotting area (see Figure 5). The controls



housed by this panel consist of three slider widgets and two button widgets. On the top-most portion of the Image Previewer window are two labels: a "Scan ID" label that displays the regex that identifies the scan and an "Image Total" label that displays a count of all the currently selected scan points. Once a user has perused through the point images and identified one or more regions of interest within these images, then they can press the "Analyze" button which changes the display to the third and final window, also known as the Function Wizard.

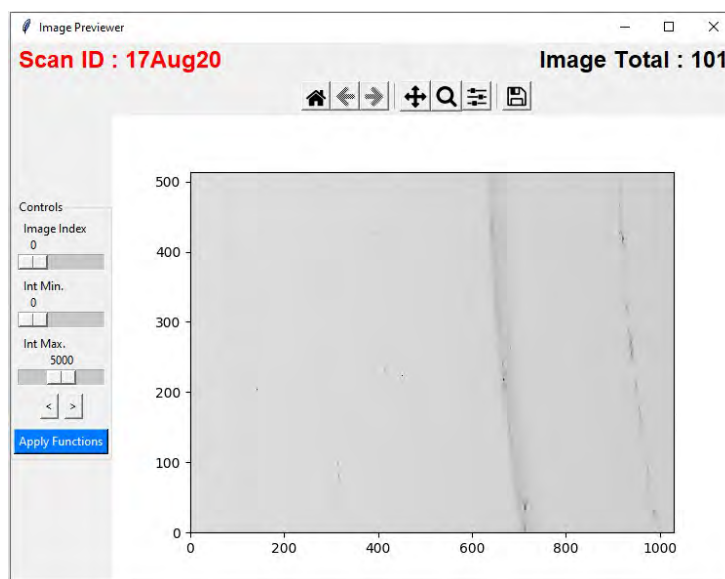


Fig. 5 – Image Previewer window for the XRD Image Viewer application with example data

The third window also has a plotting area with the point image shown when the "Analyze" button was pressed in the previous window as can be seen in Figure 6. A list box with associated title label and a series of three labeled panels appear to the left of the third window's plotting area. The list box contains all the user-written functions that have currently been incorporated into the application; to execute any one function, a user simply needs to double-click its corresponding element. In the case a user simply wishes to view function information, they can just perform a single click. The first piece of function information displayed would be the instructions on what to do when function is executed. Next would be an inventory of the output types with short associated descriptions. Finally, all function output which can

be parsed as text would be displayed. The aforementioned three pieces of information would be displayed in said order within three corresponding labeled panels

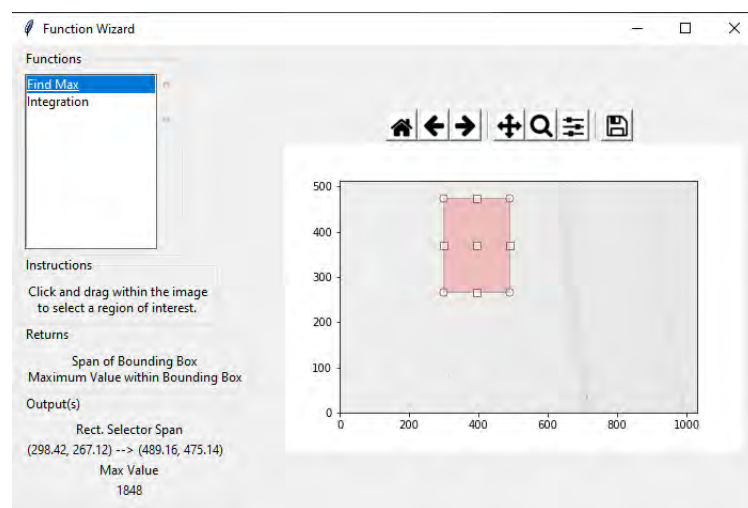


Fig. 6 – Function Wizard window for the XRD Image Viewer application with example data

Our results seem to substantiate the hypothesis that good programming practices can deliver software that is more performant and usable in scientific computing contexts, while also being easier to develop in the long run, than traditional approaches used in the field.

## ACKNOWLEDGMENTS

I am profoundly grateful to Dr. Kenneth Evans-Lutterodt for the opportunity to work on this project and to Karl Clarke for his exceptional support. I'd also like to thank Dr. Carmen Caiseda and Prof. Joel Lisboa for their help in this and many other endeavors.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## REFERENCES

- <sup>1</sup>G. Wilson, J. Bryan, K. Cranston, J. Kitzes, L. Nederbragt, and T.K. Teal, PLOS Computational Biology **13**, (2017).
- <sup>2</sup>R. Lucas, J. Ang, K. Bergman, S. Borkar, W. Carlson, L. Carrington, G. Chiu, R. Colwell, W. Dally, J. Dongarra, A. Geist, Grider, Gary Los Alamos National Laboratory, R. Haring, J. Hittinger, A. Hoisie, D.M. Klein, P. Kogge, R. Lethin, V. Sarkar, R. Schreiber, J. Shalf, T. Sterling, R. Stevens, J. Bashor, R. Brightwell, P. Coteus, E. Debenedictus, J. Hiller, K.H. Kim, H. Langston, Laros III, James Sandia National Laboratories, Leyffer, Sven Argonne National Laboratory, R.M. Murphy, Ross, Rob Argonne National Laboratory, C. Webster, and S. Wild, (2014).
- <sup>3</sup>Peter Knoke via Don Bagert. Graduate Software Engineering Program Survey Results & Evaluation. Forum for Advancing Software engineering Education (FASE), vol 8 no 9, September 15, 1998, <http://www.cs.ttu.edu/fase/v8n09.txt>.
- <sup>4</sup>Mary Shaw. Prospects for an engineering discipline of software. IEEE Software, pages 15-24, November 1990.
- <sup>5</sup>James E. Tomayko. Forging a discipline: An outline history of software engineering education. Annals of Software Engineering 6 (1998) 3-18.
- <sup>6</sup>L. Trenner, Journal of Information Science **13**, 99 (1987).

# Shielding for Accelerator Radiological Operations

Zaha Shahdad, Division of PEMaCS, Delaware State University, Dover, DE 19709

Sean Harling, Steven Coleman; ES&H Directorate; Brookhaven National Laboratory, Upton, NY 11973

## ABSTRACT

The Medical Isotope Research Program (MIRP) within the Collider-Accelerator Department (C-AD) at Brookhaven National Laboratory (BNL) produces isotopes for research purposes and to supply the national isotope user community. Proton beams from the Linear Accelerator (LINAC) at BNL irradiate a target array containing various materials, producing the desired radionuclides that are then processed and purified. The operations' primary hazard of ionizing radiation has enough energy to knock-off electrons from an electrically neutral atom, thereby ionizing it. Examples of ionizing radiation at BNL include alpha and beta particles, gamma/x-rays, and neutron particles. For protecting workers, visitors, public, and the environment, operational hazards are identified and minimized to insignificant levels with shielding and other controls. Through this internship, I collaborated with Health Physicists at BNL and familiarized with FLUKA, a general-purpose Monte Carlo code used to calculate the radioactivity of radioisotopes in MIRP targets, analyze their impact on a shielding apparatus, and perform other related analyses. These shielding applications and FLUKA results are performed to meet the requirements of DOE O 420.2C and 10 CFR 835. This report includes few examples of simple geometries modeled using FLUKA and associated Flair Graphical user interface (GUI).

## **BACKGROUND**

### *Radioisotopes and ionizing radiation*

Atoms with the same number of protons but varying number of neutrons in the nucleus are called isotopes; in a radioisotope, the nucleus is unstable and has the potential to emit radiation during decay. For example, tritium is an isotope of hydrogen with two neutrons and a proton in the nucleus. Tritium classifies as a radioisotope and decays to an isotope of helium, emitting a neutrino and a beta-particle in the process. Radioisotopes are often used for diagnosis and treatment of various medical conditions, including different types of cancer. BNL's Medical Isotope Research Program (MIRP) produces and supplies various such radionuclides, including Actinium-225, which emits alpha particles during decay. Alpha radiation is an example of ionizing radiation; the high energy particle can damage cells but have a short range of travel, making it an effective tool against certain cancers when targeted towards cancerous cells.<sup>1</sup> In the production and retention of these radionuclides, various other forms of ionizing radiation may also be produced.

Ionizing radiation is defined as radiation that can knock-off electrons from neutral atoms, thereby ionizing them. Four main types of ionizing radiation are alpha particles (made up of two protons and two neutrons), beta particles (an electron emitted from the nucleus and not bound to the atom), gamma and X-ray (two kinds of electromagnetic radiation), and neutrons (emitted from the nucleus with high speed).<sup>2</sup> If unshielded against, large doses of ionizing radiation can pose a danger to personnel and/or the environment.

### *Shielding in radiation safety*

A fundamental concept in radiation safety is the concept of Time, Distance, and Shielding: decreasing the time spent near a radiation source, increasing one's distance from the source, and placing appropriate shielding between personnel and the source are important ways of reducing exposure to ionizing radiation. Among the various controls that are established and maintained to ensure the safety of workers, visitors, public, and the environment, shielding is an important engineering control. The shielding material, density, and dimensions vary greatly depending on the type and intensity of the radiation source one is shielding against. For example, since alpha

<sup>1</sup> See References: Shea (2018).

<sup>2</sup> See References: *Radiation Basics*.

radiation has a shorter range of travel in air compared to beta radiation, the former can be shielded against using a sheet of paper, while the latter would require a denser material such as Aluminum (Al) plate. In contrast, gamma and X-rays are much higher in energy and would require an even denser material such as lead (Pb) block.

Establishing appropriate shielding is an essential part of accelerator safety. In minimizing, analyzing, and controlling hazards, compliance with DOE O 420.2C (Safety of Accelerator Facilities)<sup>3</sup> and 10 CFR 835 (Occupational Radiation Protection)<sup>4</sup> are also required. To briefly summarize the two documents, DOE O 420.2C establishes accelerator-specific safety requirements and 10 CFR 835 establishes standards, limits, and requirements for radiation protection. Along with these regulations, radiation exposure is required to be As Low As Reasonably Achievable, commonly referred to as the ALARA principle. At BNL, FLUKA analyses are frequently performed to comply with these requirements.

## **METHODS**

Effective dose (or dose equivalent) is defined as “the amount of radiation absorbed by person, adjusted to account for the type of radiation received and the effect on particular organs.”<sup>5</sup> The unit of effective dose in the S.I. system is the Sievert (Sv) and in the United States is commonly the rem. Dose rate is the amount of time over which a given dose of radiation is received and can be represented by units of rem/hr. The unit millirem (mrem) is more commonly used than rem at BNL, due to very low occupational radiation exposures. The total dose one may receive is then calculated as follows: Total dose (mrem) = Dose rate (mrem/hr) \* Exposure duration (hr).

According to 10 CFR 835, occupational dose limits for general employees include a total effective dose limit of 5 rems, or 5000 mrem. At BNL administrative control level, this limit is further reduced to 1250 mrem. In combination with the philosophy of ALARA, which does not establish a threshold, any unnecessary radiation exposure from accelerator operations can be minimized. One of the many steps to ensure that effective controls are in place is to use simulation tools to model operations and analyze potential risks. An example is to dose score operations.

<sup>3</sup> Department of Energy Order 420.2C, accessible at [www.directives.doe.gov](http://www.directives.doe.gov)

<sup>4</sup> Code of Federal Regulations, Title 10, Chapter 3, Part 835, accessible at [ecfr.gov](http://ecfr.gov)

<sup>5</sup> See References: *Radiation Terms and Units*.

FLUKA<sup>6</sup> is a general-purpose Monte Carlo code for simulating transport of particles and nuclei and their interactions with matter in complex geometries. FLUKA has been used at BNL to design the National Synchrotron Light Source-II (NSLS-II) accelerator and beamline shielding and simulate irradiation of various MIRP targets. For example, FLUKA can be used to calculate inventory of radionuclides produced in each of the targets. The Geometry feature in the associated Flair<sup>7</sup> GUI also allows users to build and visualize 3D geometry models of the targets.

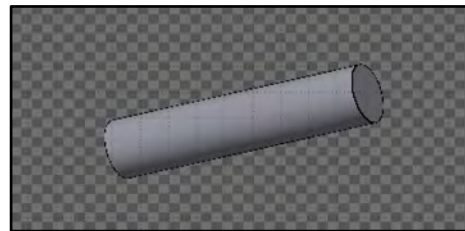
## RESULTS

### *Examples of simple FLUKA analyses*

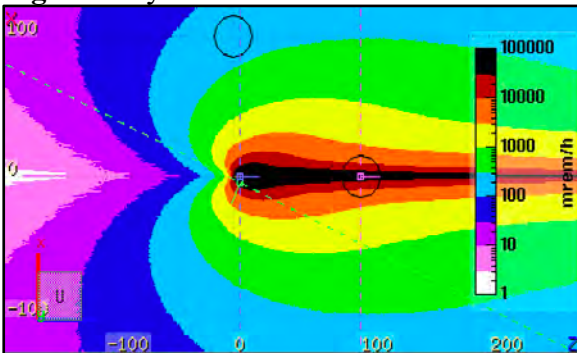
The figures below show unshielded dose scoring of a 3 GeV beam of electrons striking a copper (Cu) cylinder and lead (Pb) cylinder, which are few simple geometries modeled with FLUKA. A total of two detectors are placed: at 0 degrees and 1 meter away from the source and 90 degrees and 1 meter away from the source, respectively. The unshielded ambient dose rates (in mrem/hr) from photons, neutrons, and electrons produced from the interaction are displayed here, assuming a beam intensity of 1 nC/min. Both cylinders are identical in dimensions with a length of 10 cm and a radius of 1.2 cm, and the axes of dose rates are also given in cm.



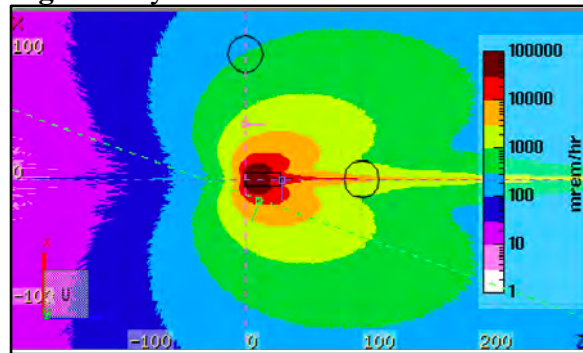
**Fig 1. Cu cylinder**



**Fig 2. Pb cylinder**



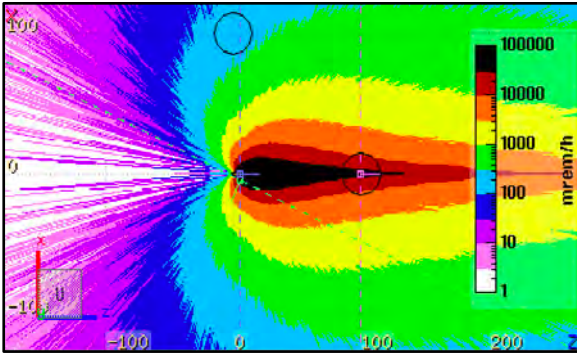
**Fig 3. Cu target. Unshielded Ambient Dose Rates- Photons**



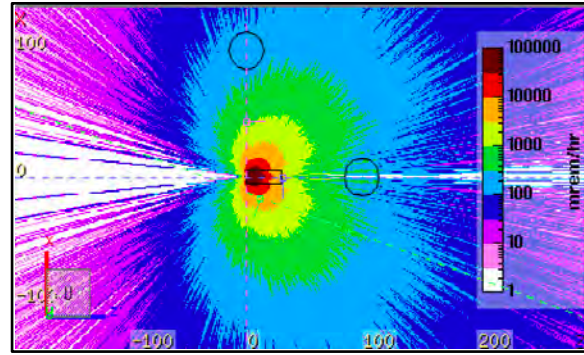
**Fig 4. Pb target. Unshielded Ambient Dose Rates- Photons**

<sup>6</sup> FLUKA Website: <https://fluka.cern>; See References: G. Battistoni, et al. and T.T. Bohlen, et al.

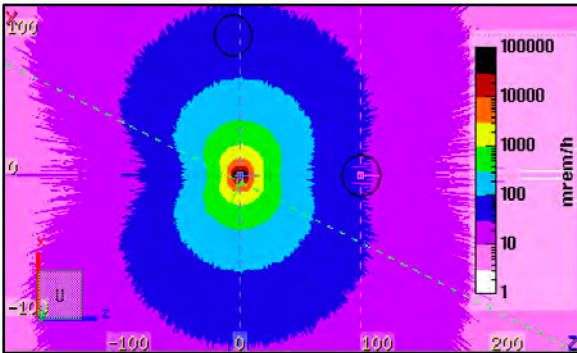
<sup>7</sup> See References: V. Vlachoudis.



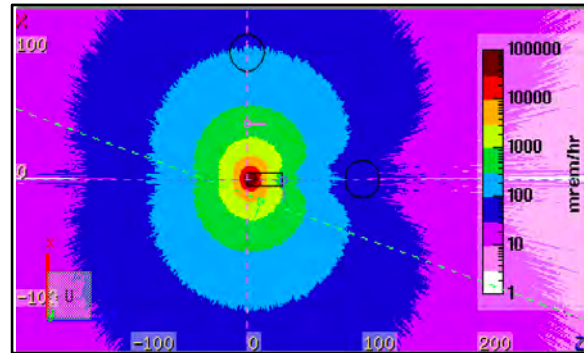
**Fig 5. Cu target. Unshielded Ambient Dose Rates- Electrons**



**Fig 6. Pb target. Unshielded Ambient Dose Rates- Electrons**



**Fig 7. Cu target. Unshielded Ambient Dose Rates- Neutrons**



**Fig 8. Pb target. Unshielded Ambient Dose Rates- Neutrons**

## DISCUSSION

It can be noted that dose rates from photons and electrons in the forward direction along the z-axis (beam propagation axis) is lower from the Pb target (figures 3 and 5) than the Cu target (figures 4 and 6). This suggests that Pb attenuate photons and electrons fluence along the beam axis more effectively than Cu. However, the higher Z-value of Pb contributes to higher ambient dose rates from neutrons compared to dose rates from the Cu target (figure 7 vs. figure 8).

## CONCLUSION

Shielding is an important engineering control for accelerator radiological operations. Tools such as FLUKA can be used not only to simulate the irradiation of targets in the MIRP program, but also to design shielding for protecting personnel and the environment by limiting exposure to ionizing radiation from the operations. Furthermore, it is an important analysis to ensure compliance with various regulatory limits on radiation exposure and other accelerator and radiological operations requirements.



## **ACKNOWLEDGEMENTS**

This project was carried out under the mentorship of Steve Coleman and Sean Harling, who taught me a lot about accelerator radiation safety. The FLUKA simulation results were made possible under the guidance of Mohamed Benmerrouche, who helped me get hands-on experience with the code. Thank you to all my mentors and the Radiological Control Division for their time and commitment in making this internship a productive experience.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## References

G. Battistoni, T. Boehlen, F. Cerutti, P.W. Chin, L.S. Esposito, A. Fassò, A. Ferrari, A. Lechner, A. Empl, A. Mairani, A. Mereghetti, P. Garcia Ortega, J. Ranft, S. Roesler, P.R. Sala, V. Vlachoudis, G. Smirnov, "*Overview of the FLUKA code*", *Annals of Nuclear Energy* 82, 10-18 (2015).

*Radiation Basics*. United States Nuclear Regulatory Commission . (2020, March 20).  
[www.nrc.gov/about-nrc/radiation/health-effects/radiation-basics.html](http://www.nrc.gov/about-nrc/radiation/health-effects/radiation-basics.html).

*Radiation Terms and Units*. Environmental Protection Agency. (2021, July 21).  
[www.epa.gov/radiation/radiation-terms-and-units](http://www.epa.gov/radiation/radiation-terms-and-units).

Shea, S. B. (2018, June 20). *The journey of actinium-225: How scientists discovered a new way to produce a rare medical radioisotope*. Energy.gov.  
[www.energy.gov/science/articles/journey-actinium-225-how-scientists-discovered-new-way-produce](http://www.energy.gov/science/articles/journey-actinium-225-how-scientists-discovered-new-way-produce).

T.T. Bohlen, F. Cerutti, M.P.W. Chin, A. Fassò, A. Ferrari, P.G. Ortega, A. Mairani, P.R. Sala, G. Smirnov, and V. Vlachoudis, "*The FLUKA Code: Developments and Challenges for High Energy and Medical Applications*", *Nuclear Data Sheets* 120, 211-214 (2014).

V. Vlachoudis, "*FLAIR: A Powerful But User Friendly Graphical Interface For FLUKA*", in *Proc. Int. Conf. on Mathematics, Computational Methods & Reactor Physics (M&C 2009)*, Saratoga Springs, New York, 2009.

Mechanical engineering & design for Coherent Electron Cooling diagnostics beamline and Electron Ion Collider Superconducting Radio Frequency test facility upgrades

Christopher Sherling

Mechanical Engineering, Stony Brook University, Stony Brook, NY 11794

Clifford Brutus

Collider-Accelerator Department, Brookhaven National Laboratory, Upton, NY 11973

## **ABSTRACT**

Coherent Electron Cooling (CeC) is a method of increasing the number of collisions that occur in the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory (BNL). The CeC experiment aims to cool the hadron beam from RHIC using its generated electron beam, reducing the energy spread and emittance of the ion beam. A Time Resolved Diagnostic Beamline (TRDBL) is implemented along the CeC beamline to ensure that the properties of the beam are as desired. A variety of components along the TRDBL were relocated to accommodate additional requirements. This required considerable reworking to their mounting and fixtures. Updates were made accordingly to the 3D model of the TRDBL and new technical drawings were created for use in manufacturing. In addition to work on the CeC TRDBL, work has been done as part of the ongoing effort to convert the RHIC to an Electron-Ion Collider (EIC). This requires a variety of new components along the beamline, including superconducting radio frequency (SRF) cavities. The effort to implement EIC is being done in collaboration with Jefferson Lab, who will build the cavities and send them to BNL to be tested at full power prior to installation. To test the cavities safely and efficiently, two test bunkers are necessary to achieve the desired testing throughput. The bunkers will be designed to shield personnel from radiation produced by the cavities. This design will also include the necessary provisions for radiation-proof doors to allow for easy swapping of the cavities.

## **INTRODUCTION**

The Coherent Electron Cooling (CeC) Time Resolved diagnostic Beamline (TRDBL) is used to ensure that the CeC beam properties are as desired. As the CeC project is still under development, it is necessary to modify and remove the diagnostic capability of the beamline to ensure that the beam's properties are ideal. Several new components were added along the

TRDBL which required mounting provisions and, in some cases, new fixtures to support the new mounts. The new mounts and fixtures needed to be incorporated into the 3D model of the TRDBL and have technical drawings created to supply the necessary information for manufacturing of the components. As of writing this paper, a few the designed components and assemblies have been fabricated.

As part of the ongoing effort to convert the Relativistic Heavy Ion Collider (RHIC) to an Electron-Ion Collider (EIC), over 50 new Superconducting Radio Frequency (SRF) cavities will be necessary to its operation. The cavities, which will be built assembled in their cryomodules at Jefferson Lab, will be sent to Brookhaven National Laboratory (BNL) to be implemented in EIC. Prior to this, it is necessary to test the cavities. To do this safely and efficiently, two radiation-shielding bunkers were designed.

## **SCOPE & OBJECTIVE**

This project consists of two major subprojects. The first is the design of mounting and fixturing for components along the CeC TRDBL. The second is the design of radiation-shielding bunkers for the testing of SRF cavities. Each subproject had numerous constraints.

The CeC TRDBL consists of numerous components with different functionalities and measuring capabilities, mounted on a variety of tables. The TRDBL is shown below:

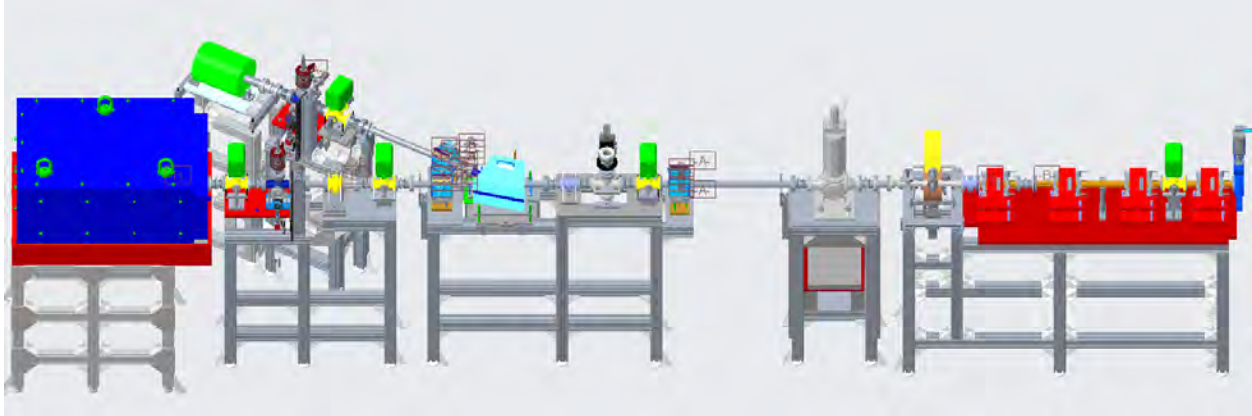


Figure 1: Full View of the Original Design for the CeC TRDBL

One of the main constraints for the design of mounting and fixturing for components along the CeC TRDBL was to minimize the number of new components. In addition to reducing costs, reusing parts greatly decreases the manufacturing and assembly time required to complete the project. An additional constraint that was necessary was to leave a 36” gap between two existing tables to enter the area at the center of the TRDBL. This gap is shown below:

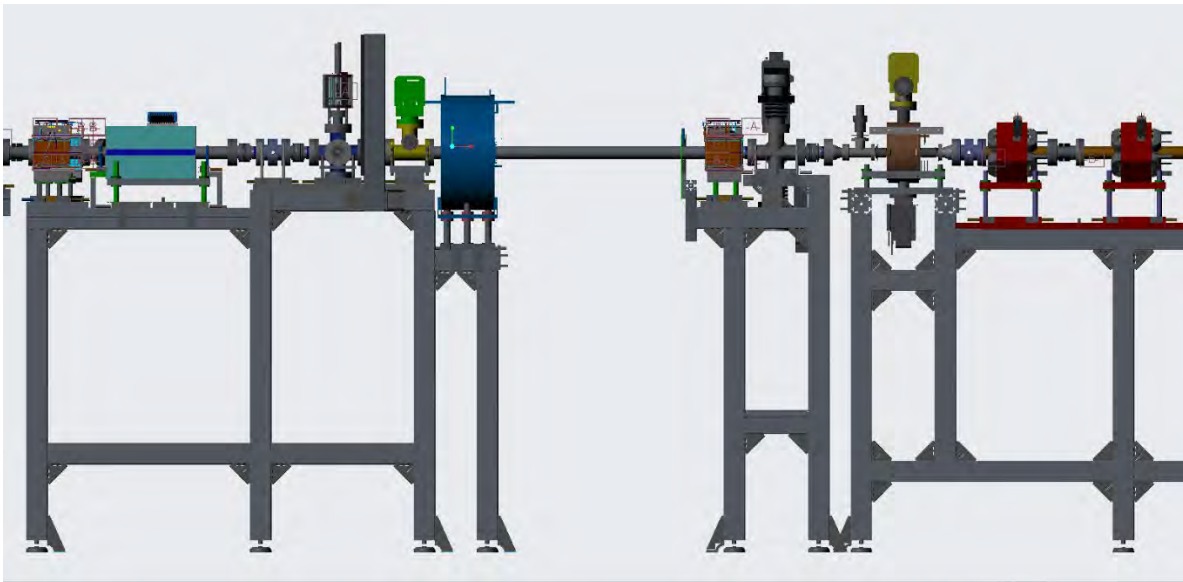


Figure 2: View of the Gap Between the Solenoid and Corrector/TROG Tables

Lastly, components that were introduced along the TRDBL need to be positioned as accurately as possible, to ensure that they are concentric with the beam. This requires all mounting provisions to be designed to allow adjustability for the components laterally and longitudinally to the beamline.

The EIC SRF test bunkers will be built and operated in building 912, where the test bunkers will be placed in the large empty region towards the top right of the image, which is shown below:

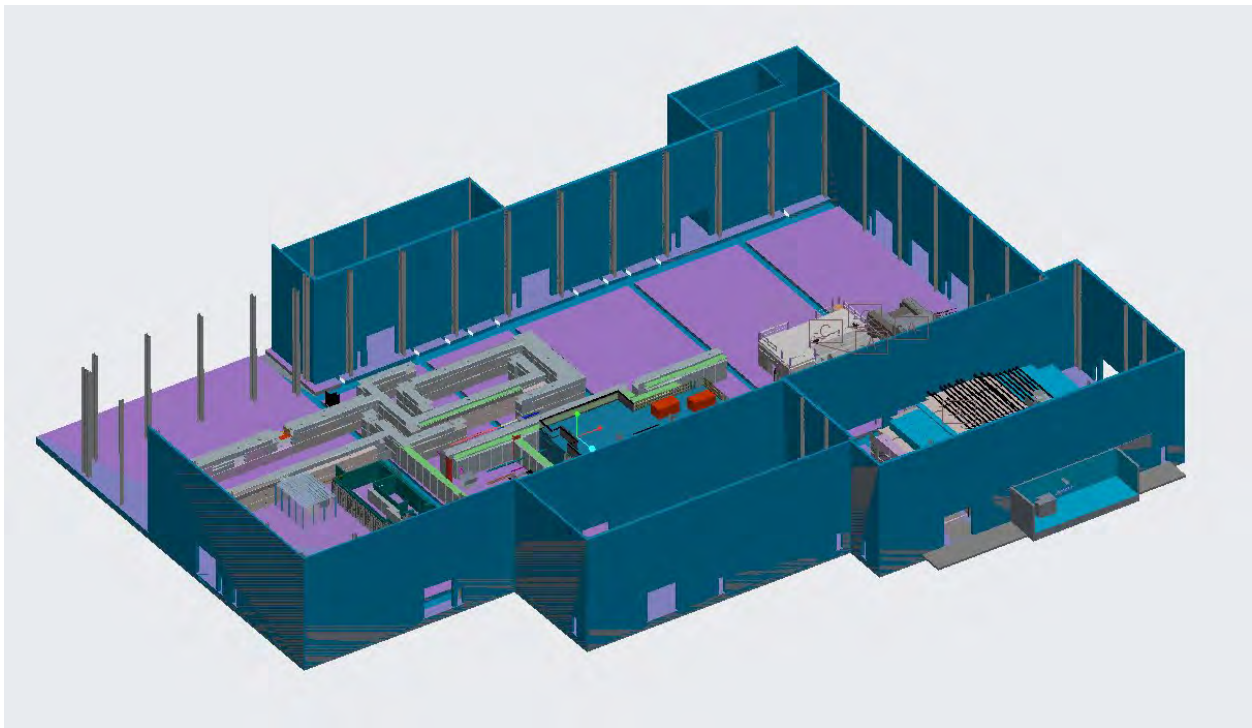


Figure 3: Building 912

The design of the EIC SRF test bunkers had several constraints on it. First, all walls of the bunker needed to be a certain minimum thickness to properly shield personnel during operation of the cavities. Shown below is a table listing the thicknesses of potential shielding materials:

Table 1: Potential Material Thicknesses for Shielding Radiation from Cryomodules[1]

<b>Material</b>	<b>Thickness (in)</b>
Water	168
Heavy Concrete	42
Steel	16
Lead	7.570
Lead, with 2 inches of steel	6.616
Steel, with 4 inches of lead	7.484

An additional constraint is that the bunker structure had to minimize the use of custom shielding blocks to reduce costs, restricting most of the bunker to use the concrete blocks listed below:

Table 2: Quantity of Available Heavy Concrete Blocks

<b>Block Type</b>	<b>Dimensions (in)</b>	<b>Quantity</b>
A	120 x 48 x 54	44
B	60 x 48 x 54	36
E	120 x 48 x 18	19
F	60 x 48 x 18	3
24 FT Roof Block	288 x 48 x 27	114

Due to the limited quantity of each block available, it is also necessary to design the bunkers within the number of available blocks.



Additionally, each bunker needs to have its own large radiation shielding door to allow each bunker to operate independently. The doors are designed to have lead cores with steel shells to minimize the thickness of the door.

The bunkers need to be able to fit cryomodules of which could potentially be as large as 5 meters long, 3 meters wide, and 3 meters tall. The bunkers also need to have sufficient space around the cryomodules to allow personnel to work and move around, and space above to allow waveguides, cable trays, and other components to be routed along the ceiling.

In terms of the placement of the bunkers in the buildings, it is ideal that they be near the trench, which is indicated by the red arrow, while avoiding the trench indicated by the yellow arrow in the image below:

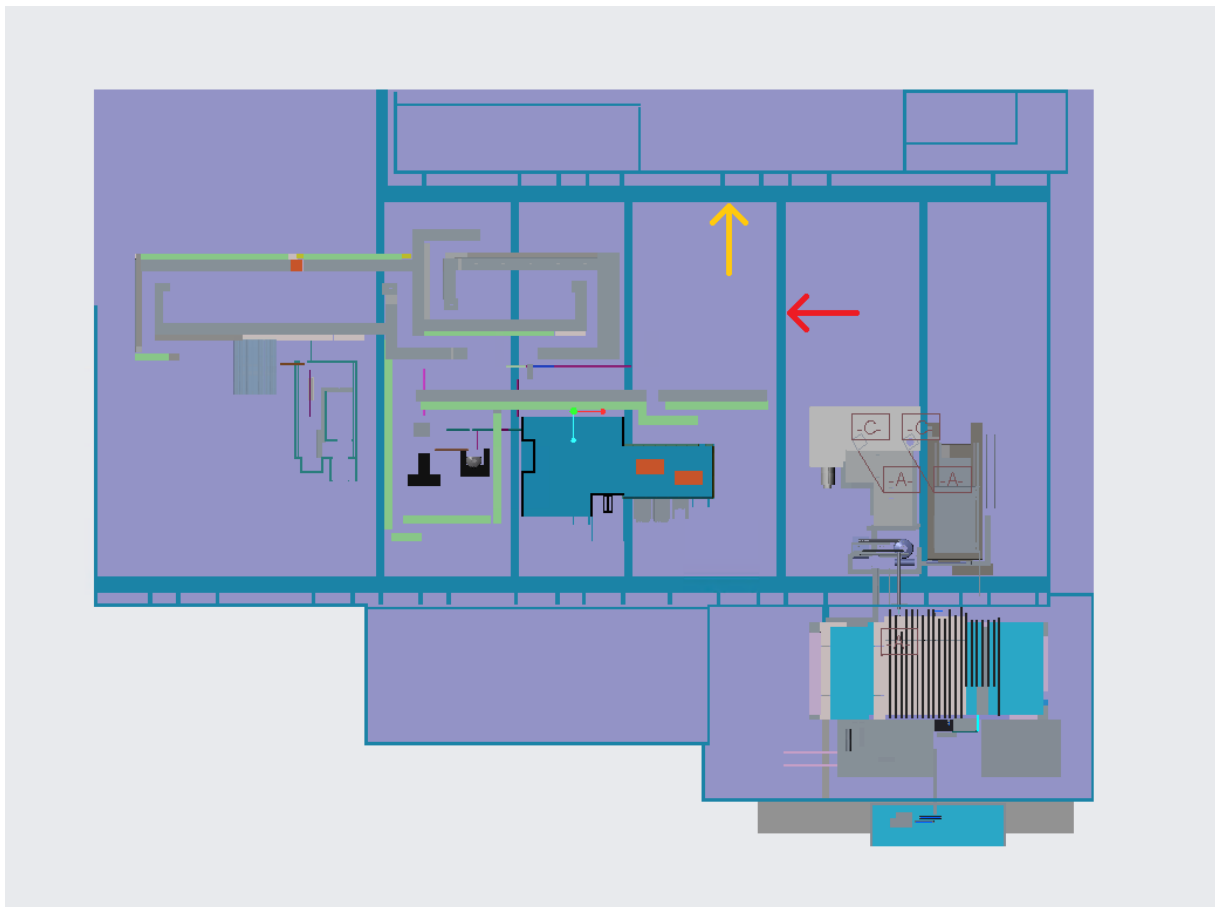


Figure 4: View of Building 912 Showing the Trenches

The proximity to the trench would allow the water and cryo-lines necessary for the operation of the cryomodules to be routed into the bunkers.

Additionally, the bunkers needed to avoid a few existing structures. The first of these are heating, ventilation, and air conditioning (HVAC) pipes which come out of one of the trenches.

These pipes are represented by the green cube in the image below:

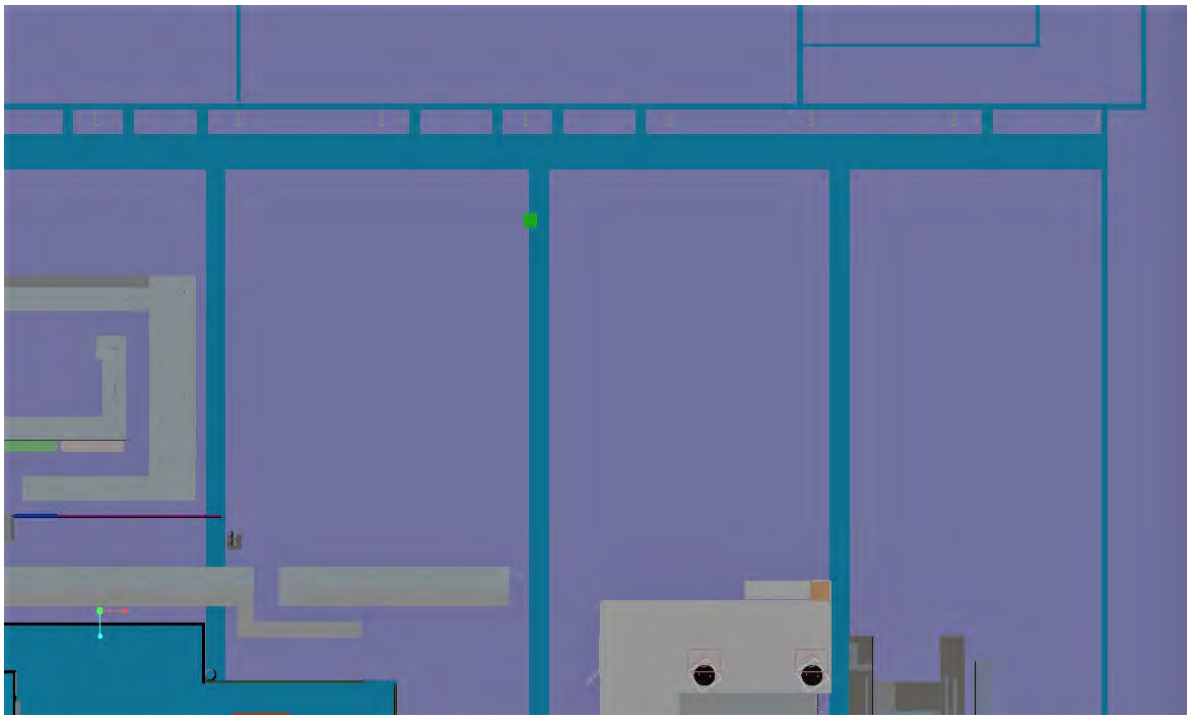


Figure 5: View of Building 912, Showing the HVAC Pipes in Green

The pipes are also shown in the bottom left of the image below:



Figure 6: View of the HVAC Pipes

There is also a fence and pit that are not accounted for in the current 3D model of the building, located in the building as shown inside the fence in the image below:

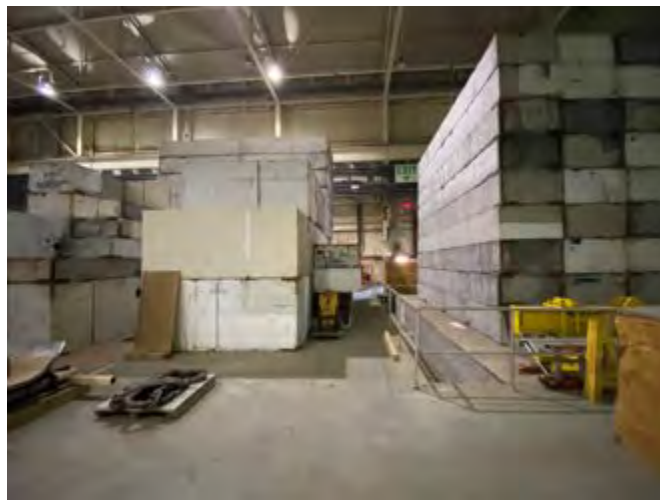


Figure 7: View from Inside Building 912 of Pit

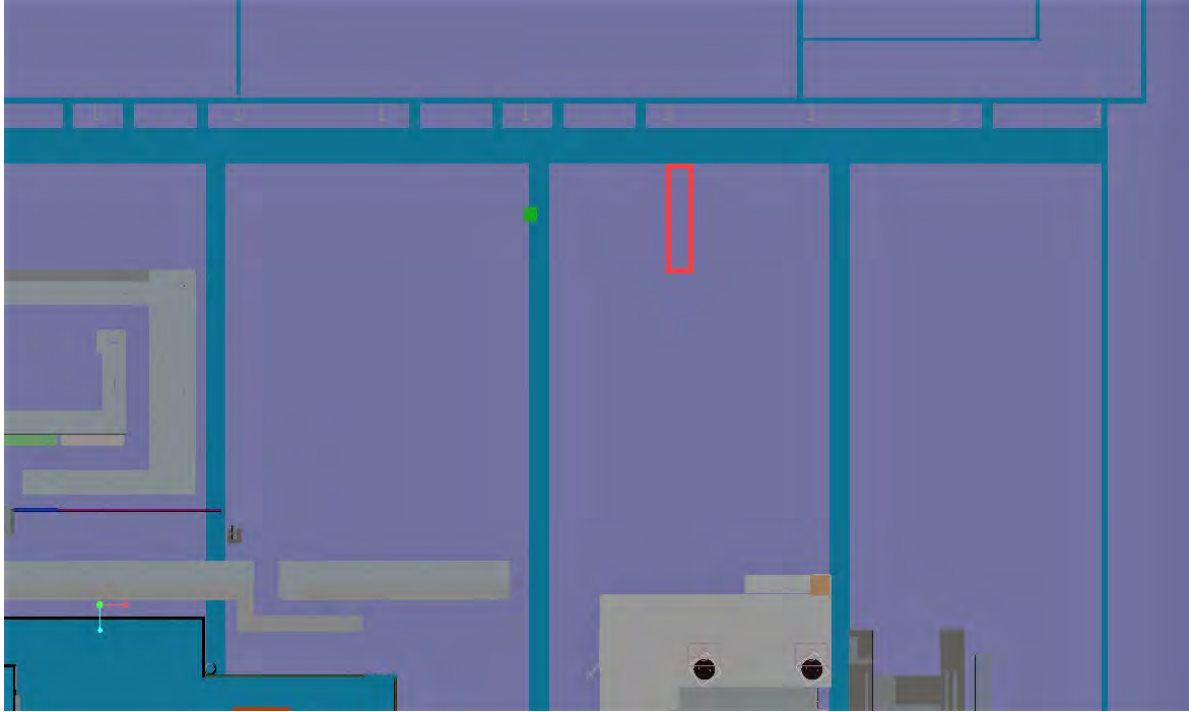


Figure 8: View of Building 912 Showing Approximate Location of Pit

## **METHODS, PROCESS, & RESULTS**

PTC Creo was used as the Computer Aided Design (CAD) software of choice for both projects. Creo was used to create the 3D models of all potential designs and to create drawings for the purposes of fabrication.

The original layout of the TRDBL is shown in Figure 1 in the Introduction. New components were progressively added to main assembly of the TRDBL, starting with a new Horizontal and Vertical (H/V) slit. This change is shown below:

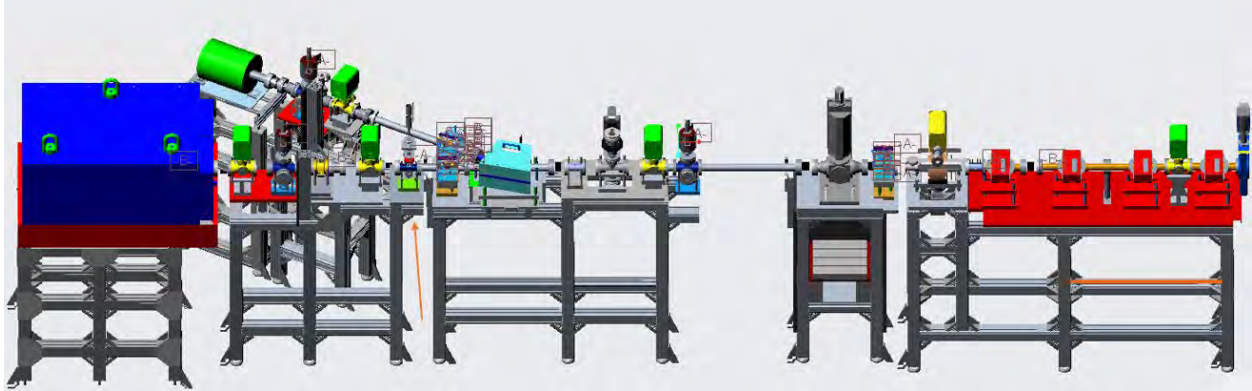


Figure 9: CeC TRDBL with New H/V Slit

The new H/V Slit required a new mounting assembly, and drawings for the assembly and the mounting plate for manufacturing which are shown below:

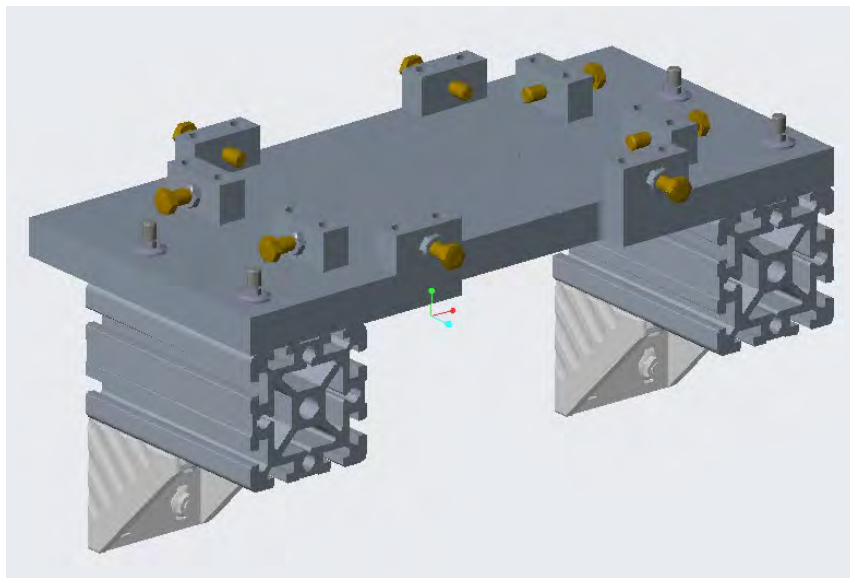


Figure 10: H/V Slit Mounting Assembly

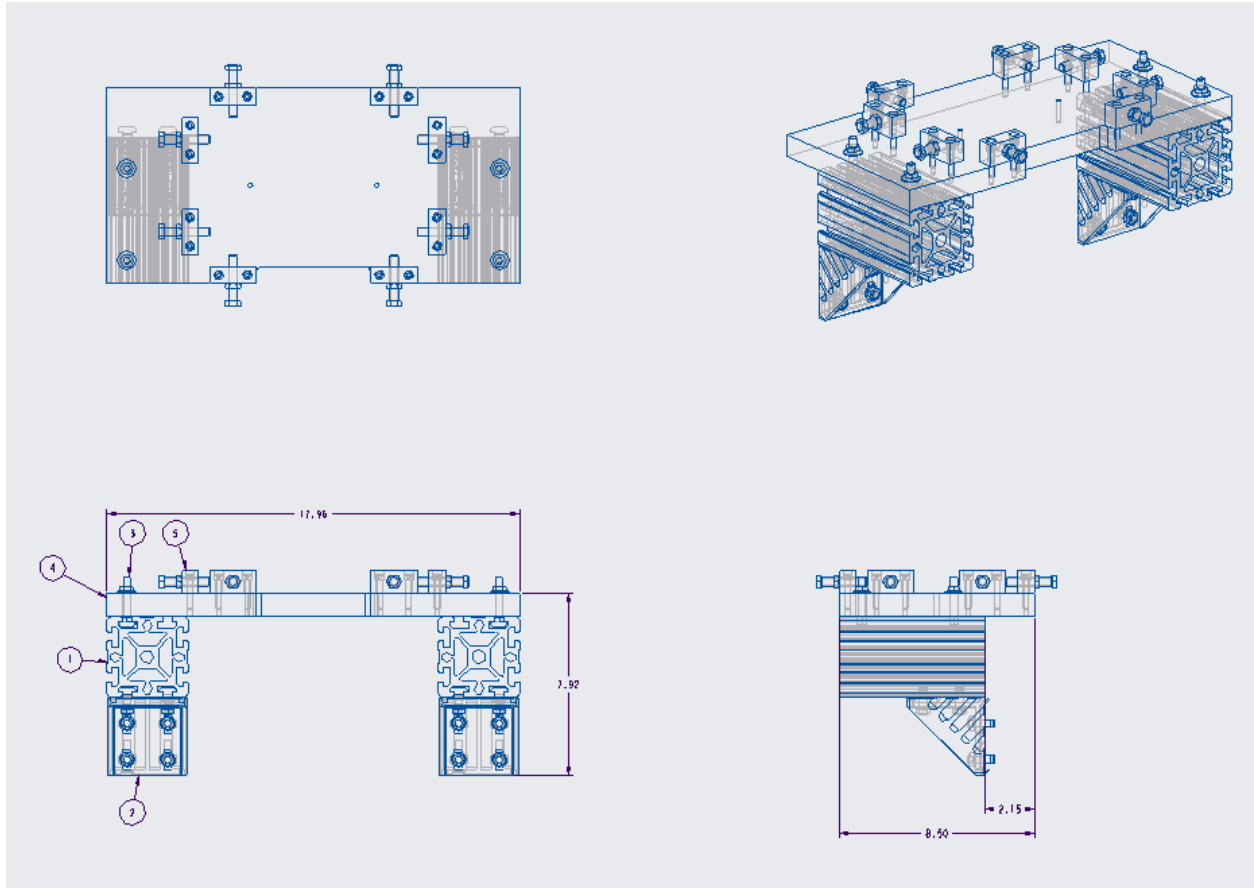


Figure 11: H/V Slit Mounting Assembly Drawing

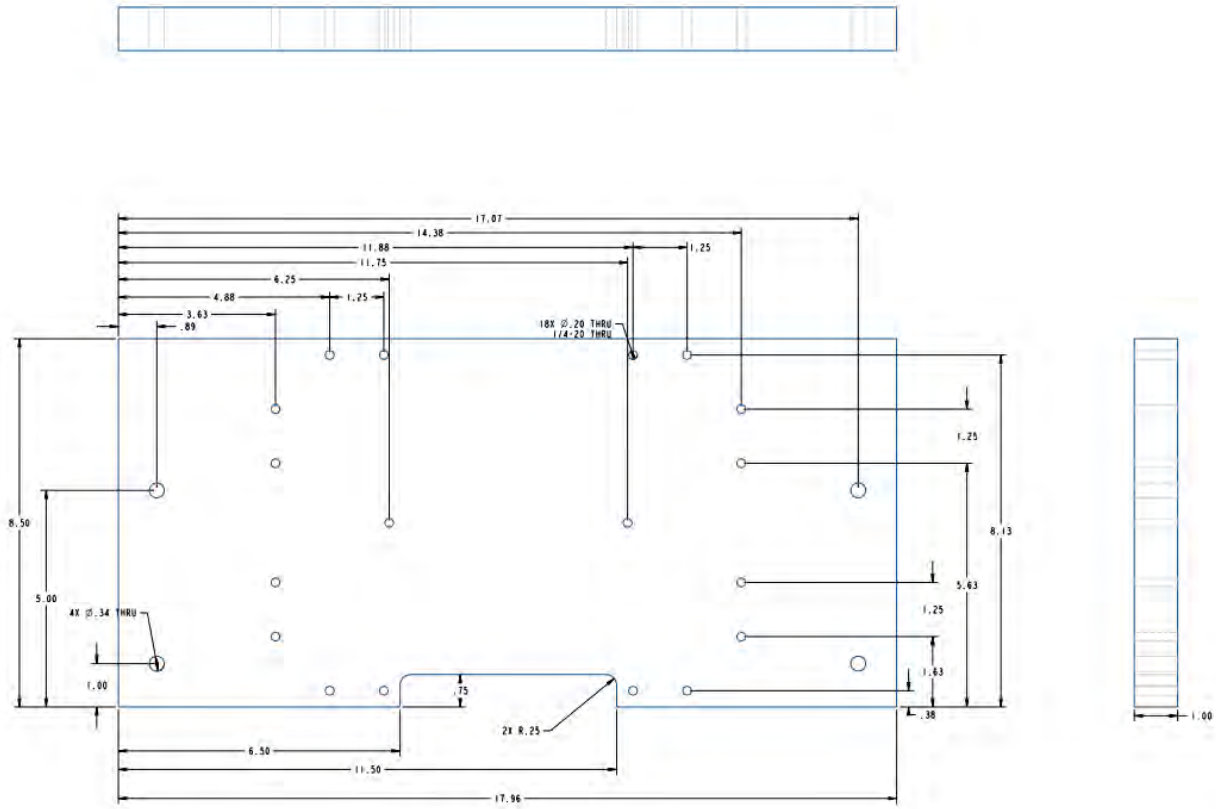


Figure 12: H/V Slit Mounting Plate Drawing

After the addition of the H/V slit, a profile monitor was added with a new mounting plate for it. This addition, as well as the drawing of the new mounting plate are shown below:

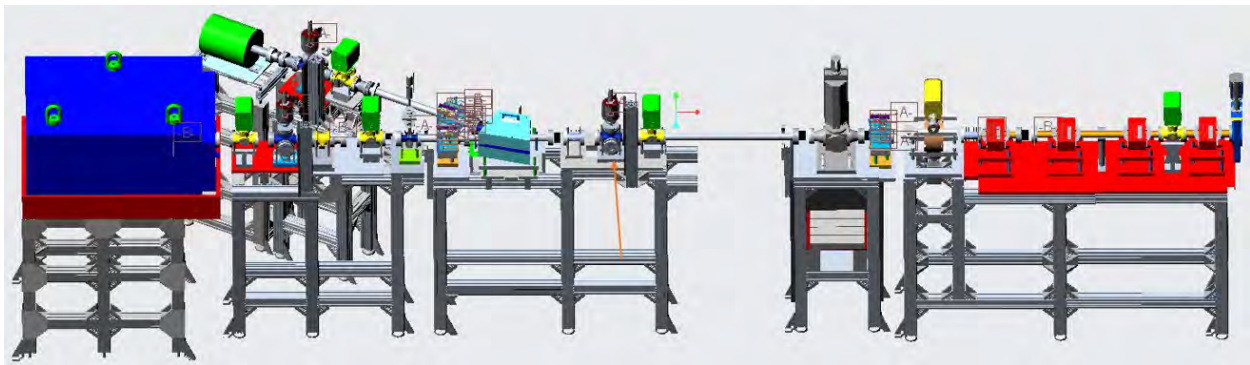


Figure 13: CeC TRDBL with New Profile Monitor



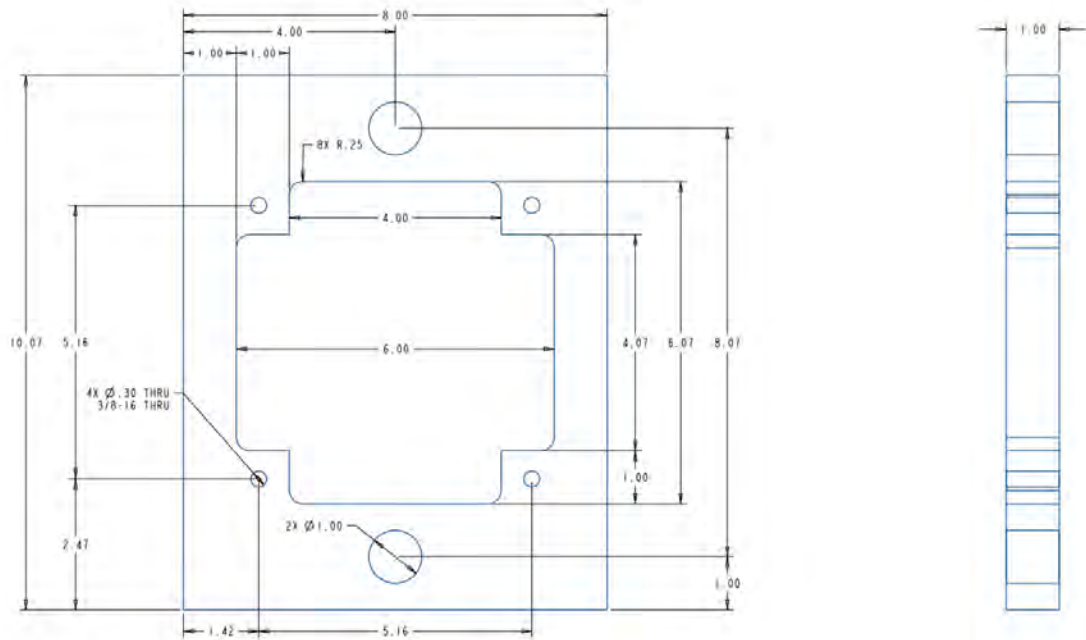


Figure 14: Profile Monitor Mounting Plate

Afterwards, a new solenoid was placed upstream of the profile monitor with new legs to support it. The addition of the solenoid, as well as a drawing of the plate, are shown below:

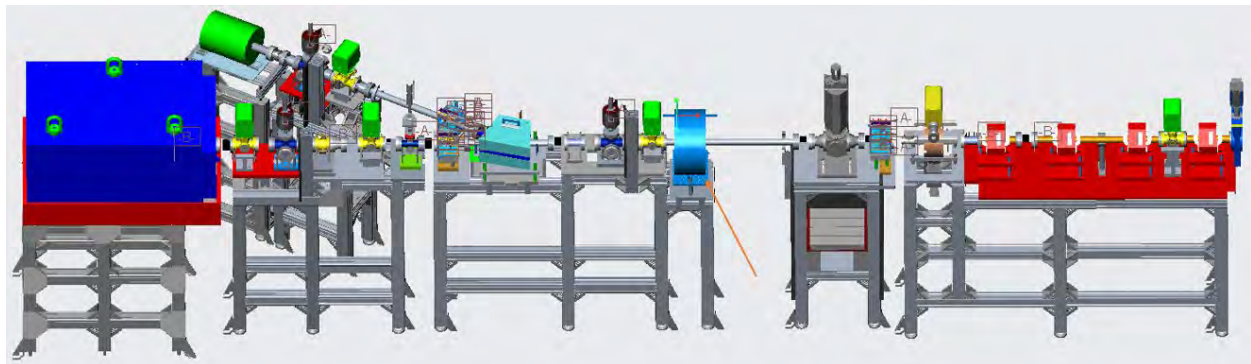


Figure 15: CeC TRDBL with New Solenoid and Mounting Stand



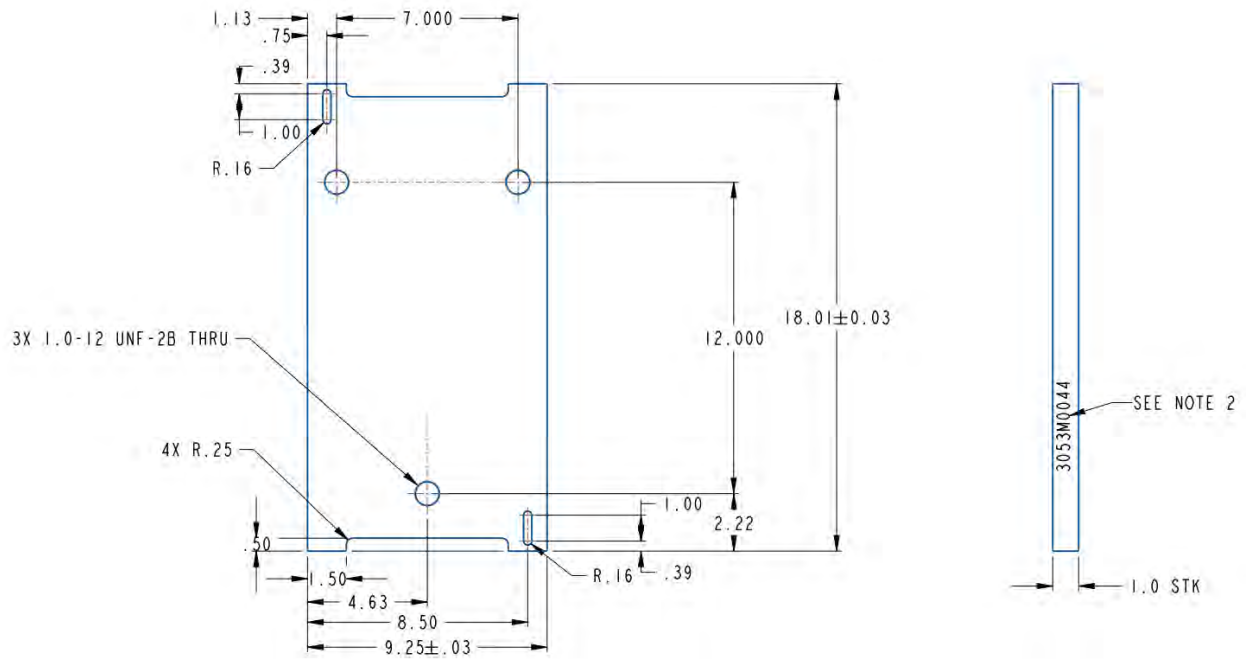


Figure 16: Solenoid Mounting Plate

The finished mounting plates for the solenoid and the H/V slit are shown below:



Figure 17: Manufactured Solenoid Plate (Left) and H/V Slit Plate (Right)

A new table had to be made to replace the existing table upstream of the solenoid, which is the table to the right of the solenoid in Figure 15. This table had to incorporate the addition of a Temporally Resolved Optical Gating monitor (TROG) and a new corrector, as shown below:

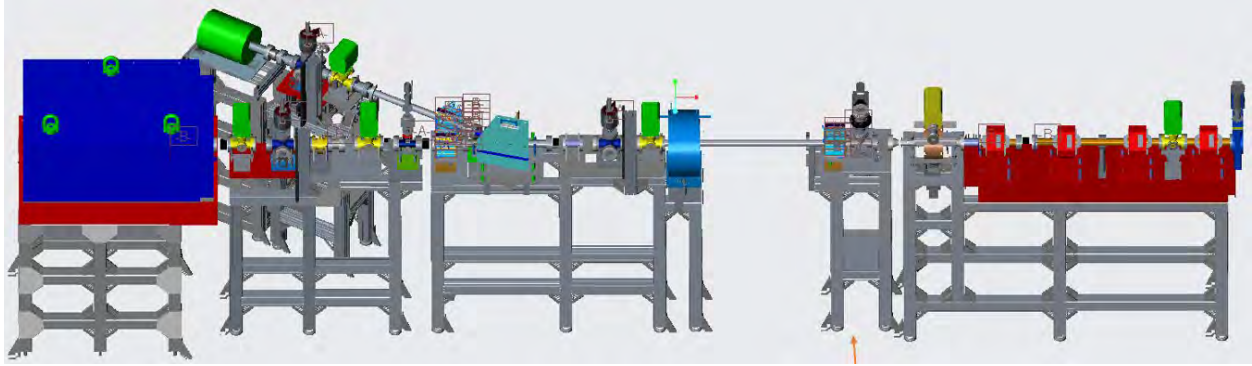


Figure 18: CeC TRDBL with New TROG, Corrector, and Mounting Table

Several measures had to be taken to maximize the space between the solenoid mounting table and the new table. The gap between these two tables, which is required to be at least 36 inches to allow personnel to safely go between the two, was previously about 28 inches. The first change was decreasing the width of the solenoid table. Furthermore, the mounting of the corrector and TROG were vertically offset allowing the TROG's mounting to slightly overlap the solenoid as shown below:

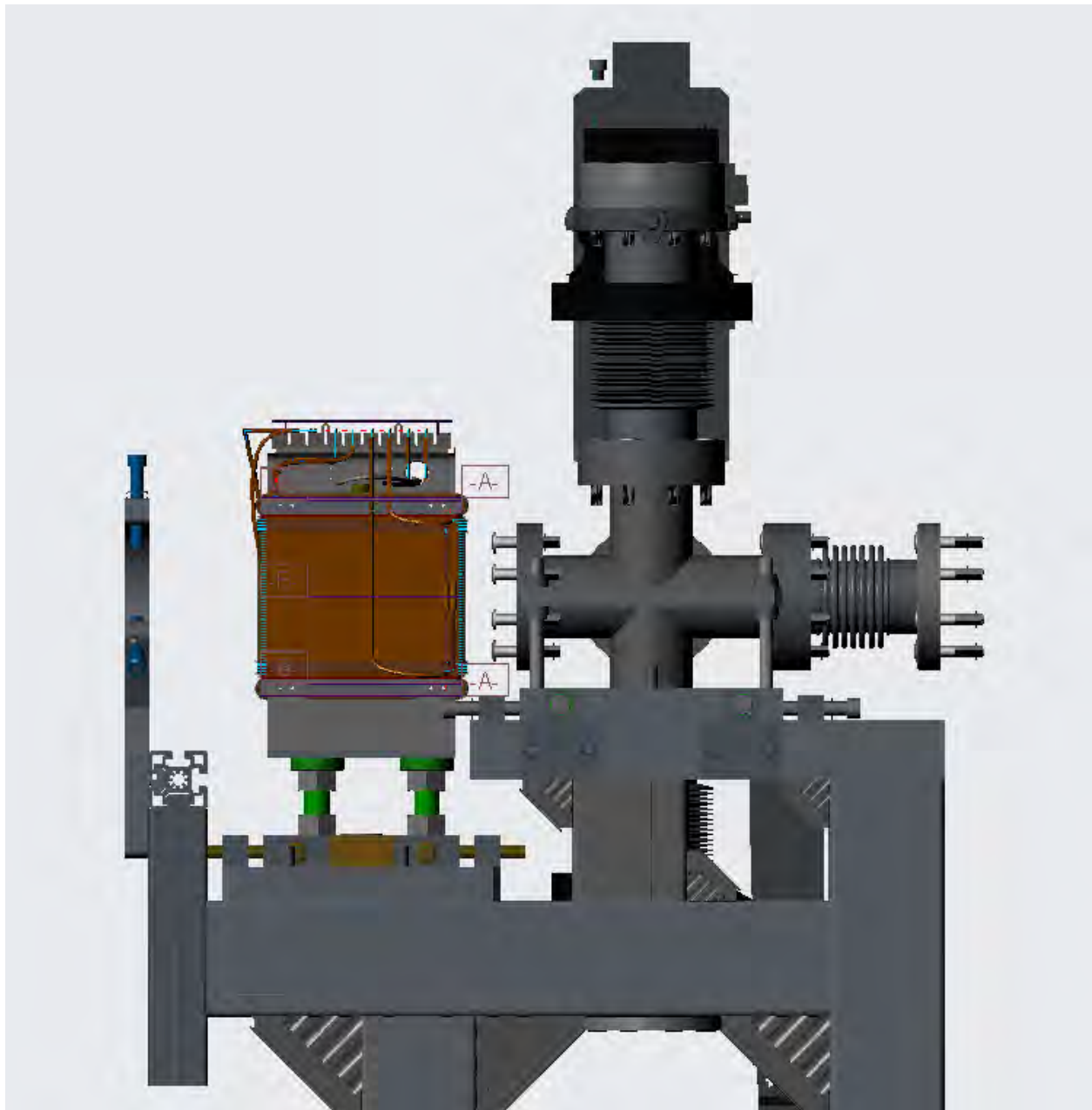


Figure 19: Side View of the Corrector (Left) and TROG (Right) Mounting

The final change for CEC was the addition of a Direct Current Current Transformer (DCCT) and the necessary mounting. The DCCT was put in to replace a 500 MHz cavity, as shown below:



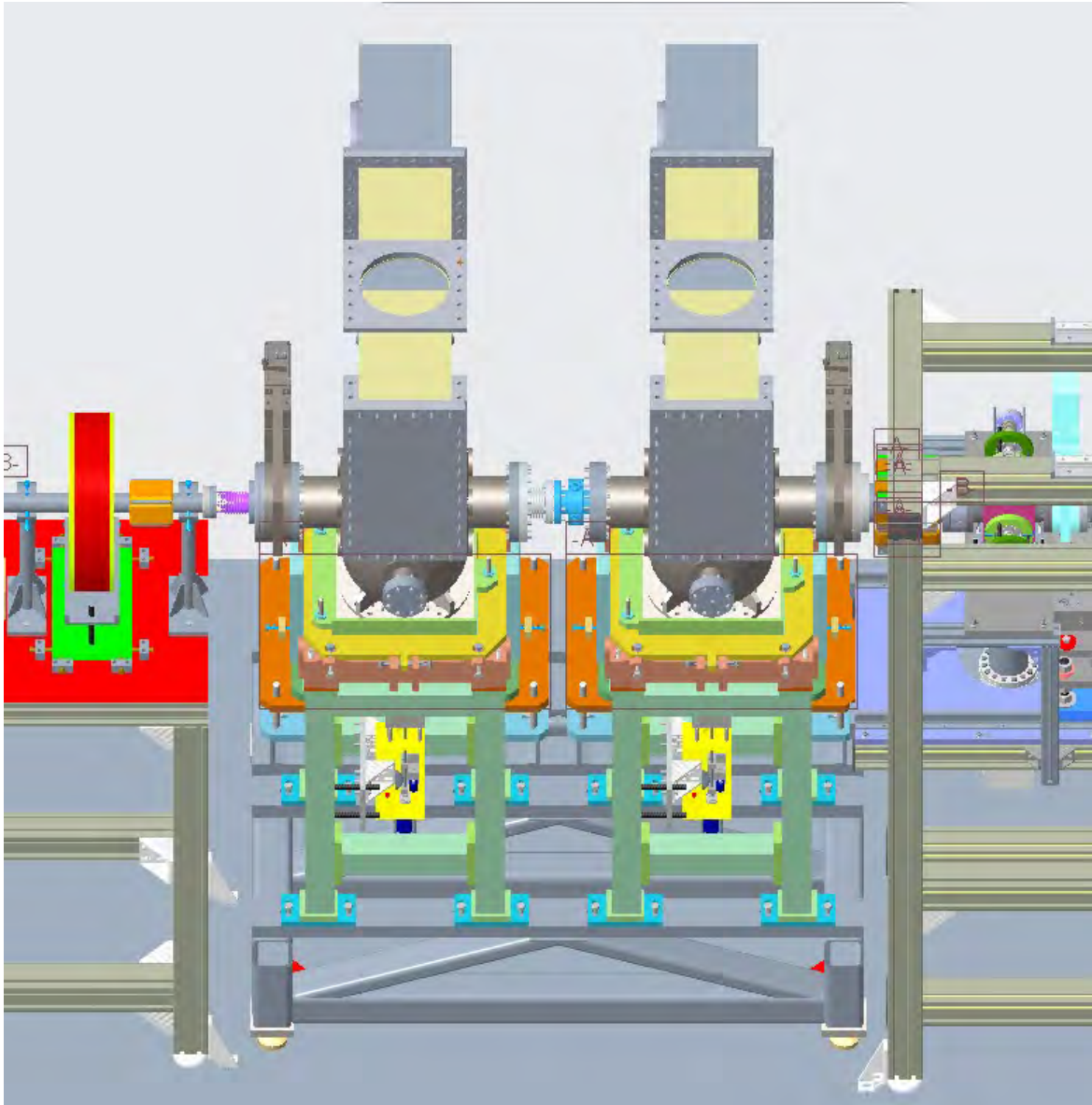


Figure 20: View of the Two Cavities

Behind the cavities, there are existing mounting plates used for a pump. The pump's mounting assembly was reused to minimize the number of new components and assemblies. The proposal for this, with the DCCT replacing the cavity, is shown below:

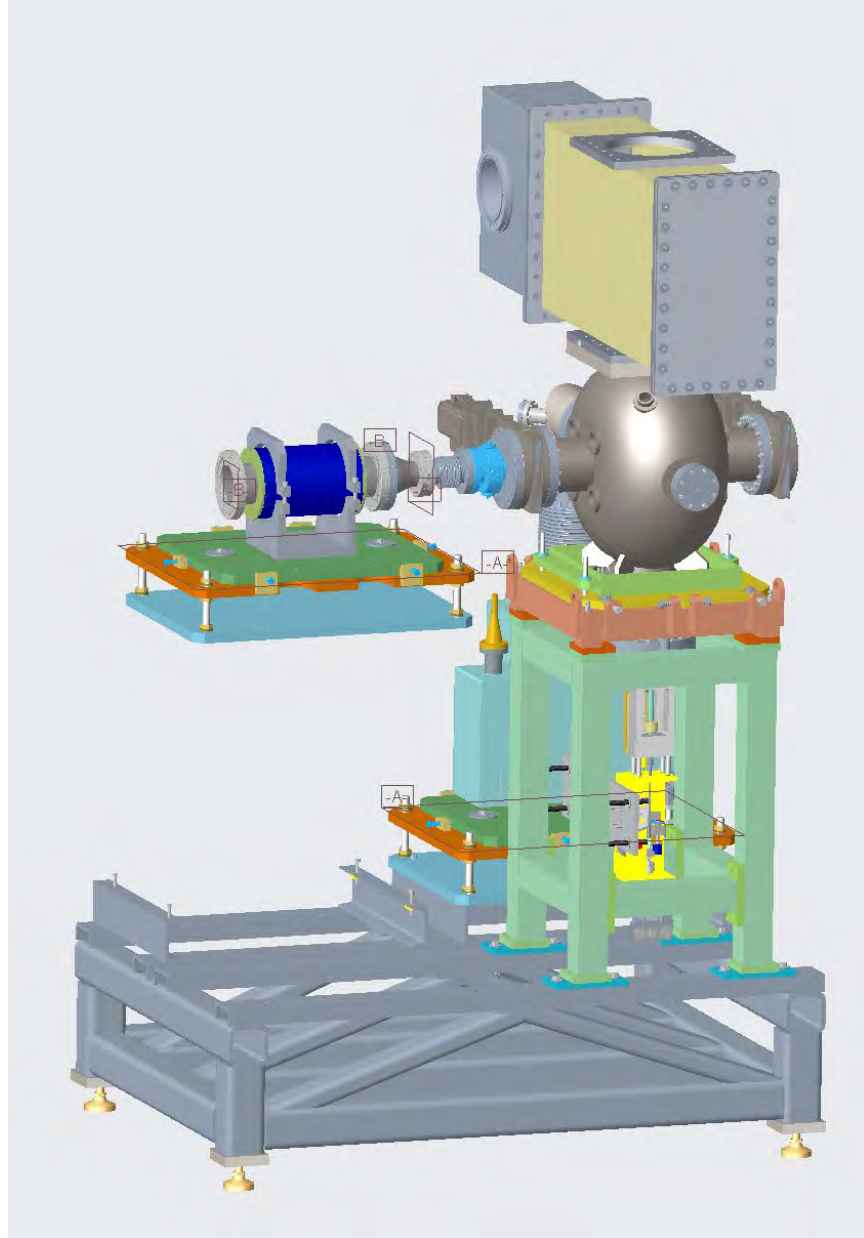


Figure 21: View of the DCCT Replacing a Cavity

This placement of the DCCT required a new table to be made. The model is shown below, as well as an imagine of the manufactured and assembled table:

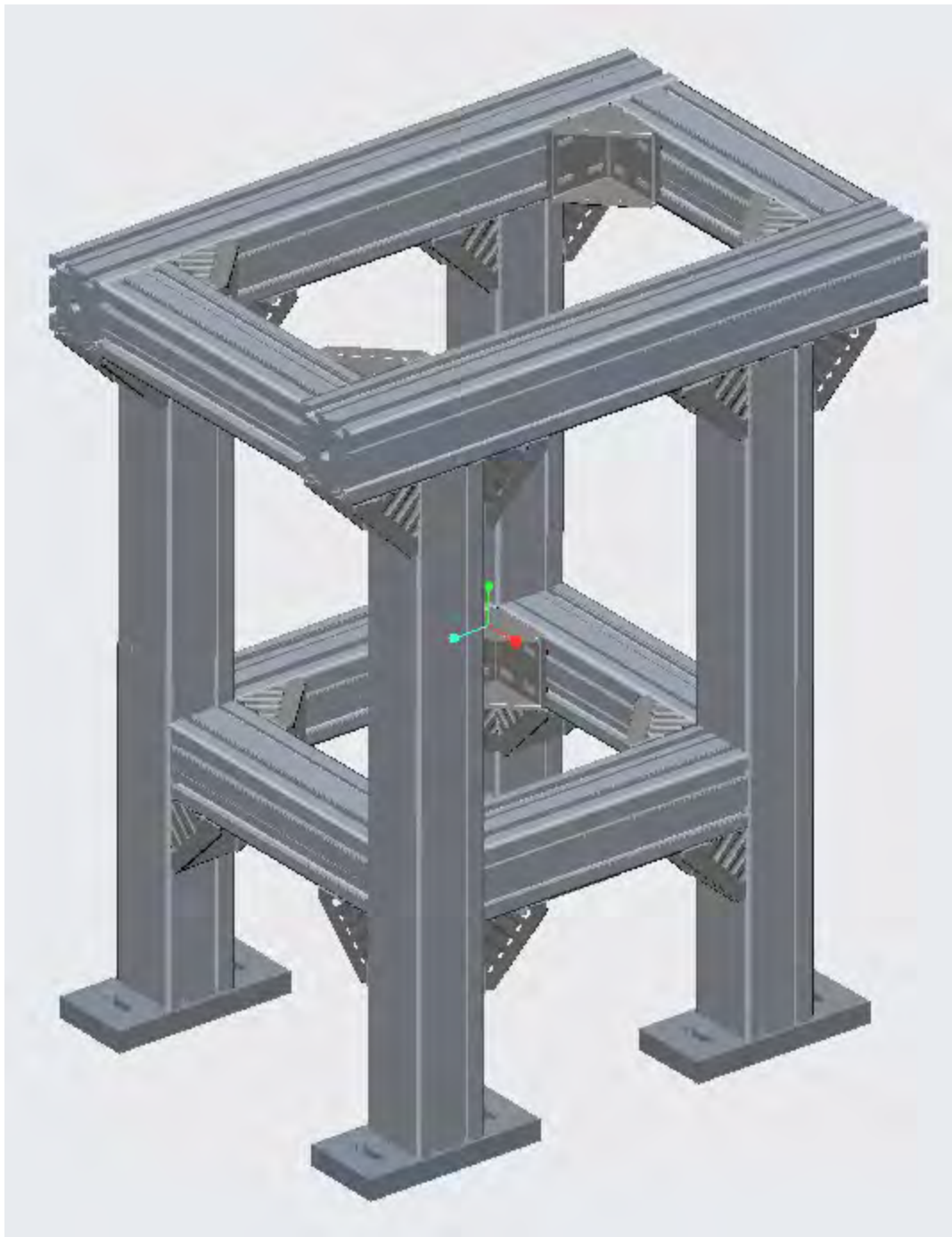


Figure 22: New DCCT Mounting Stand



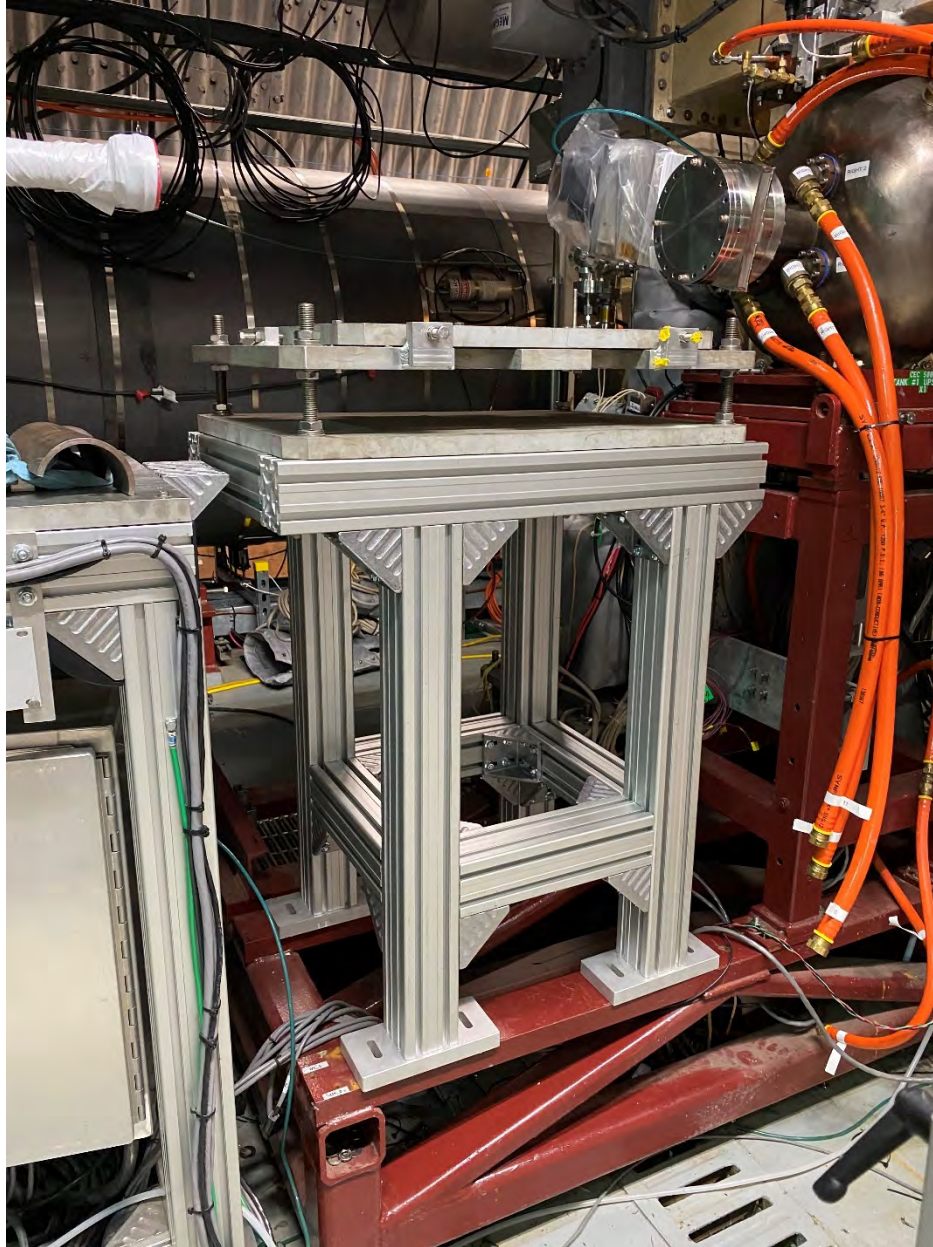


Figure 23: New DCCT Mounting Stand, Fabricated and Installed

Initially, two layouts for the test bunkers were proposed. As more requirements, constraints, and information was obtained, further iterations were created to accommodate everything accordingly. Several of the desired improvements were made based on suggestions provided on bunker layouts that were presented at meetings attended by various engineers and



physicists also working on EIC. This process of iterating on each layout resulted in a total of 6 layouts, with the sixth layout being the ideal configuration as of writing this paper.

The first layout featured four bunkers with all four of their main doors parallel. Each bunker was independent of one another but were grouped in two pairs that were centered around two trenches of the building. Various views of the layouts are shown below:

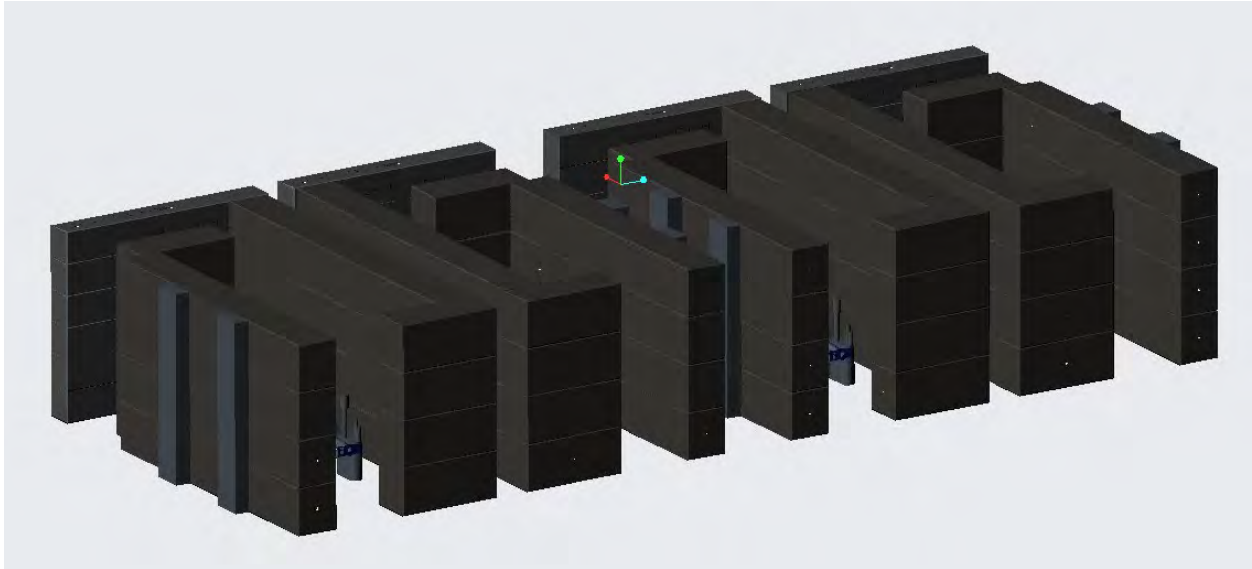


Figure 24: Test Bunker Layout 1, No Roof

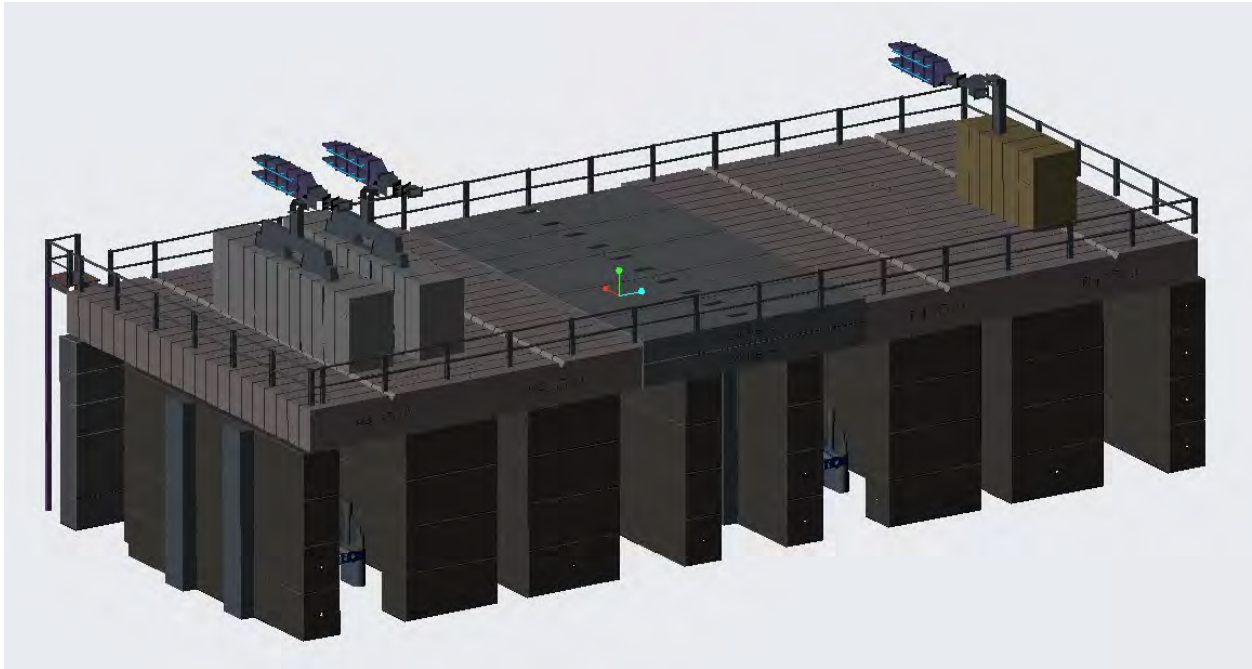


Figure 25: Test Bunker Layout 1, With Roof

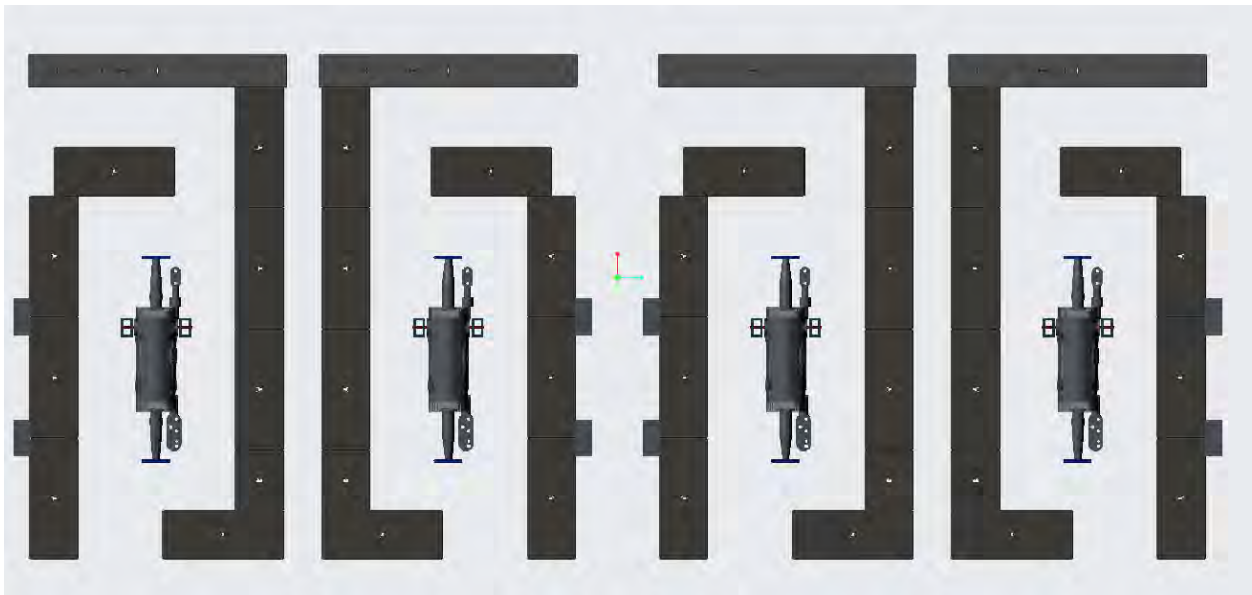


Figure 26: Test Bunker Layout 1, Top View

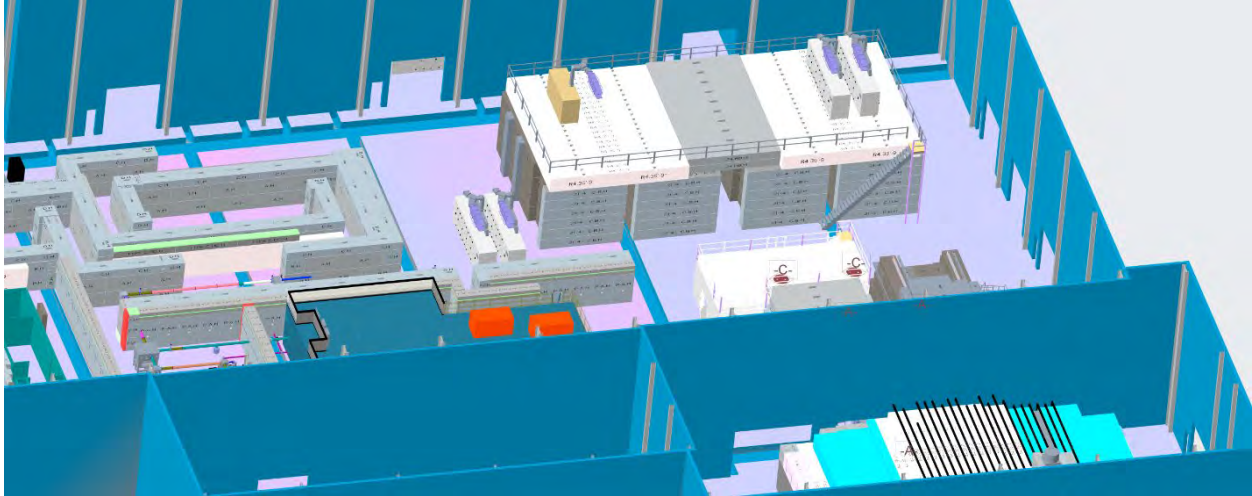


Figure 27: Test Bunker Layout Inside Building 912

The second layout featured four bunkers, but with each pair sharing a common wall and both pairs also featuring a common wall. This significantly reduced in the number of building blocks used. Various views of layout 2 are shown below:

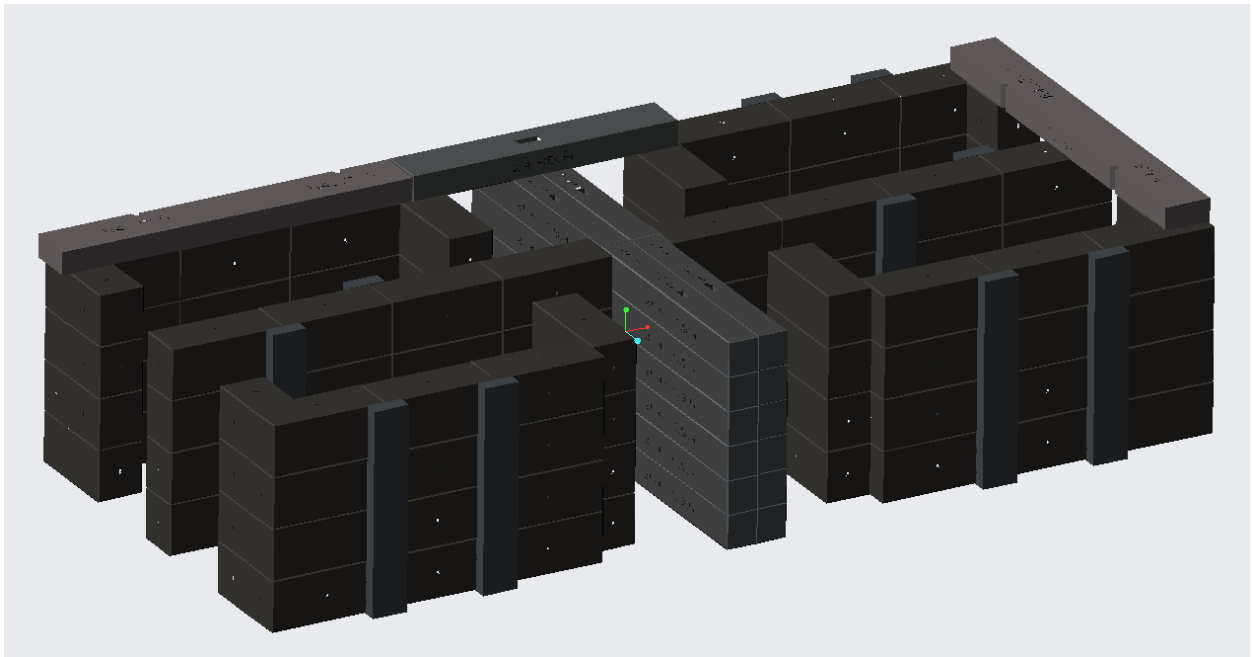


Figure 28: Test Bunker Layout 2

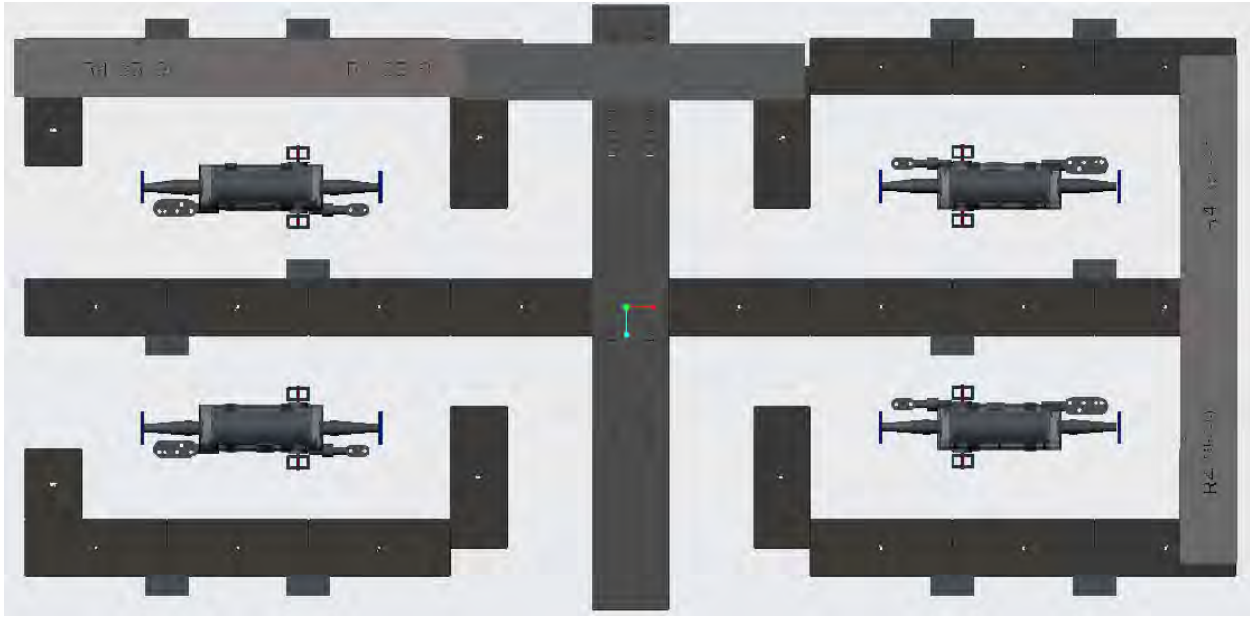


Figure 29: Test Bunker Layout 2, Top View

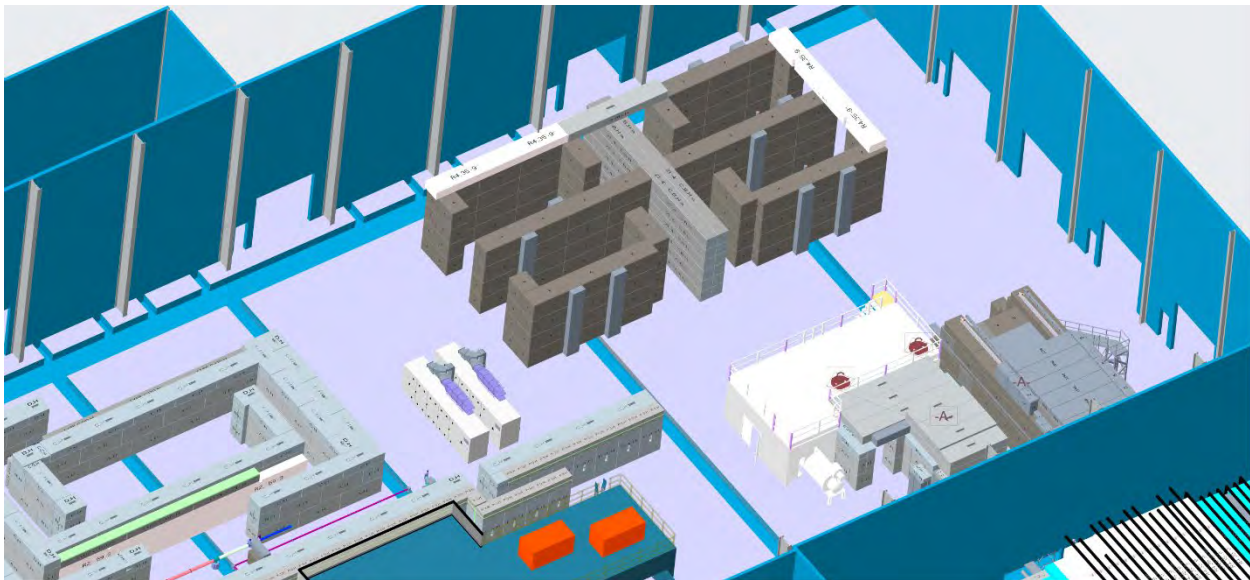


Figure 30: Test Bunker Layout Inside Building 912

For each layout, it is necessary to supply the necessary power to the cryomodules. This is done using two power amplifiers. Layouts 1 and 2 were made with the intent of having said

amplifiers located on their roofs. These roofs would have staircases leading to the tops and have railings around the perimeters for the safety of any personnel working on top of the roofs.

However, in the first of the aforementioned meetings, it was determined that the height of the power amplifiers and the components/waveguides on top of them would present difficulties as these components would need to be placed on the roof through the use of an existing crane with a limited working height. As a result, it was decided that the power amplifiers need to be located on the building floor, adjacent to the bunkers.

In the first meeting it was also determined that only two bunkers would be necessary, that the cryomodules could be larger than reflected in the model, and that the bunkers would have to be turned 180 such that the main doors would face the inside of the building. This would allow for more space to bring the cryomodules in and out. These additional requirements resulted in the creation of bunker layouts 3 and 4.

Layout 3 of the test bunkers had only two bunkers, which were slightly larger than those from layout 1 and 2. The bunkers in layout 1 had an internal size of 156 by 420 inches, whereas layout 3 has a size of 172 by 480 inches. Additionally, the roof was simplified since the amplifiers didn't need to be placed on the roof. Layout 3 is shown below:



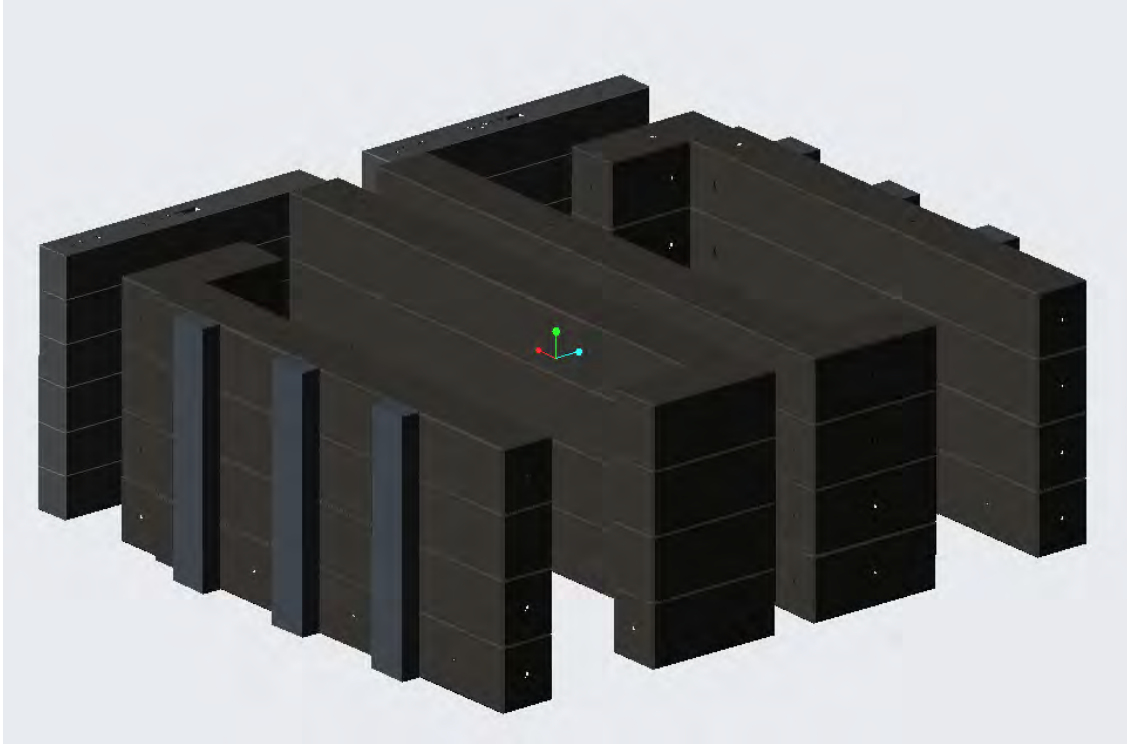


Figure 31: Test Bunker Layout 3, No Roof

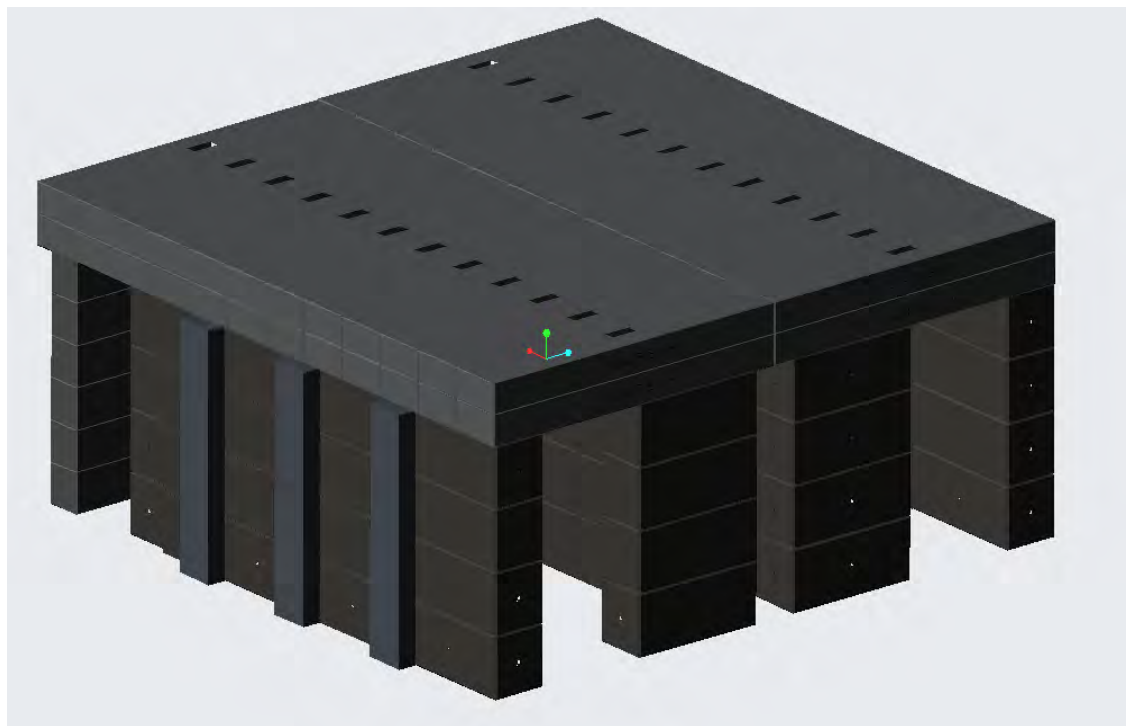


Figure 32: Test Bunker Layout 3, With Roof

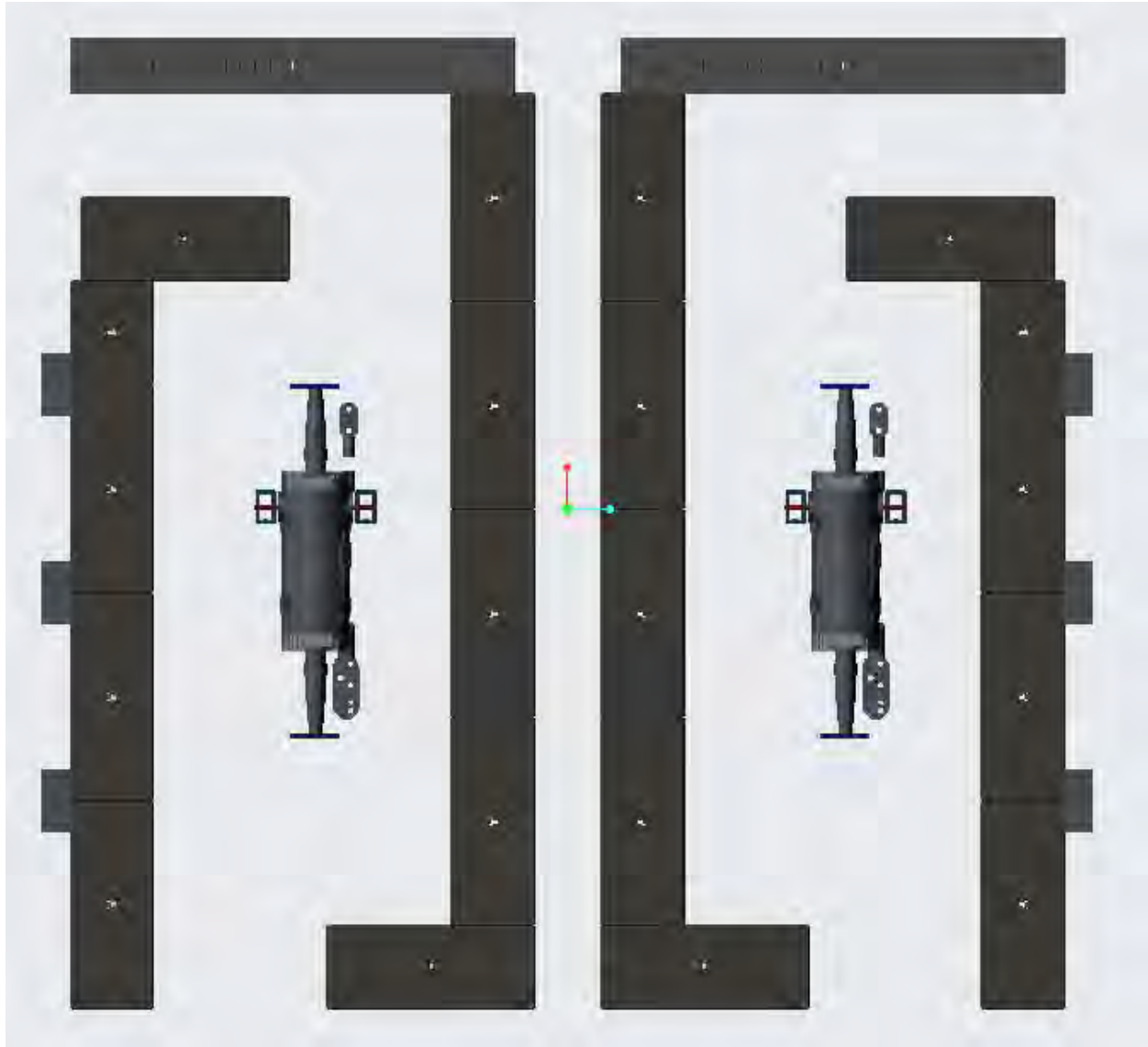


Figure 33: Test Bunker Layout 3, Top View

Layout 4 only had small differences from layout 3, except that the bunkers are slightly wider, further apart, and have slightly different blocks. This is shown below:

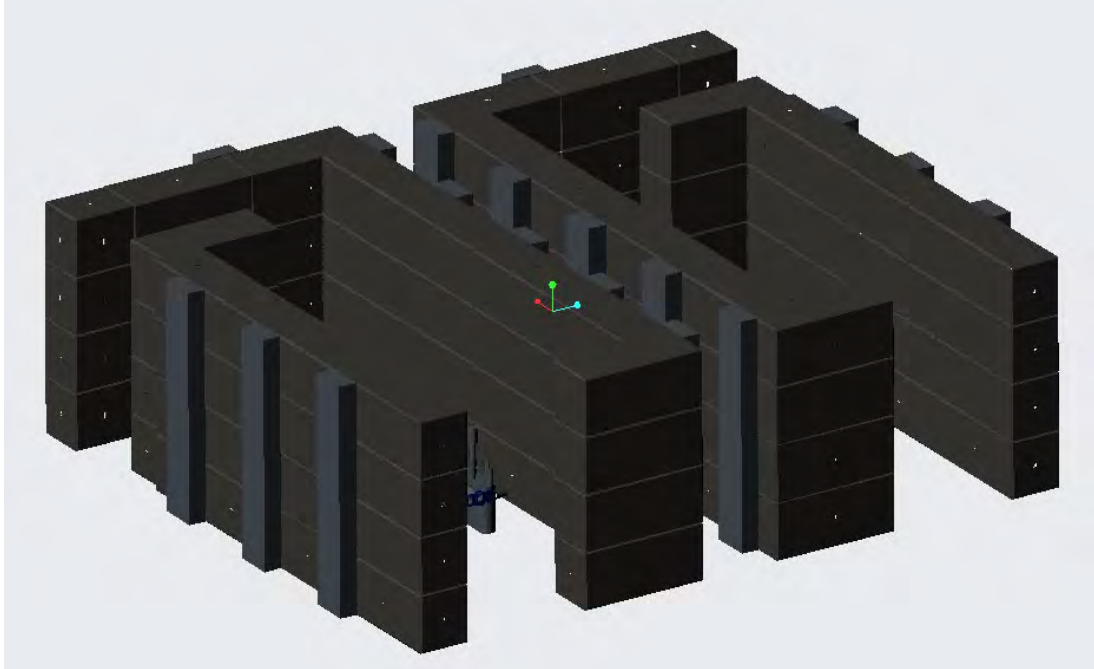


Figure 34: Test Bunker Layout 4, No Roof

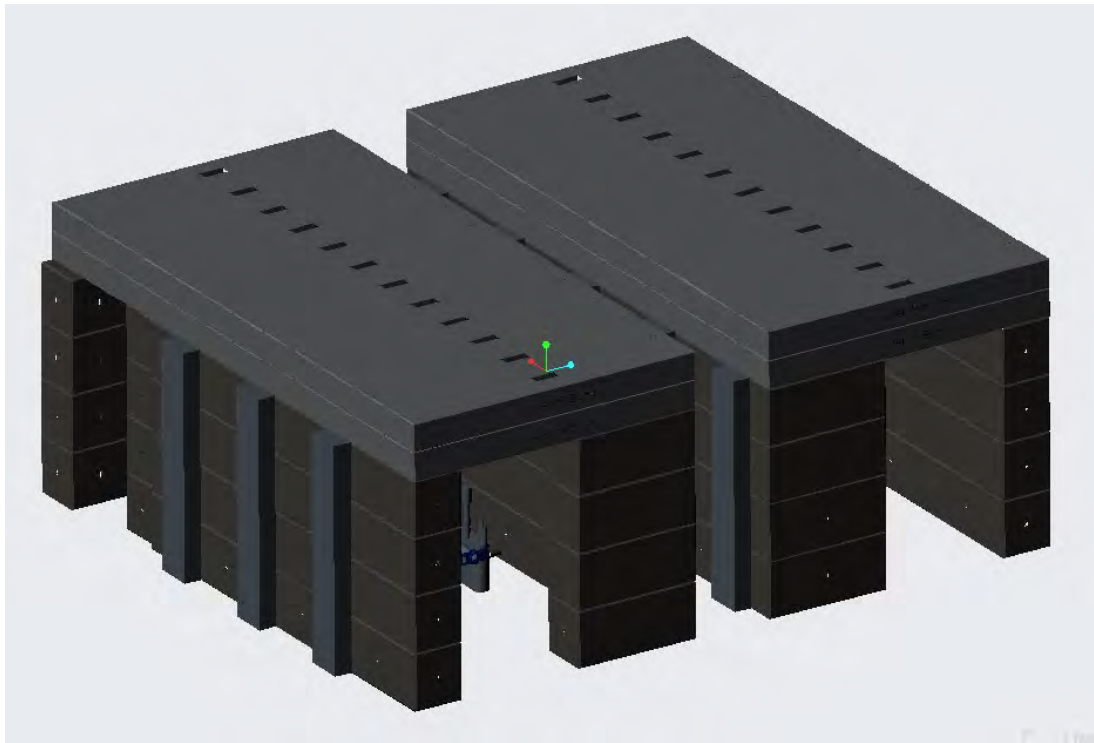


Figure 35: Test Bunker Layout 4, With Roof



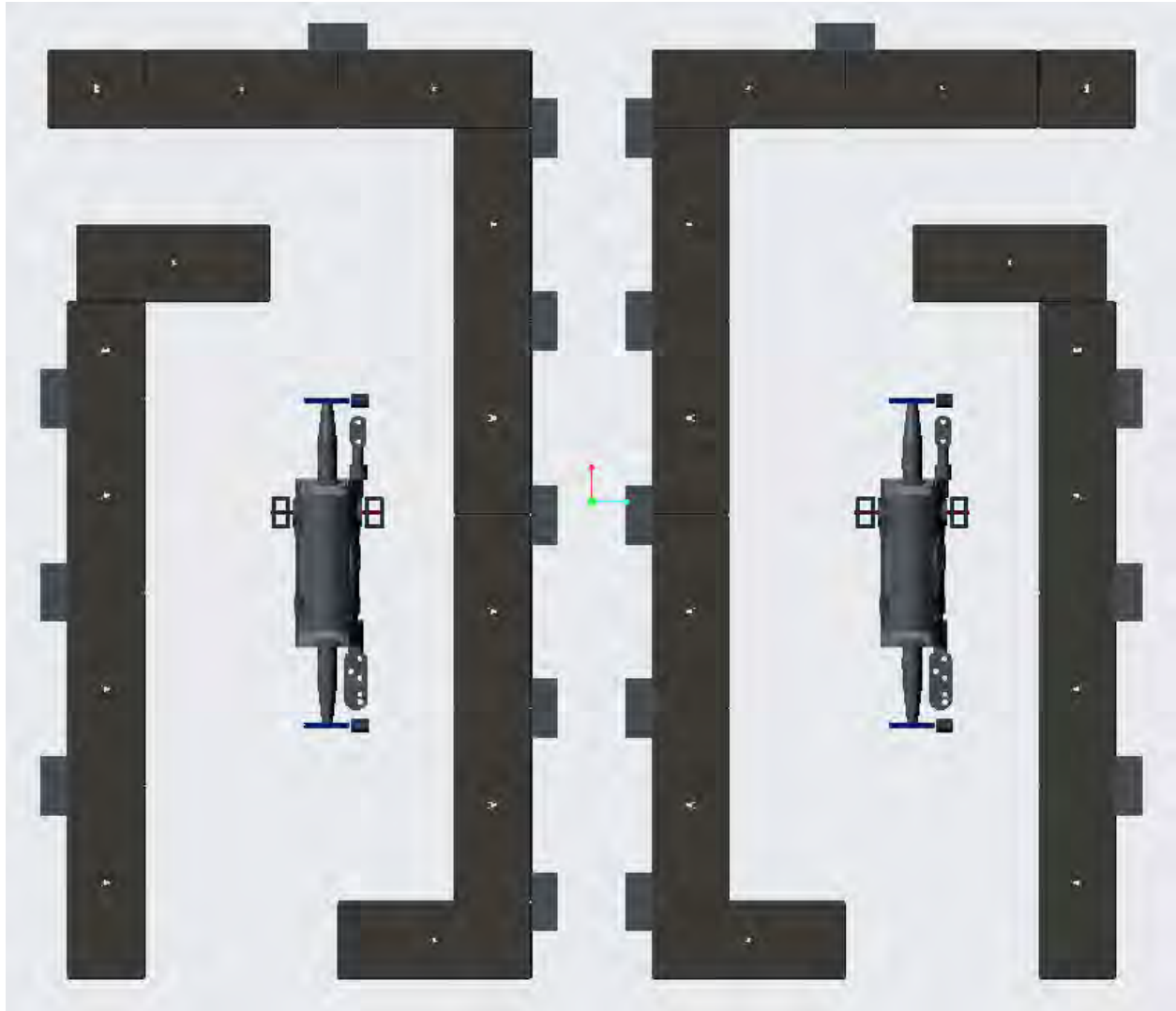


Figure 36: Test Bunker Layout 4, Top View

The biggest change from layout 4 to layout 5 was that the roof was lowered from 18 feet to 15 feet, as the extra height was not necessary. This cut down on the number of A-blocks that were necessary. In addition, the roof blocks by the main door were lowered to provide a mounting area for the doors. Layout 5 is shown below:

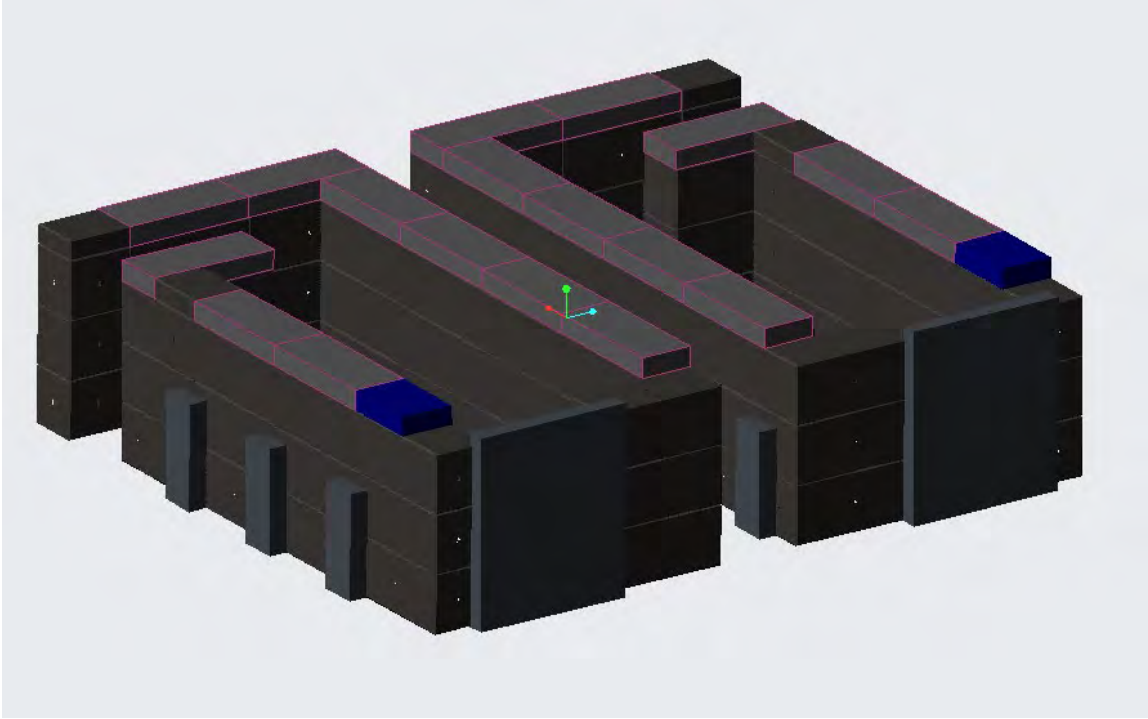


Figure 37: Test Bunker Layout 5, No Roof

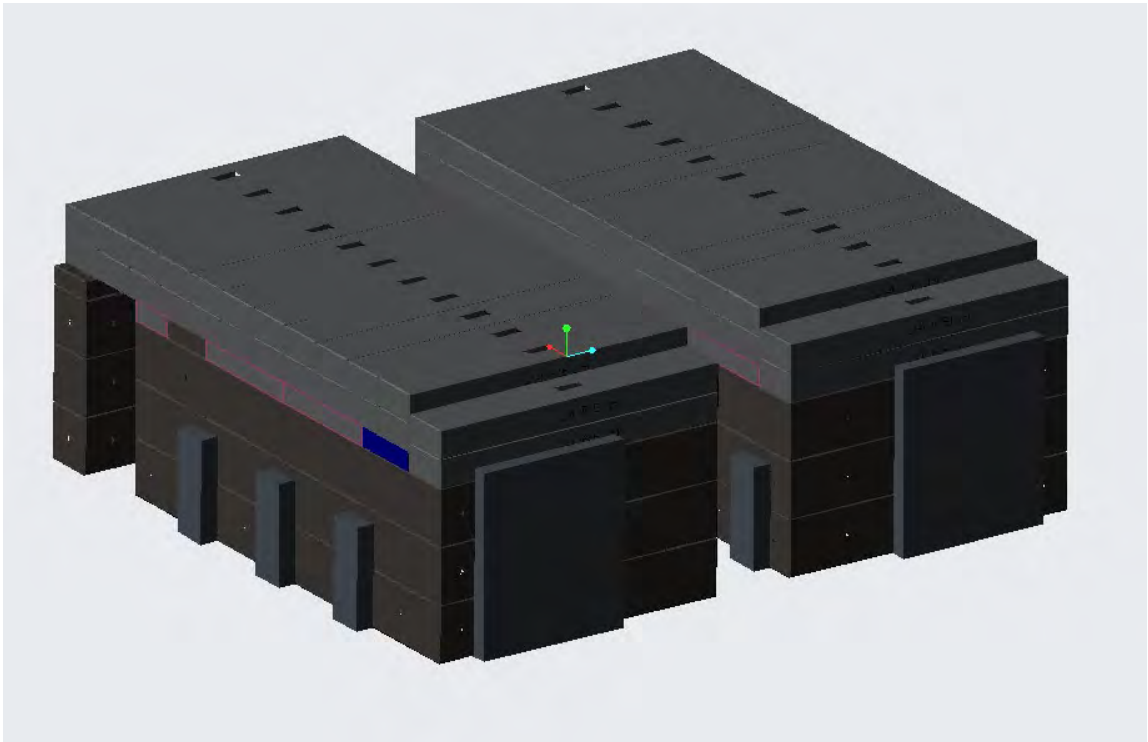


Figure 38: Test Bunker Layout 5, With Roof

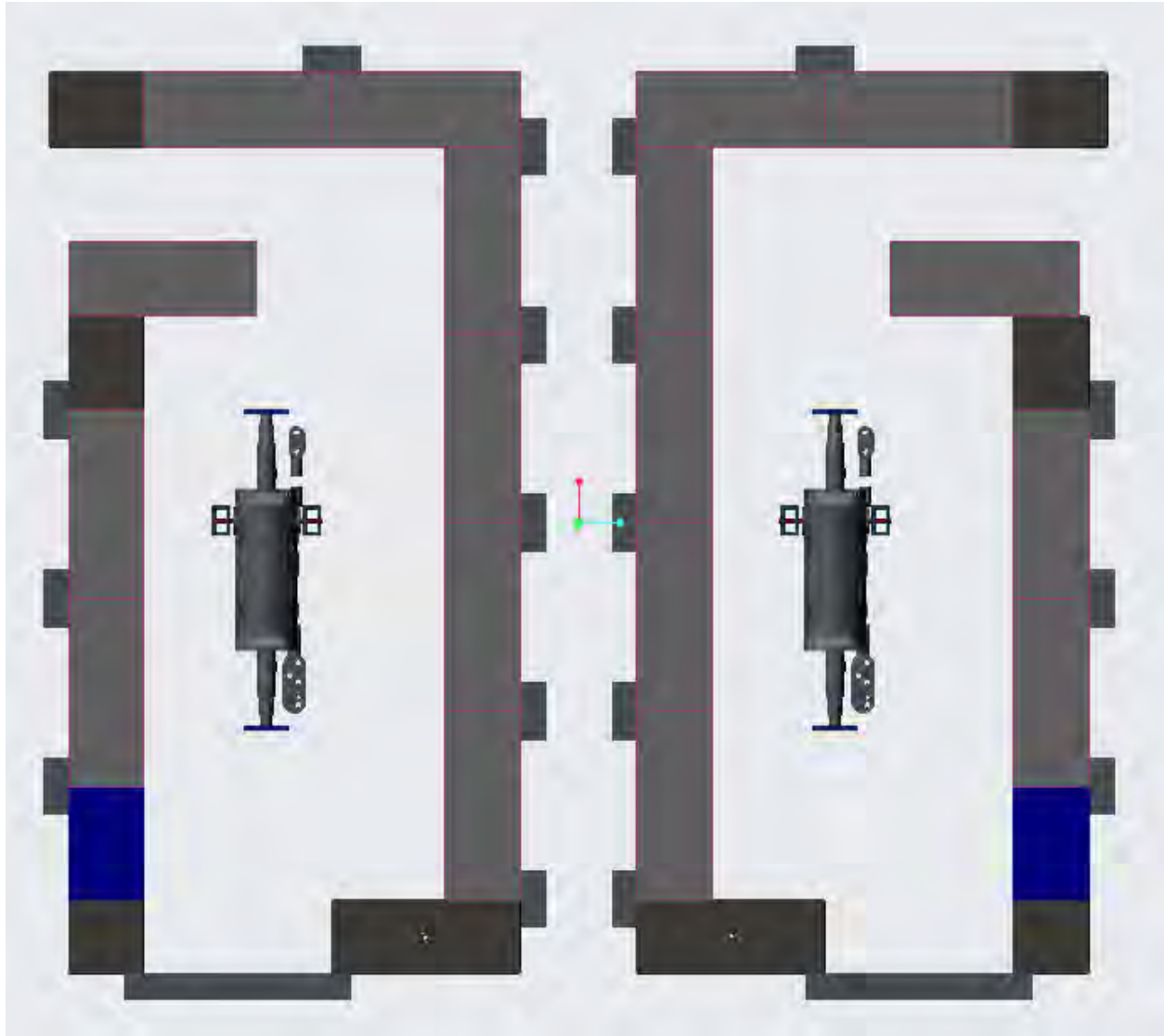


Figure 39: Test Bunker Layout 5, Top View

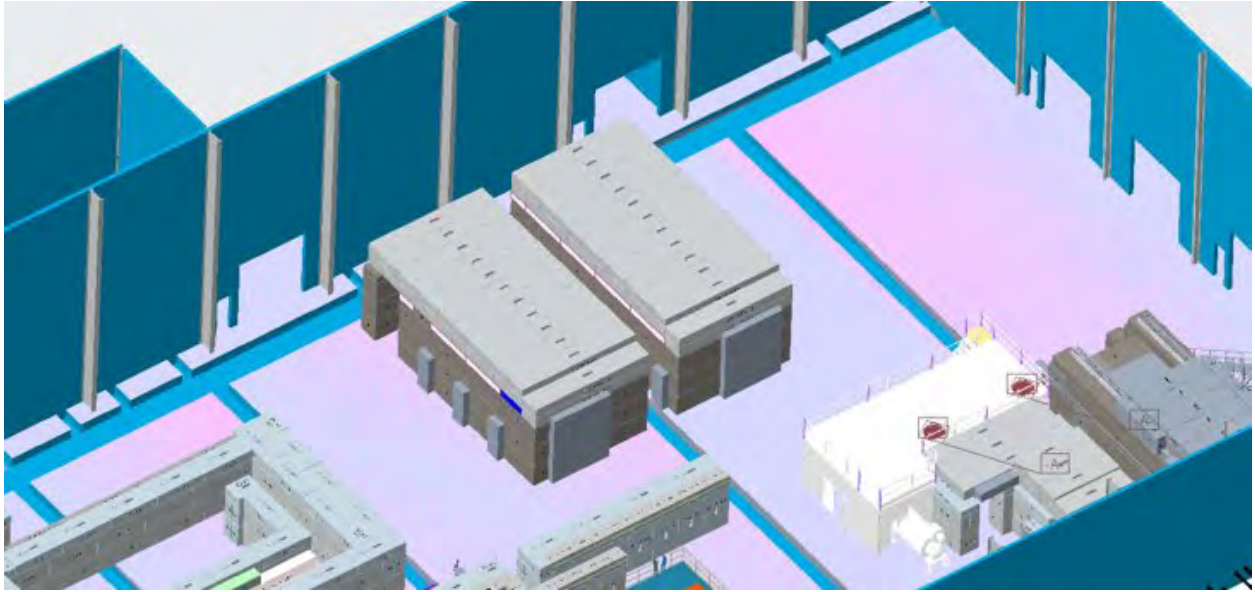


Figure 40: Test Bunker Layout 5 Inside Building

In layout 6 the two bunkers are brought together to share a common wall, reducing the number of blocks. The bunkers are turned 90 degrees to face the large open space in the building, allowing for cryomodules to be easily loaded in and out of the building. Several other provisions are also taken to reduce the number of building blocks including replacing six stacks of A blocks with roof blocks. Layout 6 is shown below:

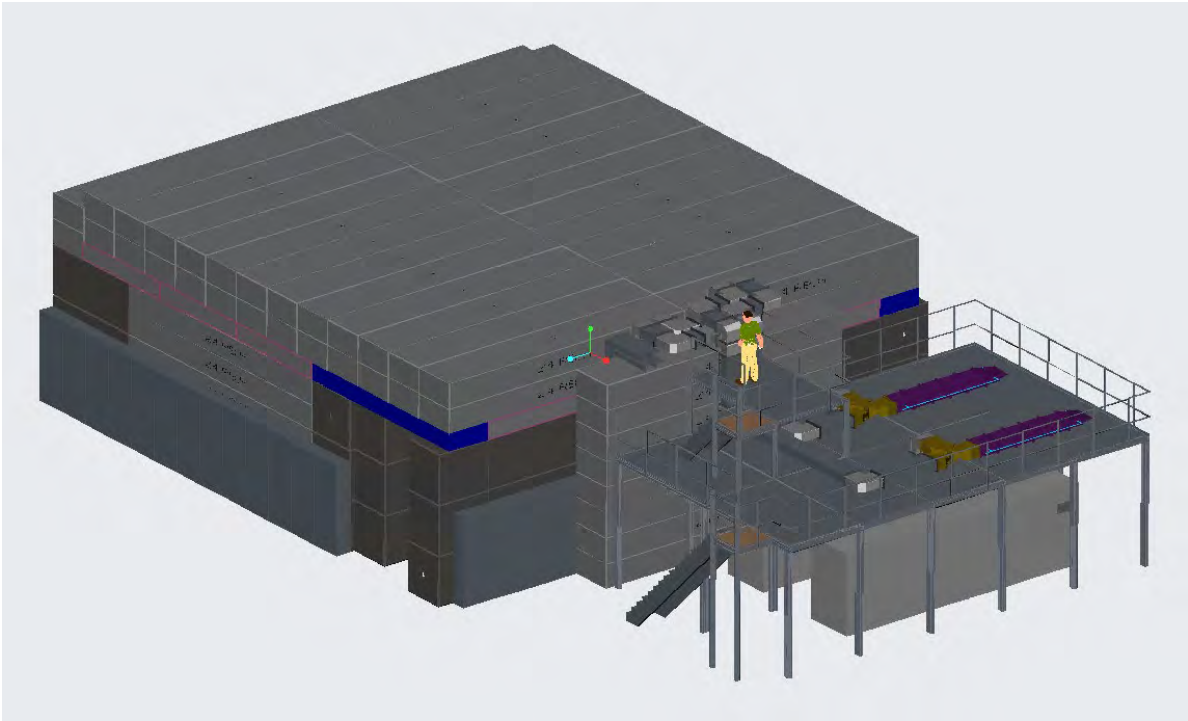


Figure 41: Test Bunker Layout 6, Full View

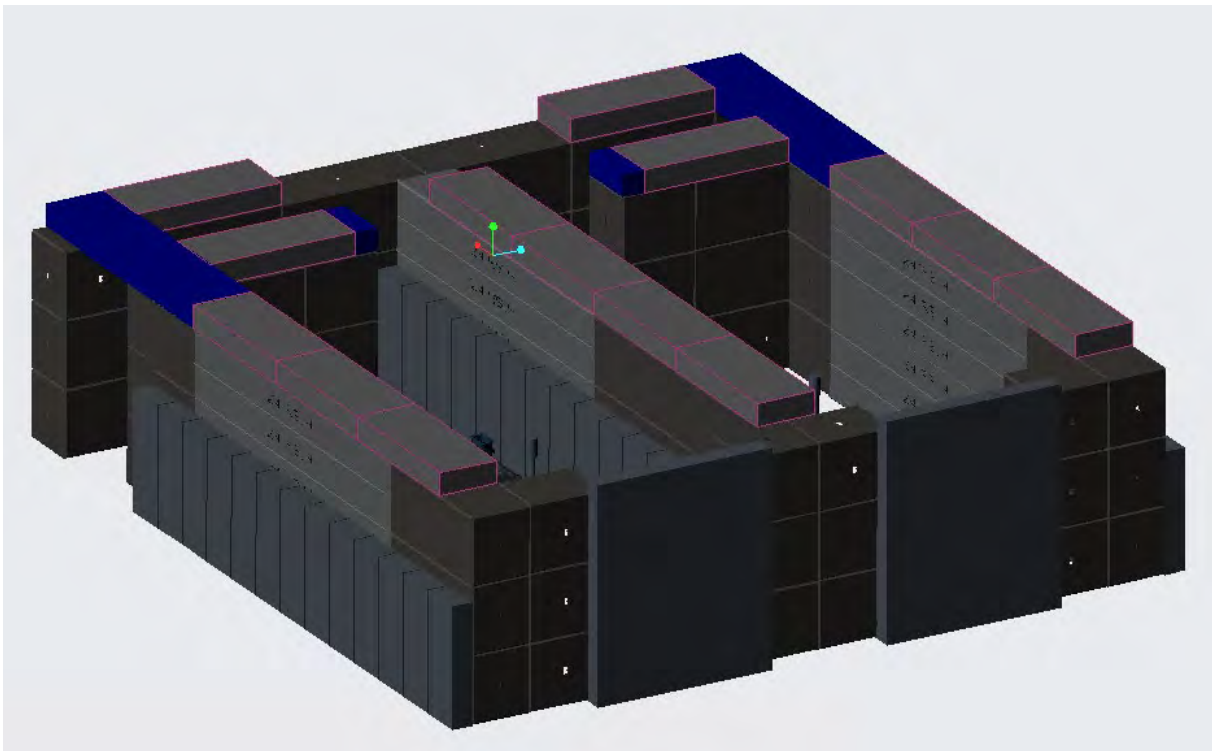


Figure 42: Test Bunker Layout 6, No Roof



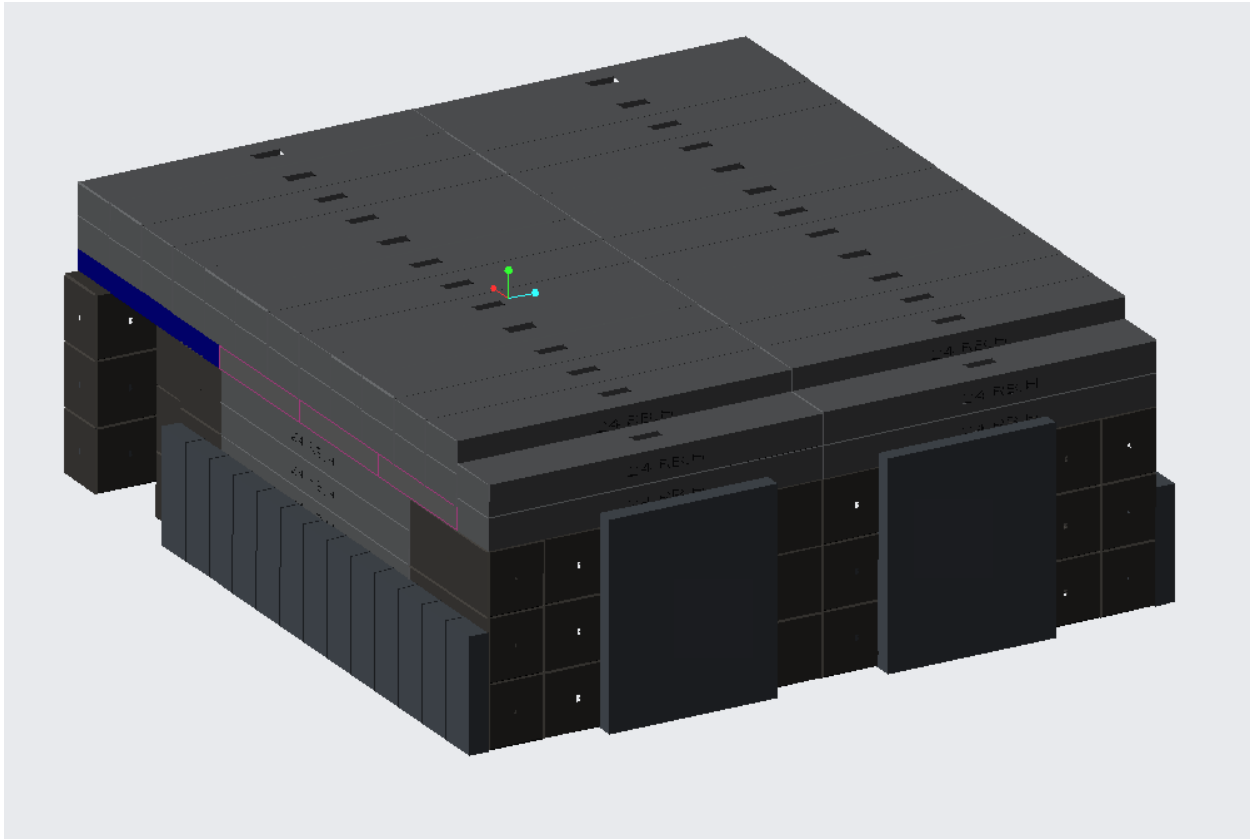


Figure 43: Test Bunker Layout 6, Front Side with Doors

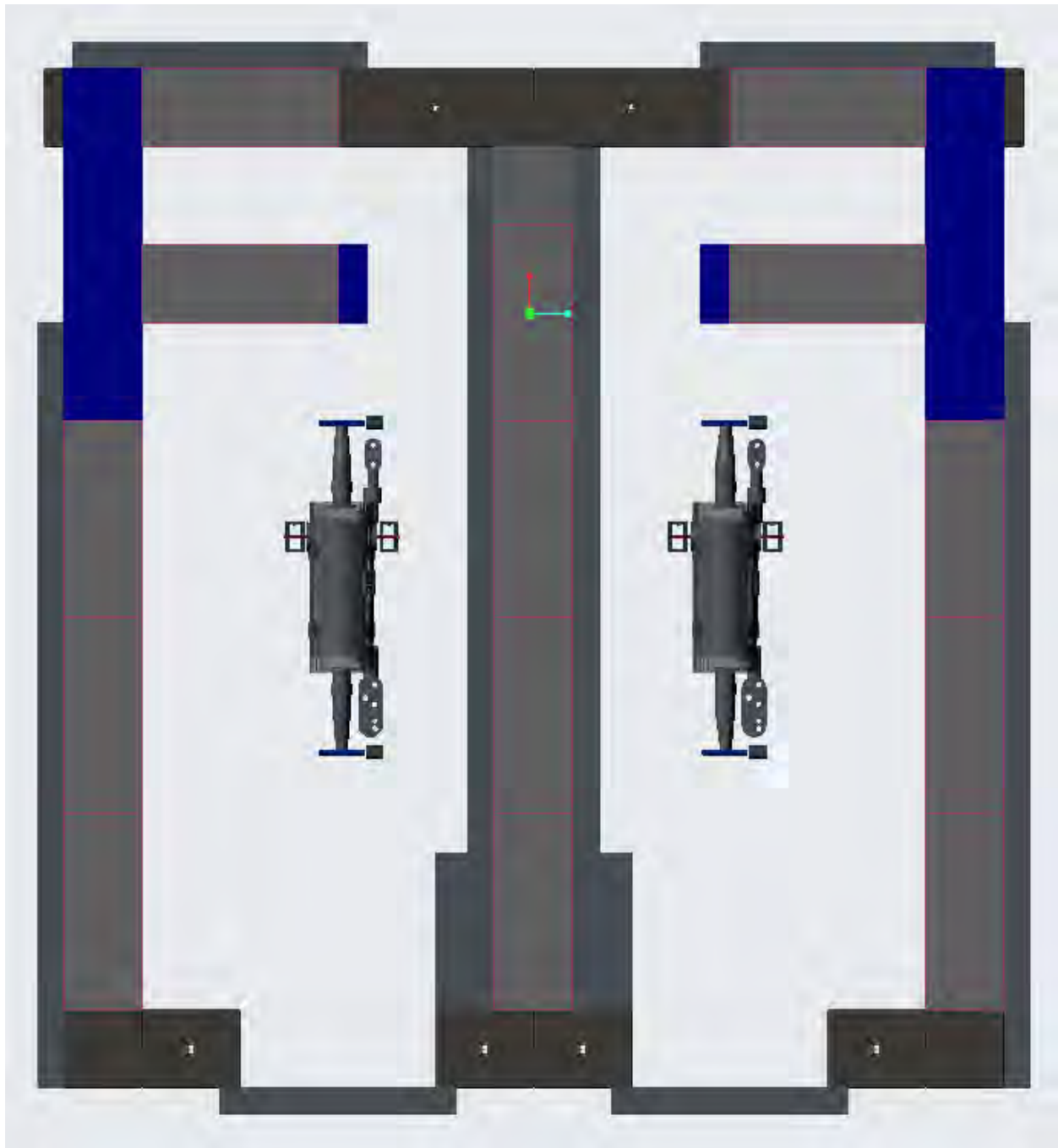


Figure 44: Test Bunker Layout 6, Top View

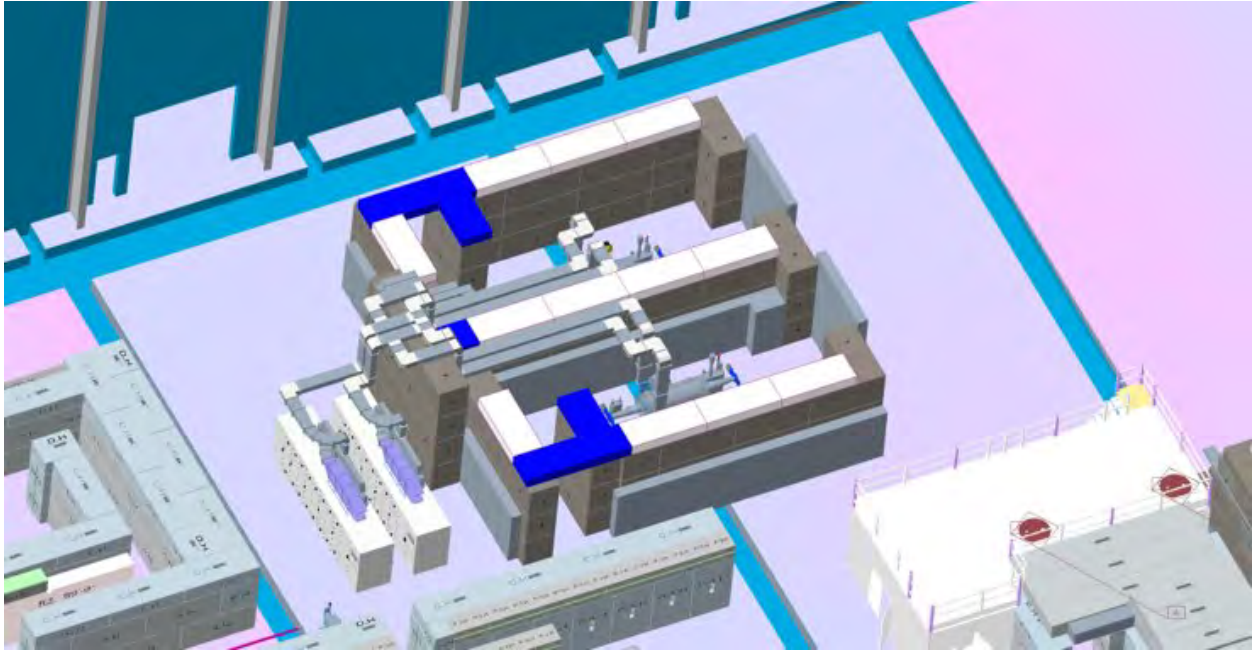


Figure 45: Test Bunker Layout 6, Inside Building 912

The model for layout 6 includes the necessary waveguides, the scaffolding necessary to install and maintain them, and cable trays. The two power amplifiers that are placed by the bunkers are only sufficient to power a single cavity, requiring that their outputs be able to be swapped between the two bunkers. As a result, the waveguides that route into the bunkers have a swap point located above a shield wall which is placed in front of the openings for the waveguides to block radiation. This swap is shown below, in both configurations:



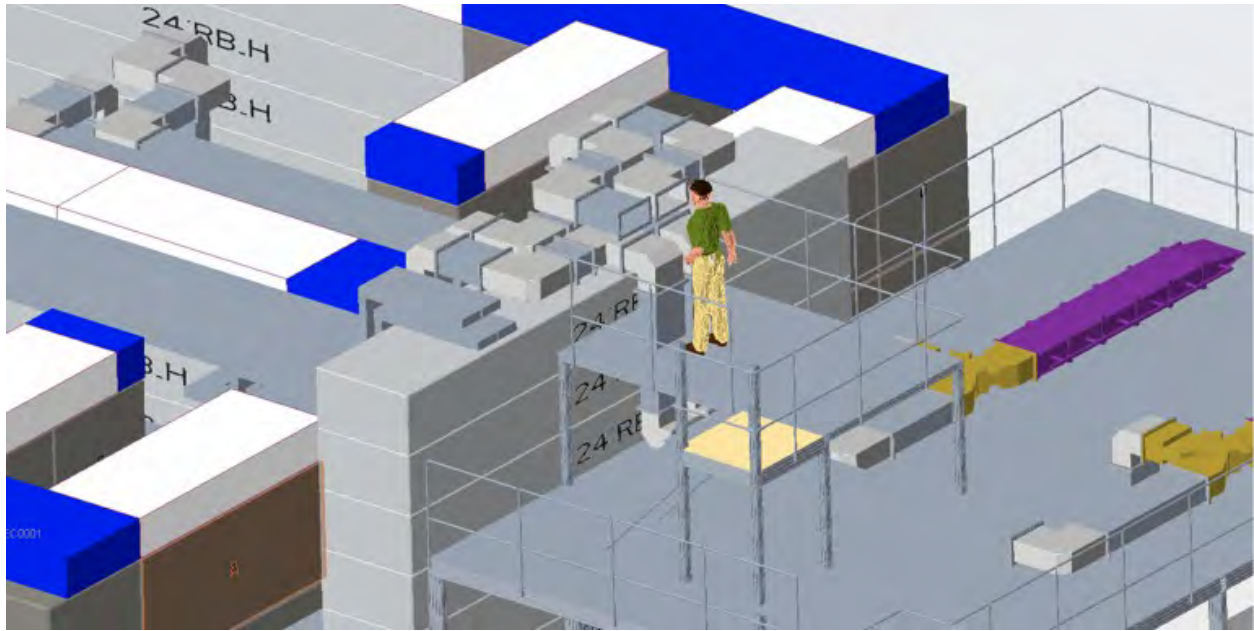


Figure 46: Test Bunker Layout 6 Waveguide Swap Configuration 1



Figure 47: Test Bunker Layout 6 Waveguide Swap Configuration 2

This swap allows both sets of the waveguides to have similar lengths. They can be accessed via a two-level scaffold, which is shown below:



Figure 48: Test Bunker Layout 6 Scaffolding

## CONCLUSION

An H/V slit, profile monitor, solenoid, TROG, and corrector were all added to the 3D model of the TRDBL, and a DCCT was added further upstream along the CEC beamline. The addition of new components required the design of new mounting provisions. All new components had drawings created with the necessary dimensions for manufacturing. As of writing this paper, several of the designed components have been manufactured.

The final chosen layout was the sixth proposed layout which featured two large bunkers 15 feet high. These bunkers are built using 48-inch-wide concrete blocks which supply the necessary thickness for radiation shielding. The bunkers share a common wall to reduce the number blocks to a total of 30 A-blocks, 30 B-blocks, 14 E-blocks, 5 custom blocks, and 72 24 ft roof blocks, and 72 custom 8 ft tall blocks. Two large power amplifiers capable of powering one cavity are placed next to the rear of the bunkers, with waveguides outputting from the amplifiers to the cavities. The connection to the waveguides going into each bunker can be swapped on top of an extra shield wall placed to prevent radiation from exiting the openings for the waveguides. A scaffolding is placed around the waveguides to allow for maintenance and installation to be done. The final layout is shown below:

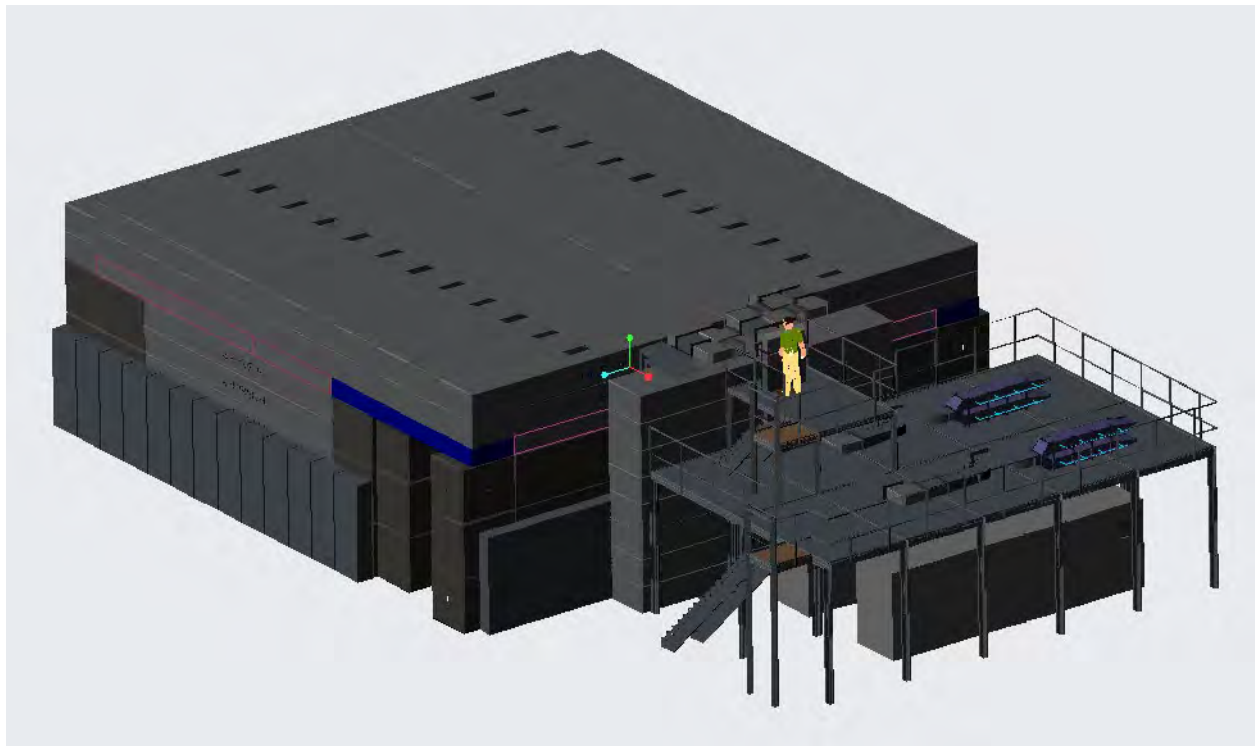


Figure 49: Test Bunker Layout 6, With Roof

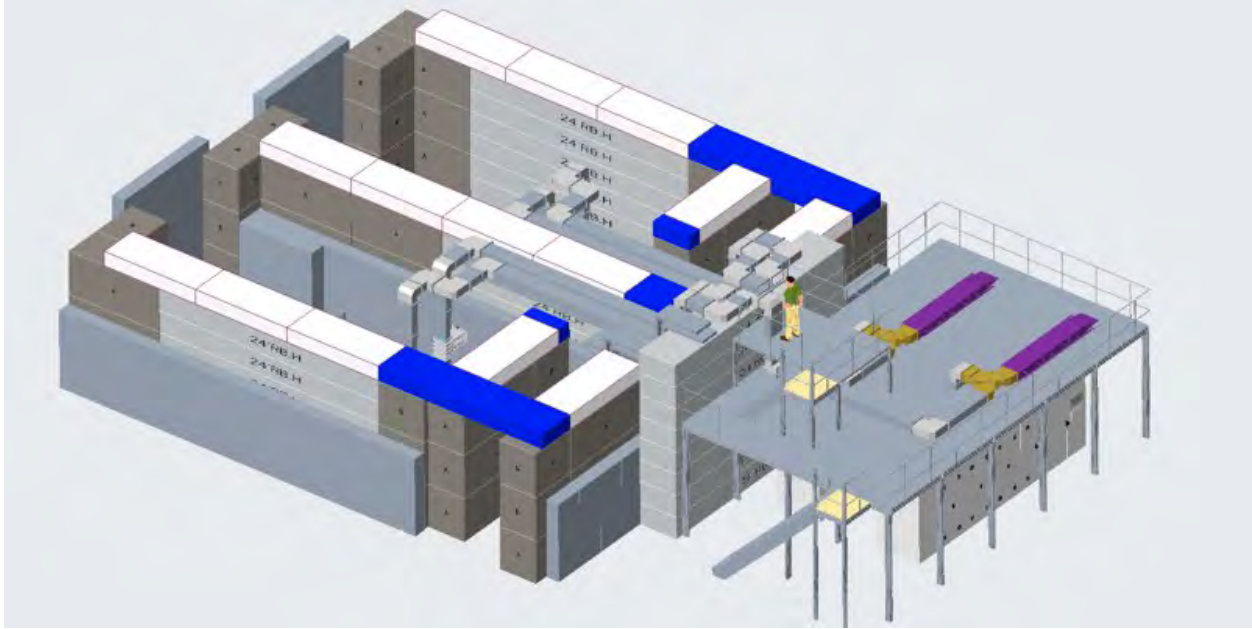


Figure 50: Test Bunker Layout 6, No Roof

The work done for CeC will allow for better monitoring and characterization of the electron beam, giving personnel more information about the properties of the beam. The work done for the EIC SRF cavity test bunkers will allow the cavities to be safely and efficiently tested prior to installation, and in doing so contribute to the implementation of the EIC along the RHIC.

## ACKNOWLEDGMENTS

I would like to thank my mentor Clifford Brutus for his help, support, and advice while working on these projects.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI)

## **References**

[1] Kin Yip, (private communication)

Spin matching for the EIC Electron Storage Ring  
Matthew Signorelli<sup>1</sup> and Jorg Kewisch<sup>2</sup>.

<sup>1</sup>) *Department of Physics, Cornell University, Ithaca, NY 14853.*

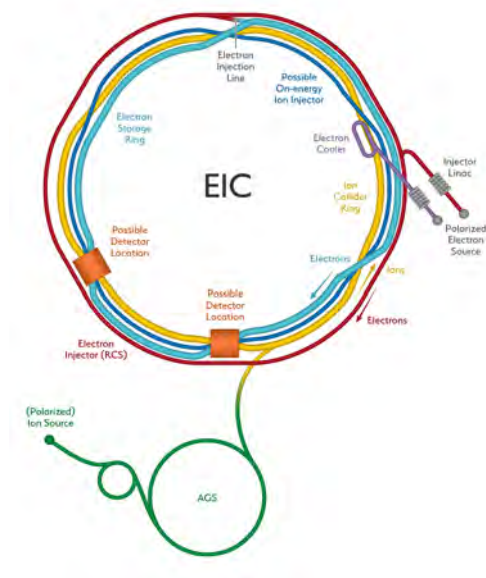
<sup>2</sup>) *Collider-Accelerator Department, Brookhaven National Laboratory, Upton, NY 11973.*

# Abstract

The Electron-Ion Collider at Brookhaven National Laboratory will provide polarized collisions of electron and ion beams. The spin rotator systems in the Electron Storage Ring (ESR) must be spin matched to prevent depolarization of the beam caused by synchrotron radiation emission. Previously, a preliminary 17.84 GeV ESR lattice with 1 interaction point was spin matched based off of an analytical derivation of the conditions that neglects terms proportional to the anomalous magnetic moment of the electron. In this project, the exact spin-orbit integral for horizontal matching was derived. Then, using Tao, the lattice was perfectly spin matched in the horizontal direction to the first order, per the implementation of the SLIM formalism in Bmad. Finally, the solution was compared to the original lattice by computing the exact spin-orbit integral for four particles, each with an initial transverse phase space displacement, through the solenoids at discrete steps.



# 1 Introduction



**Figure 1:** Proposed layout of the EIC<sup>1</sup>.

The Electron-Ion Collider (EIC) at Brookhaven National Laboratory (BNL) will allow for the next frontier of high energy physics experiments by providing collisions of spin-polarized electron and ion beams. It is necessary that the Electron Storage Ring (ESR) be designed to maintain high polarization levels for a reasonable amount of time while effectively manipulating particle spin to point longitudinally at the interaction points (IPs) for a range of energies. Electrons however, as opposed to hadrons, pose a more complex challenge due to their emission of significant synchrotron radiation when travelling through bending magnets; this changes the polarization of the beam through the Sokolov-Ternov effect and the diffusion of the closed/stable spin direction throughout the lattice caused by the instantaneous reductions in energy from photon emissions. The latter effect of spin diffusion can be remedied by spin matching – choosing lattice properties/magnet strengths to minimize or ideally completely remove the spin-orbit coupling excited by the rotator systems.

Previously, a preliminary 17.84 GeV ESR lattice with 1 IP was spin matched based off of an analytical derivation of the conditions. This derivation, however, neglects terms



proportional to the anomalous magnetic moment of the electron  $a$ , and the lattice showed an imperfect spin match in Bmad. In this project, the exact spin-orbit integral for horizontal matching was derived without neglecting terms proportional to  $a$ . Then, using the Tao interface to Bmad, the lattice was perfectly spin matched in the horizontal direction to the first order, while maintaining reasonable Twiss parameters and optics; following the SLIM formalism, the  $\mathbf{G}_x$  matrix was minimized to zero by variation of the quadrupole strengths in between the half solenoids of both the short and long solenoidal inserts used in the ESR. The solenoid fringe field effects on spin were also included in the optimization. Finally, the solution was compared with the original lattice by directly calculating the exact spin-orbit integrals for four particles, each with an initial displacement in the transverse phase space  $(x_0, x'_0, y_0, y'_0)$  at discrete steps in the solenoids.

## 2 Theory

### 2.1 Stable Spin Direction

Particle spin in an accelerator is defined as the expectation values of each component  $\vec{S} = (\langle S_x \rangle, \langle S_y \rangle, \langle S_z \rangle)$  from the quantum mechanical spinor  $\psi$ , that is  $\vec{S} = \langle \psi | \vec{\sigma} | \psi \rangle$  where  $\vec{\sigma}$  is a vector of the Pauli matrices. The Thomas-BMT equation, shown in Equation (1), defines the semiclassical motion of particle spin in the Lorentz-boosted particle rest frame as a result of laboratory frame magnetic and electric fields<sup>2,3,4</sup>.

$$\frac{d\vec{S}}{dt} = \vec{\Omega} \times \vec{S} \quad (1)$$

$$\vec{\Omega} = -\frac{q}{\gamma m} \left[ (1 + a\gamma)\vec{B}_\perp + (1 + a)\vec{B}_\parallel + \left( a\gamma + \frac{\gamma}{\gamma + 1} \right) \frac{\vec{E} \times \vec{\beta}}{c} \right]$$

The electric field term is often neglected because, in an RF cavity,  $\vec{E} \times \vec{\beta} \approx 0$ . The

Thomas-BMT equation can be linearized so that rotations of the spin  $\vec{S}$  through each magnet element are represented by a matrix multiplication (note that the spin rotation matrices are in general dependent on the phase space position  $\vec{x}$ ).

$$\vec{S}_f = R(\vec{x})\vec{S}_i \quad (2)$$

The spin rotation for 1-turn around the ring can then be obtained by multiplication of each magnet element rotation matrix for a specific particle's phase space point at that element. The invariant spin direction  $\hat{n}$  is a periodic solution to the Thomas-BMT equation where, for a specific point in phase space  $\vec{x}$  at a specific angle  $\theta$  on the azimuth:

$$\hat{n}(\vec{x}, \theta) = \hat{n}(\vec{x}, \theta + 2\pi) \quad (3)$$

Only on closed orbits do particles return to the same point in phase space for each turn; ideally, particles are injected onto the design/closed orbit of the ring (most often designed where  $\vec{x} = 0$ ), with the spin direction pointing along the invariant spin direction for the closed orbit at that point. This allows for the spin to follow a stable and periodic trajectory. The invariant spin direction on the closed orbit is denoted  $\hat{n}_0$ , and can be obtained by solving the eigenvector equation defined in Equation (4). Particles with spin at an angle to  $\hat{n}_0$  will precess around  $\hat{n}_0$ <sup>5</sup>.

$$\hat{n}_0 = \vec{R}_{1\text{-turn}}\hat{n}_0 \quad (4)$$

As such,  $\hat{n}_0$  is the stable spin direction. For the ESR, it must be manipulated to point longitudinally at the IPs and, for a periodic solution, rotated back to vertical in the arcs.

Designing a rotator system for a hadron ring would consist of choosing magnet elements that do exactly that - rotate  $\hat{n}_0$  to point in the desired directions at each point in the ring - while accounting for intrinsic and imperfection resonances. The design challenge becomes more complex when storing leptons, which have a much lower mass; electrons, as opposed to

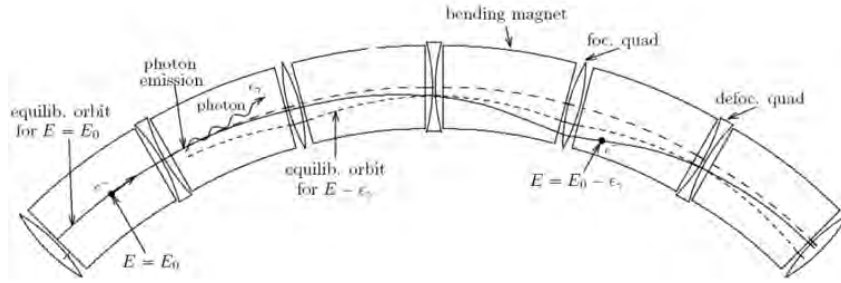
hadrons, emit significant synchrotron radiation when experiencing perpendicular accelerations, as in dipoles. Synchrotron radiation emission has several major effects on polarization, one of which is the Sokolov-Ternov effect; this is an unavoidable phenomenon, derived from the Dirac equation, where the probability of spin flip during photon emission asymmetrically prefers flipping antiparallel to the magnetic field<sup>6</sup>. This effect can be advantageous if the desired spin direction is antiparallel to the magnetic field. Baier, Katkov, and Strakhovenko derived the asymptotic polarization in a ring by including the Sokolov-Ternov effect in solving the Thomas-BMT equation; the BKS polarization is shown in Equation (5) with a corresponding time constant in Equation (6)<sup>7,8</sup>. However, the Sokolov-Ternov effect is not the whole picture; quantum emission of photons causes a diffusion of the stable spin direction, which has major effects on the asymptotic polarization and build-up time.

$$P_{bks} = P(\theta \rightarrow \infty) = -\frac{8}{5\sqrt{3}} \frac{\oint ds \frac{\hat{n}_0(s) \cdot \hat{b}(s)}{|\rho(s)|^3}}{\oint ds \frac{1 - \frac{2}{9}(\hat{n}_0(s) \cdot \hat{s})^2}{|\rho(s)|^3}} \quad (5)$$

$$\tau_{bks}^{-1} = \frac{5\sqrt{3}}{8} \frac{\hbar r_e \gamma^5}{m_e c} \oint ds \frac{1 - \frac{2}{9}(\hat{n}_0(s) \cdot \hat{s})^2}{|\rho(s)|^3} \quad (6)$$

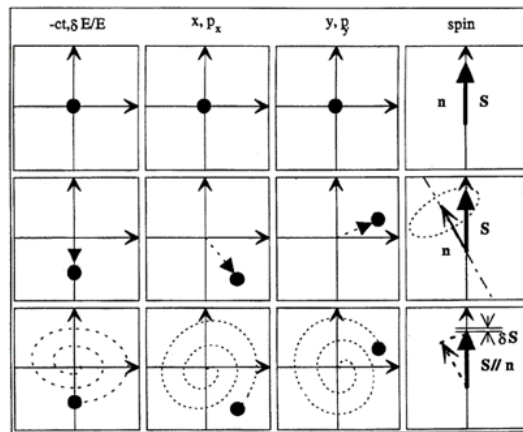
## 2.2 Spin Diffusion and Spin Matching

When electrons emit synchrotron radiation in bending magnets, they lose energy, in an effect called radiation damping. This reduction in the momentum is later restored in the RF cavities but in only in the longitudinal direction, causing a damping of the transverse motion. In a solely classical picture, this damping would result in beam emittances so small that major problematic effects would be introduced<sup>9</sup>. However, because the energy loss when emitting a photon is quantized, these statistical, quantum fluctuations in energy induce new oscillations; when a particle emits a photon and instantaneously changes its energy  $\Delta p/p_0$ , it will instantaneously have a new equilibrium orbit and have some nonzero displacement on that orbit. Thus, synchrotron and betatron oscillations around the new equilibrium orbit will be induced. This effect is shown in Figure 2.



**Figure 2:** Visual of orbit change caused by photon emission<sup>9</sup>.

When considering both radiation damping and radiative excitation, the effects on phase space in the design of an electron ring can be accounted for. However, these quantized reductions in energy also have a significant effect on spin; once the particle has a new closed orbit,  $\hat{n}_0$  changes; there is a coupling between the invariant spin direction  $\hat{n}$  and the energy of the particle. A visualization of this effect is shown in Figure 3.



**Figure 3:** Effect of photon emission on phase space and spin<sup>10</sup>.

The  $\vec{d}$  function, shown in Equation (7), defines the spin-orbit coupling, where  $\delta = \frac{\Delta p}{p_0}$ .

$$\vec{d} = \frac{\partial \hat{n}}{\partial \delta} = \gamma \frac{\partial \hat{n}}{\partial \gamma} \quad (7)$$

Derbenev and Kondratenko extended the BKS asymptotic polarization and time constant by including these quantum emission effects/spin-orbit coupling in their derivation. These

equations are shown in Equation (8) and Equation (9)<sup>11</sup>.

$$P_{dk} = P(\theta \rightarrow \infty) = -\frac{8}{5\sqrt{3}} \frac{\oint ds \left\langle \frac{\hat{n} \cdot \hat{b}}{|\rho(s)|^3} \right\rangle_s}{\oint ds \left\langle \frac{1 - \frac{2}{9}(\hat{n} \cdot \hat{s})^2 + \frac{11}{18} \left| \frac{\partial \hat{n}}{\partial \delta} \right|^2}{|\rho(s)|^3} \right\rangle_s} \quad (8)$$

$$\tau_{dk}^{-1} = \frac{5\sqrt{3} \hbar r_e \gamma^5}{8 m_e C} \oint ds \left\langle \frac{1 - \frac{2}{9}(\hat{n} \cdot \hat{s})^2 + \frac{11}{18} \left| \frac{\partial \hat{n}}{\partial \delta} \right|^2}{|\rho(s)|^3} \right\rangle_s \quad (9)$$

$\vec{d}$  has a significant effect on the polarization, and so must be minimized or ideally completely removed; this process is called spin matching<sup>12</sup>. The general spin-matching condition derivation follows<sup>13,14</sup>. The Thomas-BMT equation can be split into its constant, closed orbit components  $\vec{\Omega}_0$  and perturbative components  $\vec{\omega}$  resulting from momentum deviations and transverse displacements, so that  $\vec{\Omega} = \vec{\Omega}_0 + \vec{\omega}$  in Equation (1). Then choosing vectors  $\hat{l}_0$  and  $\hat{m}_0$  so that  $(\hat{l}_0, \hat{m}_0, \hat{n}_0)$  form an orthonormal triad, define the complex vector  $\hat{k}_0 = \hat{l}_0 - i\hat{m}_0$ . Thus, any arbitrary spin vector  $\vec{S}$  can be defined by Equation (10), where the  $\alpha_0$  is complex.

$$\vec{S} = \sqrt{1 - |\alpha_0|^2} \hat{n}_0 + \Re(i\alpha_0 \hat{k}_0^*) = \sqrt{1 - |\alpha_0|^2} \hat{n}_0 - \alpha_{0,Re} \hat{m}_0 - \alpha_{0,Im} \hat{l}_0 \quad (10)$$

$\alpha_0$  thus describes the deviation from the stable spin solution throughout the ring. Inserting  $\vec{S}$  into the Thomas-BMT equation gives the evolution of  $\alpha_0$  in terms of the spin precession perturbation vector  $\omega$  to the first order, shown in Equation (11).

$$\frac{d\alpha_0}{ds} = -i\vec{\omega} \cdot \hat{k}_0 \quad (11)$$

Integrating this over the rotator insertion gives the change in  $\alpha_0$  from the start to end. To spin match, this should be zero, and so the spin matching condition is obtained from the spin-orbit integral in Equation (12).

$$\int_{s_{in}}^{s_{out}} \vec{\omega} \cdot \hat{k}_0 ds = 0 \quad (12)$$

## 2.3 ESR Spin Matching Conditions

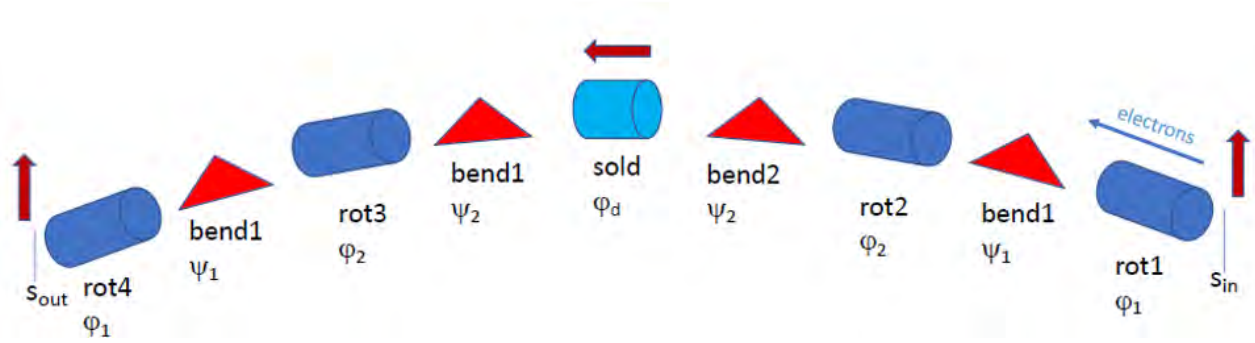


Figure 4: ESR spin rotator system<sup>15</sup>.

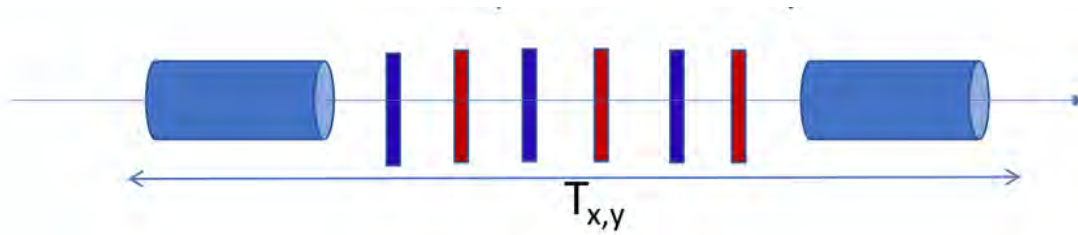


Figure 5: Individual solenoidal insert<sup>15</sup>.

The ESR rotator system utilizes two solenoidal inserts at either side of the IP - a “long” insert with 7 quadrupoles in between two half solenoids, and a “short” insert with 6 quadrupoles in between two shorter half solenoids. This design was chosen so that  $\hat{n}_0$  can be rotated to longitudinal at the IP and rotated back to vertical for energies between 5-18 GeV. A minimum of 5 quadrupoles is necessary to achieve a horizontal spin match; 3 quadrupoles for betatron decoupling ( $\beta_x = \beta_y$ ,  $\alpha_x = \alpha_y$ ,  $\varphi_x = \varphi_y + \pi$ ) and 2 more to meet the horizontal spin matching conditions. V.I. Ptitsyn derived the horizontal spin matching condition for the ESR, neglecting terms proportional to  $a$ . The spin-orbit integral can be expressed as a summation over the solenoids when neglecting terms proportional to  $a$ , and is shown in Equation (13) where each of the four indices are for the four total solenoidal inserts<sup>16</sup>.

$$\sum_i^4 \frac{\phi_i}{2} [(k_{0x}x' + k_{0y}y')_{\text{in}} + (k_{0x}x' + k_{0y}y')_{\text{out}}] = 0 \quad (13)$$

A horizontal transfer matrix across each insert that satisfies Equation (13) can be obtained by satisfying the conditions in Equations (14) and (15), with  $\mathbf{T}_x = -\mathbf{T}_y$ <sup>16</sup>.

$$m_{11} \frac{1}{2} \tilde{K}_s \sin \phi + m_{21} \cos \phi = 0 \quad (14)$$

$$1 + m_{22} \cos \phi + m_{12} \frac{1}{2} \tilde{K}_s \sin \phi = 0 \quad (15)$$

## 2.4 ESR Exact Spin-Orbit Integral

Derivation of the exact spin-orbit integral was performed by following that from V.I. Ptitsyn, but retaining terms proportional to  $a$ . In order to expedite the derivation, MATLAB was used to symbolically evaluate and simplify the spin-orbit integrand with the precession perturbation vector in Equation (16)<sup>13</sup>,  $\hat{k}_0$  vector as defined previously, and  $\hat{k}'_0$  vector in Equation (17) obtained by inserting  $\hat{k}_0$  into the first order Thomas-BMT equation.

$$\begin{aligned} \omega_x &= (1 + \nu_0)y'' + (1 + a)\tilde{K}_s x' \\ \omega_s &= (1 + a)(\tilde{K}'_y y - \tilde{K}_s p_\sigma) - (\nu_0 - a)\tilde{K}_y p_y \\ \omega_y &= -(1 + \nu_0)x'' + \left(\nu_0 + \frac{a}{\gamma_0}\right)\tilde{K}_y p_\sigma + (1 + a)\tilde{K}_s y' \end{aligned} \quad (16)$$

$$\begin{aligned} k'_{0x} &= (1 + a)\tilde{K}_s k_{0y} - \nu_0 \tilde{K}_y k_{0s} \\ k'_{0s} &= \nu_0 \tilde{K}_y k_{0x} \\ k'_{0y} &= -(1 + a)\tilde{K}_s k_{0x} \end{aligned} \quad (17)$$

After simplification and only retaining terms proportional to the complex amplitude of the first complex conjugate eigenvector pair  $A_I$  (assuming no betatron coupling at the start and exit, no vertical dispersion, and that  $\hat{n}_0$  is vertical at the start and exit), the exact horizontal spin-orbit integral is obtained in Equation (18).

$$\left( \int_{s_{in}}^{s_{out}} \vec{\omega} \cdot \hat{k}_0 ds \right)_{A_I} = \int_{sol} -\nu_0 \tilde{K}_s (1+a) \left[ k_{0y} y' \left( 1 + \frac{1}{\nu_0} \right) + k_{0x} x' \right] ds \quad (18)$$

### 3 Methods

The Tao interface to Bmad was used to perform the horizontal spin match. Bmad utilizes the SLIM formalism, formulated by Chao, to express first order spin matching<sup>14</sup>. The formalism expresses changes to  $\hat{n}_0$  by small variations in the other axes  $\hat{l}_0$  and  $\hat{m}_0$ . This gives the spin vector in Equation (19) with small values  $\alpha$  and  $\beta$ .

$$\begin{aligned} \vec{S} &= \sqrt{1 - \alpha^2 - \beta^2} \hat{n}_0 + \alpha \hat{l}_0 + \beta \hat{m}_0 \\ &\approx \hat{n}_0 + \alpha \hat{l}_0 + \beta \hat{m}_0 \end{aligned} \quad (19)$$

The usual 6-vector formalism, where  $\vec{x} = (x, p_x, y, p_y, z, p_z)$  then becomes an 8-vector formalism where the elements  $\alpha$  and  $\beta$  are tacked on to the end. In the SLIM formalism, each magnet transfer matrix is now an  $8 \times 8$  matrix, shown in Equation (20). Note that Stern-Gerlach effects are neglected.

$$\begin{pmatrix} \mathbf{X}_{2 \times 1} \\ \mathbf{Y}_{2 \times 1} \\ \mathbf{Z}_{2 \times 1} \\ \alpha \\ \beta \end{pmatrix}_f = \begin{pmatrix} \mathbf{M}_{6 \times 6} & \mathbf{0}_{6 \times 2} \\ \mathbf{G}_{2 \times 6} & \mathbf{D}_{2 \times 2} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{2 \times 1} \\ \mathbf{Y}_{2 \times 1} \\ \mathbf{Z}_{2 \times 1} \\ \alpha \\ \beta \end{pmatrix}_i \quad (20)$$

$$\mathbf{G}_{2 \times 6} = \begin{pmatrix} \mathbf{G}_x & \mathbf{G}_y & \mathbf{G}_z \end{pmatrix} \quad (21)$$



Thus,  $\mathbf{D}_{2 \times 2}$  represents the transfer of  $\alpha$  and  $\beta$  for a particle on the closed orbit (for closed orbits where  $\vec{x} = 0$ ). Each of  $\mathbf{G}_x$ ,  $\mathbf{G}_y$ , and  $\mathbf{G}_z$  are  $2 \times 2$  matrices, showing the spin-orbit coupling in  $x$ ,  $y$ , and  $z$ , respectively. A horizontal spin match to the first order would thus require setting  $\mathbf{G}_x = 0$ .

Bmad intuitively implements the SLIM formalism. An optimization scheme was set up in Tao so that each of the quadrupoles in between the half solenoids for each insert were variables, but overlaying them such that the short and long solenoidal inserts on either side of the IP are symmetrical - that is, the quadrupole strengths in the short insert on the forward side are equivalent to those on the rear side, and likewise for the long insert. The Bmad “datums” used were: no betatron coupling at the end of each of the four solenoidal inserts (weight of 10 for each coupling matrix element), and the spin  $\mathbf{G}_x$  matrix to be 0 across the entire rotator system (weight of 5 for each matrix element). A Bmad lattice with only the 17.84 GeV ESR rotator system was set up to speed computation times, using the same start beta and alpha functions, and phases, as that in the periodic lattice. The Bmad integration step size was reduced until an acceptable convergence in the G matrices was reached (reached at 1 mm for each lattice element). Then, the Levenberg-Marquardt optimization algorithm, implemented in Tao, was used to obtain the minimum. This algorithm excels at finding local minima, and since the original lattice was already close to 0, it appeared to make sense to run this particular algorithm. The optimization was run many times until no change in the merit function was obtained; at this point a lattice satisfying the datums was obtained, and the corresponding quadrupole strengths written to a file.

After obtaining a perfectly horizontally matched lattice, four different particles were propagated through both the original and matched lattices using single tracking in Bmad, each with some different initial phase space displacement. The orbit data  $(x, p_x, y, p_y)$  of each throughout the rotator system were written to a file with 10000 datapoints total. For a total length of 327.528 m, this corresponds to data points at 3.27528 cm increments. The  $p_x$  and  $p_y$  values that Bmad stores are *not* the canonical momenta, but the field-free region

momenta, and so are exactly  $x'$  and  $y'$  respectively.

After obtaining the orbit data for each of the four particles, propagated through both the original lattice and the fully-matched lattice, a MATLAB program was written to parse the data and perform the integration in Equation (18), assuming a constant solenoid strength  $\tilde{K}_s$ . To increase accuracy, a cubic spline interpolation of the data points was implemented in MATLAB to reduce the step size down to 1 mm. Finally, the spin-orbit integrals were evaluated using a trapezoidal integration of the finer data.

## 4 Results

The  $\mathbf{G}_x$  matrix before and after performing the optimization are shown in Equation (22) and Equation (23) respectively. Table 1 shows the corresponding exact spin-orbit integrals calculated for both the original lattice and matched lattice for the four particles. The difference in the beta function (matched - original) is shown in Figure 6.

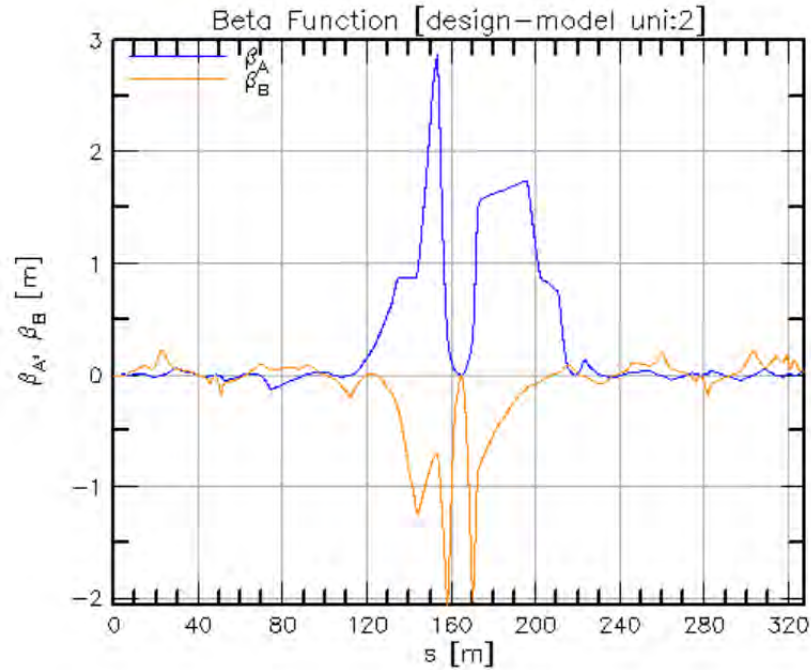
$$\mathbf{G}_{x,\text{original}} = \begin{pmatrix} 0.0000120 & 0.0005214 \\ 0.0018243 & 0.0075423 \end{pmatrix} \quad (22)$$

$$\mathbf{G}_{x,\text{matched}} = \begin{pmatrix} 0.0000000 & 0.0000000 \\ 0.0000000 & 0.0000000 \end{pmatrix} \quad (23)$$

$\vec{x}_0$	Original $\int_{s_{in}}^{s_{out}} \vec{\omega} \cdot \hat{k}_0 ds$	Matched $\int_{s_{in}}^{s_{out}} \vec{\omega} \cdot \hat{k}_0 ds$	$ \Delta S/S_0 $
(1e-04[m]) $\hat{x}$	-4.5067e-04 + 2.2452e-05i	-4.2839e-04 + 2.1598e-05i	0.0494126
(1e-05[rad]) $\hat{x}'$	-1.9300e-04 + 1.9485e-05i	-1.7120e-04 + 1.8637e-05i	0.1124720
(1e-05[m]) $\hat{y}$	-8.5301e-05 + 2.4473e-06i	-6.3169e-05 + 1.5714e-06i	0.2595533
(1e-06[rad]) $\hat{y}'$	+2.1927e-05 - 8.6941e-07i	+2.1980e-05 - 8.7141e-07i	0.0024221

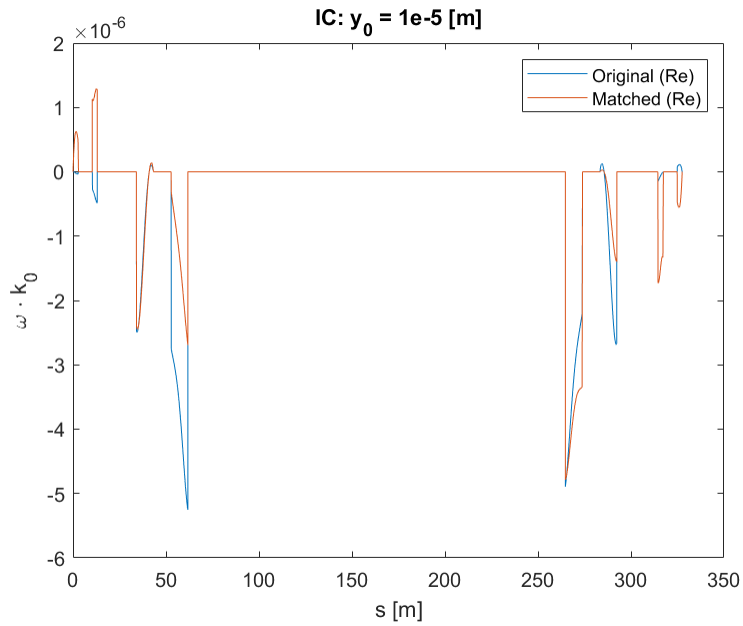
**Table 1:** Exact spin-orbit integrals for four different particles over entire rotator system.

The differences in the beta functions are not very large, only on the scale of a couple meters, which is desirable. Reductions in the spin-orbit integral were observed in all particle

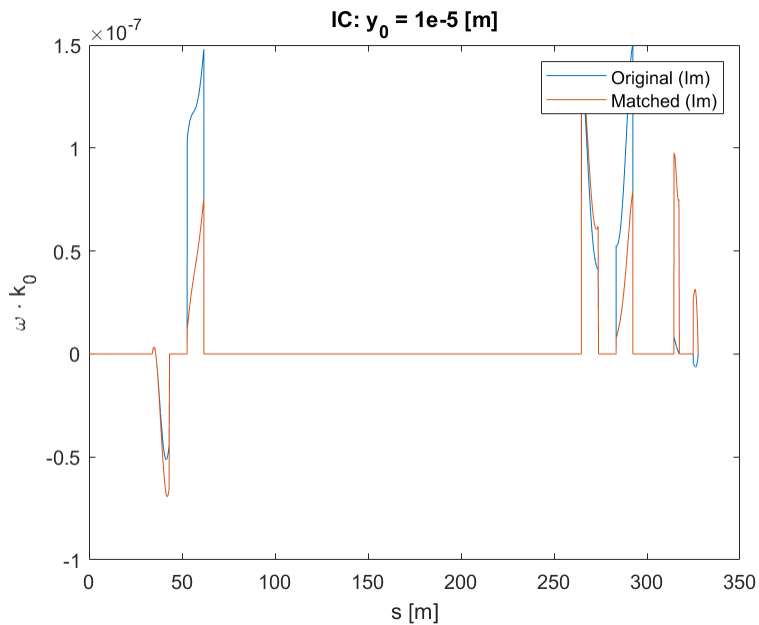


**Figure 6:** Difference in the beta function over the rotator system (matched – original).

cases, though most significantly for the  $\hat{y}$  initial offset case. The beam is assumed to be flat in the design, however this result may still be significant when considering the effects of fringe fields and higher order effects. For this case, a plot of the integrand vs.  $s$  shows a clear reduction in the integral, and so that case is shown in Figures 7 and 8.



**Figure 7:** Real part of spin-orbit integrand across the rotator system for  $\vec{x}_0 = (1e-05[m])\hat{y}$ .



**Figure 8:** Imaginary part of spin-orbit integrand across the rotator system for  $\vec{x}_0 = (1e-05[m])\hat{y}$ .

## 5 Conclusions

The exact spin-orbit integral, without neglecting terms proportional to  $a$ , was derived. The effect of this assumption in the spin matching was analyzed by obtaining a perfectly horizontally spin matched lattice in Bmad, and comparing it with the original solution. The exact spin-orbit integrals for four different particles, each with some initial displacement in phase space, were obtained by tracking the particles through the lattice, cubic-splining the data to increase accuracy, and performing trapezoidal integration. The result shows a better spin matched lattice that can be used in the further analyses necessary for the ESR.

A next step for this project includes analyzing why the spin orbit integrals are not significantly smaller than the original case, for what they were found to be in the matched case; a possible reason could be the choice of step size in Bmad, or that the 1 mm increments obtained by cubic spline interpolation are still too large. It is also possible that these are resulting from higher order effects not captured in the first order SLIM formalism. Following this, the optics in between each insert need to be matched to meet the required conditions for the ESR at the IP and between the crab cavities. Finally, spin-tracking with radiation, including Sokolov-Ternov effects, can be performed with the better-matched lattice in Bmad.

## Acknowledgements

The author would like to thank Jorg Kewisch and Georg Hoffstaetter for their mentorship and guidance in this project, as well as David Sagan for help with using Bmad/Tao. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## References

- [1] BNL, EIC schematic (with labels), 2020.
- [2] L. H. Thomas, *Phil. Mag.* **3**, 1 (1927).
- [3] V. Bargmann, L. Michel, and V. L. Telegdi, *Phys. Rev. Lett.* **2**, 435 (1959).
- [4] L. H. Thomas, *AIP Conference Proceedings* **95**, 4 (1983).
- [5] D. P. Barber, G. H. Hoffstaetter, and M. Vogt, *EPAC Conference Proceedings* (1998).
- [6] A. A. Sokolov and I. M. Ternov, *Sov. Phys. Doklady* **8**, 1203 (1964).
- [7] V. N. Baier and V. M. Katkov, *Physics Letters A* **25**, 492 (1967).
- [8] V. N. Baier, V. M. Katkov, and V. M. Strakhovenko, *Sov. Phys. JETP* **31**, 908 (1970).
- [9] L. Rivkin, *Electron dynamics with synchrotron radiation*, 2019.
- [10] J. Buon and J. Koutchouk, *CERN-SL-94-80-AP* (1994).
- [11] Y. S. Derbenev and A. M. Kondratenko, *Sov. Phys. JETP* **37**, 968 (1973).
- [12] B. W. Montague, *Physics Reports* **113**, 1 (1984).
- [13] V. I. Ptitsyn, Y. M. Shatunov, and S. R. Mane, *Nucl. Inst. and Meth.* **608**, 225 (2009).
- [14] A. W. Chao, *Nucl. Inst. and Meth.* **180**, 29 (1981).
- [15] V. I. Ptitsyn and S. Tepikian, *Spin rotator design and spin matching*, 2020.
- [16] V. I. Ptitsyn, *Spin matching derivation*, 2021.
- [17] F. Meot, *Spin dynamics*, 2021.
- [18] S. Y. Lee, *Spin Dynamics and Snakes in Synchrotrons*, World Scientific, 1997.

# Study charge carriers diffusion in field free silicon

Mikhail Smirnov, Department of Electrical and Computer Engineering,

New York Institute of Technology, New York, NY 10023

Ivan Kotov, Instrumentation Division, Brookhaven National Laboratory, Upton, NY, 11973

## **Abstract**

The National Synchrotron Light Source II (NSLS-II) Soft Inelastic x-ray Scattering (SIX beamline) at Brookhaven National Laboratory (BNL) includes two Electron-Multiplying Charge-Coupled Devices (EMCCDs) for x-ray detection. These devices work by converting beams of photons into electrons captured by pixel gates that are then converted into a voltage to create images of the incident x-rays. On the way to the pixel gates, the electrons drift in the depletion zone and diffuse in the field-free region. The diffusion of charges in the field-free region is important to model as it affects the readout. This work examines the diffusion of charges in EMCCDs using mathematical modeling and computational simulation in the C++ ROOT framework. Being able to theoretically predict charge diffusion will allow us to develop EMCCDs with better accuracy in coordinate measurements for x-rays. One way we can better describe the charge distribution shape is by modeling the electron cloud with a Gaussian shape instead of a simple point charge with the Dirac delta function. From gaining a better understanding of EMCCD charge diffusion, x-ray measurement at BNL's NSLS-II can be done more accurately, in addition to improving upon other EMCCD applications, such as in astronomy, medical imaging, and optical microscopy.

## **I. Introduction**

A CCD (Charge-Coupled Device) is an integrated circuit containing metal-oxide-silicon (MOS) structures divided up into pixels. Each pixel allows for conversion from photons into electron-hole pairs. CCDs are commonly used as sensors for digital imaging by transferring charge between their pixels. Each pixel contains gates (electrodes) in which voltages can be applied to in order to create potential wells under each electrode. The gates are used to shift the charges along columns of pixels and then get collected into a serial register. Each pixel in each line of the image shifts its charge to the sense node and the energy of the photons hitting the pixels dictates how many electron-hole pairs are generated. Since the intensity of the incident rays is proportional to



the amount of electron-hole pairs, the resulting amount of collected electrons allows for measurement of the incident rays. In this case, CCDs are used for x-ray detection at BNL's NSLS II's SIX beamline<sup>1</sup> with pixel size being  $16\ \mu\text{m}$  and the total device thickness being approximately  $14\ \mu\text{m}$ . The first  $8\ \mu\text{m}$  of width in the back side of the EMCCD is the field-free region, while the depletion layer is  $6 - 8\ \mu\text{m}$  thick. The total imaging area spans 1632 columns and 1608 rows of pixels. The electrons are collected under electrode gates where a potential well is formed. In addition, a specific type of CCD known as, Electron-Multiplying Charge-Coupled Devices (EMCCDs), multiply electrons by impact ionization. This added stage allows for CCDs with greater sensitivity and lower readout noise.

A typical CCD model consists of multiple layers: the depletion layer, the field-free region and the reflective boundary at the entrance window. An x-ray beam is incident on the depletion region and the created electron hole pairs diffuse in the field-free region until they get reflected at either the entrance window or until they reach the interface between the field-free and depletion regions. The depletion region is described as the area where charges from x-rays are absorbed and travel towards the nearest potential well. The area underneath, the field-free region, does not experience an electric field but instead the charges migrate by diffusion<sup>2</sup>. Understanding the diffusion of charges in the field-free region allows for a more accurate prediction of x-rays, resulting in better coordinate accuracy when measuring x-rays with CCDs.

## II. Scope and Objectives

This work models the diffusion of charges in the field-free region of an EMCCD by first incorporating the diffusion equation with a boundary value problem at the boundaries between the top of the EMCCD (the interface between the depletion zone and the field free region) and the bottom of the EMCCD (the reflective entrance window boundary). In previous works<sup>1,3</sup>, typically the initial charge cloud distribution was modeled by the Dirac delta function, which represented a single point of charge at the very center of the pixels. However, a more practical and realistic example is to use the Gaussian function to model the initial spread of charges. For this reason,

this work compares the equations for charge density when incorporating these two different initial conditions. Afterwards, this work aims to calculate and plot the charge distribution on the EMCCD pixels in the field-free region accurately. This is done by computing the charge distribution function, calculated with the Gaussian function initial condition, by numerically solving the equation using C++ and plotting the resulting histograms in ROOT.

### III. Methods

#### A. Diffusion Equation

In order to model the diffusion in the field-free region of CCD's, we follow the process in <sup>4</sup> and introduce the three-dimensional diffusion equation in cylindrical coordinates and assume axisymmetric symmetry,

$$D\nabla^2\rho(\mathbf{r}, z, t) = \frac{\partial\rho(\mathbf{r}, z, t)}{\partial t}, \quad (1)$$

where  $\rho$  is the charge density and  $D$  is the diffusivity, with the radial and axial directions,  $(\mathbf{r}, z)$ . The radial direction,  $\mathbf{r}$ , is defined as,  $\mathbf{r} = (x, y)$  from the origin  $(x_0, y_0)$ , while the axial direction,  $z$ , describes the depth of the CCD.

We set  $z = 0$  to be the top of the sensor (the depletion zone interface) and  $z = h$  to be the reflective entrance window boundary. We assume no accumulation of charges at the depletion zone interface and assume no flux of charges through the reflective entrance window. In total the two boundary conditions are,

$$\rho(\mathbf{r}, z, t)\Big|_{z=0} = 0, \quad \frac{\partial\rho(\mathbf{r}, z, t)}{\partial z}\Big|_{z=h} = 0. \quad (2)$$

##### i. Dirac Delta Function as an Initial Condition

To solve this boundary value problem we take the solution<sup>4</sup> which considered the Dirac delta initial condition,  $\rho(\mathbf{x}, 0) = \delta(\mathbf{x} - \mathbf{x}_0)$ . Considering the vectors are  $n$ -dimensional, the solution for

the charge density is found to be the following for each vector component,

$$\rho(\mathbf{x}, t) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} e^{-Dt \sum k_j^2} e^{-i \sum k_j (x_j - x_j^0)} \prod dk_j$$

$$= \frac{1}{(4\pi Dt)^{n/2}} \exp \left[ -\frac{(\mathbf{x} - \mathbf{x}_0)^2}{4Dt} \right],$$
(3)

where  $k = \frac{u}{\sqrt{Dt}}$ .

The boundary conditions (2) are then applied to the general solution, which results in the following equation for the solution with the Dirac delta function initial condition,

$$Z(z')T(t) = 2 \sum_{n=0}^{\infty} \sin(\alpha_n \cdot z'_0) \sin(\alpha_n \cdot z') \exp \left( -\frac{\alpha_n^2 Dt}{h^2} \right),$$
(4)

where  $\alpha_n = (n + \frac{1}{2})\pi$  and  $z'$  is the coordinate normalized by the height of the EMCCD,  $z' = \frac{z}{h}$ .

## ii. Gaussian Function as an Initial Condition

If we instead assume the Gaussian function as the initial condition,  $\rho(\mathbf{x}, t) \Big|_{t=0} = A e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\sigma^2}}$ ,

$$\rho(\mathbf{k}, 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} A e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\sigma^2}} e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x},$$
(5)

where  $\rho(\mathbf{k}, 0)$  is the Fourier transform of the initial condition of the charge density, we can plug the above equation into the following differential equation solution,

$$\rho(\mathbf{k}, t) = \rho(\mathbf{k}, 0) e^{-Dk^2 t}.$$
(6)

This results in

$$\rho(\mathbf{k}, t) = \frac{1}{2\pi} e^{-Dk^2 t} A \int_{-\infty}^{\infty} e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\sigma^2} + i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x},$$
(7)

and if  $A = \frac{1}{\sqrt{2\pi}\sigma_0}$  for the normalized Gaussian function, along with  $\sigma$  being notated as  $\sigma_0$ , the

equation simplifies to

$$\rho(\mathbf{k}, t) = \frac{1}{2\pi} e^{-Dk^2t + ix_0\mathbf{k} - \frac{k^2\sigma_0^2}{2}}. \quad (8)$$

Taking the inverse Fourier transform,

$$\rho(\mathbf{x}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-[(Dt + \frac{\sigma_0^2}{2})k^2 + (-ix_0 + i\mathbf{x})\mathbf{k}]} d\mathbf{k}, \quad (9)$$

$\sigma(t) = \sigma = \sqrt{2Dt + \sigma_0^2}$  is defined for simplicity. Normalized variables are also used which were scaled by the height of the EMCCD,  $D' = \frac{D}{h^2}$  and  $\sigma'_0 = \frac{\sigma_0}{h}$ .

We perform the integration and using the definition of the error function, ( $\text{erf}(z) = \int_0^z \frac{2e^{-u^2}}{\sqrt{\pi}} du$ ), we finally arrive at the solution for  $n$ -dimensions for the charge density using the Gaussian function initial condition,

$$\rho(\mathbf{x}, t) = \left( \frac{e^{-\frac{(\mathbf{x}-x_0)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \right)^n, \quad (10)$$

along with the normalized solution,

$$\rho(\mathbf{x}, t) = \left( \frac{e^{-\frac{(\mathbf{x}'-x'_0)^2}{2\sigma'^2}}}{\sqrt{2\pi}\sigma'} \right)^n. \quad (11)$$

Similar to the Dirac delta function initial condition, we apply the boundary conditions (2) to the general solution, this time with the Gaussian function. The Gaussian function initial condition solution is obtained as the following,

$$Z(z')T(t) = \text{erf}(1) \sum_{n=0}^{\infty} e^{-\frac{\alpha_n^2}{2}(\sigma_0'^2 + 2D't)} \sin(\alpha_n z'_0) \sin(\alpha_n z'). \quad (12)$$

### iii. Charge Distribution

By investigating the charge flux at the boundary between the field-free region and depletion region (at  $z = 0$ ) integrated over time, we are able to obtain the distribution of charges,

$$q(\mathbf{r}) = D \int_0^\infty \left. \frac{\partial \rho}{\partial z} \right|_{z=0} dt. \quad (13)$$

We take the solution from <sup>3</sup>,

$$\rho(\mathbf{r}, z, t) = \frac{Q_0}{4\pi Dth} \cdot 2 \sum_{n=1}^{\infty} \sin(\alpha_n z'_0) \sin(\alpha_n z') \exp\left(-\frac{\alpha_n^2 Dt}{h^2} - \frac{r^2}{4Dt}\right), \quad (14)$$

and take the derivative with respect to  $z$ ,

$$\left. \frac{\partial \rho(\mathbf{r}, z, t)}{\partial z} \right|_{z=0} = \frac{Q_0}{4\pi Dth^2} \cdot 2 \sum_{n=1}^{\infty} \alpha_n \sin(\alpha_n z'_0) \exp\left(-\frac{\alpha_n^2 Dt}{h^2} - \frac{r^2}{4Dt}\right) \quad (15)$$

The charge,  $q_j$ , in each pixel,  $j$ , can be obtained by integrating the flux at the depletion zone boundary over both the pixel area and over time. This is done by integrating (15) in Cartesian coordinates,  $(x, y)$  and by integrating over time,  $t$ , such as in (13). Following this process in normalized coordinates, we arrive at

$$\begin{aligned} q_j &= D \int_0^\infty \int_A^B \int_C^E \frac{Q_0}{4\pi Dth^2} 2 \sum_{n=0}^{\infty} \alpha_n \sin(\alpha_n z'_0) \exp\left(-\frac{\alpha_n^2 Dt}{h^2} - \frac{r^2}{4Dt}\right) dx dy dt, \\ q_j &= \frac{Q_0}{2} \int_0^\infty \int_A^B \int_C^E \frac{1}{\pi} \sum_{n=0}^{\infty} \alpha_n \sin(\alpha_n z'_0) \exp\left(-\frac{\alpha_n^2}{4} \sigma^2 - \frac{r'^2}{\sigma^2}\right) d\left(\frac{x'}{\sigma}\right) d\left(\frac{y'}{\sigma}\right) d\sigma^2, \end{aligned} \quad (16)$$

where  $\sigma^2 = \frac{4Dt}{h^2}$ . Note that  $A$  and  $B$  are the  $x$ -coordinates representing the start and end of each pixel while  $C$  and  $E$  are the  $y$ -coordinates representing the start and end of each pixel. The center of the grid is represented by  $(x_0, y_0)$  which is considered to be  $(0, 0)$ . In addition, the coordinates are all normalized by the height,  $(x' = \frac{x}{h}, y' = \frac{y}{h}, z' = \frac{z}{h}, \mathbf{r}' = \frac{r}{h})$ .

Looking at only the spatial integral over area, we have

$$p_j = \frac{1}{4} \left( \operatorname{erf}\left(\frac{b}{\sigma}\right) - \operatorname{erf}\left(\frac{a}{\sigma}\right) \right) \left( \operatorname{erf}\left(\frac{e}{\sigma}\right) - \operatorname{erf}\left(\frac{c}{\sigma}\right) \right), \quad (17)$$

which results in the entire integral including integration over time,

$$q_j = \frac{Q_0}{2} \sum_{n=0}^N \alpha_n \sin(\alpha_n z'_0) \int_0^\infty \exp\left(-\frac{\alpha_n^2}{4} \sigma^2\right) \cdot p_j\left(\frac{1}{\sigma}\right) d\sigma^2, \quad (18)$$

as  $N \rightarrow \infty$ ,  $\alpha_n = (n + \frac{1}{2})\pi$ , and the rectangular pixel borders are at  $x = [a, b]$  and  $y = [c, e]$ .

This charge distribution function is evaluated along a 5 x 5 grid of square pixels, with the  $x$  and  $y$  coordinates being scaled to from  $-5$  to  $5$  and a pixel size of  $2 \times 2$  dimensionless units. Equations (17) and (18) will together demonstrate the charge distribution in the field free region (at  $z'_0 = 1$ ) with the initial cloud charge,  $Q_0 = 1$ , along the EMCCD grid.

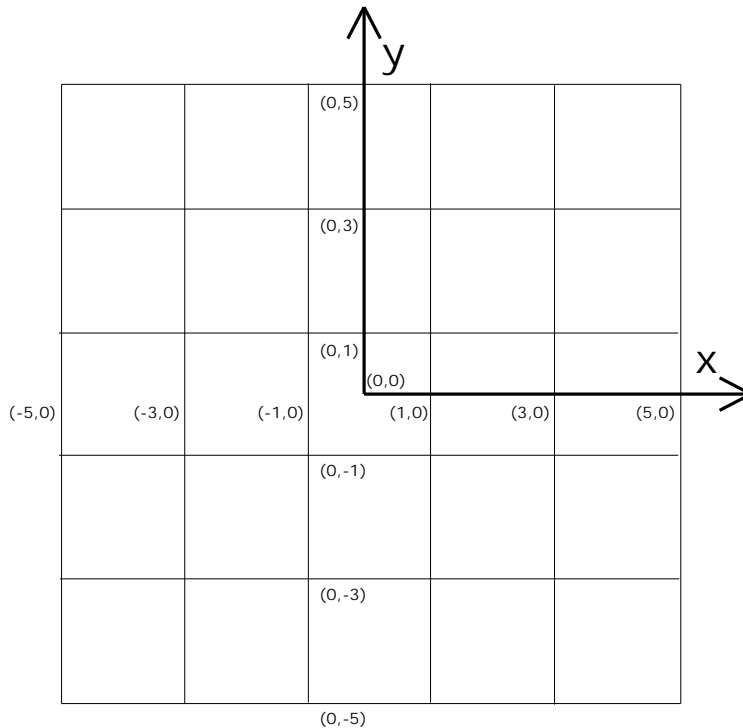


Figure 1: A 5x5 grid used to represent the pixels and their respective coordinates of the EMCCD.

## B. Numerical Integration

In order to evaluate and plot (18), we perform three different integration methods: Midpoint Rule, Trapezoidal Rule, and Simpson's Rule. In general, the Midpoint Rule approximates curves using straight horizontal lines from rectangles, the Trapezoidal Rule approximates curves using ramped straight lines from each trapezoid's leg, and the Simpson's Rule approximates curves using quadratic curves. Therefore, especially when it comes to more complicated functions, the Midpoint Rule provides the least accurate approximation while Simpson's Rule provides the most accurate approximation for an integral.

### i. Midpoint Rule

The midpoint between any two points,  $f(a)$  and  $f(b)$ , of a function,  $f(x)$ , is calculated by

$$\frac{f(a) + f(b)}{2},$$

which allows for numerical integration using the midpoints of the subintervals of a function.

The Midpoint Rule states that a definite integral over an interval,  $[a, b]$ , can be approximated by dividing the area under a function,  $f(x)$ , into  $n$  subintervals of equal width,  $\Delta x$ , and summing up the midpoints between each subinterval. The area between each subinterval is approximated by a rectangle. The Midpoint Rule is defined by

$$\int_a^b f(x)dx \approx \Delta x[f(x_1^*) + f(x_2^*) + \dots + f(x_n^*)], \quad (19)$$

where  $\Delta x = \frac{b-a}{n}$  is the width of  $n$  subintervals, and  $x_i^*$  is the midpoint of each subinterval,  $x_i$ .

### ii. Trapezoidal Rule

Similar to the Midpoint Rule, the area under a function,  $f(x)$ , is split up into  $n$  subintervals over the interval  $[a, b]$ . However, instead of rectangles, the subintervals are composed of trapezoids

whose legs try to approximate the curve with straight slanted lines. The area of a trapezoid is given by,

$$\left(\frac{A+B}{2}\right)H,$$

where  $A$  and  $B$  are the bases of the trapezoid and  $H$  is the height of the trapezoid. This concept can therefore be used to obtain the formula for the Trapezoidal Rule,

$$\int_a^b f(x)dx \approx \frac{\Delta x}{2}[f(x_0) + 2f(x_1) + 2f(x_2) + \dots + 2f(x_{n-1}) + f(x_n)], \quad (20)$$

where  $\Delta x = \frac{b-a}{n}$  is the width of  $n$  subintervals, and  $x_i$  is the point of each subinterval,  $x_i$ .

### iii. Simpson's Rule

Instead of approximating the curve with straight lines as with the Trapezoidal Rule, the function can be approximated with a quadratic function that has to agree with three points from the subintervals using Simpson's Rule.

For example, the area under the approximation on the intervals,  $[x_{i-1}, x_i]$  and  $[x_i, x_{i+1}]$ , would be given by

$$\frac{\Delta x}{3}[f(x_{i-1}) + 4f(x_i) + f(x_{i+1})].$$

By incorporating  $n$  subintervals, the general equation for Simpson's Rule is defined by,

$$\int_a^b f(x)dx \approx \frac{\Delta x}{3}[f(x_0) + 4f(x_1) + 2f(x_2) + \dots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)], \quad (21)$$

where  $\Delta x = \frac{b-a}{n}$  is the width of  $n$  subintervals, and  $x_i$  is the point of each subinterval,  $x_i$ .



## IV. Results

A 2D plot (Figure 2) of the charge distribution time integral at the center pixel is shown in order to demonstrate the overall behavior of the integral. The function is continuous along the entire domain and converges quickly past approximately  $\sigma = 4$ . The plot overall resembles a function similar to a Gaussian curve, due to the close relationship between the error function and the Gaussian function. Since the time integral is bounded from 0 to  $\infty$  and rapidly converges, the bounds from the integral are approximated to be from 0 to 10 in order to be numerically solved.

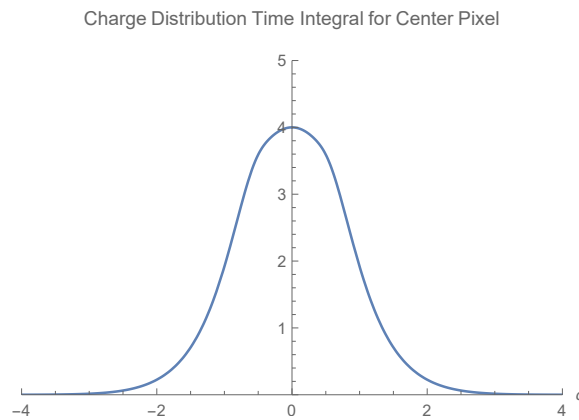


Figure 2: Plot of charge distribution time integral term for center pixel.

To compare the accuracy of the various numerical integration methods of the time integral, the charge distribution is plotted for  $n = 0$  in the summation (Figure 3). The integral is evaluated three different times from 0 to 10 with 10 subintervals using the Midpoint Rule, the Trapezoidal Rule, and Simpson's Rule. Due to the values around the center pixels rapidly decreasing, the functions were plotted on a logarithmic scale. The shapes of the plots are very similar, meaning that the different integration methods do not have a significant impact on the plot. However, based on the central pixel values, as well as some of the surrounding pixel values, Simpson's Rule produced slightly more accurate values. The Midpoint Rule and Trapezoidal Rule plots are identical in values, meaning that the quadratic approximation from the Simpson's Rule plot matches the curve more accurately. As the function shown in Figure 2 has a steep curvy shape, a quadratic approximation instead of a linear approximation for its integral would be more accurate. For the sake of

simplicity and accuracy, the rest of the results for the charge distribution will be shown using only Simpson's Rule.

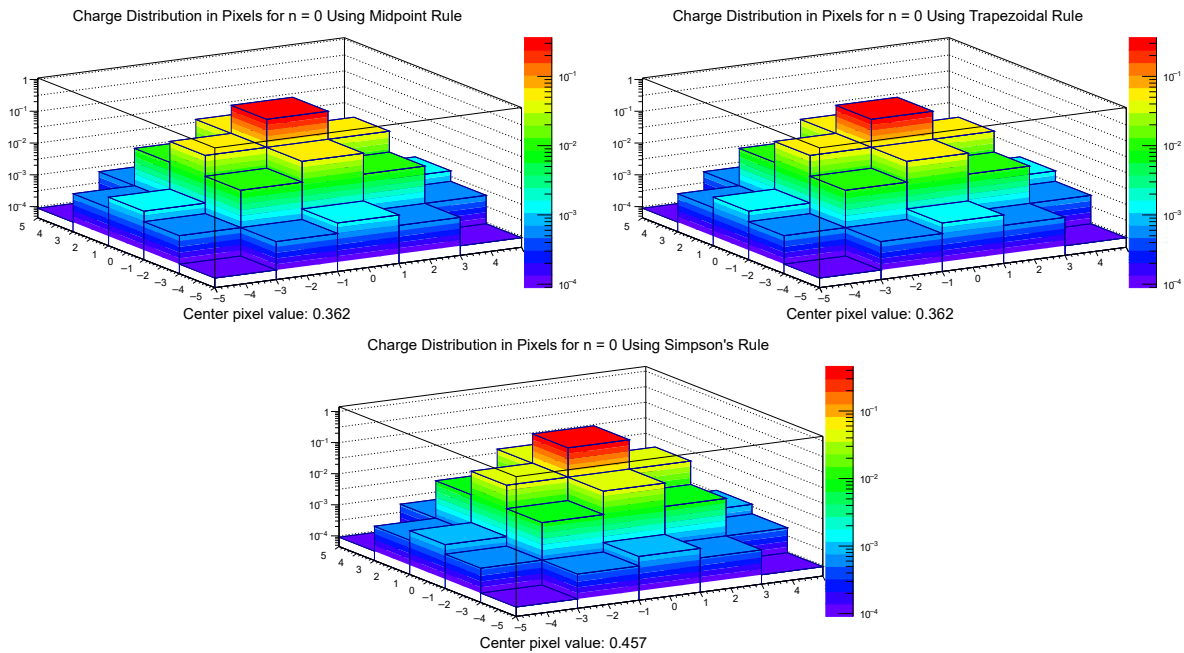


Figure 3: Comparison of evaluating the charge distribution integral with the Midpoint Rule, Trapezoidal Rule, and Simpson's Rule.

The full charge distribution function is plotted for different summation limits,  $N$ , using a logarithmic scale in Figure 4. For the purpose of demonstration, each of the values are also scaled by the center pixel value, meaning that the center pixel value is normalized to 1. For most accuracy, the number of subintervals in the numerical integration is 1000. Each summation for each value of  $n$  in the charge distribution function decreases, meaning that the summation converges very quickly. As large values of  $n$  would produce infinitesimal values, the shape and values of the plots change by only a very small factor. Due to this, the plots are not graphed past the summation limit of  $N = 100$ , as the values converge and do not produce any significant differences at higher limits. Each of the plots show an expected 3D Gaussian curve shape along the center pixel with symmetry along  $x = 0$  and  $y = 0$ . The values quickly decrease logarithmically around the center pixel, while the center pixel maintains its maximum value. The charge distribution is therefore centered along the center pixel and spreads radially with gradual values.

The full charge distribution is also plotted on a linear scale in Figure 5 to demonstrate the

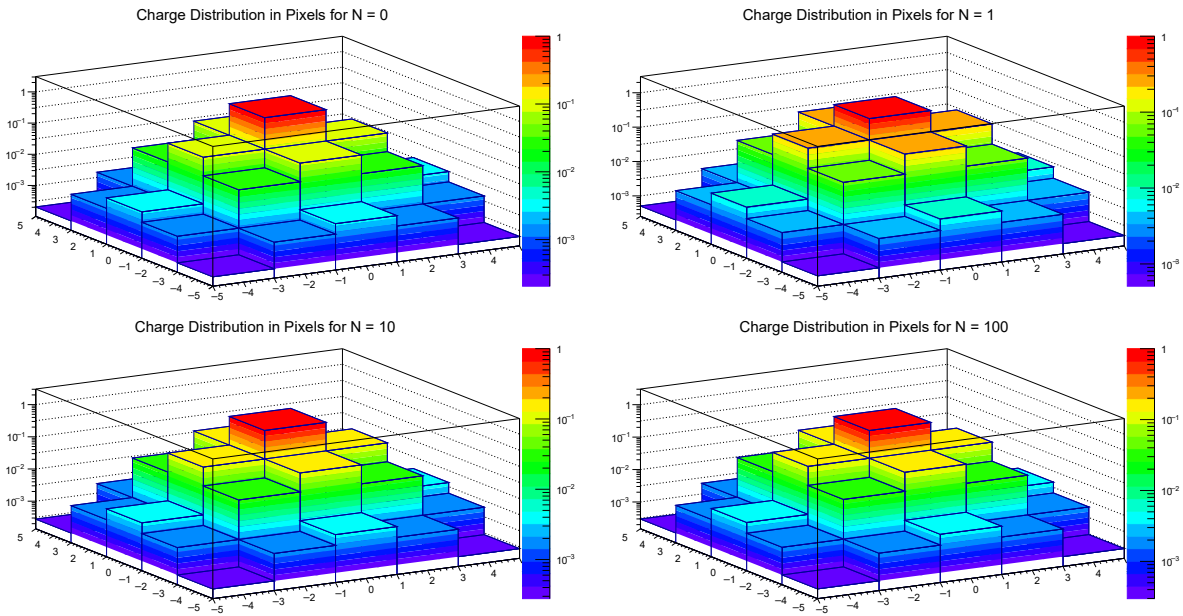


Figure 4: Charge distribution plots for different summation limits.

large concentration of charges in the center pixel. The summation limit is  $N = 100$  with 1000 subintervals used in the numerical integration and the values are scaled by the center pixel value similar to Figure 4. This plot provides a practical view of the charge distribution and shows the rapid drop in values at the pixels neighboring the center pixel. With the  $z$ -axis not being on a logarithmic scale, the 3D Gaussian shape becomes much thinner but still keeps its symmetry and behavior on a linear scale.

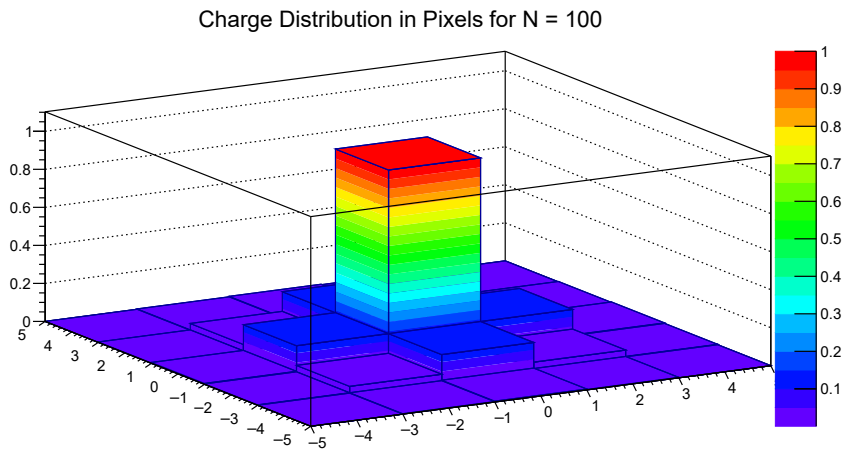


Figure 5: Charge distribution for  $N = 100$  on a non-logarithmic scale.

## **V. Conclusion**

After evaluating the charge distribution function, (18), the distribution of charges on the surface of the EMCCD sensor was able to be demonstrated. Since the initial condition for the initial electron cloud shape was modeled by a Gaussian function, the plots having a similar Gaussian shape match with the initial predictions. In addition, the charge density functions for both the Dirac delta function initial condition and the Gaussian function initial condition were solved for, which showed that a Gaussian function is a more natural and realistic approximation for the initial charge distribution instead of having the Dirac delta function at the center of the pixels model a single point charge. The time integral and summation of the charge distribution function both converge rapidly at low values, meaning that evaluating the charge distribution function does not require a lot of computation and that the function is well-behaved. Also, despite the various integration methods, all three of them produced very similar results, however Simpson's Rule was still the most accurate as it approximated the curve quadratically. Each of the histograms show that the charges are most focused in the center pixel as the x-ray beam is incident on the center of the 5x5 grid of pixels, and the charges spread out symmetrically at small values. These results provide a more accurate function to predict charge distribution and measure x-ray coordinates that can be used in applications of the NSLS-II SIX beamline, as well for other various EMCCD uses.

## **VI. Acknowledgements**

First, I would like to acknowledge my mentor, Ivan Kotov, for providing invaluable guidance throughout this project and granting me the opportunity to work on this project that allowed me to further improve my research skills. Secondly, I want to thank my intern collaborator, Andrea Houck, who provided great assistance to parts of this research, such as the mathematical modeling, as we both worked together on similar parts of the project. Lastly I would like to acknowledge the U.S. Department of Energy Office of Science and BNL's Office of Educational Programs, as this project was supported in part by the U.S. Department of Energy, Office of Science, Office of

Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

## VII. References

<sup>1</sup>I. Kotov, S. Hall, D. Gopinath, A. Barbour, J. Li, Y. Gu, K. Holland, A. Holland, I. Jarriage, J. Pellicciari, M. Soman, S. Wilkins, and V. Bisogni, "Analysis of the EMCCD point-source response using x-rays", Nuclear Inst. and Methods in Physics Research, A: Accelerators, Spectrometers, Detectors and Associated Equipment, **985**, 164706 (2021).

<sup>2</sup>G. R. Hopkinson, "Charge diffusion effects in CCD x-ray detectors", Nuclear Instruments and Methods **216**, 423-429 (1983).

<sup>3</sup>George G. Pavlov, "Charge diffusion in CCD X-ray detectors." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, **428**(2-3), 348-366 (1999).

<sup>4</sup>I. Kotov, "Diffusion in the field-free zone", (private communication) (2021).

Automated generation and machine learning analysis of transition metal containing active sites  
relevant to synthetic and biological applications

Robert Stahl, Computer and Information Sciences, University of Delaware, Newark, DE 19711

Atiya Banerjee, Department of Chemistry, Brookhaven National Laboratory, Upton, NY 11973

## **Abstract**

The geometrical makeup of metal active sites in proteins present an important opportunity for structure generation and machine learning in modern computational theory. Accurate generation of the active site geometries aid in accurate prediction of molecular and electronic level properties of reactive centers and in turn, produce better model of these systems. Current approaches for automatic generation in the chemical language makeup suffers from a lack of information related to the unique electronic aspects of transition metals. As such, the structure generation contains inaccuracies that lead to erroneous results.

In order to gain better control over the construction of these sites, the present work would be using principles of graph theory chemistry to automate the generation of these active sites efficiently. Our methodology uses connectivity theory in addition to empirical information obtained from quantum mechanical considerations to arrive at a representative structure. This approach has been successful in generating models of iron sites with multiple iron centers connected with various oxygen bridge coordinations. A software package capable of generating these models has been developed as a tool that can be integrated alongside other research. The tool is expected to benefit researchers evaluating these proteins with different computational techniques and improve the existing modeling software available to the community. The software package outputs chemical structures in the Protein Data Base (PDB) format for use by the community.

## **Introduction**

For metal active sites in proteins, very few accurate generation methods exist to create models of these complexes for calculation purposes. Chemical modeling languages like SMILES do not have implementation for active sites that contain multiple metals in the center. These languages have implementation centered on sites with one metal but do not extend to cover multiple metals. These metals in the site are connected using oxygen bridges and have ligand coordinations adjusted by these bridges. The goal of this project is to create a software package able to generate these bridge active sites given certain chemical properties. This would be accomplished by using graph theory transformations and techniques to generate these active sites according to given information. The scope of the project begins with generation of metal centers with only one metal and expands to double metal systems using the implementation of the one metal centers. These generations would act as starting models to be used by computational chemists for various experimental calculations. Ideally the software would be able to be implemented alongside machine learning analysis to generate accurate models of these active sites.



## Methods

The beginning objective of this package started out with the construction of monoiron systems with an octahedral coordination. A database containing the bond lengths between the iron and various ligands including water or pyridine in these monoiron systems was given from gathered data in another project. Ideal PDB files of these ligands were gathered from the Research Collaboratory for Structural Bioinformatics Protein Data Base<sup>1</sup> and reformatted to form a collection of usable ligands to be loaded in by the program.

The software itself started by prompting the user for the type of metal atom, the multiplicity of said atom and the name of the ligand to attach. From there, an array was filled with strings containing basic information to form an atom of the specified metal at the origin in the format given by the worldwide Protein Data Base<sup>2</sup>. Once this metal is placed at the origin, ligand information is loaded in centered at the origin from a PDB file corresponding to the ligand named in the input. This ligand information is stored in an array similar to the metal center and the ligand coordinates are rotated by constructed transformation methods to properly orient the ligand in the desired direction facing away from the metal. From there, the ligand is then translated from the origin by a distance designated by the aforementioned bond length database according to which ligand is loaded. After the rotation and translation, the ligand should be properly placed in space relative to the iron and a connection is created between the connecting atom in the ligand and the metal at the origin.

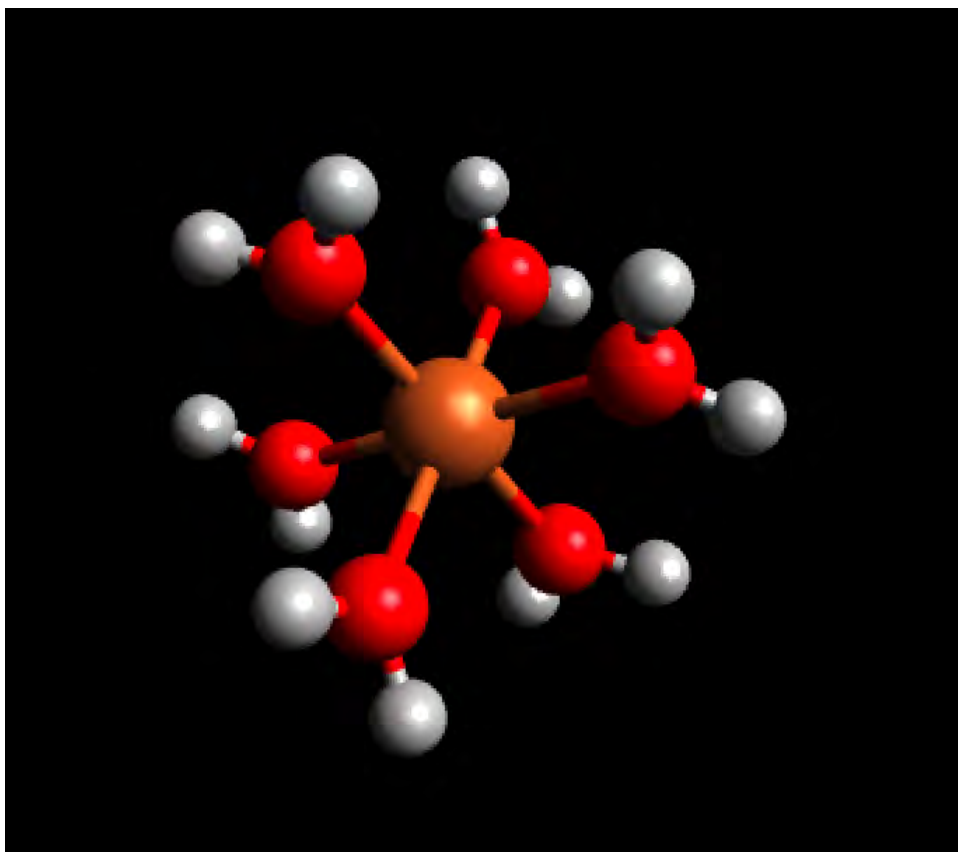
For the octahedral structure, six ligands are loaded and placed at a distance equal to the loaded in bond length along every axis in the positive and negative direction. Once all of the

---

1

2

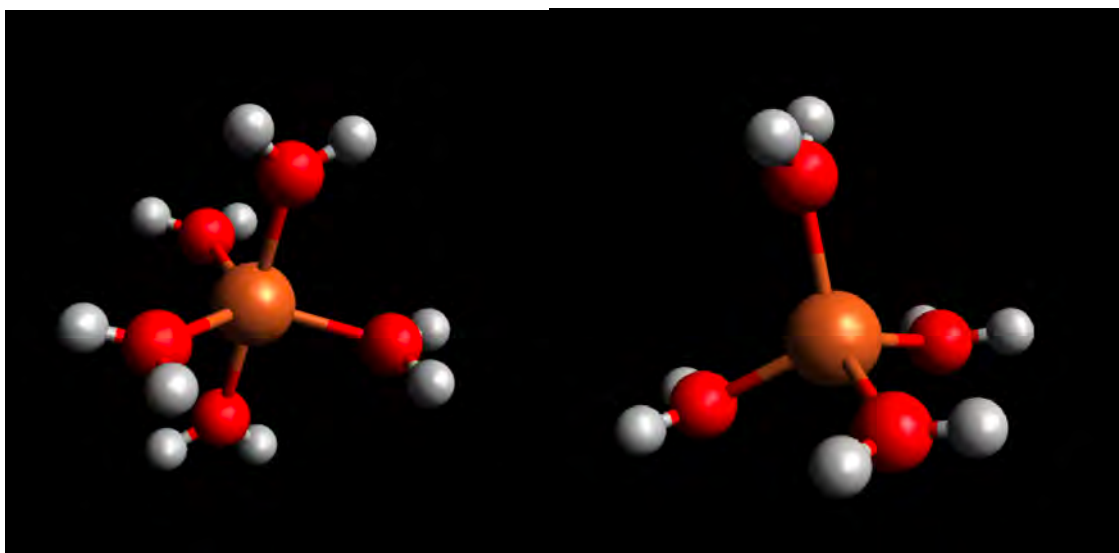
ligands are in place and connected, the information is then formatted and written to a PDB file named according to user input. This overall method was successful, leading to models of monoiron systems shown in Figure 1.



**Figure 1:** Generation of Hexacoordinated Iron with Water

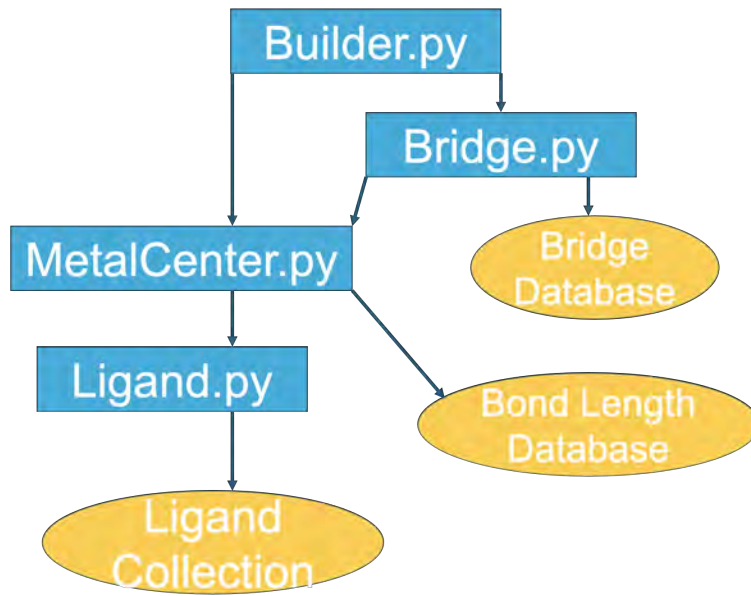
Upon successful generation, implementation was added to the software to allow bipyramidal and tetrahedral structures to be generated. To achieve bipyramidal, ligands were loaded in and placed in the positive-z and negative-z directions as well as in the positive-y direction. The last two ligands were loaded in the positive-y axis and rotated by  $120^\circ$  and  $240^\circ$  around the origin in order to achieve the proper bond angles. To achieve a tetrahedral structure, two ligands were placed initially: one in the positive-z direction and one in the positive-y direction. The ligand in the positive-y direction is then rotated around the x-axis in the negative-z

direction by  $19.5^\circ$  degrees in order to achieve a proper bond angle relative to the z-axis ligand. From there, this y-direction ligand is duplicated and rotated by  $120^\circ$  and  $240^\circ$  in order to form the final two ligands of the structure. Using these methods, successful generation of monoiron complexes with these coordinations were created as shown by Figure 2.



**Figure 2:** Generation of Bipyramidal and Tetrahedral Iron with Water

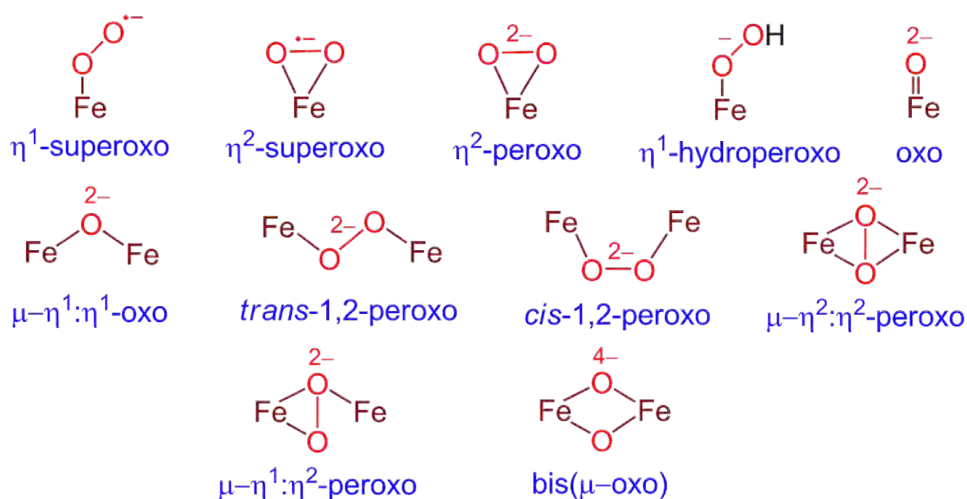
Upon being able to successfully generate monoiron centers, the next object of the package was to form bridge structures with two metals in the site. The first step toward this was to refactor the codebase into a more complex structure to allow new implementation for the bridge. The layout of this structure is shown by in Figure 3.



**Figure 3:** Project Software Structure Graph

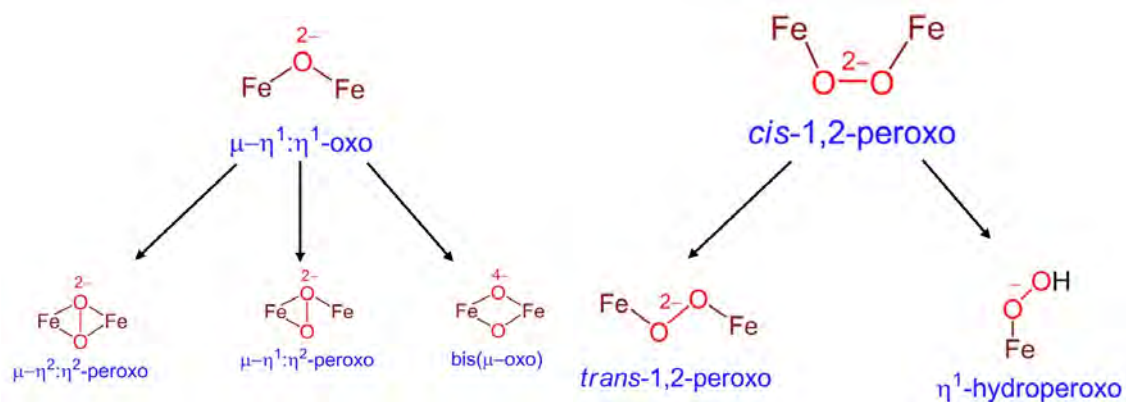
Builder.py is the central python file where the user is able to enter their inputs for their desired structure as well as input any irregularities into the structure. From Builder, either a Bridge structure object is created for sites with multiple metals or a MetalCenter object is directly created for a site with only one metal. The Bridge object generates the oxygen bridge information based on a bridge database that would load in the specifics of the bridge including bond angles and lengths. Then this Bridge object would generate two MetalCenter objects and orient them in space relative to the desired bridge structure. These MetalCenter objects would then construct their ligands as described earlier for monoiron systems using Ligand objects that directly take from the ligand PDB collection. Finally, the Bridge object adjusts the Ligand objects from the MetalCenter according to what bridge was loaded in and what coordination the centers have. Builder then writes the final output PDB file by traversing down through each layer and writing each atoms information and connections.

With this structure in place, implementation for the bridge structures was able to begin. Figure 4 shows what bridge structures were to be added to the package.



**Figure 4:** Visualization of Bridge Structures for Software Base

The top row indicates the different coordinations relative to only one iron in the complex while the bottom two rows indicate the overall bridge structure relative to both irons. To begin generating these bridges, a bridge database was extrapolated from complexes generated in previous work from chemical optimizations. Bond angles and lengths were extrapolated from optimized models and placed in an Excel sheet to be loaded in by the Bridge object. Once this was in place, general bridge methods were created such that they could be combined in order to generate the specific bridge desired. Figure 5 shows the idea behind how these general methods could be combined to generate specific bridge structures.

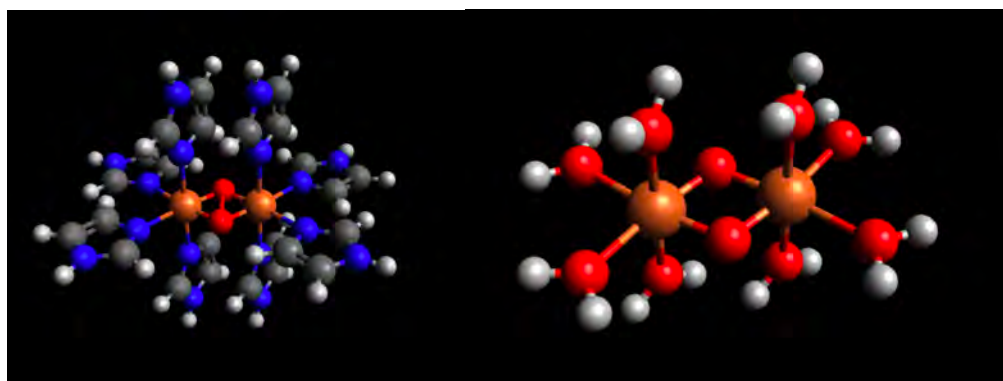


**Figure 5:** createOxo Implementation Graph and createPeroxo Implementation Graph

Firstly, a createOxo function was created to make a single oxygen connector similar to the  $\mu\text{-}\eta^1:\eta^1\text{-oxo}$  bridge. This was accomplished by loading in the desired data row from the bridge database and passing it into this function, which then places an oxygen atom on the y-axis and two irons on the x-axis in both directions from the origin. The positions on these axes are determined using trigonometric equations to extrapolate how far on each axis the atoms should be placed relative to the loaded in bond angles and distances. Once the first bridge connector is created and the metal atom coordinates are finalized, any secondary connectors such as in  $\text{bis}(\mu\text{-oxo})$  would be generated by finding its placement along the y-axis relative to the two metals and is mirrored over the x-axis in order to generate the final bridge shape.

After creating this function, a secondary function called createPeroxo was formed in order to generate bridges similar to the trans- and cis-peroxo bridges. This function works by forming two oxygen atoms and placing them both in opposite directions along the x-axis at a distance equal to half their bond length. From these oxygens, the metal atoms are placed according to the bond lengths and angles loaded in from the database using trigonometric equations to find their coordinates. This function allows the creation of not only these bridge structure but could theoretically be expanded to form hydroperoxo bridge coordinations as well.

Once these bridge structures are in place and the coordinates of the metals are found, then the coordinates are passed into the initialize functions of the MetalCenter objects and from there the Ligand objects are able to be generated off of these MetalCenter objects as described before. Once the Ligands are created, Bridge then calls a function called adjustLigands which formats the Ligands according to what type of bridge was formed. This function takes into account not only how many bridge connectors there are, but also whether a peroxo was created or not. Based on this information as well as the coordinations of the irons, the function deletes any Ligands that would be connected where the bridges are connected and then rotates the remaining Ligands to form the proper angular structure the metal should have. Once this is done adjusting, Builder traverses down each layer and outputs a PDB in the working directory that the program is run. Two of these models are shown in Figure 6.



**Figure 6:**  $\mu$ -n2-n2-peroxo with Octahedral Imidazole and Bis( $\mu$ -oxo) with Octahedral Water

These two complexes as well as two others containing similar bridge structures with different ligand placements were used to run chemical calculations in order to find the Mossbauer parameters of these complexes. The parameters were found to be similar to models of the same complex that had been hand-created and ran through the calculations. As such, these tests ran

show that the software is able to generate models that maintain the structural and electronic information of the complex and are able to be used for calculations in the future.

## **Conclusion**

Starting from monoiron implementation, the software package was able to expand and generate active site systems containing a variety of different bridges. This was accomplished by constructing multiple databases containing chemical information related to these systems and using this information alongside geometrical transformations to properly generate these active sites. These models were successfully generated and tested using calculations ran on handmade models that were ran using the same calculations. The software generated models were able to run on the calculations and were shown to be accurate compared to the handmade models. As such, this project could ideally be opened to the computational chemistry community to aid in generation of these models. This package could also be further expanded upon in multiple ways. Future work could include expanding the scope of the software to include more bridge structures as well as more complex coordinations concerning the ligands. The package could also be implemented alongside machine learning to analyze the chemical parameters of a system and generate starting models to optimize and perform further calculations on.



## References

1. Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, Philip E. Bourne, The Protein Data Bank, *Nucleic Acids Research*, Volume 28, Issue 1, 1 January 2000, Pages 235–242, <https://doi.org/10.1093/nar/28.1.235>
2. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol*. 2003 Dec;10(12):980. doi: 10.1038/nsb1203-980. PMID: 14634627.
3. F. Riaz and K. M. Ali, "Applications of Graph Theory in Computer Science," 2011 Third International Conference on Computational Intelligence, Communication Systems and Networks, 2011, pp. 142-145, doi: 10.1109/CICSyN.2011.40.
4. Hohenberger, J., Ray, K. & Meyer, K. The biology and chemistry of high-valent iron–oxo and iron–nitrido complexes. *Nat Commun* **3**, 720 (2012). <https://doi.org/10.1038/ncomms1718>

## **Acknowledgements**

I would like to give a special thanks to my mentor Atiya Banerjee for his guidance and support throughout this project. I would also like to thank Mehmed Ertem for his insights that aided me to create the project. I would like to extend my thanks to fellow interns Bryana Lopez and Ambika Natarajan for their help assembling the databases and running calculations.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

Anomaly detection algorithm research and testing with benchmark datasets for the performance analysis of High Performance Computing applications running at Exascale

Julianne Starzee, College of Arts and Sciences, Cornell University, Ithaca, NY 14853

Line Pouchard, Christopher Kelly, and Sandeep Mittal, Computational Science Initiative,  
Brookhaven National Laboratory, Upton, NY 11973

**Abstract:**

The study of certain physical phenomena, such as climate modeling, high energy physics, and molecular biology, requires simulations of these processes to be run on High Performance Computing systems at the Department of Energy. Since these simulation programs often contain thousands of lines of code and run simultaneously on thousands of nodes on a supercomputer, we need to make them as efficient as possible in terms of computing power. To do so, we extract a detailed performance trace when these simulations are running in order to analyze them for anomalies. However, the trace data that is output is often too large to be stored and analyzed at the end of the program's run. Thus, in order to detect anomalies when the programs are run at scale, we use Chimbuko, a program that uses *in situ* (in memory) anomaly detection algorithms, which detect anomalies in real time. Here, we look at two different anomaly detection algorithms, HBOS and SSTD, to determine which is more optimal for detecting global anomalies without detecting local anomalies. We use the Area Under the Receiver Operating Characteristic Curve algorithm as a metric. Determining more optimal algorithms will improve the performance of simulation programs by being able to detect all necessary outliers without detecting any false outliers.

**Introduction:**

Currently, computing speed is increasing much faster than the ability to store data on High Performance Computing (HPC) systems<sup>1</sup>. This presents a problem when attempting to analyze the trace data of large programs, as, often, the trace data is too large to store and analyze after the program has finished its run. This is why Chimbuko, a performance analysis framework that detects and analyzes anomalies *in situ* during the program's run, is needed<sup>2</sup>.

The detection of anomalously high function runtimes in a program is needed when analyzing the trace data of simulations of physical processes. For example, climate scientists use simulations to model the climate, fusion scientists simulate future experimental fusion devices, and materials scientists use simulations to study the structural properties of different materials<sup>1</sup>. These simulations are often very large programs that can run for days. Thus, it is necessary to ensure that these simulations are as efficient as possible so that they do not run for longer than needed.

To be able to increase efficiency, Chimbuko must be able to accurately detect anomalies, meaning it must be able to detect only global outliers without also detecting local outliers, or false positives. Global outliers are defined as data points in which the function runtime is much longer than usual. Global outliers thus significantly increase the runtime of the entire program and must be analyzed and removed to improve the program's efficiency. Local outliers are data points that are locally anomalous and are generally false outliers. These outliers do not affect the program's runtime and thus should not be detected so as to not waste time and storage space.

Chimbuko is currently implemented using two different anomaly detection algorithms known as HBOS and SSTD. HBOS is a histogram-based anomaly detection algorithm that employs dynamic histogram bin width<sup>3</sup>. HBOS sorts a set number of data points into each bin of the histogram of a set number of bins and determines outliers based on the height of the bin that the data point is in, with the bin height corresponding to the density of data in that area<sup>3</sup>. This means that it is not very dependent on the type of data distribution it is given. SSTD is an algorithm that determines if a point is an anomaly based on how many standard deviations away from the mean of the distribution a data point is. This is based on an input parameter, sigma. If a

data point is more standard deviations away from the mean than sigma standard deviations, it is classified as an anomaly by SSTD.

Here, we test HBOS and SSTD with different datasets and compare their performance.

### **Research Methods:**

To test the performance of HBOS and SSTD, we created multiple randomized synthetic data sets using the Python library Random to simulate trace data outputs of a computer program. The y-axis of these data sets represents a function's runtime, and the x-axis represents the time at which the function began its run. We created data sets with a single mode and with two modes, with each mode containing 2000 data points and being either a gamma distribution or normal distribution. We then manually add 2-10 outliers in random positions with runtimes between 200 and 1000 microseconds above the highest mode of the data set. We then input these data sets into a Chimbuko simulator, testing first the HBOS implementation and then the SSTD implementation.

Once the simulator has finished its run, we extract the anomalies that each of the algorithms found. We then graph the data, labeling points as either normal data points or as outliers, as found by the algorithm (Figure 1).

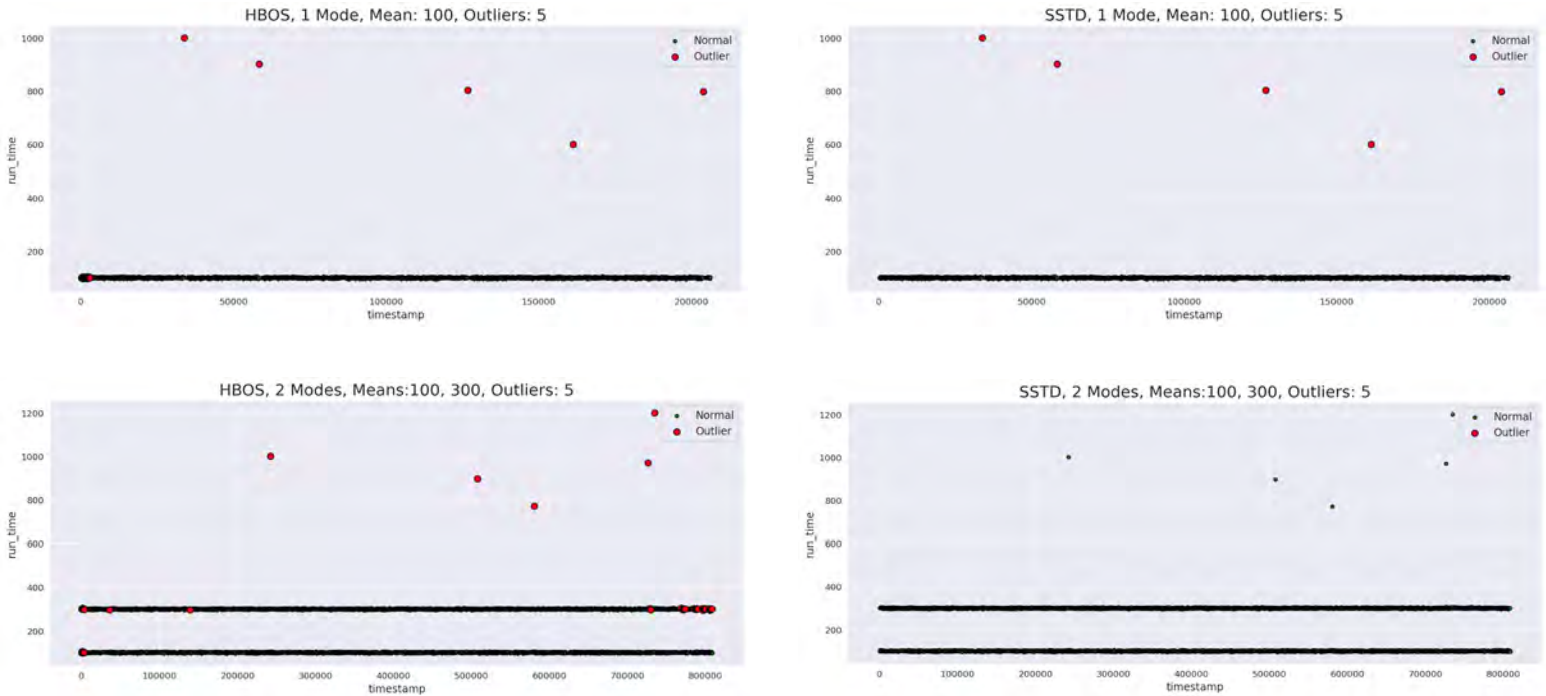


Figure 1: The red data points in these graphs represent anomalies, and the green data points represent normal data points. The graph to the top left represents a single mode data set centered at 100 put through HBOS, the graph to the top right represents the same data set put through SSTD, the graph to the bottom left represents a two mode data set with modes centered at 100 and 300 put through HBOS, and the graph to the bottom right represents the same data set put through SSTD.

For each Chimbuko run, we then calculate the Area Under the Receiver Operating Characteristic Curve (AUC ROC) score, which gives the area under the curve of the graph of the fraction of manually added outliers found by the algorithm out of the total manually added outliers vs. the fraction of local outliers found by the algorithm out of the amount of normal data points.

### Results:

We compare the AUC ROC scores in two different regards. First, we look at the difference in scores with respect to the amount of modes and type of distribution within the single mode data sets. This is shown in the table below (Figure 2).

AUC ROC Scores	Single Normal	Single Gamma	Double
HBOS	.9150500	.9086250	.9628056
SSTD	.7624833	.8707833	.6555556

Figure 2: The numbers in the chart represent the average AUC ROC scores of the different types of data sets.

HBOS performed slightly better than SSTD when processing single mode data sets. However, when processing data sets with two modes, HBOS performed significantly better than SSTD. This is most likely because SSTD assumes a normal distribution and calculates outliers based on their distance away from the mean. This means that when the algorithm received multi modal data sets, it classified both modes as being in the same distribution. Thus, the manually added outliers were not considered outliers by the algorithm because they were not far enough away from the effective singular normal distribution that the algorithm assumed. This can also be seen by the fact that, while HBOS performed very similarly when given single mode data sets of normal distributions and gamma distributions, SSTD performed worse when it was given single mode data sets of gamma distributions. This further shows that SSTD is less capable of handling single mode distributions that are not normal than HBOS.

We next compare solely the amount of true positives that the algorithms found in the table below (Figure 3).



False Positives Ratio	Single	Double
HBOS	.01588	.01883
SSTD	0	0

Figure 3: The numbers in the chart represent the average ratio of false positives detected by each algorithm to the amount of data points in the data sets for each kind of distribution.

While HBOS performed better according to the AUC ROC scores, it did find multiple false positives while SSTD found no false positives. However, in a lot of cases, most of the false positives that HBOS found were generally located at the beginning of the data set. This can be attributed to the fact that HBOS is an unsupervised algorithm, as it does not assume a certain type of distribution. This means it must observe and learn the behavior of the data set it is given before being able to accurately classify outliers. Thus, it will typically find many false positives in the beginning and then stabilize after it understands the behavior of the given data set.

The reason that the AUC ROC scores are lower for SSTD than they are for HBOS is that, though HBOS found many local outliers while SSTD found no local outliers, SSTD had more false negatives than HBOS. This is shown in the table below (Figure 4).

% True Outliers	Single	Double
HBOS	76%	95%
SSTD	76%	25%

Figure 4: The numbers in this table represent the average percent of outliers detected by the algorithms for each kind of distribution.

Even though SSTD did not detect any false positives, its AUC ROC score is much worse than that of HBOS for multi modal data sets because it detects little to no outliers for those data

sets. While detecting local outliers is wasteful, it is necessary to detect all of the global outliers because, if the algorithm used to detect outliers does not detect a global outlier in a program, that program will continue to run for an unnecessarily long time.

When decreasing the sigma input parameter of SSTD, the algorithm does begin to find the global outliers for multi modal data sets. However, it also begins to find local outliers. This is because a smaller sigma value will consider points closer to the assumed normal distribution to be outliers. Having to change the sigma value to different values depending on the data set presents a problem, however, because one does not know how the distribution of data will behave beforehand, so the algorithm should not depend heavily on input parameters.

**Conclusion:**

Overall, HBOS performs better than SSTD. Although HBOS finds false positives, in many cases, SSTD does not detect all of the global outliers of a data set. It is necessary to find all global outliers because we need to be able to remove them in order to decrease the runtimes of programs and make them as efficient as possible.

It is necessary for the algorithm used to not depend heavily on input parameters because that would require guessing the behavior of data sets before they are seen. The algorithm should also not assume a certain distribution because data sets may not be organized in a certain type of distribution. Algorithms that assume a certain distribution would thus be not as useful when analyzing a data set of a different type of distribution. In both these cases, HBOS would be the better choice than SSTD.

In order to be able to better and more efficiently study physical processes, it is necessary to find and improve algorithms that detect all of the global outliers of a data set without detecting any false positives. Currently, HBOS seems to be the best choice for implementing Chimbuko.

Further improving on this algorithm will allow scientists to study important topics more efficiently.

**Acknowledgements:**

I would like to thank my mentors, Line Pouchard, Sandeep Mittal, and Christopher Kelly, for all of their guidance on my project.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

**References:**

- <sup>1</sup> Foster, Ian et al. "Computing Just What You Need: Online Data Analysis and Reduction at Extreme Scales." European Conference on Parallel Processing. 01 August 2017. pp 3-19
- <sup>2</sup> Kelly, Christopher et al. "Chimbuko: A Workflow-Level Scalable Performance Trace Analysis Tool." ACM Digital Library. November 2020. pp 15-19.
- <sup>3</sup> Goldstein, Markus and Dengel, Andreas. "Histogram-based Outlier Score (HBOS): A fast Unsupervised Anomaly Detection Algorithm." German Research Center for Artificial Intelligence. September 2012.

Exploring Geometric Algebra for Beamline Operations

Taylor Stephens

Mathematics and Chemistry, University of Florida, Gainesville, FL 32611

Joshua Lynch

NSLS - II, Brookhaven National Lab, Upton NY, 11973

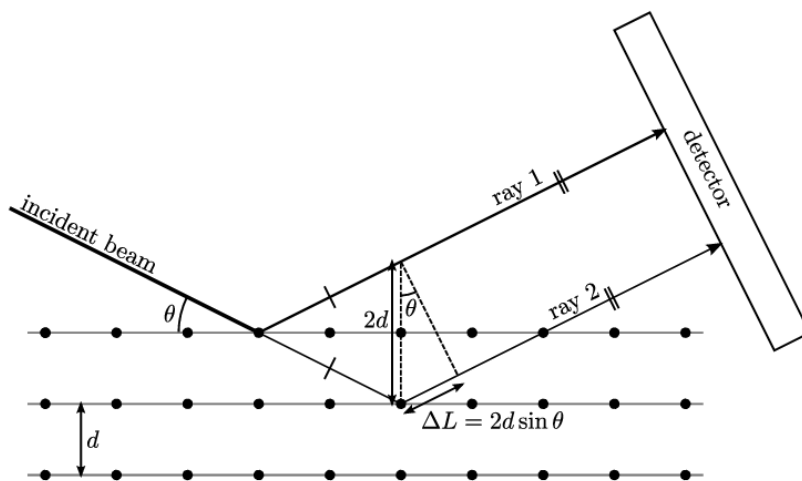
## Abstract

At Brookhaven National Laboratory, the National Synchrotron Light Source II (NSLS-II) houses 28 beamlines that use circulating electrons moving close to the speed of light to generate x-rays that are used to identify the structure and makeup of various samples and materials. As a result, crystallographic diffraction studies, commonly done at NSLS-II, rely heavily on beamlines as well as mathematics. In order to conduct these experiments, scientists must know how the crystals are oriented in their crystallographic coordinates or hkl space. Currently, the orientation and calibration of samples are calculated through matrices and vector algebra. Geometric algebra has the potential to be an effective alternative because of its ability to function in all dimensions and express algebraically common geometric objects such as points, lines, planes, spheres, etc. To investigate further, we used numerical and symbolic libraries such as SymPy and Galgebra to begin translating matrix and vector algebraic calculations to the geometric algebra system. We have successfully shown that the d-spacing formula can be expressed with only geometric algebraic operations and that rotors provide a more convenient method for completing rotations. As an aspiring mathematician this project has allowed me to explore a new system of math that is not commonly used which can serve to be a topic of research for my personal work in the future.

## I. Introduction

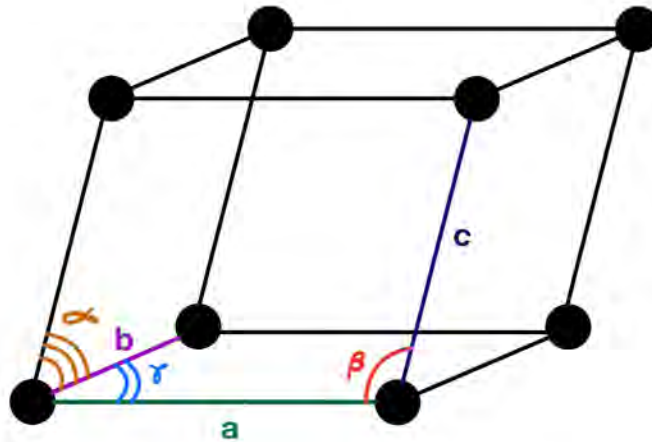
### A. D-spacing and Bravais Lattices

At Brookhaven National Laboratory, the National Synchrotron Light Source II (NSLS - II) allows scientists to identify elements with X-ray diffraction experiments. Crystallography X-Ray diffraction is used specifically to identify crystal samples based on the diffraction patterns of the crystal's unit cell.<sup>1</sup> The crystal of interest is placed on a diffractometer, or goniometer, which is built with various rotating elements such that the sample can be rotated to an appropriate position for the X-Ray beams to strike the unit cell. In order to determine the best position for the crystal sample, one must take into account the atom planes and parallel distances between each plane, known as the d-spacing, to analyze the diffraction pattern.



**Figure 1:** D-Spacing Example

The atom planes in a crystal are found with the unit cell basis vectors and the basis vectors' reciprocal vectors. In the case of a “perfect crystal” sample, the atomic planes can be described as one of the various types of Bravais lattice shapes.



**Figure 2:** Triclinic Primitive Bravais Lattice

## B. Properties of Geometric Algebra

All of these calculations are completed with traditional vector and matrix algebra, which does not always reliably remain valid in greater than three dimensions. As an alternative, we investigated the efficacy of geometric algebra for these calculations due to geometric algebra's ability to remain unchanged in all dimensions, particularly in 4D and 5D space, known as homogenous and conformal space, respectively.<sup>2</sup> Geometric algebra can also unify geometric aspects of complex numbers, quaternions, and matrices due to their shared properties. This unification allows the system to represent geometric objects such as points, lines, planes, circles, spheres, and conics as algebraic objects. Additionally, with geometric algebra one can define algorithms to find intersections and unions of geometric objects. Due to how fluidly geometric algebra can move from one dimension to another, each item in geometric algebraic space is given a grade: scalars are given the lowest grade of 0, vectors are given a grade of 1, bivectors are given a grade of 2, and so on.



### C. Operations of Geometric Algebra

In vector algebra, the dot and cross product describe an important relationship between vectors in 2D and 3D space. Because the dot product generates a scalar, it can exist in different dimensions, but the cross product does not hold value outside of 3D space. In geometric algebra, these are replaced by the inner and outer product, respectively. The inner product of vectors produces a scalar, similar to the dot product, but is a “grade lowering” operation.

Contrastingly, the outer product can be described as a “grade raising” operation that can produce bivectors, trivectors, and higher grade vectors, which are area and volume like objects, respectively. A new operation that is introduced in geometric algebra is the geometric product, which is defined by taking the sum or difference of the inner and outer products:

$$a*b = a \cdot b + a \wedge b \quad (1)$$

$$b*a = a \cdot b - a \wedge b \quad (2)$$

### D. Rotors

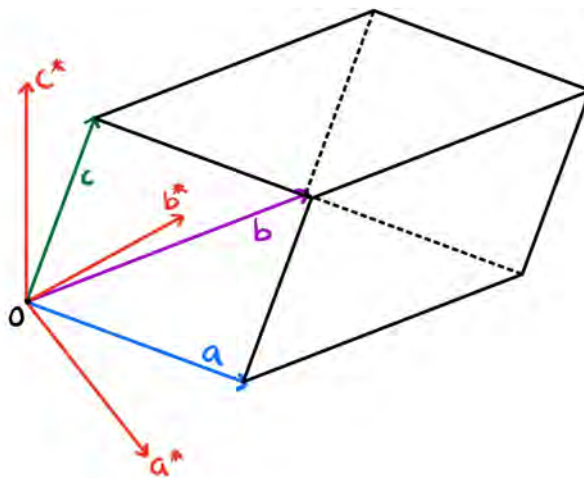
To complete rotations in vector algebra, one must use complex numbers in 2D space, and quaternions, in 3D space. An individual can also choose to complete rotations using matrices with matrix algebra. A consequence of quaternions is that it seems to function on a “left-handed” system, which can be difficult to navigate since vector algebra traditionally works on a “right-handed” system. Additionally, quaternions lose validity in higher dimensions similar to complex numbers. In geometric algebra, rotations are completed with rotors, which can be personally tailored to any desired angle or axis of rotation. Rotors can be constructed in any dimension and follow the right hand convention.

	Vector Algebra	Geometric Algebra
2D	$e^{i\theta}(a + ib)$	$R * \vec{a} * \tilde{R}$
3D	$Q * a * Q^*$	$R * \vec{a} * \tilde{R}$

**Table 1:** Summary of how rotations are completed in varying dimensions and algebraic systems

## II. Methods

With the pitfalls of vector and matrix algebra in mind, our investigation of geometric algebra began with how to use purely geometric algebra in the calculations of d-spacing. In order to calculate d-spacing, it is required to find the reciprocal vectors of the unit cell's basis vectors.<sup>3</sup>



**Figure 3:** Unit cell vectors and their reciprocal vectors

In vector algebra, to calculate the reciprocal of any vector  $\vec{a}$ , it is necessary to find the cross product of vectors  $\vec{b}$  and  $\vec{c}$ , the volume of the space created by  $\vec{a}$ ,  $\vec{b}$ , and  $\vec{c}$ , and then take the quotient of the two

$$a^* = \frac{\vec{b} \times \vec{c}}{V_c} \quad (3)$$

Normally, volumes are calculated with the triple product, which consists of the dot product between a vector and the cross product of two others

$$\vec{a} \cdot (\vec{b} \times \vec{c}) \quad (4)$$

To begin transitioning to the geometric algebraic system, we want to remove all uses of the cross product completely from the calculations of the reciprocal vector. As a result, we substituted the cross product operations with the outer product and the dual of the outer product (as this will describe a vector normal to the  $ab$  - plane). The volume was substituted with the magnitude of the triple outer product:

$$\vec{a} \wedge \vec{b} \wedge \vec{c} \quad (5)$$

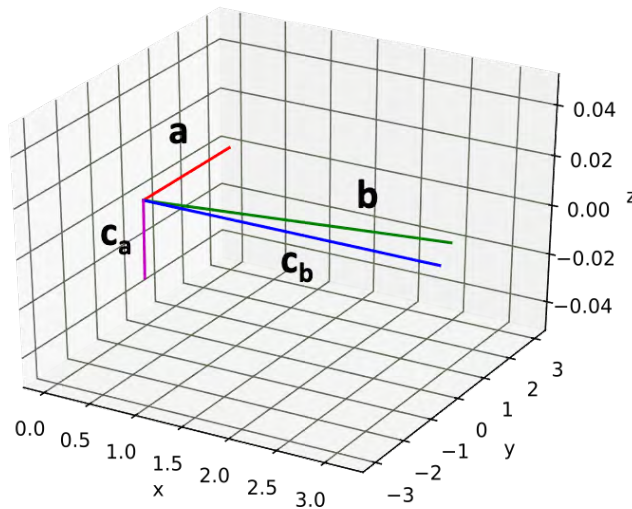
In order to confirm the validity of these substitutions, we used a Jupyter Notebook<sup>4</sup> with the SymPy<sup>5</sup> and Galgebra<sup>6</sup> symbolic libraries to test the following properties stated in Chapter 12 of *Thinking in Reciprocal Space*:

$$a^* \cdot a = b^* \cdot b = c^* \cdot c = 1 \quad (6)$$

$$a^* \cdot b = a^* \cdot c = b^* \cdot a = b^* \cdot c = c^* \cdot b = c^* \cdot a = 0 \quad (7)$$

In order to become comfortable with geometric algebra's use of rotors, we started a new Jupyter Notebook and continued to use SymPy, Galgebra, Clifford<sup>7</sup> and Numpy<sup>8</sup>. We then tasked ourselves to find the coordinates of an unknown  $\vec{c}$  within a Bravais lattice. There are various Bravais lattice formations, so we focused on the triclinic primitive lattice (as seen in Figure 2) in 3D space. A triclinic primitive lattice consists of three vectors of varying lengths as well as three different angles shared between them. When given only the lengths of two vectors and the three angles shared between all of them, we decided to use rotors and rotations of the two known vectors to find the third  $\vec{c}$  vector. We set one of the known vectors,  $\vec{a}$ , as the coordinate

axis and then, rotated the other known vector,  $\vec{b}$ , within the same plane to find the position of  $\vec{c}$  with respect to  $\vec{a}$  and  $\vec{b}$ .



**Figure 4:** Two vectors of triclinic primitive lattice and potential third vectors

The rotors and rotations were completed using the following equations where  $R$  is the rotor, and  $\tilde{R}$  is the inverse rotor.

$$b' = R * b * \tilde{R} \quad (8)$$

$$R = \frac{1 + \hat{b} * \hat{a}}{\sqrt{2(1 + \hat{a} \cdot \hat{b})}} \quad (9)$$

$$\tilde{R} = \frac{1 + \hat{a} * \hat{b}}{\sqrt{2(1 + \hat{a} \cdot \hat{b})}} \quad (10)$$

The two candidate  $\vec{c}$  vectors were then rotated around the  $\vec{a}$  and  $\vec{b}$ , which generated circles that have the potential to intersect at one or two points. These points of intersection

identified where the correct angle that  $\vec{c}$  would be from the  $ab$  atomic plane to the next plane parallel to it.

```
e1, e2, e3 = o3d.mv()
(a1, a2, a3) = sympy.symbols('a1 a2 a3', real=True)
(b1, b2, b3) = sympy.symbols('b1 b2 b3', real=True)
(c1, c2, c3) = sympy.symbols('c1 c2 c3', real=True)
(v1, v2, v3) = sympy.symbols('v1 v2 v3', real=True)
(w1, w2, w3) = sympy.symbols('w1 w2 w3', real=True)
n, m, p, g = sympy.symbols('n m p g')
```

```
a = a1*e1 + a2*e2 + a3*e3
b = b1*e1 + b2*e2 + b3*e3
c = c1*e1 + c2*e2 + c3*e3
v = v1*e1 + v2*e2 + v3*e3
w = w1*e1 + w2*e2 + w3*e3
```

```
#To create the unit vectors
ah = a/(a.norm())
bh = b/(b.norm())
ch = c/(c.norm())
```

```
#To help construct the rotors
m = ah|bh
n = 2+2*m
```

```
#Actual Rotors
R = (1+bh*ah)/sqrt(n.obj)
Ri = (1+ah*bh)/sqrt(n.obj)
R_n = (1+ah*bh)/sqrt(n.obj)
Ri_n = (1+bh*ah)/sqrt(n.obj)
```

```
#Sends vector a to b
ar = R*a*Ri
```

```
#Sends vector b to a
br = R_n*b*Ri_n
```

```
z = y.norm()
q = y/z
R1 = (cos(pi/4)- q*sin(pi/4))
R2 = (cos(pi/4)+ q*sin(pi/4))
```

**Figure 5:** Jupyter notebook of rotor calculations

To find the intersection, we used the meet operation, which is used specifically to identify the intersection of two blades.<sup>2</sup> In 2D space, the meet operation is represented by the relationship below:

$$L_1 \cup L_2 \equiv L_1^* \cdot L_2 \quad (11)$$

Outside of 2D space, the meet operation is described as an algorithm that is similar to a group of simultaneous equations. Because we worked with elements in conformal space we used the Clifford library to utilize the meet operation algorithm to locate the points of intersection.

### III. Results:

With respect to the d-spacing calculation, we found that replacing the cross products with outer products maintained the properties of reciprocal vectors and their counterparts. However, using the geometric algebraic iteration did not present a more efficient method but rather another method to calculate the reciprocal vectors. Equation 12 shows the final product of the substitutions and Figure 6 shows the Jupyter notebook used to confirm the aforementioned properties.

$$a^* = \frac{-I(\vec{b} \wedge \vec{c})}{|\vec{a} \wedge \vec{b} \wedge \vec{c}|} \quad (12)$$

```
[3]: e1, e2, e3 = o3d.mv()
(a1, a2, a3) = sympy.symbols('a1 a2 a3', real=True)
(b1, b2, b3) = sympy.symbols('b1 b2 b3', real=True)
(c1, c2, c3) = sympy.symbols('c1 c2 c3', real=True)
d, g, f, n, w, v, s, t, h, k, l = sympy.symbols('d g f n w v s t h k l')
Ar, Br, Cr = sympy.symbols('Ar Br Cr')

a = a1*e1 + a2*e2 + a3*e3
b = b1*e1 + b2*e2 + b3*e3
c = c1*e1 + c2*e2 + c3*e3
e123 = e1*e2*e3

[4]: w = -e123*(b^c)
v = a^b^c
Ar = w/(v.norm())

s = -e123*(c^a)
t = a^b^c
Br = s/(t.norm())

f = -e123*(a^b)
n = a^b^c
Cr = f/(n.norm())
```

```

[37]: (Ar|a), (Br|b), (Cr|c)
[37]: (1, 1, 1)
[38]: (Ar|b), (Ar|c), (Br|c), (Br|a), (Cr|b), (Cr|a)
[38]: (0, 0, 0, 0, 0, 0)

```

**Figure 6:** Jupyter notebook of d-spacing calculations

Our investigation of rotors allowed us to successfully identify how rotors can be used to find the coordinates of lattice vectors. The desired rotors could be generated using equations 9 and 10 or by using equations 14 and 15. Both methods are tedious so we created functions (seen in figures 7 and 8) to simplify the process.

```

def make_rotor(r1, r2, rotation_angle):
    # r1 and r2 are in the plane of rotation
    # make a vector or1 orthogonal to r1
    r1_hat = r1 / r1.norm()
    r2_hat = r2 / r2.norm()
    or1 = r2_hat - ((r1_hat|r2_hat)*r1_hat)
    or1_hat = or1 / or1.norm()
    print(f" r1_hat: {r1_hat}")
    print(f" r2_hat: {r2_hat}")
    print(f"or1_hat: {or1_hat}")
    # make a vector q1 such that the angle between r1 and q1 is one-half the rotation angle
    q1 = sympy.cos(rotation_angle / 2)*r1_hat + sympy.sin(rotation_angle / 2)*or1_hat
    q1_hat = q1 / q1.norm()
    R = r1_hat * q1_hat
    return R

```

**Figure 7:** Rotor function based on equations 9 and 10

```

def make_rotor(r1, r2, rotation_angle):
    # r1 and r2 are in the plane of rotation
    # make the normal bivector
    B = r1^r2
    Bhat = B/B.norm()
    print(f" B: {B}")
    print(f" Bhat: {Bhat}")
    # make the rotor using the exponential equation on pg. 125
    R = Bhat.exp(rotation_angle / 2)
    return R

```

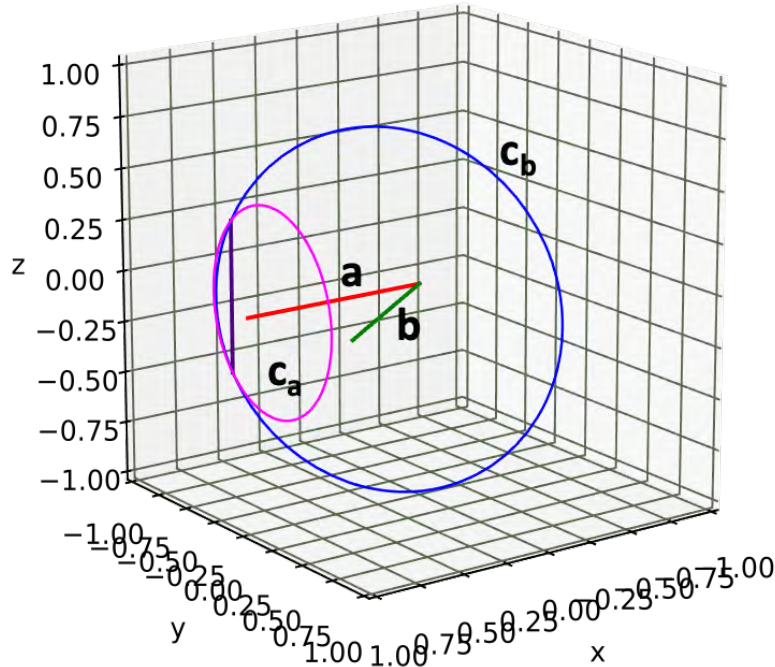
**Figure 8:** Rotor function based on equations 14 and 15

$$v' = e^{\frac{-\hat{B}\alpha}{2}} \cdot v \cdot e^{\frac{\hat{B}\alpha}{2}} \quad (13)$$

$$R = e^{\frac{-\hat{B}\alpha}{2}} = \cos \frac{\alpha}{2} - \hat{B} \sin \frac{\alpha}{2} \quad (14)$$

$$\tilde{R} = e^{\frac{\hat{B}\alpha}{2}} = \cos \frac{\alpha}{2} + \hat{B} \sin \frac{\alpha}{2} \quad (15)$$

To find the coordinates of these lattice vectors we began by making a convenient choice for  $\vec{a}$  by setting it as the coordinate axis. A convenient choice for  $\vec{b}$  was made by rotating a copy of  $\vec{a}$  by angle  $\gamma$  in the plane that both  $\vec{a}$  and  $\vec{b}$  lie in. As seen in Figure 4, the  $\vec{c}_a$  and  $\vec{c}_b$  were plotted at the angles  $\alpha$  and  $\beta$ .  $\vec{c}_a$  and  $\vec{c}_b$  are then rotated around the  $\vec{a}$  and  $\vec{b}$  respectively which generated a circle of all candidate  $\vec{c}_a$  and  $\vec{c}_b$  vectors. Alternatively, in conformal space these circles could have been generated by taking the outer product of three points that lie on the same circle.<sup>9</sup> The true position of  $\vec{c}$  is found from the intersection of the  $\vec{c}_a$  and  $\vec{c}_b$  vector circles. With the help of the meet operation and Clifford, two points of intersection were found in our test case. As seen in Figure 9 below.



**Figure 9:** Rotation of candidate third  $\vec{c}$  vectors



#### IV. Conclusion

After working with geometric algebra, it is obvious that it has both advantages and disadvantages. To begin with the disadvantages, there is a lack of efficient easy to use software due to how under-appreciated geometric algebra is. Without the appropriate software, completing calculations with geometric algebra is nearly impossible. Additionally, this is not a course commonly taught at the university level so many people continue to be unaware of this alternative system. However, when individuals are exposed to geometric algebra they are able to easily see its benefits. One of the biggest advantages found during our investigation is in geometric algebra's use of rotors. Unlike with complex numbers and quaternions, rotors allow one to conveniently define an axis of rotation. Furthermore, rotors can be used in any dimension which removes the need to transition between different methods such as having to transition from complex numbers to quaternions when going from 2D to 3D space.

Along with rotors, the meet operation is a benefit of geometric algebra because of how it conveniently solves "systems of equations" to find the intersections of geometric objects. Overall, geometric objects are effortlessly expressed as algebraic items which is helpful in the conformal model. Eventually, it could be implemented in goniometers to help determine necessary rotations to move between *hkl* planes.

As previously mentioned, geometric algebra is not commonly taught at the university level which leaves current and aspiring scientists unaware of its potential in various fields. Once people begin to recognize its benefits, this new mathematical system can be used as a new standard for completing complex computations at beamlines and beyond.

## V. References

- <sup>1</sup>Downs, B. Crystallography III, X-Ray Diffraction.
- <sup>2</sup>Vince, J. *Geometric Algebra: An Algebraic System for Computer Games and Animation* (SPRINGER LONDON LTD, 2016).
- <sup>3</sup>Williams D.B., Carter C.B. (2009) Thinking in Reciprocal Space. In: Transmission Electron Microscopy. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-76501-3\\_12](https://doi.org/10.1007/978-0-387-76501-3_12)
- <sup>4</sup>Executable Books Community. (2020). Jupyter Book (v0.10). Zenodo. <https://doi.org/10.5281/zenodo.4539666>
- <sup>5</sup>Meurer A, Smith CP, Paprocki M, Čertík O, Kirpichev SB, Rocklin M, Kumar A, Ivanov S, Moore JK, Singh S, Rathnayake T, Vig S, Granger BE, Muller RP, Bonazzi F, Gupta H, Vats S, Johansson F, Pedregosa F, Curry MJ, Terrel AR, Roučka Š, Saboo A, Fernando I, Kulal S, CimrmanR, Scopatz A. (2017) SymPy: symbolic computing in Python. *PeerJ Computer Science* 3:e103 <https://doi.org/10.7717/peerj-cs.103>
- <sup>6</sup>Bromborsky, A. Computer program, GALGEBRA, <https://github.com/pygae/galgebra>
- <sup>7</sup>Kern, R. Computer program, CLIFFORD, <https://github.com/pygae/clifford>
- <sup>8</sup>Harris, C.R., Millman, K.J., van der Walt, S.J. et al. *Array programming with NumPy*. Nature 585, 357–362 (2020). DOI: [0.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2). (Publisher link).
- <sup>9</sup>Leo Dorst, Daniel Fontijne, and Stephen Mann. 2009. *Geometric Algebra for Computer Science: An Object-Oriented Approach to Geometry*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- <sup>10</sup>Van Rossum, G., & Drake, F. L. Computer program, PYTHON, (Scotts Valley, CA: CreateSpace). <https://www.python.org>
- <sup>11</sup>W. R. Busing and H. A. Levy, *Acta Cryst.* 22, 457 (1967).

## **VI. Acknowledgements**

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

# Statistical Analysis of Solar Generation Intermittency

Justin Tondo, Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11790

Meng Yue, Department of Interdisciplinary Science, Brookhaven National Laboratory, Upton, NY 11973

## Abstract

Intermittency in solar generation due to cloud cover may cause significant increase or decrease in generation within a short time period. Such intermittency may have the similar impacts on the grid operation, therefore, can be considered as generator outages, i.e., intermittency induced outages (IIOs). The uncertainties that come with renewable energy generation are likely to be a major concern soon as more renewables are installed. A statistical analysis is to be performed to by gathering data on these transients or ramps which result in these IIOs and identify more severe common mode outages (CMOs), where transients occur at almost the same time from co-located or geographically close solar generation sites. To perform the analysis, time-series solar power and irradiance data and a MATLAB script are used to extract the transients and analyze the data for how often this reduction in generation occurs (i.e., outage frequencies) and how long it lasts (i.e., duration time). Based on the solar generation and irradiance data, the correlation between the data for the different sites has been analyzed and compared with the distances between the site locations. It has been found that the correlation increases as the distance between locations decreases, and as expected, these solar farms with a greater correlation result in a greater number of concurrent transients, and thus a greater potential for CMOs. This study is based on publicly available solar power and irradiance data. A guide to obtaining the solar power and irradiance data has also been created, as well as a map of the locations used in the data. This research will help to provide us with a better understanding of the uncertainties that come with renewable energy. As a result of this internship, I have gained more experience with finding and processing applicable solar power and irradiance data, writing programs in MATLAB, using MATLAB and Excel together, and understanding how location and distance between solar farms along with solar generation variability affects the potential for system problems.

# I Introduction

The main purpose of this project is to analyze solar generation and irradiance data to better understand and identify IIOs and CMOs when they occur. Intermittency in solar generation can cause rapid decreases or increases in power output due to passing clouds, and may cause concurrent changes in generation in co-located solar generation sites. Installing solar power introduces these uncertainties that are different from those of conventional grid components. Contingency analysis (CA) is one of the major transmission planning technique that consider the impacts of credible outages on the transmission reliability in the power grid. These non-conventional, solar related uncertainties due to the randomness of the cloud cover and other factors are not included in the current deterministic or probabilistic CA practices, i.e., DCA that considers the consequences of outage occurrences or PCA that considers both the consequences and likelihood of outage occurrences. Statistics of solar related IIOs or CMOs will be needed to include them in PCA. The need to identify these outages will escalate as the amount of solar energy added to the grid increases quickly. For example, solar PV accounted for 58% of all new electricity-generating capacity additions in Q1 2021.<sup>1</sup>

One of the main goals for this project is to obtain solar power and irradiance time-series data to analyze, as well as to sanitize and clean the data-set. Another goal is to analyze the correlation and distance between different locations of solar generation sites. Finally the relationship between the occurrence of concurrent transients and correlation/distance for multiple solar generation sites is to be analyzed. The methods behind accomplishing these goals and the end results are shown in this report.

## II Methodology

### II.I Time-series Data

Fast intermittency induced transients/ramps that can be considered as outages are investigated for solar generation data, as well as the correlation between solar generation and intermittency data at different solar farm locations. Time-series solar generation data is used to extract the transients, and perform the statistical analysis, while both solar generation and irradiance data is used to observe the correlation between the data and location.

The time-series solar generation was obtained from the Solar Power Data for Integration Studies data set from NREL.<sup>2</sup> The data set is for the entirety of the year 2006, taken at 5-minute intervals for the state of New York. The time-series irradiance data was taken from the National Solar Radiation Database (NSRDB) for a chosen section of New York at 30-minute intervals for the entirety of the year 2006.<sup>3</sup> A small section of southern New York was chosen to obtain the irradiance data which contained several of locations. Both datasets include the latitude and longitude coordinates of each solar generation site location, and the solar generation dataset labels each location with the installed capacity in MW.

Although the solar power data data set from NREL includes data for all states in the U.S.A., a small sample size was chosen with the locations in close proximity to each other. Only the data for New York was obtained, and from that 27 unique locations in the same general area of southern New York were chosen to be analyzed. The latitude and longitude coordinates, as well as the installed capacity in MW for each location is shown in table 1. A map of the 27 locations was also made, showing the discarded data as red markers, the 27 power locations as blue markers and the 27 irradiance locations as yellow markers as shown in figure 1. The reason the irradiance data has slightly different locations is because the data from the NSRDB did not use the exact same coordinates as the power data, and so the closest locations were chosen.

Table 1: Solar generation data latitude, longitude and installed capacity

Plant Number	Latitude	Longitude	Installed Capacity (MW)
1	41.25	-73.55	38
2	41.25	-74.15	29
3	41.25	-74.25	126
4	41.35	-74.45	23
5	41.45	-73.85	29
6	41.45	-74.55	88
7	41.55	-73.65	101
8	41.55	-73.75	76
9	41.55	-73.95	50
10	41.55	-74.15	50
11	41.55	-74.25	101
12	41.55	-74.45	88
13	41.65	-73.55	25
14	41.65	-73.65	101
15	41.65	-73.75	50
16	41.65	-74.25	76
17	41.75	-74.05	76
18	41.75	-74.15	35
19	41.85	-74.25	13
20	41.95	-73.85	50
21	41.95	-74.15	63
22	41.95	-74.35	88
23	42.05	-73.75	38
24	42.05	-73.85	88
25	42.05	-74.25	13
26	42.15	-73.75	13
27	42.15	-74.05	13

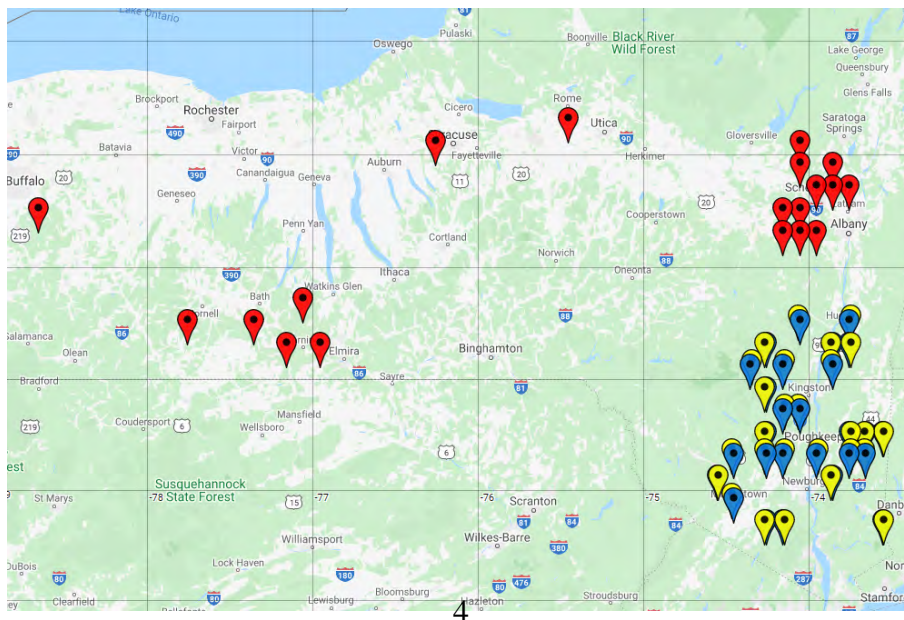


Figure 1: Map of 27 Solar Power/Irradiance Locations<sup>4</sup>  
Office of Educational Programs, 2021 Compilation of Internship Reports

The solar power data was filtered down to these 27 locations by eliminating day ahead and 4 hour ahead forecast data, while keeping actual real power output data as the NREL site provides all 3 types. There are also utility scale PVs (UPVs) and distributed PVs (DPVs), and UPV data was kept while DPV data was discarded as UPVs have more impacts on the grid and contain more high frequency variability, and thus more transients. Finally, any data too far from the main cluster of locations was discarded as seen by the red markers on the map so that co-located solar generation sites are the focus.

The solar irradiance data was gathered by selecting the area where the 27 solar power locations are located, and then finding the closest matches in terms of location as the latitude and longitude may differ by 1 or 2 hundredths of a degree like 41.45, -73.85 compared to 41.45, -73.86. The maximum difference between the solar power and irradiance data locations is about 2.78 km.

## **II.II Data Visualization Examples**

An example of the time-series power data is shown in figure 2, showing the displaying the output power of a 33 MW over the course of a single day in mid-January of 2006. There are obvious peaks and valleys showing how much the output of a solar plant can vary over the course of a day.



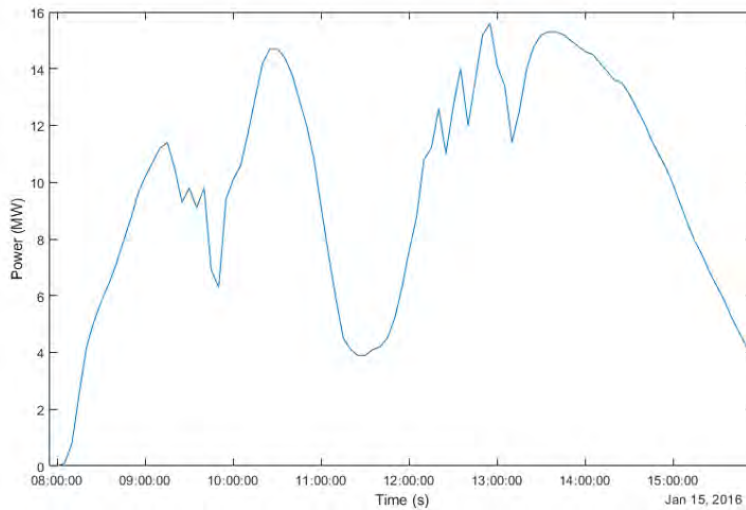


Figure 2: Solar Power Generation Over the Course of a Day Example

A more important thing to note is that co-located solar generation sites can experience the same drop or rise in production at the same time, which will result in a more drastic change in output overall. When this occurs between 2 co-located plants it is a double transient, and when it occurs between three co-located plants it is a triple transient. This is what would cause a common mode outage as mentioned earlier, and an example of this concurrent change for multiple plants is shown for the same day in figure 3. Such transients can be extracted by developing and applying MATLAB scripts. Note that the ramping events at different generation sites do not match up perfectly as the plants are not in the same exact location, but there are clear trends where all 3 have a drop or rise in output almost at the same time.

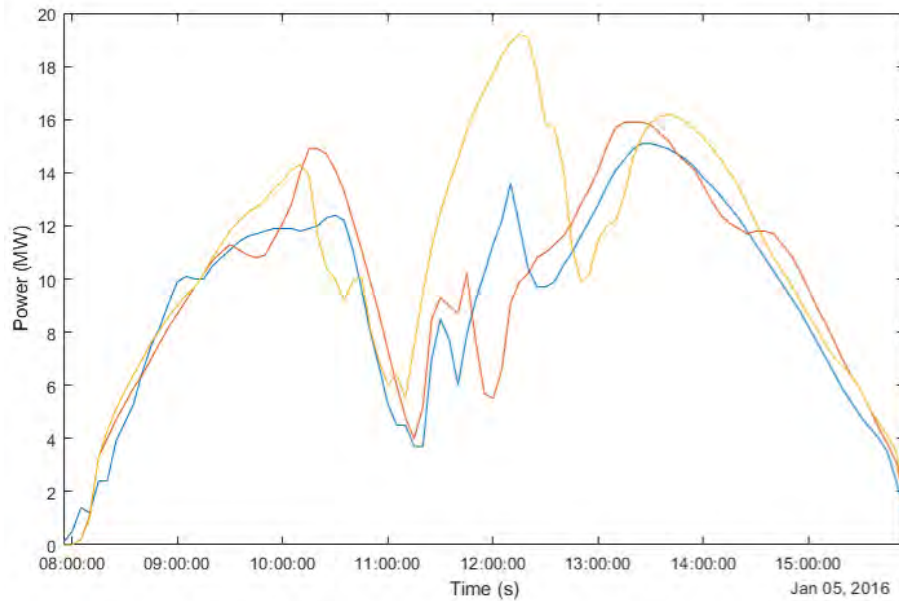


Figure 3: Solar Power Generation Over the Course of a Day for 3 Co-located Plants Example

Correlation between the two sets of time-series data can be used to measure the tendency of the data changes together. If the power output data for these 3 plants continued to closely follow each other in this way, then the correlation between the plants would be found to be closer to 1 than most. In this study, we will investigate the relationship between correlation and distances of solar plants and the correlation and the number of fast ramping events at multiple solar plants.

### II.III Transient Extraction/Analysis Method

The other subject of this project requires extracting the transients for the entire year of solar power data, and a quick explanation of that process is shown in figure 4 by describing a generic IIO.

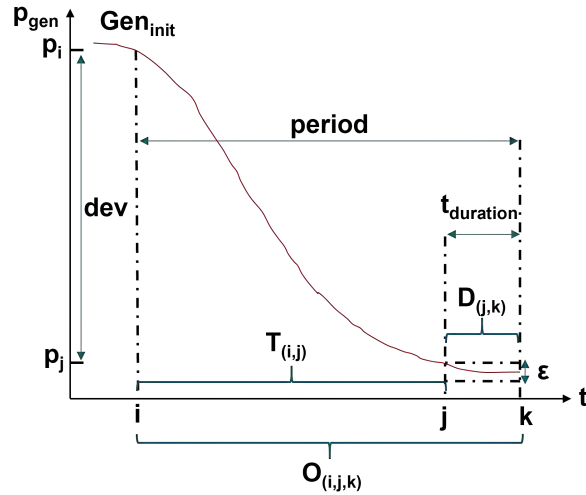


Figure 4: Transient Extraction Methodology

The IIO  $O_{(i,j,k)}$  consists of the transient,  $T_{(i,j)}$  which is the main sloping portion, and a flat period  $D_{(j+1,k)}$  which continues until the output generation rises or drops again. The output generation must vary by a certain amount to be counted as a transient which is denoted by  $\delta$  and called the deviation. This parameter can be lowered to include more transients in the final results, or raised to exclude more transients. For this project, the deviation  $\delta$  was set to 35%.  $\epsilon$  is the range of the highest or lowest generation level once the  $D_{(j+1,k)}$  portion is entered. Note that only monotonically ramping events that satisfy the given criteria will be extracted because fast generation changes are of particular interest in contingency analysis.

A MATLAB script was developed to extract fast transients from wind farm data, and has now been slightly modified to do the same for solar farm data. The script first extracts single transients, then double and triple transients separately. The solar power data is in an excel file with each solar farm's data in its own column, so in this project there are 27 columns and about 105,050 rows of 5-minute intervals data for the entire year of 2006. The data for the double and triple transients are exported to an output folder where they can then be used in another MATLAB script developed to analyze and visualize the data. This other script will be referred to as the processing script to avoid confusion.

The processing script takes the .mat files which contains the double and triple transient

data, and separates the columns of data into individual arrays, such as the SerialID array which assigns a number to each unique combination of double or triple transient plant groupings. The other arrays are the individual plants that make up each combination, and the correlation arrays which hold the correlation between the plants. The average distance between the plants is calculated for each grouping and also stored in an array. The number of transients for each unique grouping is found by counting how many times the unique grouping appears in the data. Once the data is cleaned and all repeating rows are removed, the data is visualized by plotting the amount of concurrent transients for each grouping versus the average distance/correlation between the plants. For the double transients, just the regular correlation and distance are plotted as there is no need to average the data since there are only 2 plants per grouping.

## II.IV Correlation/Distance Analysis Method

The processing script also calculates the correlation and distance between the individual solar power and irradiance measurement locations. This requires the excel files that contain the power and irradiance data, as well as the latitude and longitude .mat files for the locations to calculate the distance.

The correlation is calculated in the form of a matrix with dimensions equal to the number of locations by the number of locations. The code to calculate the correlation matrix was taken from the original wind transients extraction script and used in this processing script. The distance between plants is also calculated in the form of a matrix of equal size, and uses the Haversine distance formula and equations shown below, where the  $r = 6371$  (radius).

$$\delta_{lat} = \frac{lat_2}{180} - \frac{lat_1}{180} \quad \delta_{lon} = \frac{lon_2}{180} - \frac{lon_1}{180} \quad (1)$$

$$a = \sin\left(\frac{\delta_{lat}}{2}\right)^2 + \cos(lat_1) \cdot \cos(lat_2) \cdot \sin\left(\frac{\delta_{lon}}{2}\right)^2 \quad (2)$$

$$distance = r \cdot (2 \cdot \text{atan2}(\sqrt{a}, \sqrt{1-a})) \quad (3)$$

The output of this part of the processing script is correlation versus the distance for all of the plants using the power data and the irradiance data in 2 separate plots, as well as the irradiance correlation data versus the power correlation data, and the irradiance distance data versus the power distance data. Since the correlation and distance data is in the form of 27x27 matrices for this case, the data is either plotted in 27 different plots each with 4 subplots, or for one specific plant in 4 separate plots.

One final added feature of the processing script developed for this project is the ability to plot the power and irradiance data for a specific plant, and specific day of the year.

### **III Results/Discussion**

The following results are outputs from the processing MATLAB script previously mentioned, for the 27 solar power and irradiance data locations. The correlation/distance analysis and double/triple transients analysis results are included.

#### **III.I Correlation/Distance Analysis Results**

The correlation/distance comparison for location 5 is shown below for the solar power data and solar irradiance data in figures 5 and 6 respectively. Location 5 was chosen arbitrarily to demonstrate what kind of output is to be expected from this analysis. The output data is similar between all of the plants, with some slightly better or worse than the others.

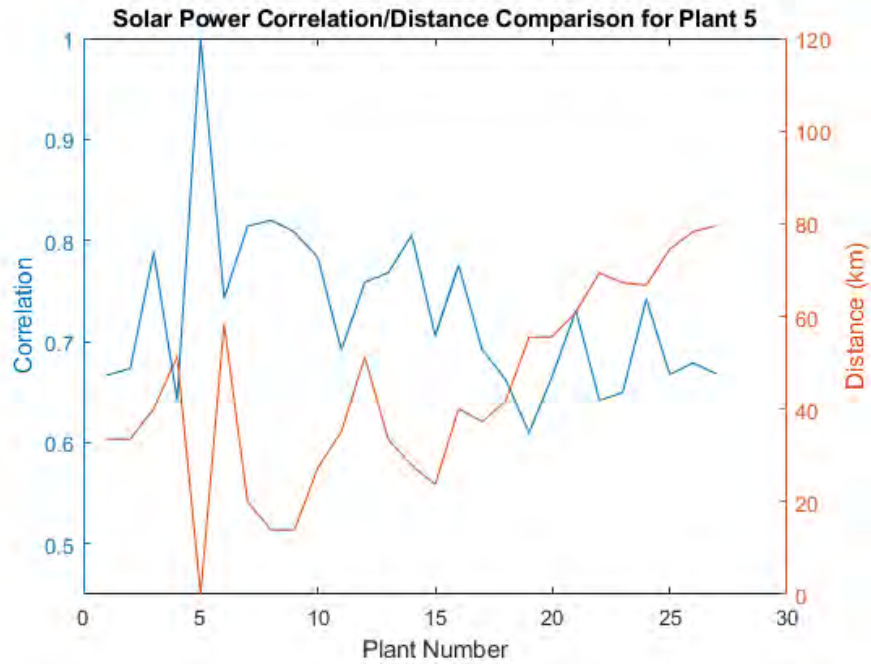


Figure 5: Distance vs Correlation for Power (Plant 5)

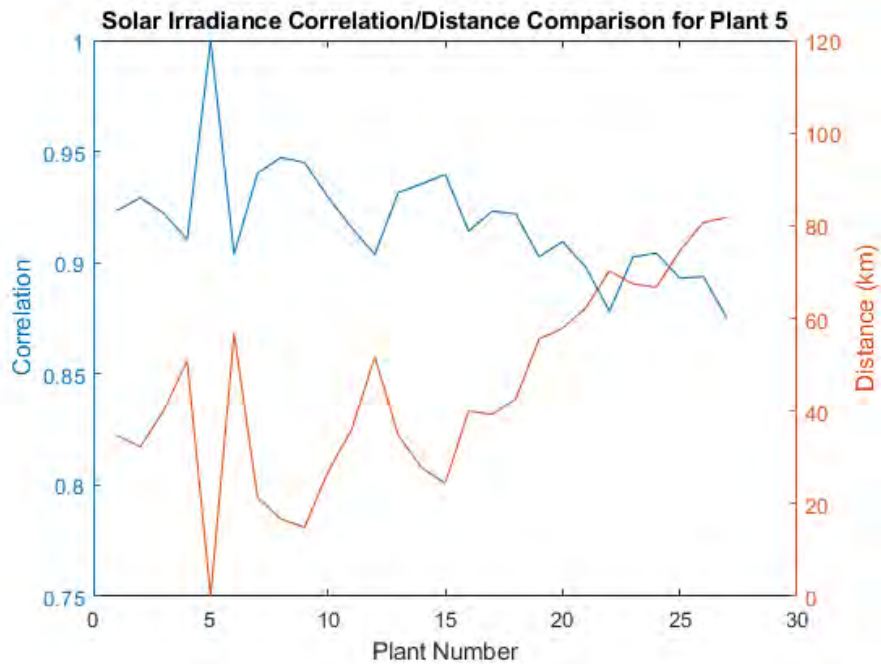


Figure 6: Distance vs Correlation for Irradiance (Plant 5)

As expected, the correlation between plants increases as the distance between plants decreases for the most part. Since this specific output is for location 5, the correlation goes to 1

for 5 and the distance goes to 0 for 5. This output shows that as the distance between plants gets smaller, the correlation between them increases and most likely the potential for concurrent transients also increases. Another thing to note is the difference between the power correlation and the irradiance correlation, which is better shown in figure 7.

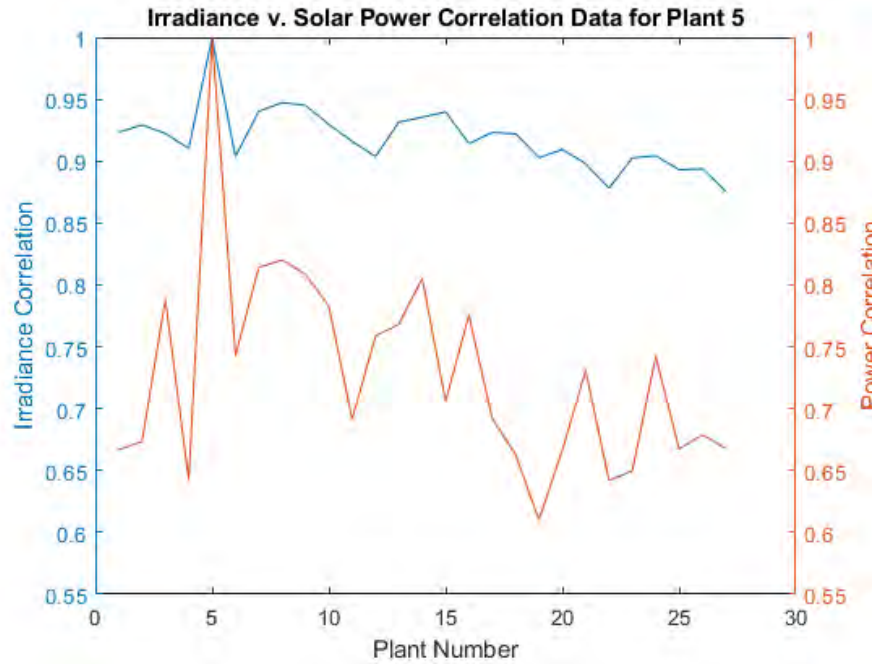


Figure 7: Irradiance Correlation vs Power Correlation (Plant 5)

The correlation for the irradiance data seems to change much less drastically from location to location, while the correlation for the power data has much more noticeable differences between locations. What this means is that the irradiance over the area where the 27 plants are located changes much less than the actual power produced by the plants themselves. This indicates that irradiance is a determining factor but other factors may also contribute to the solar generation such as temperature, tilt angles, the direction the arrays in the plant face, or other factors like shading from trees and other objects that affect some plants but not others. Regardless, the different correlation outputs seem to generally follow the same pattern. The similarity between the distance output for the power and irradiance data should also be noted, as better shown in figure 8.

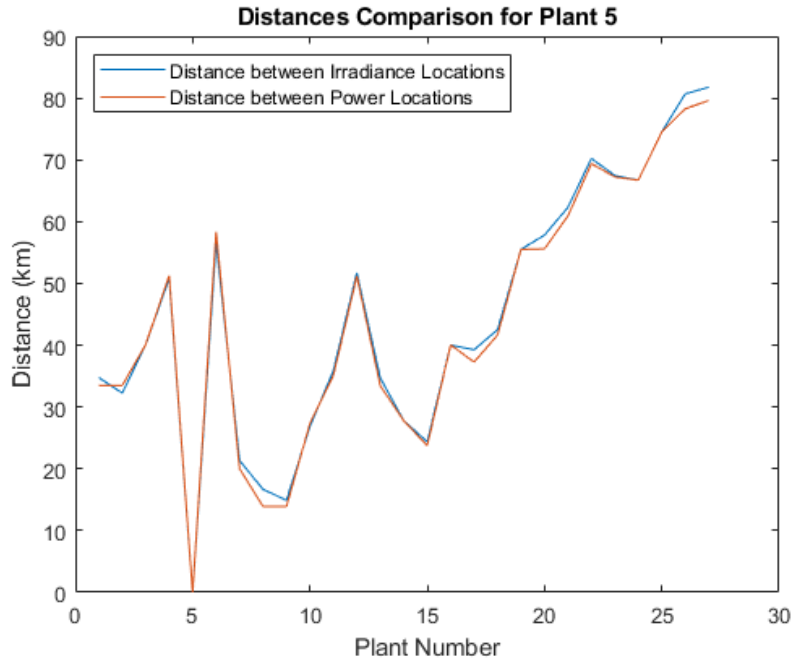


Figure 8: Distances Between Irradiance Locations vs Power Locations(Plant 5)

The slight difference in distance data is due to the fact that the irradiance data could not be obtained for the same exact locations as the power data as mentioned earlier. Once again the latitude and longitude coordinates may be off by as much as 1 or 2 hundredths of a degree, resulting in a maximum of 2.78 km difference between power and irradiance locations. Though, the difference is very small as shown in figure 8, and figure 9 shows just how close the two locations really are when separated by the maximum difference.

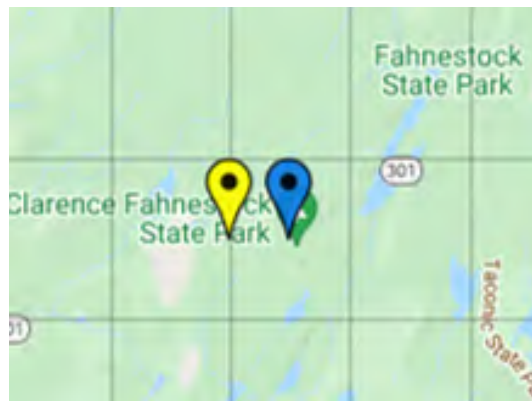


Figure 9: Difference Between Power Location and Irradiance Location



### III.II Double Transients Analysis Results

The results for the concurrent transients analysis is shown in the next 2 sections. It should be noted that the results for double and triple transients was found for a **deviation value of 35%**. The effect of a greater deviation value on the number of concurrent transients found is addressed later. First, the results for the double transients are shown below, starting with the number of transients per unique grouping compared with the distance between the 2 plants that make up each group.

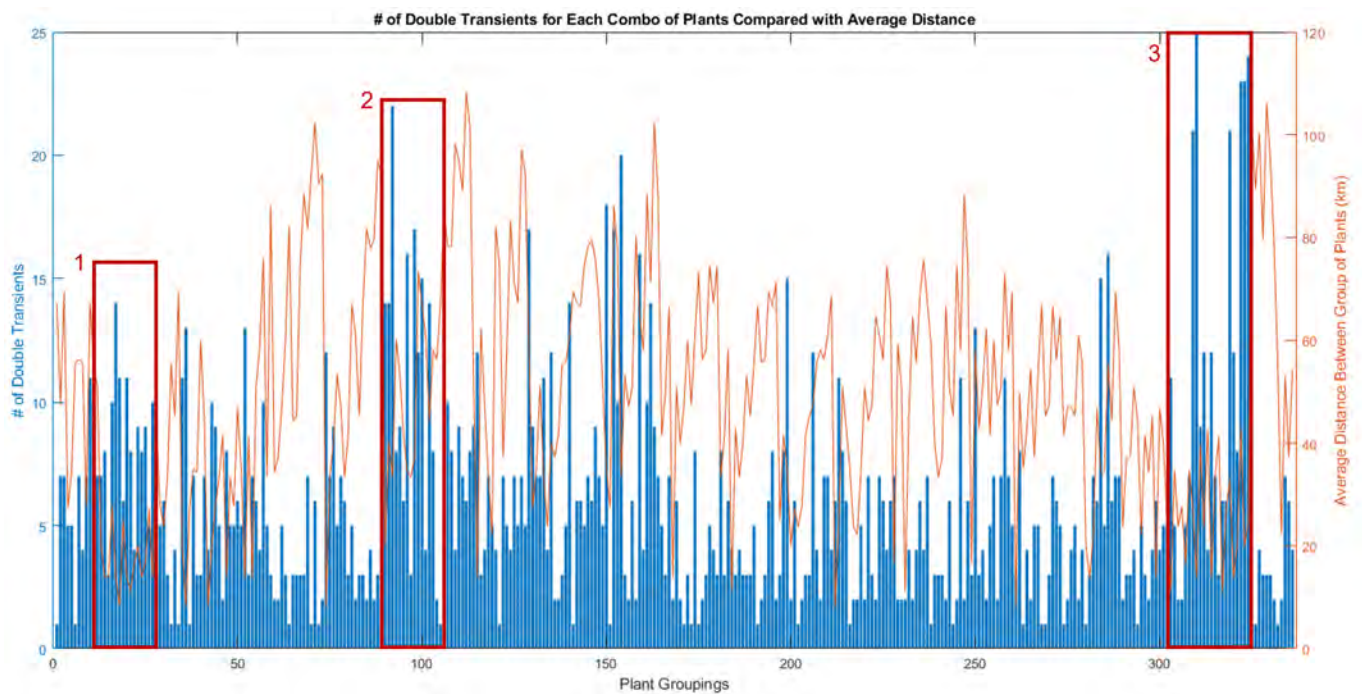


Figure 10: Double Transients Compared with Average Distance

In general, it can be seen in figure 10 that for portions where the location between plants is relatively small, the number of transients in those groups is large. This can best be seen in the areas marked with the red boxes. Of course there are some outliers but for the most part this data makes sense and shows that closely located plants have a better chance of resulting in double transients.

Figure 11 shows the same double transient count for plant groupings, but now compared with the correlation between the 2 plants rather than the distance between them.

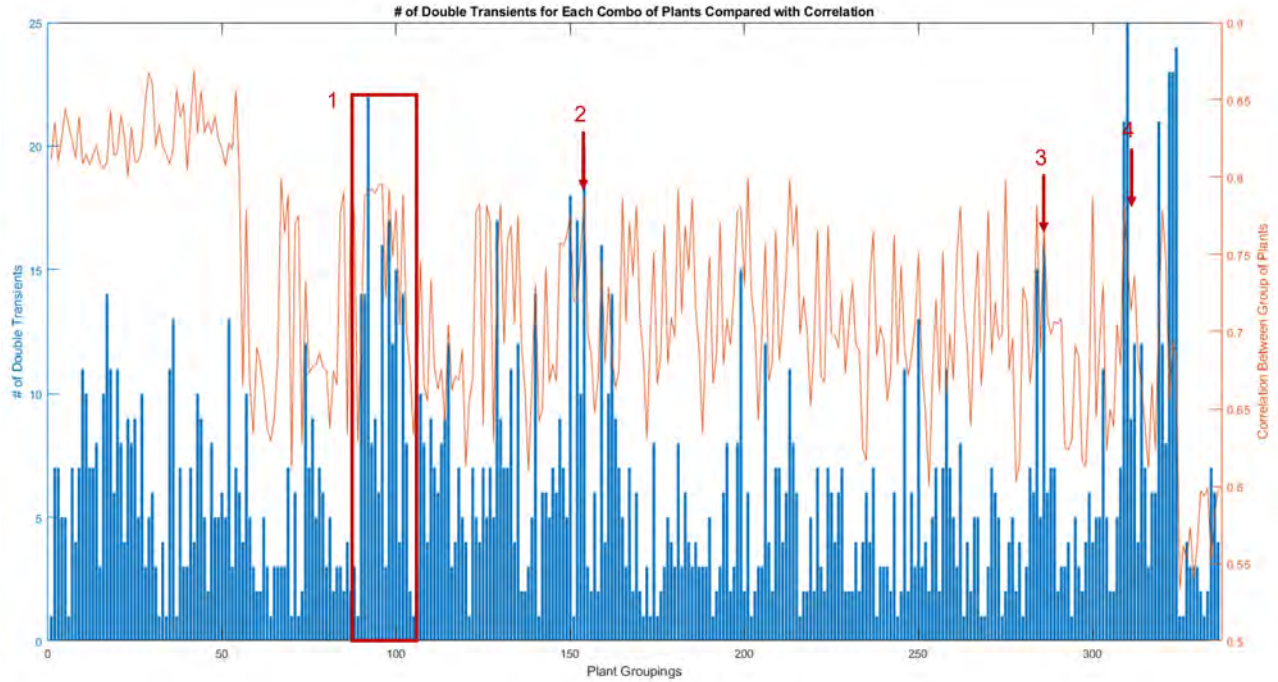


Figure 11: Double Transients Compared with Average Correlation

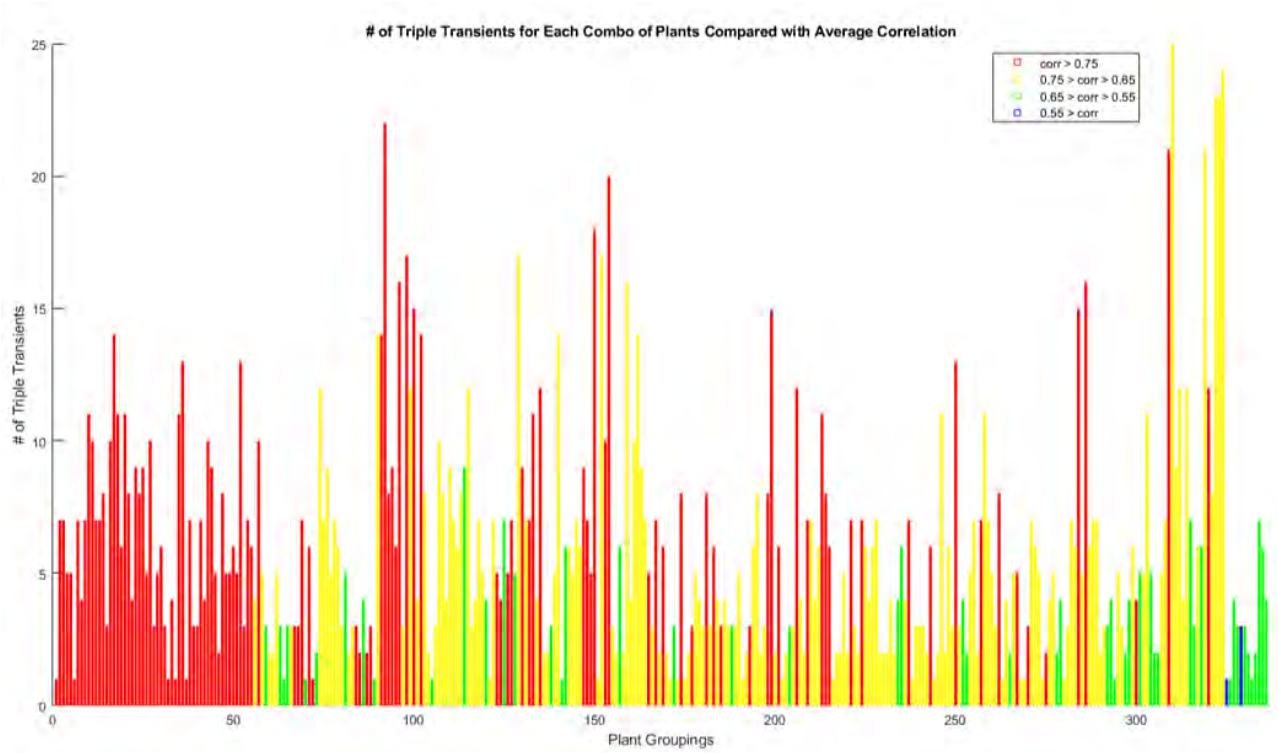


Figure 12: Number of Double Transients Color Coded by Correlation (0.75,0.65,0.55)

The first part of the plot in figure 11 shows an area of high correlation with most of the groupings in that section resulting in 5 or more double transients. Section 1 in this figure is the same as section 2 in figure 10, showing that there are indeed sections where large correlation and small distance both result in more concurrent transients. For the correlation, there are less large chunks or sections of groupings within a section of high correlation, and more small peaks that occur such as where arrows 2, 3 and 4 are pointing.

Figure 12 shows the number of double transients per grouping color coded based on the correlation range the group falls under. The ranges are ( $\text{corr} > 0.75$ ), ( $0.65 < \text{corr} < 0.75$ ), ( $0.55 < \text{corr} < 0.65$ ) and ( $\text{corr} < 0.55$ ). Table 2 shows how many double transients were found for each correlation range, as well as the average number of double transients per group. As expected, the average increases as correlation increases.

Table 2: Number/Average of Double Transients per Correlation Range (0.75,0.65,0.55)

Correlation Range	# of Double Transients	Average # of Double Transients per Group
$\text{corr} > 0.75$	896	7.467
$0.65 < \text{corr} < 0.75$	903	5.5061
$0.55 < \text{corr} < 0.65$	175	3.5
$\text{corr} < 0.55$	4	2.0

Figure 13 and table 3 show similar results but for different correlation ranges. These correlation ranges are ( $\text{corr} > 0.80$ ), ( $0.60 < \text{corr} < 0.80$ ) and ( $\text{corr} < 0.60$ ). There is one less correlation range than the previous results, and the values are shifted up by 0.05, so less double transients are in the upper range.

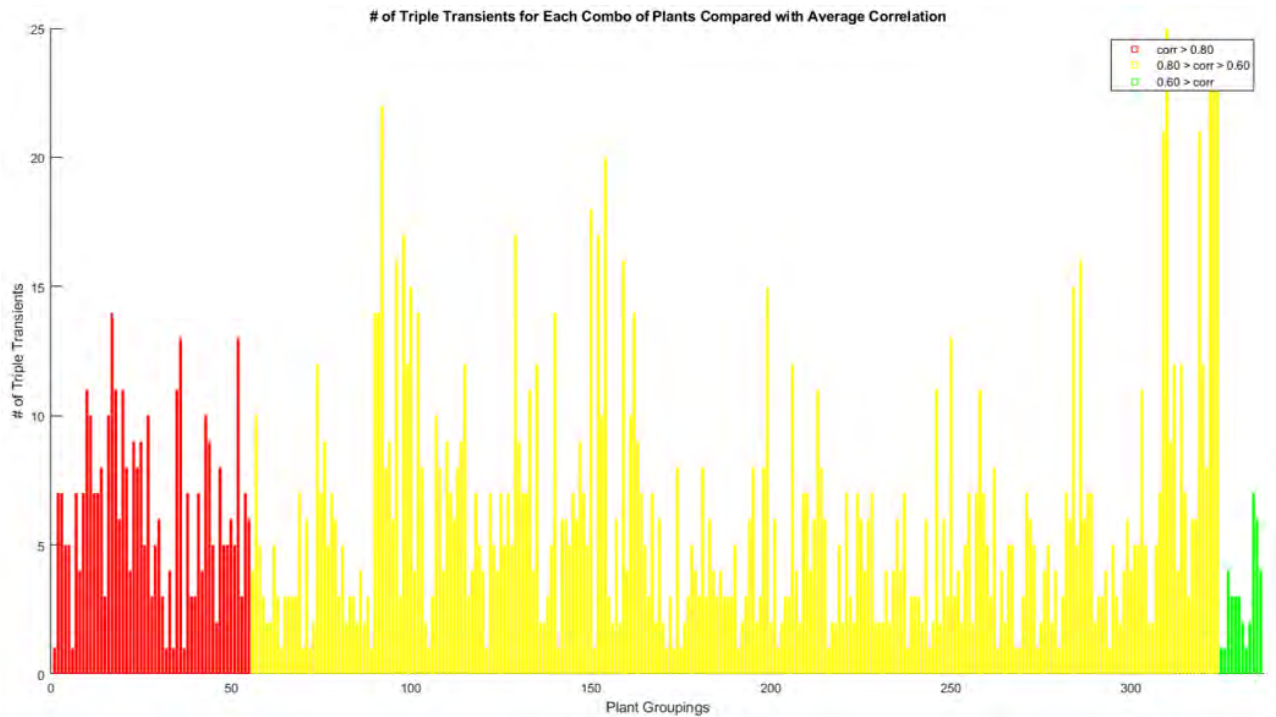


Figure 13: Number of Double Transients Color Coded by Correlation (1.0,0.8,0.6)

Table 3: Number/Average of Double Transients per Correlation Range (1.0,0.8,0.6)

Correlation Range	# of Double Transients	Average # of Double Transients per Group
corr > 0.80	351	6.3818
0.60 < corr < 0.80	1590	5.9108
corr < 0.60	37	3.0833

In this case, less of the groupings in the middle fall under the highest correlation range, and there are many more double transients present in the middle range than previously. However, the average number of double transients per group still increases as the correlation increases as shown in table 3.

A comparison between the distance and correlation can be seen in figure 14, where they generally behave opposite to one another as expected, showing that the correlation increases as distance decreases and vice versa for the most part. In general, these results show that co-located solar generation sites will result in more double transients.

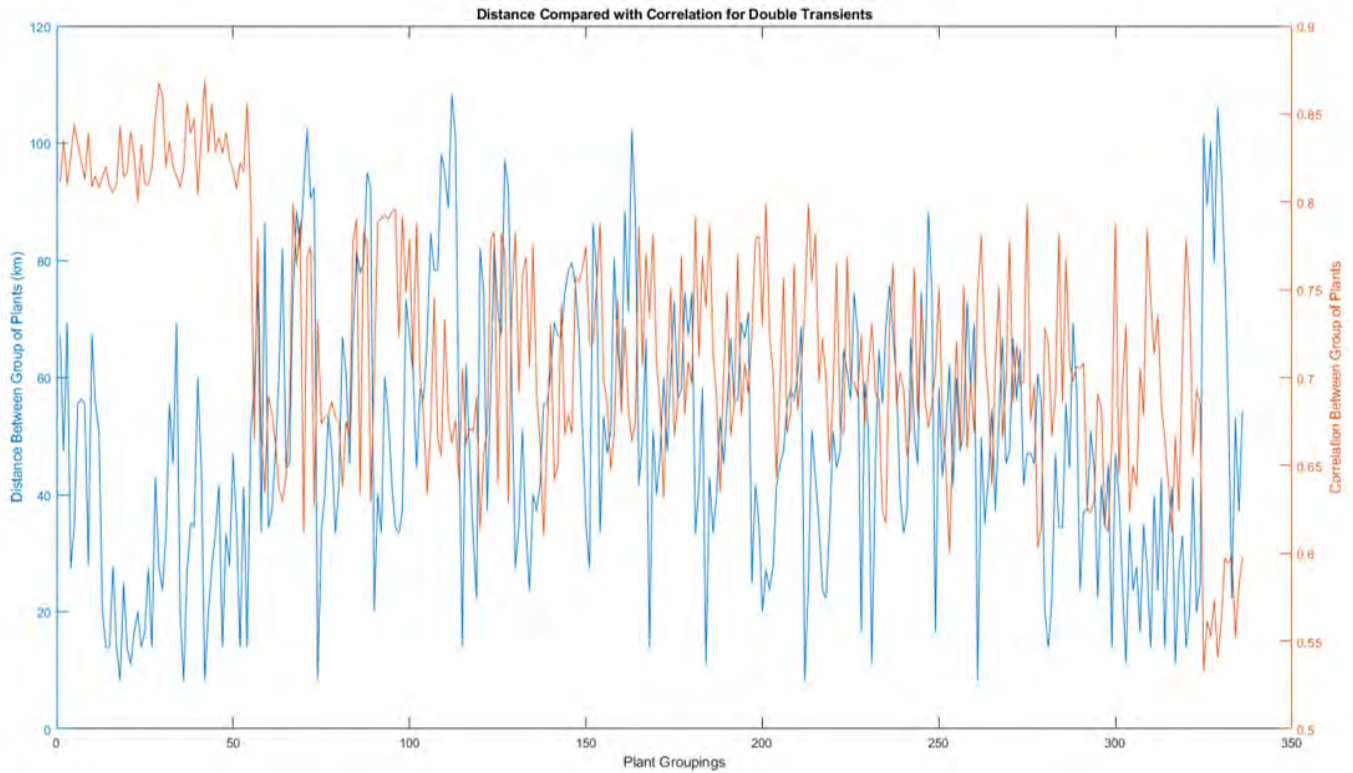


Figure 14: Double Transients Distance vs. Correlation

### III.III Triple Transients Analysis Results

This section shows the results for the triple transient analysis and how it differs from the double transient analysis. Instead of the number of transients per grouping being plotted against the location between the plants, now it is plotted against the **average** location between the plants. This is because every grouping is now made up of 3 plants, so there is the distance between plants 1 and 2, between plants 2 and 3, and between plants 1 and 3. These 3 distances are averaged for each grouping resulting in the plot in figure 15.



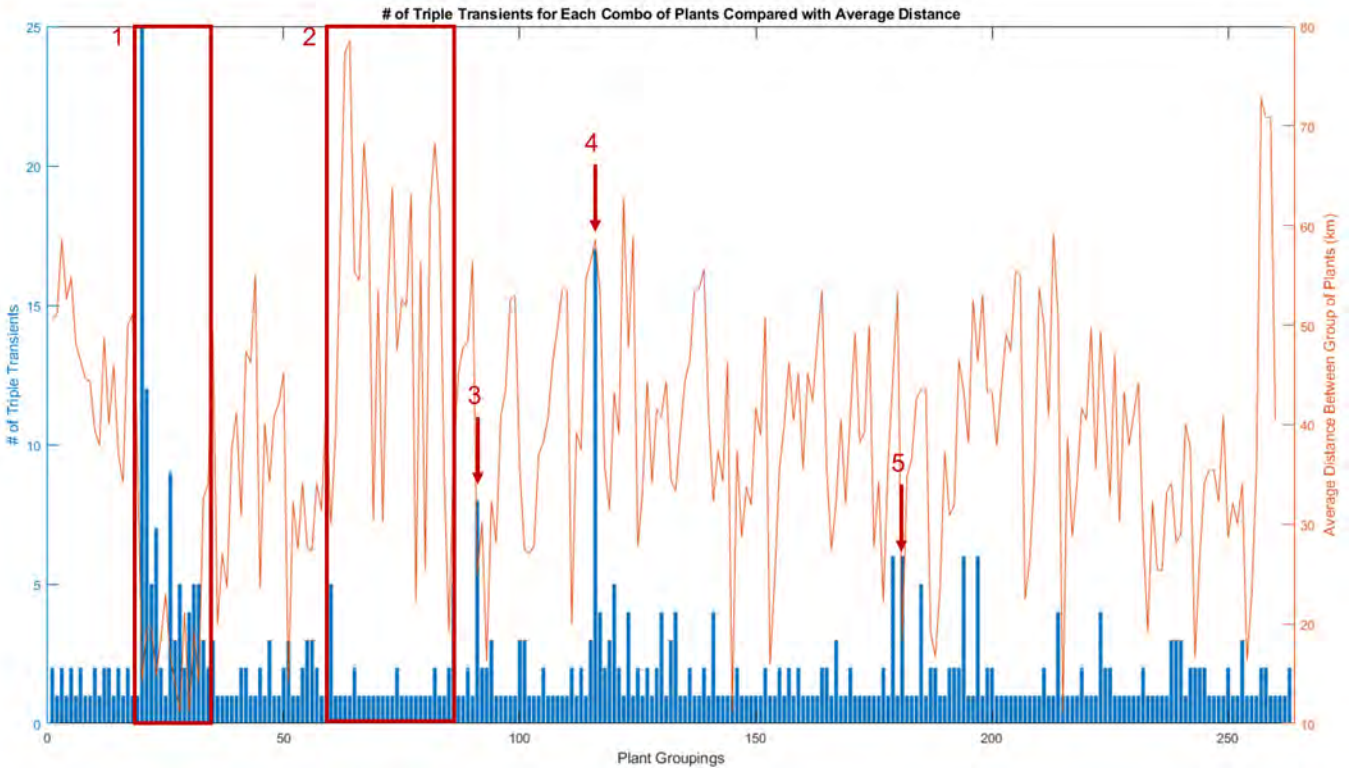


Figure 15: Triple Transients Compared with Average Distance

There are less triple transients than there were double transients, as it is less likely that 3 plants will experience a concurrent change in generation than 2 plants. For this project, 511 triple transients were found, as well as 263 unique groupings.

The highlighted section 1 shows an area where the greatest concentration of triple transients are, as well as where the distance between plants is the lowest. So once again, this shows that less distance between plants seems to result in more concurrent transients. The highlighted section 2 shows an area where the distance between plants is particularly large in general, and as a result there are not many triple transients present in those groupings. Arrows 3 and 5 shows groupings that have a fair amount of triple transients due to a drop in distance between plants, and arrow 4 shows an outlier that seems to do the opposite.

Next is the comparison between the number of transients per grouping and the **average** correlation between the 3 plants that make up each grouping.

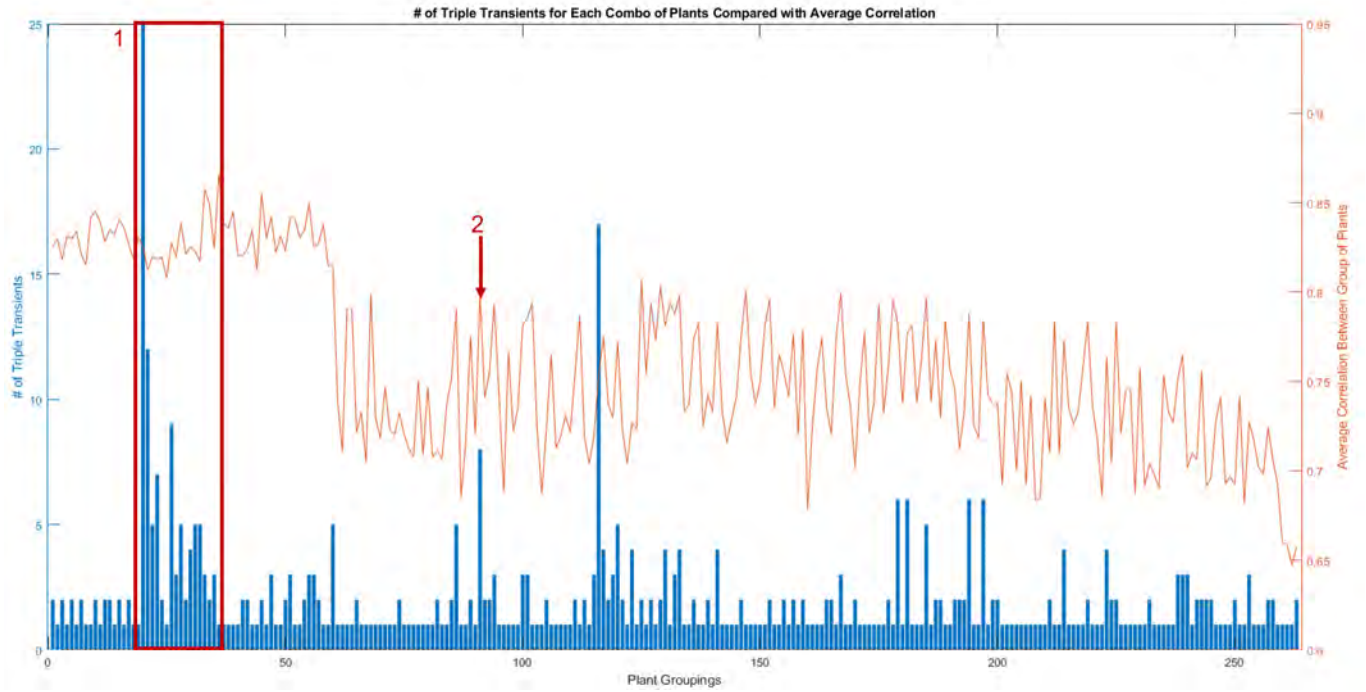


Figure 16: Triple Transients Compared with Average Correlation

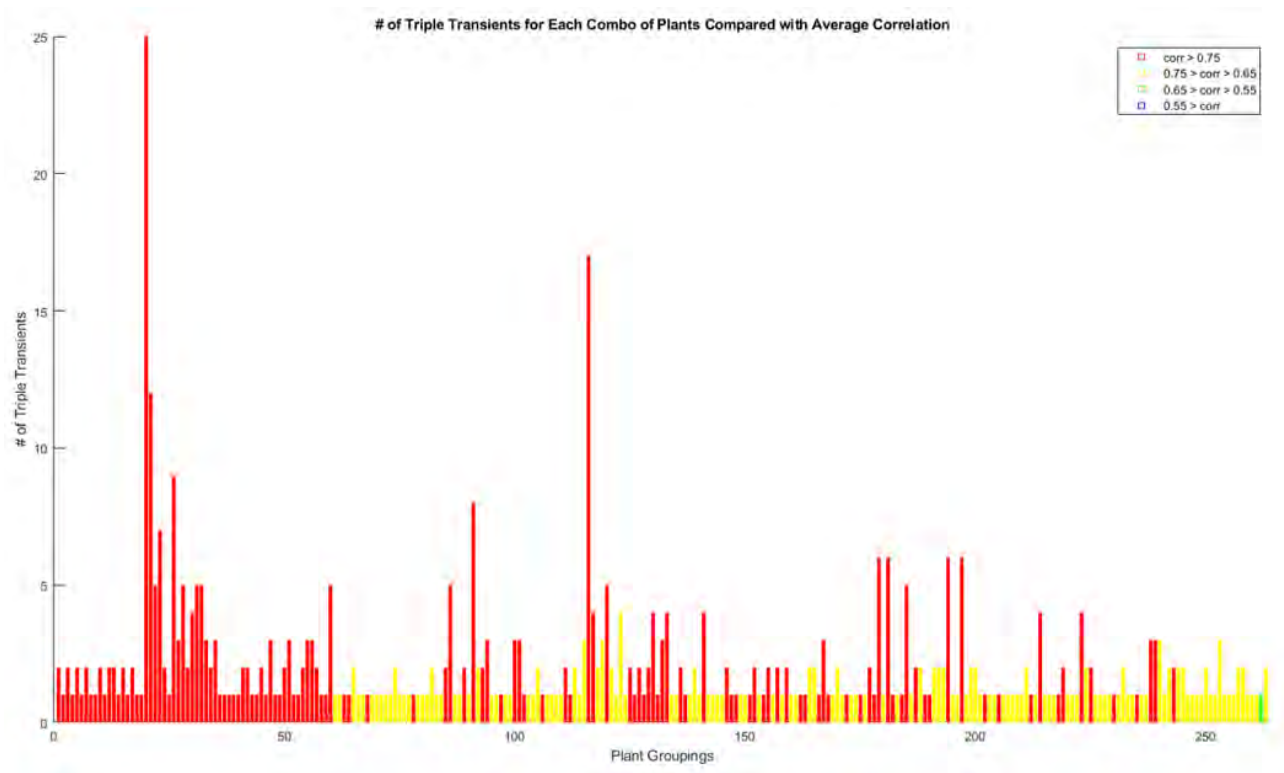


Figure 17: Number of Triple Transients Color Coded by Correlation

Similarly to the distance, the correlation between the 3 plants for each grouping was averaged to produce a single average correlation for each. As seen in figure 16, the highlighted group 1 from before falls under the section where the average correlation is highest. Besides this section, there are only small peaks where correlation is high and the number of transients rises slightly as well, such as where arrow 2 is pointing, and the outlier from before still remains.

Once again, a color coded plot is shown in figure 17 for the different correlation ranges the groupings fall under. Table 4 below shows the specifics for this data, as the average number of triple transients increases with increasing correlation, just as the double transients did.

Table 4: Number/Average of Triple Transients per Correlation Range

Correlation Range	# of Triple Transients	Average # of Triple Transients per Group
$\text{corr} > 0.75$	342	2.5714
$0.65 < \text{corr} < 0.75$	168	1.3023
$0.55 < \text{corr} < 0.65$	1	1.0
$\text{corr} < 0.55$	0	N/A

This time there less concurrent transients in the two lower correlation ranges, as there is only 1 triple transient in the 0.55 to 0.65 range, and 0 in the less than 0.55 range. The concentration of triple transients is much higher in the highest correlation range, whereas for double transients there were slightly more in the second highest range. This shows that triple transients grouping fall under a higher correlation range in general.

Figure 18 and table 5 show the triple transient data for the second group of correlation ranges as was shown for the double transients.



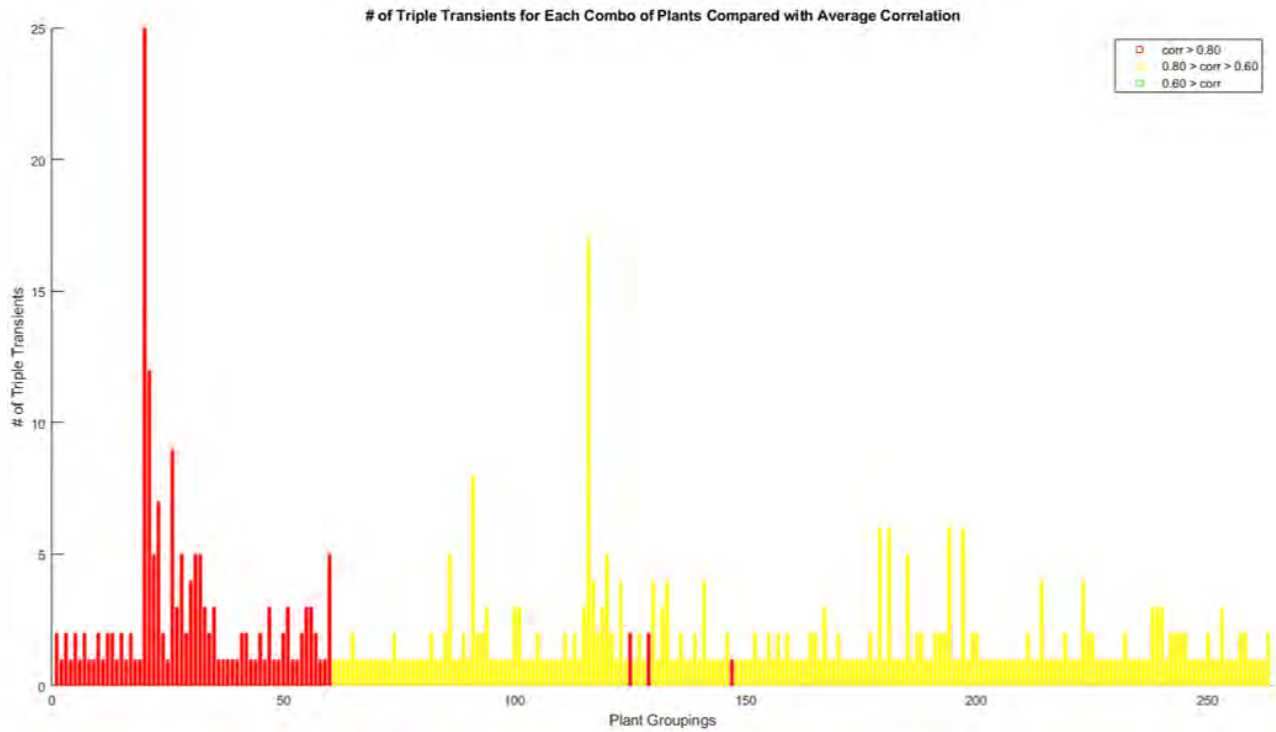


Figure 18: Number of Triple Transients Color Coded by Correlation (1.0, 0.8, 0.6)

Table 5: Number/Average of Triple Transients per Correlation Range (1.0, 0.8, 0.6)

Correlation Range	# of Triple Transients	Average # of Triple Transients per Group
corr > 0.80	169	2.6825
0.60 < corr < 0.80	342	1.71
corr < 0.60	0	N/A

Similarly to the double transient results, the concentration of triple transients has shifted away from the highest correlation range into the middle correlation range since the ranges values were increased by 0.05. Though, the average number of triple transients per group still increases as the correlation increases, showing higher correlation results in more concurrent transients on average per group.

In general, the correlation and distance data seems to act as it has previously, and show that the number of transients increases as correlation increases and distance decreases. The comparison between the distance and correlation data is shown for triple transients in figure 17 where this trend can be seen once again.

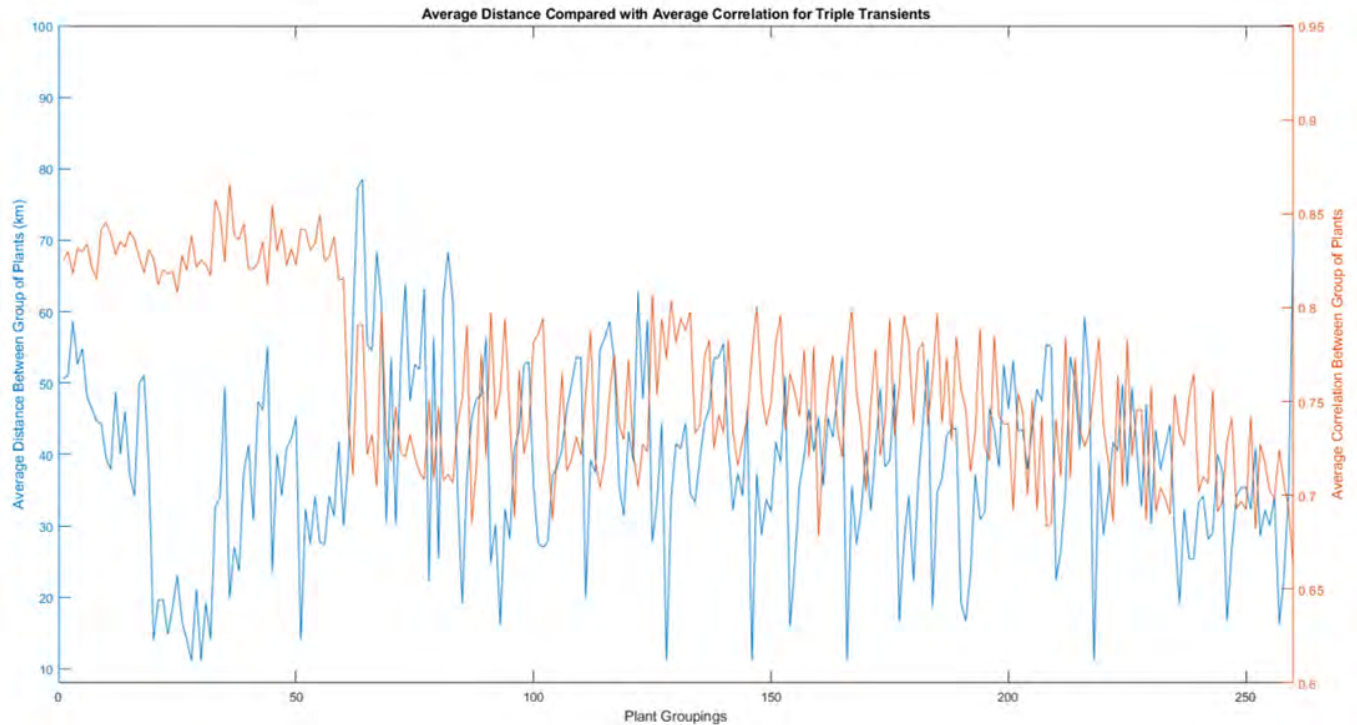


Figure 19: Triple Transients Average Distance vs. Average Correlation

Table 4 displays the number of transients for each case of single, double and triple transients, as well as the number of unique groupings. In total, there were 2,896 transients found for the 27 plants over the entire year of 2006. That number was reduced down to 1,978 for double transients, and even lower to 511 for triple transients. The drop from 336 unique grouping to 263 unique grouping for double and triple transients respectively is much less drastic than from single to double, showing that many of the same plant groupings that resulted double transients resulted in triple transients as well.

Table 6: Amount of Double/Triple Transients and Unique Groupings

Transient Type	# of Transients	# of Unique Groupings
Single Transients	2896	2896
Double Transients	1978	336
Triple Transients	511	263

The deviation value used to extract the transients for the data shown above is 35%, but increasing this value has an effect on the number of transients extracted as the requirements for being a transient become more strict. Table 5 below shows the effect increasing the deviation value has on the number of transients extracted for the single, double and triple cases.

Table 7: Number of Transients Found for Different Deviation Values

Deviation	# of Single Transients	# of Double Transients	# of Triple Transients
35%	2896	1978	511
40%	2300	1193	206
45%	1693	642	83
50%	1152	319	28

As expected, the number of transients extracted decreases as the deviation increases. Once the deviation value is set to 50%, only 28 triple transients are found, and 319 double transients are found, compared to 511 and 1978 respectively for 35% deviation.

Tables 6 and 7 in the Appendix show the specific plant numbers involved in each grouping for double and triple transients respectively. The grouping numbers correspond to the numbers at the bottom of the plots shown in the double and triple transients results above.

## IV Conclusion

The goals for this project were to obtain time-series solar power and irradiance data, clean and sanitize that data, analyze the correlation and distance between different solar power and irradiance locations, and to analyze the relationship between the occurrence of concurrent transients and the correlation/distance for multiple solar generation sites. Using the two MATLAB scripts mentioned previously, those goals were accomplished with the results displayed.

It has been shown that concurrent transients occur between co-located solar generation sites using time-series solar generation and irradiance data. In general, as distance and correlation decrease, the number of triple and double transients increases as well. More importantly, as correlation increases, the average number of concurrent transients per grouping of plants increases. This may be interpreted as plants with higher correlation will experience more concurrent transients on average per year.

From the relationship between location/correlation and the number of concurrent transients, it can be inferred that sites located closer to one another than the sites used in this project would produce an even greater correlation and number of concurrent transients, and thus pose a larger risk. Each of the 27 locations used are no greater than 8 km from the next closest location, so an area with solar generation sites even closer to each other certainly exists and should also be studied.

## V Acknowledgements

- **Science Undergraduate Laboratory Internships (SULI)**

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

- Thank you to my mentor Dr. Meng Yue, those in the “Development of a Planning, Operation, and Control Framework for Hybrid Energy Storage and Renewable Generation Systems” research group and others I have met during my time in this program. This was a great learning experience and I am thankful for my time here.

## References

<sup>1</sup> Solar Market Insight Report 2021 Q2 (SEIA, Wood Mackenzie Power & Renewables, 2021).

<sup>2</sup> Solar power data for integration studies. <https://www.nrel.gov/grid/solar-power-data.html>,(Accessed 8 June 2021).

<sup>3</sup> National solar radiation database. <https://nsrdb.nrel.gov/>,(Accessed 8 June 2021).

<sup>4</sup> Solar Power Data Map. <https://maps.co/map/60f713787454a058188831jqtf36d7e>,(Accessed 29 July 2021).

## VI Appendix

Table 8: Double Transient Grouping with Corresponding Plant Numbers

Grouping Number	Plant1	Plant2
1	1	11
2	1	15
3	1	17
4	2	4
5	2	11
6	2	15
7	2	17
8	2	18
9	4	11
10	4	15
11	4	17
12	4	18
13	5	7
14	5	8
15	5	9
16	5	14
17	6	12
18	7	8
19	7	9
20	7	13
21	7	14
22	8	9
23	8	13
24	8	14
25	9	10
26	9	14
27	10	16
28	11	15
29	11	17
30	11	18
31	11	19
32	11	20
33	11	22
34	11	23
35	12	16

Grouping Number	Plant1	Plant2
36	13	14
37	15	17
38	15	18
39	15	20
40	15	22
41	15	23
42	17	18
43	17	19
44	17	20
45	17	22
46	17	23
47	18	19
48	18	20
49	18	22
50	18	23
51	19	20
52	19	22
53	20	22
54	20	23
55	22	23
56	1	3
57	1	4
58	1	5
59	1	6
60	1	7
61	1	8
62	1	9
63	1	10
64	1	12
65	1	13
66	1	14
67	1	18
68	1	19
69	1	20
70	1	21
71	1	22
72	1	23
73	1	24
74	2	3
75	2	5



Grouping Number	Plant1	Plant2
76	2	6
77	2	7
78	2	8
79	2	10
80	2	12
81	2	13
82	2	14
83	2	16
84	2	19
85	2	20
86	2	21
87	2	22
88	2	23
89	2	24
90	3	4
91	3	5
92	3	6
93	3	7
94	3	8
95	3	9
96	3	10
97	3	11
98	3	12
99	3	13
100	3	14
101	3	15
102	3	16
103	3	17
104	3	18
105	3	19
106	3	20
107	3	21
108	3	22
109	3	23
110	3	24
111	3	25
112	3	26
113	3	27
114	4	5
115	4	6

Grouping Number	Plant1	Plant2
116	4	8
117	4	9
118	4	10
119	4	12
120	4	13
121	4	14
122	4	16
123	4	19
124	4	20
125	4	21
126	4	22
127	4	23
128	4	24
129	5	6
130	5	10
131	5	11
132	5	12
133	5	13
134	5	15
135	5	16
136	5	17
137	5	18
138	5	19
139	5	20
140	5	21
141	5	22
142	5	23
143	5	24
144	5	25
145	5	26
146	5	27
147	6	7
148	6	8
149	6	9
150	6	10
151	6	11
152	6	13
153	6	14
154	6	16
155	6	17

Grouping Number	Plant1	Plant2
156	6	18
157	6	19
158	6	20
159	6	21
160	6	22
161	6	24
162	6	25
163	6	26
164	6	27
165	7	10
166	7	11
167	7	12
168	7	15
169	7	16
170	7	17
171	7	18
172	7	19
173	7	20
174	7	21
175	7	22
176	7	23
177	7	24
178	7	25
179	7	26
180	7	27
181	8	10
182	8	11
183	8	12
184	8	15
185	8	16
186	8	17
187	8	18
188	8	19
189	8	20
190	8	21
191	8	22
192	8	23
193	8	24
194	8	25
195	8	26

Grouping Number	Plant1	Plant2
196	8	27
197	9	11
198	9	12
199	9	13
200	9	15
201	9	16
202	9	17
203	9	18
204	9	19
205	9	20
206	9	21
207	9	22
208	9	23
209	9	24
210	9	25
211	9	26
212	10	11
213	10	12
214	10	13
215	10	14
216	10	15
217	10	17
218	10	18
219	10	19
220	10	20
221	10	21
222	10	22
223	10	23
224	10	24
225	10	25
226	10	26
227	10	27
228	11	12
229	11	13
230	11	14
231	11	16
232	11	21
233	11	24
234	11	25
235	11	27

Grouping Number	Plant1	Plant2
236	12	13
237	12	14
238	12	15
239	12	17
240	12	18
241	12	19
242	12	20
243	12	21
244	12	22
245	12	24
246	12	25
247	12	26
248	12	27
249	13	15
250	13	16
251	13	17
252	13	18
253	13	19
254	13	20
255	13	21
256	13	23
257	13	24
258	13	25
259	13	26
260	13	27
261	14	15
262	14	16
263	14	17
264	14	18
265	14	19
266	14	20
267	14	21
268	14	22
269	14	23
270	14	24
271	14	25
272	14	26
273	14	27
274	15	16
275	15	19

Grouping Number	Plant1	Plant2
276	15	21
277	15	24
278	15	25
279	15	26
280	16	17
281	16	18
282	16	19
283	16	20
284	16	21
285	16	22
286	16	24
287	16	25
288	16	26
289	16	27
290	17	21
291	17	24
292	17	25
293	17	26
294	17	27
295	18	21
296	18	24
297	18	25
298	18	27
299	19	21
300	19	23
301	19	24
302	20	21
303	20	24
304	20	25
305	20	26
306	20	27
307	21	22
308	21	23
309	21	24
310	21	25
311	21	26
312	21	27
313	22	24
314	22	25
315	22	27

Grouping Number	Plant1	Plant2
316	23	25
317	23	26
318	23	27
319	24	25
320	24	26
321	24	27
322	25	26
323	25	27
324	26	27
325	1	26
326	2	25
327	2	27
328	4	25
329	4	26
330	4	27
331	11	26
332	18	26
333	19	25
334	19	26
335	19	27
336	22	26

Table 9: Triple Transient Grouping with Corresponding Plant Numbers

Grouping Number	Plant1	Plant2	Plant3
1	1	2	11
2	1	2	15
3	1	2	17
4	1	11	15
5	1	11	17
6	1	15	17
7	2	4	17
8	2	4	18
9	2	11	15
10	2	11	17
11	2	11	18
12	2	15	18
13	2	17	18
14	4	11	15
15	4	11	17
16	4	11	18
17	4	15	17
18	4	15	18
19	4	17	18
20	5	7	8
21	5	7	9
22	5	7	14
23	5	8	9
24	5	8	14
25	5	9	14
26	7	8	9
27	7	8	13
28	7	8	14
29	7	9	14
30	7	13	14
31	8	9	14
32	8	13	14
33	11	15	17
34	11	15	18
35	11	15	22
36	11	17	18
37	11	17	19



Grouping Number	Plant1	Plant2	Plant3
38	11	18	19
39	11	18	20
40	11	19	20
41	11	19	22
42	11	20	22
43	11	20	23
44	11	22	23
45	15	17	18
46	15	17	22
47	15	18	20
48	15	18	22
49	15	18	23
50	15	20	22
51	17	18	19
52	17	18	23
53	17	19	20
54	17	20	22
55	17	20	23
56	18	19	20
57	18	20	22
58	18	20	23
59	18	22	23
60	19	20	22
61	1	2	3
62	1	2	6
63	1	2	22
64	1	2	23
65	1	11	12
66	1	11	14
67	1	11	21
68	1	15	19
69	2	4	12
70	2	11	13
71	2	11	16
72	2	11	21
73	2	11	24
74	2	15	16
75	2	17	21
76	2	18	21
77	2	18	24

Grouping Number	Plant1	Plant2	Plant3
78	4	11	12
79	4	11	13
80	4	11	16
81	4	15	21
82	4	15	24
83	4	18	24
84	5	7	11
85	5	7	15
86	5	7	16
87	5	7	19
88	5	7	23
89	5	7	24
90	5	7	25
91	5	8	10
92	5	8	11
93	5	8	15
94	5	8	16
95	5	8	17
96	5	8	19
97	5	8	21
98	5	8	25
99	5	8	26
100	5	9	12
101	5	9	13
102	5	9	16
103	5	9	18
104	5	9	19
105	5	9	20
106	5	9	21
107	5	9	23
108	5	9	25
109	5	9	26
110	5	9	27
111	5	14	15
112	5	14	16
113	5	14	18
114	5	14	22
115	5	14	25
116	6	12	13
117	6	12	14

Grouping Number	Plant1	Plant2	Plant3
118	6	12	17
119	6	12	18
120	6	12	21
121	6	12	22
122	6	12	23
123	6	12	25
124	6	12	27
125	7	8	10
126	7	8	11
127	7	8	12
128	7	8	15
129	7	8	16
130	7	8	21
131	7	8	24
132	7	9	12
133	7	9	16
134	7	9	18
135	7	9	20
136	7	9	21
137	7	9	24
138	7	9	25
139	7	9	26
140	7	9	27
141	7	13	16
142	7	13	17
143	7	13	18
144	7	13	20
145	7	13	26
146	7	14	15
147	7	14	16
148	7	14	17
149	7	14	18
150	7	14	20
151	7	14	21
152	7	14	24
153	7	14	25
154	8	9	15
155	8	9	17
156	8	9	20
157	8	9	21

Grouping Number	Plant1	Plant2	Plant3
158	8	9	22
159	8	13	16
160	8	13	19
161	8	13	20
162	8	13	21
163	8	13	24
164	8	13	26
165	8	13	27
166	8	14	15
167	8	14	16
168	8	14	17
169	8	14	18
170	8	14	19
171	8	14	20
172	8	14	21
173	8	14	22
174	8	14	23
175	8	14	24
176	8	14	25
177	9	10	11
178	9	10	12
179	9	10	13
180	9	10	18
181	9	10	21
182	9	10	24
183	9	10	26
184	9	14	15
185	9	14	16
186	9	14	20
187	9	14	21
188	9	14	23
189	9	14	24
190	10	16	17
191	10	16	18
192	10	16	19
193	10	16	20
194	10	16	21
195	10	16	22
196	10	16	23
197	10	16	24

Grouping Number	Plant1	Plant2	Plant3
198	10	16	25
199	10	16	26
200	10	16	27
201	11	15	25
202	11	17	24
203	11	18	24
204	11	18	25
205	11	20	24
206	11	22	27
207	11	23	24
208	11	23	25
209	11	23	27
210	12	16	18
211	12	16	19
212	12	16	21
213	12	16	23
214	12	16	24
215	12	16	25
216	12	16	26
217	12	16	27
218	13	14	15
219	13	14	16
220	13	14	17
221	13	14	18
222	13	14	19
223	13	14	21
224	13	14	22
225	13	14	24
226	13	14	25
227	13	14	26
228	15	17	21
229	15	18	27
230	15	20	24
231	15	20	25
232	15	20	26
233	15	20	27
234	15	23	27
235	17	18	24
236	17	19	21
237	17	19	24

Grouping Number	Plant1	Plant2	Plant3
238	17	20	21
239	17	20	24
240	17	20	25
241	17	20	26
242	17	22	25
243	17	23	24
244	17	23	25
245	17	23	27
246	18	19	21
247	18	20	21
248	18	20	25
249	18	20	27
250	18	22	27
251	18	23	24
252	18	23	25
253	19	20	24
254	19	22	24
255	20	22	25
256	20	22	27
257	20	23	26
258	20	23	27
259	22	23	25
260	2	11	26
261	4	11	26
262	4	18	26
263	19	22	26

---

# THE APPLICATION OF CURRENT PHYSICS RESEARCH IN AN EDUCATIONAL SETTING

---

August 11, 2021

Nkeiru Ubadike

Department of Physics, Stony Brook University, Stony Brook, NY 11790

Michele Darienzo

Office of Educational Programs, Brookhaven National Lab, Upton, NY 11973

## 0.1 ABSTRACT

The Office of Educational Programs (OEP) at Brookhaven National Lab hosts various summer workshops for students in the K-12 educational level. OEP twice virtually hosted new week-long Modern Physics Workshops which introduced high school students to areas in physics of particular relevance to Brookhaven researchers today. These topics were particle physics, medical isotopes and quantum computing. In particular, I prepared a two day segment on quantum computing. I motivated the lesson by discussing the classical computer and leading an activity that demonstrates the difference between classical and quantum data. Then, I explained various quantum phenomena meanwhile introducing the basics of Dirac notation. Finally, I showed students how to implement circuits that demonstrate a number of quantum effects such as superposition and interference on the IBM Quantum Composer. Quantum computing is a nascent field at the intersection of physics, math and computer science and this workshop serves to expose the next generation of scientists to this emergent field and to inspire confidence in America's upcoming STEM workforce to pursue quantum computing studies further. As a result of my activities, I have significantly developed my capability to expound complex scientific topics to a general, non-expert audience. Additionally, I have become familiar with IBM Quantum Composer.

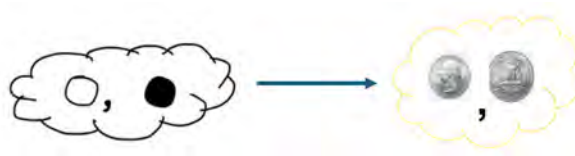
## 0.2 INTRODUCTION

This report details the quantum computing section of the Office of Educational Program's Summer Explorations workshop entitled "Modern Physics featuring LEGO" delivered virtually on July 26th-30th 2021 and on August 2nd-6th. I aimed to create and deliver an engaging learning experience, incorporating hands on activities, analogies, mini-quizzes and video resources along with lecture and mathematics. By presenting content through these various media, I hoped to mitigate some of the challenges brought by teaching in a virtual environment and make the topic of quantum computing more accessible.



### 0.3 WORKSHOP PREPARATION

To prepare lesson material, I reviewed a variety of sources to self-educate on the topic of quantum computing, in particular, quantum information science. The crucial resources that contributed to my workshop plan are presented here along with each resource's most relevant contribution to the lesson plan. The resource that contributed the most to my lesson was a book entitled "Q is for Quantum" by Terry Rudolph.[5] To explain the qubit, the carrier of quantum information, it made use of an analogy involving balls either black or white which represented the states  $|1\rangle$  and  $|0\rangle$  respectively and the superposition state represent by a cloud containing both states. I adapted this analogy by using coins instead, in the workshop to give students a visual aid.



**Figure 1:** Analogy adapted from "Q is for Quantum [5]

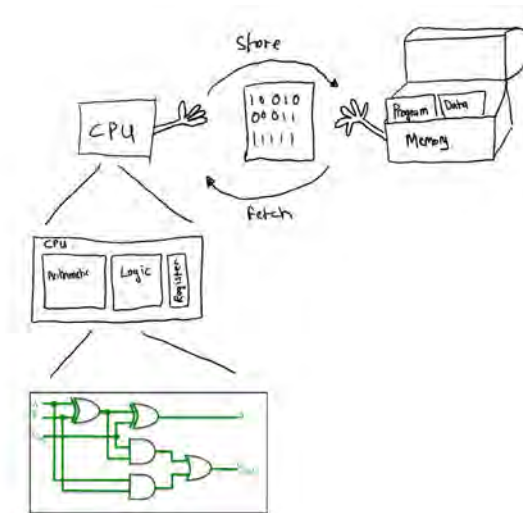
I obtained a more rigorous understanding to quantum computing from an eight part video and lecture course by Dr. Chris Ferrie [2]. This resource provided a thorough treatment of Dirac notation and calculations. I introduced Dirac notation along with Rudolph's ball analogy. To make quantum circuits, I used IBM Quantum Composer[4] which allows the general public to run circuits on a real quantum hardware. Each quantum effect presented was accompanied with a hands on demonstration using this resource. Steve Specklemire's video on Deutsch's problem [6] was of tremendous utility in preparing workshop content on the aforementioned problem. Each of these resources, in addition to others which were not directly included in workshop content were studied daily for a number of weeks. Additionally, I prepared a Tiktok video [1] using ice cream as a pedagogical tool to help explain the idea of superposition, a concept I anticipated that students would find unintuitive.

### 0.4 WORKSHOP FORMAT

My segment of the Modern Physics workshop took place virtually over the course of two days with each day containing a two hour session. The workshop content can be categorized as follows: classical computing, the qubit and quantum effects, "Quantum

Bank Robbery” activity [5], Deutsch’s algorithm and quantum no-gos.

To provide context for quantum computing, classical computing was introduced first. The role of the central processing unit (CPU) and the arithmetic and logic unit (ALU) were discussed.



**Figure 2:** Cartoon demonstrating CPU and ALU role in classical computer. (Source: Nkeiru Ubadike)

Logical operation such as NOT and OR were introduced. An exercise entitled "Bit Swapping with XOR" which involved swapping the value of two variables using an algorithm making use the XOR operation gave participants hands on practice performing logical operations. Anticipating the section <sup>1</sup> in which the students would build their own quantum circuits, great emphasis was made on the fact that the ALU consisted of transistor circuits each of which performed a specific arithmetic or logical operation.

A two part group activity [2] served as a prelude to the section on the qubit and quantum effects. The first part, entitled "1 bit channel", the students were each secretly given an integer. Each integer adds up to a collective integer. The student assigns a bit value to their integer depending on if it was odd or even. Students were assigned a place in a virtual line. The first person in line passes their bit value to the second via a direct message. The second person performs the XOR operation between the bit passed to them and their own bit and passes on the result. Subsequent participants repeat the aforementioned process. The last participant reports the final bit value. If the bit is 1, the collective integer was odd and vice versa. The participants win if the last person correctly reports if the collective integer was odd or even. In the second part of the group

<sup>1</sup>The qubit and quantum effects

activity entitled "1 qubit channel", participants were each secretly given a real number. Each real number adds up to a collective integer. Instead of assigning bit value (which is not possible if the assigned number is a decimal, irrational number etc), students were given a "disc" conceptually representing a qubit. The first person "rotates" the disc by value of their assigned number multiplied by  $180^\circ$  and passes on the "disc". Then, the second person rotates the disc by their assigned number multiplied by  $180^\circ$ . Virtually, this corresponds to performing the following calculation:

$$\theta_{current} = \theta_{previous} + (x_{assigned} \times 180^\circ)$$

where  $\theta_{current}$  is current angular orientation of the disc,  $\theta_{previous}$  is the angular orientation passed from the previous participant and  $x_{assigned}$  is the real number secretly assigned to each participant.  $\theta_{current}$  is passed to the next participant via a direct message. The last person reports the final orientation. If this final value is odd multiple of  $180^\circ$ , then the collective integer was odd and vice versa. The participants win if the last person correctly reports if the collective integer is odd or even. This group activity gives intuition about the difference between quantum and classical information processing.

After, the qubit was introduced, first, in an abstract sense from the standpoint of quantum information science using Dirac notation[2]. Then, from a physical standpoint, I briefly discussed the objects such as photons, ions and superconducting circuits that physicists use as qubits in actual quantum hardware. Finally, I introduced the coin analogy of a qubit (See figure 1). Using IBM composer [4], two circuits (pictured below) were then demonstrated.



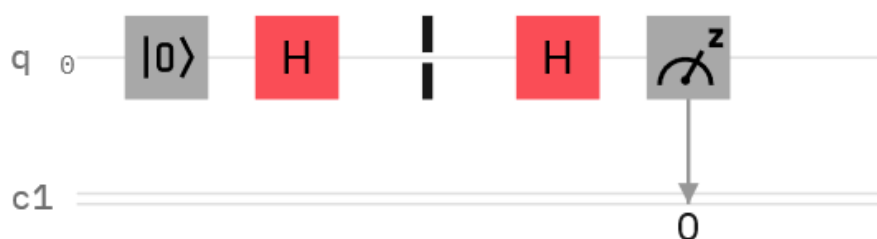
**(a)** "First Quantum Circuit"  
(Source: IBM Composer Field Guide)

**(b)** Bit Flip  
(Source: IBM Composer Field Guide)

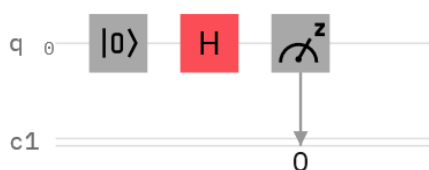
"First Quantum Circuit" and "Bit Flip" prepares a qubit in the  $|0\rangle$  and  $|1\rangle$  states respectively. Running these circuits on real quantum computers provided an opportunity to introduce the notions of quantum error, measurement and probability. After this introduction the following quantum effects and relevant quantum gates were discussed: su-

perposition, Hadamard gate, interference and entanglement. I presented video resources [3], [1] to complement lecture. Each effect made use of the cloud analogy [5] along side Dirac notation to demonstrate simple calculations.

The students were instructed how to build a circuit on IBM Quantum Composer demonstrating each quantum effect that was introduced (interference, entanglement etc)



(a) Interference circuit

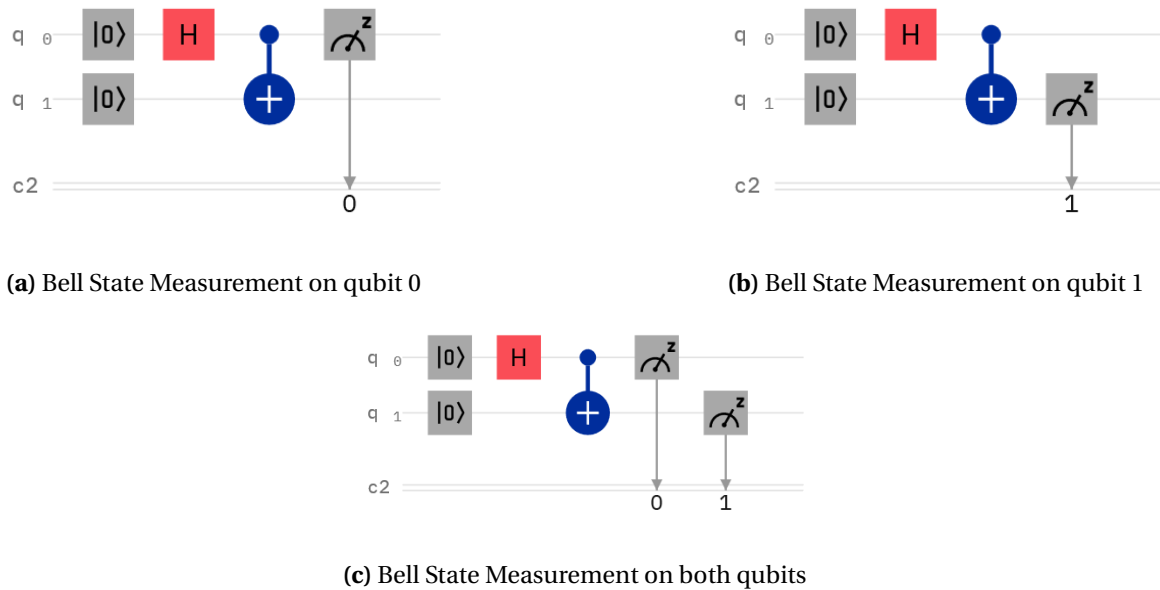


(b) Superposition of  $|0\rangle$  qubit



(c) Superposition of  $|1\rangle$  qubit

**Figure 4:** Circuits demonstrating interference and superposition (Source: IBM Quantum composer Field Guide [4])

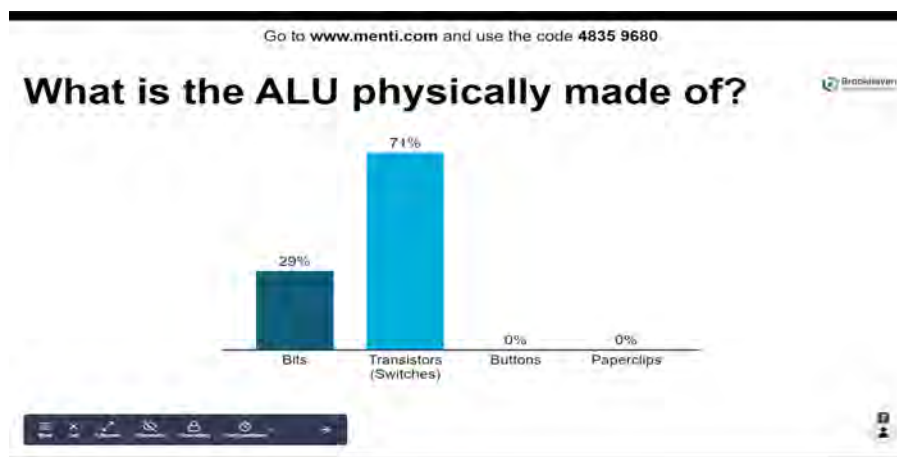


**Figure 5:** Circuits demonstrating entanglement (Source: IBM Quantum composer Field Guide [4])

Throughout the lesson, the students understanding was gauged using mini-quizzes.

#### Mentimeter Check In Questions

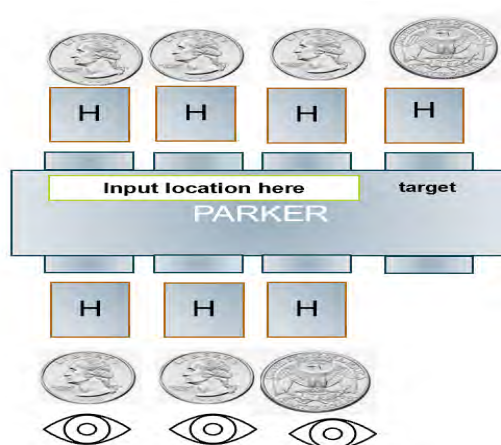
In an effort to promote engagement and gauge student's understanding, a series of check in questions were delivered virtually to students via the real time voting and feedback platform, Mentimeter.



**Figure 6:** Example Mentimeter Check in Question

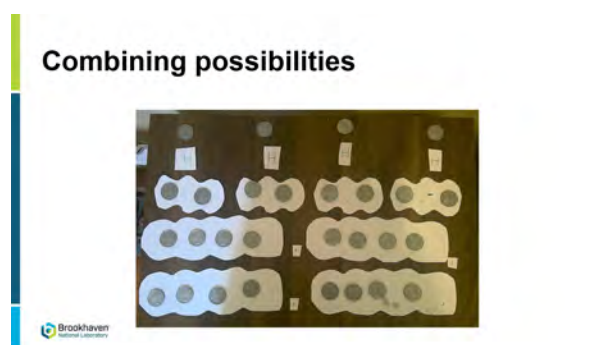
The following content was unable to be included in the workshop session and thus will only be briefly discussed here.

The “Quantum Bank Robbery” activity [5] was intended to expose participants to quantum algorithms. The premise is that a "bank robber" could use quantum effects to figure out whether a vault contained genuine or fake bars. The figure below shows the in-vault computer named "Parker"<sup>2</sup> with the Hadamard gates that the "bank robber" has added.



**Figure 7:** “Quantum Bank Robbery” Activity: Parker Computer (Source: Nkeiru Ubadike)

As shown, four coins are inserted. If any of the first three coins comes out as a tails, then the vault contains genuine bars. The activity revolves around making the calculation to prove that this is the case. In an effort to make the workshop engaging, I adapted the original calculation [5] into a hands on activity instead using paper clouds and coins to demonstrate each step of the calculation.



**Figure 8:** Excerpt of “Quantum Bank Robbery” Activity (Source: Nkeiru Ubadike)

<sup>2</sup>To incorporate some physics history content and promote awareness, I named it for the first African American women known to have completed a physics graduate degree

The students were to perform the last step of the activity in groups giving them hands on "calculating" a quantum algorithm.



**Figure 9:** The group activity section of the “Quantum Bank Robbery” activity(Source: Nkeiru Ubadike)

The activity is based on the Deutsch- Jozsa algorithm.

The last section was supposed to provide a more rigorous description [6] of the algorithm that the “Quantum Bank Robbery” activity was based on, heavily relying on Dirac notation. It also intended to dispel the notion that quantum computers will ultimately replace all classical computers by presenting some of the operations a quantum computer cannot perform, collectively known as "quantum no-go" theorems [2]

## 0.5 WORKSHOP EVALUATION METHOD

A pre-workshop survey was given to the participants at the start and a post-workshop survey at the end of the week long session. The pre-workshop survey anonymously assessed student’s academic background, their perception of particle physics and quantum computing, the interest in pursuing physics-based careers and their level of physics engagement outside the classroom. The pre-workshop survey also anonymously collected basic demographic data. The post-workshop survey aimed to assess if participant’s perceptions of particle physics and quantum computing had changed. It also asked participants to rate the difficulty of the entire workshop and identify their favourite and least favourite day and activity.

In addition to the surveys, the engagement level of participants was recorded. An engagement was counted each time a participant asked a question or made an unprompted comment.

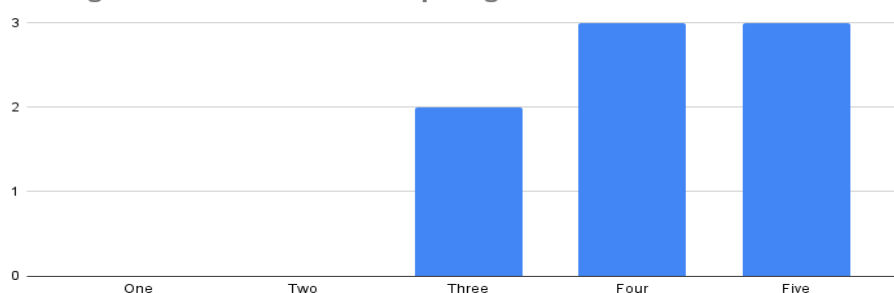
## 0.6 WORKSHOP EVALUATION RESULT

Only data from participants(identified via anonymous codenames) who had completed both surveys<sup>3</sup> were included in this analysis.

### Perception of Difficulty of Quantum Computing

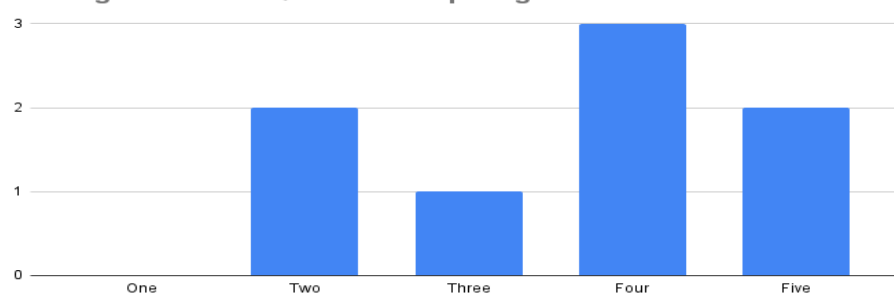
In the pre-workshop and post- workshop survey, participants were asked, on a scale of 1 to 5, how much they agreed with the following statement: quantum computing is difficult to understand. The results of workshop 1 are presented below. The average difficulty perception score was reduced from 4.1 to 3.6. A t-test performed on the pre survey and post survey data showed that the shift was not statistically significant.

**Pre survey: On a scale of 1 to 5, how much do you agree with the following statement: Quantum computing is difficult to understand.**



**(a) Pre Survey Results: Left skewed distribution with a skewness value of -0.28**

**Post survey: On a scale of 1 to 5, how much do you agree with the following statement: Quantum computing is difficult to understand.**



**(b) Post Survey Results: Less skewed with a skewness value of -0.39**

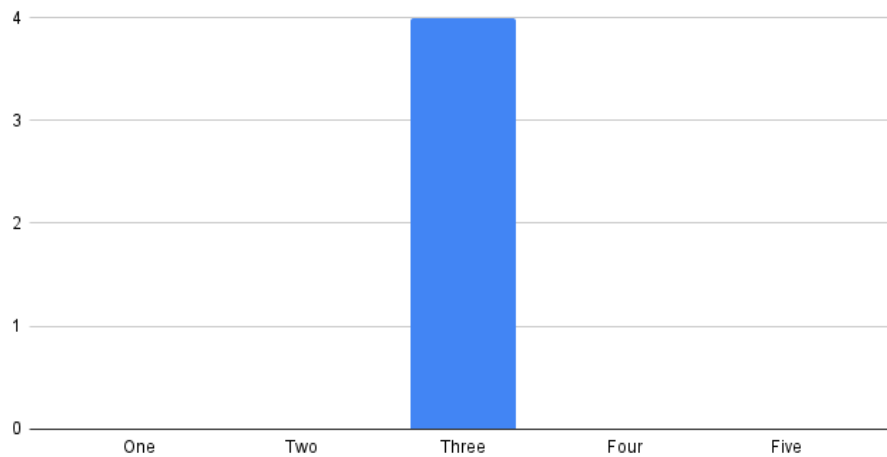
**Figure 10: Workshop 1**

<sup>3</sup>Workshop 1: 8 respondents, Workshop 2: 4 respondents



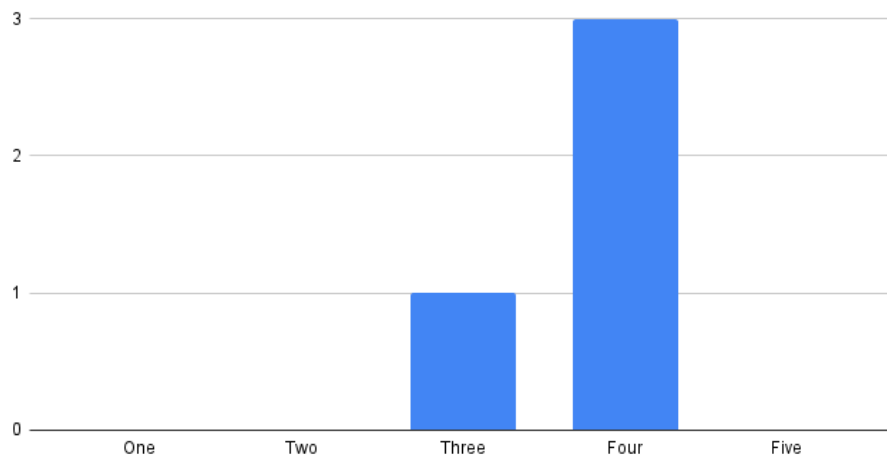
The results of workshop 2 are presented below. The average difficulty perception score increased from 3 to 3.75.

**Pre survey: On a scale of 1 to 5, how much do you agree with the following statement: Quantum computing is difficult to understand.**



**(a) Pre Survey Results**

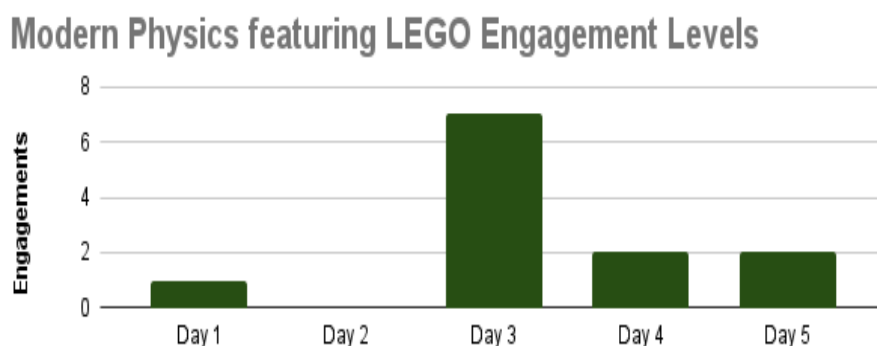
**Post survey: On a scale of 1 to 5, how much do you agree with the following statement: Quantum computing is difficult to understand.**



**(b) Post Survey Results**

**Figure 11: Workshop 2**

**Engagement Levels** The engagement levels for workshop 1 are presented below. The spike in engagement levels on day 3 can be accounted for by the inclusion of engagement with the invited speaker, radiation chemist Theresa Grimaldi working with medical isotopes at BNL. Excluding day 3, the daily average engagement level was 1.75. For my segment of the workshop, Day 4-5, the daily average engagement level was 2. No engagement data was taken for workshop 2.



**Figure 12**

**Written Feedback** Participants were asked to select their favourite and least favourite days/activities. The comments pertaining to my segment are presented below. Positive comments are highlighted in green and negative comments are highlighted in red.

<b>Workshop 1 Feedback</b>
4th day ; I liked when we did the rotation gates
The group activity we did on Thursday and the lego assembling
5th ; Entanglement was a little confusing
Thursday, the quantum computer.
I felt like the last day (friday) was difficult to understand and making different circuits just didn't pique my interest.
I don't think I'm very interested in computer science, but doing math with Exor and bits was easy.
It was a bit less interesting learning about quantum computing without a visual guide. Near the end some of the quantum computing was a little confusing due to quantum physics following laws I'm not used to.
I found a lot of the quantum computing, especially with IBMq, very confusing.

**Table 1:** Feedback from Workshop 1 (July 26- 30)

<b>Workshop 2 Feedback</b>
I enjoyed the lego modeling as well as the quantum computing block-chain code.
I think day 5 was probably my least favorite day, since the activities just felt like an extension of day 4 with a few extra topics tacked on.
Friday, I did not really enjoy the Quantum Composer website.
I didn't enjoy thursday and I wasn't a fan of the quantum computing because I didn't quite understand
Anything about quantum computing or quantum physics in general; that's my weakest subject.

**Table 2:** Feedback from Workshop 2 (Aug 2-6)

## 0.7 CONCLUSION

Both segments (particle physics and quantum computing) of the two workshop sessions were characterized by low engagement levels despite most respondents indicating in the pre-survey that they engaged with STEM outside the classroom. A delay in/lack of response to questions I asked participants made for an awkward teaching environment. I conclude, at least for this iteration, that presenting content in a variety of media is not enough to mitigate the problem of low engagement in virtual environments. From observing numerous other workshops hosted by OEP, this is a persistent problem faced by OEP educators.

For the first workshop, although the reduction in the perception of difficulty was not statistically significant, it is an encouraging sign that it was reduced by 1.5 points. Also encouraging was the response to the "1 bit channel" and "1 qubit channel" group activity. There was a 100% participation in this activity. There was also engagement with the other exercises such as "Bit Swapping with XOR" and Mentimeter quizzes.

Analysing the feedback leads me to the conclusion that this iteration of the workshop was unable to make quantum computing more understandable. Many students mentioned feeling confused. Additionally, around half of my prepared content was not delivered. This is a common occurrence in educational settings.

I would improve this virtual workshop by doing more time trials of my prepared content so that I could prioritize adding the Quantum Bank Robbery activity. I believe the students would have appreciated the hands on nature of this activity as well, the group work component as well as the variety in content delivery it would have provided. I would add more group work components particularly involving the IBM Quantum Composer. Finally, I would incorporate a section about the very latest research being conducted

in quantum computing to give students an idea of what a possible career in quantum computing could look like for them.

# Bibliography

- [1] Nkeiru Ubadike (Tiktok: @nqubits). *Explaining superposotion using ice cream*. URL: <https://drive.google.com/file/d/1wnPxY0BrfLC4P-SE5qFm1UoBW8JzleuK/view?usp=sharing>.
- [2] Chris Ferrie. *Introduction to Quantum Computing*. URL: <https://csferrie.medium.com/introduction-to-quantum-computing-df9e1182a831>.
- [3] Parth G. *Quantum Entanglement Explained for Beginners | Physics Concepts Made Easy*. URL: <https://youtu.be/-WSWz1H3mJg>.
- [4] IBM. *IBM Quantum Composer Field Guide*. URL: <https://quantum-computing.ibm.com/composer/docs/iqx/guide/>.
- [5] Terry Rudolph. *Q is for Quantum*. Terence Rudolph, 2017. ISBN: 9780999063507.
- [6] Steven Spicklemire. *Lesson 38 Quantum Computing, Deutsch's Problem*. URL: [https://www.youtube.com/watch?v=5xsyx-aNC1M&list=PLUQUmchCCkOKQ7c-U\\_tUGkT9yKCG3D0bI&index=27](https://www.youtube.com/watch?v=5xsyx-aNC1M&list=PLUQUmchCCkOKQ7c-U_tUGkT9yKCG3D0bI&index=27).

# Developing tools for analysis of spectro-microscopic data obtained from 2D materials and surfaces

Jeannet Vargas

Department of Mathematics and Computer Science, St. Joseph's

College, Patchogue, NY 11772

Jurek Sadowski, Zhongwei Dai

Center for Functional Nanomaterials, Brookhaven National Laboratory,

Upton, NY 11973

**8 August 2021**

## **Abstract**

This research project, with the Electron Spectro-Microscopy (ESM) beamline at The National Synchrotron Light Source II (NSLS-II) in Brookhaven National Laboratory, is a study on improving a python generated software program designated for X-ray photoemission electron microscopy and low-energy electron microscopy (XPEEM/LEEM) research. Data extracted from 2D nanomaterials and surfaces allow for comprehensive investigation of their topographic, structural, chemical, and magnetic properties, on the nanometer scale. With the use of the XPEEM/LEEM microscope on the ESM beamline, data of the atomic layers on a surface is obtained as images, while the electron, or photon energy is varied. There are two successful open-source software programs that allow us to visualize the structure of a surface and the ranges in electron intensity within its layers: the Python Low-Energy Electron Analysis Suite (PLEASE), and ImageJ. While comparing these two methods, adjustment of the image's contrast and sample area selection size were found as two important aspects for accurate data exploit and analysis, and these changes will be added to the PLEASE software. For such, I have been examining the python source code of the PLEASE program and was able to successfully add an input section and button for users to control the selection size and began an image menu bar that will open a window for users to enhance the loaded image to their liking. Both allow for precise data extraction from a sample area, with its varying sizes in domain, and the improved ability to identify key features and location of interested sample areas.

## **I. Introduction**

As studies within nanotechnology progress, the need for effective software programs to process this bulk data is also on an extremely high demand. We experiment on two-dimensional nanomaterials and their characteristics and aim to develop efficient programs for effortless execution and data extraction. Nanotechnology involves the manipulation of matter at the atomic level to create new functions and properties. Until the Electron Spectro-Microscopy beamline (ESM) at the National Synchrotron Light Source II (NSLS-II), it has been difficult to examine a material's properties since they are held in such miniscule areas. Presently, through studies with the beamline, we were able to understand the electronic, chemical, and physical properties of many 2D materials by their link to their surface structure.

2D materials have only one dimension within the nanoscale (1-100 nm), where its atomic structure resembles a large, thin sheet of paper. These nanomaterials make it possible to reduce the thickness of material down to a single atom.<sup>2</sup> These properties are what interest the study with the ESM beamline and promoted experiments of X-ray photoemission electron microscopy (XPEEM), and Low energy electron microscopy (LEEM).

The XPEEM/LEEM microscope that is available on the ESM beamline, exerts varied electron intensities towards a sample surface for interaction, and then redirects them to an imaging software and received as data. To process this data for understanding of a surface's properties, a designated software program that is accustomed for examining the images extracted from the beamline was needed. Thus, the Python Low-energy Electron Analysis Suite (PLEASE) was developed by Dr. Maxwell Grady in 2017. This program allowed for an easy interface for visualization of data sets and a convenient "point-and-click" method for extracting electron intensity voltage (IV) curves at selected areas on the image of a surface.



Though with new programs also comes new developments and improvements to further an enhancement in data extraction and results. As the PLEASE program is written in the Python coding language, I was able to implement a few changes that allow this. Therefore, my work with this program aimed to better the results taken from the PLEASE program, enhance this designated software for continued use in the future, and continue to make the program user-friendly. This report defines the modifications done to the PLEASE program for these enhancements, and the next steps for further development.

## **II. The Python Low-Energy Electron Analysis Suite (PLEASE)**

Developed by former Brookhaven National Laboratory employee in 2017, and currently a software developer at Enthought software company, Dr. Maxwell Grady, the Python Low-energy Electron Analysis Suite (PLEASE) was created to enable rapid analysis and visualization of “[Low-energy electron microscopy] LEEM and [low-energy electron diffraction]  $\mu$ LEED data sets. LEEM and  $\mu$ LEED are powerful techniques for surface analysis of many types of novel materials and are especially useful for the study of two-dimensional (2D) materials.”<sup>1</sup> With his program, specific emphasis is placed on the electron Intensity-Voltage (IV) data sets that can be extracted from LEEM data through its microscope on the Electron spectro-microscopy beamline (ESM). These IV curves map the scattered or reflected electron intensity as a function of incident energy. Through the connection of electronic properties and the surface structure of a material, LEEM and  $\mu$ LEED provide authentic methods for study of material properties in reduced dimensions and aim to guide future applications of these novel materials.

PLEASE is a software package that supplies an open-source cross-platform graphical user interface available for easy download and installation from GitHub.com. It provides a user-friendly “point-and-click” method for extracting IV curves from a loaded experiment, and basic

adjustment tools for smoothing the IV curves, reducing instrument noise, and subtract background signal from the IV data.

Both LEEM and LEED data sets that are downloaded with the program are loaded as images and displayed on the left-hand side of the interface. To the right, has the plotted IV curves that correspond with the mouse movement along the loaded experiment. **Figure 1** shows a screenshot of the full PLEASE interface, and a loaded experiment for reference. The user could toggle the IV curves to measure reflectivity rather than intensity voltage, and with clicking, the user can see the curve at their desired point and output their data for easy download.

As all this information is available on GitHub, especially the python source code that exposes the inner functions of the program, I was able to start my project in adding modifications to better the program, while maintaining its user-friendliness and simplicity.

### **III. ImageJ**

Changes made to the PLEASE program were inspired by a second open-source software program, Fiji's ImageJ. This program is also available to download and installation and is rather coded in Java. It is an image processing and analysis program that is essentially used in the biomedical field. It does geometric transformations, image enhancements, and the user can choose the shape and size of where they extract data from the image displayed.

I chose two of ImageJ's capabilities to be coded into the PLEASE program, as they would help in further research within LEEM data sets. These were: the change in selection size when clicking to output IV data, and the ability to change the contrast and brightness of the image. **Figure 2** shows ImageJ and its functions that were chosen to be added into the PLEASE source code.

### **IV. Selection Size Patch Width**

An examination of the PLEASE source code files was needed to implement these changes. There are 14 python scripts used to run the PLEASE program, with the main ones needed for my modifications being the please.py script and the LEEMFUNCTIONS.py python script. The please.py file consists of 2,555 lines of code that call the other scripts as methods, creates the visual graphical user interface, and connects user events to error messages, other actions and more. What I am interested in this script is the creation, and calling of the selection area patch width, which is what is used to print a circle and extract IV curve data from a user's desired point. Within a method defined as "handleLEEMClick" Grady hard-codes, or defaults, the radius of the patch appended when the mouse is clicked to an integer of 8 pixels. This aspect is what I needed to change to have the user adjust the size of the circle's radius.

To maintain feasibility, I found that the most efficient method to change the patch width radius is through adding a section, input text box and button in the configuration tab that has already been coded into the interface. To my advantage, Dr. Grady's code is extremely organized and commented, so I was able to find the methods I needed without too much strain in 2,555 lines of code. With a method that defined all the information to view in the configuration file, I could mimic his code and add what was needed for this modification. **Figure 3** and **Figure 4** show the before and after my addition in the configuration tab.

Once the input section and button were connected, along with specific messages to keep the user input an even integer for a circle, I was then able to go into the handleLEEMclick method and assign the radius to equal the user's input, rather than a default size. This completed function allows for the user to change the size of their selection area to their liking, which overall allows for the ability to see into very small features that could be present on a surface, and better's the IV curves extracted at those points.

## V. Image Enhancement

The second desired alteration, an image enhancing feature, was found as necessary for clarification of loaded experiments, focusing of the image, and a method to adjust the brightness or contrast if needed. This ability would allow for clearer view of a surface's characteristics for further data extraction. This method had more information in multiple python scripts compared to the patch width change function solely in the please.py script. For this function to work I needed to access the LEEMFUNCTIONS.py script, qthreads.py script, and create my own python script which I named adjImage.py to start modification. The first script had the methods to call the images from the user's directory and transfers those data files into a 3D NumPy array to be outputted. While the data we receive from the XPEEM/LEEM microscope is in the form of images, they are downloaded as binary (.dat), rather than an image file (.jpg, .jpeg, .png). Because of this, when loading the experiment into the PLEASE program, to process that data and upload the image, it is converted into an array.

My own adjImage.py python script was needed to test uploading an image and implementing editing modules. Imported imaging modules such as OpenCv, PILLOW, and Matplotlib were tried to execute this before adding them to the PLEASE program, though I found that this method needs to be within the program to function correctly. Therefore, I linked my script to the please.py file that calls the other python files, and in there added an Image menu bar, a button, and window for the user to open when prompted to enhance an image.

## VI. Results

The change in selection size was successfully programmed into the PLEASE software and allows for finer extraction of outputted IV data compared to the previously defaulted patch size. **Figure 5, Figure 6, and Figure 7**, show a comparison of three IV curve plots, done through the

terminal graphing software Gnuplot, that prove enhancement in data taken from patch size that is inputted to fit a desired selection area. As this change is added and uploaded to GitHub, users will be able to download Grady's program with my advancement.

The results from the image enhancement function are currently inconclusive. The image after loading into my own python script outputs distorted or not at all, which adds challenges to implementing an imaging module. The reason for this is the need for all the data files to be loaded together, and the size of the completed array once converted. The XPEEM/LEEM tools result in data in the form of images, as stated before, though rather than a few image files, they download together as a couple hundred. Therefore, when the process of converting the data to a 3D array has executed, the resulting size of the array is extremely large, and difficult to execute for the imaging modules.

The next steps for the image editing process are to continue research in 3D NumPy arrays on whether I would be able to change the values in the array and what algorithms are needed to implement a contrast change to arrays. To do so, a trial is to be done by copying the variable that holds the already uploaded images, and attempt to open and edit that image in the window I created in the Image menu bar. Once these trials result in an output, the image enhancing method will be further developed to and allow for superior data analysis.

## **VII. Conclusion**

The development of a designated, LEEM processing software program is necessary for stronger data analysis of 2D nanomaterials, and overall allow for clear examination of a surface's electronic, chemical, and physical properties. Through research in the developed Python Low-energy Electron Analysis Suite (PLEASE) and its source code, a new development in the selection size patch width, and the start of an imaging enhancement function, inspired by

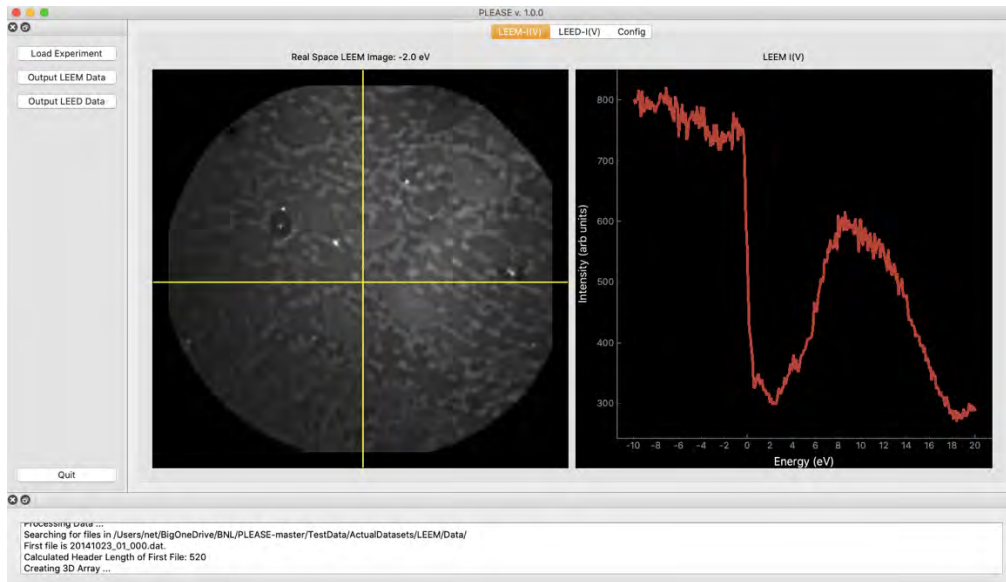
methods within a former software program, ImageJ, were added and result in advanced output electron intensity voltage (IV) data extraction. To continue improvements in this program, research is to be done on the formed arrays, their ability to change values and effects with contrast change and adjustments, and how the image's pixels are modified.

## VIII. References

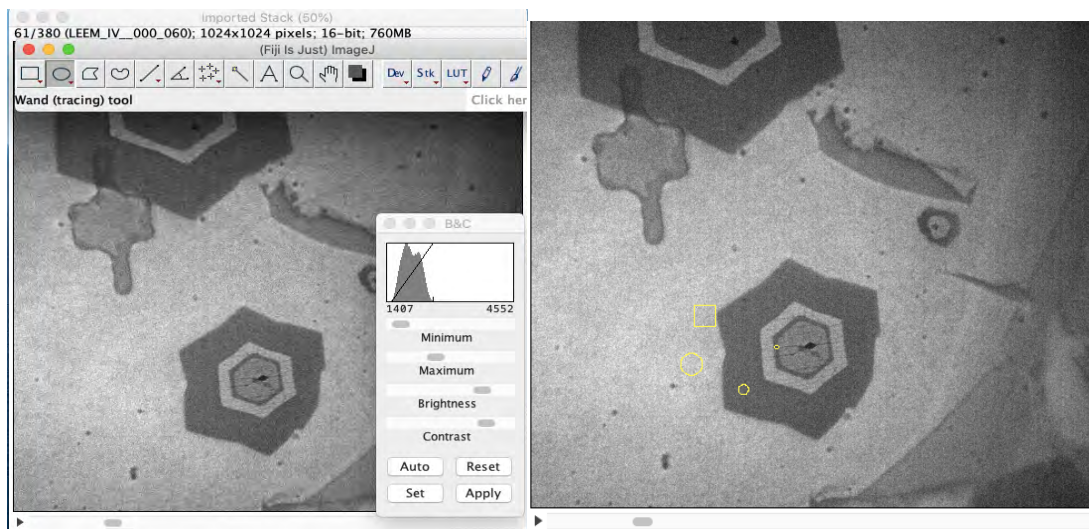
<sup>1</sup>mgrady3. “mgrady3/PLEASE: Please: The PYTHON Low-Energy Electron Analysis Suite - Enabling Rapid Analysis and Visualization Of LEEM AND LEED Data.” *GitHub*, [github.com/mgrady3/PLEASE](https://github.com/mgrady3/PLEASE)

<sup>2</sup>“2D Materials: An Introduction to Two-Dimensional Materials.” *Ossila*, Ossila, [www.ossila.com/pages/introduction-2d-materials](https://www.ossila.com/pages/introduction-2d-materials).

## IX. Figures

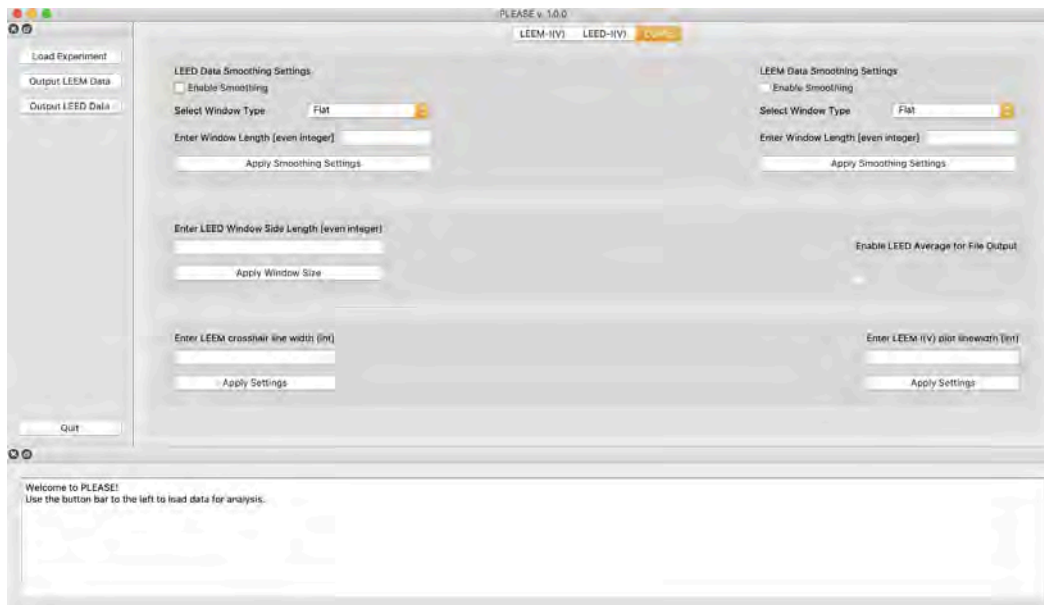


**Figure 1:** A screenshot of the complete PLEASURE program with a loaded experiment. The IV plot is to the right of the program and the loaded image data to the left. The plot corresponds with the crosshair position on the loaded experiment.

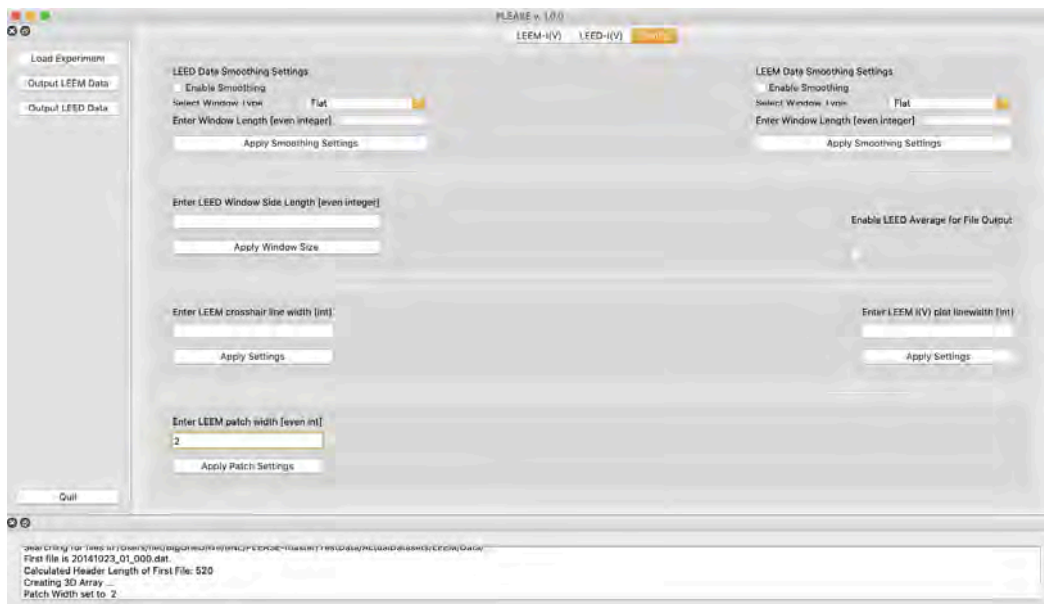


**Figure 2:** ImageJ with a loaded raw data set, including the adjustment abilities on the left and the selection size variation available on the right.

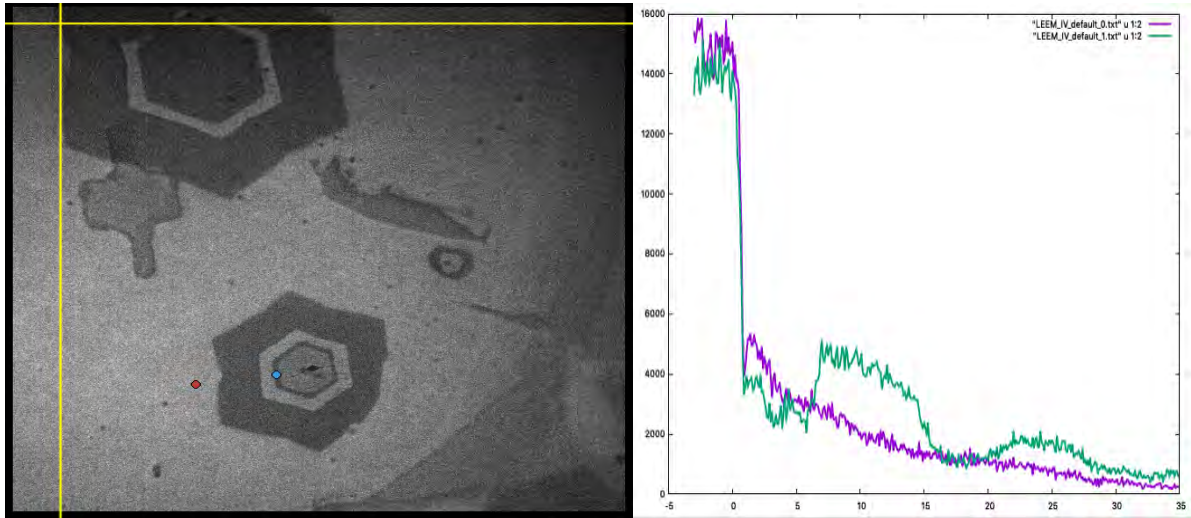




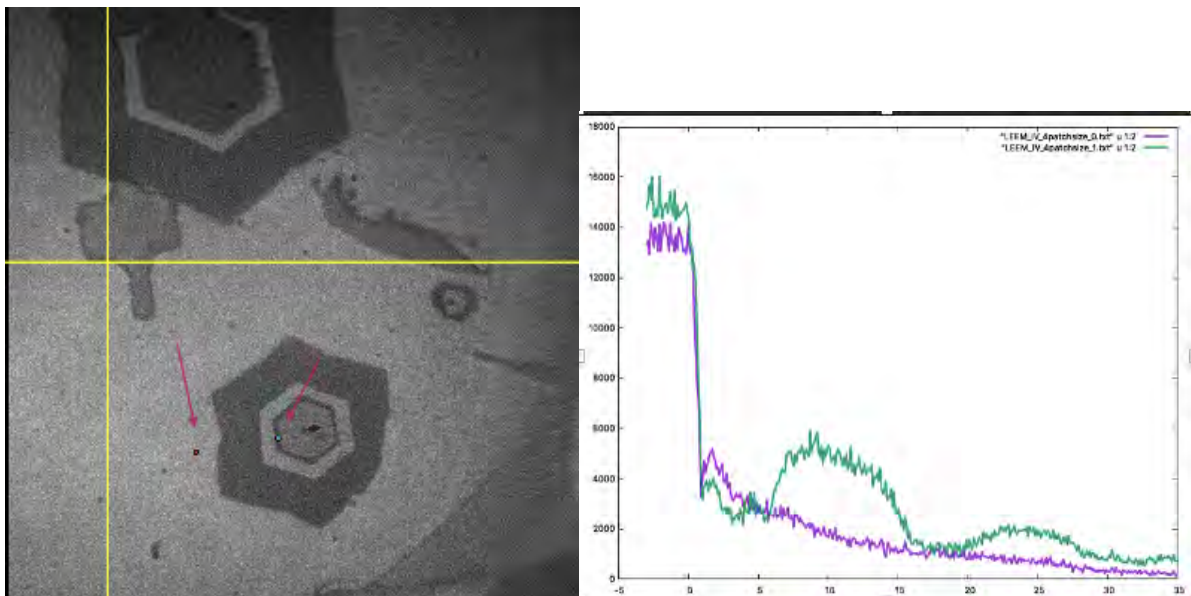
**Figure 3:** Dr. Maxwell Grady’s PLEASURE program downloaded directly from GitHub. The configuration file where the user can adjust aspects in the program before my addition.



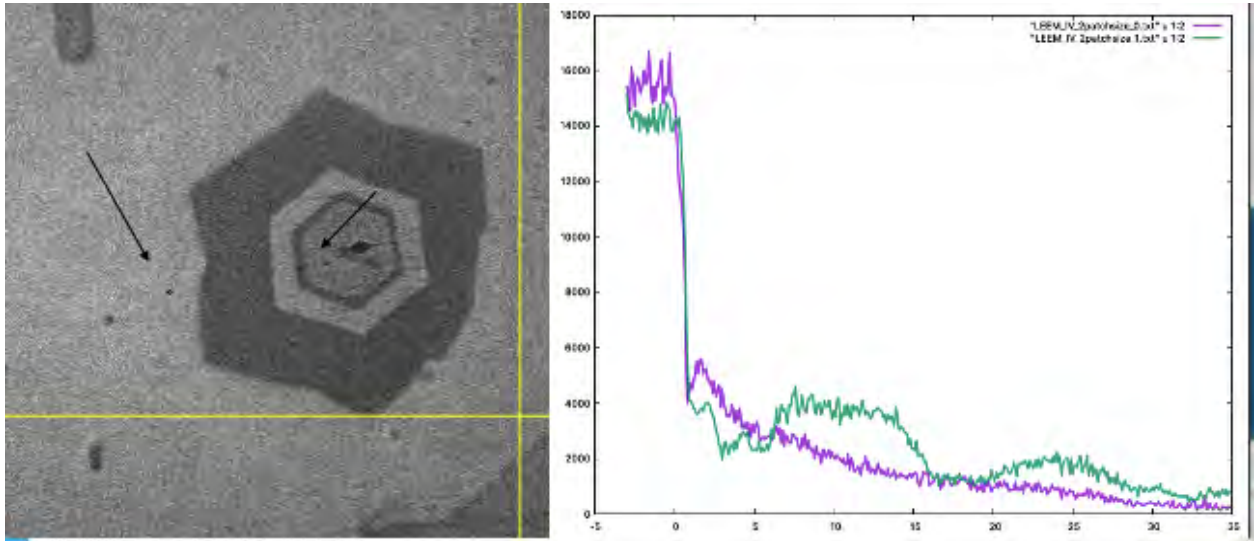
**Figure 4:** The configuration tab with my addition of patch size input. User is prompted to enter an even integer, click the “Apply Patch Settings,” button, and results in a changed circle when image is clicked on.



**Figure 5:** Comparison of the first and fourth layers of surface's image. Screenshot from the PLEASE program on the left with the default patch size of 8, and on the right Gnuplot graph of outputted IV data from patch points.



**Figure 6:** Added modification of patch width change. Size 4 set in the configuration tab, compared to size 8 it is a lot smaller and precise into the finer fourth layer. Gnuplot graph to the right shows differences in peaks compared to the size 8 curve.



**Figure 7:** Last changed patch size width with new development to size 2. More definite in the image's surface. Able to clearly pinpoint the fourth layer's IV curve without strain of blending layers together. Gnuplot graph to the left shows increased differences in first and fourth layers- prove as enhanced method of extracted output IV compared to patch size 8.

## **X. Acknowledgements**

### **Science Undergraduate Laboratory Internships (SULI)**

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

**Center of Fundamental Nanomaterials (CFN) Department of Brookhaven National  
Laboratory, Upton NY 11973**

Dr. Jurek Sadowski , Dr. Zhongwei Dai

Department Name  
Building #  
P.O. Box 5000  
Upton, NY 11973-5000

[www.bnl.gov](http://www.bnl.gov)

# **Civil Engineering Design for the SRF Support Facility**

Oksana Vysochanska

School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853

Joseph E Tuozzolo

Collider-Accelerator Department, Brookhaven National Laboratory, Upton, NY 11973

## **Abstract**

The Electron Ion Collider (EIC) at Brookhaven National Laboratory (BNL) will allow scientists to create snapshots of the internal structure of atoms by colliding hadron and electron beams at very high energies, 275 GeV and 17 GeV, respectively. 400 kW solid-state power amplifiers (PA) for the superconducting radio frequency (SRF) cavities that accelerate and store these beams will be housed in a two-story SRF support facility outside of the ten o'clock interaction region (IR) of the EIC. A 30 MVA transformer field outside of this facility will power the PA's and all the additional equipment. However, around 19 MW of power is lost to heat energy in the radio frequency (RF) equipment, requiring water cooling to maintain ideal performance temperatures. A water pump room, external to the support facility to prevent vibration of the sensitive RF system, will provide the cooling water to both the support facility and the IR. The IR will contain SRF cavities cooled to cryogenic levels, focusing magnets, and vacuum chambers for the EIC particle beam along with the supporting equipment for water cooling, electric power, HVAC, liquid helium, and other essential equipment. The aim of this project is to draft a 3D model using Creo Parametric to develop a viable infrastructure plan to further understand the civil engineering systems interfaces with the scientific equipment in the IR and the SRF support facility. Physical space constraints, electrical systems, water cooling, helium cooling, radiation shielding, and personnel safety requirements were major considerations throughout the project. Through collaboration with a multidisciplinary team, I gained practical experience applying civil engineering design principles to an ongoing project and developed my skills with Creo Parametric.

## I. Introduction

Within nuclear physics, there is a large gap in knowledge about quarks, gluons, and the strong force that bind these building blocks of our universe. To understand the pieces of this puzzle, the Electron-Ion Collider (EIC) is being designed and built using repurposed infrastructure from the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory. This report will focus on the design and layout of Building 1010, a new superconducting radio frequency (SRF) support facility that will house solid-state power amplifiers and supporting equipment that provide the power to accelerate and store high energy electron and hadron beams in the EIC.

Within the EIC, a hadron beam and an electron beam travel in opposite directions in short bunches at near relativistic speeds. There are about 300 bunches in each ring, each bunch circulating a 2.4-mile circumference ring 76,000 times a second. The 17 GeV electron beam and the 275 GeV hadron beam will be accelerated and stored until they reach a collision point at two of six interaction regions (IR). These IR's are expanded tunnel sections where the RHIC accelerator originally collided heavy ions. Large detectors at the two collision points will be able to collect a three-dimensional snapshot of all the particles scattering, providing clues to how the particles interact with each other and their internal structures. However, before the collision can happen, the beams need to be accelerated and maintained at high energy as they travel along a 2.4-mile circumference ring.

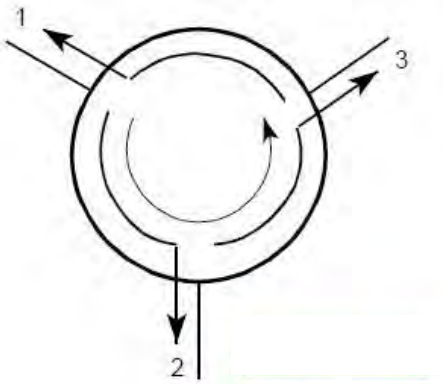
One of the technical challenges is the loss of energy in the electron beam as the beams travel along the circumference of the tunnel. The electrons beams lose synchrotron light energy, slowing them down from nearly relativistic speeds. By injecting



radiofrequency (RF) energy into the electron storage ring, the energy of the stored electron beam is maintained at high energy. At the ten o'clock IR, the RF energy will be sent through waveguides from solid state power amplifiers into superconducting RF (SRF) cavities through which the beam travels. At the beginning of this project, there were 22 SRF cavities that would require 40 power amplifiers (PA). This includes 18 400 kW electron storage ring (ESR) cavities, 3 75kW rapid cycling synchrotron (RCS) cavities, and 1 75 kW hadron storage ring (HSR) cavity. However, the design of the ESR cavities is currently being studied as all 18 cavities do not fit within the IR. A temporary solution is used in this project that will be detailed within this report. Spacing is a challenge in the IR as each SRF cavity has one or two waveguide ports that need large, bulky tuners placed as close to the port opening as possible. Helium supply lines, 2' in diameter, are also needed inside the IR because the cavities are superconducting for high efficiency and must be cooled to cryogenic temperatures with liquid helium. Sextuple, quadrupole, triplet, and corrector magnets are also distributed along the 3 beam lines.

The new SRF facility, known as Building 1010, is needed to house all of the 40 PA's with waveguide runs connecting them to the SRF cavities. Each PA needs a RF circulator and dummy load located along the waveguide run. A circulator is a device that controls the direction of the RF signal transmission. Illustrated in Figure 1, RF energy is inputted into one of the three ports and can only exit out of the next port in either a clockwise or counterclockwise direction. The circulator protects the PA from damage when RF power is not absorbed by the Hadron or electron beams. The SRF cavities reflect the RF power back along the waveguide. This high RF power returning signal can

only be transmitted into the third port which is connected to the dummy load. The dummy load absorbs the RF energy that is bounced back.



**Figure 1.** A counterclockwise circulator diagram showing the possible paths that the RF power passing through can take from “How RF Circulator Works.” *RF Wireless World*, RF Wireless World, 2012, [www.rfwireless-world.com/Terminology/How-RF-Circulator-Works.html](http://www.rfwireless-world.com/Terminology/How-RF-Circulator-Works.html).

All the major RF equipment needs to be water-cooled. Being only 50% efficient, the PA's create 400 kW of excess heat that needs to be removed. Under normal operations with beam, only 53.4 kW of heat is absorbed by the dummy load from RF waves bouncing back along the waveguides. However, there is a possibility that when there is no electron beam in the SRF cavities, each of the dummy loads must reabsorb the 400 kW of power that is sent to the SRF cavities by the PA's. All of this energy is converted into heat and needs to be cooled to prevent a complete meltdown of the equipment. The circulator is also not 100% efficient in transferring the RF power so 10 kW of heat energy is continuously dissipated, even during testing or beamline failure cases. The cooling water will come from a water pump room separated from the main building because the RF equipment performance can be negatively impacted by vibrations from the pumps.

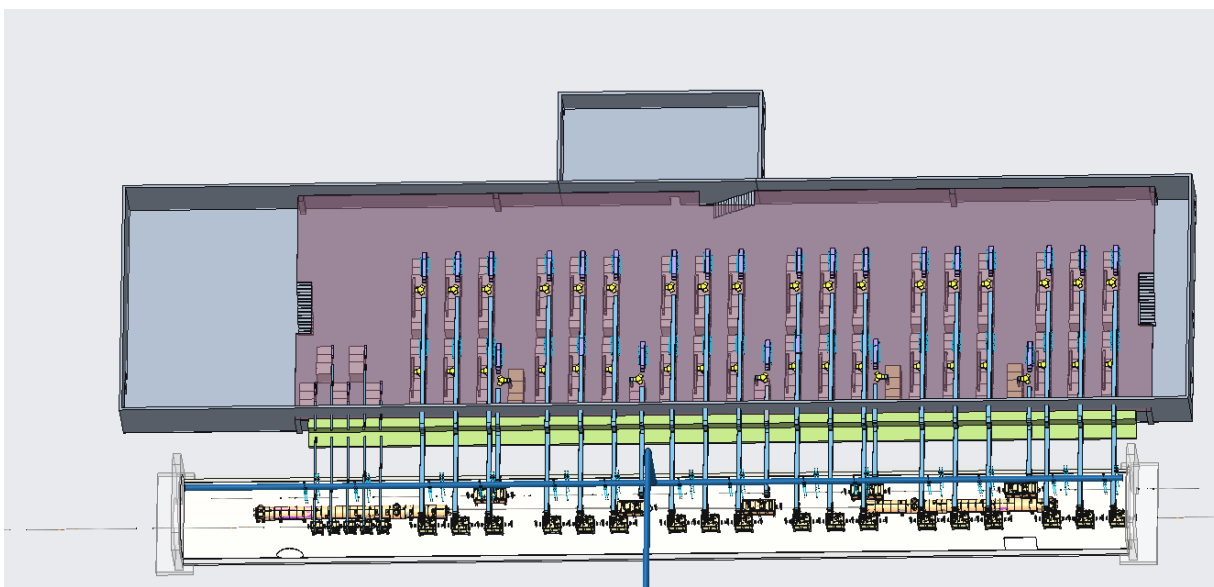
High energy beamlines require a lot of electrical power off the electrical grid to maintain EIC operations. This power will come from a dedicated transformer field outside of the 1010 SRF support facility building. The electron storage ring (ESR) alone needs 7.2 MVA of power. However, this number does not account for the additional four SRF cavities for the hadron storage ring (HSR) and the rapid cycling synchrotron (RCS), the electrical needs of the building (HVAC, lights, controls electronics, etc.), or the inefficiencies of the RF equipment. Altogether, the transformer field will provide 30 MVA of power into the building. At times, power may need to be cut off from a PA for repairs or maintenance. To do this safely, 480 V power contactors and disconnect switches will be distributed throughout the building for each PA. The disconnect switches protect workers from direct exposure to potential arc flashes. Cable trays are also needed to distribute the 480 V 3-phase electricity to the PA's. Three 535 mcm cables will be carrying nearly 1000 amps of current into each ESR PA.

To determine the layout of the RF system, water cooling pipes, electrical system, and infrastructure, Creo Parametric was utilized to create a detailed 3D model of Building 1010. 3D modeling allows for the space limitations of the project to be precisely defined.

## II. Methods

To create a 3D model of Building 1010 and the associated IR, Creo Parametric was employed. A 3D model of this facility already existed from a previous summer's project. Figure 2 shows the Creo model that was used as a reference for this project. During the interval between summers, the design of key equipment and the infrastructure layout was changed, although the model did not reflect this.

The outdated Creo model of Building 1010 did not correctly reference the origin point or the coordinate system that is a standard in the EIC. Beginning with a new, properly referenced file, the layout of equipment in the IR was first determined. Using drawings from a third-party design firm, the building base was laid into the model. The locations of the SRF cavities in the IR determined the location of the PA's and other RF equipment within the building. Having these major components in the model, the electrical system was laid in next, followed by the water-cooling pipes. All of these steps were done with guidance from engineers, physicists, and fellow interns.

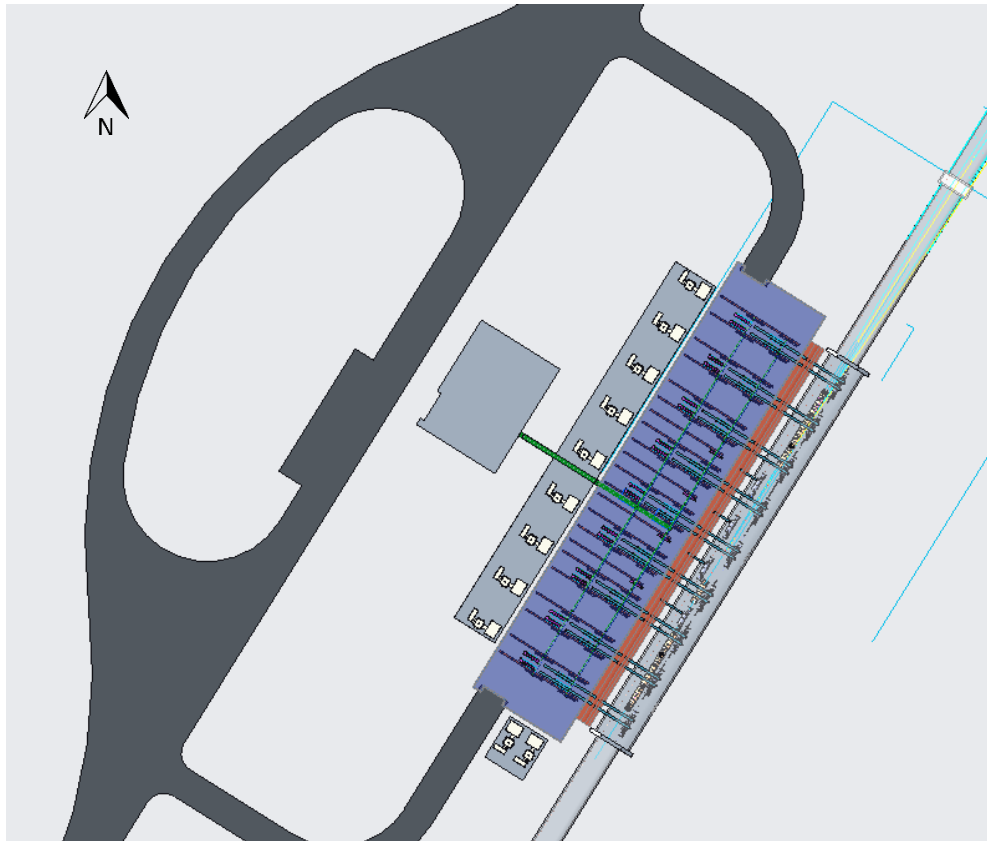


**Figure 2.** Final design of Building 1010 (gray) from previous summer's internship. Many components are outdated along with the floor plan of the building.

### III. Design Development

#### a. Overview

Figure 3 shows the entire Creo model of the 10 o'clock region of the EIC. It includes the RF, water, and electrical systems and a service road for construction and maintenance access.

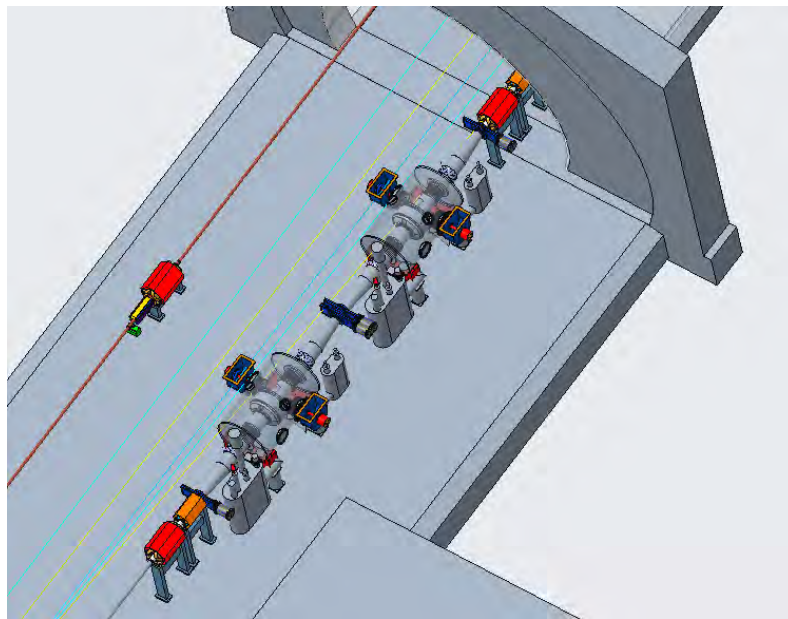


**Figure 3.** 10 o'clock region of EIC. Building 1010 is in purple with a transformer field and water pump room on the northeast side and IR10 on the southwest side.

#### b. Interaction Region RF System and Magnet Layout

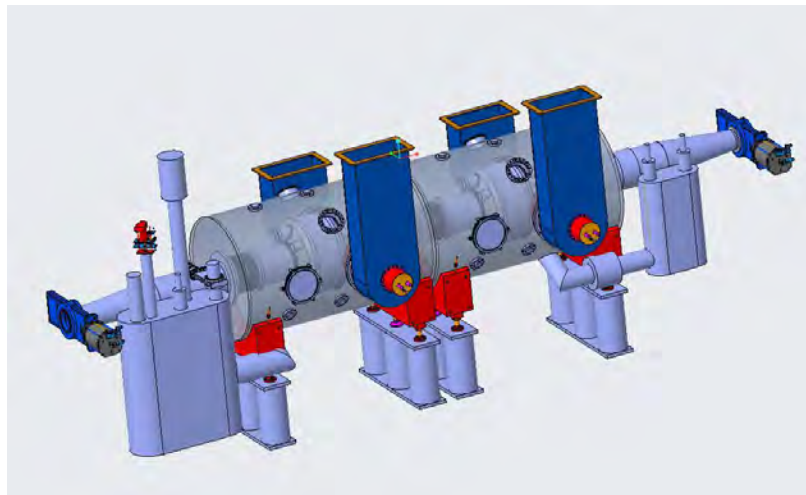
The layout of the IR at 10 o'clock was a key driver of the layout of Building 1010. The first challenge was to model the most current version of the ESR lattice, Version 5.4. The lattice determined the precise location and rotation of all the beam

line focusing magnets, cavities, and other equipment throughout the entire ring. At the 10 o'clock region, quadrupole and sextuple magnet assemblies were spaced out throughout a large diameter 90 m space in the tunnel that has the size needed for the SRF cryomodules, the RF wave guides, and the large cryogenic helium supply line. A major issue with Version 5.4 was that 18 of the ESR SRF cryomodule assemblies each with one single cell RF cavity would not be able to fit into the limited space between the magnets. However, even when the distance between centers of quad magnets was set to a constant 10 m, the 18 SRF cavities still did not fit in the space. Figure 4 shows how tight the fit is between the quadrupoles. The layout does not allow for any additional space which is required for vacuum equipment, beam tube diameter transitions, and beam instrumentation along the beamline.



**Figure 4.** North-east end of IR with two ESR SRF cavities placed between quadrupole-sextuple (orange - red) magnets that are 10 m apart. The southmost end of the cavities overlaps with the sextuple magnet.

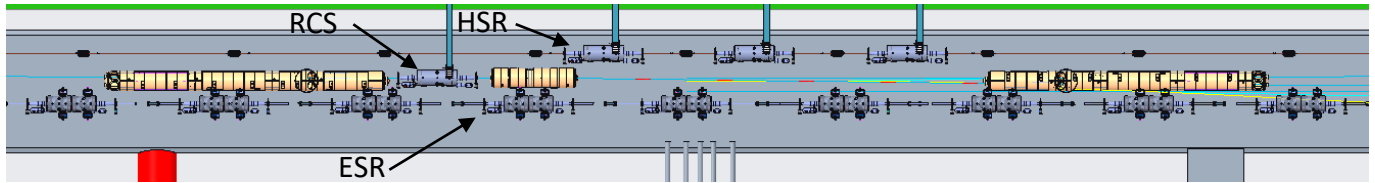
The ESR SRF cavities in Figure 4 are single cell cavities with a cylindrical cryomodule from Jefferson Lab. For a niobium cavity to be superconducting, it is placed into a cryomodule with cooling liquid helium that brings the temperatures down to 2 K. A potential solution to the space limitations in the ESR is to place two single-celled cavities into one cryomodule. This would effectively remove a higher order mode (HOM) absorber and conical transition that prevent high order frequencies from passing into the next cavity. Although this may be a viable solution, it needs to be studied in detail to determine if the lattice can accommodate high order frequencies that needs different aperture sizes. A physical representation of the new design is shown in Figure 5.



**Figure 5.** Solution to spacing issue in IR10. Two single cell cavities are placed into one cryomodule.

Creo models of the RCS and HSR lattices were previously developed, making the process of placing these lattices into the IR much simpler. Both of these rings use a 5-cell, 75 kW cavity that only has one waveguide port, unlike the ESR cavities which have two waveguide ports. The HSR lattice also has two large triplet magnets, enclosed in 4.2' diameter cryostats. All of the SRF cavities need clearance on the sides so that the fundamental power couplers (FPC) can slide into the cavity as needed. FPC's transmit RF

power from the waveguides into the cavities. Since there are only three HSR cavities, these can be distributed in the IR to avoid interfering with the cryostat. Figure 6 shows the layout of the cavities in each of the rings. The large gap between the DX and triplet magnets easily allows 3 HSR cavities to fit in the space provided.

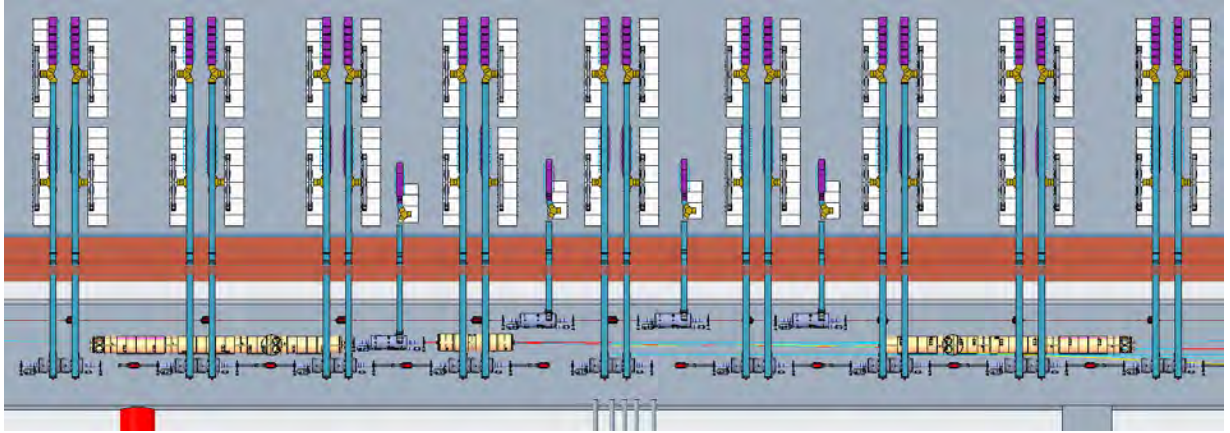


**Figure 6.** Plan view of IR10 showing ESR, HSR, and RCS lattice layouts. The yellow tube-like assemblies in the RCS are triplet and DX magnets.

### c. Building 1010 RF System

During the previous summer, the general layout of the PA's shown in Figure 7 was developed to optimize equipment performance. However, the spacing between PA's changed as the design of the cavities, and subsequently the location, had to be updated. There are 18 single celled ESR cavities, requiring 36 400 kW PA's. Each group of four 400 kW PA's is connected to the cavities housed in a singular cryomodule. One concern was that the aisle widths between PA's would not meet the 42" minimum clearance required by building codes. However, the narrow aisles within a group of four PA's are 102" wide and the wider aisles between groups are 180" wide. The wider aisles even allowed for the smaller, 75 kW PA's to be placed between groups, with the narrowest aisle at a width of 50". The wider aisles are ideal for storing the additional racks of supporting equipment, although they are not modeled here.



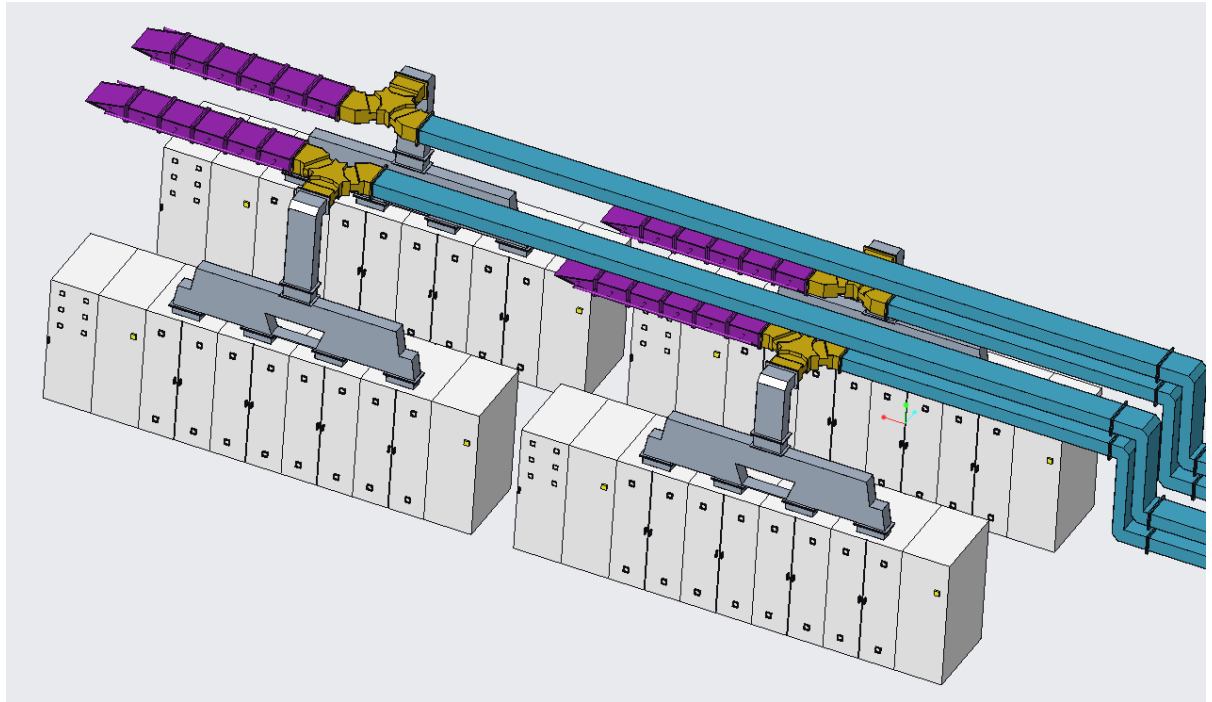


**Figure 7.** Plan view of RF system layout in Building 1010. PA's are the white components, dummy loads are the purple components, and the circulators are the small yellow components connected to the dummy loads.

Part of the equipment performance optimization meant minimizing the waveguide run length. This layout has a maximum waveguide run length of 82', which is an 8% decrease from previous layouts. Bends in waveguides also contribute to inefficiencies and require concrete radiation shielding to protect workers from being in direct radiation shine paths. By aligning cavities and PA's, additional bends in the horizontal direction in Figure 6 were avoided. Concrete shielding is still necessary, although it only needs to block radiation from entering the building due to a dog leg in the waveguide shown on the far right of Figure 8.

Within the group of four PA's the circulators and dummy loads could not be at uniform heights or orientations. The four waveguides on the SRF cavities shown in Figure 4 are too close to have all the circulators in the group of four PA's oriented in the same direction. Shown in Figure 8, the circulators face each other, allowing for the aisle created by the PA's to be wide enough to meet the minimum clearance requirement. Also, the waveguides overlap each other as we get closer to the IR,

requiring the waveguides, circulators, and dummy loads in the row of PA's closest to the IR to all be at a lower height than the row further away.

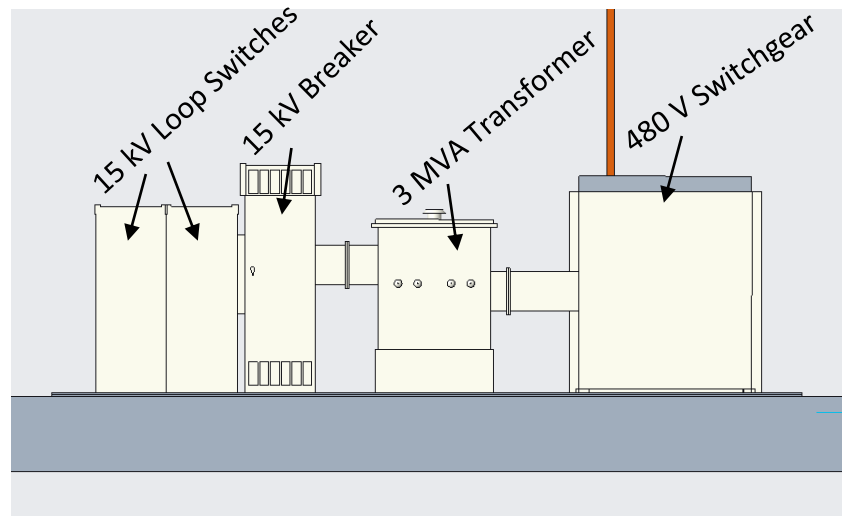


**Figure 8.** Group of four PA's connected to the cavities enclosed in the same cryomodule. 400 kW PA's are the white assemblies, dummy loads are the purple components, and the circulators are the small yellow components.

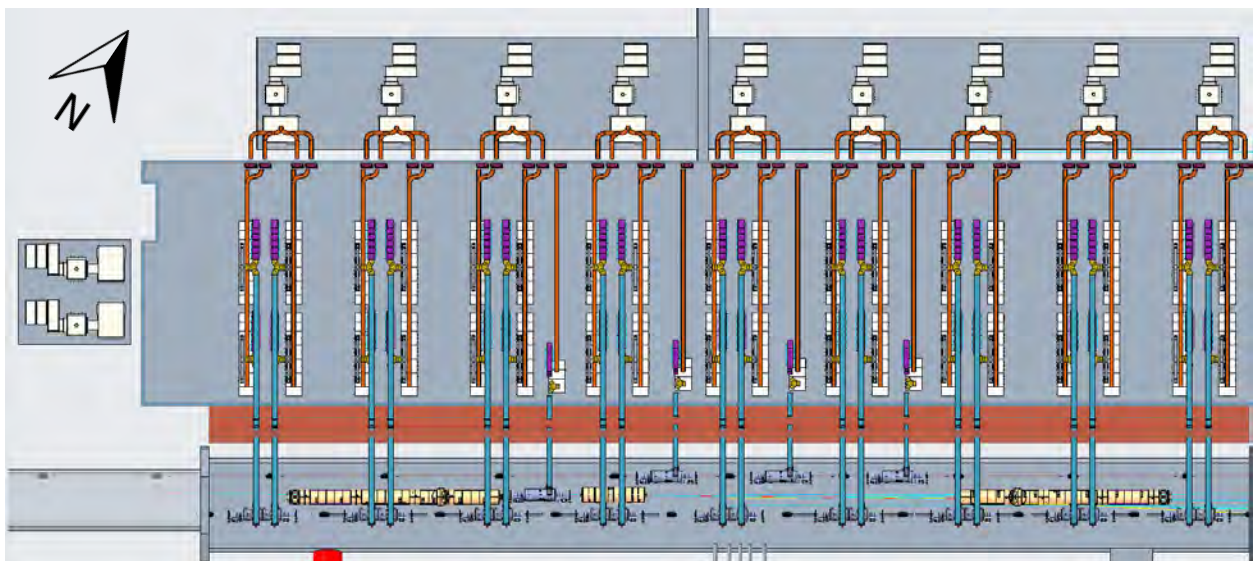
#### d. Electrical System

The electrical system is a critical part of Building 1010. A 30 MVA transformer field is needed outside of the support facility. Nine substations consisting of the components labeled in Figure 8 will be placed along the long end of the building. Each of these provides the electrical power to the four ESR PA's in each group. Each of these PA requires 800 kW to produce the 400 kW of RF power needed to maintain the high energy beams. The four 75 kW PA's for the HSR and RCS cavities and the rest of the building equipment will be powered by a substation on the southwest side

of the building. A spare transformer will be stored next to this substation. Figure 10 illustrates the locations of all the substations.



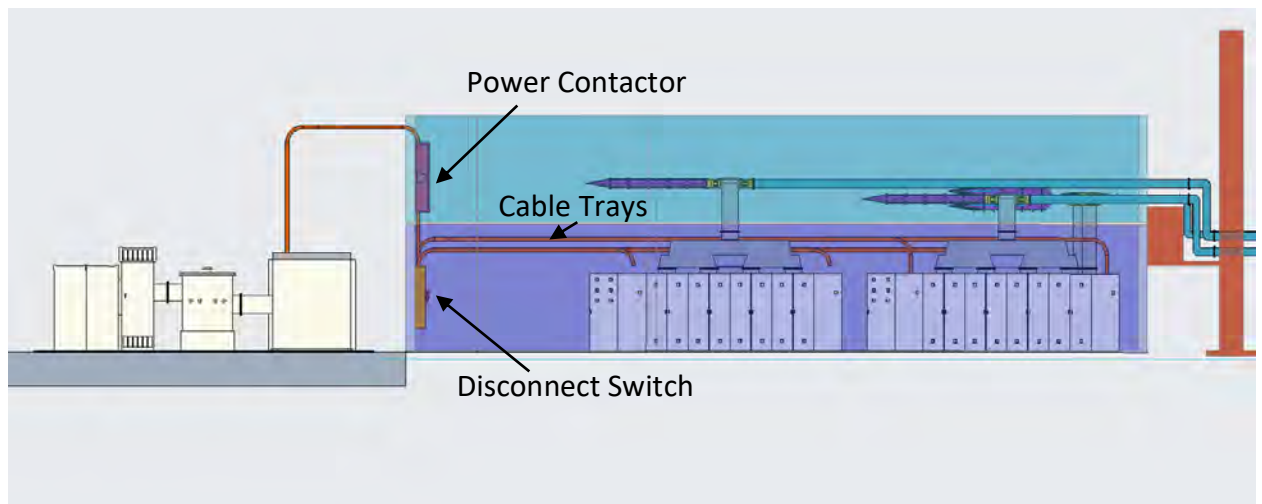
**Figure 9.** Substation outside of Building 1010.



**Figure 10.** Transformer field layout outside of Building 1010. Nine substations feed 36 400 kW PA's. A substation for the rest of the buildings electrical requirements and a spare is at the southwest end of the building.

The disconnect switches and the power contactors are very large devices that need to be mounted on the wall. The power contactor was placed on the wall on the second floor as shown in Figure 11. Except during installation and repairs, easy access is not

necessarily required for the power contactor. The second floor is mostly a utility floor, holding the circulator and dummy loads. The disconnect switch was also placed on the wall, below the power contactor on the first floor. It needed to be at a comfortable height for a person to be able to throw the switch when working on a PA. Cable tray carrying the 535 mcm cable connect the disconnect switch to the two power racks in the PA. Cable trays also connect the substation to the power contactor.



**Figure 11.** Side view of Building 1010 with some electrical components labeled.

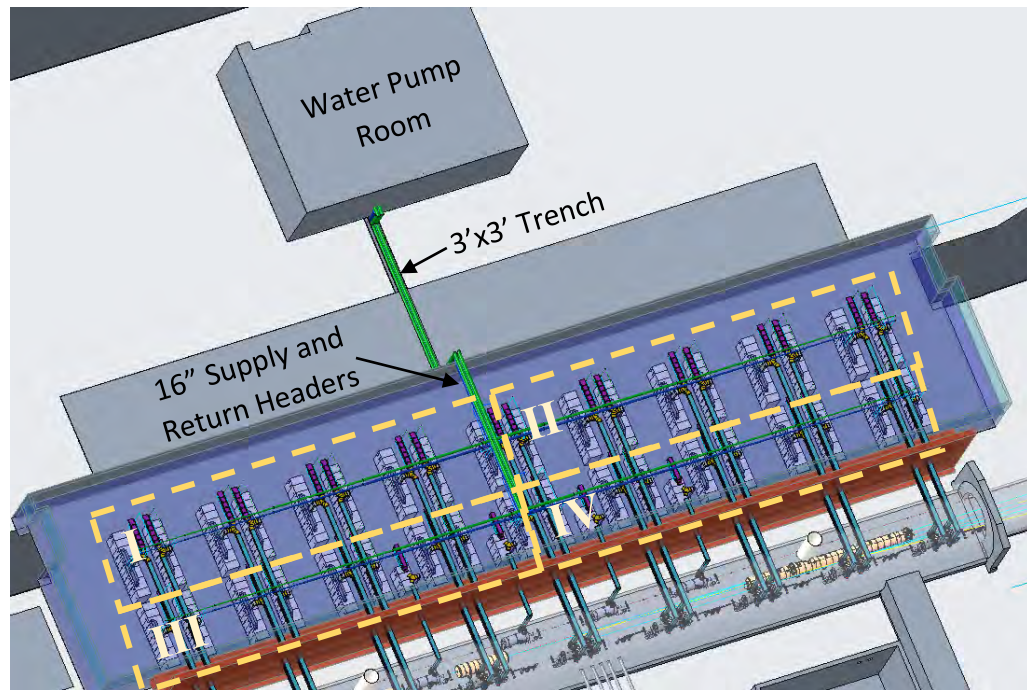
#### e. Water Cooling System

The RF system in Building 1010 produces 19 MW of heat. According to the State of New Jersey Energy Master Plan, a home will use only 2 to 4 kW on average at given moment in time. Since this heat load is so large, each RF component has to be water cooled.

The first component that was laid in was the water pump room. However, modeling the pumps and other equipment in the room was beyond the scope of this project so only a shell of a building is shown in Figure 12. An optimized piping

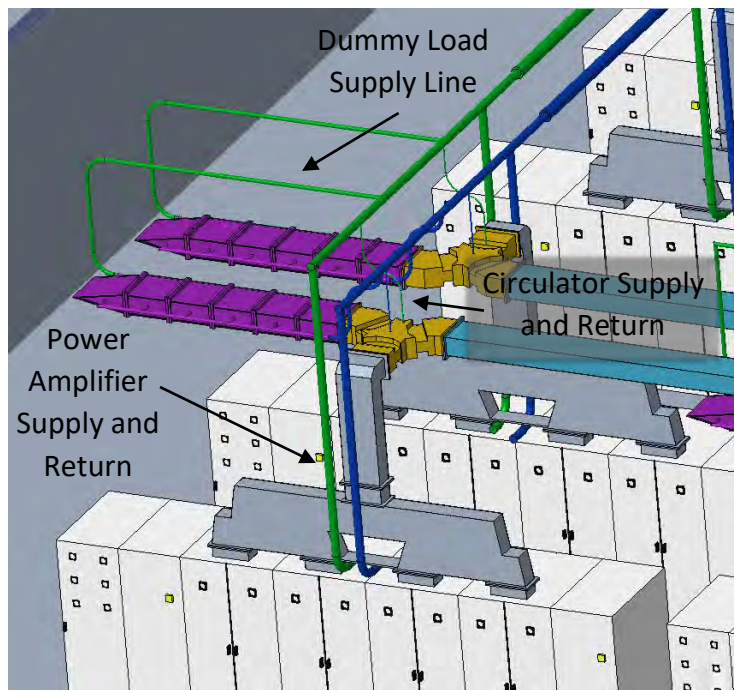


layout that maintains a velocity between 5 and 10 ft/s was developed by a fellow intern, Thea Kumar. The water is supplied at 30°C and returns at a maximum of 45°C. Two supply and two return headers, bring water between Building 1010 and the water pump room. 9138 GPM of water is supplied through these 16” pipes.



**Figure 12.** Water cooling system in Building 1010. Each quadrant represents the RF equipment supplied by one supply line. Note: Electrical system has been hidden from view for clarity.

Each 16” header feeds two supply lines that decrease from 12” to 4” in diameter as it gets further away from the header. Because there are two supply headers, there are a total of four supply lines which provide the water to the four different quadrants shown in Figure 12. From the supply lines, 4” lines connect to the PA, 2” lines connect to the dummy loads, and ¾” lines connect to the circulator. Return pipes of the same size return the water back into the water pump room. Figure 13 shows a close up of the return and supply pipes in Quadrant I (see Figure 12 for quadrants).

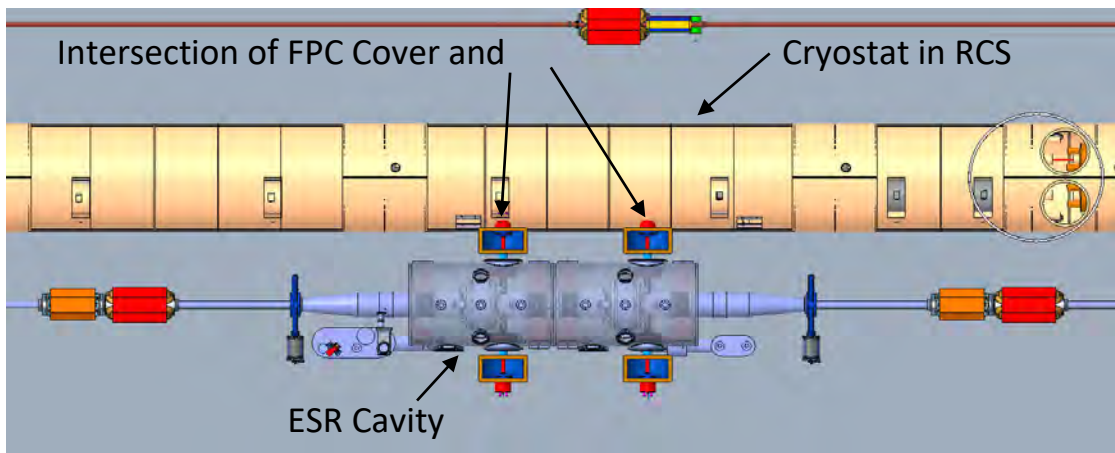


**Figure 13.** Close-up of water-cooling lines. Blue represents return pipes and green represents supply lines.

#### IV. Future Work and Conclusion

Throughout the modeling process, there were several issues discovered that require redesign of some components. Along with these discoveries, there are many design decisions that need to be made down the road. As a result, this model is a working model, representing the current state of the project but is subject to change.

One of the first issues that 3D modeling uncovered was the ESR lattice layout in the ten o'clock IR. To accommodate all of the SFR cavities, the quadruple and sextuple magnets may need to be shifted from their original location in Version 5.4 of the ESR lattice. However, magnet locations affect the properties of the beam, so a deeper analysis needs to be done by physicists. Already discussed in Section 3.b, the entire ESR cryomodule and cavity needs to be redesigned so that it fits in the IR. The current solution is to combine two single-celled cavities into one cryomodule, but this is being studied to see if it is a possibility. The ESR cavities pose another issue where they are next to the cryostats in the RCS. In the cavity, the FPC's are normal to the cylindrical body of the cavity and have a red cover (see Figure 14). Not only is there is not enough clearance for the FPC's to slide out, but the FPC covers intersect with the cryostat.



**Figure 5.** Close-up plan view of IR10 showing where the FPC cover of the ESR cavity intersects with the cryostat.

There are several systems and components of systems that are missing from this model. The water-cooling system is incomplete within the building. Only the water pipes are laid in, but they are rudimentary placeholders. Valves, flanges, and meters are not shown in this initial model. Additionally, the water pump room needs to be laid out so any issues within that building can be addressed too. The building infrastructure is not fully developed in the model either. The location of beams, the size of beams, and many other components are yet to be specified. Equipment like instrumentation cable trays, HVAC systems, tool racks, etc. needs to be modeled in to accurately convey spacing limitations in the building.

Creating a 3D model of the SRF facility at the ten o'clock IR of the EIC provided valuable insight into what challenges need to be addressed. Doing detailed design work in the early stages of the EIC project can prevent costly changes down the road. The EIC is an expensive and challenging endeavor, yet it has the potential to unlock the secrets of the forces that binds the entire universe together.



## V. References

“How RF Circulator Works.” *RF Wireless World*, RF Wireless World, 2012, [www.rfwireless-world.com/Terminology/How-RF-Circulator-Works.html](http://www.rfwireless-world.com/Terminology/How-RF-Circulator-Works.html).

“Frequently Asked Questions About Energy.” *State of New Jersey Energy Master Plan*, State of New Jersey, 2021, [www.nj.gov/emp/energy/faq.html](http://www.nj.gov/emp/energy/faq.html).

## VI. Acknowledgements

This project would not have been possible without all those I worked at Brookhaven National Laboratory with including my mentor, BNL engineers, and my fellow interns who are listed below:

Mentor:

- Joesph Tuozzolo, EIC Technical Systems Division Director

EIC Engineers:

- Douglas Holmes, EIC Mechanical Engineering Group Leader
- Kevin Smith, EIC Technical Systems Division Deputy Director
- Alex Zaltsman, EIC RF Group Leader
- Ram Srinivasan, EIC Cooling & Mechanical Systems Group Leader
- Tom Nehring, EIC Electrical Power Systems Group Leader

Interns:

- Angela Zavala Yanes, CCI Intern
- Thea Vijaya Kumar, SULI Intern

I would also like to thank the Supplemental Undergraduate Research Program and the Office of Education Programs for funding and providing support throughout the duration of my internship.

**Advancing computer aided design skillset to grow professionally and contribute to the National Synchrotron Light Source-II mechanical design team**

Matthew Warner, CCI Intern, Department of Photon Sciences, Suffolk County  
Community College, Selden, NY, 11738

Greg Fries, Program Mentor, Department of Photon Sciences, Brookhaven National  
Laboratory, Upton, NY, 11973

**Abstract:**

The National Synchrotron Light Source II (NSLS-II) is comprised of a linear accelerator (LINAC), booster ring, electron storage ring, and currently twenty-nine beamlines for nanoscale resolution x-rays. The electron ring consists of numerous components, instrumentation, and utilities each requiring regular maintenance, repair, or replacement. Many of these items are manufactured onsite through traditional fabrication or 3D printing, and these must be designed to ensure maximum efficiency. The design team at NSLS-II utilizes Autodesk's Inventor program to facilitate the design of required components. My role within this internship session has been familiarization with Inventor and assisting with the design of simple components, to learn ways of incorporating interns and new employees for future endeavors. To acclimate to Inventor, I utilized educational sources from the internet including readings, videos, and in-program tutorials. Additionally, I met with lead designers of the NSLS-II team to discuss usage and application of the program within the facility. To assist with learning and begin testing application of the new knowledge, I was assigned the design of an adaption to the spring component of a vertical stage stack assembly utilized on the HXN-beamline. In application, I have found that the knowledge gained from self-education methods of adapting to the new design interface to be challenging, but with one-on-one sessions have made strong progress towards contributing to the design team. Over-all it appears that instructional learning from an established design team member greatly outweighs that of independent study. Personally, learning a new design software has allowed me the flexibility to advance beyond the structural and aesthetic usage of computer aided design and incorporated mechanical aspects, which expands my professional skillset.

The National Synchrotron Light Source II is one of the most advanced synchrotron facilities in the world. A synchrotron is a large machine that accelerates electrons to near lightspeed through a storage ring utilizing magnets to control and steer the beam. As the beam deflects through magnetic fields it produces x-rays which are channeled through beamlines for basic and applied research for scientists from around the world.

The NSLS-II is utilized “in biology and medicine, materials and chemical sciences, geosciences, and nanoscience” [1]. The facility has been and will continue to be key for evolving medicines, including work done to create the Pfizer vaccine for COVID-19 [2], battling climate change via green energy techniques, advance energy storage, and utilized to research / develop many other areas of industry [1].

The NSLS-II Storage ring utilizes a system of over 900 magnets, vacuum systems / components, advanced instrumentation, and an array of mechanical utilities for high beam stability and minimal electron emittance, while maintaining electron speed close to the speed of light [2]. Beamlines have similar components in addition to other mechanics for the individual needs of areas of research. With extremely tight control over conditions of operation (i.e. the storage ring cannot fluctuate more than .1 degree Celsius [2]), regular monitoring and maintenance is required.

The electron storage ring and various beamlines are highly complicated machinery utilizing numerous parts and assemblies to operate. On occasion portions of this machine must be repaired, replaced, and or adapted to maintain functionality. To ensure efficiency and manage cost, many of the machine's components are fabricated in house through machining or 3D printing. Due to the inflexible requirements of optimal

operation for the facility, in-house parts must be thoroughly designed and engineered to ensure that there will not be any hindrances from installation.

To produce the required components, NSLS-II carries an in-house design team who work with the requests of various engineers, technicians, and scientists, to fully plan and fabricate pieces. Most of this work is completed utilizing AutoDesk's Inventor, a highly complicated computer aided design program utilizing two- and three-dimensional modeling, which is specialized for mechanical design.

This internship is an experiment in understanding the best methods of training a new team member in the use of Inventor. I, the subject of this experiment, have certification in drafting and design through Suffolk County Community College, though no previous experience with Inventor specifically. The methods were divided into self-education and instructional learning, and each were used at the discretion of the trainee.

The internship began with an immediate setback. Due to the fact that my educational license expired following my graduation from Suffolk County Community College, I was unable to get full access to the program until early July almost halfway through the experiment window. To continue development, I utilized the month-long trial version of Inventor which allowed simple practice measures but not saving, sharing, or accessing files. To make the most of time, self-education was shifted to the initial priority, and a catalog of resources was created featuring text, video, and hands on tutorials.

Beginning education with a familiar resource, accessing textbooks was an initial priority. While there were many options available online for sale, to ensure accessibility in the experiment texts were sourced from the local library. Unfortunately within this constraint, the selection was limited to only a few options with most being written on outdated versions of the software. The best of these options was a tutorial book titled *Autodesk Inventor 2020 For Beginners: Part Modeling, Assemblies, and Drawings* written by the author group “Tutorial Books”. This book mostly featured simple lessons on the creation of mechanical parts, beginning with how to create sketches and utilize said sketches for three dimensional tools such as Extrude, Revolve, and Sweep. Most of the provided lessons fell within these categories though it did have some instruction on the utilization of these created parts for patterning and assemblies.

The largest advantage to utilizing the text was a sense of familiarity that came from similar studies being done with previous computer aided design programs. This style of tutorial book is convenient for identifying uses, features, and commands of these programs to be utilized in general professional settings. Overall it instills a foundation for furthering individual skill with the program. With these three attributes in mind text is considered a strong source for beginning learning to a self-starting individual.

There are however disadvantages to utilization of text for self-education of Autodesk’s Inventor program. The format carries limitations unless a collection of books are purchased and the financial accessibility is concerning. Additionally, the specific tutorial book, and others of this nature, are unadaptable to NSLS-II’s advanced needs without oversight. As with all new learning there are set backs and confusion and the

text alone does not benefit with trouble shooting errors. Overall, I personally found the text rather ineffective without instruction.

Upon the realization that instruction was preferred but still falling within self-education I turned to video instruction, through YouTube lessons. The idea of YouTube's effectiveness as an educational resource has gained traction in recent years, especially in the virtual based era of COVID-19. To build upon lessons already learned, I found videos based on commands taught in the text and then found subsequent tutorials. There are two primary ways of navigating YouTube, which are finding user channels and surveying their collection or utilizing key word searches. Skimming collections and individual videos for knowledge was a truly time-consuming endeavor and the key word search was far more impactful. Due to this, the video resource was far greater with supplemental learning, and best paired with the other educational methods rather than as a primary source of knowledge.

Though limited, YouTube's catalogs were beneficial in learning the program. The primary advantage was commentary and instruction provided by those working within industry. Being familiar with the website allows for easy navigation, removing hindrances of learning outside of the desired knowledge. The greatest attribute of YouTube is the massive instructional library, featuring lessons in essentially every aspect of the program.

While remaining a useful tool, YouTube does have known drawbacks as an educational vessel. Concerning Inventor, the database has a lot of redundant information and locating quality sources can be a struggle. Navigating user collections did not prove fruitful, and the keyword search requires preliminary knowledge of the

program and which features are to be utilized. In general video-based learning is an extremely time-consuming endeavor, and inefficient when switching between instruction and practice. The largest drawback to utilizing YouTube for self-education is the well known “YouTube Hole” of following diverging videos and getting further and further distracted from tasks at hand. Overall video tutorials are a quality resource to supplement other forms but would not be recommended as a primary method.

Within the Inventor program there are a series of user made tutorials which review beginner, intermediate, and advanced skills. These tutorials walk the user through various challenges, step by step, familiarizing one with an assortment of commands and applications. The tutorials include video of each step, highlight necessary items in the browser and toolbar, and feature premade components. As far as self-education this feature was of great benefit to myself and was by far the largest contributor to my progress of the utilized methods.

There were a few key elements that made these tutorials so valuable. Most important was the hands-on aspect of this learning method. While text and video were helpful it is difficult to understate the impact of using the program directly to learn and familiarize oneself with it. The tutorials are largely based on required key features, and with premade components efficiently focused on such. When following through written steps, accompanied with video, the toolbar commands were highlighted and built familiarity with ease.

However even the best tools have limitations. The scope of the tutorials was limited, and I was not able to find items available for all of different features identified to be useful. While the thoroughness was largely beneficial it did hamper the ability to



experiment or adapt. Overall, the largest drawback to tutorials is that after each step, the users work was discarded, and the design was reset. This prevented the ability to have compounded mistakes that would need to be worked through and limited the mental development of usage.

The self-education portion of this developmental experiment featured many trials and tribulations. Being limited to the trial version of the program for the beginning portion, forced the hand but did not dampen results overall. While resources have been identified for further development in the future, the over-all concept of single handedly learning a program ended up being an unrealistic challenge for me in the limited time frame.

This was particularly identified when forwarded an assignment from Brian Mullany, design engineer and supervisor for NSLS-II, which originated amongst beamline scientists. The task at hand was to develop an adaption to the vertical stage utilized on the HXN beamline, one of multiple stages with the primary function of height adjustments for the sample holder, as it was employed upside down and had not been designed to be. The primary issue with the inverted installment was that the constant force spring did not work properly, designed to push weight up but was now pulling weight up, and would need to be reversed and reattached for optimal function. To do so properly, we had to design plates to attach to the vertical stage which would hold the now inverted spring. Attempting to decipher the project and develop the components on my own had minimal progress and required teaching sessions one on one for success.

Mr. Mullany led multiple work sessions utilizing Microsoft Teams Screen Share feature, and together we developed the required components for the vertical stage

stack. Cutting out time consuming methods, we utilized the projected geometry feature on existing faces to create sketches. From these sketches the plate adapters were extruded, and holes were placed for proper attachment. If we were to continue this project, proper dimensioning would be required and then the new assembly could be constructed.

Beyond an understanding of the task at hand there were numerous benefits from these teaching sessions. Like the tutorials, all learning was completed in a hands on method. The sessions themselves were both educational and helpful to the team as we were working on an actual prescribed task of the design team. Mr. Mullany's knowledge of the program highlighted features commonly used professionally particularly by NSLS-II's design team. Outside of the educational and functional portions was the fact that this time was one of the best networking opportunities provided to me in this internship. The only negatives found in the teaching sessions were that they required more of Mr. Mullany's time, and on occasion there were scheduling issues.

While conclusions can be made about this venture, there are also points of discussion that must be brought to attention. For preliminary purpose, what level of knowledge is required to be considered self-sufficient? What is an acceptable time table that the development of skills is considered efficient? In what ways were the results affected by the virtual work environment? Did the time taken away from a senior employee out-weigh the benefit of the teaching sessions? Lastly, and most important for further use, would results vary with other individuals?

Overall, this experiment in professional development has led to multiple conclusions beneficial for future use at the lab. Primarily it has been determined that

hands on teaching session from the design team and the Autodesk inventor tutorials were the most constructive and helpful. A small catalog of helpful learning resources has been created. The hindrances of self-education seemingly outweigh the benefits, and working with the design team would be best for future employees and interns to learn the program. Additionally, I personally gained considerable knowledge of Autodesk's inventor, to the point drafting Vertical Stage adaptation utilizing the program.

## ACKNOWLEDGEMENTS

Community College Internships (CCI) This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program (CCI)

Big thanks my mentor, Greg Fries, for his professionalism, inclusiveness, and kindness during the CCI program. I'd like to proclaim my gratefulness of Brian Mullany for his time and effort spent to assisting with this experiment and building my personal skillset. Additionally, my sincerest gratitude to the Office of Education Programs for assisting myself and all interns along this path of personal development. Finally, I would like announce my appreciation of Brookhaven National Laboratory and the Department of Energy for this fantastic opportunity.

References:

- 1- “About National Synchrotron Light Source II.” *BNL Photon Sciences* | *About NSLS-II*, Photon Science Division, [www.bnl.gov/ps/nsls2/about-NSLS-II.php](http://www.bnl.gov/ps/nsls2/about-NSLS-II.php).
- 2- Ongoing discussion and various presentations, G. Fries, Deputy Director for Projects, NSLS-II Accelerator Division, Brookhaven National Laboratory, Internship Mentor

Environmental safety and health- business and environment

Miniya Williams, School of Business & Management, Morgan State University, Baltimore, MD  
21251

Joy Haskins, Environmental Health and Safety Department, Brookhaven National Laboratory,  
Upton, NY 11973

Steven Coleman, Environmental Health and Safety Department, Brookhaven National  
Laboratory, Upton, NY 11973

## Abstract

The Environmental Protection Division is concerned with the future energy supply of the nation. Some of the ways they work on this are by reducing energy consumption, buying green products, and improving health safety. During my internship with Brookhaven National Laboratory (BNL), I expanded my knowledge on ultra-low temp freezers and fume hoods that are in the research laboratories. Most labs consume an average of 5 to 10 times more energy than commercial buildings because of these big energy consumers. A lot of that energy is wasted due to poor maintenance and improper ventilation systems. Many of the laboratories at BNL have fume hoods and freezers that stay on constantly and can draw as much energy as an average domestic household. The improper use of energy can be expensive and harmful to the labs. The research was done by contacting administration and collecting inventory on the big energy consumers and seeing which of those don't fit the energy efficiency criteria. I created a cost of benefit analysis, which will show how cost can be efficiently utilized. I gathered tips for improvement from the SmartLabs program and the Annual Conference of the International Institute for Sustainable Laboratories. I also was able to use the SFTool product search, Staples, and Office Max to search for environmentally safe products. The purpose of this research has been to examine energy efficiency, environmental safety, and the improvements that should be made. As a result of this research, I added the importance of energy efficiency to my repertoire of knowledge. I have become familiar with the Environmental Preferable Purchasing program (EPP), sustainability, and ventilation systems.

## I. Introduction

### A. Freezers and fume hoods energy use

In many, research laboratories the ultra-low temp freezers (ULT) and fume hoods are the largest users of energy. In 2015, the CEEL report stated that the United States has at least 580,000 ULT freezers, which consume 4 billion kWh/year. The ultra-low temp freezers' temperatures can range between -80°C and -86°C. Samples of DNA, RNA, bacteria, viruses, and cells can all be stored in these freezers. Many institutions and researchers are suggesting raising the temperature to -70°C, to save energy without compromising the samples. Energy Star® is a program run by the EPA that supports energy efficiency by providing certified products that can save laboratories money and energy. Another big energy consumer is Fume Hoods, which are high-power fans that help with collecting and removing vapors, dust, gases, and fumes out of the laboratories exhaust system. The air that is being sucked through the fume hood is what consumes so much energy, as well as the electricity that is being used to power the large fans inside of the fume hoods. When the fume hoods are not in use shutting the sash and turning off the occupancy switch can help with reducing the energy consumption.

### B. Environmentally Preferable Purchasing

The Environmental Preferable Purchasing program (EPP) encourages use of products from recycled materials or lower impact compared to environmentally friendly products. The products can be recycled material, bio-based, non-ozone depleting, or non-toxic. EPP aids in pollution prevention efforts by reducing greenhouse gas and hazardous waste generation and this can help improve the health of communities. Product searches can be time consuming which drives low participation in the program. Administrative staff are tasked with finding products on top of their current workloads, and request for office products are made 2-3 x a week.

## II. My contributions

With this project, I focused on ways to improve energy consumption and green purchasing at BNL. I attended different webinars and read research reports to expand my knowledge on the benefits of energy-efficient labs. I was able to look at BNL's Annual Internal Assessment of the Environmental Management System and Compliance Assessment of Sustainability Requirements for 2021 and I found that there can be improvements for sustainability purchasing. I made notes of all the suggestions for improvements that I found and shared them with my mentor. My mentor told me that a lot of the freezers and fume hoods at BNL use up so much energy because they are old. I created a cost of benefits analysis, that contrasts both the older and newer fume hoods and freezers. I reviewed energy use data for multiple different buildings to compare newer more efficient equipment. Due to lack of time, I was unable to gather all factors needed to make an accurate comparison and some of the information couldn't be found.

I used the SFTool product search to look up non-paper office equipment that follows the EPP guidelines. Staples and Office Max are approved vendors and sites that they order from. I was also told that it is hard for the staff at BNL to find non-paper office products that fit the criteria of EPP. So, my mentor asked me to document how long it took me to find products on SFTool. In an hour and thirty-three minutes, I was able to find pens, markers, highlighters, building furniture, organizers, glue, and tape. I couldn't find batteries, clipboards, binders, or correction fluid. In the BNL stock, I found that some items need to be reordered. I created steps for this process for the staff to follow, that will be included in an operator aid. Sometimes when they do find environmentally friendly products, they can be overpriced that they continue to purchase the products that don't fit the criteria. I found file folders on Office Max that are environmentally



friendly, and it was a pack of 100 for \$18.99. At Staples, I found similar folders that were environmentally friendly, and it was a pack of 50 for \$50.99.

### III. Recommendations and Results

After looking at the electrical consumption data for buildings 363,555, and 734 for 2020 and 2021, pictures of the “Freezer Farm”, and doing a virtual tour of one of the labs. I determined that the reasoning a lot of these freezers are consuming so much energy is because they haven’t been defrosted, the temperatures can be raised, and/or they need to be upgraded. Some of these freezers have been sitting around for so long that defrosting them would consume too much energy. When new freezers are purchased monthly defrosts would help increase the lifespan and preserve the quality of these freezers. I also noticed a range of different temperatures on multiple freezers. The set temperature for all ULT freezers should be between  $-70^{\circ}\text{C}$  and  $-75^{\circ}\text{C}$  instead of  $-80^{\circ}\text{C}$ . Each freezer should be evaluated to determine raising the temperature and how it will impact the material storage. Upgrading can be expensive, but it would be cost-effective. Upgrading to Energy Star® freezers that are more energy-efficient can help with saving at least 70% of energy.

When reviewing the energy consumption for Bldg. 743 I noticed that the energy use was a lot higher than Bldg. 363 and 555. Bldg. 743 is a newer building and most of the equipment should be Energy Star® or equivalent. My mentor told me that this could be because of the number of occupants in the building and/or there are more fume hoods and other equipment in use compared to the older buildings. So, another recommendation is that the number of big energy consumers in one building should be limited.

Shutting the sashes on fume hoods when not in use can help with reducing the airflow, which lowers energy use. Also putting them on standby mode and/or making sure the occupancy switches are turned off when not in use. With variable air volume fume hoods, the building's fan speed and the volume of air can be reduced when the sash is lowered or shut. The older fume hoods don't have automatic sashes or sashes that can be shut, so they use up a lot of energy. These fume hoods should be replaced with fume hoods that have automatic closure systems. Also, fume hoods shouldn't be used just to store chemicals or materials. At least 70% percent of energy can be reduced if these recommendations are followed.

Better participation for the EPP program can be for the staff to create an operator aid to make the process easier and creating a database of commonly purchased products that will make ordering easier. There are not a lot of vendors or companies that have products that fit the EPP criteria, so the process can be time-consuming, but the results are beneficial. Different barriers can make it hard to purchase these products, either from administration or being skeptical of the products either because they are too expensive or not knowing if they follow the criteria.

#### IV. Conclusion

The research laboratories at BNL have the potential to be cost-effective and energy efficient if they follow the recommendations I have suggested. In the future, there should be more information collected on the energy use of individual equipment. This will help with identifying which big energy consumers should be replaced. Participation with the EPP program can increase if tools are related for the staff to easily access EPP products. As a result of this summer, I learned about topics that are outside of my field of study, but I was still able to

incorporate my business skills into this research project. I leave this internship with a better understanding of energy efficiency and the purchasing of environmentally safe products.

V. Appendix A

<b>Fume Hoods</b>	<b>Without automatic closure system</b>	<b>With automic closure system</b>
<b>Company</b>	Kewaunee MYG.C.O.	Labconco
<b>Purchase of the machine</b>	N/A	\$8,410
<b>Size and Type</b>	60"/ Ducted/Not Hipa Filtered	60"/ Ducted
<b>Hours on per day</b>	24 hours	When in use/ 10-12 per day
<b>Material</b>	Wood	Coated Steel
<b>Air Flow monitor</b>	Included	Included
<b>Sash</b>	Sash doesn't shut	Manual or Automatic
<b>Remote Fan</b>	Enclosed	Enclosed Direct Drive
<b>Cost for one year</b>	\$ 9,100.00	\$ 5,600.00

Table I: Cost of Benefits Analysis for Fume Hoods



Figure 1. Older fume hood without automatic sash



Figure 2. Upgraded/New fume hood with automatic sash

Freezers	Old Freezers	ENERGY STAR® Freezers
Company	N/A	Phcbi/VIP ECO
Hours used per day	24 hours	Average of 8 hours/ average of 6 door openings
Electricity used per day	0.67 kWh	0.29 kWh/day/cu-ft
Typical Temperature	"-68°C	"-80°C
Power consumption	5kW	7.86 Kwh/day
Average cost for one year	\$ 2,808.00	\$ 1,814.39
Average cost for five years	\$ 13,806.00	\$ 9,071.95

Table II- Cost of Benefits analysis for Ultra-low temp Freezers



Figure 3: Inside of a new freezer from the “Freezer Farm”



Figure 4: “Freezer Farm”/Upgraded Freezers



Figure 5: An example of a freezer with ice build up

## VI. Acknowledgments

First, I would like to thank both of my mentors Joy Haskins and Steven Coleman for their knowledge, patience, understanding, and generosity. I also would like to thank Brookhaven National Laboratory and the Department of Energy under the Science Undergraduate Laboratory Internships Program for allowing me the opportunity to do this internship.

## VII. References

“About the Environmentally Preferable Purchasing Program.” *EPA*, Environmental Protection Agency, [www.epa.gov/sustainable-marketplace-greener-products-and-services/about-environmentally-preferable-purchasing](http://www.epa.gov/sustainable-marketplace-greener-products-and-services/about-environmentally-preferable-purchasing).

Davis, Carla. “Saving Energy in LABS Looks like This.” *Sustainability*, 5 Oct. 2015, [sustainability.ncsu.edu/blog/changeyourstate/save-energy-in-lab/#:~:text=Ultra%2Dlow%20temperature%20freezers%20can,Purchase%20green%20lab%20equipment](http://sustainability.ncsu.edu/blog/changeyourstate/save-energy-in-lab/#:~:text=Ultra%2Dlow%20temperature%20freezers%20can,Purchase%20green%20lab%20equipment).

*Fume Hoods Provide Key to Laboratory Energy Savings*. PG&E, July 2012, [www.pge.com/includes/docs/pdfs/mybusiness/energysavingsrebates/incentivesbyindustry/biotech/fs\\_FumeHood.pdf](http://www.pge.com/includes/docs/pdfs/mybusiness/energysavingsrebates/incentivesbyindustry/biotech/fs_FumeHood.pdf).

“Purchasing Energy-Efficient LABORATORY-GRADE Refrigerators and Freezers.” *Energy.gov*, Energy Efficiency & Renewable Energy, [www.energy.gov/eere/femp/purchasing-energy-efficient-laboratory-grade-refrigerators-and-freezers](http://www.energy.gov/eere/femp/purchasing-energy-efficient-laboratory-grade-refrigerators-and-freezers).

Savage, Luke. “Reducing Fume Hood Energy Consumption.” *Labconco*, Oct. 2019, [www.labconco.com/articles/reducing-fume-hood-energy-consumption](http://www.labconco.com/articles/reducing-fume-hood-energy-consumption).

“Sftool Product Search Training for Federal Procurement & Project Teams.” *Ecomedes*, [www.ecomedes.com/sftool-product-search-training-for-federal-procurement-project-teams](http://www.ecomedes.com/sftool-product-search-training-for-federal-procurement-project-teams).

Wirdzek, Phil. “Thinking Outside the Icebox on Lab Sustainability.” *Nature News*, Nature Publishing Group, [www.nature.com/articles/d42473-018-00223-9](http://www.nature.com/articles/d42473-018-00223-9).

Convective cloud lifecycle over the Houston region using a cell tracking algorithm

Brian Wittemann, Department of Earth and Atmospheric Sciences, State University of New York at Oneonta, New York 13820

Michael Jensen, Department of Environmental and Climate Sciences, Brookhaven National Laboratory, Upton, New York 11973

Dié Wang, Department of Environmental and Climate Sciences, Brookhaven National Laboratory, Upton, New York 11973

Grace Kowalski, Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, Indiana 47907

## Abstract

The Cloud Processes Group at the Brookhaven National Laboratory (BNL) aims to progress the understanding of microphysical and dynamical processes that impact the lifecycle of clouds. Particularly, the focus is currently on the Tracking Aerosol Convection interactions Experiment (TRACER) that aims to improve the knowledge of the processes that drive the life cycle of isolated deep convective clouds. In preparation for TRACER, the objective of the internship is to develop an enhanced understanding of the variability in convective cloud properties under different synoptic and mesoscale conditions. Specifically, for my project, I applied a cell tracking algorithm to National Weather Service (NWS) Next Generation Radar (NEXRAD) observations to develop statistics of cell characteristics under different synoptic regimes. The method consists of multiple python scripts, including an open-source cell tracking algorithm, TiNT, to generate statistics of radar observed properties of isolated convective cells on possible sea and bay breeze days over the last 10 years. Each individual cell is required to have a minimum lifetime of 30 minutes and to be classified as isolated (not connected to nearby cells) at least once over that period. The synoptic regimes were combined into clusters based on similar weather patterns and a mean of the maximum radar reflectivity, cell area, maximum echo top height and cell volume were determined on a timescale normalized around the time each cell reaches its maximum reflectivity. A secondary objective refers to the forecasting preparation for the TRACER campaign. Forecasting exercises occurred daily to improve knowledge of the forces driving the isolated convective systems and aerosol conditions in the region.



## 1. Introduction

### *1.1 Tracking Aerosol Convection interaction ExpeRiment (TRACER)*

With a goal to improve the understanding of the lifecycle of deep convective clouds and their interactions with aerosols, the TRACER campaign is scheduled to take place in the Houston, Texas region from October 2021 through September 2022. To accomplish these goals observations of clouds, aerosols, precipitation, atmospheric state, and radiation will be collected through the employment of more than 50 instruments including vertically pointing and scanning radars and lidars, passive microwave radiometers, and radiosondes<sup>1</sup>. The observations collected will vary based on a range of environmental and aerosol conditions.

To inform daily field operations, meteorological forecasts are expected to take place multiple times daily during the TRACER campaign. Specifically, forecasters will rotate through positions of validator, forecaster, and shadow forecaster. The forecasts will determine the mode of operations to take place on each particular day.

### *1.2 Coastal Circulation*

Sea and bay breezes are a thermally driven circulation pattern driven by differential heating over adjacent land and water surfaces<sup>2</sup>. The development of sea breeze circulation generates a mesoscale phenomenon called a sea breeze front. A sea breeze frontal boundary extends about 100 – 150 kilometers (km), whereas a synoptic scale cold front is in the order of 1000 – 2000 km. For comparison, the empire state building is about a half of a kilometer tall. Along the Gulf of Mexico, the tendency is for a sea breeze front to be parallel to the coast with slight curvature near the bay if a bay breeze is evident (Figure 1). In the Houston region the sea and bay breezes are most prominent in the warm weather months of June, July, August, and

September. The same weather months selected for the intense operational period (IOP) of TRACER.

### *1.3 Self Organizing Maps (SOMs)*

Weather regimes are developed through the use of SOMs, a machine learning technique, that groups days based on similar synoptic scale weather patterns<sup>3</sup>. The weather regime classification is generated from 10 years of 700 hPa geopotential height anomalies during the IOP months. The outputs contain 16 synoptic regimes (nodes) with 4 primary regimes including Bermuda High, Continental Anticyclone, Pre-Trough, and Post-Trough. The 16 regimes were grouped based on similar synoptic conditions and assessed throughout each level of the atmosphere. For instance, nodes 0 and 1 are associated with a westward expansion of the Bermuda High anticyclone that is centered over the North Atlantic Ocean. A deep, isolated high pressure system forms near the coast of the Gulf of Mexico along with the migration of the Bermuda High. At the 1000 hPa and 850 hPa levels, the Houston region is placed at the western ridge of the Bermuda High with strong southerly flow. At the 500 hPa level, a clear difference in the wind flow patterns can be observed where the western edge of the Bermuda High meets a trough to the east. Thus, SOM-derived-regimes provide a broad understanding of the synoptic weather patterns in the localized region.

## **2. Data and Methods**

### *2.1 National Weather Service (NWS) Next Generation Radar (NEXRAD)*

Historical NEXRAD radar observations and statistics from 2010 through 2019 in approximate 5-minute scan intervals was provided by the NWS at site KHGX (Figure 2). Specifically, applying an open-source storm tracking algorithm (TiNT), the maximum radar

reflectivity, storm area, maximum echo top height (maximum altitude) and storm volume over the storm's lifecycle are extracted for the time frame of 16 UTC – 21 UTC<sup>4</sup>.

## *2.2 Sea Breeze Assessment Between Nodes*

An analysis of mean individual storm statistics of the 4 variables was performed for sea breeze days to build an increased understanding of isolated storm lifecycle. Sea breeze days were classified by NEXRAD observations as well as visible satellite imagery. We developed multiple thresholds to separate each individual storm into independent datasets. Each individual storm must have a lifetime longer than 30 minutes and be isolated (not connected to nearby cells) at least once during its lifetime. These constraints were applied to each sea breeze day within a regime. The nodes were then combined into clusters based on similar synoptic scale patterns. The clusters consisted of maritime anticyclone (Nodes 0, 1, 4, 8), continental anticyclone (Nodes 5, 9, 12, 13), post trough (Nodes 7, 11, 14, 15), and trough (Nodes 2, 3, 6, 10).

## *2.3 Sea Breeze v. Synoptic Scale Systems*

Similar to the sea breeze day comparison between clusters, the synoptic days were classified by NEXRAD and visible satellite imagery. The differences between sea breeze (i.e., mesoscale) and synoptic scale statistics were analyzed for 4 variables among 2 nodes within different clusters. Nodes 8 and 15 were analyzed due to the anticyclonic nature versus trough pattern. The threshold remained the same for the sea breeze days within each regime, whereas the isolated constraint was removed for the synoptic days.

### 3. Results

#### *3.1 Sea Breeze Node Assessment*

The storm lifetime as related to mean maximum radar reflectivity, storm area, maximum echo top height and storm volume were determined in the timescale normalized around the time each storm reaches its maximum reflectivity (Figures 3 – 6). An observed tendency is for a limited number of storms to live prior to -2 hours and longer than 1.5 hours. Storm initiation does occur prior to -2 hours and dissipate following 1.5 hours, but there is a small sample size of storms that exist in the timeframe, so we disregard those storms for the purposes of this study. A consistent peak is observed at 0 hours (consistent with the defined normalization) with a substantial decrease 15 minutes after the maximum reflectivity (Figure 3). Thus, signifying the weakening of the deep convective cloud or radar echo. The anticyclonic regimes in mean maximum echo top height show a decrease following the time maximum reflectivity is reached (Figure 4). However, the trough patterns exhibit an increase in maximum echo top height following 1 hour (Figure 4). The result leads to two different explanations, (1) that there is a mixture of synoptic scale systems and sea breeze days, or (2) trough regimes vary from anticyclonic regimes in terms of maximum echo top height following the time maximum reflectivity is attained. Similarly, the trough regimes experience a large increase in area and volume compared to the anticyclonic regimes (Figures 5 – 6).

#### *3.2 Sea Breeze v. Synoptic Scale Systems Assessment*

A comparison between the sea breeze days and days with synoptic scale systems for the two regimes demonstrated a similar mean maximum reflectivity in terms of radar echoes (Figure 7). A hypothesis for this comparison was an expectation to observe an increase in

overall maximum reflectivity due to a corresponding increase in area and volume. Although the increases in storm area and volume were shown, the maximum reflectivity did not support the hypothesis (Figure 7). Maximum echo top height, storm area, and storm volume demonstrate increased values in synoptic conditions compared to sea breeze days. On days with synoptic conditions the minimum area is observed to be at the time of maximum reflectivity, which leads to the analysis of storm counts per hour. At -2 hours the monitored storm count is approximately 20,000 less than the count at the time maximum reflectivity is reached. The large variation between the storm counts suggests that outliers are affecting the small sample size. The assessment of storm area is consistent with storm volume (Figure 7).

#### **4. Discussion**

An analysis of the NWS NEXRAD observations revealed primarily similar outputs for the maximum radar reflectivity, storm area, maximum echo top height and storm volume on sea breeze days. However, one exception was observed in the contrast between clusters for mean maximum echo top height. Variability in the amount of sea breeze days was noticed with the continental anticyclone regime containing the most sea breeze days with 66, whereas the subsequent largest number of sea breeze days were 32 in the maritime anticyclone regime. A decrease in the number of sea breeze days is shown with the trough regimes. For instance, the post trough regime has 13 days, while the trough regime has 7 days. The variability of the number of sea breeze days contributes to the broad spectrum for the number of storms within each hour for each cluster. The trough regime at the low end of the spectrum provides possible reasoning to the exception observed in maximum echo top height. Although the number of storms varied per cluster, the assessment offered insight into the understanding of isolated mean storm lifetime.

## **5. Setbacks**

A time constraint faced during this study occurred while extracting the statistics from NWS NEXRAD using TiNT. The amount of time to extract one day differed based on the number of storms but was in the timescale of 10 – 20 minutes. The timescale lessened the depth of the assessment due to the minimal amount of data able to be obtained in a 10-week span. An additional setback was the construction of various python scripts. The construction of each script allows for more flexibility and ease going forward but caused a delay in the analysis of the statistics. Although a delay occurred, an opportunity revealed itself in the ability to learn new aspects of evaluating real datasets.

## **6. Future research**

Future research in preparation to the TRACER campaign could increase the observations of each node to all days classified by the SOM regime. Possible comparisons consist of sea breeze days, synoptic days, pop-up convection, and clear sky days. Future research will take place during the TRACER campaign from October 2021 through September 2022 in the Houston, Texas region.

## **7. Acknowledgements**

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). We thank the TRACER science team, TRACER forecasting team, and my fellow interns for their participation in the background campaign. Lastly, we thank the NWS for providing high-resolution radar meteorological observations.

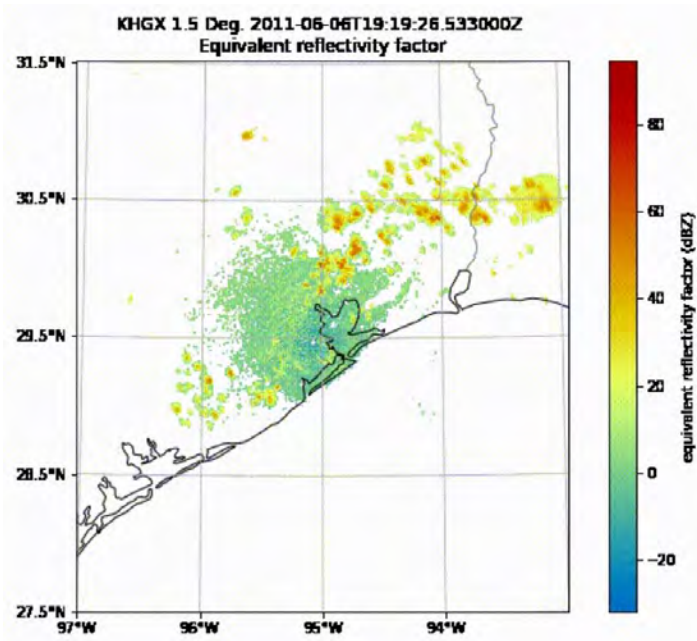
## References

<sup>1</sup> Jensen, Michael et al., 2021: Tracking Aerosol Convection interactions Experiment (TRACER) Backgrounder. DOE/SC-ARM-19-024.

<sup>2</sup> Kocen, Michele, 2013: Observations of Sea-Breeze Fronts Along the Houston Gulf Coast.

<sup>3</sup> Kohonen, T. 2001. Self-Organizing Maps. Third, Extended Edition. *Springer Series in Information Sciences*, vol. 30, Berlin, Germany: Springer-Verlag, ISBN 978-3-540-67921-9.

<sup>4</sup> Fridlind, A. M., van Lier-Walqui, M., Collis, S., Giangrande, S. E., Jackson, R. C., Li, X., Matsui, T., Orville, R., Picel, M. H., Rosenfeld, D., Ryzhkov, A., Weitz, R., and Zhang, P.: Use of polarimetric radar measurements to constrain simulated convective cell evolution: a pilot study with Lagrangian tracking, *Atmos. Meas. Tech.*, 12, 2979–3000, <https://doi.org/10.5194/amt-12-2979-2019>, 2019.

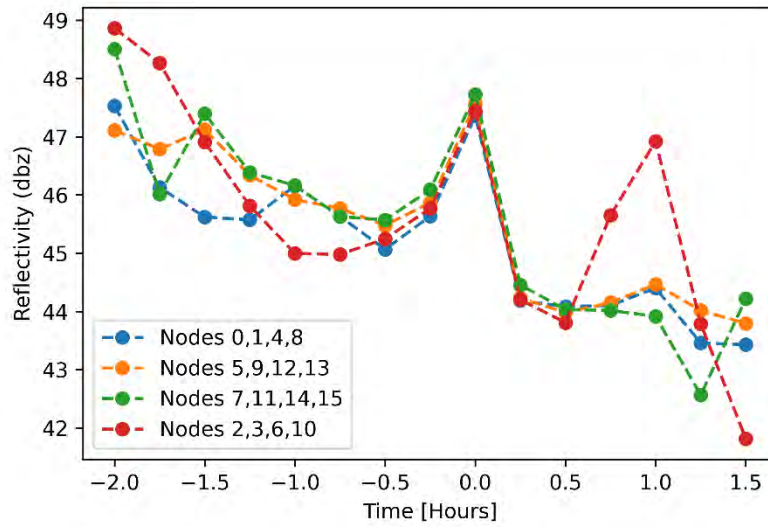


**Figure 1.** Sea / Bay breeze radar animation from June 6, 2011. The red signifies deeper convection as well as heavier precipitation.

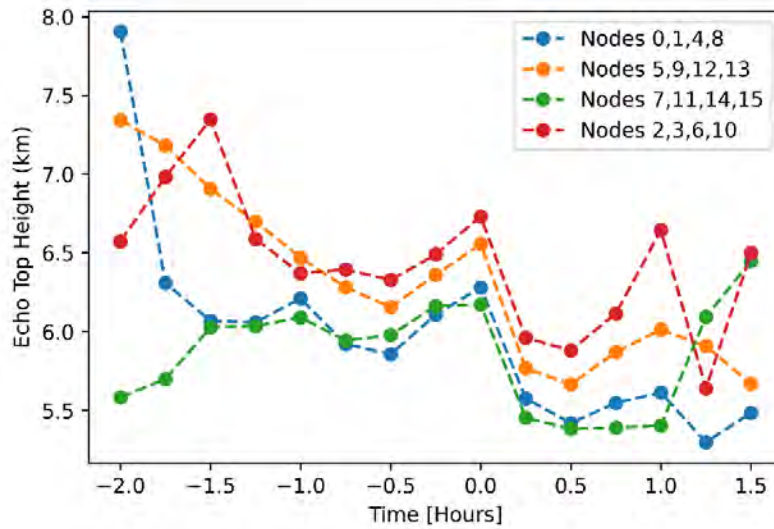


**Figure 2.** NWS NEXRAD site KHGX (Houston). The orange dot represents the location of the KHGX; orange circle is the scan domain.

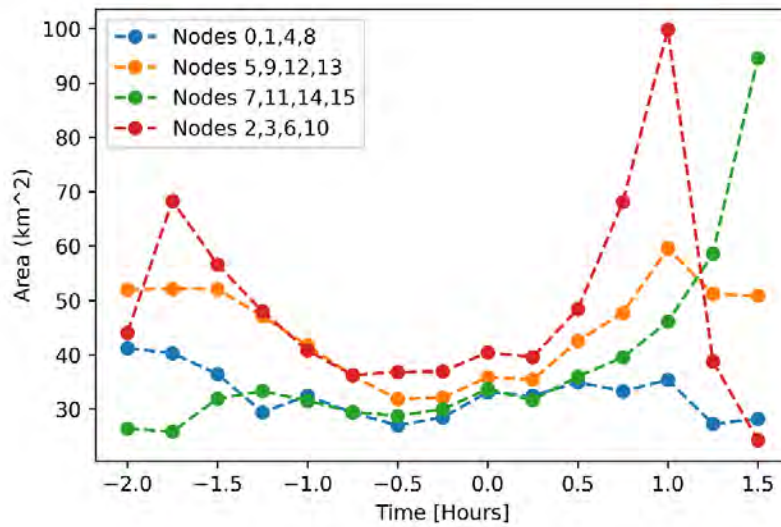




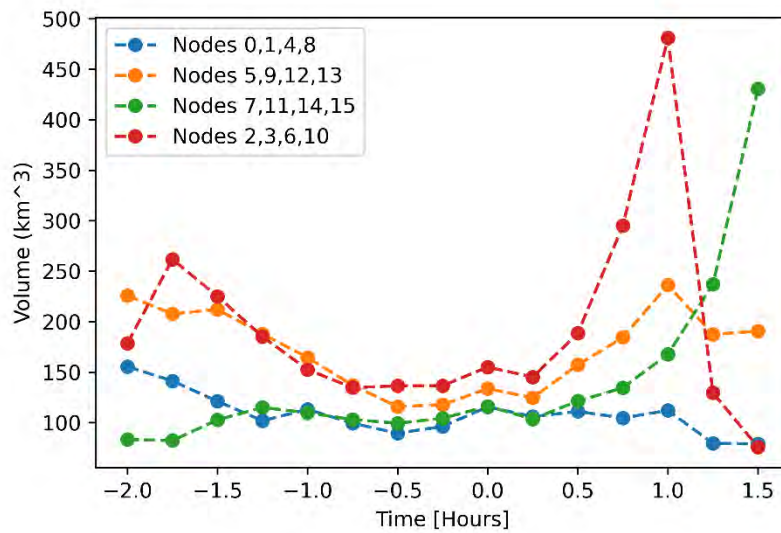
**Figure 3.** Mean Maximum Reflectivity vs. Time on possible sea breeze days.



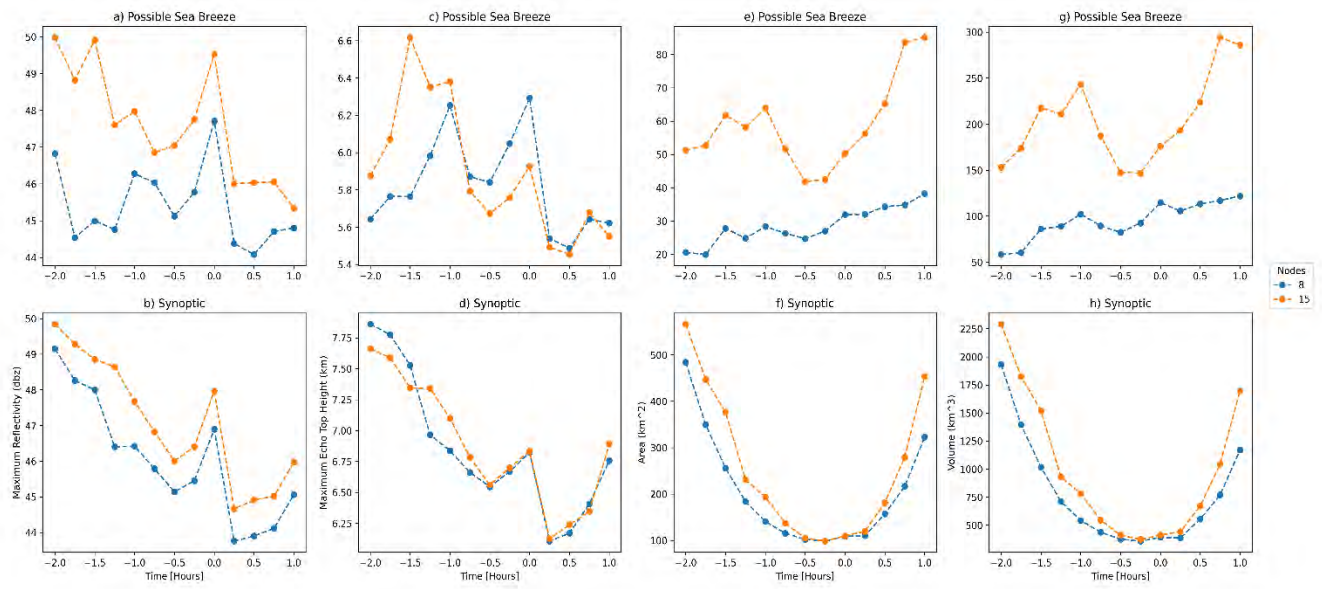
**Figure 4.** Mean Echo Top Height vs. Time on possible sea breeze days.



**Figure 5.** Mean Area vs. Time on possible sea breeze days.



**Figure 6.** Mean Volume vs. Time on possible sea breeze days



**Figure 7.** Possible Sea Breeze Days vs Synoptic Days for Nodes 8 and 15

Creating a software-defined emergency power off button

Samuel Woronick

Computer Science Department, Suffolk County Community College, Selden, NY 11784

Justine Haupt

Instrumentation Division, Brookhaven National Laboratory, Upton, NY 11973

## **Abstract**

As part of an overarching parent project known as LIQuiDNet, the Instrumentation Division at Brookhaven National Laboratory is involved in several smaller projects implementing quantum information science and technology. One of which, the free space optical link for entangled photon distribution, has the goal of demonstrating a quantum network established between two endpoints over free space. Under the guidance of my mentor, my role is to provide classical information technology and networking support to facilitate project development. My most prominent task this appointment was to produce an emergency power off solution, given the models of two different power distribution units and the fact they would be powering equipment during tests. After a short investigation period, I determined I could create a software-defined emergency power off solution that would shut off predetermined outlets on each device. I subsequently created a proof of concept using the Python scripting language. After demonstrating this to my mentor and other department members involved in her project, I received important feedback and used it to guide the development of a verbose, responsive, and reliable application. For this, I used a general-purpose programming language, Java, with which I already had some familiarity. In doing this project, I was able to enhance my skillset while simultaneously aiding my mentor and her colleagues.

## **I. Background**

The plan at the beginning of this internship was to work on a project that required site access. However, with the rise of the COVID-19 Delta variant, that would not be possible. As a result, the virtual project I worked on during the interim (until I could gain site access) became the prominent one of the internship.

Initially, my mentor asked me to learn about two different Power Distribution Units (PDU) and figure out how to interface with them. Referencing their respective manuals, this was a swift task. Shortly after, she told me they would be powering equipment during tests in a hallway and tasked me with producing an emergency power off solution.

The model numbers of the PDUs discussed in this report are omitted for simplicity. However, their collective characteristics will still be discussed.

## **II. Introduction**

While reading the manuals for each PDU, I learned that they are managed devices. In networking, intermediary devices (e.g., switches and routers) are typically categorized as managed or unmanaged, referring to their ability to be configured. In the case of the PDUs, this means they have a command line interface and support various remote management protocols. The protocol that sparked my interest is Secure Shell (SSH), which is an encrypted, remote command line. It occurred to me that I could easily automate the execution of commands to shut off predetermined outlets on each device. (Such automation is commonly done by system administrators, using a scripting language of their choice and asymmetric key pairs.) The only thing left to be verified was whether the operating system on each PDU provided commands to enable and disable specific outlets, which they do provide.

The next step was to create a proof of concept. This was done quickly using the Python scripting language. The program written is primitive and consists of about sixty lines of code. It presents a red button that maximizes on the screen and, when pressed, momentarily opens an SSH session to execute commands. (I did not have access to the PDUs during the internship, so I used SSH servers on my home network for testing.) While it lacks the verbosity and reliability of the production application that would be developed around this idea, using the Java programming language, it demonstrates the viability of the solution.

### **III. Virtual Button Design**

During the design process for the production application, the PDUs became known as target devices for two primary reasons. The first of which being that while the SSH servers are PDUs in my mentor's particular use case, that will not necessarily be true in all situations. The SSH servers could be hosted on any device capable of running the service, so it would not be appropriate to call them PDUs. The second reason is that referring to the PDUs as servers, which is the next logical thing to call them, may confuse users that are not familiar with the client-server model.

The EPO button is said to be primed (i.e., ready for use) when the application is in a connected state. This is because it maintains a session with all target devices and the physical button while primed.

The philosophy behind making the EPO button more reliable has two main aspects; the application should demonstrate to the user in an unquestionable manner that it is working properly, and it should fail loudly if a problem arises while it is in a connected state. The combination of these two things makes the application reliable because the user is confident the

application will work when the EPO button is pressed. In practice, this requires implementing a verbose feedback system. The production application informs the user of background processes via a console and additionally alerts them using audible notifications if a problem arises. This way, the user does not need to check the computer screen for problems. There is also an animation that indicates the current connection status, so the user does not need to read the console to verify connectivity, either.

To prime the EPO button, the user must provide valid credentials, socket information, test commands, and EPO commands for each target device. (Optionally, the user may also provide information to authenticate the server to the client, which is desirable in networks that are not private.) Before the EPO button is primed, and every five seconds afterwards, the application verifies the connectivity and functionality of the EPO button by simultaneously executing test commands on all target devices. If the execution of test commands fails for any target device, the user is continuously alerted via audible notifications until they either disconnect the application or the problem is resolved. The application will automatically attempt to restore itself to a connected state attempting to prime the button again every five seconds. (It is important to note that the method used to execute the EPO commands is the same one used to continuously execute the test commands.) Similarly, if the application does not receive input from the physical button for half a second, it will begin alerting the user in the same way.

#### **IV. Physical Button Design**

After demonstrating the proof of concept for the virtual button to my mentor and some of her colleagues, there were concerns about whether a panicked user would know how to click on a virtual button during an emergency. To alleviate this concern, a physical button was created for



simultaneous use with the virtual one. Unfortunately, due to time constraints, a formal design for the button was not created, only one for demonstrative purposes. However, its design is simple and will still be discussed.

The physical button consists of an Arduino Nano, USB Type-A to Micro-B cable, and a traditional EPO button. The USB Type-A connector plugs into the computer that will be running an instance of the EPO application, and the Micro-B connector plugs into the Arduino, which fits inside of the physical button. This is how the Arduino will be powered and how it will communicate with the EPO application. When the physical button is reset, a constant 5V pin on the Arduino will be connected to a digital RX pin, and when the physical button is pressed, the RX pin will be connected to a GND pin instead.

The code for the Arduino is also simple; it continuously checks whether the digital RX pin is receiving a HIGH or LOW voltage. The result of this check is sent through the serial cable as a one or zero to the computer running the EPO application. In this situation, the EPO application is a proxy program, which interprets the output of the Arduino and decides what to do accordingly. If the voltage is LOW, the EPO application will execute the EPO commands every five seconds; if the voltage is HIGH, nothing happens. This corresponds to the pressed and reset states of the physical button, respectively.

## **V. Pros and Cons**

The approach of using a software-define EPO solution has various pros and cons that should be considered carefully. Traditionally, an EPO solution is hardware-based, simple, and in-line with the circuit it controls. This is the strongest design philosophy because it needs to work

in an emergency. The software-defined solution violates that notion. By increasing the complexity of the design, the reliability of the solution is decreased.

On the other hand, using a software-define solution has unique benefits. It allows the user to centralize it for relatively little effort when multiple target devices are involved. It also allows them to easily duplicate and distribute it without any additional cost. In ad-hoc situations, it may also provide an advantage by requiring less time to set up.

Personally, I believe the bottom line is whether human lives are involved. If not, then a software-defined solution is probably acceptable. Otherwise, a hardware-based solution should be used whenever possible. Regardless of the confidence of the programmer, myself, there are simply too many variables to consider when human lives are at stake. While this may feel inappropriate for this project report, it is an important closing note, especially if anyone is considering deploying the solution described in this report.

## **VI. Closing Notes**

The next step would be to deploy the application and receive feedback about it. Much like by demonstrating the proof of concept to my mentor and hear colleagues, I would expect valuable feedback from individuals attempting to deploy it. I hypothesize there may be questions about the set-up process, so perhaps a user guide should be created. If it is decided this solution is needed in the future, it will be available.

Working on this project has allowed me to enhance my knowledge of the Java programming language and refresh that of the Python scripting language. I also learned about topics outside of my normal education and got valuable work experience that will aid me in the future.

## **Acknowledgement**

I want to thank my mentor, Justine Haupt, for giving me this opportunity and encouraging me to always keep learning. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program (CCI).

**The Civil Engineering Development of the Electron Ion Collider (EIC) Support Infrastructure**

Angela Zavala Yanes, Civil Engineering, Suffolk County Community College, Ammerman Campus, Selden, NY 11784

Joseph E. Tuozzolo, Collider-Accelerator Department, Brookhaven National Laboratory, Upton, NY11973.

## **Abstract**

The Electron Ion Collider (EIC) at Brookhaven National Laboratory (BNL) will utilize much of the existing Relativistic Heavy Ion Collider (RHIC) infrastructure. The EIC will allow polarized electrons at high energy to travel through a storage ring. Likewise, there will be another storage ring that carries positive charged particles. Eventually, the electrons and ions will collide with each other providing scientists a 3D image of the inner structure of matter. The image allows us to see and better understand the arrangement of quarks and gluons. Since electrons can individually grab the gluons; scientist study how the gluons recombine with inner structure of electrons<sup>1</sup>. For my project, Creo Parametric engineering software will be used to generate a 3D layout of new accelerator systems to be housed in a new equipment service building and beam line tunnel at IR02. This layout will be used to better visualize civil engineering infrastructure requirements for water cooling piping, electrical equipment, and equipment space. This facility will house an Energy Recovery LINAC (ERL) with 12 superconducting radiofrequency cavity cryomodules (SRF cryomodules) (**Appendix 1**) to accelerate electron beams for the EIC. Using existing models of similar SRF cryomodules, I created a representation of the harmonic H3 SRF cryomodules that accelerate the high-power electron beam. The design model includes the radio-frequency power amplifiers (RFPA) that will provide electrical energy to the SRF cryomodules and the WR1500 waveguides that transmit the power from the RFPA to the SRF cryomodules. Once, the waveguides, RFPA's, and the beamlines are assembled, I can use the 3D model to analyze the space needed for safe access for maintenance, and the routing of the 3 phase 480VAC supply system, and the water-cooling system. As a result of this internship, I will improve my skill level using Creo parametric and

better understand foundation systems, water-piping systems, electrical power distribution, and safety systems.

## **I. Introduction**

The Collider-Accelerator Department at BNL focuses on the continued development of the IR02 service building and the new tunnel at the 2 o'clock region of the EIC. This project's objective is to study the space constraints at the first and second floor of the service building to place the water-piping system and electrical equipment. By making use of a 3D software known as Creo Parametric I was able to gather, and designed parts such as a waveguide to coax adapter required to complete a final product of the ERL's power amplifier. The ERL is located at the first floor of the service building. The ERL is composed of 12 cryomodules that contained superconducting radiofrequency cavities made of an expensive material known as niobium. The ERL's cavities are known as superconductive because liquid helium is used to cool them down to 2°K for better performance. Each cryomodule has two power amplifiers located at the second floor that power the cavity inside. I assembled the waveguides and the power amplifiers for each of the ERL cavities. In addition, the 400MeV LINAC (**Appendix 2**) was also placed at the first floor. I found the copper cavities from last year and assembled six of them. I designed the waveguides used and found their respective power amplifiers. These two LINACS accelerate the electrons traveling through the beamline by providing electrical energy sent by the power amplifiers<sup>2</sup>. Nevertheless, the ERL LINAC sends its electrons to the Hadron Storage ring to cool positive charge ions to create denser particles more likely to collide with electrons at the RHIC tunnel<sup>3</sup>. On the other hand, the 400MeV LINAC accelerates its electrons at room temperature. The accelerated polarized electrons send by the warm LINAC enter the RHIC ring with the objective of colliding with dense hadrons to capture a 3D image representing the

bonding between quarks and gluons<sup>4</sup>. Scientists and engineers plan to understand what matter is made of by studying the arrangement of this plasma.

## **II. Methods/ Results**

The ERL LINAC assembly was provided to me by Douglas Holmes in which he used an older version of the 1773 MHz harmonic H3 1-cell cryomodule. Being that the cryomodule was an imported feature I used flexible modeling to modify it. I was instructed to make the cavity's diameter a 100 mm. As I scaled the cavity down to 100 mm, the body length of the part became smaller (**Appendix 3**). As a result, I created an offset plane to use as a reference to enlarge the length of the body to its original length. Furthermore, the model had an unnecessary cone shape funnels at both ends that were removed. Since, the cavity had been scaled down, the fundamental power couplers were required to be scaled down also. I used the same scaling factor on the power couplers I had use on the cavity.

I designed the WR1500 waveguide at both sides of the cryomodule where the fundamental power coupler system connects to the cavity to power it. I was also tasked to rotate the gate valves at the end of the 1773 MHz cryomodules 90 degrees to provide a different point of view and perhaps make more space (**Appendix 4**). Once the harmonic H3 cryomodule was changed I decided to replace the outdated harmonic H3 cavities by the updated version I remodel to make the assembly look more realistic. In continuation, I was also assigned to design the ERL injector buncher warm copper cavity (**Appendix 5**) placed by the upstream end of the ERL LINAC. Furthermore, the 1-cell cavities and 5-cell cavities at the ERL are made of niobium. Niobium does not absorb the electrical energy provided by the power amplifier but instead reflects it back<sup>5</sup>.

Oksana Vysochanska designed the three-port circulator (**Appendix 6**) being use at IR02. The circulator has three available ports, and one port is connected to a 90-bend waveguide (**Appendix 7**) that connects to a power amplifier. The second port is connected to a 30-bend waveguide and the third port is connected to a waveguide to coax adapter (**Appendix 8**) I designed that connects to a dummy load also known as an absorber. For the power amplifiers closer to the edge of the floor on top of the ERL assembly the 30-bend waveguide (**Appendix 9**) is replaced by a 45-degree bend waveguide (**Appendix 10**) with the purpose of leaving about one foot of space between the dummy load and the waveguide above used for the second power amplifier behind the first power amplifier.

The electrical equipment outside the building is composed of two 15KV loop switch gears, a 15KW breaker, a 3000KVA transformer and a 480V switch gear that powers all the ERL and warm LINAC power amplifiers<sup>6</sup>. The 480VAC three phase electrical power (**Appendix 11**) provides a total of 720KW of electrical energy to each one of 591MHz cavities' power amplifiers<sup>7</sup>. There are two 591MHz cavities after the electron source. Since the power amplifier is only 50% efficient, the power amplifier absorbs 360KW of the RF power and the other 360KW is sent to the cavity<sup>8</sup>. The absorbed RF power is converted into water heat load<sup>9</sup>. Since the power amplifier is sensitive to temperature, the water supply is 30°C and returns at 36°C. Delta T may only be 6°C otherwise the centrifugal pumps in the pump room will require a larger electrical demand<sup>10</sup>. We do not want to use anymore electrical energy because we are already using lots of it. Each cavity has two fundamental power couplers thus each power coupler obtains 180KW of RF power<sup>11</sup>. Being that niobium does not absorb RF power, some power is reflected to the dummy load. Only 2.5% of power is absorb by the circulator thus it's considered negligible, however the dummy load absorbs 15% of the RF power in normal



conditions<sup>12</sup>. The dummy load is not as sensitive to temperature because it only absorbs 15% of energy unlike the power amplifier in which it absorbs 50% of the energy<sup>13</sup>. As a result, the quantity for gallons per minute (GPM) is a smaller quantity. If the GPM is gets smaller then less electrical energy will be required to pump the water through the pipes thus saving energy. Nevertheless, the ERL injector booster's dummy load (**Appendix 12**) designed by Oksana Vysochanska is planned to absorb 100% of the RF power reflected even if it doesn't absorb it all. Moreover, to obtain the RF power left by one cavity to accelerate the electrons in the beamline may be found by subtracting 360KW by 15% of 360KW<sup>14</sup>. The total power that goes into the beamline in the ERL injector may be found by adding both cavities' result of the subtraction between the 360KW minus the 15% of the 360KW (**Appendix 13**).

The transformer sends 70KW of RF power to one harmonic H1 power amplifier. Since the power amplifier is only 50% efficient it absorbs 35KW of RF power and the other 35KW is sent to the cavities. The 35KW absorbed becomes water heat load. Just like the injector booster delta T is 6°C for the power amplifiers and it cannot be larger than that. Otherwise, the GPM will have a large value and more energy will be in demand to pump the water. The 591MHz cryomodule has two cavities hence it contains two fundamental power couplers<sup>15</sup>. Each power coupler gets 17.5KW of RF power. Since only 15% of the cavity's RF power is absorb by the dummy load unlike 50% absorb by the power amplifier delta T is 15°C thus it has a smaller GPM. Being that the dummy load absorbs less power than the injector booster I decided to use the 40KW RF water-cooled absorber (**Appendix 14**) I designed. The competency of this dummy load goes as far to absorb not only 15% of the RF power reflected but 100% of it. To obtain the amount of RF power used to accelerate the electrons in the beamline subtract 35KW minus the 15% of the 35KW.

The same procedure is use for the harmonic H3 cryomodules. Each cryomodule contains 1 cavity, and each cavity contains 1 cell<sup>16</sup>. The 480VAC electrical energy sends 40KW of RF power to the power amplifier. Due to its 50% inefficiency, the power amplifier absorbs 20KW of RF power. The absorbed 20KW of power is then converted into water heat load. Once again, delta T is 6°C. The delta T may only be 6°C otherwise the centrifugal pumps will require a larger electrical input. In this case, the absorbed power by the amplifier is smaller than all the other cavities thus the GPM will be smaller. Furthermore, 10KW of power is fed to each power coupler and 15% of the total 20KW fed to the cavity is absorbed by the 40KW RF dummy load. Since, the dummy load only absorbs 15% of 20KW then the amount of energy converted into water heat load is smaller thus the dummy load can have a delta T of 15°C. In any case, the GPM for the dummy load will be smaller than the GPM for the power amplifier. The RF power that goes into the beamline may be found by subtracting 20KW minus 15% of the 20KW.

The ERL 197MHz warm injector buncher cavity is made of copper. Copper does not reflect RF power because copper acts as an absorber<sup>17</sup>. Like all other cavities, this warm injector is also powered by the 480VAC three phase electrical power. The power amplifier is fed 120KW. The power amplifier absorbs 60KW and feeds 60KW to the cavity. The 60KW energy absorb by the power amplifier turns into water heat load to prevent overheating. In my Creo assembly there is no dummy load nor a circulator because the copper absorbs the RF energy sent to it by the power amplifier. The energy absorbed by the copper cavity is then converted into water heat load also.

In comparison to the ERL the 400 MeV LINAC also known as the warm LINAC works similar but differently at the same time. The cavities are made of copper instead of superconductive material because it is cheaper, and it gets the job done. The assembly is located

at the first floor 102 inches from the ERL LINAC. The LINAC is also next to a concrete shielding block to prevent radiation on the other side. The size of the shielding block is still subject to change. The power amplifiers of the 400MeV LINAC are located at the first floor but on the opposite side of the shielding block. Each of the power amplifiers are connected to waveguides that connect to the copper cavities. However, there is no circulator nor dummy load on the power amplifiers. The reason why there is no need for a circulator, nor a dummy load is because the copper cavities act as absorbers. The warm LINAC power amplifiers are supplied energy by the same electrical equipment that the ERL LINAC is powered by. The electrical equipment provides 14KW of energy to the power amplifier. However, the power amplifier absorbs 4KW of RF power. These 4KW of RF power turns into water heat load. The delta T is still 6°C but the GPM is smaller because the water heat load is only 4KW. The 10KW of RF power is send through the waveguides to copper cavity. The copper cavity absorbs the 9KW of RF power and 1KW of RF power is used to accelerate the electrons<sup>18</sup>. In this case, the delta T is 10°C but the water heat load is 9KW. Thus, the GPM of the dummy load is larger than the power amplifier. In contrast to the ERL, the warm LINAC's power amplifier has smaller GPM thus it demands less electrical energy to pump the water unlike the dummy load that has a larger GPM, and it does require more electrical energy for pumping water through the pipes.

### **III. Conclusion**

In conclusion, the 3D Creo Parametric model did help better visualize the building. The way some power amplifiers had been placed at the 2D plane layout in AutoCAD did not leave enough space once it was placed in the building using Creo Parametric. I rotated the power amplifiers 90 degrees to make more space available for maintenance. The typical distance

between the Harmonic H1 cryomodels was about 80 inches which is enough space for a person. In addition, the AutoCAD layout showed the ERL waveguides simply connected to the power amplifiers at the second floor when in reality some waveguides obstructed other waveguides space. The IR02 3D building will allow for better analysis to place the water-cooling system pipes and the electrical conduits. However, the IR02 building is still subject to change to find better ways to take advantage of space and reduce the cost of this project. I wish I had more time to continue looking for new ideas to assemble the waveguides for the ERL assembly to make more space. Not to mention, I would have love to design the water-piping system and the electrical equipment.

## References

<sup>1</sup> EIC Science. (n.d.). Retrieved July 2020, from <https://www.bnl.gov/eic/science.php>

<sup>2</sup> Tuozzolo, Joseph. Internship meeting. July 2021.

<sup>3</sup> Tuozzolo, Joseph. Internship meeting. July 2021.

<sup>4</sup> EIC Science. (n.d.). Retrieved July 2021, from <https://www.bnl.gov/eic/science.php>

<sup>5</sup> Tuozzolo, Joseph. Internship meeting. July 2021.

<sup>6</sup> Nerhing, Tom. Internship meeting. July 2021.

<sup>7</sup> Smith, Kevin. Internship meeting. July 2021.

<sup>8</sup> Smith, Kevin. Internship meeting. July 2021.

<sup>9</sup> Srinivasan, Ram. Internship meeting. July 2021.

<sup>10</sup> Tuozzolo, Joseph. Internship meeting. July 2021.

<sup>11</sup> Smith, Kevin. Internship meeting. July 2021.

<sup>12</sup> Tuozzolo, Joseph. Internship meeting. July 2021.

<sup>13</sup> Tuozzolo, Joseph. Internship meeting. July 2021.

<sup>14</sup> Tuozzolo, Joseph. Internship meeting. July 2021.

<sup>15</sup> Tuozzolo, Joseph. Internship meeting. July 2021.

<sup>16</sup> Tuozzolo, Joseph. Internship meeting. July 2021.

<sup>17</sup> Tuozzolo, Joseph. Internship meeting. July 2021.

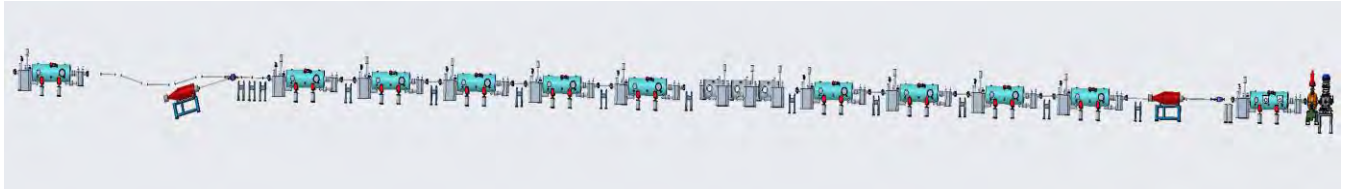
<sup>10</sup> Tuozzolo, Joseph. Internship meeting. July 2021.

## **Acknowledgements**

I would like to thank my mentor Joe Tuozzolo for giving me an opportunity to learn new skills and grow as a professional. I would also like to thank Kevin Smith, Alexander Zaltsman, Ram Srinivasan, Vito De Monte, Steven Bugros, Peter Braunius, Douglas Holmes and Michael Stegman for guiding me throughout this internship. My colleagues Oksana Vysochanska, Joseph DeRienzi, and Thea Vijaya Kumar. In addition, this project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program.

## Appendix

### Appendix 1

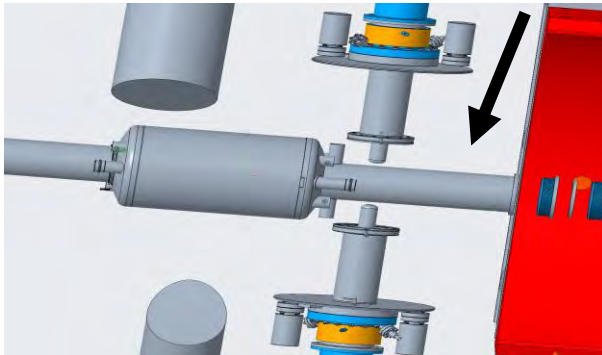


### Appendix 2

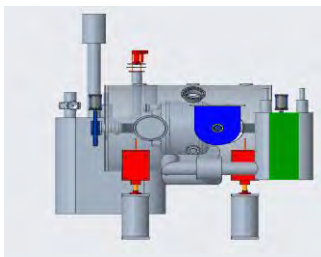


### Appendix 3

100 mm

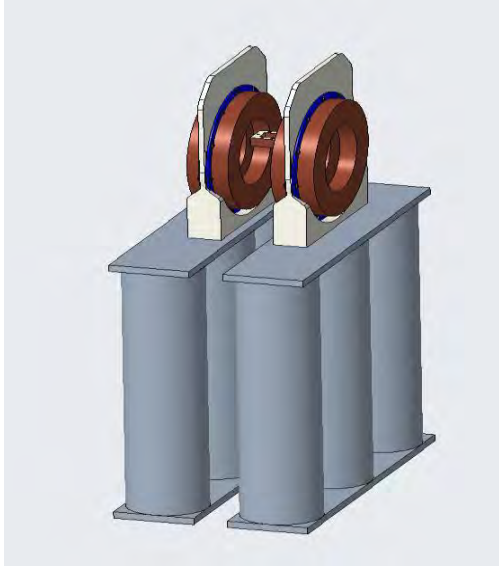


### Appendix 4

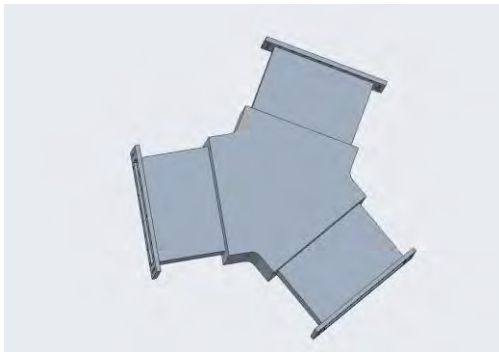


### Appendix 5

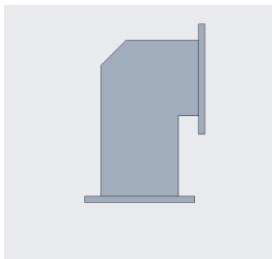




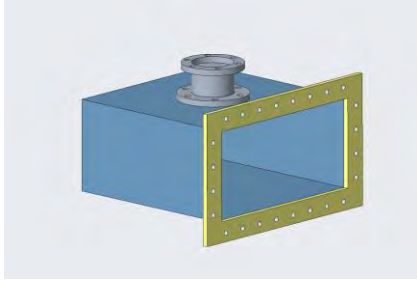
Appendix 6



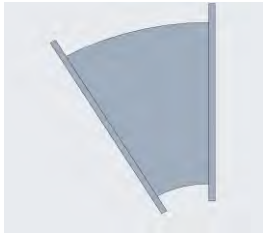
Appendix 7



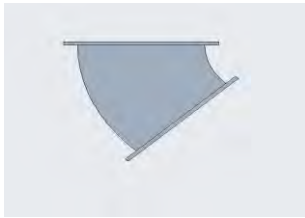
Appendix 8



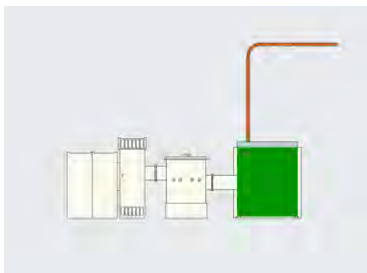
Appendix 9



Appendix 10



Appendix 11



Appendix 12



## Appendix 13

Row	Subsystem	Equipment Description	PA RF Output Power kW	AC Power kW/2	AC Voltage	Notes	Special Needs Water	Water Heat Load kW/2	Supply Temp
ERLG 1	SHC esource SRF System (591 MHz)	Power Amplifier	360	720	480	3 phase	Filtered, Temp	360	30
ERLG 1	SHC esource SRF System (591 MHz)	2 Dummy Loads for 2 waveguides				WR1500	No	54.0	30
ERLG 1	SHC esource SRF System (591 MHz)	2 Circulator (transmission loss)				WR1500	No	9.0	30
ERLG 2	SHC esource SRF System (591 MHz)	Power Amplifier	360	720	480	3 phase	Filtered, Temp	360	30
ERLG 2	SHC esource SRF System (591 MHz)	2 Dummy Loads for 2 waveguides				WR1500	No	54.0	30
ERLG 2	SHC esource SRF System (591 MHz)	2 Circulator (transmission loss)				WR1500	No	9.0	30
ERLG 3	SHC esource warm buncher (197 MHz)	Power Amplifier	60	120	480	3 phase	Filtered, Temp	60	30
ERLG 3	SHC esource warm buncher (197 MHz)	Dummy Load				WR1500	No	9.0	30
ERLG 3	SHC esource warm buncher (197 MHz)	Circulator (transmission loss)				WR1500	No	1.5	30

## Appendix 14



***Upgrading the Area Radiation Monitor Infrastructure at the National Synchrotron Light Source-II***

Robert Zinser - Intern  
Suffolk County Community College, Selden NY 11784

Matthew McCarthy - Intern  
Suffolk County Community College, Selden NY 11784

Robert Rainer - Mentor  
Controls Engineer, National Synchrotron Light Source II, Upton, NY 11973

## I. ABSTRACT

The Personal Protective System used at the National Synchrotron Light Source II (NSLS-II) [Figure 1] is complex and filled with various redundancies and safety measures. The purpose of the PPS is always to ensure the safety of all personnel who either work at or visit the facility. Within the PPS you have the Area Radiation Monitoring System (ARMs), which was the primary focus of this project. In order to improve the reliability and workflow of the system, the focus for the project became the Input/Output Control system (IOC). Currently the ARMs are run off one global IOC. While the global IOC functions without issue, a weakness is when maintenance on individual devices are needed. Work on an individual device results in global interruptions of the entire system. To prevent this, a template IOC was created for access to common binaries, executables, and database files that individual IOCs can access for Operations. This new IOC structure will allow for updates common to the common directories to be propagated throughout all IOCs, without the risks associated with having to update each individual subsystem. Work on this project has given exposure to a range of new applications and operating systems that are not included within normal school curriculum. Software and Operating Systems involved include: LINUX, Experimental Physics Industrial Control System (EPICS) and the Graphical User Interface platform Control System Studio (CSS).

## II. INTRODUCTION

The National Synchrotron Light Source II (NSLS-II), in its entirety, is an international research facility [1] located in Upton, NY. The Personal Protective System (PPS) used at the National Synchrotron Light Source II (NSLS-II) facility is complex and filled with various redundancies and safety measures. The purpose of the PPS is to always ensure the safety of all personnel who either work at or visit the facility. Within the PPS is the Area Radiation Monitoring System (ARMs), which uses HPI 6030 monitoring devices [Figure 2]. These are low level ion chamber detectors capable of reading radiation doses from 0.01mR/h to 100mR/h and they connect to digital displays, HPI 6016s, in order to communicate directly with a programmable logic controller (PLC) that is tied into the PPS. This data is then output through servers that convert the signals into what is known as Process Variables (PVs) in the Experimental Physics Industrial Control System (EPICS). This information is displayed through graphical user interfaces (GUIs) in the Control System Studio (CSS) platform which can be found on systems across the facility. The Input/Output Control System (IOC) architecture that is currently in use is undergoing an upgrade, moving away from a single IOC that controls all the ARM devices. As it stands now, any time a device needs to be replaced it requires an IOC reboot. In this event, the entire ARM system throughout the facility is affected. Our goal with this project is to create a template IOC for access to common binaries and executables. This will allow future updates of the common directories to be pushed out in mass rather than going one by one across more than 50 different active ARMs. This new workflow will reduce the likelihood of human error and conserves server resources in the process. The devices will then receive their own IOCs targeted to the specific needs of that device. As a result of this spring, I have become much more familiar with range of topics and application in software to include EPICS, Linux, CSS, and IOC development.

### III. FACILITY OVERVIEW



*Figure 1 – National Synchrotron Light Source-II*

The particle accelerator portion of National Synchrotron Light Source II is owned and run by Brookhaven National Laboratory [Figure 1]. It is a state-of-the-art medium-energy electron storage ring designed to produce x-rays of high intensity and brightness. Owing to the radioactive nature of the x-rays, shielding and radiation monitors are an integral part of the facility. The walls of the Storage Ring Tunnel are comprised of lead-based cement and have extra radiation shielding throughout the complex. The complex is broken down as follows: Firstly, there are 5 equally distance adjacent sections known as “Pentants”, each of which contains 6 cells. The experimental floor can be accessed through the main lobby, as well as one of the five the Lab Office Building (LOBs). The mezzanine contains many of the power supplies that various devices in the accelerator rely on. Alongside this, there are 5 Support Buildings (SBs) which have access-way to the Storage Ring Tunnel. “Maintenance Doors” at each cell allow access to the Storage Ring Tunnel from the Experimental Floor. All beamlines are located on the experimental floor with a few exceptions that have satellite buildings. Lastly, there is an Injector Service Building (ISB), which houses the LINAC and Booster [2]. At each maintenance door and beamline, an Area Radiation Monitor [ARM – Figure 2] is placed. The model of ARMs used at NSLS-II is HPI 6016, with which the ISB has several placed at strategic areas to carefully monitor any radiation that is created beyond the shielding.

### IV. UNDERLYING SOFTWARE/HARDWARE

Area Radiation Monitors [Figure 2] are useful in maintaining levels of radiation in a specific area that are safe for human contact. These monitors emit real-time observations of radiation levels, as well as records those levels in order to be references in the case of a threshold being exceeded, as well as any time an outlier is detected. The ARMs used at NSLS-II are equipped with 3 lights in relation to *high*, *low*, and *ok* levels of radiation. With this, there is a screen that displays the specific radiation level, which later on can be grabbed by an Input/Output controller to be further examined from different remote areas. Every controls system needs a set of tools specific to their needs. Brookhaven National Laboratory uses EPICS (Experimental Physics and Industrial Control System) underlying framework to rely on regarding transferring values and thresholds into legible text in a Graphical User Interface (GUI). More specifically, NSLS-II uses the Debian Linux package in order to integrate the controls software(s) into a Linux-based Operating System (OS). This allows the user (a.k.a. any given controls engineer) to perform specific tasks through a terminal. A common advantage of a Linux-based OS is the availability to have multiple terminals open at once. This gives way for a more efficient

workflow – different Input/Output Controllers can be open in different terminals and referencing each other is more efficient. More specifically EPICS allows for use of hundreds, if not more, of computers to be run under a centralized control network using the Channel Access (CA) network protocol. Due to the sufficient level of bandwidth that the CA network protocol is built to handle, these computers are all capable of perform real-world Input/Output tasks at the same time, as well as access the state of said IOCs and their underlying ARMs from virtually anywhere. The reliability of EPICS thus far exceeds the Laboratory’s needs, being able to run nonstop for as long as is needed. This reliability ensures that any Researchers borrowing time and space at the beamline will get the most out of their period, and the potential for countless IOCs under one network is essential in the idea of upgrading NSLS-II’s IOC architecture.

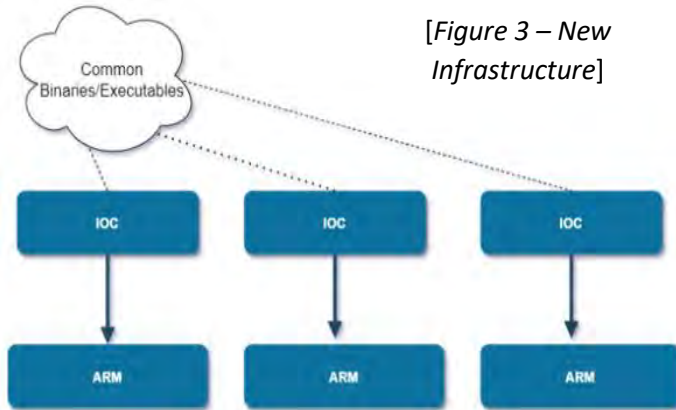


Figure 2 - HPI 6016 Area Radiation Monitor

## V. NEW INFRASTRUCTURE

To ensure the safety of those borrowing space at the National Synchrotron Light Source-II, there is one Area Radiation Monitoring device implemented outside each cell around the beamline. These Area Radiation Monitors, otherwise known as ARMs, all need to be read by an IOC in order for an end user to pull and modify more information pertaining to the device it is controlling. The IOC infrastructure at NSLS-II is outdated and rather redundant, and only holds one global Input/Output Controller for *all* Area Radiation Monitors. While this structure has worked for some time, the irritability comes when one ARM needs to be referenced. In the situation that a threshold value needs to be changed or added to an Area Radiation Monitor, the IOC that is overseeing that Area Radiation Monitor needs to be rebooted. In this case, when the global IOC is rebooted for maintenance, every ARM under it is affected. Those borrowing time and space at the beamline may be interrupted in their experiments, which is one of the center priorities at NSLS-II. Upon realization of this, a new infrastructure [Figure 3] was proposed in order to control each ARM more efficiently. Within this new infrastructure, a new Input/Output Controller will be implemented for *each* Area Radiation Monitor. This will ensure that when modifying files or threshold values related to one ARM, every other ARM will have their own set of files that will not be affected. Alongside this, the Binaries and Executables that are not often





[Figure 3 – New Infrastructure]

modified are still kept global, however they will now apply to each of the separate IOCs, rather than one. Essentially, the new standard will be for each Area Radiation Monitor to have its own Input/Output Controller.

**⊞ Subsection: The Workflow**

To go about modifying the infrastructure, terminal commands are inputted through the Linux-based Operating System running EPICS (as mentioned in *Section IV*). As it sits, the global Input/Output Controller (IOC) is held in one single directory, `/RadMonACC/`. Within this directory is the `/hpi6016/` child directory, holding all of the database files for Radiation Monitors circumventing National Synchrotron Light Source II. More specifically, there is a deeper child directory within `/iocs/RadmonACC` that accompanies the same name, `/iocBoot/iocRadMonACC/`, and holds the `st.cmd` file [see Figure 4 for the full path]. This file in specific is the command file used to run the IOC. Given a communication failure to a particular device, network fidelity drops, or power dips, the `st.cmd` file would need to be run again, interfering with all of the macros (which reference each ARM individually) that lay under the parent IOC. For the purpose of being able to pinpoint modifications to one ARM without interrupting others, each macro (which pertains to each Area Radiation Monitor) that currently lies in a file under the global RadMonACC IOC, will be converted into a secluded copy under its own new IOC. This will require creating new directories for each individual IOC, as well as modifying the RadMonCommon file [Figure 4] to load records and environmental values from each separate IOC [Figure 5]. It does so by calling for a variable within the portion of script that loads the records themselves. RadMonCommon will then grab the information from each new IOC that reflects the variable. This allows for all database files and environmentals to be loaded separately under a single parent command file.

To go about modifying the infrastructure, terminal commands

```

rrainer@dev-acc: /epics/iocs/HPI_Linac_1
rrainer@dev-acc: /epics/iocs/HPI_Linac_1 74x26
GNU nano 2.7.4 File: st.cmd Modified
#) /epics/iocs/RadMonCommon/bin/linux-x86_64/hpi6016
#the line above (#!...) is not a comment: it actually executes the IOC bin$
# load common settings from commonSt.cmd file
</epics/iocs/RadMonCommon/commonSt.cmd

epicsEnvSet("SYS", "LN-AM")
epicsEnvSet("DEV", "RadMon:1")
epicsEnvSet("PORT", "HPI_Linac_1")
epicsEnvSet("DEV-NAME", "HPI_Linac_1")
epicsEnvSet("IP", "10.8.135.188:4001")
]
# initialize the IOC with Access security control
</epics/iocs/RadMonCommon/commonInitWithAutoSave.
  
```

[Figure 5 – Global Area Radiation Monitor Common file]

```

rrainer@dev-acc: /epics/iocs/RadMonCommon
rrainer@dev-acc: /epics/iocs/RadMonCommon 76x26
GNU nano 2.7.4 File: commonSt.cmd Modified
# Common support components
dbLoadDatabase "/epics/iocs/RadMonCommon/dbd/hpi6016.dbd"
hpi6016_registerRecordDeviceDriver(pdbbase)

# Environment Variables
epicsEnvSet("ENGINEER", "Rainer x5473")
epicsEnvSet("LOCATION", "Bldg. 740")

epicsEnvSet("STREAM_PROTOCOL_PATH", "${TOP}/proto")
epicsEnvSet("EPICS_BASE", "/usr/lib/epics")
epicsEnvSet("EPICS_CA_AUTO_ADDR_LIST", "NO")
epicsEnvSet("EPICS_CA_ADDR_LIST", "10.0.153.255")

# Load record instances for db/HPI6016, CS5/reccaster, iocAdminSoft/Reboot
dbLoadRecords("/db/HPI6016.db", "SYS=${SYS}, DEV=${DEV}, PORT=${PORT}")
dbLoadRecords("${EPICS_BASE}/db/reccaster.db", "P=OP-CT{IOC:${IOCNAME}}RecS$")
dbLoadRecords("${EPICS_BASE}/db/iocAdminSoft.db", "IOC=OP-CT{IOC:${IOCNAME}}$")
  
```

[Figure 4 – Single Area Radiation Monitor IOC]



## VI. SUMMARY

The National Synchrotron Light Source II (NSLS-II) is considered essential by many researchers in bringing their experimental ideas and questions to light. This liability is one of the centers of importance at the facility and is one of the main reasons that the Personal Protective System (PPS) was implemented in the first place. Ensuring the safety of all personnel who either work at or visit the facility is essentially the big concern that this project accounts for. The Area Radiation monitors within the PPS directly contribute to that safety by providing an output that is legible to anyone accessing it. The low-level ion chamber detectors output 'high', 'low', and 'ok' radiation doses from 0.01mR/h to 100mR/h on the HPI 6016s digital display. By completion of the entire infrastructure implementation, specific Area Radiation Monitors (ARMs) will be able to be modified, turned off, changed, and/or replaced without affecting its surrounding ARMs and their respective HPI 6016s/HPI 6030s. In its entirety, the updated infrastructure makes sure that anyone who is borrowing a cell can make the most use of their time and space by not having to leave the cell or be subject to unknown conditions if another device caused that cells arm to be unresponsive. The new structuring will inflict each IOC to be run through a main common file, allowing for separate access to the specific IOC if needed. It will also impose a new workflow that will be easier to move through, as well as reduce human error when modifying certain directories and files.

### *Acknowledgements*

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internships Program (CCI).



- [1] *BNL Photon Sciences* | “About National Synchrotron Light Source II.” *About NSLS-II*, 30 Mar. 2021, [www.bnl.gov/ps/nsls2/about-NSLS-II.php](http://www.bnl.gov/ps/nsls2/about-NSLS-II.php)
  
- [2] *Proceedings of PAC07* | “Conceptual Design of the NSLS-II Injection System.” *Albuquerque, NM*. 6 April. 2021, <https://accelconf.web.cern.ch/p07/PAPERS/TUPMS083.PDF>
  
- [3] *EPICS Community Documentation* | “EPICS 3.13 Record Reference Manual” 26 May. 2009, [https://wiki-ext.aps.anl.gov/epics/index.php/RRM\\_3-13](https://wiki-ext.aps.anl.gov/epics/index.php/RRM_3-13)
  
- [4] *Argonne National Laboratory* | “Experimental Physics and Industrial Control System” 28 February. 2021, <https://epics.anl.gov/index.php>
  
- [5] *Epics Documentation* | “EPICS Related Documentation” 2019, <https://docs.epics-controls.org/en/latest/index.html#>

# Monte Carlo simulations of radiological issues for Electron-Ion Collider around Interaction Point 12 of Relativistic Heavy Ion Collider

J. Zweifel, Department of Physics, University of Rochester, New York 14627

K. Yip, Collider-Accelerator Department, Brookhaven National Lab, Upton, New York 11973

# **I Abstract**

The future installation of the Electron-Ion Collider (EIC) at Brookhaven National Lab (BNL) has brought forth a need to assess the current radiation shielding at the 12 o'clock site of the Relativistic Heavy Ion Collider (RHIC), which will be converted into the EIC. Here, we examine the resulting radiation due to the hadronic collimator, located in the 12 o'clock experimental hall, and the electron beam dump, located east of 12 o'clock, between magnets Q3 and Q4. Each equipment site has unique geometry and each experience different particle loss patterns. Using Monte Carlo simulation package MCNP6.2, we constructed detailed models of the collimator and the beam dump including their respective surroundings. We then ran simulations to estimate the radiation doses outside the shielding structures. Simulation results suggest that current shielding is sufficient, but soil activation is a serious concern that requires an impenetrable cap to prevent leakage into the water table. Additional shielding can be added if it is found to be necessary.

# **II Background**

The RHIC and the future EIC are potential sources of significant radiation emissions as trillions of particles will circulate the EIC during operation. While the RHIC has already been in operation for over two decades, the EIC has yet to be built and therefore no data exists to suggest the safety of the EIC in regards to radiation emissions. Thus, simulation of areas of high potential for radiation emission ensures that safety standards are followed for the well-being of BNL employees, the general public, and the environment.

New accelerator instruments, which may potentially produce high radiation, will be placed at two different locations around the 12 o'clock interaction point of RHIC. Running simulations ensures existing radiation shielding is sufficient to meet BNL radiation dose emission standards.

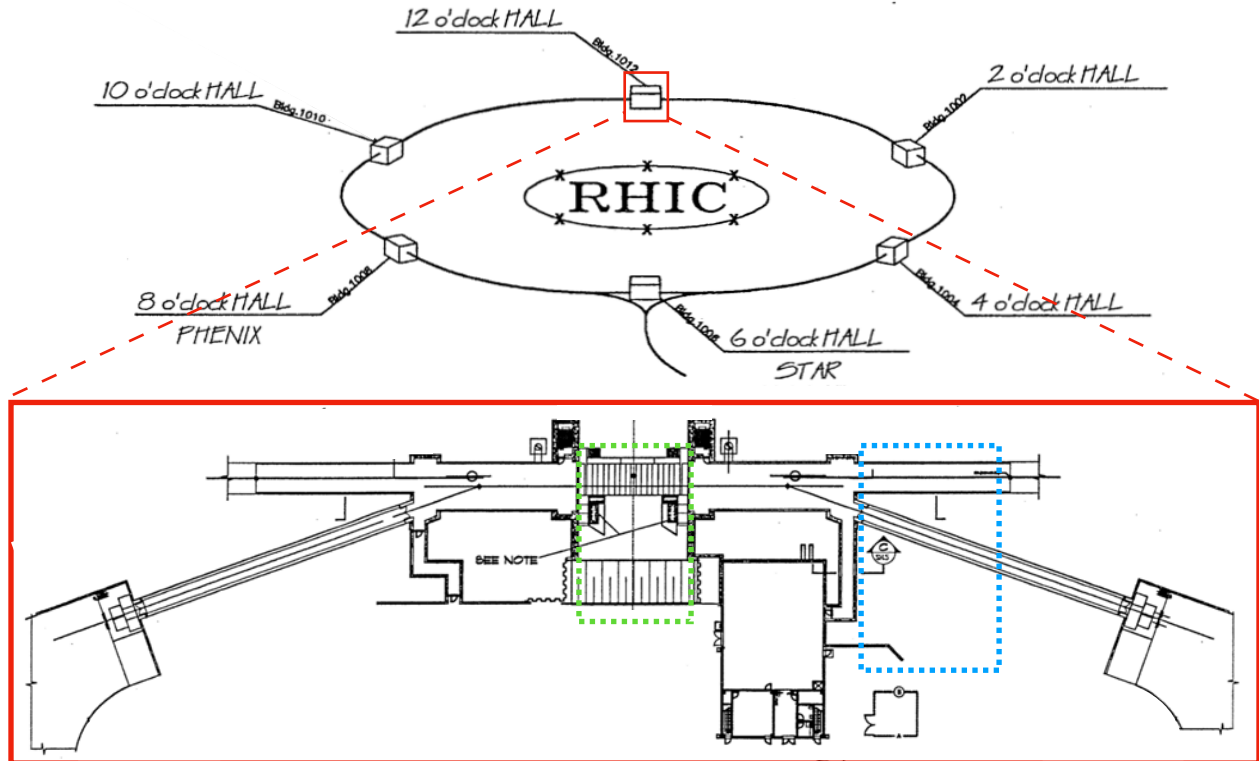


Fig. 1 A map<sup>1</sup> of the RHIC collider showing the areas of interest for simulations shown in this report. Area in green is an approximate regions simulated for collimator simulations. Area in blue is approximate region simulated for beam dump site.

The first simulation created was at the future site of the EIC hadronic collimator, located at the 12 o'clock interaction point site of RHIC within the area shown in green in Figure 1. A collimator's role is to narrow the particle beam (analogous to removing the fuzzy halo surrounding a laser beam or flashlight) thus reducing possible background for particle detectors at the interaction points. In private communications with Angelika Drees<sup>2</sup>, an expert on collimators, the routine loss of hadrons due to a collimator is 1% loss/hour, with a maximum loss

of 10% loss/hour. This makes the collimator site critically important to simulate as radiation of this magnitude could be detrimental to one's health.

The second simulated area was the site of the electron beam dump which was chosen because excess electrons will be disposed of here constantly during operation. EIC's electron beam dump is expected to be positioned between the Q3 and Q4 magnets, which lie between the 12 o'clock and 1 o'clock interaction points; this is the roughly the region within the blue box in Fig. 1. As many as 1.24 quadrillion ( $1.24 \times 10^{15}$ ) electrons/hour will be dumped each hour—supporting calculations shown in section V. This source of electron radiation can cause neutrons and photons to be produced under nuclear interactions, which can more easily pass through the shielding.

Both areas were simulated to test that dosage in areas of possible exterior foot travel; these regions include parking lots and key points outside the berm. Additionally, there is a substantial concern for radiation contamination in the soil surrounding the electron beam dump, called soil activation, which can leech into the water table and thus cause serious damage to the water supply. This occurs when radiation in the form of free neutrons enters nearby soil and causes stable isotopes to become unstable.

To create the simulations, this project relied upon MCNP6.2 for all simulations which is a transport code that uses Monte Carlo methods to simulate the procession of particles traveling through a simulated geometry<sup>3</sup>. While traveling through the simulated geometry, particles interact with their surroundings probabilistically, either continuing on their original path or causing new particles to spawn. Only two types of detectors were used in the simulations. The

first utilized MCNP6.2 detector tallies, which act as detectors at a single point<sup>3</sup>. For these simulations, output from detector tallies gives a count of the measure of rem/particle at the tally location. The second type of detector, MCNP6.2 mesh tally, generates 2D histograms that give a more comprehensive look at the distribution of radiation intensity (flux) at various locations of interest in a color spectrum<sup>3</sup>. Only neutron and photon tallies are used in the simulation because they are the only two relevant radiological dosages posing potential safety concerns at these locations.

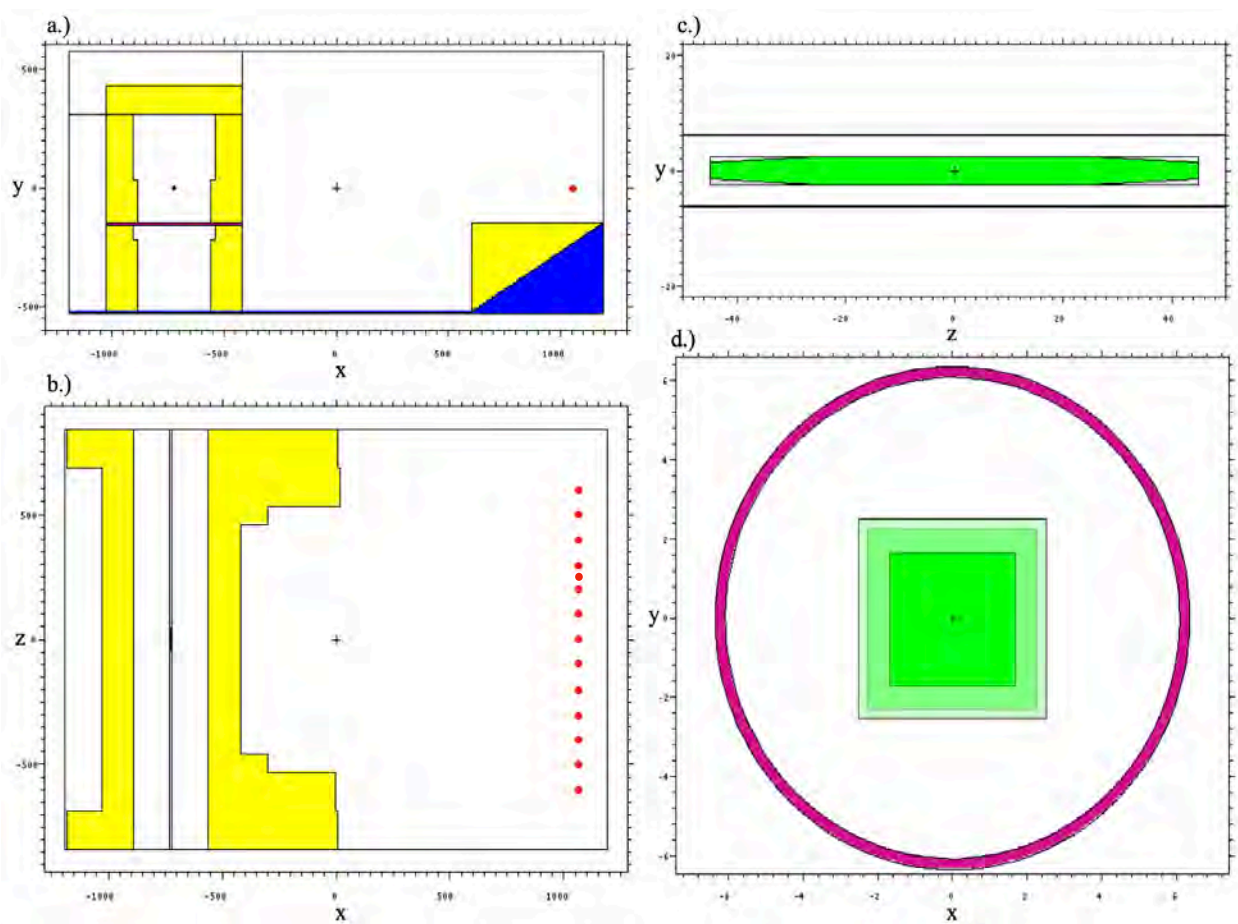


Fig. 2 Four different views of the collimator site—red dots represent detector tallies.  
 Fig. 2a. shows a screenshot of the xy plane of the experimental hall and parking lot at  $z=0$ .  
 Fig. 2b shows the xz plane at  $y=0$  and the detector tallies.  
 Fig. 2c shows the collimator design in the yz plane.  
 Fig. 2d shows three different z positions of the collimator within the beam pipe in the xy plane.

### III Methods

The general construction of the two geometric models followed a similar process. Blueprints were carefully studied to understand the originally planned physical geometry of the area. Since the blueprints are upwards of two decades old, photographs of the exterior of the 12 o'clock experimental hall were also obtained to compare the current state of the shielding against the original blueprints. From this combination of blueprints and photographs, measurements were obtained and wherever organic geometry occurred, approximations were taken to favor a slight overestimate of the radiation (e.g. the berm). All measurements in the engineering blueprints were given in imperial units. Since MCNP6.2 input requires cm, all distances were converted using standard conversions.

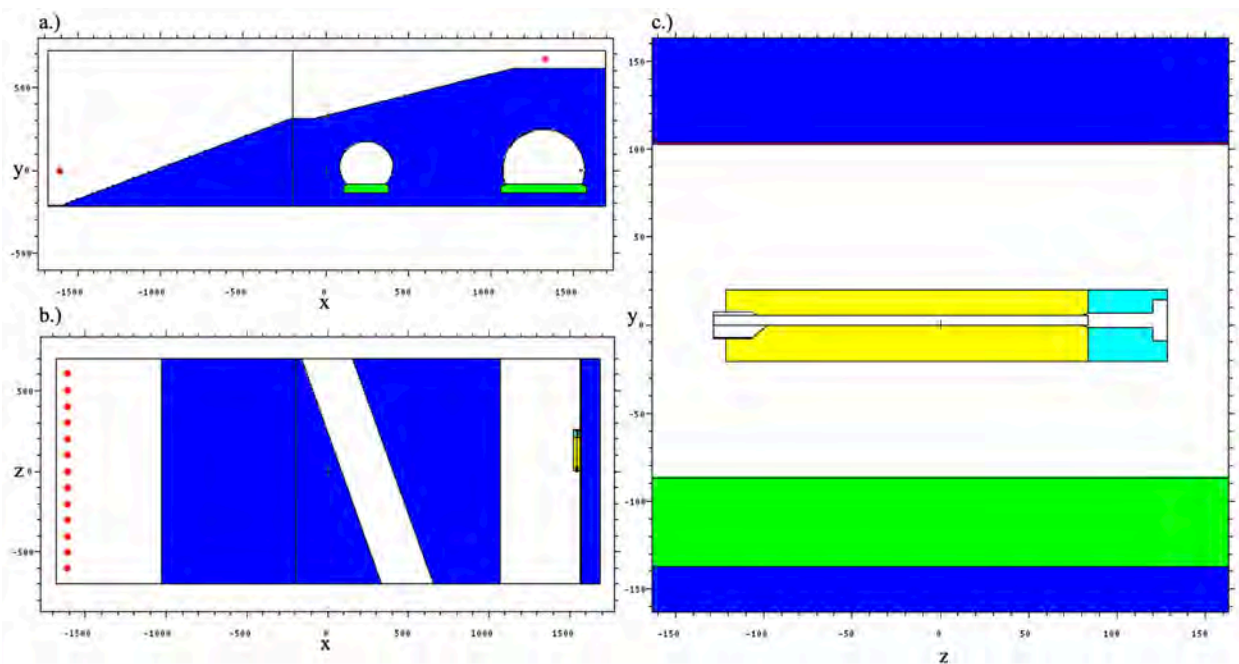


Fig. 3 Three different views of the beam dump site—red/pink dots represent detector tallies. Fig. 3a. shows a screenshot of the xy plane of the service tunnel (left) and the RHIC collider tunnel (right) at  $z=0$ . Note there are detectors at  $y=0$  and  $y\sim 600$ . Fig. 3b shows the xz plane at  $y=0$  and the approximate of detector tallies. Fig. 3c shows the beam dump design in the yz plane.



With an understanding of the geometry, models were then constructed in MCNP6.2 using a simplified list of materials. The list of all materials used for the simulations included aluminum, brass, concrete, copper, reinforced carbon, soil, and steel. Upon successful creation of the geometric model and thorough examination for any undefined regions, simulations began. Tallies were used to gauge where the regions of high radiation occurred within certain areas of interest. Most tally simulations were used to gauge the radiation dosage at the closest areas persons would be present outside the shielding. After determining where the regions of high radiation occurred, mesh tallies were used. Both tally and mesh tallies counted only a single type of particle for any given simulation. Photon or neutron radiation extracted from the simulation and analyzed.

## **IV Results**

### **Hadronic Collimator Results**

The simulations for the hadronic collimator sought to estimate the dosage outside the experimental hall. Two different collimator materials were compared to see which would produce higher doses: copper and reinforced carbon. Calculations were done using F5 neutron tally detectors placed at beam height ( $y=0$ ) with an x-distance of  $\sim 1777$  cm every 100 cm in the z-direction, as indicated in Fig. 2b. One-hundred thousand (100,000) particles were simulated. The results of these simulations are plotted in Fig. 4. Standard flux-to-dose calculations were used to convert from flux to rem/particle.

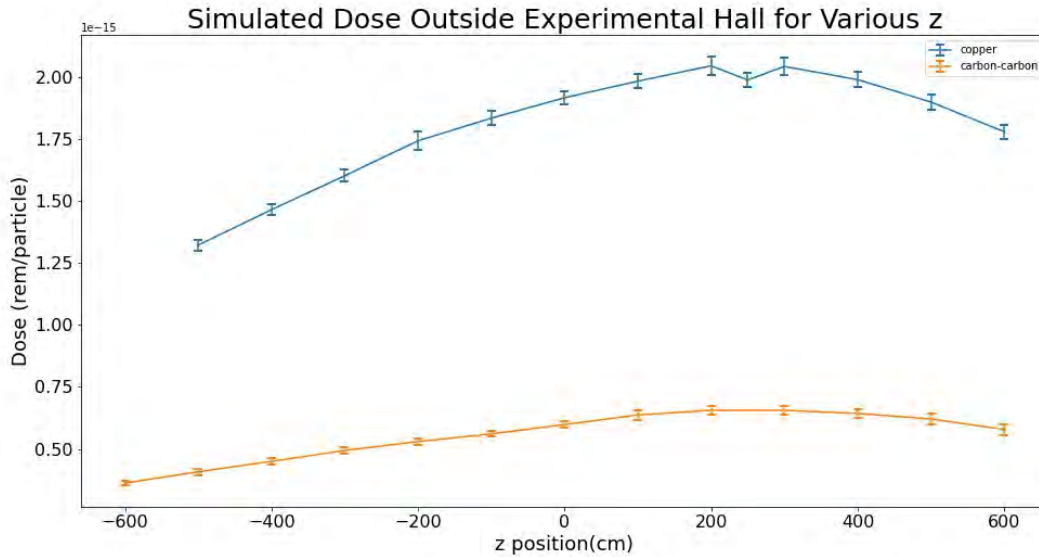


Fig. 4 Results of simulations run with one-hundred thousand particles simulated. All F5 tallies were placed with the same xy coordinates, (x,y)=(1777.30875, 0).

## EIC Electron Beam Dump Results

Simulations to estimate the radiation dose for the electron beam dump included two regions: inside the collider ring beyond the berm (red dot in Figure 3a.) and above the topsoil directly over the EIC tunnel (pink dot in Figure 3a.). For both sets of simulations, F5 photon and neutron detector tallies were placed every 100 cm in the z-direction. Tallies for the inside of the collider ring were placed at beam height (y=0) and an x-distance of -2910 cm. Similarly, tallies for above the topsoil were placed at the geometric center of the RHIC collider tunnel (x=0) and a height of y=714.9788 cm, exactly 100 cm above the berm. Standard flux-to-dose calculations were used to convert from flux to rem/particle for both photons and neutrons. Only results for neutron dose results are shown in Figure 5. because photon doses were significantly smaller than neutron doses.

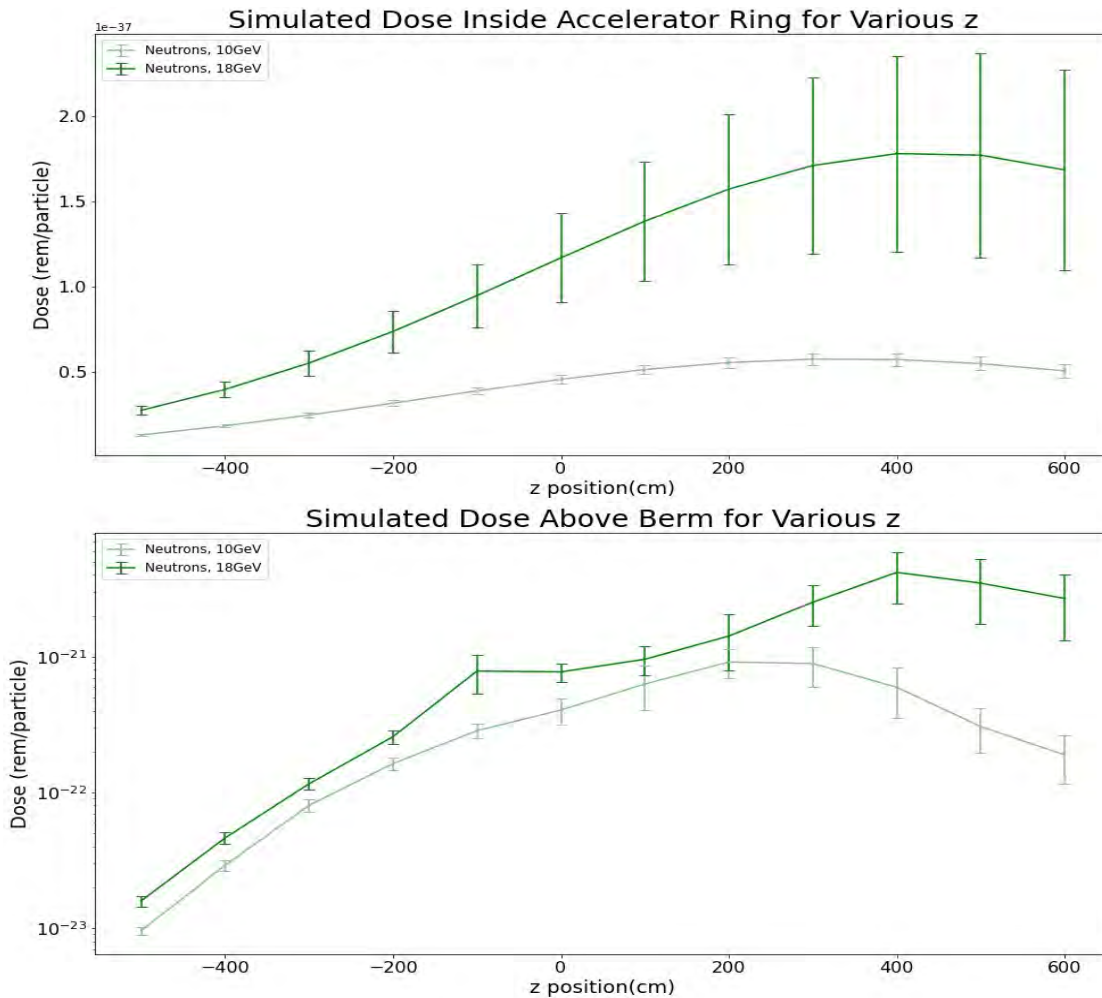


Fig. 5 Simulation results for dosage exposure within RHIC collider ring. Photons results are not displayed as all results were considerably lower than those of the neutrons. For both regions, the 18 GeV source, shown in dark green, sees higher dosage results.

The final simulations run estimated the soil activation near the beam dump using rectangular mesh tallies. The xy mesh tally was given the parameters given in Table 1. One-hundred million ( $1 \times 10^8$ ) particles were simulated twice: once with a 10 GeV source and another with an 18 GeV source. Results are plotted in Figure. 6.

Table 1

	X Range, Bins	Y Range, Bins	Z Range, Bins
<b>XY Mesh Tally</b>	[-500, 371.48], 100	[-220, 500], 100	[95, 105], 0

Table 1 The ranges used for the XY mesh tally to examine soil activation outside the beam dump.

## Simulated Soil Activation Around Electron Beam Dump in XY Plane

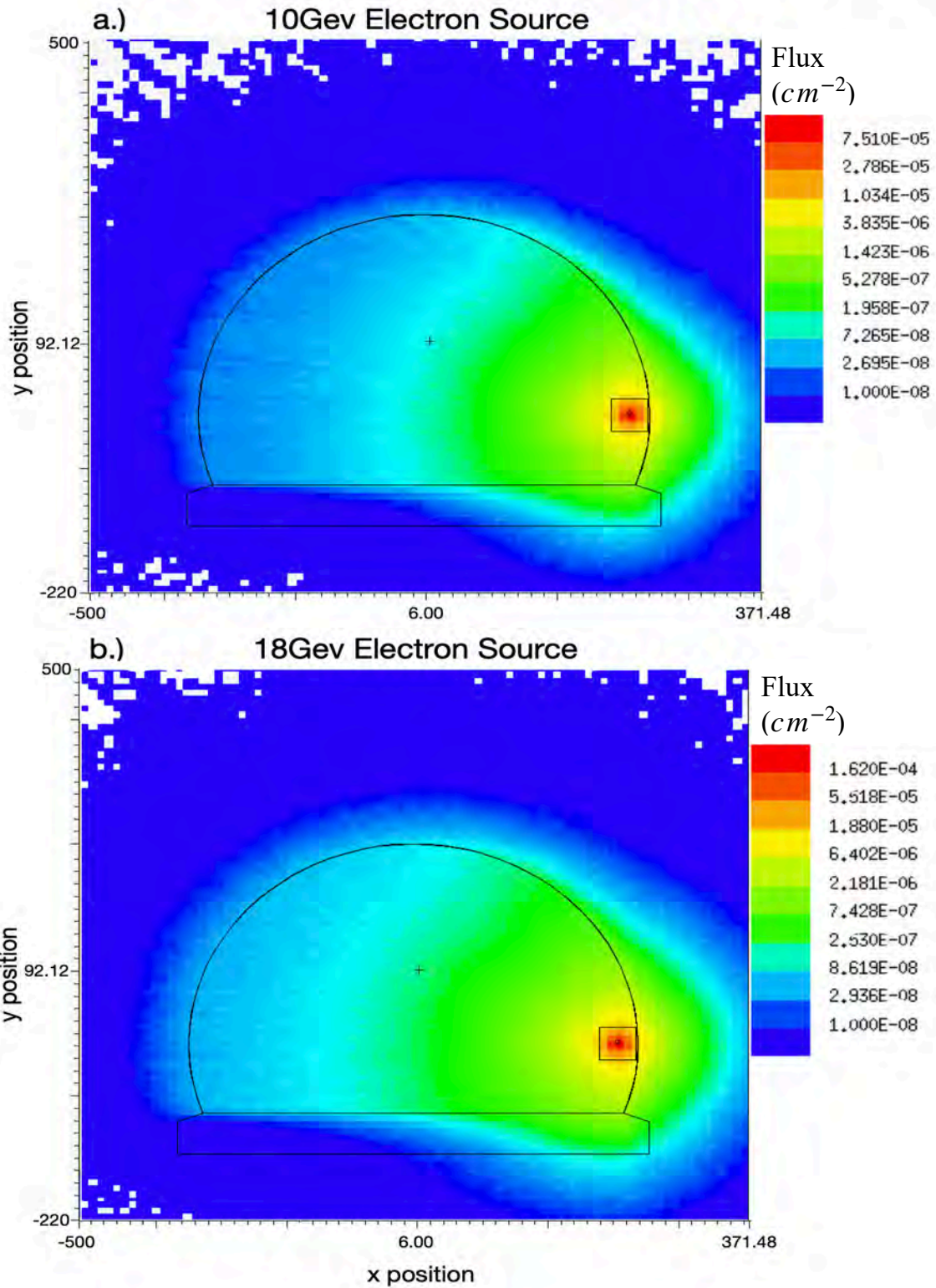


Fig. 6 Dosage displayed as a heat map in the xy plane overlaid onto the RHIC tunnel where Fig 6a. shows results with a 10 GeV source and Fig 6b. with an 18 GeV source.

## V Discussion

The values resulting from these simulations give general estimates for the radiation resulting from conditions outside the experimental hall area and various regions of the beam dump site. Using estimates for the number of particles present at each location, these values become more meaningful.

The dose rate around the experimental hall at interaction point 12 can be calculated with additional information courtesy communications with Angelika Drees<sup>2</sup> which says the maximum number of protons passing the collimator will be  $5 \times 10^{13}$  protons of subject to 1% routine loss per hour to 10% maximum loss per hour.

$$n_p l r = 5 \times 10^{13} \cdot 0.01 \cdot 2.04 \times 10^{-15} = \boxed{1.02 \text{ mrem/hour}}, \quad \text{Eq. 1}$$

where  $n_p$  is the number of protons,  $l$  is the percent loss/hour, and  $r$  is the radiation dose in rem/particle. The maximum simulated radiation dose is used for this calculation to place an approximate upper bound on the expected dose. With a routine dose of  $\sim 1$  mrem/hour, the shielding around the experimental hall can be said to be sufficient as this is similar to receiving a dental x-ray (around 1.5 mrem)<sup>4</sup>. Even if a fault were to occur (a worst case, one-time 10% loss), a dose of  $\sim 10$  mrem would be expected, which is the equivalent of receiving a chest x-ray<sup>4</sup>.

The number of electrons dumped per hour can also be calculated thanks to information courtesy of Christoph Montag<sup>5</sup> stating that,

“The stored current in the electron ring is 2.5 A at 10 GeV and below, and 250 mA at 18 GeV. Up to 10 GeV we have 1160 bunches, and 290 bunches at 18 GeV. We replace 2 bunches per second at any energy.”

Using this information and the period of revolution approximately is  $1.27888 \times 10^{-5}$  seconds with 2 bunches replaced per second, the number of dumped electrons/hour at 10 GeV can be easily calculated as shown in Eq. 2.

$$n_e = 2 \left( \frac{3600IT}{q_e n_b} \right) = 2 \left( \frac{3600 \cdot 2.5 \cdot 1.27888 \times 10^{-5}}{1.6 \times 10^{-19} \cdot 1160} \right) = 1.24 \times 10^{15} \text{ electrons/hour, Eq. 2}$$

where  $n_e$  is the number of electrons/hour, I is the stored current in the electron ring, T is the period of revolution,  $n_b$  is the number of bunches present for a given energy, and  $q_e$  is the fundamental charge constant. The same can also be done for 18 GeV yielding  $4.95 \times 10^{14}$  electrons/hour. Using Eq. 1 with  $l=100\%$ , the doses are calculated in Table 2. These doses are smaller than simulated doses at the collimator.

Table 2

	10 GeV (mrem/hour)	18 GeV (mrem/hour)
Above Berm	$1.13 \times 10^{-3}$	$2.06 \times 10^{-4}$
Inside Ring (parking lot)	$7.16 \times 10^{-20}$	$8.82 \times 10^{-20}$

Table 2 Simulated dosage results surrounding the beam dump for 10 GeV and 18 GeV. “Above Berm” refers to the pink dot in Figure 3a. and “Inside Ring” refers to red dot in the same figure.

Finally, soil activation surrounding the beam dump can be estimated using values approximated from the mesh tally plots shown in Fig. 6. The estimate for maximum flux outside the steel collider tunnel is approximately  $5 \times 10^{-6} \text{ cm}^{-2}$  at 10 GeV and  $1 \times 10^{-5} \text{ cm}^{-2}$  at 18 GeV. Then the soil activation in atoms/cc/yr can be calculated.

$$\frac{\phi}{\lambda} \cdot N_l \cdot N_e = \frac{5 \times 10^{-6}}{40} \cdot 0.075 \cdot (n_e \cdot 24 \cdot 180) \approx 5.02 \times 10^{10} \text{ atoms/cc/yr, Eq. 3}$$

where  $\lambda$  is the interaction length (cm),  $\phi$  is the flux,  $N_i$  is the number of atoms per interaction (0.075 for  $^3\text{H}$ ) and  $N_e$  is the number of electrons/year. Eq. 3 assumes the collider runs 24 hours a day for 180 days out of the year. For 18 GeV, the soil activation is  $4.01 \times 10^{10}$  atoms/cc/yr. Using BNL's standard leeching model of tritium production<sup>6</sup>, the soil activation values can be easily converted into the standard pCi/L/yr values<sup>7</sup>. The soil activation values, both of which violate the 5% of drinking water or 1000 pCi/L standard, are listed in Table 3.

Table 3

	10 GeV (pCi/L/yr)	18 GeV (pCi/L/yr)
Soil Activation	$2.64 \times 10^6$	$2.11 \times 10^6$

Table 3 Soil activation results calculated taken from results featured in Fig 6. Flux values are the best visual approximations of the highest flux values occurring within the soil outside the tunnel.

These results indicate that best estimates suggest dosages at the closest areas persons might pass during operation (parking lot, above berm, and inside collider ring) are within safe ranges. While shielding can be added if desired, simulations indicate that the current shielding is satisfactory given current expectations of the running parameters of EIC operation. However, soil activation requires additional measures to be taken. Construction of an impenetrable cap to divert water away from contaminated areas and prevent radioactive isotopes from seeping into the water table is the easiest way to prevent radiation from the beam dump from causing significant harm to the populous and the environment.

## VI References

- [1] R. Tulipano: RHIC Construction Major Facility at 12:00, blueprints A3001B0101 and A3001B0102, c/o Kin Yip private communications, 1997
- [2] Angelika Drees: private communications, 2021
- [3] Christopher J Werner: MCNP® USER'S MANUAL Code Version 6.2, chp. 1-3, 5, 2018.
- [4] NRC: Doses in Our Daily Lives, <https://www.nrc.gov/about-nrc/radiation/around-us/doses-daily-lives.html>, 2021
- [5] Christoph Montag: private communications, 2021
- [6] SBMS: Design Practice for Known Beam-Loss Locations (BLLs), <https://sbms.bnl.gov/Uploads/SA/40/10.2/1R09E011.PDF>, 2012
- [7] Kin Yip: Calculators for Various Radiation Considerations, <https://www.c-ad.bnl.gov/kinyip/Radiation/Calculators.html>, 2019

## VIII Acknowledgements

I want to extend a special thank you to my mentor, Kin Yip, for the invaluable knowledge and insight, both personal and career, he offered throughout this project. Additionally, I would like to thank the Hills/Thurber family for hosting me during the first half of this project and my family in South Dakota for putting up with me for the second half.

This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).



## Deployment and Repository Deposition of Multiple Software Packages Using Automation

### Student Intern

John Gedell, Computer Science, St. Joseph's College, Patchogue, NY 11772

### Lab Mentor

Maksim Rakitin, Data Science and Systems Integration Program of NSLS-II, Brookhaven National Laboratory, Upton, NY 11973

## Abstract

The Data Science and Systems Integration Program of NSLS-II deploys software to the experimental floor of NSLS-II. The entire network consists of more than 100 workstations and servers. In the past, Brookhaven National Lab used individual custom collection environments for each workstation or server. This method of package deployment is a very inefficient use of both cost and time and ultimately leads to an increased workload for both the scientists and the programmers. Additionally, individual deployments are far more apt to failure than a singular, central deployment. It is far more efficient to create a single environment for all workstations and servers. Automation of this entire process would be even more efficient and can allow the developer to focus on other tasks. Furthermore, there is no detailed reporting tool to show the results, in terms of testing and errors, of each step of the deployment process. Therefore, a tool has been created that automates the creation of a single conda environment for deployment of various packages. The creation of such a tool would allow for a consistently dependable process that would greatly benefit researchers and their ability to collect and input data. It would also allow for a smoother deployment process. The design of this process can be completed by using a Virtual Box with a Vagrant environment and a Linux server and then using Python and Docker to package the files. The files were completed in Python, Docker, and Linux shell languages and then templated using Jinja. After these steps are completed, they are sent off to Zenodo for storage. Additionally, the images can be stored on hubs like Dockerhub, Quay, and GHCR. The project was successfully completed this summer. Multiple files can now be successfully deployed from a central location with minimal user input. While I did have experience with Python and Linux prior to this summer, I have now added all of the other aforementioned tools

to my repertoire of automation techniques. I previously had no knowledge of or experience with Docker or Jinja. Furthermore, this project was the first large-scale automation tool that I've developed.

## **Background**

### **A. BNL Experimental Flaws and the Downside of a Decentralized Network**

The NSLS-II experimental floor contains 28 experimental beamline stations and over 100 total workstations and servers.<sup>1</sup> To keep the data acquisition and analysis software available and up-to-date on the beamline workstations, the deployment routine is performed by the DSSI program during each shutdown in the NSLS-II operation, typically 3 times per year. In the past, each deployment was operated as a single, individual process that was disconnected from the rest of the work stations. This occurred for several reasons. First, the workstation local configurations may be slightly different, and the same conda installation command is not consistent and sometimes may result in failing packages. Also, the conda dependency resolution mechanism is really slow, so our goal is to do it once when we create the conda-packed environment, and then just unarchive it on the machines. This cuts the time from up to 30 minutes for a single deployment to ~3 minutes per machine. If the machines are not very powerful in terms of CPU/RAM, the difference can be even larger. A decentralized deployment process consisting of an individual, manual process can lead to multiple system failures. Furthermore, a decentralized network can lead to significant inefficiencies in both time and cost and may lead to an increased workload for both scientists and programmers.<sup>2</sup>

### **B. Tools that can be used to Create an Automated Centralized Network**

An alternative to a decentralized network is a centralized network. One benefits of using

this kind of network is that it can be easily automated with various scripting tools. One such tool is the Python programming language. Python has become popular in recent years, primarily due to its ease of use and other important functionalities. Because Python code does not require compilation, it can be easily set up on different machines and then quickly tested.<sup>3</sup> Additionally, Python is known as a flexible programming language that does not conform to specific programming paradigms. It allows non-programmers to read and understand the code much better than Java or C++. A final benefit of using Python is its wide network of libraries that can be easily accessed and understood. Since Python is a mature language, it has very good documentation and resources as well.

A second scripting tool that is commonly used is Docker. Docker is an open source platform for building, deploying, and managing containerized applications.<sup>4</sup> With Docker, a developer can package applications into containers. Containers combine the source code with the operating system and dependencies that allow the files and processes to be executed in any environment. Containers have become increasingly popular due to their ease of use in cloud environments.

Docker containers have several benefits over traditional methods of storing multiple resources for use, specifically virtual machines. First, they are much smaller than virtual machines. Virtual machines tend to store resources in gigabytes, whereas Docker containers store resources in megabytes.<sup>4</sup> It also reduces time cost by allowing programmers to develop, deploy, and provision these resources far more quickly. Therefore, these programmers will have time to work on other projects. This is an important aspect of continuous integration, which has also been growing in popularity in the software development field. Docker containers also allow

for greater resource efficiency in that several copies of the program can be run on the same machine at the same time without taking up significant space. Furthermore, developers and CEO have reported that applications have made it to market 76% times faster than when they used virtual machines. Developers and users have reported a 73% decrease in application downtime as the Docker containers continue to improve in function.<sup>4</sup>

In addition to containers, Docker has several other tools, including Docker images, Dockerfile, and the DockerHub repository. Docker images are basically single instances of Docker containers. Dockerfiles automate the process of creating Docker images. DockerHub is a public repository for Docker images where users can freely upload, download, and share these images.

The third scripting tool that was used is called Jinja. Jinja is a fast templating engine that similar to Python in syntax.<sup>5</sup> In fact, Jinja is a templating language that was built specifically for Python.<sup>6</sup> Jinja templates can be created from various files and are then passed data to render the final document. This templating language can be used to create elegant code very quickly.

The fourth scripting tool is Linux. Linux has been in use for a very long time, originating as a free UNIX-like operating system in the early 1980s.<sup>6</sup> It has been evolving since then and became widely used in the mid-1990s. Linux has several derivatives, one of which is called Ubuntu, which is the primary operating system used in the development of the system described in this paper. The primary way that a user interacts with a Linux system is through a command-line interface called a bash shell.

There are several reasons why Linux is important in the creation of automation tools.

Linux's command-line interface is the basis of that of more modern operating systems.<sup>6</sup>

Therefore, many systems will use Linux shells. Additionally, many modern companies are incorporating full Linux distributions in their network equipment. Therefore, it is likely that Linux will become even more relevant in both networking and automation as a whole. Many Python scripts, as well as other systems, such as Ansible, usually run their networks on Linux, as opposed to their own networks directly.

### C. The Process by which we Created Our Centralized Automation System

In order to achieve our goals of creating a centralized automated system, we have to combine all of the aforementioned tools, along with a few others. I was able to accomplish our ultimate goal on my Windows operating system by downloading VirtualBox, a virtual machine software, and Vagrant, open-source software containing portable environments. I created the Vagrant environment through Ubuntu 20.04, the latest fully developed option. I used these software products in unison with Windows Powershell.

Once the environment was established, I installed Docker. Next, I created a Dockerfile Jinja template that, when rendered, becomes a regular Dockerfile. The Dockerfile serves to import a specific environment from Conda Forge. Conda Forge is a subset of the Anaconda software that contains community-created recipes, build infrastructure, and distributions for the conda package manager.<sup>7</sup> Conda-pack is a command line tool for creating archives of conda environments that can be installed on other systems and locations. A tool like conda-pack is necessary because conda environments are not relocatable. This Conda Forge environment is then packaged together with a set of other required packages. Two linux shell scripts run consecutively to create a Docker image within a Docker container. This is a two-step process

where a Docker image is created based on the Dockerfile and a Docker container, based on the Docker image, is run to create artifacts. This Docker image can then be uploaded to different repositories. This can allow the image to be automatically uploaded to DockerHub, Quay, and GHCR (Github packages) if configured. The primary file that is created is a tar.gz file. Additionally, a .yaml file is created. This file is the spec for the conda environment used inside of that packaged environment. This way, users can easily access the contents without installing the actual environment. Other files that are created are text files containing checksums that are calculated with sha256 and md5 algorithms and used for validation. The files are then automatically uploaded to repository using a configuration file (that is entered via the command line) that corresponds to the particular conda environment being built and published. The Zenodo deposition is used for automated testing of the beamline software via profile collections CI. This repository, Zenodo, is an international repository that allows researchers to deposit research papers, data sets, research software, and many other research related items with sizes up to 50 GB. It creates a static, unremovable deposition with a DOI (digital object identifier) that can be cited in papers.

One of this project's creations, the Zenodo uploader, has various functions. It has a built-in search function where the user is able to parse the metadata of a file by its title, creator, or owner. This way, the file can be located quickly and easily. Additionally, it can upload up to three files at one time to the Zenodo repository. Finally, given the fact that the user may not wish to automatically publish content that is uploaded to Zenodo, it gives the user the option to do so via the command-line interface. If the user chooses to immediately publish the upload, they are able to without further input.

## Works Cited

1. “National Synchrotron Light Source II.” *Brookhaven National Laboratory*, [www.bnl.gov/nsls2](http://www.bnl.gov/nsls2). Accessed 1 Aug. 2021.
2. X. Wang, X. Liu, B. Zhang, H. Wu and S. Chen, "Research on Evaluation Methods of Decentralized and Centralized Operation Quality of Computer Networks," *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 2020, pp. 300-305, doi: 10.1109/ITNEC48623.2020.9084686.
3. Tantra, Juki Wirawan. “Experiences in Building Python Automation Framework for Verification and Data Collections.” *The Python Papers Monograph 2:17 Proceedings of PyCon Asia-Pacific 2010*, 2010, [citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.686.9485&rep=rep1&type=pdf](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.686.9485&rep=rep1&type=pdf).
4. IBM Cloud Education. “Docker.” *IBM*, 23 June 2021, [www.ibm.com/cloud/learn/docker](http://www.ibm.com/cloud/learn/docker).
5. “Jinja.” *Jinja*, [jinja.palletsprojects.com/en/3.0.x](http://jinja.palletsprojects.com/en/3.0.x). Accessed 1 Aug. 2021.
6. Edelman, Jason, et al. *Network Programmability and Automation*. O’Reilly Media, Inc., 2018.
7. “Conda-Forge.” *Anaconda*, [anaconda.org/conda-forge](http://anaconda.org/conda-forge). Accessed 1 Aug. 2021.





Office of Educational Programs  
Building 438  
P.O. Box 5000  
Upton, NY 11973-5000

[www.bnl.gov](http://www.bnl.gov)