

SAMPLE PROPOSAL

Student Name
Faculty Sponsor
On-Site Supervisor, Title, Organization
Date
Experiment Websites:
<https://astroparticle.rice.edu/>
<http://www.xenon1t.org/>

Fast Integration of Poisson Distributions for Dead Sensor Marginalization

Define exactly what the student will be doing as an intern. Specify goals and objectives and how background and responsibilities will help fulfill these objectives.

In this internship, I will be continuing the work I did this past summer for my IRIS-HEP fellowship, where I worked on writing a data analysis program for the XENON project. The XENON project is an experiment that uses a particle detector to attempt to detect dark matter. The particle detector used in this research records the position of particles passively traveling through the xenon. The detector uses many different sensors to track photons and their locations in relation to one another by recording the electrical signals the particles trigger. My project involves writing a program to figure out what correlations exist between the number of photons detected by each sensor in the particle detector based on their position in the array of sensors. This information will be used to help discern where in the detector the interaction took place, and will help to be able to fill in gaps of information left by sensors that have broken.

Calculating these correlations requires finding a numerical integral over a Poisson distribution in one or more dimensions. These computations can be very complex, and as the graph may consider correlations with at least the seven sensors adjacent to each sensor, it can be a high dimensional problem. Computing this is unfortunately very costly and would take a long time to do with the amount of data being collected within this experiment, and potentially require an unreasonable amount of data (up to 38,000,834 Gigabytes for 6 broken sensors).

I spent most of last summer working on reviewing concepts required for this project, including math (Bayesian networks, Poisson distributions, multidimensional integration), found and tested function approximations to improve run-time and avoid issues such as overflow and underflow while maintaining the necessary amount of precision. I also learned optimization techniques beyond what we've used in classes, such as code profiling methods like snakeviz, learning to time programs and using matplotlib to compare them. I ended the summer with a mostly functional program, but did not have time to properly research and implement two additional parts of the program, which I would like to work on adding this fall. One of them is zarray, a Python package used to compress large multidimensional arrays (which is what we are using to store the large Poisson distributions), as well as xarray, which we want to use to simplify operations involving multidimensional arrays. We also want to implement a look-up table to speed up the Poisson distribution generation, but did not have time to work on finding the most efficient way to do this. I would also like to do a more in-depth analysis of how the program runs after I have made these changes. Lastly, once the program is at a good point, I will write documentation to explain what the package does and how to use it.

I will meet with my on-site supervisor once a week on zoom to go over my progress, and can use slack/email to answer any questions that come up in between meetings, and report my working hours.

Indicate how the internship will be integrated with a liberal arts education and how it will contribute to intellectual and personal growth.

This is a great learning experience, both in terms of practical Computer Science knowledge and getting an opportunity to do interdisciplinary work involving data analysis in a research field (something I've been interested in pursuing after college) and getting programming advice from those in other disciplines has been surprisingly different than I'm used to, and I'd like to continue learning from them. This project is the first I've worked on with this level of independence and the first with real data.

Specify how much and by what means communication will occur with the faculty sponsor during the internship (remember that a mid-point written communication between the sponsor and the student is required).

I will write weekly updates about my progress to send to my faculty sponsor, Professor Eckmann.

Describe the exhibit, project, term paper, or other materials that will be presented to the faculty sponsor for evaluation of the completed internship. This material will usually include some expository writing. The amount of the assigned coursework should be commensurate with the number of credit hours for the internship.

The material I will have to present at the end of this will be documented code for a program that described above and a final write-up of the entire project.