IRIS-HEP Undergraduate Fellowship Proposal

Emily Tsai, University of Texas at Austin

Project: OpenCL based implementation of a graph neural network on an FPGA
Supervisor: Dr. Savannah Thais, Princeton University
Duration: January 11, 2021 - May 28, 2021 (20 weeks)

The CMS experiment at the LHC receives data at an extremely high rate (upwards of hundreds of gigabytes per second in Run 2), and this rate will significantly increase with the High-Luminosity LHC. As a result, sophisticated triggering systems are needed to quickly filter and then reduce the data down to manageable amounts. The CMS triggering system has a low-level trigger, which consists of mostly hardware, and a high-level trigger, which consists of software. Reconstructing particle trajectories, or tracking, is a crucial, but slow, step in the high-level trigger. Machine learning models can speed up the tracking process.

Hits in the CMS detector are geometric. Thus, a graph made up of nodes and edges can represent detector hits well. A machine learning model called a graph neural network (GNN) can be applied to these graphs to get the probabilities of each edge. A post-processing algorithm can then link the edges into tracks based on their probabilities. FPGAs can be used to accelerate GNN tracking. However, GNNs have yet to be implemented fully on an FPGA.

This project focuses on creating an OpenCL implementation of a GNN completely on an FPGA, rather than the current implementation of the GNN on CPU and FPGA coprocessors. After becoming familiar with the project setup, I will first investigate reducing the size of the GNN in order to fit the entire GNN on an FPGA. Next, I will focus on implementing the GNN fully on an FPGA. Finally, I will analyze the performance of the GNN implemented on an FPGA versus the performance of the GNN implemented on CPU and FPGA coprocessors.

I am currently an undergraduate senior. I have previously done work on jet reconstruction algorithms for the ATLAS experiment's low-level trigger and on detector simulations for the proposed KLEVER experiment at the CERN SPS. This project will further my software skills and research experience in preparation for my goal of earning a Ph.D. in particle physics with a focus on scientific computing. The knowledge I will gain in using FPGAs and machine learning will be applicable to work I can contribute to during graduate school and beyond. Additionally, slow tracking time is a prevalent issue in many HEP experiments, and working to reduce tracking time can help HEP experiments other than CMS. Finally, work towards optimizing machine learning calculations would help many fields outside of particle physics that use machine learning.

Proposed Timeline: This fellowship will take place part-time in the Spring 2021 semester alongside a part-time student course load.

- **Week 1-4**: Set up project and become familiar with OpenCL implementation of GNNs on CPU and FPGA coprocessors. Repeat previous work done on the coprocessors.
  - **Deliverable**: Replication of previous work.
- **Week 5-9**: Investigate and optimize size reduction of GNN to fit fully on an FPGA.
  - **Deliverable**: Report on results of GNN size reduction.
- **Week 10-14**: Implement the GNN on an FPGA.
  - **Deliverable**: Report on results of GNN implementation on FPGA.
- **Week 15-19**: Analyze the performance of the GNN implemented on an FPGA versus the performance of the GNN implemented on CPU and FPGA coprocessors.
  - **Deliverable**: Analysis of GNN performance on FPGA versus on coprocessors.
- **Week 20**: Document results and give presentation(s).
  - **Deliverable**: A written report and presentation on work done in the fellowship.