Contribution ID: **14** Type: **not specified**

Intelligent experiments through real-time AI: Fast Data Processing and Autonomous Detector Control for sPHENIX and future EIC detectors

Tuesday, 4 October 2022 14:15 (15 minutes)

The upcoming sPHENIX experiment, scheduled to start data taking at the Brookhaven National Laboratory (BNL) Relativistic Heavy Ion Collider in 2023, and the future Electron-Ion Collider (EIC) experiments will employ sophisticated state-of-the-art, high-rate detectors to study high energy heavy ion and electron-ion collisions, respectively. The resulting large volumes of raw data far exceed available DAQ and data storage capacity.

The application of modern computational techniques, in particular recent developments in artificial intelligence (AI) and machine learning (ML), has the potential to address these new challenges and revolutionize our approach to collecting, reconstructing and understanding data, and thereby maximizing the discovery potential. Our proposal seeks to transform future flagship detectors into "intelligent experiments"and "smart data acquisition and control"through the integration of next generation AI/ML hardware, electronics and algorithms into these detector systems.

We propose to develop a selective streaming readout system, built on state-of-the-art readout electronics and computing, to achieve continuous readout and inspection of all essential detector signals. AI algorithms will be used to reduce the raw data volume on the fly by identifying events with heavy flavor jets, through real-time detector control, reconstruction and event selection.

The tracking detectors will all use the BNL-designed Front-End Link eXchange (FELIX) FPGA card as a data aggregator. With the development of advanced deep neural networks, a parallel strategy is needed to ensure that these networks can be designed to operate at low latency and high throughput on the FELIX FPGA cards. Developing and implementing these techniques in the context of scientific low-power, low-latency, or resource constrained use cases is a major goal of this research program.

In this proposal we will use hls4ml to integrate AI models into the streaming system of sPHENIX. hls4ml will take a neural network model generated in a standard ML format (keras/tensorflow/pytorch/onnx) and translate this to an FPGA/ASIC synthesizable high level synthesis code. The generated code in high level synthesis language (HLS) is C++ based.

The challenge in creating an optimal digital design is to balance available resources with achieving the power, latency, throughput goals of the target application. In this talk, we will present details of the implementation and the latest progress of this project.

Primary author: RIGATTI, Micol (Fermi National Accelerator Lab. (US))

Presenter: RIGATTI, Micol (Fermi National Accelerator Lab. (US))

Session Classification: Contributed Talks