10th International Workshop on Boosted Object Phenomenology, Reconstruction and Searches in HEP (BOOST 2018)

Contribution ID: **44** Type: **Talk**

Fast inference of deep neural networks in FPGAs for particle physics

Thursday 19 July 2018 12:00 (25 minutes)

Recent results at the Large Hadron Collider (LHC) have pointed to enhanced physics capabilities through the improvement of the real-time event processing techniques. Machine learning methods are ubiquitous and have proven to be very powerful in LHC physics, and particle physics as a whole. However, exploration of the use of such techniques in low-latency, low-power FPGA hardware has only just begun. FPGA-based trigger and data acquisition (DAQ) systems have extremely low, sub-microsecond latency requirements that are unique to particle physics. We present a case study for neural network inference in FPGAs focusing on a classifier for jet substructure which would enable, among many other physics scenarios, searches for new dark sector particles and novel measurements of the Higgs boson. While we focus on a specific example, the lessons are far-reaching. We develop a package based on High-Level Synthesis (HLS) called \hlsfml˜to build machine learning models in FPGAs. The use of HLS increases accessibility across a broad user community and allows for a drastic decrease in firmware development time. We map out FPGA resource usage and latency versus neural network hyperparameters to identify the problems in particle physics that would benefit from performing neural network inference with FPGAs. For our example jet substructure model, we fit well within the available resources of modern FPGAs with a latency on the scale of 100˜ns.

Author: HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US))

Presenter: NGADIUBA, Jennifer (CERN)

Session Classification: Machine Learning