Contribution ID: **42**   Type: **Talk**

SONIC: Coprocessors as a service for deep learning inference in high energy physics

Wednesday, 2 December 2020 15:40 (6 minutes)

In the next decade, the demands for computing in large scientific experiments are expected to grow tremendously. During the same time period, CPU performance increases will be limited. At the CERN Large Hadron Collider (LHC), these two issues will confront one another as the collider is upgraded for high luminosity running. Alternative processors such as graphics processing units (GPUs) can resolve this confrontation provided that algorithms can be sufficiently accelerated. In many cases, algorithmic speedups are found to be largest through the adoption of deep learning algorithms. We present a comprehensive exploration of the use of GPU-based hardware acceleration for deep learning inference within the data reconstruction workflow of high energy physics. We present several realistic examples and discuss a strategy for the seamless integration of coprocessors so that the LHC can maintain, if not exceed, its current performance throughout its running.

Primary authors: KRUPA, Jeffrey (Massachusetts Institute of Technology);   HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US))

Presenter: RANKIN, Dylan Sheldon (Massachusetts Inst. of Technology (US))