Contribution ID: **219** Type: **not specified**

Coprocessors as aservice for accelerated inference of DL algorithms

Friday, March 19, 2021 1:20 PM (15 minutes)

DL algorithms have been widely adopted in LHC analyses. Use of DL in triggering, however, has been limited because of stringent throughput limitations and limited hardware availability. This limitation is compounded as LHC physicists look for new particles below existing trigger thresholds. We begin to address these challenges by developing SONIC, a tool for running DL algorithms into CMS workflows. SONIC leverages highly optimized DL inferences on GPUs, and it allows application-specific integration of DL algorithms onto heterogeneous clusters with minimal changes to workflows. In this talk, we will present the first scaled tests of realistic DL algorithms integrated in a HEP High Level Trigger system using GPUs and FPGAs as a service, and discuss ongoing efforts in HEP and beyond.

Primary author: KRUPA, Jeffrey

Presenter: KRUPA, Jeffrey

Session Classification: TDAQ

Track Classification: TDAQ