Contribution ID: **17**   Type: **Standard Talk**

Fast ML inference in FPGAs for the Level-1 Scouting system at CMS

Monday 25 September 2023 15:45 (15 minutes)

A novel data collection system, known as Level-1 (L1) Scouting, is being introduced as part of the L1 trigger of the CMS experiment at the CERN LHC. The L1 trigger of CMS, implemented in FPGA-based hardware, selects events at 100 kHz for full read-out, within a short 3 microsecond latency window. The L1 Scouting system collects and stores the reconstructed particle primitives and intermediate information of the L1 trigger processing chain, at the full 40 MHz bunch crossing rate. Demonstrator systems consisting of PCIe-based FPGA stream-processing boards and associated host PCs have been deployed at CMS to capture the intermediate trigger data. An overview of the new system, and results from Run 3 data taking will be shown. In addition, a neural-network based re-calibration and fake identification engine has been developed to improve the quality of the L1 trigger objects for online analysis. We utilise new solutions for creating portable, flexible, and maintainable ML inference implementations that are accessible to those without hardware design knowledge. Tools such as High Level Synthesis languages, soft core libraries, and custom ML inference compilers will be discussed and explored. The results of these strategies as used for the CMS L1 trigger scouting system will be presented.

Author: JAMES, Thomas Owen (CERN)

Presenter: JAMES, Thomas Owen (CERN)

Session Classification: Contributed Talks

Track Classification: Contributed Talks