Contribution ID: **290** Type: **presentation**

Fast Boosted Decision Tree inference on FPGAs for triggering at the LHC

Tuesday, 10 July 2018 14:45 (15 minutes)

Boosted Decision Trees are used extensively in offline analysis and reconstruction in high energy physics. The computation time of ensemble inference has previously prohibited their use in online reconstruction, whether at the software or hardware level. An implementation of BDT inference for FPGAs, targeting low latency by leveraging the platform's enormous parallelism, is presented. Full pipelining provides high throughput, and the use of a high-level programming language adds flexibility to construction of new ensembles. Classification latencies of tens of nanoseconds are achieved within the FPGA, as well as a 600 times speedup over inference on a single threaded CPU with the FPGA as a coprocessor. A use case within the CMS Level 1 track trigger is presented, demonstrating use of the implementation on a real classification problem, including resource and latency aware hyperparameter tuning.

Primary author: SUMMERS, Sioni Paris (Imperial College Sci., Tech. & Med. (GB))

Presenter: SUMMERS, Sioni Paris (Imperial College Sci., Tech. & Med. (GB))

Session Classification: T1 - Online computing

Track Classification: Track 1 - Online computing