Contribution ID: **303** Type: **Oral**

FPGA-accelerated machine learning inference as a solution for particle physics computing challenges

Tuesday 12 March 2019 19:00 (20 minutes)

Resources required for high-throughput computing in large-scale particle physics experiments face challenging demands both now and in the future. The growing exploration of machine learning algorithms in particle physics offers new solutions to simulation, reconstruction, and analysis. These new machine learning solutions often lead to increased parallelization and faster reconstructions times on dedicated hardware, here specifically Field Programmable Gate Arrays. We explore the possibility that applications of machine learning simultaneously also solve the increasing computing challenges. Employing machine learning acceleration as a web service, we demonstrate a heterogeneous compute solution for particle physics experiments that requires minimal modification to the current computing model. First results with Project Brainwave by Microsoft Azure, using the Resnet-50 image classification model as an example, demonstrate inference times of approximately 50 (10) milliseconds with our experimental physics software framework using Brainwave as a cloud (edge) service. We also adapt the image classifier, for example, physics applications using transfer learning: jet identification in the CMS experiment and event classification in the Nova neutrino experiment at Fermilab. Solutions explored here are potentially applicable sooner than may have been initially realized.

Primary authors: TRAN, Nhan Viet (Fermi National Accelerator Lab. (US)); PIERINI, Maurizio (CERN); HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US)); PEDRO, Kevin (Fermi National Accelerator Lab. (US)); DUARTE, Javier Mauricio (Fermi National Accelerator Lab. (US)); KREIS, Ben (Fermi National Accelerator Lab. (US)); RANKIN, Dylan Sheldon (Massachusetts Inst. of Technology (US)); JINDARIANI, Sergo (Fermi National Accelerator Lab. (US)); LIU, Miaoyuan (Fermi National Accelerator Lab. (US)); TSARIS, Aristeidis (Fermilab); NGADIUBA, Jennifer (CERN); Dr HOLZMAN, Burt (Fermi National Accelerator Lab. (US)); WU, Zhenbin (University of Illinois at Chicago (US))

Presenter: NGADIUBA, Jennifer (CERN)

Session Classification: Track 1: Computing Technology for Physics Research

Track Classification: Track 1: Computing Technology for Physics Research