

Increasing the LHC Computational Power by integrating GPUs as a service

Monday, 3 October 2022 14:05 (5 minutes)

Over the past several years, machine learning algorithms at the Large Hadron Collider have become increasingly more prevalent. Because of their highly parallelized design, Machine Learning-based algorithms can be sped up dramatically when using coprocessors, such as GPUs. With increasing computational demands coming from future LHC upgrades, there is a need to enhance the overall computational power of the next generation of LHC reconstruction. In this talk, we demonstrate a strategy to port deep learning algorithms to GPUs efficiently. By exploiting the as-a-service paradigm to port algorithms to GPU, we are able to optimally use GPU resources, allowing for a path towards efficient GPU adoption at the LHC as more algorithms become parallelizable. In this talk, we present this path and demonstrate an end-to-end workflow with current reconstruction using the Compact Muon Solenoid.

Primary authors: KHODA, Elham E (University of Washington (US)); DUARTE, Javier Mauricio (Univ. of California San Diego (US)); PEDRO, Kevin (Fermi National Accelerator Lab. (US)); LIU, Miaoyuan (Purdue University (US)); TRAN, Nhan (Fermi National Accelerator Lab. (US)); THOMAS, Nirmal; HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US)); KANSAL, Raghav (Univ. of California San Diego (US)); HSU, Shih-Chieh (University of Washington Seattle (US)); ROTHMAN, Simon (Massachusetts Inst. of Technology (US)); PIPEROV, Stefan (Purdue University (US)); MCCORMACK, William Patrick (Massachusetts Inst. of Technology (US)); FENG, Yongbin (Fermi National Accelerator Lab. (US))

Presenters: MCCORMACK, William Patrick (Massachusetts Inst. of Technology (US)); FENG, Yongbin (Fermi National Accelerator Lab. (US))

Session Classification: Contributed Talks