



Contribution ID: 51

Type: **Standard Talk**

Portable Acceleration of CMS Mini-AOD Production with Coprocessors as a Service

Monday 25 September 2023 14:30 (15 minutes)

Computing demands for large scientific experiments, such as the CMS experiment at CERN, will increase dramatically in the next decades. To complement the future performance increases of software running on CPUs, explorations of coprocessor usage in data processing hold great potential and interest. We explore the novel approach of Services for Optimized Network Inference on Coprocessors (SONIC) and study the deployment of this as-a-Service approach in large-scale data processing. In this setup, the main CMS Mini-AOD creation workflow is executed on CPUs, while several machine learning (ML) inference tasks are offloaded onto (remote) coprocessors, such as GPUs. With experiments performed at Google Cloud, the Purdue Tier-2 computing center, and combinations of the two, we demonstrate the acceleration of these ML algorithms individually on coprocessors and the corresponding throughput improvement for the entire workflow. We also show that this approach can be easily generalized to different types of coprocessors, and even deployed on local CPUs without performance decrease. We emphasize that SONIC enables high coprocessor usage and brings the portability to run workflows on different types of coprocessors.

Authors: Dr HOLZMAN, Burt (Fermi National Accelerator Lab. (US)); DUARTE, Javier Mauricio (Univ. of California San Diego (US)); KRUPA, Jeffrey (Massachusetts Institute of Technology); PEDRO, Kevin (Fermi National Accelerator Lab. (US)); GRAY, Lindsey (Fermi National Accelerator Lab. (US)); ACOSTA FLECHAS, Maria (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)); 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))

Presenter: MCCORMACK, William Patrick (Massachusetts Inst. of Technology (US))

Session Classification: Contributed Talks

Track Classification: Contributed Talks