TDAQ group report

A. Gioiosa (Pisa), G. Pezzullo (Yale)

for the Mu2eII tdaq subgroup

Architectures under study

Two TDAQ architectures proposed so far:

1. **2-level Trigger (L1 Trigger + HLT)**
2. Software Trigger using GPUs

L1 + High Level Trigger

- Aggregate the data into a board equipped with multiple FPGAs
- Run the early stage of the track reconstruction + full calorimeter reconstruction
- Apply a L1 decision and move data to the HLT farm which runs full track reco

- **Main focus:**
  a. Do some processing on FPGA and the remainder on software
  b. Can we make a L1 trigger decision at FPGA level?
  c. Track pattern-recognition on FPGA?
  d. **Need to develop FPGA algorithms**

L1 + High Level Trigger: where do we stand?

- Simulated datasets from the Mu2e-II software group are now ready

Main focus:

- Estimate the bandwidth and test the online reconstruction algorithms performances

L1 + High Level Trigger: where do we stand?

- **Jinyuan Wu** illustrated a possible algorithm that can be implemented on FPGA
  - Need to access performance with simulated data
- Ryan illustrated to Richie and Giani how to use Vivado for doing development using High Level Synthesis
  - It will start with the new simulated data
- We made a contribution to a white paper: **Applications and Techniques for Fast Machine-Learning in Science** – N. Tran et al., to be submitted on Big Data and AI in High Energy Physics
  - Paper under circulation for final review
- **Robert Soleti** showed interest in doing development with HLS4ML
- Potential help from a graduate student

Software Trigger on GPUs

- We have been in contact with Gianluca Lamanna from Pisa
- He suggested us a way to make preliminary tests using **OpenAcc** to parallelize KinKal package:
  Referring to Amdahl's Law, we could consider the KinKal parallelizable if the time execution on GPUs is at least one order of magnitude lower than the execution with CPUs only
- OpenACC and the KinKal track-fitter algorithm are installed and ready to use on a workstation for tests

Main focus:

- We are able to use the KinKal package with CUDA compiler. Preliminary we want to use OpenAcc pragmas
- We are fixing few minor issues to run the **C++** compiler with **OpenAcc** so will be ready soon to start tests on an NVIDIA GPU, also thanks to the help and effort of the KinKal developers (Dave Brown and Roberto Soleti)
- Interface with artdaq? Starting referring to the paper:
  - GPU-accelerated machine learning inference as a service for computing in neutrino experiments

LDRD Proposal

- DAQ without borders: Approaching the increasingly diverse and heterogenous future with intelligence and coherence

Main focus:

- The innovative idea is the reduction of all network communication between different and complex devices (without standard network protocols) to simple load/store and put/get memory operations
- Exploiting the new memory-semantic fabrics based on intelligent cache-coherent links, like Gen-Z and CXL, to demonstrate a proof-of-concept (POC) that is applicable to the next generation accelerator controls for PIP-II and DAQ systems for experiments like Mu2e and DUNE

Expected timeline

- Simulated data from the Mu2e-II software group are ready
    - This allow us to start some preliminary tests
- We will restart soon our meeting to organize the work
    - We also would like to hear from the subgroup their ideas for their ROC development
- In the same period, we expect to move forward the GPU benchmark studies