Instrumentation BRN: TDAQ Midterm Report

D.Acosta[*], T.Bose[†], G.Karagiorgi[‡], W.Ketchum[§], P.O'Connor[¶], M.Williams[‖], J.Zhang[**]

November 2019

1 TDAQ

Below is the list of **draft** priority research directions in the Trigger and DAQ area. Developments for the medium term to long term include fast machine learning for large datasets, tool development that leverages new technology from industry, and automating DAQ systems. These span across all Frontiers. Longer term needs include rad-hard data links for future hadron collider experiments, which are necessary for future high granularity tracking detectors (pixel, 4D tracking) to enable streaming data and/or first-level triggering from such devices. Similarly, precision timing in the trigger enhances the Cosmic and Energy Frontiers, and 4D tracking using timing would enhance the capability at future hadron and lepton collider experiments.

- **Streaming DAQ:** In the future, several energy, intensity and cosmic frontier experiments including experiments at future facilities will need to deal with very large data volumes and/or more intense beams and higher interaction rates. The data acquisition and trigger systems of these experiments will have to be developed to handle the data effectively and efficiently. After preliminary discussions, we consider R&D towards "streaming DAQ" architecture solutions to be a priority. This is the case where data is streamed out from the detectors of an experiment at its natural full rate (or nearly so) with essentially no pre-selection. This is sometimes referred to as "trigger-less" designs for Energy Frontier experiments, and it is also the operating mode for the DUNE Intensity Frontier experiment. It requires high bandwidth from the detectors to the next data processing layer and/or compression of the data being sent. Experiments such as

[*]University of Florida, acostad@ufl.edu
[†]University of Wisconsin, Tulika.Bose@cern.ch
[‡]Columbia University, georgia@nevis.columbia.edu
[§]Fermi National Accelerator Laboratory, wketchum@fnal.gov
[¶]Brookhaven National Laboratory, poc@bnl.gov
[‖]MIT, mwill@mit.edu
[**]Argonne National Laboratory, Jinlong.Zhang@cern.ch

LHCb have successfully moved to such a readout by absorbing the large data volume directly into software-based triggers running on next generation CPUs (central processing units) and GPUs (graphical processing units). In the Cosmic Frontier area, several proposed experiments make extensive use of "software defined radio" techniques for implementing channelization and correlation of RF signals either for frequency-domain multiplexing or for 21cm intensity mapping. The high channel count and extreme digitization rates result in data rates and volumes approaching (or exceeding) those in collider detectors.

Significant R&D effort needs to be put in for exploring streaming DAQ architecture options for future intensity and cosmic frontier experiments as well as for future collider experiments. Advancements in data transmission technology will be vital for getting data off the detector so R&D in this area is critical. Cost-effective and power-effective solutions for the processing architecture will be needed. Several recent developments related to commodity networking and computing also will need to be considered. This includes developments in GPUs, hybrid CPU-FPGA (CPUs integrated with field-programmable gate arrays) and GPGPU (general purpose computing on GPUs) solutions, storage, high speed optical and electrical communication. Extremely precise (sub-ps) synchronization across thousands of processors is also required for some applications.

- **Fast Machine Learning for large datasets:** Both with and without "streaming DAQ" instrumentation, future experiments will need an unprecedented level of sophistication for data reduction and data processing informing trigger decisions. We anticipate the need for increased usage of machine learning techniques for real-time or online data processing purposes. R&D is needed from the community for the development of machine learning algorithms that work with low latency and with large data volumes, as well as for the development of instrumentation for the acceleration of machine learning algorithms which are already successfully demonstrated for offline use, for real-time or online deployment. Both R&D avenues will serve to maximize the physics sensitivity of future experiments.

- **Challenges of reading out a pixel detector at a reasonable rate for trigger:** For future hadron collider experiments at the FCC and e^+e^- experiments at future linear or circular colliders an important challenge is to extract the data at high bandwidth from high granularity pixel detectors. At hadron colliders, the high occupancy from the collisions and pileup will necessitate high granularity detectors, yielding anticipated large data rates such as about 800 TB/s from the tracking detectors. Readout options in such a detector include trigger-less, single level triggered, multi-level triggered and regional readout. If there are no significant advances in radiation tolerant optical fiber technologies, this implies more than one million detector readout optical fibers at 10 Gb/s and a 10 Pb/s event builder network. This will have significant implications for the detector material budget to extract the data and provide power and cooling as well as for the off-detector farm's requirements for CPUs, power and cooling. Therefore, an important challenge is to extract the data at high bandwidth from the tracking detectors without adding a prohibitive burden of material due to the large number of fiber drivers and the power and cooling they would need. Hermetic detectors like the current CMS and ATLAS provide no path outside of the active region for the data and trigger fibers and drivers which must survive in a harsh radiation and magnetic field environment.

- **Development of tools that leverage new technology:** The development of new technology that can facilitate applications such as machine learning algorithms for data processing, and in particular real-time data processing in FPGA platforms for power-constrained and/or data-intensive applications, will also require the development of tools for design exploration and optimization. Those tools will be needed in order to facilitate efficient and effective translation of algorithms from high level language to hardware-interpretable logic, as well as to facilitate efficient exploration of parameters for performance optimization, e.g. in terms of latency, hardware resources, and power consumption. While some improvements in tools may come from industry itself, we anticipate still needs from the HEP community given its specialized needs.

  For FPGA platforms where resources are constrained, the deployment of DNNs will be particularly challenging for low-latency applications involving large networks and large input data sizes. Performance optimization will require thorough exploration of DNN deployment on customizable and efficient hardware accelerator designs for the entirety of the network or individual stages of the networks, which can be done using High Level Synthesis-based design flows. The proliferation of High Level Synthesis (HLS) tools for Machine Learning will be needed in order to meet anticipated development needs.

  Tool development will also need to keep up with advances in network and interconnect models and technology as well as advances in heterogeneous computing platforms.

- **Automated, self-running DAQ systems:** Increases in detector size, complexity, and data-taking rates, along with requirements for high detector uptime and efficient collection of good-quality data, place high demands on control and monitoring systems. While better design may increase fault-tolerance in DAQ systems, fast collection and processing of conditions data will be necessary to enable detection of problems as they appear. Further development of technologies (like high-performance distributed databases and messaging, for example) and their application to real-time systems that will handle the scale of future detectors is needed.

  Increasing automation in the detection of anomalous conditions and performing corrective measures, rather than relying on manual interventions, will also be needed to improve data-taking efficiency, but it requires some key development efforts both in performing real-time or near-real-time detection of anomalies and problems, and quickly determining and executing an appropriate response. This is in many ways similar to the problems faced in triggering and data selection, except that in many cases automation of a self-running DAQ system may need to be more contextually adaptive: continually collecting and analyzing data, making appropriate adjustments, and reevaluating under new conditions. Advances in AI and machine learning (e.g. applications of unsupervised and reinforcement learning) may enable new types of algorithms to be put in use for improving the performance and stability of automatized error detection and prevention, and even self-calibration, and are an important avenue for future development.

Many of these same ideas may be applied to improved control of other complicated systems, like particle accelerator operation, and present an opportunity for cross-cutting initiatives.

- **Precision timing in trigger, and precision synchronization:** For experiments at colliders, detectors have traditionally derived timing references from the machine RF signals which in turn are synchronized to a master oscillator with exceptionally high stability. Derived clocks from the central machine timing are used to generate the detector timing and synchronization messages which are then distributed as needed to the front- and back-end electronics over optical fiber links. For future colliders, collision frequencies will increase, and it will become necessary to implement more precise timing detectors to improve event selection. Furthermore, the trend towards triggerless or "streaming" DAQ will bring new challenges for synchronization and event tagging.

  Detectors for cosmic and intensity/neutrino frontier experiments face different issues. In the absence of a signal from machine RF systems, the detectors will be responsible for defining their own master timing reference. Many proposed experiments involve arrays of detector stations spread over kilometer distances, including hostile (underwater, desert, under-ice etc.) environments. Timing distribution over cables (copper or optical) subject to temperature variation will suffer phase drifts and timing systems will need to develop compensating techniques to preserve synchronicity. For multi-messenger studies, where e.g. detection of a transient astrophysical event (gravitational wave, supernova, UHE cosmic ray) in one detector initiates followup operations at other sites, it will be essential to make accurate absolute timestamps available across networks that may span thousands of kilometers. Future cosmic survey experiments that use software-defined radio have stringent requirements for phase coherence across multiple receiving stations. In some applications (e.g. hydrogen intensity mapping), the sampling clocks for GSa/s digitizers must be phase-coherent to within $< 1\%$ of the RF period, or around 1ps. Clock distribution systems used in present experiments may be able to reach this level of accuracy in laboratory-scale systems but advances will be required to extend the range of synchronicity over kilometer scales.

- **4D tracking with spatial and timing readout:** Precision timing detectors historically have been very useful for particle identification. More recently, the incorporation of timing detectors into the hadron collider experiments for the HL LHC are planned to help mitigate the effects of pileup, adding a time dimension to the spatial tracking to further disentangle the matching of tracks to vertices at high pileup. Silicon sensor technology (e.g. LGAD) can be used for such purposes and achieve timing resolutions of order 50 ps. A potential next leap in tracking systems would be to embed this timing capability into future silicon tracker designs for future collider experiments, providing 4D tracking capability. This would further improve the performance at high pile-up by reducing the combinatorics in building tracks in a high density environment, for example at a Future Circular Collider. Along with the sensor and front-end electronics R&D required to achieve this would be the need to develop methods to read out the increased amount of data from the tracker. Moreover, with silicon tracking now on the cusp of inclusion into Level-1 electronic trigger systems at the HL LHC, a further next step would be how to extend such trigger designs to take leverage this 4D capability to improve overall performance in the presence of high pileup.

- **Rad-hard high bandwidth data links, high bandwidth wireless transmission:** Radiation-hardened, high bit-rate optical transmitter module will be required for future HEP experiments at hadron colliders. The tolerance to total ionizing dose (TID) should be on the order of 10 Mrad for a transceiver to be suitable. Though some commercial Silicon Photonics devices (modulator based) demonstrated a data transmission rate at >10 Gbps after irradiation, some modifications of components were needed. It will be crucial to design a silicon Photonics-based chip which could be used in integrated module to support data rates of 10 Gbps, 25 Gbps or higher per link in high radiation environment, such as the innermost area of the tracking detector for the future HEP collider experiments. This will require investment from the community.

  The feasibility of wireless readout of large instrumentation systems, like the silicon tracker for the future collider experiments should also be studied. Some very preliminary work with use of the 60 GHz or higher bands was demonstrated. The commercial chipsets can be used to construct a prototype. A transceiver chip need be developed suitable for high data rate and short distance applications, which can provide wireless Multi Gigabit per second radial data transmission inside the detector, making readout and processing all hit data feasible.