FERMILAB-CONF-12-302-CD

NOνA Event Building, Buffering and Data-Driven Triggering From Within the DAQ System

M Fischler, C. Green, J. Kowalkowski, A. Norman, M. Paterno, R. Rechenmacher[1]

Abstract.
To make its core measurements, the NOνA experiment needs to make real-time data-driven decisions involving beam-spill time correlation and other triggering issues. NOνA-DDT is a prototype Data-Driven Triggering system, built using the Fermilab **artdaq** generic DAQ/Event-building tools set. This provides the advantages of sharing online software infrastructure with other Intensity Frontier experiments, and of being able to use any offline analysis module–unchanged–as a component of the online triggering decisions.

The NOνA-**artdaq** architecture chosen has significant robustness advantages, including graceful degradation if the triggering decision software fails or cannot be done quickly enough for some fraction of the time-slice "events." We have tested and measured the performance and overhead of NOνA-DDT using an actual Hough transform based trigger decision module taken from the NOνA offline software. The results of these tests–98 ms mean time per event on only 1/16 of the available processing power of a node, and overheads of about 2 ms per event–provide a proof of concept: NOνA-DDT is a viable strategy for data acquisition, event building, and trigger processing at the NOνA far detector.

1. Introduction

NOνA [1] is a program to investigate the properties of neutrinos. The experiment is designed to probe the oscillations of neutrinos and anti-neutrinos to determine the transition probabilities for

$$\nu_\mu \rightarrow \nu_e \text{ and } \bar{\nu}_\mu \rightarrow \bar{\nu}_e$$

These transition rates will allow NOνA to determine the neutrino mass hierarchy and highly constrain the CP violating phase δ_{CP}. The NOνA measurements will improve our understanding of the mixing angles θ_{13} and θ_{23}.

The NOνA project and experiment includes:

- Doubling of the Fermilab NuMI beam power to 700 kW
- A 15,000 ton totally active surface detector, 14 mrad off axis at 810km (which is the first oscillation maximum for 2GeV neutrinos)
- A 220 Ton totally active near detector

The detectors have been optimized as a segmented low Z calorimeter/range stack to reconstruct EM showers, measure muon tracks, and detect nuclear recoils and interaction vertices.

[1] Fermi National Accelerator Laboratory

Operated by Fermi Research Alliance, LLC under Contract No. De-AC02-07CH11359 with the United States Department of Energy.

The enormous far detector has over 368,000 channels of synchronized readout electronics, operating in a "free-running" trigger-less mode. There is no deadtime; the entire raw data stream for the detector is actively buffered in a large computing farm. Somehow, the subset of this data stream containing events corresponding to the beam pulses (and neutrino creation) at Fermilab must be identified for permanent retention.

The large volume of data readout, combined with the difficulty of synchronizing any trigger for saving data with the beam pulses at Fermilab, presents a challenge for modern computing and data acquisition. An approach to meeting this challenge is to integrate triggering algorithms based on the data, with the data acquisition and event building systems. This Data Driven Triggering (DDT) approach needs to be implemented in an efficient and reliable way.

Since the DDT approach risks degrading the desirable "separation of concerns" between trigger algorithm code and DAQ/Event Builder software, it becomes important to implement it in a framework that deals with the key data movement issues and provides organizational stability and system robustness. We have tested the performance of the NOνA-ART DAQ framework in this context, providing measure of the performance level which can be achieved in a specific well-structured environment providing access to the stream of data.

1.1. A Brief Overview of NOνA Triggering Issues

The 368,000 channels in the NOνA far detector are continually digitized and the data are accumulated through a multi-tiered readout chain into a large farm of high-speed computers. Since the full data rate is too large for any practical scheme of retaining all the data for every event, some form of trigger is needed to filter the data logging.

To make its core physics measurements, NOνA needs to be able to:

(i) Send a narrow 10μs pulsed beam of mainly ν_μ to the near detector and far detector (massive) where the beam will interact.

(ii) Correlate interactions in each detector with the beam timing.

(iii) Compare the rates of interaction to determine appearance/disappearance rates and energy spectra.

For the core NOνA beam-related physics topics ($\theta_{13}, \theta_{23}, \delta_{CP}, \Delta m^2$), the only "triggers" needed are correlation with the beam spill, and a calibration pulser. Both are open zero-bias selections.

In principle, restricting ourselves to considerations of the neutrino oscillation physics goals, the trigger desired would be time-coincidence with the arrival of neutrinos from the beam pulse at Fermilab (plus some small fraction of off-timed calibration triggers). Since the beam pulse is of very short duration, and the cycle time is 2 seconds, a suitably correlated active window of (say) 30μs every 2 seconds would capture all events associated with the beam, and would reduce the volume of data by about five orders of magnitude.

1.1.1. Motivation for a Data-Driven Trigger (DDT) While a pure timing trigger, correlated with the beam pulses, would be perfect for the neutrino oscillation studies, there is no way to provide hard timing signals (which would allow correlation of data triggering with the beam spill) to the far detector 810km away. Nor is there any way to transmit or "predict" the beam spill timing. Spills are slotted into each accelerator super-cycle, but the time-line structure for those spills changes according to which experiment is running, and the time alignment of the super-cycles will vary based on that time-line structure. Therefore, it is impossible (at the far detector) to verify the timing of a spill until it is far too late for immediate triggering.

Another approach would be to use the gross level of activity in the far detector as a signal to determine the timing of the beam pulses. (While this approach would tie the triggering to some gross property of the data, it is not what we mean by "Data Driven Triggering," which involves a global and detailed examination of the full event to decide whether to trigger retention

of that event.) Since the far detector is a surface detector, there is always significant activity, so a trigger based on gross activity level won't solve the problem.

Thus NOνA needed to design a system that goes beyond simple traditional triggers, and can tolerate extremely long latencies in readout. The solution is continuous readout with absolute synchronization. That is, there is no distinct "active" time for the detector; each channel provides time-stamped information whenever it is hit. To handle this, the system must:

(i) Instrument every channel with TDC.

(ii) Synchronize every channel of the detector to every other channel (global synchronization).

(iii) Readout the detectors continuously.

(iv) Time stamp every hit and store every hit in deep buffers.

(v) Time stamp the beam spill at FNAL and send the information to the far detector – thus the spill information arrives at the detector later than the corresponding data is taken. The system specifications include the ability to cope with these latencies, to avoid data loss.

(vi) Analyze the hit data to characterize neutrino interactions.

(vii) Match the hit data with the beam spills.

(viii) Record the data that correlates with the deduced beam spill timing.

In this paper we wil go into some detail about the deep buffering and the last four of these items; details of the timing, synchronization and continuous readout technique are in reference [2].

The strategy is that beam spill information from NuMI[3] (Neutrinos at the Main Injector) is generated independently at FNAL and transmitted to the Ash River NOνA far detector site. (The NuMI beamline is a facility at Fermilab which uses protons from the Main Injector ring to produce an intense beam of neutrinos for NOνA and for the MINOS experiment.) When the beam information is received, it is combined with information from high-level analysis of the events, to determine the actual beam spill timing at the far detector. Then the relevant data blocks that match the beam spill are extracted with zero bias and recorded to permanent storage.

This strategy requires Data Driven Triggering, based on algorithmic examination of the full event, running on a processor in a farm. (An "event" in the context of the asynchronous readout consists of hits that are close enough in time to be associated with tracks from particles produced by a common primary neutrino.) The data for each event in the full stream is examined and categorized, and information pertinent to the beam spill timing correlation is gathered. An accept/reject decision is made before the the buffer space occupied by the event data is needed for a later event; this is a "point of no return" on saving or discarding the complete data. It's important to note that *all* the data ($\approx 4.3GB/s$) is buffered into a farm of "buffer nodes," and that the buffering and temporary retention of the full data is adequate to allow a single processing node to run a sophisticated filtering algorithm on one event, before that event is permanently saved or discarded. This places a baseline buffering requirement of 20 seconds of data (long enough to execute anticipated complex decision algorithms); the actual system exceeds this requirement by almost two orders of magnitude.

Being able to examine the full data stream in real time while it is in the buffers gives us the opportunity to generate various Data Driven Triggers. The most important, of course, are triggers based on determination of the deduced beam pulse window. But because the detector is 100% live, other triggers have tremendous potential for calibration and for extending the topics addressed by NOνA to non-accelerator physics:

- Improve detector calibrations using "horizontal" muons
- Verify detector/beam timing with contained neutrino events
- Search for exotic phenomena in non-beam data (magnetic monopole, WIMP annihilation signatures)

- Search for directional cosmic neutrino sources
- Detect nearby supernovae

2. NOνA-artdaq – a Platform For the NOνA Data Driven Triggering

Data Driven Triggering (DDT) requires use of a great deal of computing power, integrated into the system that acquires, moves and logs the event data. Implementing an efficient system to coordinate with a data acquisition system, moving its data to build events and sharing those events with a collection of general-purpose computers, is a non-trivial endeavor. NOνA-DDT is implemented using the **artdaq** data acquisition and event processing framework[4] for Intensity Frontier and other medium-scale experiments, which handles the common organizational aspects of that task.

NOνA-**artdaq** is a variant of the **artdaq** framework. The **artdaq** project at Fermilab has been established (and is under active development) to design a generic toolkit for the construction of robust event filtering and analysis programs as part of an integrated data acquisition system. **artdaq** allows sharing of DAQ infrastructure between experiments, and supports the use of commodity computers and hardware accelerators (GPU's) as close to the data source as possible. **artdaq** focuses particularly on efficient data communication among data collection nodes and nodes running data processing algorithms (including compression and triggering). Flexibility is stressed: The two functions might be done in the same or different sets of nodes, and reconfiguring the system (for example, adding more processing capacity) can be done without reprogramming.

To facilitate integration of online and offline event-processing code, **artdaq** makes use of the **art** [5, 6, 7] event-processing framework, which is used as the offline framework for many of the newer Fermilab experiments. Thus code to be run in online triggers can be verified and profiled for performance improvements in the offline environment. A "filter module" used in the offline code can be used (without change) in the online system running **artdaq**. The advantages of shared algorithms include transparent simulation of triggers, accurate measurement of efficiencies, and leveraging work on offline reconstruction and analysis modules, which become tools available to the trigger as well. This commonality simplifies development, testing, and validation cycles, and avoids the concern inherent in the comparison of two different implementations of (what is intended to be) the same algorithm.

These features make **artdaq** a natural environment for implementation of the NOνA Data Driven Trigger. Starting from the fundamental framework, NOνA-**artdaq** has been tailored to read from shared memory event buffers, which are populated directly by the NOνA event builder.

In NOνA-**artdaq** the processing of complex trigger decisions is partitioned off from the primary DAQ chain; this improves system stability. For example, if for any reason the DDT processes doing triggering analysis fall behind in the analysis with respect to the rate of event data taken, the consequence for the main global trigger is only that the overall decision as to what constitutes the beam spill timing becomes based on a large fraction of the time slices, instead of the entire event data stream. Since the zero-bias trigger window is somewhat wider than the actual beam pulse, the efficacy of triggering degrades very gradually, if at all.

2.1. The NOνA DAQ Readout

The NOνA event builder is integrated with its DAQ system. Data Concentrator Modules (DCM) are designed to integrate the continuous data stream from 64 front end boards (2048 detector cells) into partially time-sorted data blocks. Each DCM handles the data for one geographic quadrant on the detector. The DCM combines a large FPGA with a single-board computer to build events (more accurately, data slices) based on 5 ms of data. (Since the actual time scale for tracks produced by an incoming neutrino is much smaller than 5 ms, the fraction of data lost because the actual event straddles a boundary of these slices is insignificant.) The DCM's run

an embedded linux with custom support for NOνA firmware. Demonstrations of readout on the near detector have shown each DCM can handle a 24 MB/s data rate.

The data from DCM's are sent to a farm of buffer nodes. Every 5ms the data stream from ALL DCM's on the detector is transmitted into a single buffer. This gives a complete 5ms snapshot of the detector in that buffer. The destination buffer node is then rotated in a round-robin pattern through the entire cluster (with a nominal farm of 200 buffer nodes, there is a one second time to return to the first node). Since each node can deal with multiple buffers, the total trigger decision latency tolerated is much more than one second: The design point is at least 20 seconds, and at projected data rates we can actually buffer up to 69 minutes of raw, minimum bias readout system-wide.

2.1.1. DAQ Topology The NOνA DAQ system supports continuous readout of over 368,000 detector channels by creating a hierarchical readout structure that transmits raw data to aggregation nodes which perform multiple stages of event building and organization before retransmitting the data in real-time to the next stage of the DAQ chain. In this manner data moves from Front End digitizer boards (FEBs) to DCM's to Event Building and Buffering nodes and finally into a data logging station.

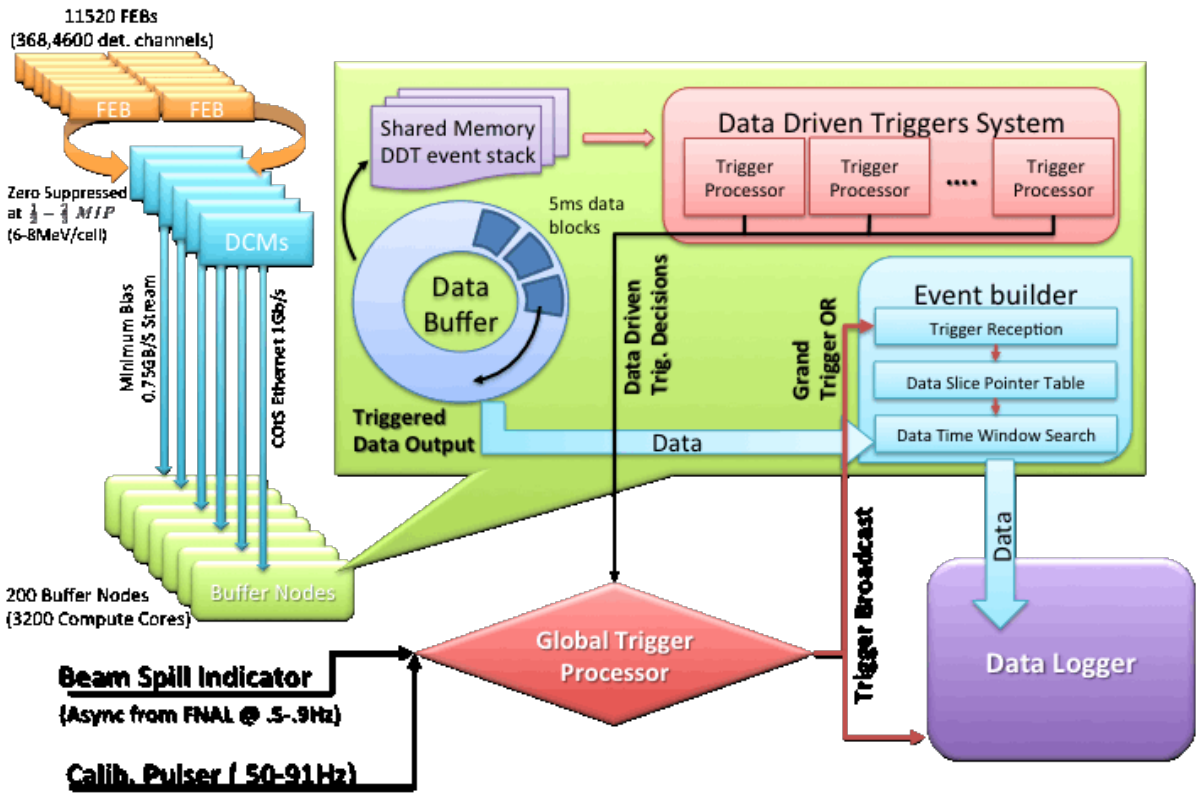

Figure 1: DAQ topology – An overview of the DDT design.

The flow of data out of the buffer nodes is controlled by a Global Trigger process that issues extraction directives based on time windows identified to be of interest. These windows can represent beam spills, calibration triggers, or data driven trigger decisions.

The buffer nodes used for DAQ processing are a farm of high performance commodity computers. Each note has a minimum of 16 cores and 32 GB of memory available for data processing and buffering. This permits them to run both the event builder process and the full

multi-threaded Data Driven Triggering suite, and provides sufficient interim storage for triggering based on full, continuous, zero-bias data readout.

2.2. DAQ Event Building and DDT Buffering

The NOνA-DDT system has two distinct buffer pools: the Global Trigger Pool (shown as the circular data buffer in figure 1), and the DDT pool. Data in the 200 buffer nodes used for the Global Trigger Pool also goes into the shared memory DDT event stack, whence it can be accessed by trigger processors in a Data-Driven Trigger system. This system can be as large as is needed to keep up with the long-term computational needs of making trigger decisions. Each trigger processor pulls out a 5 ms "slice" of data at a time and makes decisions based on analysis of that slice; results and accept/reject decisions are communicated to a Global Trigger Processor, which also receives trigger information packets sent from FNAL over the Internet to the far detector, (these have an average one way flight time of 11ms). The Global Trigger Processor can cause any given event in the data buffers either to be built and logged, or to be discarded. (Although the hits of interest within a slice of data are typically concentrated in time, the data logging accepts or rejects the entire 5 ms slice.)

Because the DAQ system performs a 200 to 1 synchronized burst transmission of data between the Data Concentrator Modules and the event builder buffer nodes, a large amount of network and data transmission buffering is required to handle the data rate. This buffering has been tuned by adjusting the size of data packets to match a single complete time window (this reduces the number of transmission and queue-entry overheads). The receive window on the buffer nodes has also been tuned to exploit the buffering capability of the high-speed switched fabric (a Cisco 4948 network switch array) between the DCM's and the buffer nodes; see figure 2.

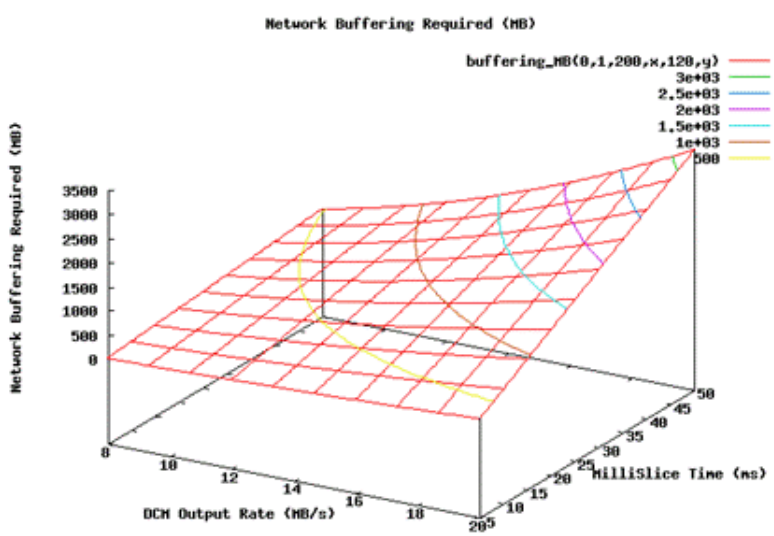

Figure 2: Network (switch) buffering optimization as a function of far detector average DCM rates and acquisition window time.

The system can retain time-slices written to the Global Trigger pool (the circular data buffer in figure 1) for at least 20 seconds if necessary, before they are potentially extracted by a global trigger and sent to the datalogger. The DDT pool, treated as input queues for data driven trigger analysis processes, is a separate set of IPV4 shared memory buffers. The DDT shared memory interface is optimized as a one-writer-many-readers design. The shared memory buffer has N

"large" (fixed size) segments which are rotated through by the writer to reduce collisions with reader processes. And the writing process uses guard indicators before and after the detector data, which allow reader processes to tell if the data segment being read may have been over written during read-out.

Note again that the asynchronous aspect of the system means that even if the Data Driven Triggers system were to get behind on its processing, the data buffer would not block, the key zero-bias beam spill triggering decisions can degrade very gradually, and logging of events based on that key trigger and on other individual triggers generated can still proceed.

3. Performance of NOνA-artdaq DDT

The NOνA Data Driven Trigger (NOνA-DDT) prototype system is a customized application using the **artdaq** framework. It contains a demonstrator "filter module" (in the **art** framework sense), shown in the darker green box in figure 3. The system uses the high speed event feeding and event filtering components of **artdaq** to provide a real world proof of concept application that meets the NOνA trigger requirements.

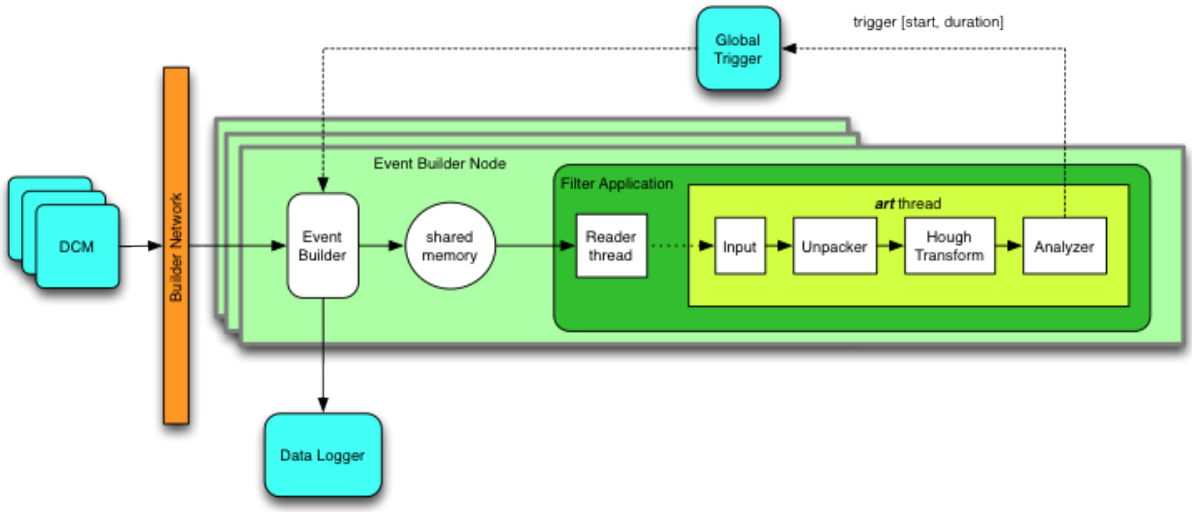

Figure 3: The NOνA Data Driven Trigger (NOνA-DDT) prototype system

3.1. Real-time Event Filtering

NOνA-DDT takes already-built 5ms wide raw data windows from shared memory, and injects them into the **art** framework where trigger decision modules are run. A decision message is broadcast to the NOνA Global Trigger System.

The **art** framework allows the usage of standard NOνA reconstruction components to be applied to the online trigger decision process. This additionally allows for online reconstruction and filtering algorithms to be reproduced and applied in the offline environment. Thus the identical code used in the DAQ can also be run in offline simulation and testing, for determination of selection efficiencies, reconstruction biases, and overall trigger performance.

3.2. Performance Measurements on the Prototype System

The NOνA-DDT prototype system implements, in the filter module, a real world trigger algorithm designed to identify tracks in the NOνA detector[8]. This provides a realistic test of the Data Driven Trigger performance. The algorithm calculates the pair-wise intersections of the Hough

trajectories for the detector hits, and identifies the resulting track candidate peaks in the Hough space.

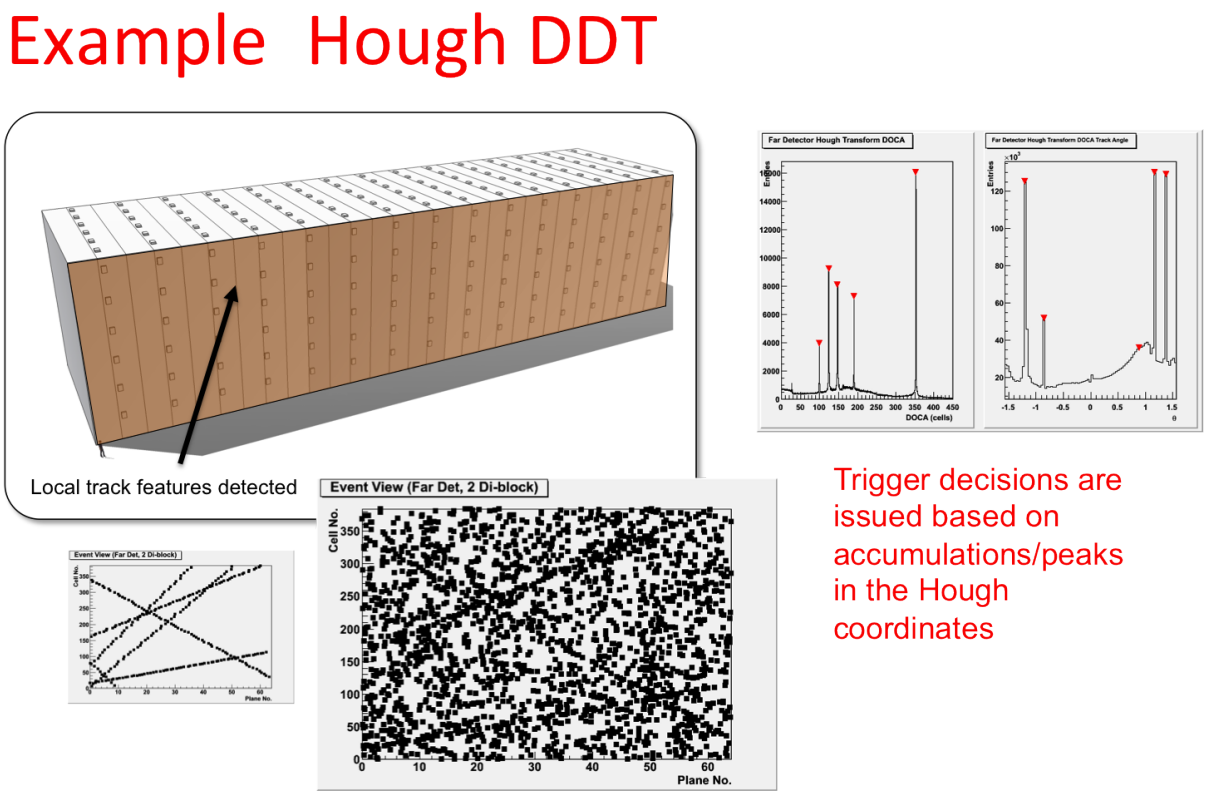

Figure 4: A Data Driven Trigger based on a 2D Hough transform.

This X/Y calorimeter track finding using a Linear Hough transform in 2D is a real world example of an $O(N^2)$ complexity algorithm and is of particular interest due to its ability to identify candidate slow particle tracks (e.g., Magnetic Monopoles candidates with $\beta > 10^{-5}$) in the far detector. Thus when performance measurements are made on the prototype system, it is running a demonstration module which has the same characteristics as will be present in actual trigger code.

(The Hough transform methodology can be expanded into more than two dimensions, incorporating timing and pulse height information. This, or course, increases the computational burden. Against that is the fact that these methods are well-suited for massively parallel calculation, and for use of GPU's to accelerate the processing.)

The target goal in the proof-of-concept demonstration was to run the trigger decision algorithm in an average of 60 ms. The NOνA-DDT Hough algorithm, without parallelization, was run on a single core in a node in the NOνA near detector DAQ trigger cluster. The DDT system was run on the near detector prototype data stream, with zero suppression threshold set to match far detector data rates. Total readout of each 5ms slice of data (500 times the beam spill window) were processed as a single block (no partitioning).

The near detector DAQ cluster is identical in architecture to the far detector DAQ, and has the huge advantage over artificial simulations of generating actual events. Thus the test measurements were made in a very realistic context, other than not exploiting the available computation power of the remaining 15 cores on the nodes running the algorithm.

The algorithm was able to reach decisions in on average 98 ± 14 ms. (The distribution of processing times is shown in figure 5). It was able to achieve this level of performance using only $\frac{1}{16}$ of the available CPU resources of each cluster node. This initial test shows that the system, after data partitioning, parallelization and algorithm optimizations, will be able to meet and exceed the NOνA trigger requirements.

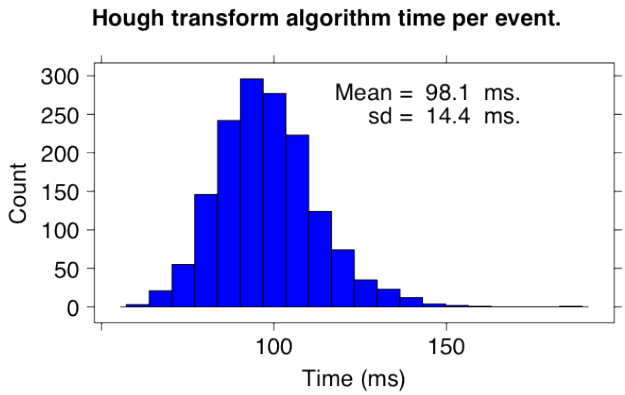

Figure 5: Hough transform execution time for 5ms time windows of NOνA near detector readout. No parallelization was applied to the algorithm.

A general concern when embedding time-critical calculations in an organizing framework, is the overhead incurred due to the framework and due to any data movement involved. Our performance tests measured this overhead for the **artdaq** framework. The average overhead for a 5 ms time window, including all actual framework code and including unpacking and formatting of "DAQHit" objects used as input to the Hough transform filter algorithm, is a very acceptable 2.5% of the processing time.

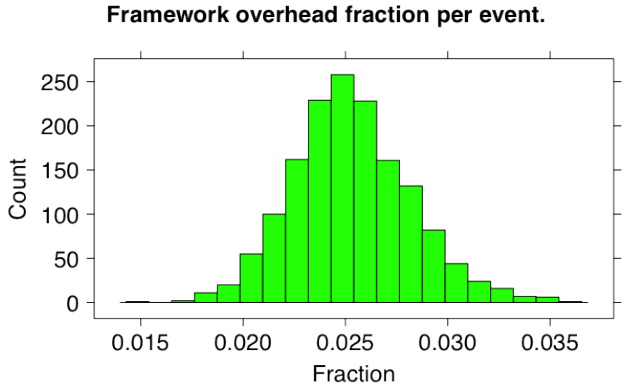

Figure 6: The **artdaq** framework overhead for processing of 5 ms time windows of NOνA near detector readout.

Having in hand a realistic and robust prototype DDT system positions the system builders well for larger scale tests, and will facilitate the experiment's algorithm development.

4. Summary

We have built the foundation for a unique free-running DAQ system, suitable for the purposes of NOνA and other Intensity Frontier or medium-scale experiments. Using this **artdaq** foundation,

we have build a prototype system which can solve the problem of beam-spill time correlation and synchronization. The Data Driven Triggering prototype system (NOνA-DDT) reliably perform continuous readout and buffering.

Tests using actual data from the DAQ system running on the near detector clearly indicate that the NOνA-DDT architecture can handle their triggering and event-building needs. Further work needs to be done, to exploit the available parallelism of the system.

We have begun to explore the full power and opportunities that open up when Data Driven Triggering can be done using a well-structured framework with all the experiment's offline components available. The NOνA-DDT prototype system might only scratch the surface of the space of possibilities.

Acknowledgments

Scientific Computing Division, Fermi National Accelerator Laboratory, Batavia, IL, USA. This work was supported in part by the U.S. Department of Energy, Office of Science, HEP, Scientific Computing.

[1] Nova neutrino experiment URL `www-NOvA.fnal.gov`

[2] Norman A *Procedings of CHEP2012* (IOP) in progress

[3] The minos experiment and numi beamline URL `www-numi.fnal.gov`

[4] Biery K, Green C, Kowalkowski J, Paterno M and Rechenmacher R 2012 **artdaq**: An event filtering framework for fermilab experiments Tech. Rep. FERMILAB-CONF-12-278-CD Fermi Antional Accelerator Laboratory

[5] Kutschke R 2010 *J. Phys. Conf. Ser.* **331** 032019 URL `http://stacks.iop.org/1742-6596/331/i=3/a=032019`

[6] cet-is art documentation wiki URL `https://cdcvs.fnal.gov/redmine/projects/art/wiki`

[7] Brown W, Green C, Kowalkowski J and Paterno M *Procedings of CHEP2012* (IOP) in progress

[8] Norman A *Procedings of CHEP2012* (IOP) in progress