Fast Tracker (FTK): A Hardware Track Finder for the ATLAS Trigger

Takashi Mitani on behalf of the ATLAS Collaboration

Abstract—**During 2010-2012, the LHC provided the first proton-proton collisions to its experiments. The ATLAS trigger system was successfully operated and it contributed to several important results such as observation of the Higgs boson with a mass of about 125 GeV/c^2. Starting in 2015, the collision energy will increase to 13-14 TeV and the instantaneous luminosity will reach 1-3 $\times$ 10^{34}cm^{-2}s^{-1} with a 25 ns bunch crossing period. Due to the energy increase, the cross sections for SM processes are expected to increase. Additionally, the number of overlapping proton-proton interactions per bunch crossing, which is referred to as pile-up, is expected to increase significantly up to about 80. Therefore it will be challenging to control trigger rates while keeping good efficiency for interesting physics events.**

This document summarizes the development of the Fast Tracker and its tracking performance for the ATLAS experiment. The Fast Tracker is a custom electronics system that will operate at the 100 kHz level-1 trigger rate and provide high quality track information at the beginning of the high level trigger. This will be performed by a hardware based track reconstruction system with massive parallelism using associative memories and FPGAs. The availability of the full tracking information from Fast Tracker will ensure robust selection in the high level trigger within its limited latency, including the mitigation of effects due to pile-up.

I. Introduction

I N order to select interesting physics processes for LHC proton-proton collisions, the ATLAS experiment [1] uses a three-level trigger system. The level-1 trigger (L1) is a custom hardware system, which selects high momentum muons detected in the muon spectrometer and energy depositions in the electromagnetic and hadronic calorimeters. It reduces the trigger rate from 40 MHz to around 100 kHz, and identifies the Regions of Interest. These are seeds for the High Level Trigger (HLT), which reduces the event rate to a few hundred Hz. Trigger selection will become much more challenging in the 2015-2017 run than it was in the very successful 2010-2012 run. The increase in beam energy and instantaneous luminosity will raise the trigger rate for all processes, and events will be further complicated by the large number of overlapping proton-proton collisions (pile-up). Because of its fine resolution and granularity, tracking information is critical to effectively select events that should be kept for the HLT. However, extensive tracking is prohibitively expensive in terms of processing resources.

The Fast TracKer (FTK) [2] will provide within 100 μs global reconstruction of all tracks with p_T above 1 GeV/c. FTK will use the full readout from the silicon inner tracker detectors: silicon pixels (including the Insertable B-Layer, IBL [3]) and silicon strip detectors (SCT). FTK is a hardware track reconstruction system with massively parallel processing to

handle the large L1 trigger rate with a small latency. The core functionality of the FTK system consists of pattern recognition and track fitting. FTK tracks, freed from the CPU constraints of HLT tracking, will be an important tool for the ATLAS HLT, allowing it to improve event selection.

In this document, the architecture of the FTK system is presented. The design and development of each FTK component is described. Finally the simulated performance of the full system and the integration with the HLT are reported.

II. Architecture of the FTK System

Figure.1 shows the architecture of FTK system. It employs one specially designed custom chip for the associative memory (AM) and a lot of FPGAs.
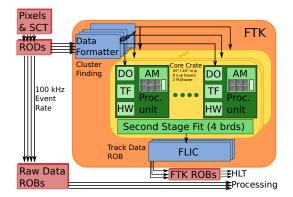

Fig. 1. Architecture of FTK system. AM is the Associative Memory, DO is the Data Organizer, FLIC is the FTK-to-Level-2 Interface Crate, HW is the Hit Warrior, ROB is the ATLAS Read Out input Buffer, ROD is a silicon detector Read Out Driver, and TF is the Track Fitter. Second Stage Fit is referred to in the text as the Second Stage Board.

FTK uses the information from 12 logical layers, corresponding to 4 pixel detector layers and 8 from SCT, including both axial and stereo sides. The FTK is organized as a set of independent engines to achieve large processing power. Each engine works on a different region of the silicon tracker: 64 η-ϕ towers, 16 in ϕ (azimuthal angle) by 4 in η (pseudorapidity, $|\eta| < 2.5$). To avoid inefficiency at tower boundaries, the towers must overlap because of the finite size of the beam luminous region along the beam axis and the curvature of charged particles in the magnetic field.

The pixel and SCT data are transmitted from the detector Readout Drivers (RODs) on S-LINK fibers and received by the Data Formatters (DF). DF mezzanine cards, the FTK_IM, perform cluster finding, two-dimensional for the pixel layers, and centroid calculations. Clusters consist of hits connected side-by-side or diagonally. A moving window of programmable size

is used for the clustering algorithm. Clusters extending beyond the window boundary are truncated. The DFs receive data from FTK_IMs, organize the data into the 64 FTK η-ϕ towers and transmit the cluster centroids to the core crates containing the pattern recognition and track fitting hardware.

In the core crates, the Data Organizers (DO) store the full resolution hits and convert them to coarse resolution (Super-Strips or SS) for pattern matching. DOs store the full resolution hits so that when a pattern has been found the hits in the pattern can be immediately sent to the Track Fitter (TF).

The AM boards contain a very large number of preloaded patterns corresponding to the possible combinations of a SS in each layer. Only 8 layers are used in this pattern-recognition step. The patterns are determined in advance using simulation of single tracks in the ATLAS detector. In the AM chip, the matching of a set of hits with all the preloaded patterns is made in one single clock cycle.

When a pattern with 7 or 8 hits is matched (such patterns that contain track candidates are referred to as roads), it is sent back to the DO which retrieves the associated full resolution hits and sends them to the TF. Since each road is quite narrow, the TF can provide an accurate estimate of the track χ^2 (and helix parameters in the Second Stage Board) with a linear approximation. Therefore fitting a track is extremely fast, and approximately 10^9 track candidates can be fit per second in a modern FPGA. The Hit warrior (HW) algorithm removes duplicate tracks among those 8-layer tracks in a road.

A second step fit is executed in the Second Stage Boards (SSB), where hits on all 12 logical layers are used.

SSB output tracks, each of which consists of the hits on the track, the quality of fit and the helix parameters, are sent to the FTK-to-Level2 Interface Crate (FLIC). The FLIC organizes the tracks and sends them to the HLT Read Out Systems (ROS) using the standard ATLAS protocols, and carries out monitoring functions. The FLIC crate is organized so that additional functions can be added in the future, including global event functions such as primary vertex reconstruction.

III. Design and Development

The FTK_IM functions are implemented in a mezzanine card that connects to the DF main board. A FTK_IM board receives 4 inputs from inner detector RODs. The data goes directly to the two FPGAs on the board, where the clustering algorithm and centroid calculation are performed.

The DF system uses a full-mesh Advanced Telecom Computing Architecture (ATCA) backplane interconnect. Each board is connected to every other board with multiple point-to-point links in the full mesh ATCA backplane. A DF board has a Rear Transition Module (RTM) which is used to send the data downstream as well as to perform inter-crate data sharing.

Pattern recognition and track fitting are performed in the Processor Units (PU) in the 9U VME crates (core crates). Each PU consists of an AM board with a large auxiliary board (AUX) behind it. The full FTK system will have 128 PUs in eight 9U VME crates.

The hits from the DFs are received by two Input FPGAs on an AUX card. Within the Input FPGAs, the SS number for each hit is generated and sent to the AM board. In addition to Input FPGAs, the AUX card contains four processor FPGAs to perform the DO, TF and HW functions. After the TF and HW operations are carried out, the track information is sent to the SSB.

An AM board (AMB) is a VME board on which 4 local associative memory boards (LAMB) are mounted, each of which holds 16 AMchips. The AMchip is an ASIC specifically designed for FTK pattern recognition.

The SSB boards are also mounted in the core crates. Each SSB receives the output from AUX cards and the hits on the additional silicon detector layers from the DFs. A FPGA is mounted for board configuration and diagnostics, read/write access to memory used to store the 12-layer constants, and for readout of the SpyBuffer memory. In addition, the SSB holds 3 FPGAs which extrapolate the 8-layer tracks into the other 4 layers and carry out the TF and HW functions. To remove duplicate tracks in the tower overlap regions, track information is shared among the SSBs. After the HW function, the SSB sends the tracks to the FLIC.

The FLIC is implemented in a single ATCA shelf with full mesh backplane. An ATCA Input Card receives the data from the core crates, formats each event for the ATLAS DAQ, and passes the data to a rear transition module, the Output Card. The Output Cards are connected to the HLT ROSs. To facilitate optional trigger processing in the future, such as primary vertex finding or beam spot determination, the event data streams are copied and can be collected into a full record by a trigger processor blade in the shelf.

The functionality of each prototype board has been tested intensely using test data from FTK simulation. Mass production of the FTK boards has now begun, and they will be installed into the ATLAS trigger system starting in the summer of 2015.

IV. FTK Performance

A. Timing Simulation

A simulation tool for estimating the FTK latency has been developed for tuning the system architecture to ensure that the FTK can handle a 100 kHz L1 trigger rate at high luminosity. The tool estimates the FTK latency by dividing the system into functional blocks: DF, DO write mode (receiving hits from the DF and sending SSs to the AMB), AMB, DO read mode (receiving roads from the AMB and sending roads and hits to the TF), TF, HW, and SSB. The most time consuming steps are processing from DO write mode through TF. The DF adds little to the overall latency since each cluster found is immediately sent to the DO. Therefore the DF and DO execution times almost completely overlap. Similarly, the HW has a relatively short latency because each track that enters is either sent to the output or discarded. For each functional block, the time of transfer of the first and last words into and out of the block are calculated. Because each core crate operates independently, the FTK event execution time depends on the busiest crate. The execution time for a block depends on the number of input words, the processing time per word, and the number of output words. The left figure of Fig.2 shows an example timing chart for individual Z$\rightarrow$ $\mu\mu$ events with average pile-up of 69. The timing of the slowest core crate is shown. If the L1 input rate

were too large for the FTK system, there would be a backlog of track reconstruction and the event latency would steadily increase. This does not happen as shown in the right figure of Fig.2; thus the FTK system operates well with pile-up of 69.
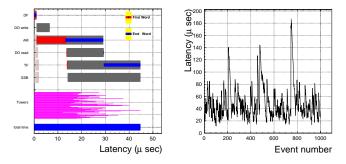

Fig. 2. FTK latency for Z$\to \mu\mu$ events with 69 pile-up. The left plot shows an example of 1 event. The timing of the functional blocks is given for the core crate that takes the most time. The time for each of the 64 regions is shown below that, with the total latency shown in the bottom bar. The right plot shows the event-by-event latency. For each event, the execution time starts when the event is available (10 μs after the previous event, corresponding to a 100 kHz level-1 trigger rate) and ends when the FTK has completed analyzing that event [2].

B. Physics Performance

The full simulation framework in the ATLAS detector is used to produce physics data samples for FTK performance studies. The FTK has a reasonable reconstruction efficiency with respect to offline tracks. The efficiency is above 90% for $p_T > 10$ GeV/c, and is flat over the full rapidity region. In addition, the FTK reconstruction is robust against pile-up, which is important for future high luminosity LHC operation.
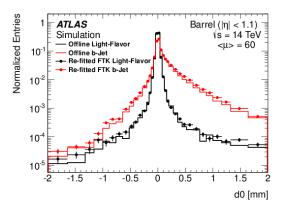

Fig. 3. The transverse impact parameter for tracks associated to light-flavor (black) and heavy-flavor (red) jets. The solid lines show the distribution for the offline tracks, whereas the points show the re-fitted FTK tracks [4].

The availability of full tracking information from the FTK will allow more sophisticated HLT selection. It opens the possibility of applying b-tagging and τ selection algorithms with near offline performance as the very first steps of the HLT selection, when the event rate is still very high. The transverse impact parameter is the basic input for b-tagging. This is sensitive to the long lifetime of particles within the

b-jet and enables algorithms to separate b-jets from light jets. Figure.3 shows the signed transverse impact parameter for re-fitted FTK tracks and offline tracks associated to light-flavor and heavy-flavor jets. There is a longer positive tail in the distributions for heavy-flavor jets as expected for particles from B-hadron decays. Identification of hadronically decaying τ leptons is based on the presence of 1 or 3 tracks in a very narrow cone with little track activity in a surrounding isolation region. Figure.4 shows the track multiplicity of tau candidates computed with FTK tracks in an isolation region. The HLT can rapidly reject the multijet backgrounds by requiring that the FTK track multiplicity in the isolation region is less than or equal to 2. In addition, since the processing time of track reconstruction is greatly reduced, the tau identification algorithm potentially can be improved in the future, such as by adding the reconstruction of neutral pions from the calorimeter data. The FTK also makes possible the reconstruction of the primary vertices in HLT. In MC simulation studies, the number of vertices obtained with offline reconstruction and that with FTK tracks have a linear correlation. Applying corrections based on the number of primary vertices provides improved robustness of the HLT selections against pile-up.
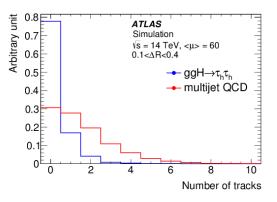

Fig. 4. Track multiplicities of tau candidates computed with FTK tracks in an isolation region with 0.1< ΔR <0.4 ($\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ from the leading FTK track). A signal sample of gluon fusion H$\to \tau\tau$ events is shown in blue, and multijet background sample is shown in red. Both signal and background samples are produced with pile-up of 60 at $\sqrt{s}$ =14 TeV [4].

V. CONCLUSION

The FTK system will provide global tracking information for events accepted by the L1 trigger. The timing simulation of the FTK shows that it performs well with pile-up of 69. MC simulation studies show that FTK allows more sophisticated algorithms in the HLT. The inclusion of the FTK system into the ATLAS trigger system is planned to start in summer 2015.

REFERENCES

[1] ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, 2008 JINST 3 S08003

[2] ATLAS Collaboration, Fast TracKer (FTK) Technical Design Report, CERN-LHCC-2013-007

[3] ATLAS Collaboration, ATLAS Insertable B-Layer Technical Design Report, CERN-LHCC-2010-013

[4] http://twiki.cern.ch/twiki/bin/view/AtlasPublic/FTKPublicResults