The FTK: A Hardware Track Finder for the ATLAS Trigger

The ATLAS FTK Upgrade Project [1]

Abstract–**The ATLAS experiment trigger system is designed to reduce the event rate, at the LHC design luminosity of 10^{34} cm^{-2} s^{-1}, from the nominal bunch crossing rate of 40 MHz to less than 1 kHz for permanent storage. During Run 1, the LHC has performed exceptionally well, routinely exceeding the design luminosity. From 2015 the LHC is due to operate with higher still luminosities. This will place a significant load on the High Level Trigger system, both due to the need for more sophisticated algorithms to reject background, and from the larger data volumes that will need to be processed. The Fast TracKer is a hardware upgrade for Run 2, consisting of a custom electronics system that will operate at the full rate for Level-1 accepted events of 100 kHz and provide high quality tracks at the beginning of processing in the High Level Trigger. This will perform track reconstruction using hardware with massive parallelism using associative memories and FPGAs. The availability of the full tracking information will enable robust trigger selection within the affordable latency available at the High Level Trigger, with only a limited degradation in performance arising from the additional pileup from higher luminosity running.**

I. Introduction

IN order for the ATLAS experiment [2] to reduce the event rate to the level at which only interesting events will be fully reconstructed, a three-level trigger system [3] has been deployed. The level 1 trigger (L1) reduces the rate to 100 kHz using custom, pipelined electronics and identifies Regions of Interest (RoI) worthy of further study in the trigger. The Region of Interest Builder (RoIB) delivers the RoI records to the level 2 trigger (L2) which runs selection algorithms on a farm of commodity processors to further reduce the rate to approximately 4.5 kHz. Finally the Event Filter (EF) reduces the rate to approximately 400 Hz for permanent storage. The time budget for processing events at L2, and the EF - referred to collectively as the High Level trigger (HLT) - is approximately 40 ms and approximately 1 s respectively. The track reconstruction at L2 therefore cannot be as detailed as that of the EF.

The Fast TracKer (FTK) [4] will perform global track reconstruction immediately following the L1 accept, the results being available at the start of HLT processing. It is a system of custom electronics that will rapidly find and fit tracks in the ATLAS inner tracking detectors for all events accepted by the L1 trigger for the full tracking volume - not just within the RoIs identified by L1. The core functionality consists of pattern recognition and a track fit. Pattern recognition is carried out by dedicated Associative Memory (AM) devices [5], which find track candidates using coarse resolution patterns, referred to as roads. This part uses massive parallelism to perform what is usually the most CPU intensive aspect of tracking, by processing approximately 10^9 roads simultaneously as the silicon data pass through the system. When a road has hits in enough silicon layers, an initial track fit is performed by DSPs running on FPGAs with the full resolution hits from the road to determine the track helix parameters and goodness of fit. Since an architecture which uses all four Pixel detector (PIX) layers and eight SemiConductor Tracker (SCT) layers for pattern matching is not attainable, eight of the 12 silicon layers are used to perform pattern recognition and perform the initial track fitting. Tracks from the first stage pass to a second stage where they are extrapolated into the other 4 silicon layers not used in the first stage. Nearby hits in those layers are found and the tracks are refitted using the hits in all 12 layers.

II. System Overview

The FTK is organized as a set of independent processing engines, each covering a different detector region. The detector coverage is divided into azimuthal regions further into η regions. The potential inefficiency at region boundaries is remedied by having overlap between regions. At the luminosities up to 3×10^{34} cm^{-2}s^{-1}, a segmentation of 64 η-φ towers, 16 in φ by 4 in η, ensure high efficiency for tracks with p$_T$ of 1 GeV/c and above. With such a detector segmentation, the data for one tower can be distributed on designated parallel buses at the full 100 kHz rate.
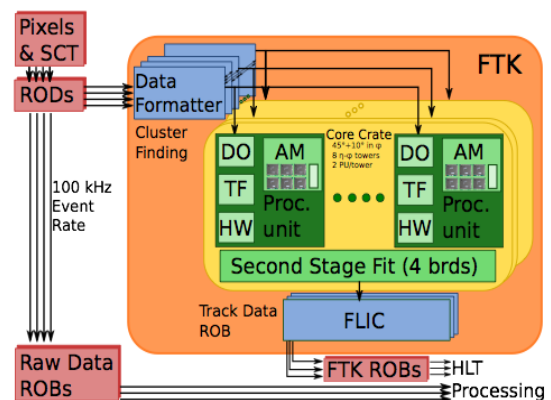
A schematic illustrating the system function is shown in Fig. 1. The PIX and SCT data are transmitted from the detector Readout Drivers (RODs) and received by the Data Formatters (DF). Mezzanine cards on each DF perform cluster finding before the data are reorganized into η−φ towers and the DF transmits the cluster centroids of the eight layers to the processing units.

The Data Organizers (DO) store hits from the DFs at full resolution and also convert them to coarse resolution superstrips (SS) appropriate for pattern recognition in the AM system. The DOs hold smart databases where full resolution hits are stored in a format that allows rapid access based on the pattern recognition road ID and then retrieved when the AM finds roads with the requisite number of hits.

The AM system contains AM chips which contain a very large number of preloaded patterns, corresponding to the possible combinations for real tracks passing through a SS in each silicon layer. These are determined in advance from a full ATLAS simulation of single tracks using detector alignment extracted from real data. The AM system is a massively parallel system in that each hit is compared with all patterns simultaneously.

After being found by the AM system, roads are returned to the DOs, which immediately fetch the associated full resolution hits and send them, together with the road, to the Track Fitter (TF). Because each road is quite narrow, the TF can obtain helix parameters with high resolution from a linear fit using the local coordinates in each layer. Such a fit is extremely fast and a modern FPGA can fit approximately 10^9 track candidates per second.

Following fitting, duplicate track removal is carried out by the Hit Warrior (HW) for those 8-layer tracks in a road.

When a track passes the quality cuts of the 8-layer fit, the road number and hits are sent to the Second Stage Board (SSB). The track is extrapolated into the 4 additional layers, nearby hits are found, and a full l2-layer fit is carried out. Duplicate track removal is again applied.

SSB output data consisting of the hits on the track and the helix parameters of the track are sent to the FTK Level2 Interface Crate (FLIC). The FLIC formats and organizes the tracks and sends them to the HLT using the standard ATLAS data transmission protocols, and carries out monitoring functions.

III. DESIGN AND DEVELOPMENT

The DF design is based on Advanced Telecom Computing Architecture (ATCA) technology. Each DF board is connected to the others with multiple point-to-point links in the full mesh ATCA backplane. Each DF board contains two FPGAs, to process two towers. Eight boards are assigned to each 14-slot ATCA shelf and the full system consists of 32 boards in four ATCA shelves. Each DF board has a Rear Transition Module (RTM) which supports up to eight QSFP+ and six SFP+ transceivers. These transceivers will be used for the DO, SSB and inter-shelf communication.

Pattern matching and first stage fitting are performed in the Processor Units (PUs) in the 9U VME crates (core crates). The full FTK system will have 128 PUs in eight 9U VME crates. Each PU consists of an AM system with a large auxiliary board (AUX) behind it. The AUX card holds four FPGAs to perform the DO, TF and HW functions. The hits from the DFs enter through two QSFP+ connectors and the SS number for each hit is generated in one of the two Input FPGAs. The SS numbers are sent through the VME J3 connector to the AM system.

The AM system consists of two types of boards, the 9U VME board (AMB) on which are mounted four local associative memory boards (LAMB), with each holding up to 32 AM chips. The current prototypes of AMB and LAMB have been designed to utilize the current version of AM chip (AMChip04) and comply with the power consumption requirement. The final version of AMB and LAMB will be totally based on serial communication. The next version of AM chip (AMchip06) will have capacity of 128k patterns and lower power consumption per unit. The full FTK system will require 8192 AMChip06 chips.

Fig. 2. The prototype of individual FTK components: DF with RTM, AMB with one LAMB mounted, AUX, SSB, FLIC.

The SSBs reside in the same core crate. Each SSB receives through SFP+ transceivers of its own RTM the output from AUX cards and the hits on the additional layers from the DFs. Four FPGAs on the main board will handle the extrapolation, the TF and HW functionalities. After processing the SSB merges the data in a core crate which is sent to the FLIC.

The FLIC is implemented in a single separate 6-slot ATCA shelf with full mesh backplane. The core crates use SFP links to transfer data from the SSBs to the FLIC Input Card. The data from the core crates are processed on the Input Card and then passed via the ATCA Zone 3 connectors to the Output Card, which is a RTM. The Output Card holds SFP connections to the Readout Systems (ROSs). To facilitate optional trigger processing in the future, the event data

streams are copied to ATCA Zone 2 fabric thus can be collected into a full record by a processor blade in the shelf.

All boards are being prototyped and the intensive tests are being performed. Fig. 2 shows the prototypes of some major components.

IV. PERFORMANCE

A dedicated simulation tool for estimating the FTK latency has been developed for tuning the system architecture and parameters and ensuring that the FTK can handle at least 100 kHz L1 rate at the higher luminosities. The tool emulates the major FTK 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. Detailed studies have been performed for the most time consuming steps, from DO write mode through TF. The DF adds little to the overall latency since each cluster found is immediately sent to the DO thus the DF and DO execution times almost completely overlap. Similarly the HW has a relatively short latency for each track processed.

For each functional block, the time of the transfer of the first and last words into and out of the block are calculated. The numbers of input and output words for each block come from simulated events. 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 processing time per word for each block type is estimated from the architecture and the experience with prototypes. As shown in Fig. 3, the FTK can finish global tracking in approximately 25 μs for a typical ATLAS event at the luminosity of 3×10^{34} cm^{-2}s^{-1}, in contrast to more than 100 ms required for the full detector tracking in the current HLT.
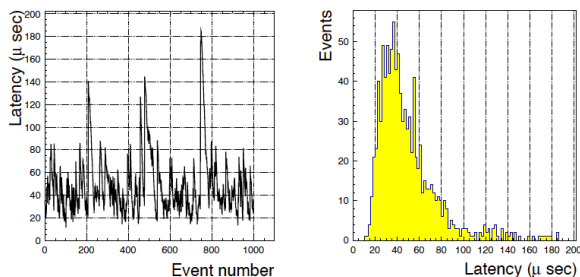

Fig. 3. The FTK latency: the event-by-event latency (left) and a distribution histogram (right).

The standard ATLAS Monte Carlo (MC) simulation framework [6] is being used to produce physics data samples for FTK performance studies. The FTK has a reasonable efficiency with respect to tracks from the offline reconstruction [4]. Therefore the immediate availability of FTK tracks for the full detector immediately following a L1 accept would free significant resources from the HLT to allow more sophisticated tagging algorithms and open the possibility of b-tagging performance and rejection close to that of offline in events where HLT tracking is too costly. This will be particularly important at very high luminosities. In hadron collider experiments τ jet identification is greatly based on the presence of 1 or 3 tracks in a very narrow cone with little or no track activity in a surrounding isolation cone. With FTK tracks the HLT can perform rapid rejection of the QCD background for τ selection. The preliminary results based on MC physics sample studies show that FTK tracking does nearly as well as offline tracking for b tagging and τ selection, and gives similar efficiencies and background rejection at high luminosities.

V. SUMMARY

The ATLAS trigger system will be greatly enhanced by the FTK. The FTK project will complete the component design, test each board and integrate with the ATLAS systems and prepare a system for data taking in 2015. As the LHC luminosity increases, the availability of tracks from the FTK will be essential to ensure that the excellent physics performance of the ATLAS detector can be maintained.

REFERENCES

[1] ATLAS Collaboration, "The FTK Authorlist", ATL-DAQ-PUB-2014-001, CERN, Geneva, 2014, https://cds.cern.ch/record/1709835.
[2] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", 2008 JINST 3 S08003.
[3] ATLAS Collaboration, "ATLAS High-Level Trigger Data Acquisition and Controls Technical Design Report", CERN/LHCC/2003-022 (2003).
[4] ATLAS Collaboration, "Fast TracKer (FTK) Technical Design Report", CERN-LHCC-2013-007.
[5] A. Annovi et al., "A VLSI Processor for Fast Track Finding Based on Content Addressable Memories", IEEE Trans. Nucl. Sci. 53 (2006) 2428.
[6] ATLAS Collaboration, "The ATLAS Simulation Infrastructure", Eur. Phys. J. C70 (2010) 823.