FERMILAB-CONF-11-868-PPD

PROCEEDINGS OF SCIENCE

ATLAS FTK: Fast Track Trigger

S. Amerio[j,*], A. Andreani[gh], A. Andreazza[gh], A. Annovi[f], M. Beretta[f], V. Bevacqua[no], G. Blazey[ai], M. Bogdan[c], E. Bossini[q], A. Boveia[c], V. Cavaliere[s], F. Canelli[cd], F. Cervigni[no], Y. Cheng[c], M. Citterio[g], F. Crescioli[no], M. Dell'Orso[no], G. Drake[a], M. Dunford[c], P. Giannetti[n], F. Giorgi[b], J. Hoff[d], A. Kapliy[c], M. Kasten[s], Y. K. Kim[cd], N. Kimura[r], A. Lanza[k], V. Liberali[gh], T. Liu[d], D. Magalotti[m], A. McCarn[s], C. Melachrinos[c], C. Meroni[g], A. Negri[kl], M. Neubauer[s], B. Penning[d], M. Piendibene[no], J. Proudfoot[a], M. Riva[gh], C. Roda[no], F. Sabatini[g], I. Sacco[p], M. Shochet[c], A. Stabile[gh], F. Tang[c], J. Tang[c], R. Tripiccione[e], Joseph Tuggle[†c], V. Vercesi[k], M. Verzocchi[d], M. Villa[b], R.A. Vitillo[n], G. Volpi[f,‡], J. Webster[c], J. Wu[d], K. Yorita[r] and J. Zhang[a]

[a] *Argonne National Laboratory*
[b] *INFN and University of Bologna*
[c] *University of Chicago*
[d] *Fermi National Accelerator Laboratory*
[e] *University of Ferrara*
[f] *INFN LNF*
[g] *INFN and* [h] *University of Milan*
[i] *Northern Illinois University*
[j] *INFN Padova*
[k] *INFN and* [l] *University of Pavia*
[m] *INFN and University of Perugia*
[n] *INFN and* [o] *University of Pisa*
[p] *Institut für Technische Informatik der Universität Heidelberg*
[q] *INFN and University of Siena*
[r] *Waseda University*
[s] *University of Illinois at Urbana-Champaign*
E-mail: jtuggle@uchicago.edu

A track reconstruction system for the trigger of the ATLAS detector at the Large Hadron Collider is described. The Fast Tracker is a highly parallel hardware system designed to operate at the Level-1 trigger output rate. It will provide high-quality tracks reconstructed over the entire inner detector by the start of processing in the Level-2 trigger. The system is based on associative memories for pattern recognition and fast FPGA's for track reconstruction. Its design and expected performance under instantaneous luminosities up to $3 \times 10^{34}/\mathrm{cm}^2/\mathrm{s}$ are discussed.

The 20th Anniversary International Workshop on Vertex Detectors
June 19 – 24, 2011
Rust, Lake Neusiedl, Austria

[*] Marie Curie Fellow.
[†] Speaker.
[‡] Marie Curie Fellow.

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike Licence. http://pos.sissa.it/

PoS(Vertex 2011)040

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

1. Introduction

The Fast Tracker (FTK) [1] is a trigger upgrade for the ATLAS detector at the Large Hadron Collider (LHC). As the LHC luminosity approaches its design level of $10^{34}/\mathrm{cm}^2/\mathrm{s}$, the combinatoric problem posed by charged particle tracking becomes increasingly difficult. The FTK is a highly-parallel hardware system intended to provide high-quality tracks of transverse momentum above 1 GeV. Its speed and physics performance have been estimated at simulated luminosities up to $3 \times 10^{34}/\mathrm{cm}^2/\mathrm{s}$, corresponding to 75 interactions per bunch crossing. An overview of the ATLAS detector and trigger system is given in the next section, followed by a discussion of the motivations behind building the FTK, an overview of the system, the details of pattern recognition and track fitting, and the predicted performance. A new scheme to introduce variable-width patterns is discussed before concluding with an outlook and summary.

2. The ATLAS Detector and Trigger

ATLAS is one of the general-purpose detectors at the LHC [2]. It is designed to detect the products of proton-proton collisions at a center of mass energy of 14 TeV. Particles are detected by the inner tracking detector, electromagnetic and hadronic calorimeters, and a muon system. The inner tracker is inside a 2 T solenoidal magnetic field. Since 2010 ATLAS has been taking data with the LHC running at a center-of-mass energy of 7 TeV; a peak instantaneous luminosity of $2 \times 10^{33}/\mathrm{cm}^2/\mathrm{s}$ has been achieved.

The FTK will use information from the two inner tracker silicon systems (Chapter 4 in [2]): the pixel detectors and the silicon strip detectors (SCT) . An illustration is shown in Figure 1. The pixel system consists of three cylindrical barrel layers and three forward disks on each side of the barrel. The pixels are $50x400\mu\mathrm{m}^2$, comprising a total of 80 million channels. The SCT consists of eight barrel layers grouped into four pairs. Each pair has one layer of strips parallel to the detector axis and another layer with strips offset by an angle of 40 mrad. There are 9 pairs of disk layers on each side of the barrel. The strip pitch is on average 80μm, and there are 6 million readout channels.

To select interesting collisions from up to 40 MHz of bunch crossings, a three-level trigger system is used [3]. The first level is hardware-based, and has a maximum output rate of 100 kHz. It provides regions of interest – portions of the detector in azimuth and pseudorapidity – to the next two levels. Level 2 and the Event Filter are implemented in software and are colletively known as the High Level Trigger (HLT). The HLT runs on a farm of several thousand CPUs. Level 2 operates within regions of interest with an output rate of 2 kHz, while the Event Filter has access to all information in the event, with an output rate of 200 Hz.

3. Physics with FTK

Although the source of electroweak symmetry breaking is not known, it couples to the mass of particles. This makes processes involving the third generation of matter more likely to yield crucial information. Top quark decays are identified in large part though the reconstruction of b quark jets, which are in turn identified by tracks coming from a secondary vertex. Decays of τ leptons are

PoS(Vertex 2011)040

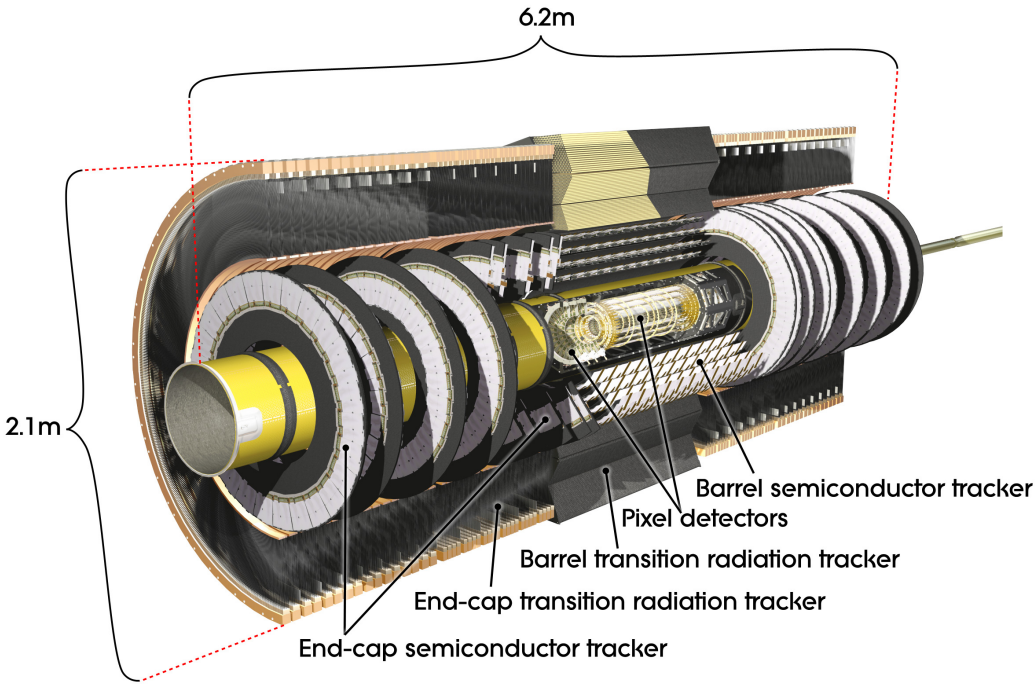

Figure 1: Cut-away view of the ATLAS inner detector.

distinguished by comparing the number of tracks inside a small signal cone with the number found inside a larger isolation cone.

The W and Z gauge bosons are also important in the search for new physics. These can be identified by isolated muons and electrons (hereafter referred to generically as leptons). Isolation is typically determined using information from the calorimeter. However, the presence of multiple interactions per bunch crossing (pile-up) adds energy that is uncorrelated with the interaction of interest, leading to increased background. If energy thresholds are simply raised, signal efficiency suffers. Alternatively, the isolation can be determined by tracking, considering only tracks coming from the same vertex as the lepton in question. This can improve performance by filtering out irrelevant information from pile-up.

These examples demonstrate that tracking can play an important role in identifying interesting events. Because of the need for massive online data reduction, the use of tracking information becomes more important as the luminosity increases. The FTK will operate at the Level-1 trigger output rate, reconstructing all tracks at near-offline quality for use in the HLT. The CPU time that would have been used for track reconstruction in the HLT can be used to further optimize the software triggers.

4. FTK Overview

The FTK is a system based on FPGAs and a highly-parallel Associative Memory (AM) chip [4]. It is modeled on the successful Silicon Vertex Trigger at the Collider Detector at Fermilab [5, 6]. The system is divided into eight core crates, each covering a range of detector azimuth, with suf-

PoS(Vertex 2011)040

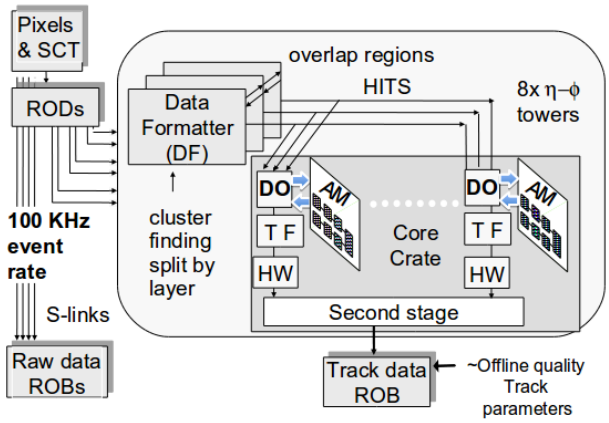

Figure 2: A schematic showing the FTK components and their connection to the ATLAS data acquisition system.

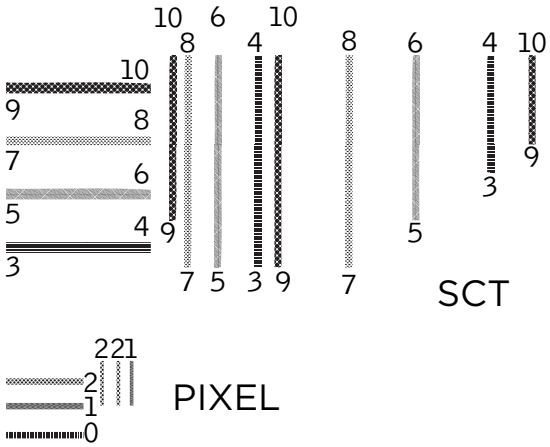

Figure 3: The logical layer assignments (0-10) for each of the physical silicon layers.

ficient overlap to reconstruct tracks with at least 1 GeV of transverse momentum (p_T). The core crates are subdivided into four regions of pseudorapidity and two regions of azimuth, yielding 8 $\eta - \phi$ towers per crate. Each $\eta - \phi$ tower is served by two *processing units*, which consist of the AM chips for pattern recognition, a Track Fitter, a Data Organizer to coordinate between pattern recognition and track fitting, and a Hit Warrior to remove duplicate tracks from the output. Each of these components and an external Data Formatter board are described below. A schematic is given in Figure 2.

Silicon hits enter the FTK system through the Data Formatter board, which performs hit clustering, maps the physical detector layers to the 11 logical layers of the FTK system (Figure 3), then routes the clusters to the relevant processing units. Within a processing unit, Data Organizers coarse-grain the clusters into *superstrips* to be used in pattern recognition. The resolution of the superstrips is approximately an order of magnitude larger than that of the incoming clusters. The full-resolution clusters are stored for later use. The Data Organizer then passes the superstrips to the AM chips for pattern recognition.

PoS(Vertex 2011)040

The AM processors operate in parallel to compare the incoming superstrips to precomputed patterns. For the luminosity expected after the LHC Phase-I upgrades, several hundred million precomputed patterns per core crate will be required. As patterns are found, they are sent back to the Data Organizer, which looks up the full-resolution hits within a pattern and sends them to the Track Fitter board. Instead of a computationally-intensive helix fit, the FPGA-based Track Fitter board performs a fast linear fit, described in the next section. Every combination of hits within a found pattern is fitted, and only those which pass goodness-of-fit criteria are kept. The final stage of the processing unit is the Hit Warrior. It removes duplicate tracks based on the number of shared hits between track candidates and the goodness-of-fit.

5. Pattern Recognition and Track Fitting

The FTK will perform tracking in two stages in order to keep the combinatorics to a manageable level. The first stage uses 8 out of 11 silicon layers for pattern recognition and track fitting, while the second stage refits the tracks found in the first stage using all 11 layers.

For the first stage, the detector layers used in the barrel are all three pixel layers, all four axial SCT layers, and two of the small-angle stereo SCT layers. In the forward disks, the mapping between physical and logical layers is such that tracks will cross eight logical layers across the full rapidity range. Seven out of eight hits will be required to match a pattern.

The associative memories are custom-built chips that store 50,000-100,000 precomputed track patterns. Incoming silicon clusters are compared to all stored patterns simultaneously. Groups of 32 AM chips are linked through a local associative memory board (LAMB). There are four LAMBs on each AM board. Once a LAMB has found all matched patterns (roads) in an event, it sends the information to a Data Organizer, four of which (one for each LAMB) sit on an auxiliary (AUX) card attached to the AM board. The AUX card also has four copies each of the Track Fitter and Hit Warrior. Upon receiving the found roads from a LAMB, the Data Organizer looks up the full resolution hits within a road and sends them on to the Track Fitter.

The Track Fitter unit is expected to perform one track fit per nanosecond. It achieves this high rate by performing a simple linear fit for each set of N hit coordinates x_j:

$$p_i = \sum_j^N c_{ij} x_j + q_i. \tag{5.1}$$

The p_i are the five helix parameters and the $N-5$ χ^2 components, all of which are determined by the constants c_{ij} and q_i. The constants for each fit are selected depending on the detector modules in which the hits are found. Approximately 100,000 sets of constants per core crate are expected to be used.

If a fit combination is above the configurable χ^2 threshold, a so-called majority recovery is performed. The combination is refit several times, ignoring one of the hits each time to check whether a subset of hits will yield a good χ^2. This allows a track to be recovered if one of the hits in the combination was due to noise or a different track.

The second stage of fitting uses the good tracks from the first stage to look for expected hits in the remaining 3 layers, using an inversion of Equation 5.1. If at least 10 out of 11 hits are found, the track is refit with all of the found hits and subjected to a final goodness-of-fit test.

PoS(Vertex 2011)040

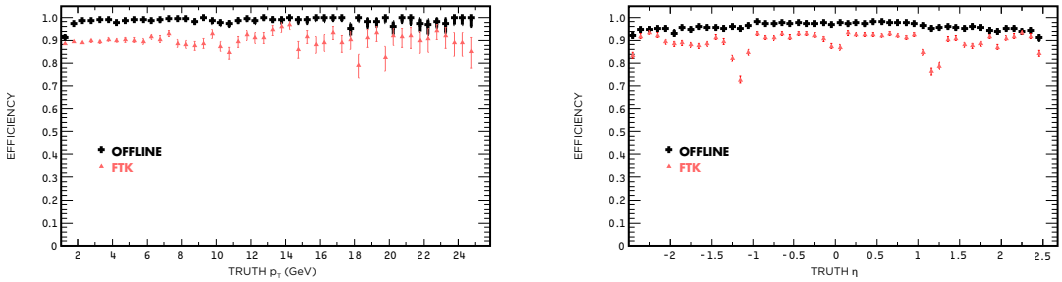

Figure 4: Comparison of FTK and offline efficiency to reconstruct muons as a function of p_T (left) and pseudorapidity (right).

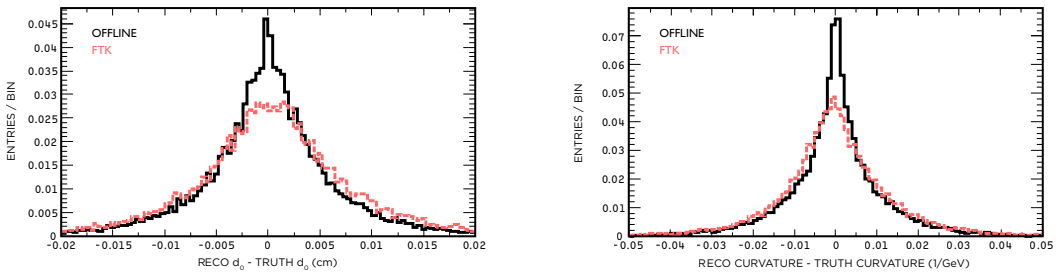

Figure 5: Comparison of FTK and offline resolution for the transverse impact parameter (left) and the curvature (right).

6. Simulated Performance

Each of the FTK components has been simulated in software to tune the system and determine its expected performance. FTK tracks are compared with tracks reconstructed using the ATLAS offline tracking algorithm [7].

Figure 4 shows the efficiency of the FTK and offline algorithm to reconstruct muons of $p_T > 1$ GeV. The FTK configuration uses pattern banks and constants suitable for up to 75 pile-up interactions. The efficiency dips for FTK tracks around pseudorapidity $|\eta| = 1$ are at the transition between the barrel and forward disks. This will be improved by relaxing the number of required hits on a track for this region. The resolution of track parameters is shown in Figure 5 for tracks with $|\eta| < 1$, demonstrating that the FTK resolution is comparable to that of the offline tracking.

The potential for b jet tagging using FTK tracks at high luminosity is evaluated using an algorithm based on the transverse impact parameter d_0 of tracks in jets. Template histograms of the impact parameter divided by its uncertainty are created for light-flavor jets and b jets. These are combined to form a likelihood, the expected performance of which is shown in Figure 6 for FTK and offline tracks. The simulated event sample is WH with $H \to b\bar{b}$ at $3 \times 10^{34}/\text{cm}^2/\text{s}$. The operating point for the trigger system is at 80% efficiency.

The performance is also evaluated for τ lepton identification. To separate τ decays from multijet backgrounds, one looks for 1 or 3 tracks inside a narrow cone with few or no tracks in a surrounding isolation cone. A simulated sample of vector boson fusion Higgs production with $H \to \tau\tau$ at $3 \times 10^{34}/\text{cm}^2/\text{s}$ is used to test the FTK performance. Although the selection criteria

PoS(Vertex 2011)040

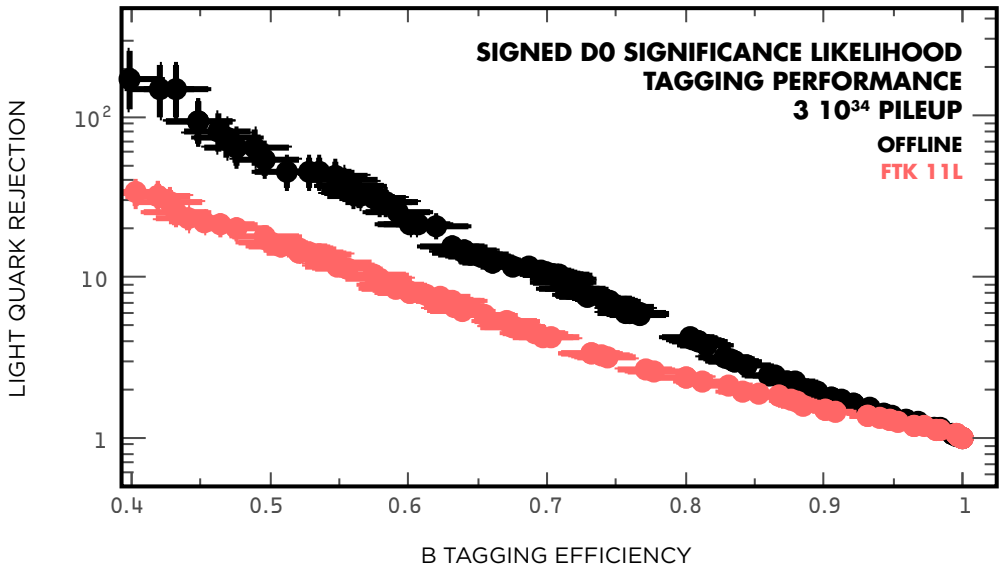

Figure 6: Light-flavor jet rejection versus b jet signal efficiency for the FTK and offline tracking on a simulated sample of WH events ($H \to b\bar{b}$) at $3 \times 10^{34}/\text{cm}^2/\text{s}$.

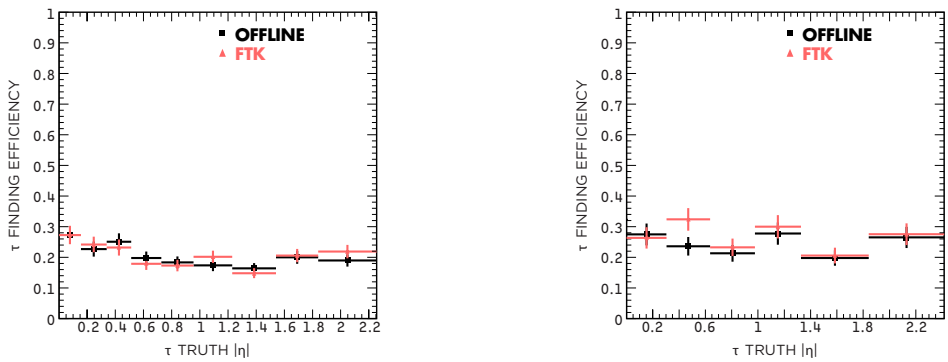

Figure 7: Comparison of FTK and offline τ selection efficiency versus pseudorapidity for 1-prong (left) and 3-prong (right) τ decays. The sample is VBF Higgs production with $H \to \tau\tau$ at $3 \times 10^{34}/\text{cm}^2/\text{s}$.

for τ identification at high luminosity have not been optimized, both FTK and offline tracks yield similar τ efficiency. Figure 7 shows the efficiency as a function pseudorapidity for 1-prong and 3-prong τ decays.

Single high-p_T lepton identification will become more challenging as the luminosity increases. To select leptons from heavy vector boson decays, isolation criteria are normally used to discriminate against backgrounds. With additional interactions per bunch crossing the total amount of energy in the calorimeter increases, which decreases the effectiveness of isolation criteria if one uses only calorimeter information. A study on isolated muon identification has been performed in the context of the FTK system. As shown in Figure 8 (left), the effectiveness of calorimeter-based

PoS(Vertex 2011)040

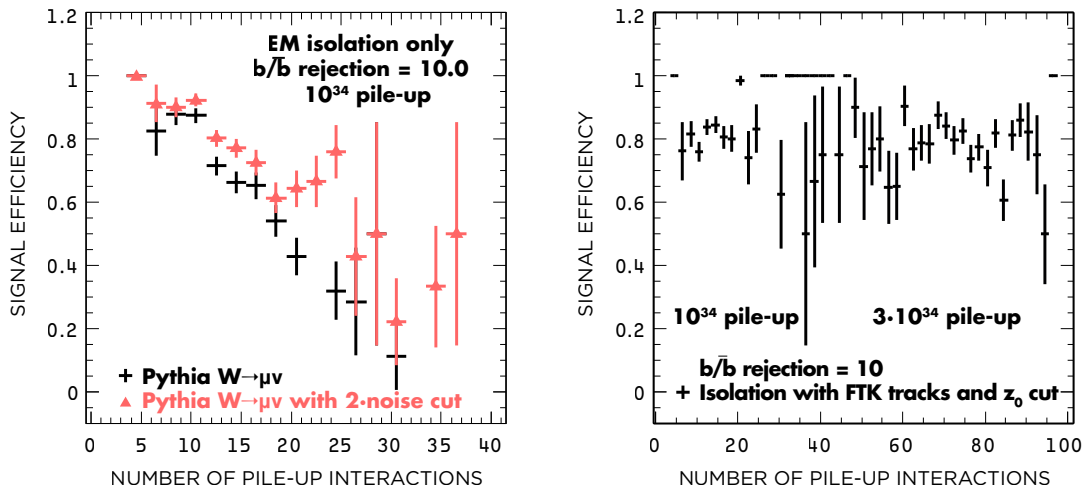

Figure 8: Left: isolated muon identification efficiency as a function of the number of pile-up events for a constant background rejection factor, using only calorimeter information. Right: The identification efficiency using FTK tracking information to ignore tracks not coming from the muon vertex.

isolation drops rapidly with the number of pile-up interactions, for a constant rejection factor of 10 against a background of non-isolated leptons from b-hadron decays. Increasing the calorimeter energy threshold by a factor of two does not significantly improve the efficiency. However, with tracking, one can consider only those tracks pointing to the production vertex of the muon. Figure 8 (right) shows that isolation based on FTK track information yields stable performance even up to 100 pile-up interactions.

The timing of the FTK system has been simulated using a sample of WH events at $3 \times 10^{34}/\mathrm{cm}^2/\mathrm{s}$. By accounting for the time to process each word of data for each FTK component (Data Organizer, Associative Memory, Track Fitter) operating at the full Level-1 trigger output rate of 100 kHz, it is estimated that FTK will take on average 25 μsec per event to perform global tracking, as shown in Figure 9. For comparison, the Level-2 software tracking is estimated to take approximately 1000 times as long per region of interest for the same data. Although the Level-2 software is not yet optimized for high luminosity, the speed advantage offered by FTK is clear.

7. Variable Resolution Patterns

One of the recent challenges in the design of the FTK was the large number of matched roads expected at $3 \times 10^{34}/\mathrm{cm}^2/\mathrm{s}$. Although the system was already segmented into independently-operating $\eta - \phi$ towers, the simulated number of matched roads out of the AM might still overwhelm the Data Organizer. To protect against this, a strategy of variable-resolution patterns was developed. Finer-resolution patterns lead to a lower fake rate at the cost of more required AM space, while coarse resolution maintains higher efficiency at the cost of increased combinatorics for the Track Fitter. The solution described here maintains a balance between these considerations.

The logic of the AM chip design has been upgraded to include local subdivisions of each superstrip in a pattern. During the pattern bank training, for example, if tracks leave hits in both

PoS(Vertex 2011)040

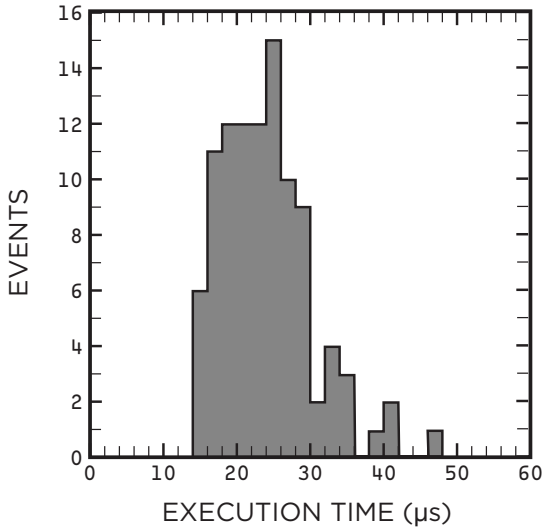

Figure 9: The total FTK execution time for WH events at $3 \times 10^{34}/\text{cm}^2/\text{s}$.

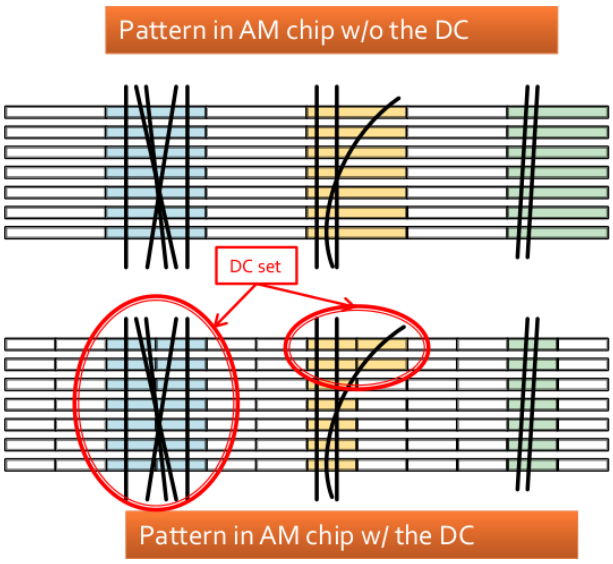

Figure 10: Illustration of the types of patterns that are allowed without (top) and with (bottom) a single "don't-care" bit.

halves of a superstrip, a so-called "Don't-Care" (DC) bit can be set. This means for a given layer, the AM will not care about which half of the superstrip has hits. However, if that bit is not set, another bit is set to indicate which half of the superstrip is important for the pattern. In this second case, the effective resolution for a layer in a pattern is improved by a factor of two. Figure 10 shows the different types of patterns that can arise from this scheme. In the chips currently being designed, up to three DC bits can be set, offering a division in the superstrip width by a factor of up to $2^3 = 8$. This scheme effectively allows the creation of a pattern bank with fine patterns that occupies the same amount of memory as a bank with only coarser patterns.

Simulation studies on events with 40 pile-up interactions have been performed, corresponding

PoS(Vertex 2011)040

to conditions expected after the LHC Phase-I upgrade. Using a pattern bank with a single DC bit in the AM logic, the total dataflow through the system has been evaluated. At that level of pile-up, a fairly uniform 1200-1400 silicon clusters are expected per logical layer per core crate. The output of the AM with the DC bit is about 100,000 matched roads per core crate, leading to 700,000 fit combinations. Timing studies indicate that this is an acceptable level of dataflow, and the potential for more DC bits will mitigate increases in dataflow at high luminosities.

8. Conclusion

It is essential to maintain high efficiency and effective background rejection for events containing b quarks, τ leptons, and isolated, high-p_T electrons and muons. Track reconstruction plays a key role in this, but the necessary computational power increases rapidly with luminosity. The massively-parallel FTK system offers a solution to this problem by performing full-detector track reconstruction before the Level-2 software triggers run. The time saved by FTK can then be used to make more sophisticated trigger decisions.

The AM chip will be submitted for prototyping in late 2011, and tests are ongoing for the AM board development. A full prototype of the latter is expected in 2012. The AUX card is under design with a partial prototype expected in 2011. In winter 2011-2012, the silicon readout will be modified so that a parallel stream of information can be sent to the FTK without distrupting the current system. This is in preparation for commissioning an $\eta - \phi$ slice of the FTK in 2012, with the goal of covering the full barrel during the LHC Phase-I upgrade period, then completing the coverage in stages as soon as possible afterwards.

In summary, the FTK is expected to play a significant role in the ATLAS trigger system. It has performed well in simulations with up to 75 pile-up events, while 40 are expected by the end of this decade. Prototypes of all components of the FTK are expected to be completed within two years, and the full system is foreseen to be in place during the run following the Phase-I upgrades.

References

[1] A. Annovi *et al.*, *IEEE Trans. Nucl. Sci.* **51** (2004) 391.

[2] The ATLAS Collaboration, G. Aad *et al.*, *JINST* **3** (2008) S08003.

[3] The ATLAS Collaboration, G. Aad *et al.*, arXiv:0901.0512 [hep-ex] (2009) 549.

[4] M. Dell'Orso and L. Ristori, *Nucl. Instr. and Meth.* **A 278** (1989) 436.

[5] J. Adelman *et al.*, *Nucl. Instr. and Meth.* **A 572** (2007) 361.

[6] A. Annovi *et al.*, *IEEE Trans. Nucl. Sci.* **53** (2006) 2428.

[7] T. Cornelissen *et al.*, ATLAS Note, ATL-SOFT-PUB-2007-007.

PoS(Vertex 2011)040