

The FastTracker Real Time Processor and Its Impact on Muon Isolation, Tau and b-Jet Online Selections at ATLAS

A. Andreani, A. Andreazza, A. Annovi, M. Beretta, V. Bevacqua, G. Blazey, M. Bogdan, E. Bossini, A. Boveia, V. Cavaliere, F. Canelli, F. Cervigni, Y. Cheng, M. Citterio, F. Crescioli, M. Dell'Orso, G. Drake, M. Dunford, P. Giannetti, F. Giorgi, J. Hoff, A. Kapliy, M. Kasten, Y. K. Kim, N. Kimura, A. Lanza, H. L. Li, V. Liberali, T. Liu, D. Magalotti, A. McCarn, C. Melachrinou, C. Meroni, A. Negri, M. Neubauer, J. Olsen, B. Penning, M. Piendibene, J. Proudfoot, M. Riva, C. Roda, F. Sabatini, I. Sacco, M. Shochet, A. Stabile, F. Tang, J. Tang, R. Tripiccione, J. Tuggle, V. Vercesi, M. Villa, R. A. Vitillo, G. Volpi, J. Webster, K. Yorita and J. Zhang

Abstract—As the LHC luminosity is ramped up to $3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ and beyond, the high rates, multiplicities, and energies of particles seen by the detectors will pose a unique challenge. Only a tiny fraction of the produced collisions can be stored offline and

immense real-time data reduction is needed. An effective trigger system must maintain high trigger efficiencies for the physics we are most interested in while suppressing the enormous QCD backgrounds. This requires massive computing power to minimize the online execution time of complex algorithms. A multi-level trigger is an effective solution to meet this challenge. The Fast Tracker (FTK) is an upgrade to the current ATLAS trigger system that will operate at full Level-1 output rates and provide high-quality tracks reconstructed over the entire inner detector by the start of processing in the Level-2 Trigger. FTK solves the combinatorial challenge inherent to tracking by exploiting the massive parallelism of associative memories that can compare inner detector hits to millions of pre-calculated patterns simultaneously. The tracking problem within matched patterns is further simplified by using pre-computed linearized fitting constants and relying on fast DSPs in modern commercial FPGAs. Overall, FTK is able to compute the helix parameters for all tracks in an event and apply quality cuts in less than 100 μs . The system design is defined and the performance presented with respect to high transverse momentum (high- p_T) Level-2 objects: b jets, tau jets, and isolated leptons. We test FTK algorithms using the full ATLAS simulation with WH events up to $3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ luminosity and compare the FTK results with the offline tracking capability. We present the architecture and the reconstruction performance for the mentioned high- p_T Level-2 objects.

Manuscript received June 13, 2011. This work was supported in part by the U.S. Department of Energy, the U.S. National Science Foundation, Istituto Nazionale di Fisica Nucleare and the European Commission with the FP7-PEOPLE-2009-IOF Programme (G. Volpi is a Marie Curie fellow).

G. Drake, J. Proudfoot, J. Zhang are with Argonne National Laboratory, Argonne, Illinois, USA (e-mail: jinlong@mail.cern.ch).

F. Giorgi and M. Villa are with INFN and University of Bologna, Bologna, Italy.

M. Bogdan, A. Boveia, Y. Cheng, M. Dunford, H. L. Li, A. Kapliy, C. Melachrinou, M. Shochet, F. Tang, J. Tang, J. Tuggle, J. Webster are with University of Chicago, Chicago, Illinois, USA (e-mail: shochet@hep.uchicago.edu).

J. Hoff, T. Liu, J. Olsen, B. Penning are with Fermi National Accelerator Laboratory, Batavia, Illinois, USA (e-mail: thliu@fnal.gov).

F. Canelli and Y. K. Kim are with Fermi National Accelerator Laboratory and University of Chicago, Chicago, Illinois, USA.

R. Tripiccione is with University of Ferrara, Ferrara, Italy.

A. Annovi, M. Beretta, G. Volpi are with INFN LNF, Frascati, Italy.

M. Citterio, C. Meroni, F. Sabatini are with INFN, Milan, Italy, (e-mail: mauro.citterio@mi.infn.it, chiara.meroni@mi.infn.it)

A. Andreazza, A. Andreani, V. Liberali, M. Riva and A. Stabile are with INFN and University of Milan, Milan, Italy (e-mail: valentino.liberali@unimi.it, alberto.stabile@unimi.it).

G. Blazey is with Northern Illinois University, DeKalb, Illinois, USA and with Argonne National Laboratory.

TA. Lanza and V. Vercesi are with INFN, Pavia, Italy (telephone: +39-0382 987430, e-mail: agostino.lanza@pv.infn.it, valerio.vercesi@pv.infn.it)

A. Negri is with INFN and University of Pavia, Pavia, Italy

D. Magalotti is with INFN and University of Perugia, Perugia, Italy

P. Giannetti, R.A. Vitillo are with INFN, Pisa, Italy (telephone: +39-050-2214000, e-mail: paola.giannetti@pi.infn.it)

V. Bevacqua, F. Cervigni, F. Crescioli, M. Dell'Orso, M. Piendibene, C. Roda are with INFN and University of Pisa, Pisa, Italy (e-mail: mauro.dellorso@pi.infn.it).

I. Sacco is with Institut für Technische Informatik der Universität Heidelberg, Mannheim, Germany

E. Bossini is with INFN and University of Siena, Siena, Italy

N. Kimura and K. Yorita are with Waseda University, Tokyo, Japan (e-mail: kohei.yorita@waseda.jp).

V. Cavaliere, M. Kasten, A. McCarn, M. Neubauer are with University of Illinois at Urbana-Champaign, Urbana, Illinois, USA (e-mail: msn@illinois.edu)

I. INTRODUCTION

THE most interesting processes at hadron colliders are likely to be rare and hidden under an extremely large background. Implementing the most powerful selections online is therefore essential to fully exploit the physics potential of experiments where only a very limited fraction of the produced data can be transferred to offline storage. Enormous real-time data reduction must be realized. A multi-level trigger [1] is an effective solution for this task.

Real-time track reconstruction can be an important element in triggering at CERN's Large Hadron Collider (LHC), and even more so after the planned luminosity upgrade. There are numerous examples of the importance of tracking in the trigger. The source of electroweak symmetry breaking couples in proportion to mass. Thus heavy fermions are likely in the

final state, in particular b quarks and τ leptons. High trigger efficiency for these processes requires sensitivity to the generic hadronic decays of the heavy fermions. The challenge comes from the enormous background from QCD-produced light quark and gluon jets, which can be suppressed using tracking information. Tracks coming from a secondary vertex or not pointing to the beam line identify b quark jets, while τ jets can be separated from background using the number of tracks within a narrow “signal cone” and the number in a larger “isolation region”.

Electron and muon triggers can also be improved, particularly at high luminosity, by using track information. Traditionally background is suppressed by applying an isolation requirement using the calorimeters. At high luminosity the energy added by the additional collisions results in either decreased lepton efficiency or increased background contamination. The effect can be greatly ameliorated with track-based isolation only using tracks pointing to the vertex from which the lepton candidate originated.

The Fast TracKer (FTK) [2] will provide global track reconstruction immediately after the ATLAS Level-1 hardware trigger. The near-offline-quality tracks and increased available execution time will allow for the improvement of downstream trigger algorithms. This is only possible with a hardware processor. The addition of tracking to the hardware trigger at hadron collider experiments has been shown, for example at CDF, to be an effective method of improving trigger. The CDF experiment has exploited hardware-based tracking for trigger in both RUN 1 [3] and RUN 2 [4]. FTK is based on the very successful CDF Silicon Vertex Trigger (SVT) [5], [6].

An FTK R&D program has been carried out for a number of years, and prototypes have been built. Details of their design and performance are given in [2], [7] and in the references in those papers. The latest generation of prototypes [8] is used today for a small FTK demonstrator, called the “vertical slice”, which allows early integration in the ATLAS experiment in order to perform real tests and measurements before the final production.

The final system architecture and timing are described in section II. The FTK vertical slice is briefly described in section III. The system performance is described in section IV, including resolution and efficiency for individual tracks and identification capabilities for physics objects. For the first time, it is possible to estimate FTK performance at a luminosity of $3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, with the expected average pile-up of 75 minimum bias interactions. Throughout this paper we will refer to two luminosity scenarios: the LHC design luminosity of $1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ and a “high” luminosity of $3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. We present results obtained with a full simulation of W and Higgs bosons. These channels contain all the signatures we are interested in: b-jets, taus and muons.

II. SYSTEM DESCRIPTION

The ATLAS trigger system [9] consists of three levels. The hardware Level-1 Trigger quickly locates regions of interest in the calorimeter and muon system, operating at rates up to 100 kHz. The subsequent trigger levels, Level-2 and the Event

Filter (EF), are collectively known as the high-level trigger (HLT). They consist of software algorithms running on a farm of commercial CPUs. The Level-2 algorithms may request track information in a Level-1 region of interest while the EF has access to information throughout the entire detector. The final EF output rate is limited to 200 Hz.

FTK is an electronics system that rapidly finds and fits tracks in the ATLAS inner detector silicon layers for every event that passes the Level-1 Trigger. It uses all 11 silicon layers over the full rapidity range covered by the barrel and the disks. It receives a parallel copy of the pixel and silicon strip (SCT) data at the full speed of the readout from the detector front end to the read-out subsystem following a Level-1 Trigger. After processing the hits, FTK sends over S-LINK [10] to the read-out subsystem the helix parameters of all tracks with transverse momentum p_T above a minimum value, typically 1 GeV/c.

The FTK algorithm consists of two sequential steps. In step 1, pattern recognition is carried out by a dedicated device called the Associative Memory (AM) [11], which finds track candidates in coarse-resolution roads. When a road has hits on at least all but one silicon layers, step 2 is carried out in which the full-resolution hits within the road are fit to determine the track helix parameters and a goodness of fit. Tracks that pass a χ^2 cut are kept.

The first step uses massive parallelism to carry out what is usually the most CPU-intensive aspect of tracking by processing hundreds of millions of roads nearly simultaneously as the silicon data pass through FTK. The road width must be optimized. If it is too narrow, the needed size of the AM and hence the cost is too large. If roads are too wide, the load on the track fitters can become excessive due to the large number of uncorrelated hits within a road. This increases both the number of roads the track fitter must process and the number of fits within the road due to the hit combinatorics.

A. System Segmentation

The FTK input bandwidth sets an upper limit on the product of the Level-1 Trigger rate and the number of silicon hits per event. We mitigate this limitation by transporting the silicon hits on multiple 100 MHz buses within FTK. Nevertheless, in order to sustain a 100 kHz Level-1 Trigger rate, it is necessary to organize FTK as a set of independent engines, each working on a different region of the silicon tracker. The first step is to divide the detector into azimuthal regions. The potential inefficiency at region boundaries is avoided by allowing some overlap between regions. At luminosities up to $3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, we plan to use 8 ϕ regions (45° wide) with an overlap of 10° to guarantee high efficiency for tracks with p_T of 1 GeV/c and above. Each region will have its own core processor contained in a 9U VME crate (core crate), for a total of 8 engines working independently. The eight FTK regions are further segmented into η - ϕ towers, each with its own AM and track fitters. The η range of each region is divided into four intervals, and the region’s ϕ range is divided again by two (22.5° plus 10° overlap). A tower receives substantially fewer silicon hits, and the track fitters have substantially fewer

candidates to process than the original regions. The towers are constructed with enough overlap to maintain high efficiency. With this detector segmentation, we can distribute the data on 8 parallel buses at the full 100 kHz Level-1 Trigger rate with the detector occupancy expected at high luminosity.

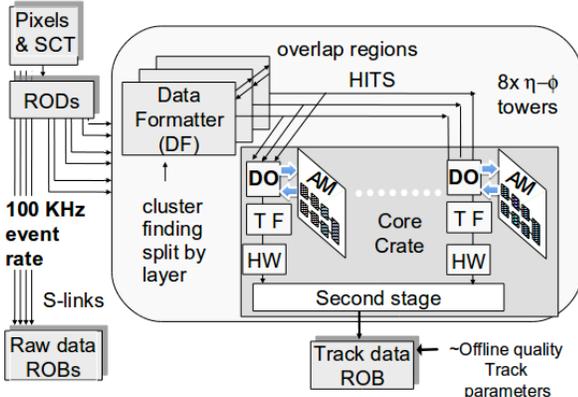


Fig. 1. Functional sketch of an FTK core crate plus its data connections.

B. Dataflow

The pixel and SCT data are transmitted from the front end ReadOut Drivers (RODs), on S-LINK fibers and received by the Data Formatters (DFs) which perform cluster finding (see Fig. 1). The DFs are not partitioned into regions. They organize the detector data into the FTK η - ϕ tower structure for output to the core crates, taking the needed overlap into account. The cluster centroids in each logical layer are sent to the Data Organizers (DOs). The barrel layers and the forward disks are grouped into logical layers so that there are 11 layers over the full rapidity range.

The DO boards are 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. In addition to storing hits at full resolution, the DO also converts them to a coarser resolution, referred to as super-strips (SS), appropriate for pattern recognition in the AM.

The AM boards 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 [12]. The AM is a massively parallel system in that each hit is compared with all patterns nearly simultaneously. When a pattern has been found with the requisite number of hit layers, it is then labeled as a road, and the AM sends the road back to the DOs. They immediately fetch the associated full resolution hits and send them and the road to the Track Fitter (TF). Because each road is quite narrow, the TF can provide high resolution helix parameters using the average parameters across the relevant tracking modules and applying corrections that are linear in the actual hit position in each layer. Fitting a track is thus extremely fast since it consists of a series of multiply-and-accumulate steps. A modern FPGA can fit

approximately 10^9 track candidates per second [13].

C. Duplicates and Fakes

Both the pattern matching and track fitting functions produce duplicates or fakes. A fake road or track is one in which the fraction of hits contributed by a single track is below some threshold, typically 70%.

Duplicate roads occur because we use majority logic in the pattern matching stage. FTK can require N fired layers among the total of M layers (N/M) to declare a road successfully matched. We plan to allow one missed layer in order to keep track-finding efficiency high. The use of the $(M-1)/M$ matching criterion generates duplicate roads. For each real track, it is possible to find a single M/M road (a hit found on each layer) and/or several $(M-1)/M$ roads (with a hit missing on one layer). However, we find duplicate roads are negligible compared to the number of fake roads in high detector occupancy conditions.

Duplicate tracks, on the other hand, are also a source of fake tracks. A pair of duplicate tracks will share most, but not all, of their hits. They occur when a real track has non-associated nearby hits, either from noise or from another track. If the replacement of real track hits by other hits still produces a satisfactory χ^2 , then there will be duplicate tracks. The Hit Warrior (HW) function is applied after track fitting, and reduces the duplicate track rate by keeping only the best χ^2 track among those tracks that share a minimum specified number of hits.

D. FTK Architecture Evolution

The FTK R&D program has steadily evolved towards a more parallel architecture to cope with increasingly difficult working conditions proposed for the LHC. The initial FTK structure [7] was proposed to be organized into only 8 pipelines. Each one contained a single DO and TF, aimed at the low and baseline LHC luminosity. However, the LHC upgrade program has prompted changes to the FTK design.

The major change since the high luminosity simulation became available is the further segmentation and parallelization inside the core crates. The data flow is parallel through η - ϕ towers. Moreover each tower contains multiple DOs, TFs, and HW logic, and only processes the hits within its range. Only the tracks exiting the HW are sent out of the tower hardware logic. The final system will have 64 towers, each one containing 8 DO-TF pipelines working in parallel to process the enormous number of roads produced by the AMs.

The rapid advancement in FPGA technology allows the DO, TF, and HW functions (which were previously 9U VME boards) to fit into a single inexpensive FPGA (Xilinx Spartan or Altera Cyclone families). A clear example of such an advancement was the GigaFitter at CDF [13]: three Virtex 5 chips were able to fit all the combinations previously executed by four 9U TF boards.

A half-tower in FTK, including its own AM, 4 DOs, 4 TFs, and HW function, fits into a single core crate slot, with a main board and an auxiliary card on the back of the crate. We call the logic in a single slot a ‘‘Processor Unit’’. As a result, the

number of crates in the system does not increase, despite the increased number of engines. The ability to place multiple functions in a single slot provides a significant advantage for the data flow between functions. These high rate transfers now occur on short PCB lines rather than across a complex high-speed custom backplane as needed with the previous version of the FTK architecture [7].

E. FTK timing

A simulation tool for calculating FTK execution time was developed to tune the system architecture and parameters, and to ensure that FTK can handle a 100 kHz Level-1 Trigger rate at high luminosity. Within the timing tool, the system is represented by the following functional [pipelined](#) blocks: DF, DO write mode (to receive hits from DF and to send SS to AM), AM, DO read mode (to receive matched roads from AM and to send roads and hits to TF), TF, and HW. [Each block has been made powerful enough to process events with an average time of 10 microseconds. Pipeline stages are separated by derandomizing FIFO memories to allow fluctuations on event complexity.](#) The focus so far has been on the most time-consuming steps, between DO write mode and TF, where the large number of produced roads and needed fits and their fluctuations dominate the processor timing. After the TF the amount of data flowing inside FTK is strongly reduced due to the fact that the χ^2 cut is very effective against fakes, so all steps after the TF can easily work inside the pipeline constraints. Moreover: (a) the DF and DO write mode execution times completely overlap, since each cluster found is immediately sent to the DO; (b) the HW has a constant latency (a few tens of clock cycles) for each track that enters before the track is either sent to the output or discarded; (c) the time required to read and send data to and from the FTK system over SLINKs is parallelized to avoid bottlenecks, so that the transfer is executed in pipeline with the other functions.

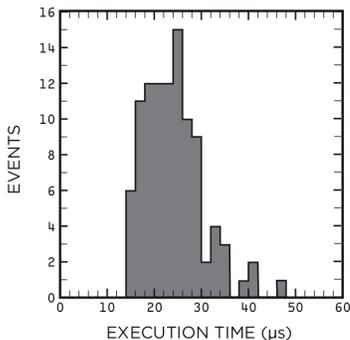


Fig. 2. The FTK execution time for 100 WH events at $3 \times 10^{34} / \text{cm}^2 / \text{s}$.

For each functional block, the time of the first and last words into and out of the block are calculated. Since each core crate operates independently, the FTK event execution time ends when the last word exits the busiest crate for that event. The execution time for a block depends on the number of input words, the processing time per word, and the number of output words. We estimate the processing time per word for each block type from the architecture, our experience with

prototypes (see next section), and the available chips on the market. The numbers of input/output words for each block come from a simulation of the FTK hardware. The results use WH events simulated with high pile-up.

Although each η - ϕ tower has its own unique set of AM roads, the hits from the DF often have to be sent to more than one tower due to the curvature of $p_T \geq 1$ GeV/c tracks, the length in z of the LHC luminous region, and multiple scattering. The amount of hit duplication was determined from the FTK system simulation and used by the timing simulation.

These studies assume very large input FIFO buffers. During the engineering design phase, we will perform queuing studies to determine the necessary depth of each input buffer.

To determine whether FTK can handle the 100 kHz Level-1 Trigger rate, we analyzed a sample of 100 events at high luminosity. If the rate was too high for our system, we would see the event execution time steadily increase as FTK falls behind, working on a stack of previous events before getting to the current one. This does not happen. Some events take longer than others to do global tracking, but after such an event the execution time quickly returns to the typical range. Fig. 2 shows the timing histogram ([the pipeline latency](#)) for FTK operating on high-luminosity events. Although there are uncertainties in the minimum-bias event simulation, this study represents our best estimate of the likely conditions at a high-luminosity LHC. Given this, the system is expected to operate well at a 100 kHz Level-1 Trigger rate at $3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. The results from the vertical slice described in the next section will be used to tune the simulation for future studies.

III. THE VERTICAL SLICE: MEASUREMENTS & TESTS

To understand the system and to test the control software, we plan early parasitic commissioning of a small proto-FTK, based on existing prototypes. The system will reconstruct tracks inside a narrow azimuthal slice (tower) of the detector. Parasitic commissioning means that there will be no impact on normal ATLAS data taking, thanks to an additional output fiber provided for FTK by the tracking front-end. The data flow check can be disabled on the FTK channels, allowing ATLAS to take data regardless of the FTK status. The FTK output can be written to the calibration stream for off-line studies. We call this proto-FTK a “vertical slice” because it will be small (operating on a slice of the detector) but functionally complete from the detector inputs up to the track output available for the Level-2 CPUs.

The vertical slice is installed in a CERN laboratory, in a standalone configuration, using the latest generation of available prototypes [8] (Figure 3): the EDRO board (Event Dispatch and Read-Out) that performs the DF function and the AM board. The input mezzanine (FTK Input Mezzanine, FTK_IM), compatible with both the EDRO and the final FTK DF board, receives data on S-LINKs from the detector and performs hit clustering. The FTK_IM is the final FTK board: it is based on two Spartan VI FPGAs from Xilinx, each receiving four S-LINK optical channels from the silicon detector RODs. It sustains a 40 MHz input data rate over the

S-LINK channels and delivers clusters in real time to the host motherboard. In the laboratory, the detector data are produced by a CPU acting as a pseudo front-end. The clusters are transferred to the AM board that finds roads, which are then provided back to the EDRO. After tests in the laboratory, the vertical slice will be moved to the experiment and will spy on real data during normal data taking.



Fig. 3: The Vertical Slice at CERN. The AM board (right) and the EDRO (left) are shown inside the crate

The AM board is based on the CDF associative memory chip (AMchip03 [6]) and it is compatible with the local associative memory board (LAMB) mezzanines from CDF that can handle 16 or 32 AM chips. It has a maximum capacity of 640k patterns if loaded with 128 AM chips. The final FTK AM board will require a chip with a much larger pattern capacity [14]. Like the final one, the prototype AM board receives a separate bus for each input layer, loading hits from different layers in parallel. The AM board needs a custom backplane with an additional 4-pin 48V power source to provide power to the 128 AM chips per board. An extension of the board in the front has been necessary to allocate large DC-DC converters from 48 Volts down to 1.8 V, the core AM chip voltage. The board has 6 DC-DC converters, each one providing a maximum of 25 A at 1.8 Volts, for a total of 150 A and a maximum power of 270 W. The test stand also includes the hardware necessary to send data to the Level 2 farm (Figure 3, on top of the crate).

We measured the latency times inside the boards and validated the numbers used in the timing simulation described in Section II. We measured 22 A as the core (1.8 V) current sink by a 32-AM-chip LAMB when the system is running at the maximum possible activity. We designed the rack infrastructure to provide the necessary power, based on this measure, and we plan realistic cooling tests for these prototypes.

Several software tools were developed for the Vertical Slice, (most of them will also be useful for the final FTK project) in a modern client/server architecture with plug-ins. With our architecture we are able to write the low level control and monitor routines and automatically export their functionalities to high level programs, both stand-alone and within the ATLAS infrastructure. This software model allows us to add new hardware (e.g., a new prototype AM board, with

different registers or different AM chips) or new functionality (e.g., a new quantity to monitor, a new histogram) with a minimal amount of new code to write [8].

IV. PERFORMANCE AT HIGH LUMINOSITY

A. Single track efficiency and resolution

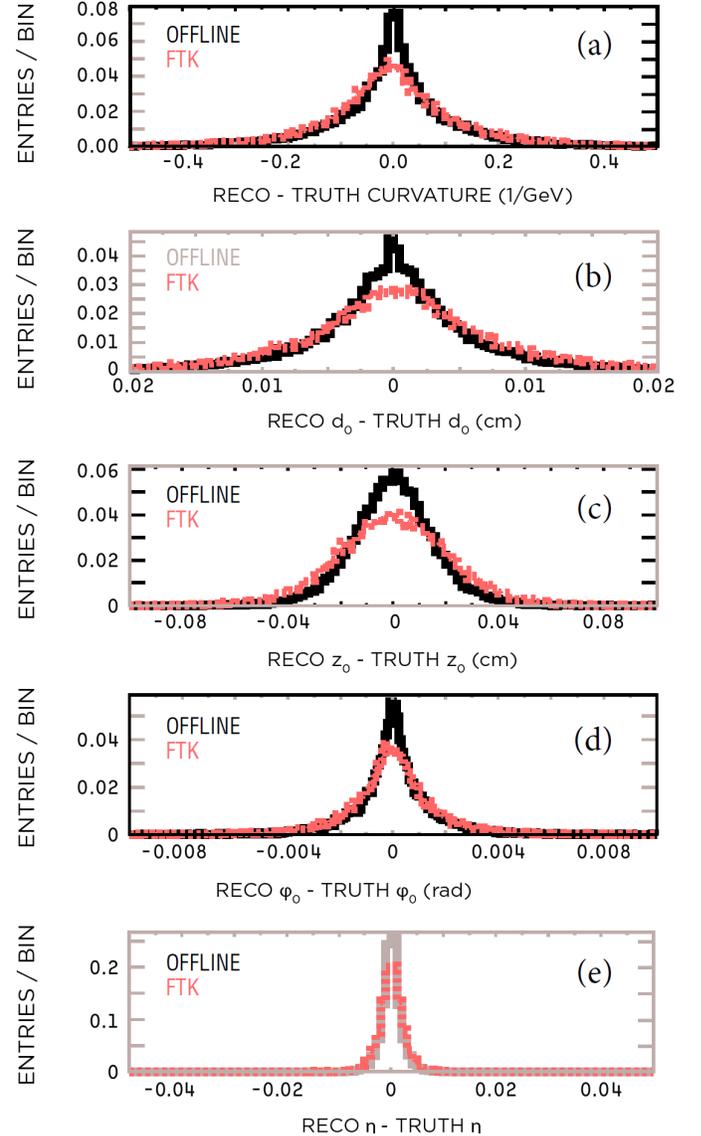


Fig. 4. Comparison of FTK (red) and offline (black) helix parameter resolutions in the barrel region: (a) curvature, (b) d_0 , (c) z_0 , (d) ϕ_0 , and (e) η . The resolution is calculated as the reconstructed (RECO) quantity minus the generated (TRUTH) value.

If FTK tracking is to be useful in e , μ , b , and τ triggering, it must perform well on single tracks. In this section, we show the resolution and efficiency for single-muon events without pile-up, using the FTK settings that would be used at high luminosity. Comparison is made with the ATLAS offline tracking. Our requirements include a minimum of 9 silicon hits for the offline track reconstruction, as expected for high-luminosity operation, and a minimum of 10 hits for FTK, as discussed above.

FTK helix parameter resolutions (the transverse impact parameter d_0 , the longitudinal distance of closest approach z_0 , the azimuth of the momentum direction ϕ_0 , the pseudorapidity η , and the curvature, defined as the inverse of the particle p_T multiplied by the charge) are compared to the offline ones in Fig. 4 for the barrel region. Performance is only slightly degraded with respect to the offline reconstruction.

To calculate efficiency, a reconstructed track is considered to be matched with a generated track if at least 70% of the silicon hits used in the fit were left by the original particle. We consider tracks with generated $p_T > 1 \text{ GeV}/c$.

The track efficiency is shown in Fig. 5 as a function of pseudorapidity and transverse momentum. We are currently implementing an improvement that nearly eliminates the efficiency dip near $|\eta|=1.2$, at the transition between the barrel and forward regions

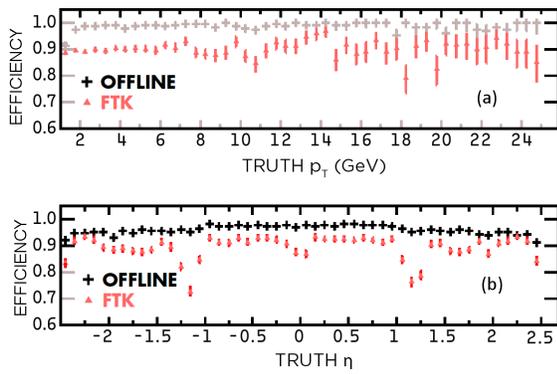


Fig. 5. Single muon tracking efficiency as a function of (a) pseudorapidity and (b) transverse momentum, comparing FTK with offline tracking.

B. *b*-tagging

New physics that couples to heavy fermions may be rich in final-state b quarks but, depending on the process, not necessarily leptons or very high- p_T jets. Given the enormous QCD production of light quarks and gluons, it is important that the ATLAS trigger be able to efficiently select b jets while providing a large rejection factor against other jets. Offline secondary vertex b taggers do well at separating b jets from light jets because they are able to fully exploit very high quality track information. Time constraints in the Level-2 Trigger make both the tracking and tagging difficult, and one must make compromises to perform both for all regions of interest within the current 40 ms Level-2 decision time. Since FTK tracks have near-offline efficiency, helix parameter resolution, and fake rates, the immediate availability of high quality FTK tracks following a Level-1 Trigger would allow the entire 40 ms to be used for more sophisticated tagging algorithms. This could allow the Level-2 Trigger to have an operating point close to that of the offline b -tagging algorithm, increasing the light jet rejection factor at Level-2. Although the current Level-2 Trigger system is expected to perform tracking and b -tagging well up to the design luminosity, the above improvement in background rejection may prove crucial at very high luminosities where the tracking environment will

be much more complex. In the following, we compare FTK and offline tracks [15] in aspects key to b tagging.

We fully simulate WH production ($M_H = 120 \text{ GeV}/c^2$, with $H \rightarrow b\bar{b}$ and $H \rightarrow u\bar{u}$) to provide samples of signal b -quark and background u -quark jets. To suppress the substantial rate of fake tracks in offline reconstruction at high luminosity, we apply requirements recommended for very high luminosity: the tracks must leave hits on at least 9 layers, there must be hits on all the pixel layers, and no more than 2 SCT layers can be missing, while the distance between the track impact parameter and the beam center must be less than 1 mm in the transverse plane and 15 cm in the longitudinal direction.

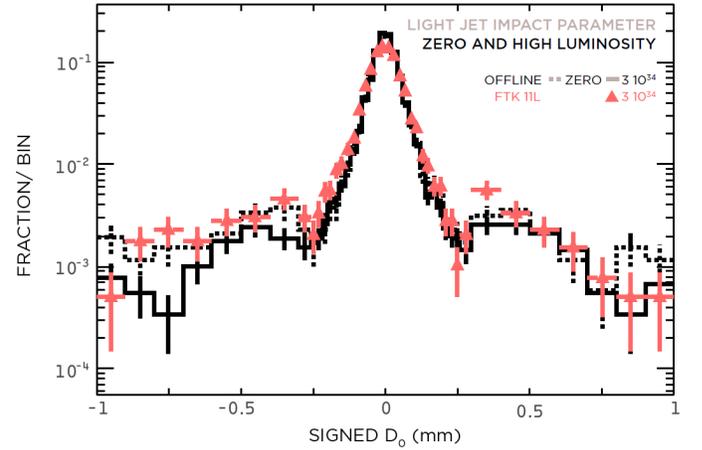


Fig. 6. The signed impact parameter distributions from offline (dark) and FTK (light) for tracks in light-quark jets without pileup (dotted), and at high luminosity (solid lines). The distribution is stable as luminosity increases.

Jet b -tagging starts by identifying the tracks associated with a particular jet. Existing Level-1 and offline calorimeter clustering algorithms will have to be retuned for high luminosity. To be independent of calorimeter performance, we begin by clustering generated tracks into jets of cone size $\Delta R \equiv \sqrt{(\Delta\phi^2 + \Delta\eta^2)} \leq 0.4$. We then associate with the jet all reconstructed tracks of momenta within $\Delta R \leq 0.4$ from the generated jet centroid. Finally, we apply a Δz_0 cut relative to the highest- p_T track to reduce impact of pile-up, and recompute the jet direction as the p_T -weighted average η and ϕ . A jet is labeled a b -quark jet if, within $\Delta R \leq 0.3$ of the jet direction, there is a b quark from the Higgs decay. If a jet is not labeled as a b -quark jet, it is labeled as a light jet.

Fig. 6 shows the distribution of transverse impact parameter significance for tracks in light quark jets without pile-up (dotted lines) and at high luminosity (solid lines). The FTK d_0 resolution, while about 30% larger than that of offline, is still well suited for discriminating prompt tracks from those of b -jet decay (see Fig. 7), and both offline and FTK resolutions are stable up to luminosities of $3 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Fig. 8 shows the number of tracks in light jets with $d_0/\sigma_{d_0} > 3$ at high luminosity for both FTK and offline tracks, demonstrating that the rate of spuriously high impact parameter tracks is roughly the same for both tracking algorithms.

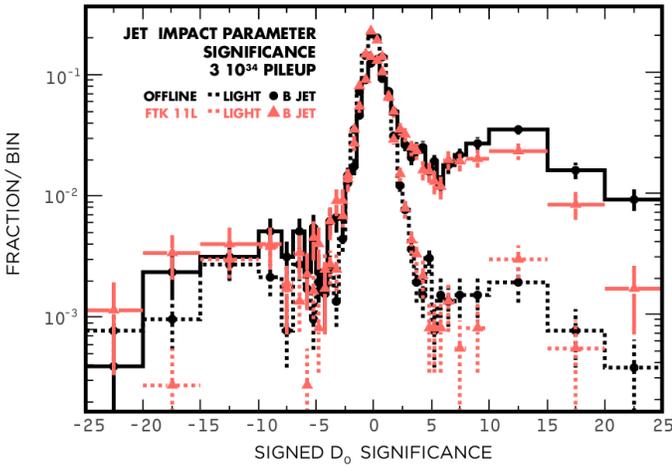


Fig. 7. The signed impact parameter distributions at high luminosity for b jets (solid) and light jets (dotted) for both offline tracking (dark) and FTK (light). The increase at significance between 5 and 15 is due to the variable size bins.

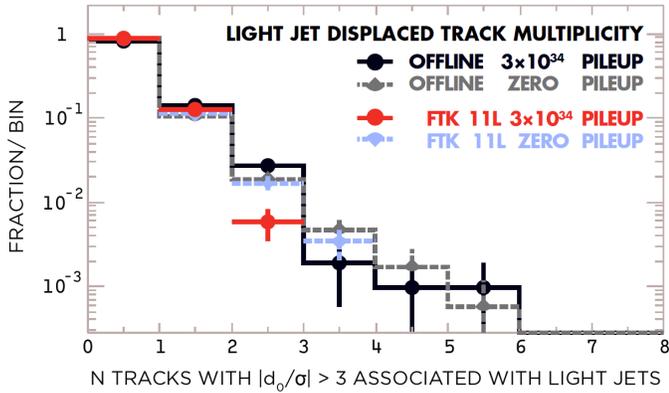


Fig. 8. The number of light-quark tracks with signed impact parameter significance greater than 3 for offline tracking (dark) and FTK (light) without pile-up and at high luminosity.

With similar single track efficiencies, resolutions, and fake rates, we expect b-tagging performance using FTK tracks to be similar to that using offline tracking. To test this, we compare FTK with offline using a tagging algorithm equivalent to the baseline ATLAS Level-2 d_0 likelihood tagger. The algorithm uses likelihood functions determined from the signed d_0 significance distributions of tracks in a sample of b and light jets (labeled as described above). The ratio, L_b/L_u , of the product of the likelihood functions evaluated for each associated track is used as a discriminant between the two types of jets, and one can tune the performance of the algorithm by varying a cut on this ratio.

Fig. 9 shows the light-quark rejection power, the inverse of the probability to tag a light-quark jet as a b jet. It is plotted as a function of the b -tagging efficiency for a sample of WH events independent of the one used to obtain the likelihood functions. The likelihood functions were obtained for each track type (either FTK or offline) and for each luminosity. FTK-based tagging performs well at all tested luminosities.

We are currently studying a secondary vertexing tagger that uses the same underlying constrained vertex fitting code as the

offline secondary vertexing.

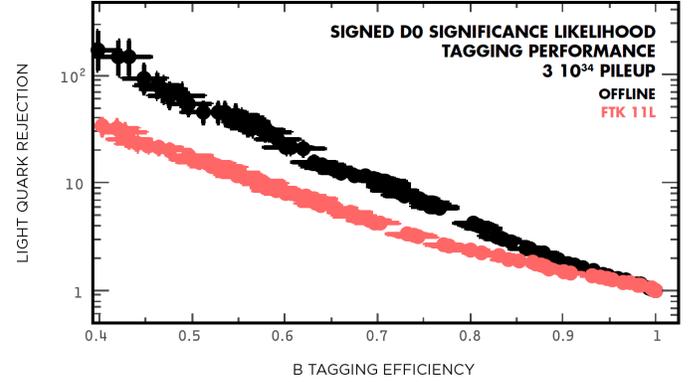


Fig. 9. The light-quark rejection vs. b-jet efficiency for offline tracks (dark) and FTK tracks (light) at high luminosity.

C. τ -identification

The heaviest charged lepton, τ , is often among the decay products in electroweak symmetry breaking scenarios. To maintain good sensitivity to these as well as other new physics processes that produce τ leptons, high trigger efficiency for τ leptons down to the lowest possible τp_T is important. Of necessity, this means good efficiency and high background rejection for hadronic τ decay, i.e. τ jets.

In hadron collider experiments, τ -jet identification has usually been track-based. Signal and the large QCD jet background are separated by the presence of 1 or 3 tracks in a very narrow cone with little or no track activity in a surrounding isolation cone. When tracking is not available in early trigger levels, calorimeter-based selection requires a narrow isolated jet. This is the essence of the ATLAS Level1 τ trigger. Here we propose rapid rejection of the QCD background by using FTK tracks at the beginning of the Level-2 τ selection process. We show that FTK tracking does nearly as well as offline tracking when applied to τ selection.

Tau leptons almost always decay to 1 or 3 charged particles (plus neutrals). The latter presents a greater challenge because of its smaller branching ratio, the size of the QCD background, and the requirement that track reconstruction be efficient for tracks very close to each other. But for some processes, important τ polarization information can be extracted from 3-prong decays.

The τ -tagging algorithm is based on the offline algorithm which utilizes conical η - ϕ regions to search for tracks around the Level-1 narrow τ -cluster candidate and count tracks within signal and isolation regions.

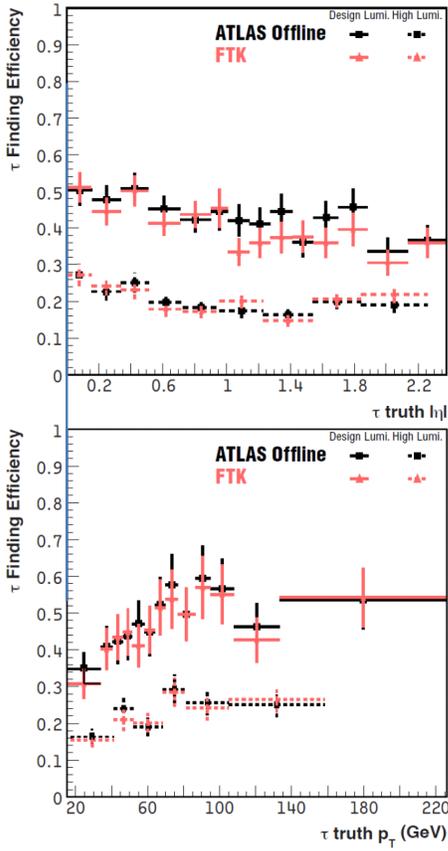


Fig. 10. Single-prong τ efficiency vs. rapidity and transverse momentum, the latter for $|\eta| < 0.8$. FTK tracking (light) is compared with offline (dark) at design and high luminosity.

Tracks with $p_T > 1.5$ GeV/c and within $\Delta R = 0.35$ of the Level-1 τ cluster are considered. Within this cone, the FTK track with the highest p_T is found. If it has $p_T > 6$ GeV/c, a signal cone of $\Delta R = 0.13$ is defined around it. Within the signal cone, there must be exactly 1 track in the case of single-prong τ decays, and 2 or 3 tracks in the case of triple-prong τ decays. An isolation cone of $\Delta R = 0.26$ is defined around the highest- p_T track, and there must be no tracks of $p_T > 1.5$ GeV/c found between the signal and isolation cones. The cone sizes and kinematic ranges are chosen to maximize S/\sqrt{B} , where S is the expected number of τ signal events and B is the expected number of background events from dijets.

Figures 10 and 11 show the τ reconstruction efficiency for 1-prong and 3-prong τ decays respectively. The data sample used is vector-boson-fusion Higgs production at design luminosity, with the Higgs decaying into two τ leptons, each of which decays hadronically. The denominator used to calculate the efficiency is the number of generated τ leptons that are successfully matched to a Level-1 τ cluster. FTK and offline tracking give similar efficiencies for both 1-prong and 3-prong τ reconstruction.

The 1-prong and 3-prong fake probabilities for FTK and the offline are at the level of few per thousand. The fake rate is defined as the ratio of the number of jets passing the τ reconstruction to the total number of reconstructed jets.

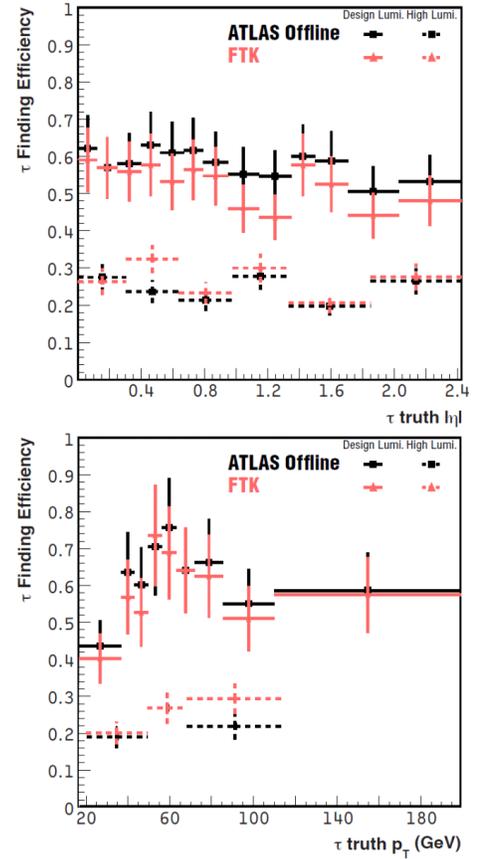


Fig. 11. Three-prong τ efficiency vs. rapidity and transverse momentum, the latter for $|\eta| < 0.8$. FTK tracking (light) is compared with offline (dark) at design and high luminosity.

The τ identification at high luminosity has the same dependence as a function of η and P_T but it is lower than at design luminosity since tracking efficiency drops while pile-up tracks in the isolation cone cause τ decays to be lost. The efficiency is roughly 30% for both 1 and 3-prong decays, for FTK and offline track reconstruction.

D. μ triggering at high luminosity

Traditionally, single lepton triggers rely on calorimeter isolation to suppress backgrounds from real or fake leptons in hadronic jets. This works well at low luminosity. However at high luminosity, when there is substantial calorimeter energy due to the pile-up interactions, the performance of this strategy deteriorates. If the isolation threshold is kept low to maintain background rejection, the efficiency drops for leptons of interest. If the isolation threshold is raised, lepton efficiency can be maintained, but at the price of increased background.

An alternative strategy for high luminosity is to apply isolation based on reconstructed tracks. If all tracks in the event are used, pile-up remains a serious problem. However for tracking, unlike calorimeter deposition, the pile-up and hard-scatter particles can be separated. Here we analyze a track-based isolation using only those FTK tracks that have a z_0 close to that of the lepton candidate.

The selection of isolated muons is critical to the searches for new physics, such as SUSY cascades involving high- p_T muons or a high mass Z' decaying into two muons, as well as

the study of Standard Model processes such as $W \rightarrow \mu\nu$ or $Z \rightarrow \mu\mu$. In the ATLAS HLT, both calorimeter-based and tracking-based isolation are used to reject major backgrounds like muons from semileptonic b-quark decays in QCD events or punch-through to the muon spectrometer by light jets. Since the rates for QCD-produced $b\bar{b}$ pairs and light-quark jets are significantly higher than those for isolated muon processes, it is critical to suppress the background rate in the trigger while maintaining high efficiency for isolated muons.

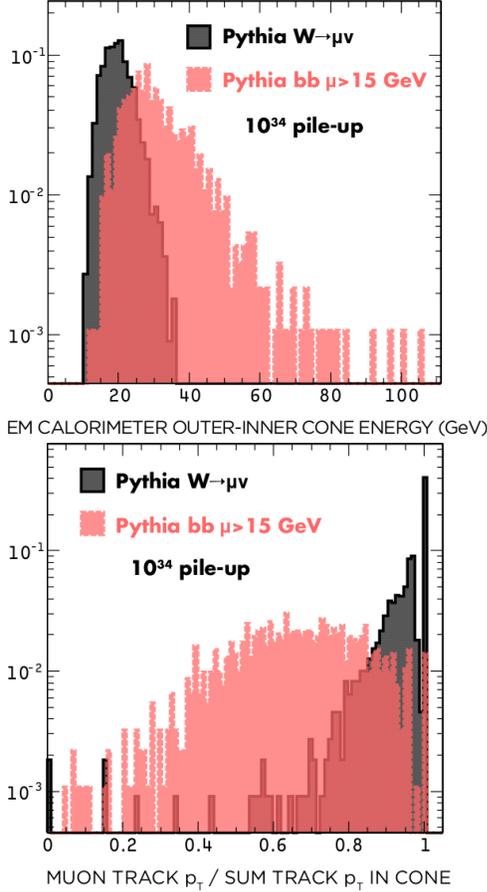


Fig. 12. EM calorimeter isolation (top) and track isolation (bottom) for signal and background at design luminosity.

As previously mentioned, at high luminosity, calorimeter-based isolation loses its effectiveness in selecting real isolated muons. With early tracking information available from FTK, one can calculate isolation using only tracks pointing to the z_0 of the muon track. This tracking-based isolation removes any requirements on the calorimeter cell energies, while maintaining high efficiency for isolated muons in environments with significant pile-up.

In the current implementation of the HLT, the isolated muon trigger requires that the muon candidate pass isolation cuts based on the energy in electromagnetic (EM) and hadronic calorimeter cells and the p_T of inner detector tracks. EM isolation is defined as the sum of the energy in all cells with energy above a 60 MeV threshold within a ring of ΔR between 0.07 and 0.4 around the muon track. For hadronic

isolation, the cell energy threshold is also 60 MeV and the isolation ring extends from ΔR of 0.1 to 0.41. The EM calorimeter isolation energy is shown in Fig. 12 (top) for isolated muons from $W \rightarrow \mu\nu$ and non-isolated muons from $b\bar{b}$ jets at design luminosity. Tracking isolation is defined as the p_T of the muon track, measured by the inner detector, divided by the p_T sum of all inner detector tracks within $\Delta R < 0.2$ of the muon track, including the muon track.

As shown in the Fig. 12, isolated muons tend to have lower energy in the EM isolation ring. However the EM isolation distribution for muons from background events has significant overlap with the isolated signal muon distribution. For the same events, tracking isolation is shown in Fig. 12 (bottom). For isolated muons, this distribution is centered close to one. In contrast to EM calorimeter isolation, the track isolation distribution for $b\bar{b}$ events is better-separated from the isolated muon peak.

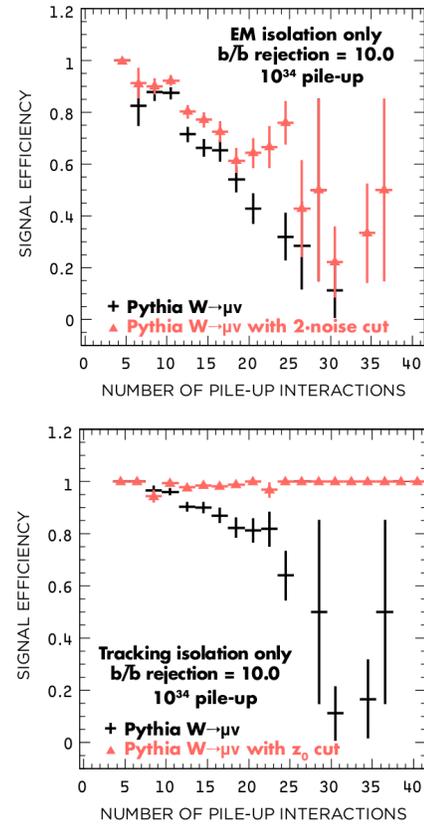


Fig. 13. Isolated muon efficiency using the EM calorimeter with two cell energy thresholds (top) or tracking isolation without (dark) or with (light) a Δz_0 cut as a function of the number of pile-up interactions in the event (bottom). The isolation cut is selected to provide a $b\bar{b}$ rejection factor of 10.

To quantify the trigger efficiency for isolated muons at design luminosity, the calorimeter and tracking isolation cut values are set so that the trigger rejection factor for $b\bar{b}$ events is 10. Fig. 13 (top) shows the isolated muon trigger efficiency as a function of the number of pile-up interactions using only EM calorimeter isolation. The trigger efficiency quickly deteriorates with increasing pile-up. Also shown is the trigger

efficiency when the EM cell energy threshold is increased by a factor of two. The efficiency degradation with increasing pile-up is still clearly visible. Fig. 13 (bottom) shows the isolated muon efficiency when tracking isolation is used instead. Here, inner detector tracks selected using offline reconstruction must have $p_T > 1 \text{ GeV}/c$ and must have at least one hit in either the pixels or SCT. In contrast to calorimeter isolation, tracking isolation is less sensitive to pile-up. The trigger efficiency can be further improved by using only inner detector tracks with z_0 within 10 mm of the muon z_0 .

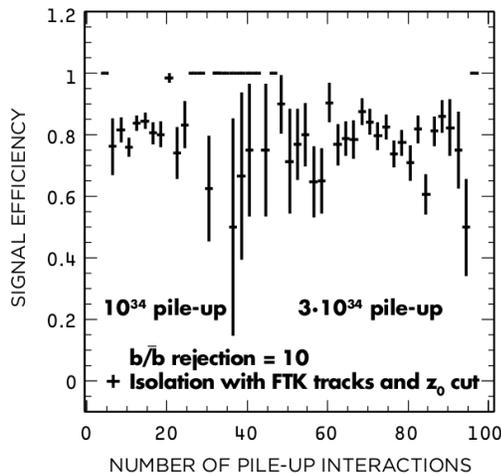


Fig. 14. The $W \rightarrow \mu\nu$ trigger efficiency with FTK track isolation as a function of the number of pile-up interactions.

Using FTK instead of offline tracks to calculate the tracking-based isolation yields similar trigger efficiencies for $W \rightarrow \mu\nu$ events. As shown in Fig. 14, the FTK efficiency as a function of pile-up is constant even for events with 100 pile-up interactions. The plots include both the design and high luminosity samples. The track-based isolation cut was tuned to give a $b\bar{b}$ rejection factor of 10. The overall efficiency for isolated muons is approximately 80% at both luminosities.

To summarize, at high luminosities, the isolated-muon trigger efficiency deteriorates dramatically when calorimeter-based isolation is used. In contrast, tracking-based isolation with fast tracking from FTK is insensitive to the amount of pile-up, maintaining high trigger efficiency for signal events and good rejection of $b\bar{b}$ background.

V. CONCLUSIONS

Events rich in heavy quarks, τ leptons, and isolated muons and electrons are particularly important for testing the limits of the SM and studying its possible extensions. Tracking devices play an essential role in this, in particular the silicon devices that are becoming the predominant tracking technology. We are constructing a very powerful parallel processing system for tracking, and we report its expected performance in simulation up to a luminosity of $3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, with the expected average pile-up of 75 minimum bias interactions.

The simulated FTK has been shown to be very successful in addressing the most difficult technological challenges for online tracking in a high-luminosity environment at the LHC. The Vertical Slice shows the FTK capabilities on a small scale. It will be used to commission the FTK with real data in 2012.

ACKNOWLEDGMENTS

The Fast Tracker project receives support from Istituto Nazionale di Fisica Nucleare, from the US Department of Energy and National Science Foundation, from Grant-in-Aid for Scientific Research by Japan Society for the Promotion of Science, from European community FP7 (Marie Curie OIF Project 254410 - ARTLHCFE), and from Ministero degli Affari Esteri - Direzione Generale per la Promozione e la Cooperazione Culturale (Italy-Japan cooperation program).

REFERENCES

- [1] W. Smith, "Triggering at LHC Experiments", *Nucl. Instr. and Meth. A*, vol. 478, pp. 62–67, 2002.
- [2] A. Annovi et al. "Hadron Collider Triggers with High-Quality Tracking at Very High Event Rates", *IEEE Trans. Nucl. Sci.*, vol. 51, pp.391, 2004.
- [3] G.W. Foster et al., "A fast hardware track finder for the CDF central tracking chamber", *Nucl. Instr. and Meth. A*, vol. 269, pp. 93–100, 1988.
- [4] E.J. Thompson et al., "Online track processor for the CDF upgrade", *IEEE Trans. Nucl. Sci.*, vol. 49, pp. 1063–1070, 2002; A. Abulencia et al., "The CDF II eXtremely Fast Tracker upgrade", *Nucl. Instr. and Meth. A*, vol 581, pp. 482–484, 2007.
- [5] J. Adelman et al., "The Silicon Vertex Trigger upgrade at CDF", *Nucl. Instr. and Meth. A*, vol. 572, pp. 361–364, 2007.
- [6] A. Annovi et al., "A VLSI Processor for Fast Track Finding Based on Content Addressable Memories", *IEEE Trans. Nucl. Sci.*, vol. 53, pp. 2428, 2006.
- [7] A. Annovi et al., "The Fast Tracker Processor for Hadron Collider Triggers", *IEEE Trans. Nucl. Sci.*, vol. 48, pp. 575, 2001.
- [8] A. Annovi et al., "The EDRO board connected to the associative memory: a 'baby' Fast Tracker processor for the ATLAS experiment", Proceedings of TIPP 2011, Chicago, USA, to be published in *Nucl. Instr. and Meth. A*.
- [9] G. Aad et al [ATLAS Collaboration], "Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics", arXiv:0901.0512 [hep-ex], pp. 549, 2008.
- [10] H. C. van der Bij et al, "S-LINK, a data link specification for the LHC era", *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 398-402, 1997.
- [11] M. Dell'Orso, L. Ristori "VLSI Structures for Track Finding", *Nucl. Instr. and Meth. A*, vol. 278, pp. 436, 1989.
- [12] E. Brubaker, et al., "Performance of the Proposed Fast Track Processor for Rare Decays at the ATLAS Experiment", *IEEE Trans. Nucl. Sci.*, vol. 55, pp. 145-150, 2008.
- [13] S. Amerio et al., "GigaFitter: Performance at CDF and perspective for future applications", *Nucl. Instr. and Meth. A*, vol. 623, pp. 540–542, 2010.
- [14] S. Amerio et al., "Associative memory design for the FastTrack Processor (FTK) at ATLAS", *IEEE NSS/MIC Conference Record*, Valencia, Spain, 23–29 October, 2011.
- [15] T. Cornelissen et al., "Concepts, Design and Implementation of the ATLAS New Tracking (NEWT), ATLAS Note, ATL-SOFT-PUB-2007-007.