

OPEN ACCESS

Comparison of two hardware-based hit filtering methods for trackers in high-pileup environments

To cite this article: J. Gradin *et al* 2018 *JINST* **13** P04019

View the [article online](#) for updates and enhancements.

Related content

- [An FPGA based track finder for the L1 trigger of the CMS experiment at the High Luminosity LHC](#)
R. Aggleton, L.E. Ardila-Perez, F.A. Ball et al.
- [A self seeded first level track trigger for ATLAS](#)
A Schöning
- [L1 track triggers for ATLAS in the HL-LHC](#)
E Lipeles

Comparison of two hardware-based hit filtering methods for trackers in high-pileup environments

J. Gradin,^{a,b,1} M. Mårtensson^a and R. Brenner^a

^a*Uppsala Universitet,*

Lägerhyddsvägen 1, 752 37 Uppsala, Sweden

^b*Universite Grenoble Alpes, LPSC,*

53 Avenue des Martyrs 38026 Grenoble cedex, France

E-mail: joakim.gradin@physics.uu.se

ABSTRACT: As experiments in high energy physics aim to measure increasingly rare processes, the experiments continually strive to increase the expected signal yields. In the case of the High Luminosity upgrade of the LHC, the luminosity is raised by increasing the number of simultaneous proton-proton interactions, so-called pile-up. This increases the expected yields of signal and background processes alike. The signal is embedded in a large background of processes that mimic that of signal events. It is therefore imperative for the experiments to develop new triggering methods to effectively distinguish the interesting events from the background.

We present a comparison of two methods for filtering detector hits to be used for triggering on particle tracks: one based on a pattern matching technique using Associative Memory (AM) chips and the other based on the Hough transform. Their efficiency and hit rejection are evaluated for proton-proton collisions with varying amounts of pile-up using a simulation of a generic silicon tracking detector. It is found that, while both methods are feasible options for a track trigger with single muon efficiencies around 98–99%, the AM based pattern matching produces a lower number of hit combinations with respect to the Hough transform whilst keeping more of the true signal hits. We also present the effect on the two methods of increasing the amount of support material in the detector and of introducing inefficiencies by deactivating detector modules. The increased support material has negligible effects on the efficiency for both methods, while dropping 5% (10%) of the available modules decreases the efficiency to about 95% (87%) for both methods, irrespective of the amount of pile-up.

KEYWORDS: Data reduction methods; Online farms and online filtering; Trigger algorithms; Particle tracking detectors

ARXIV EPRINT: [1709.01034](https://arxiv.org/abs/1709.01034)

¹Corresponding author.



Contents

1	Introduction	1
2	Triggering with tracks	2
3	Simulation	3
4	Hough transform	5
5	Pattern matching	8
6	Comparison of the hit filtering	9
6.1	Study of nominal performance	10
6.2	The effect of detector material	14
6.3	The effect of dropping detector modules	14
7	Summary	17

1 Introduction

At the Large Hadron Collider (LHC), experiments such as ATLAS [1] and CMS [2] are analyzing proton-proton interactions to study the nature of matter. The rare phenomena that are searched for are hidden in an enormous background from well known physics interactions. To increase sensitivity to interesting phenomena, the instantaneous luminosity in the LHC will be increased [3] to $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, which is 5 times higher than the design value of the current LHC. The drawback of the higher luminosity is that it will also increase the number of interactions per bunch crossing, referred to as *pile-up*, by an equal amount. This presents new technical challenges for the experiments.

The sheer amount of data generated by the collisions at the High Luminosity upgrade of the LHC (HL-LHC) makes it impossible to read out and store all events. That is why large experiments such as ATLAS and CMS use a *trigger system* to select the most interesting events. It is particularly important to be able to trigger on high transverse momentum (p_T) particles, since they provide a clean signature for a variety of interesting processes in and beyond the Standard Model. However, it is vital to keep the trigger p_T thresholds as low as possible since the acceptance fraction of interesting processes, such as Higgs bosons decaying to tau leptons, decreases rapidly with increasing p_T thresholds.

The trigger in large general-purpose experiments like ATLAS and CMS is usually organized in at least two levels: a hardware trigger based on analogue electronics and logic in Field Programmable Gate Arrays (FPGAs) and Application-Specific Integrated Circuits (ASICs) followed by a software trigger running on a computer farm. The hardware trigger in the two large LHC experiments reduces the event rate from the bunch-crossing rate of 40 MHz to the order of hundreds of kHz. It only has a

few microseconds to make a decision and must be located on or close to the detector. The software trigger uses sophisticated reconstruction algorithms and makes a trigger decision on a timescale of a second. After the software trigger the event rate to offline storage is a few hundred Hz.

In order to maintain low trigger thresholds, high efficiency and low rates, the two large LHC experiments are developing methods to use tracking information in the hardware trigger. This has not been achieved in previous and current large general-purpose experiments because of insufficient bandwidth to read out the tracker and resources to process tracking data fast enough for an early trigger decision.

In this paper we compare two pattern recognition methods that can be used in the hardware trigger for fast hit filtering of tracking data. In section 2 we present the motivation and challenge of using tracking in the trigger. In section 3 we present the simulation framework. In sections 4 and 5 we present the two methods for fast hit filtering: pattern matching using custom made Associative Memory (AM) chips and the Hough transform implemented in FPGAs. Section 6 presents the results from simulation and compares the two methods. A summary is given in section 7.

2 Triggering with tracks

The tracker is the most granular detector system in the collider experiment. It consists of 10–20 layers, located at increasing distance from the interaction point, equipped with silicon detector modules. The layers closest to the interaction point are the most granular and are equipped with pixellated sensors giving two-dimensional information of where highly energetic particles cross the sensor. Sensors with strips are used at larger distance from the interaction point. These are less granular than the pixel sensors and give only one-dimensional information but require less bandwidth to read out than for the pixels. Two-dimensional information can be obtained from the strip layers using two layers of silicon strip sensors that are separated by a small distance and rotated by an angle of 40 mrad with respect to each other. To read out data from the tracker requires high bandwidth. To give a scale, at a 1 MHz readout rate the ATLAS ITk strip detector will generate 465 GB of data per second. In current trackers the services needed to meet the required bandwidth exceed the available space and power budget. Hence, methods to reduce the quantity of data to be read out are needed. This is done by introducing data reduction on-detector as developed by CMS [4] or using readout in Regions of Interest (RoIs) seeded by the muon and calorimeter triggers, as developed by ATLAS [5]. (In ATLAS, an RoI is defined as a region in the track parameter phase-space that spans a maximum of 10 % of the tracker volume.)

In this paper the model for processing of tracker data is done in two main steps. The first is to identify hits from high- p_T particles originating from primary interactions. This is done with coarse resolution in order to save time and computing resources. The filtered hits with full resolution are forwarded to a fitting step where the track parameters are determined. The performance of the hit filtering step is essential to keep the required number of fits at a low level. The hit filtering step must have a high efficiency so that potentially interesting events are not lost.

The two methods for hit filtering studied in this paper are the Hough transform using FPGAs and pattern matching using AM chips. The methods are studied in this paper with software can be implemented in above mentioned hardware. The study is influenced by the development in ATLAS but the performed studies are general and valid for similar problems. We have decided to

limit the study to an eight layer system to make it easier to compare the two methods. A system based on eight select layers in the tracker is a reasonable choice for a hardware based system since limited bandwidth in the detector makes readout of layers close to the interaction point difficult and in general the total amount of data that can be processed is constrained by the cost of computing resources. Eight layers do still offer sufficient redundancy for tracking.

3 Simulation

A generic tracking detector, similar in layout to those of the ATLAS and CMS experiments, is modeled in GEANT4 [6]. The detector has a barrel section, with sensors arranged in cylindrical shells around the beam axis, and endcaps with disks of sensors in the transverse plane. A schematic view of the layout is shown in figure 1, and the amount of tracker material measured in radiation lengths are shown in figure 2. The nominal setup has a radiation length in the central regions similar to that of the ATLAS inner detector after the Phase-II upgrade, as described in [7]. In addition, a setup with 50 % more dense support material was tested. The barrel region contains five layers with pixel sensors and five double layers of strip sensors, with details given in table 1. This table also shows the subset of eight layers used for the patterns and Hough transform. In addition to the sensors, all layers contain material in the form of 3 mm thick carbon supports. The double layers have strip sensors on each side of the stave rotated with a stereo angle of 20 mrad in opposite directions. The stereo angle provides fine resolution in two dimensions for offline data, but at the hardware track trigger stage this information is not made available. The sensors are modeled as rectangular boxes of silicon, 320 μm thick in the radial (beam axis) direction in the barrel (endcaps). The hit positions in the sensors are recorded modulo the readout segmentation, i.e. the local (x, y) coordinates are turned into discrete hit positions of $(\text{row}, \text{column})$. Figure 3 shows the layout and local coordinates of the sensors on the support material. The sensors with one dimensional segmentation, so-called *strip* detectors, divide the local x coordinate in steps of 80 μm . Sensors with two-dimensional segmentation, so-called *pixel* detectors, divide the local x -coordinate by 25 μm and the local y -coordinate by 250 μm .

The simulation uses a right handed coordinate system centered on the interaction point and $\hat{z}$ along the beamline. Tracks are described by helical parameters $\phi_0, \eta_0, z_0, d_0, p_T$ where the subscript refers to that the value is taken at the production vertex. The angle ϕ is the azimuthal angle in the transverse plane, η is the pseudorapidity defined as $\eta = -\ln(\tan(\theta/2))$ where θ is the angle between the track and the z -axis in the $r - z$ plane. The impact parameters z_0 and d_0 are the distances from the center along the beamline or the $r - \phi$ plane respectively.

Single muons were generated directly in GEANT4 using a particle gun, while minimum bias events were generated with PYTHIA8 [8] using the SoftQCD:inelastic setting for proton beams with a 14 TeV center of mass energy. The propagation of the particles through the detector was simulated in GEANT4 with the FTFP_BERT physics list. The muons were generated with impact parameters sampled from uniform distributions of $|z_0| < 150 \text{ mm}$ and $|d_0| < 2 \text{ mm}$. The transverse momentum ranged from 4 to 400 GeV and was sampled from a uniform distribution of $1/p_T$. In addition, the muon η_0 and ϕ_0 was restricted to one RoI, $0.1 \leq \eta_0 \leq 0.3$ and $0.3 \leq \phi_0 \leq 0.5$.

The GEANT4 simulation is followed by a *digitization* step. In this step a single muon event is overlaid with minimum bias events to emulate various degrees of pile-up. We will present our results both by fixing the number of minimum bias events to 200 and by varying the number of minimum bias events from 40 to 260. For evaluating the hit filtering methods, a total of 10,000 unique muon

Table 1. The number of modules in z and ϕ , and the active area on the modules for the barrel layers. The layers with bold numbers are the ones used for the AM patterns and Hough transform. For the strip layers, the two numbers with the same radius corresponds to the stereo layers rotated with respect to each other. The width (SSW) and length (SSL) of the superstrips for the patterns are also listed.

Layer	Radius [mm]	N_z	N_ϕ	length [mm]	width [mm]	SSW [mm]	SSL [mm]	Type
0	35	8	12	40	20	—	—	pixel
1	98	10	22	40	40	—	—	pixel
2	165	10	30	40	40	—	—	pixel
3	233	10	42	40	40	—	—	pixel
4	300	10	52	40	40	1	20	pixel
5,6	400	60	28	24	100	3.2	24	strip
7,8	550	60	38	24	100	3.2	24	strip
9,10	700	30	50	48	100	3.2	48	strip
11,12	850	30	60	48	100	3.2	48	strip
13,14	1000	30	74	48	100	3.2	48	strip

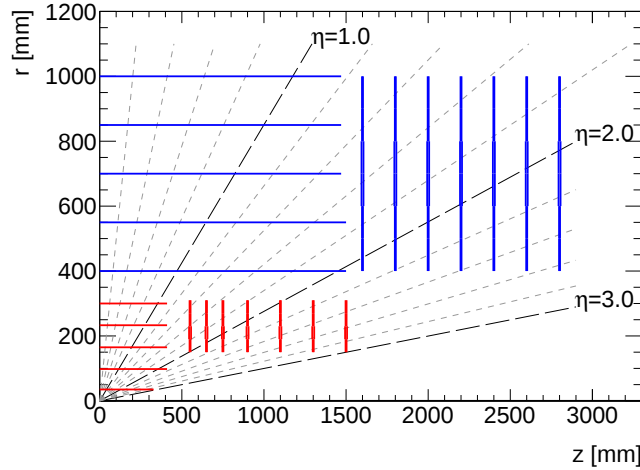


Figure 1. Layout of the detector with layers of pixel sensors in red and double layers of strip sensors in blue.

event were overlayed with minimum bias event drawn at random from a sample with 350,000 events. The reason not to use a larger minimum bias sample is the long computation time needed to generate such events. In the digitization step, muons with primary hits outside the RoI are discarded. The hits are merged by adding together the total energy deposited in each pixel or strip and requiring that the total energy deposited corresponds to more than 1 fC of collected electric charge. It is these merged hits that are referred to as *hits* in the rest of the paper. Hits coming directly from GEANT4-generated muons are labeled as signal and everything else as background. The average hit occupancy for the layers used in the hit filtering methods range from 300 hits per RoI in the pixel layer to 100 hits per RoI in the outer strip layer for single muons overlayed by 200 minimum bias events.

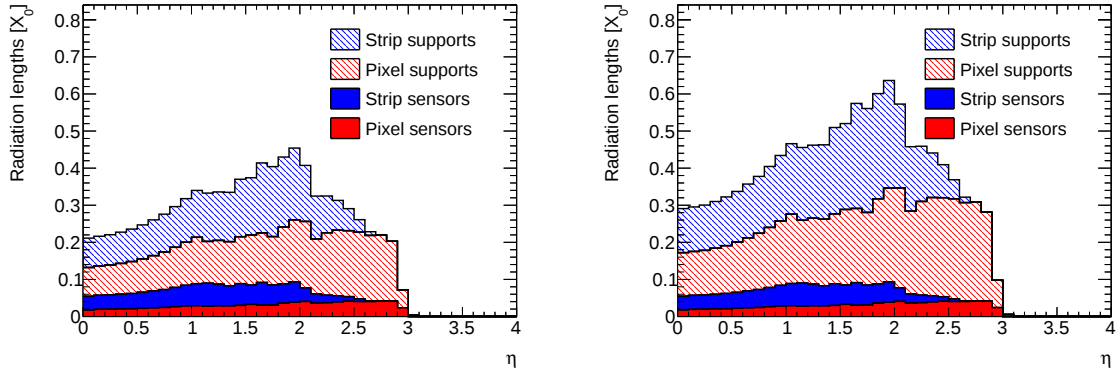


Figure 2. Number of radiation lengths in the tracker material as a function of pseudorapidity in the nominal setup (left) and with more dense support material (right).

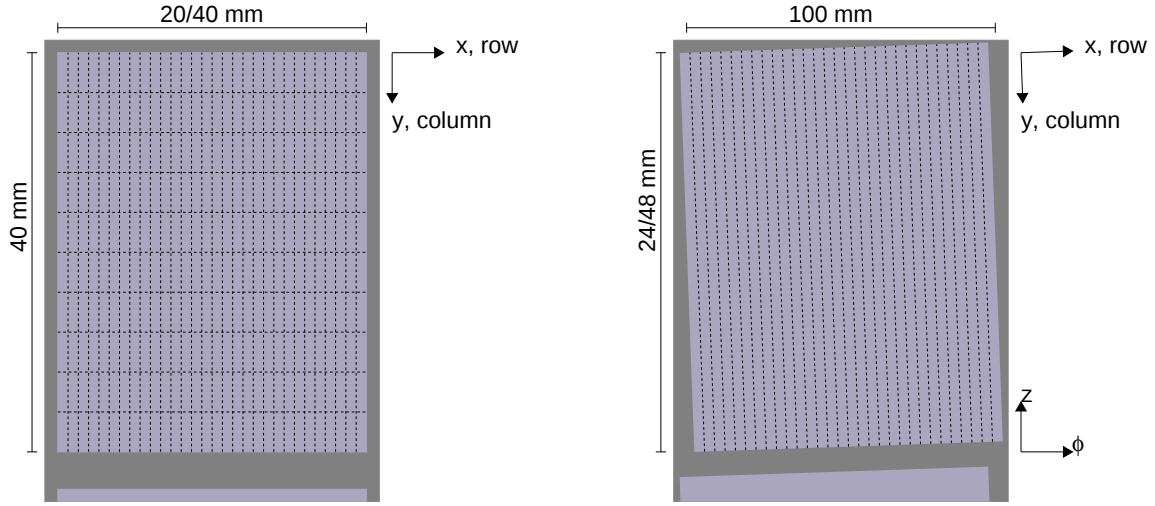


Figure 3. Schematic depiction of the pixel (left) and strip (right) sensors. The dashed lines show the orientation of the individual strips and pixels (not to scale).

The number of connected strips or pixels in ϕ that have a hit, the so-called *cluster width*, is also calculated for each hit. The cluster width is later used in the Hough transform and the AM pattern matching to ignore hits with cluster width larger than 3. The physical motivation for this is that the high- p_T tracks are straighter and produce smaller clusters than the more bent low- p_T tracks from minimum bias.

4 Hough transform

The Hough transform was originally developed for particle physics to detect particle tracks in photographic plates from bubble chambers in the late 1950s [9]. Since then the method has been generalized and is commonly used to detect features that can be represented by a few parameters, such as lines and circles, in a wide range of imaging applications [10]. The Hough transform is actively used in offline tracking, e.g. in ALICE [11], and is gaining interest with recent advancements

in parallel computing; see for example [12]. To be used in triggering, the Hough transform must be implemented in hardware, for instance using FPGAs. In this paper, the Hough transform is simulated using C++, but a very similar implementation has been written in OPENCL and run on an FPGA.

The track of a charged particle in the transverse plane of a uniform magnetic field, e.g. in a tracking detector, is described by a circular arc which can be parameterized using the p_T and azimuthal angle ϕ_0 of the track. The Hough transform calculates all combinations of track parameters consistent with detector hit and casts votes in a histogram-like object called an *accumulator* where one axis gives the track p_T and the other ϕ_0 . The coordinates of the points in the accumulator with the most votes corresponds to a candidate track. If the interaction vertex is constrained to the origin of the coordinate system and the approximation $\sin(x) \approx x$ is made, the track through a point with polar coordinates (r, φ) can be described by

$$A \frac{qB}{p_T} = \frac{\phi_0 - \varphi}{r}, \quad (4.1)$$

where $A \approx 1.5 \times 10^{-4} \text{ GeV c}^{-1} \text{ mm}^{-1} \text{ T}^{-1}$ is the unit conversion constant if r is measured in millimeters, B is the magnetic field in T, p_T is the transverse momentum of the track in GeV c^{-1} , and q is the sign of the electric charge. The approximation $\sin(x) \approx x$ is valid when looking at tracks with low curvature and when using detector layers that are not too close to the interaction point.

The accumulator is implemented as a two-dimensional histogram with $A \frac{qB}{p_T}$ on one axis and ϕ_0 on the other. The accumulator is filled by entering the hit coordinates into equation (4.1) and calculating $A \frac{qB}{p_T}$ for each discrete ϕ_0 value of interest. However, each bin does not simply contain the number of hits. Instead, the bin consists of an 8-bit number to keep track of which of the eight layers have been hit in the small parameter space that it spans. It also has a list of all the hits in the bin. After the accumulator has been filled, bins with less than 6 layers hit are discarded. The surviving bins are track candidates with rough track parameters ϕ_0 and p_T given by the location of the bin in the accumulator and a list of hits that can be sent to the track fitter.

The proton-proton collisions in ATLAS at the HL-LHC are expected to be spread out uniformly over 300 mm along the beam line, i.e. z -axis, making the RoI very wide in z_0 [13]. Since the Hough transform, as implemented here, is looking at a projection in the transverse plane, this makes it difficult to separate tracks with similar ϕ_0 and p_T that originate from different points along the beam line. The two conventional approaches of separating tracks in z_0 using the Hough transform is either to perform a four-dimensional Hough transform in z_0 , η_0 , ϕ_0 , and p_T , or make a second pass of the Hough transform in the r - z plane to find z_0 and η_0 after finding track candidates in the transverse plane. The problem with the first approach is that it requires to sweep ϕ_0 , z_0 , and η_0 which adds two dimensions to the accumulator and increases the number of computations needed. The problem with the second approach is that the Hough transform in the r - z plane would have to be performed after completing the Hough transform in the transverse plane and would increase the execution time proportionally to the number of hits passing the first stage, approximately 30 % for single muons embedded in 200 minimum bias events, since the hits are read out serially. Also, the track finding efficiency is found to be poor due to the low resolution of the strip layers in z .

Our approach to help separating tracks in z_0 is to slice up the RoI in several parts in z_0 as illustrated in figure 4 and fill separate accumulators for each slice by sorting the hits according to their z -coordinate. The RoI slices overlap with each other slightly in z because of the η_0 range of

the RoI and due to the non-zero detector resolution in z , especially for the strips. The z_0 -slicing approach is found to reduce the number of hits passing the Hough transform by 70 % without affecting the track finding efficiency for single muons embedded in 200 minimum bias events.

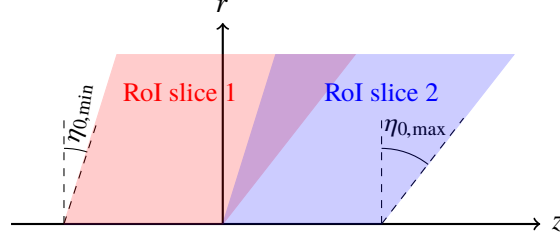


Figure 4. The RoI can be sliced up in z_0 to reduce the occupancy in the Hough accumulator. In this case the RoI is split into two slices.

The number of bins in $A \frac{qB}{p_T}$ and ϕ_0 , as well as the number of slices in z_0 , needs to be tuned to find a configuration that provides a high-enough track finding efficiency while suppressing as much of the low- p_T background as possible. For single muons embedded in minimum-bias event, the optimum number of bins in the accumulator depends strongly on the detector geometry, the detector resolution, and the layer configuration used. The optimal number of z_0 -slices is mostly affected by the pile-up conditions.

The Hough transform can be implemented in hardware using an FPGA. The FPGA receives the hits serially and fills the whole ϕ_0 range and all z_0 -slices in parallel. This does not result in memory write clashes since the mapping of ϕ_0 to $A \frac{qB}{p_T}$ is one-to-one and all z_0 -slices are independent. There are N_z accumulators, one for each z_0 -slice, with N_{p_T} bins in $A \frac{qB}{p_T}$ and N_ϕ in ϕ_0 , each consisting of 8 bits, one for each layer. The total memory requirement for the accumulators is $N_z \times N_{p_T} \times N_\phi \times 8$ bit. When all hits have been processed, bins with at least 6 hits in unique layers are considered track candidates with rough track parameters ϕ_0 , p_T , and z_0 . At this point full-resolution hits need to be associated with the track candidates and sent to the track fitter. This is expected to be done in a separate FPGA, similar to associating full-resolution hits to superstrips in the AM approach. It could also be done in the same FPGA, for instance using a shift register that encodes each hit as a 11-bit number and holds 10 hits per bin. This would require an additional 110 bits per bin.

The available hardware constrains the size of the accumulator used. For the particular geometry and layer configuration used in this study (described in section 3), the accumulator typically has $N_z = 6$, $N_{p_T} = 120$, and $N_\phi = 300$, amounting to a memory requirement of 216 kB. The low latency requirement of a hardware track trigger prevents the use of external memory; hence, the accumulator must be stored internally in the FPGA. Adding a shift register with the associated hits to each bin as discussed in the previous paragraph would increase the memory requirement to 3186 kB, which is still within reach of modern FPGAs. However, the amount of logic resources needed is unclear and can potentially pose a challenge.

The vertex constraint imposed when deriving equation (4.1) is necessary to reduce the computational complexity and the time needed to perform the Hough transform. Without the vertex constraint each hit has to be paired with every other hit and the amount of computation grows as the square of the number of hits in the event. However, imposing a vertex constraint has an important

drawback: it limits the ability to find tracks with large impact parameters, i.e. those which originate far from the beam line, especially if using detector layers located close to the interaction point.

5 Pattern matching

Pattern matching is a commonly employed technique in particle track finding; a review of its use can be found in [14]. A pattern usually consists of a set of hit positions in the detector, which correspond to the track of a high energy particle. The actual hits in the detector can be compared to these patterns, and only the hits that match a pattern are propagated to the track fitting step. The pattern matching step can be made very fast by using a subset of the available detector hit positions, a coarser resolution, or both. ATLAS is currently commissioning a hardware track trigger, the Fast TracKer (FTK) [15], which performs the pattern matching in AM chips [16]. This technology has been used in a wide variety of applications outside of particle physics, such as network switches and artificial neural networks. The AM chips can compare a detector hit to all the stored patterns in parallel, making for a very fast system. This method is considered for a fast hardware track trigger in the HL-LHC era by both ATLAS and CMS. A constraint on the design of the trigger comes from the number of patterns that can be stored on a single AM chip, which also limits the number of layers that can be used. The current estimation for the AM chips under development for the ATLAS upgrade is that they can hold about half a million patterns using eight layers. Here we assume that two such chips can be used to cover the RoI phase space of 0.2×0.2 in $\Delta\eta_0 \times \Delta\phi_0$ based on the cost and power requirements of these chips.

In this study, each pattern consist of a set of hit positions with coarse resolution in a subset of eight silicon detector layers. The coarse resolution hit positions, so-called *superstrips*, are formed by groups of contiguous silicon strip or pixel elements from a single chip. The superstrip width in ϕ is either 40 strips or 40 pixels, and the length in z is the full strip length (24 or 48 mm) or 80 pixels. The patterns therefore have a coarse resolution in both the $r - \phi$ and $r - z$ planes. A unique superstrip index (SSID) is created from the module and superstrip number, and stored in the pattern. The patterns are generated by simulating single muons and recording the hit positions in the selected layers. About 30 million muons are used to build a bank of a few million patterns. A large sample is needed to get good efficiency, but some events produce the same pattern which is why the bank typically contains less than five million patterns. The patterns are ordered by their *use count*, i.e. how many events produce that same pattern. To respect the hardware constraints, only the first one million patterns are used to match hits in a separate sample of muons overlaid with minimum bias events. During the matching all hits in the detector are converted to their superstrip index and compared to each entry in the pattern bank. Each pattern can have multiple detector hits within the superstrips which are propagated to the track fit.

Each pattern use eight separate layers and a match requires seven of these layers to have a superstrip hit. Since a realistic detector can have inefficiencies due to the lack of coverage, e.g. dead space between sensors, the efficiency of the pattern matching is increased by using patterns with *wild cards*. If a muon lacks a hit in any of the layers during the pattern generation, the layer can be marked as a wild card for this pattern. In the pattern matching, a pattern with a wild card always considers that layer to have been hit. With a maximum of two wild cards per pattern it is possible to have a matching pattern with only five real hits in separate layers, which increases the rate of fake matches from

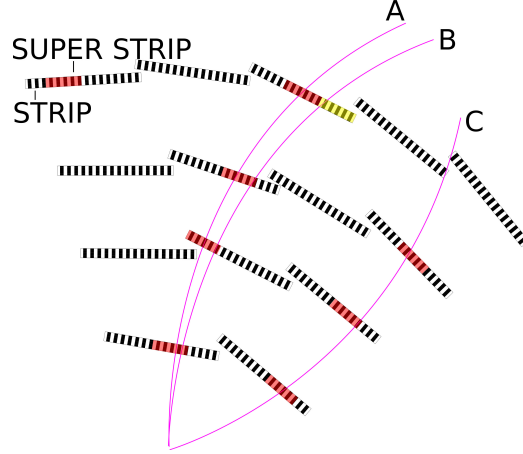


Figure 5. Depiction of the patterns formed for single muon trajectories (purple). Track A and B differ only by one neighboring superstrip (yellow) and can be combined with a DC bit in the outermost layer. Track C is missing a hit and would be assigned a wild card for the last layer.

combinations of background hits and degrades the track fit performance. For these reasons, patterns with two wild cards are required to have hits in all of the six non-wild carded layers to be matched.

The number of patterns in a bank is mainly driven by the size of the superstrips, as the number of possible patterns grow with smaller sizes. Using larger superstrips is not an ideal option, since it increases the amount of fake matches. A compromise can be found by using *don't care* (DC) bits. A DC bit on the SSID in the pattern bank effectively doubles the size of the superstrip, as it disregards the value of the least significant bit in the SSID. Many patterns share the same SSID in several layers and/or have neighboring SSIDs. These patterns can then be combined into one pattern with one or more DC bits. Since a DC bit can be set individually for each pattern and layer, this creates a bank of patterns with variable resolution. This reduces the total number of patterns in the bank but only marginally increases the number of fake matches. The concept of DC bits and wildcards are shown in figure 5.

6 Comparison of the hit filtering

The complexity of the track fitting grows rapidly with the number of hits selected by the preceding hit filtering stage. A one-pass linear fit would have to perform one fit for each hit combination. Therefore, the number of combinations of hits after the hit filtering stage is an important measure of performance. Both the Hough transform and the AM approach group hits together. In the Hough transform the hits are grouped into bins in the accumulator, while in the AM pattern matching, each matched pattern is associated with a group of hits. The number of hit combinations, N_{comb} , is calculated by taking the product of the number of hits in each layer and then summing over the number of groups:

$$N_{\text{comb}} = \sum_g \left(\prod_{l=1}^8 n_{g,l} \right), \quad (6.1)$$

where $n_{g,l}$ is the number of hits in layer l of group g .

The other important measure of performance is the track finding efficiency, which has to be in the order of 99 % for the hit filtering stage. A muon track is considered to be successfully found if at least 6 out of 8 hits from the primary muon are found in unique layers. This definition is motivated by that the subsequent track fit, e.g. the one planned for the ATLAS Phase-II hardware track trigger, is foreseen to fit tracks with 8, 7, and 6 out of 8 hits. The efficiency is defined as the number of events with a muon found divided by the total number of events. One thing to note is that the track finding efficiency of the AM pattern matching is limited by the number of patterns that can fit in the hardware. The Hough transform, on the other hand, the efficiency can be increased by having a coarser-granularity accumulator which makes it less sensitive to scattering. However, a higher efficiency comes with the cost of a larger number of possible hit combinations. For the results presented here, the Hough transform was tuned to provide similar efficiency as the AM method. However, an exact match in the efficiency is difficult to obtain since the Hough transform only can be tuned in discrete steps.

In section 6.1 we present and compare the performance of the hit filtering for AM pattern matching and the Hough transform in terms of the number of hit combinations after the hit selection, the efficiency of finding muon tracks, and how many hits originating from the true muon that pass the hit filters. In section 6.2 we present the effect of adding 50 % more material, and in section 6.3 we present the effect of randomly turning off 5 % or 10 % of the modules.

6.1 Study of nominal performance

Figure 6 shows the number of hit combinations as a function of the number of hits in the RoI when both methods are providing roughly the same track finding efficiency. The number of hit combinations required for the Hough transform is higher than that required for the AM method. Both show the same exponential increase as a function of the number of hits. The distributions of the number of hit combinations are not symmetrical, and have tails that extend to a very high number of combinations, as seen in figure 7.

The efficiency of finding tracks from primary muons is plotted against the number of hits in figure 8 and against the true muon p_T in figure 9. As expected, both methods have a flat efficiency as a function of the number of hits in the RoI. Having high efficiency for leptons with p_T in the 4–20 GeV range is important to maintain a low trigger p_T threshold and performing track-based isolation. As shown in figure 9, the difference between the track finding efficiency at 4–8 GeV and 64–400 GeV is only a fraction of a percent for the AM method. The Hough transform shows a larger difference, approximately 1 %.

It is important to verify that the hit filtering methods are not biased to any particular layer. The inner layers are expected to have more hits since tracks from low- p_T charged particles are bent more by the magnetic field. As expected, the number of hits in each layer decreases with increasing radius, as shown in figure 10. After the AM pattern matching or Hough transform all layers have roughly the same number of hits, as seen in figure 11. This is expected for tracks of high- p_T particles, since they on average produce hits with smaller cluster sizes, and verifies that neither method is biased towards any particular layer.

The number of hits from primary muons sent to the fitter will have an effect on the resolution of the track parameters, with the assumption that the best performance is obtained when all 8 hits can be used. Figure 12 shows the distributions of the number of muon hits found per event. The

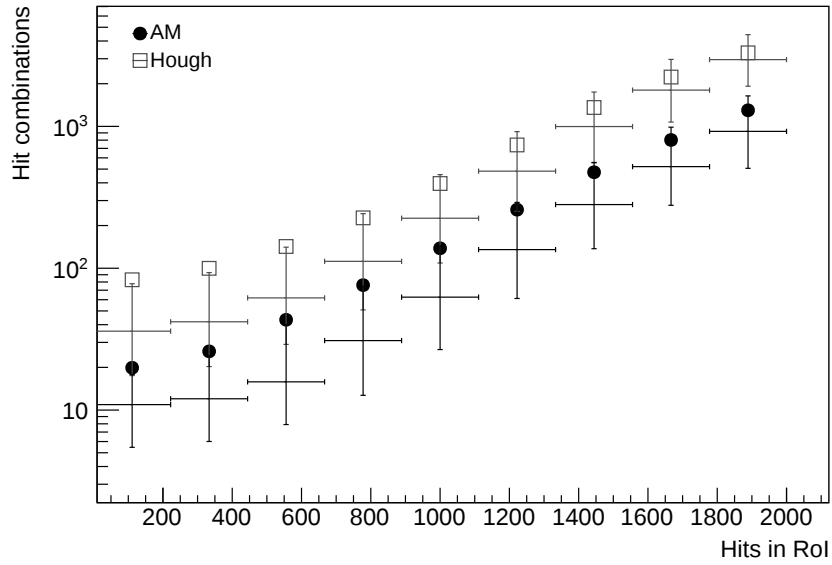


Figure 6. Number of hit combinations after the AM pattern matching and Hough transform as a function of the number of input hits for single muons overlayed with 40 to 260 minimum bias events. The circles and squares mark the average number of combinations for the AM method and the Hough transform respectively. The crosses mark the 50 % (median) with the vertical bars marking the 25 % and 75 % percentiles of the distribution.

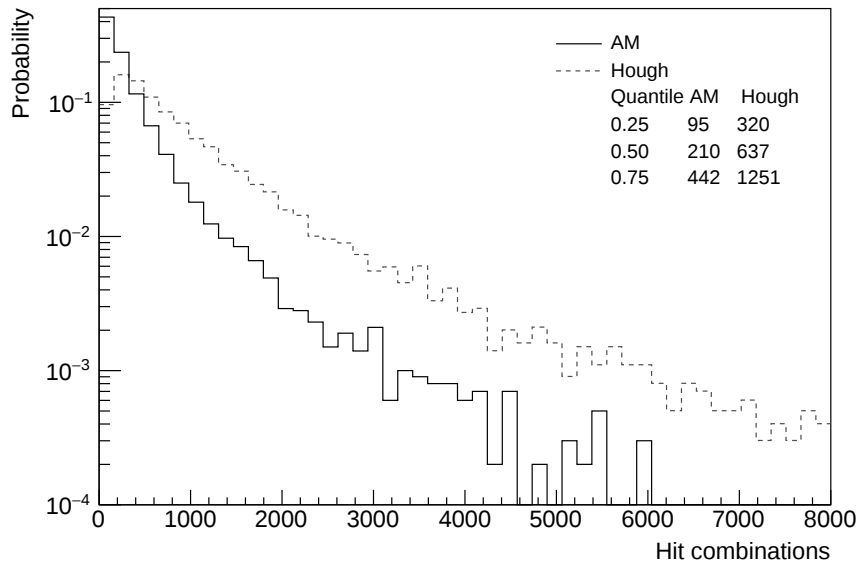


Figure 7. Distribution of the number of hit combinations after the AM pattern matching and Hough transform for single muons embedded in 200 minimum bias events.

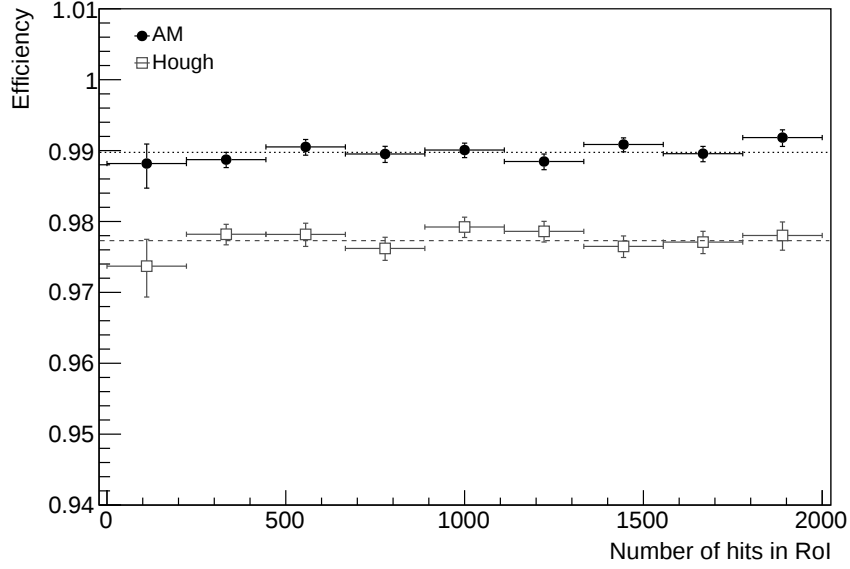


Figure 8. Muon track finding efficiency of the AM pattern matching and Hough transform as a function of the number of hits in the RoI for single muons overlayed with 40 to 260 minimum bias events. The dotted and dashed lines show the average efficiency for the AM method and Hough transform respectively.

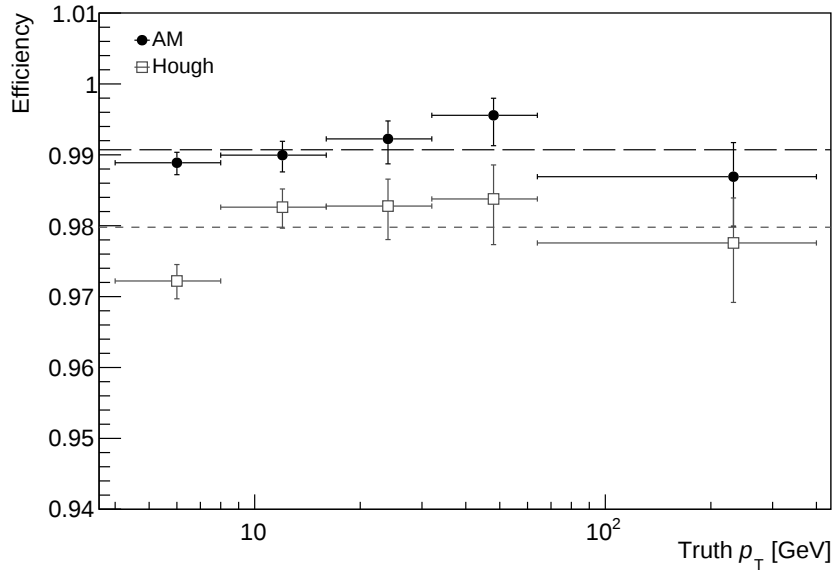


Figure 9. Muon track finding efficiency of the AM pattern matching and the Hough transform as a function of the true muon p_T for single muons embedded in 200 minimum bias events. The long and short dashed lines show the average efficiency for the AM method and Hough transform respectively.

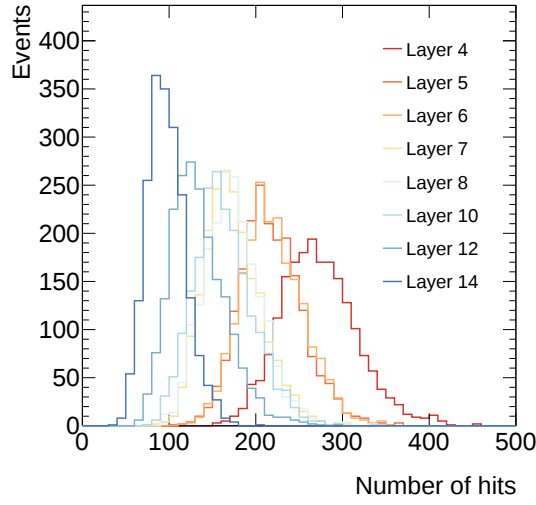


Figure 10. Number of hits in the RoI for single muon events overlaid with 200 minimum bias events, before applying the AM pattern matching or Hough transform.

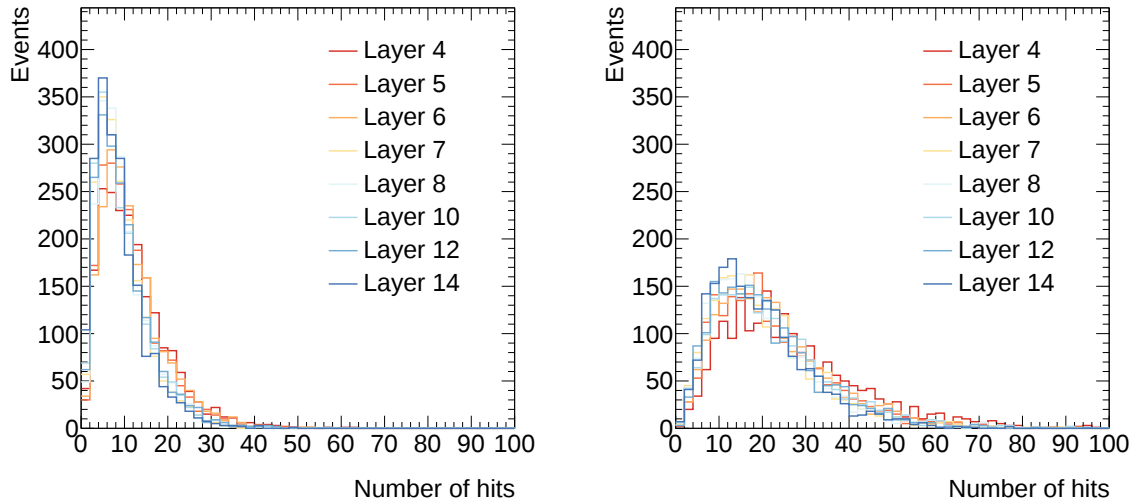


Figure 11. Number of hits in the RoI for single muon events overlaid with 200 minimum bias events, after applying the AM pattern matching (left) or Hough transform (right). Each hit is only counted once even if it is included in several patterns or bins in the Hough transform.

Hough transform and the AM pattern matching find the same number of muon hits in 40 % of the cases. In 55 % of the events, the AM matching finds more muon hits than the Hough transform. The Hough transform finds more hits in only 5 % of the events. This result clearly favors the AM pattern matching for the subsequent track fitting stage.

The results shows that the Hough transform fails to find hits from the primary muon even though it leaves more hits per layer. This can be explained by a combination of the limited detector resolution and the vertex constraint imposed making some hits end up in the wrong bin in the accumulator. In particular, the ϕ and z resolution of the silicon strip detector is very coarse since

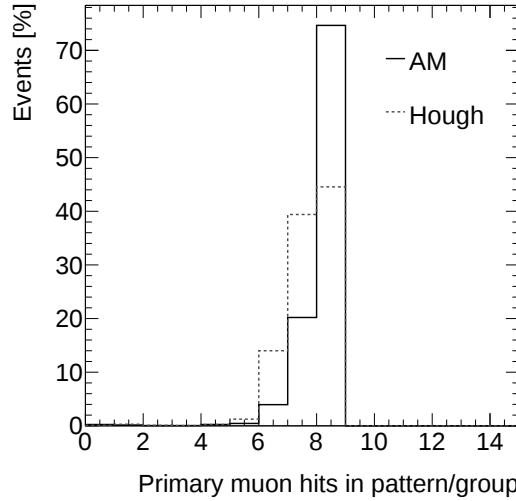


Figure 12. Number of primary muon hits found by the AM pattern matching and the Hough transform for single muon events overlaid with 200 minimum bias events.

the fine resolution given by reconstruction of the crossing strips is not available at this early stage in the trigger chain. This causes some hits to end up in wrong bin in the accumulator. One way to mitigate this could be to implement a more refined algorithm for finding peaks in the accumulator.

6.2 The effect of detector material

The effect on the track finding efficiency of adding 50 % more service material as a function of the number of hits in the RoI is shown in figure 13. The AM methods shows no difference in efficiency while the Hough transform shows an overall decrease on the order of 0.3 %. Looking at the p_T dependence, shown in figure 14, the AM method shows no difference in efficiency. The Hough transform shows a decrease in efficiency of approximately 1.5 % in the 4–8 GeV bin. This drop in efficiency at low p_T is probably due to the muon track deviating too much from a circular arc, making hits end up in the wrong accumulator bins.

6.3 The effect of dropping detector modules

Based on experience with the current trackers at the LHC, a failing module is a more likely cause for loss of efficiency than a general reduction in quantum efficiency of the silicon sensors. In this study, inefficiencies are introduced by randomly removing single detector modules from the simulation. Figure 15 shows the muon track finding efficiency as a function of the number of hits per RoI. Removing 5 % of the modules reduces the efficiency to approximately 93–94 % for the Hough transform and to approximately 95–96 % for the AM pattern matching. Removing 10 % of the modules reduces the efficiency to approximately 87–88 % for both methods, with a slightly higher efficiency observed for the AM pattern matching.

The muon track finding efficiency as a function of the true muon p_T is shown in figure 16 for single muons embedded in 200 minimum bias events. There is no significant p_T dependence seen.

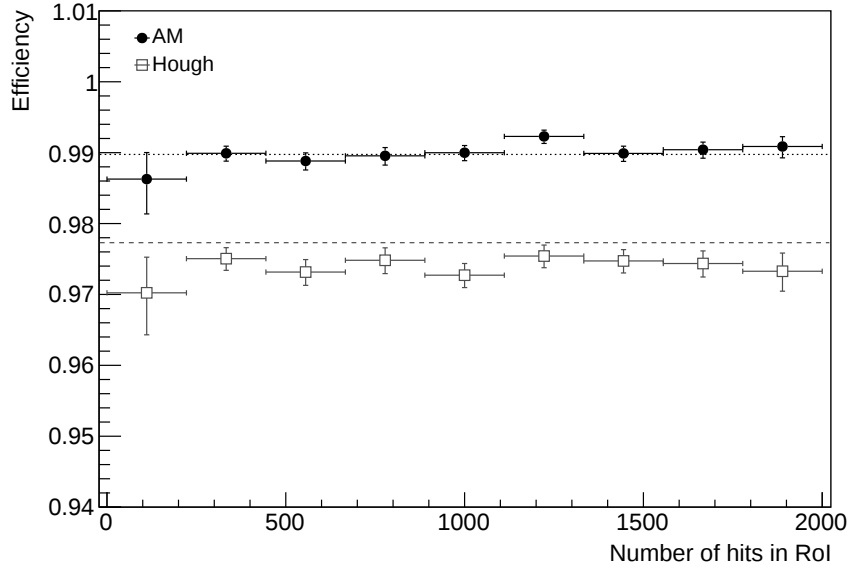


Figure 13. Muon track finding efficiency of the AM pattern matching and Hough transform as a function of the number of hits in the RoI when adding 50 % more material. The sample consists of single muons overlayed with 40 to 260 minimum bias events. For reference, the dotted and dashed lines show the average nominal efficiency for the AM method and Hough transform respectively.

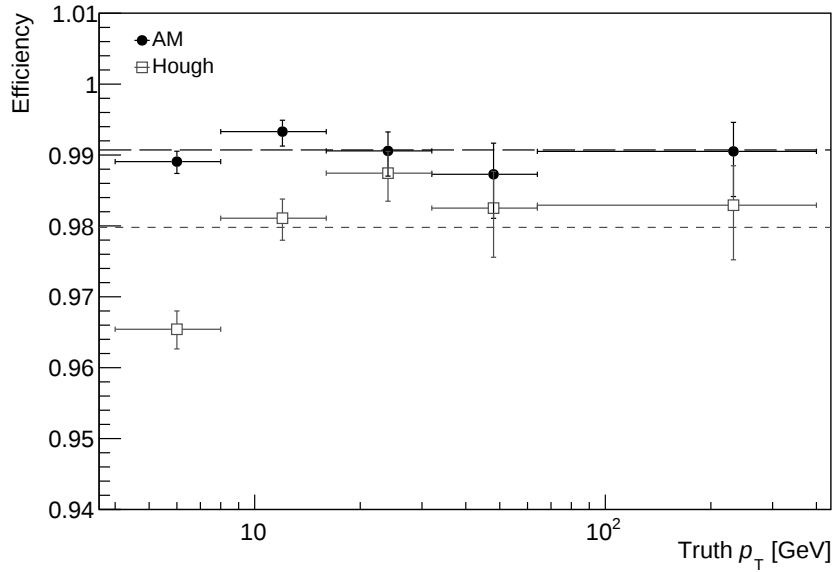


Figure 14. Muon track finding efficiency of the AM pattern matching and the Hough transform as a function of the true muon p_T for single muons embedded in 200 minimum bias events when adding 50 % more material. For reference, the long and short dashed lines show the average nominal efficiency for the AM method and Hough transform respectively.

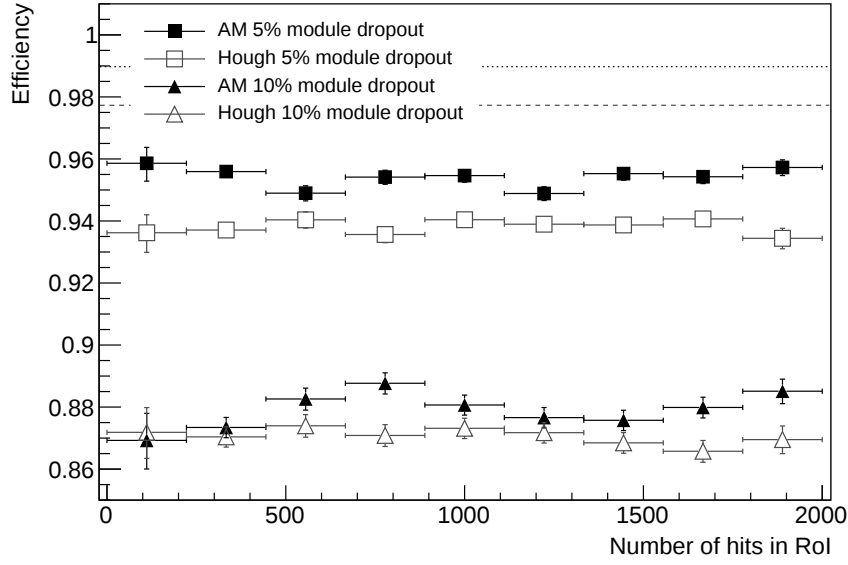


Figure 15. Muon track finding efficiency of the AM pattern matching and Hough transform as a function of the number of hits in the RoI when randomly dropping 5 % and 10 % of the detector modules. The sample consists of single muons overlayed with 40 to 260 minimum bias events. For reference, the dotted and dashed lines show the average nominal efficiency for the AM method and Hough transform respectively.

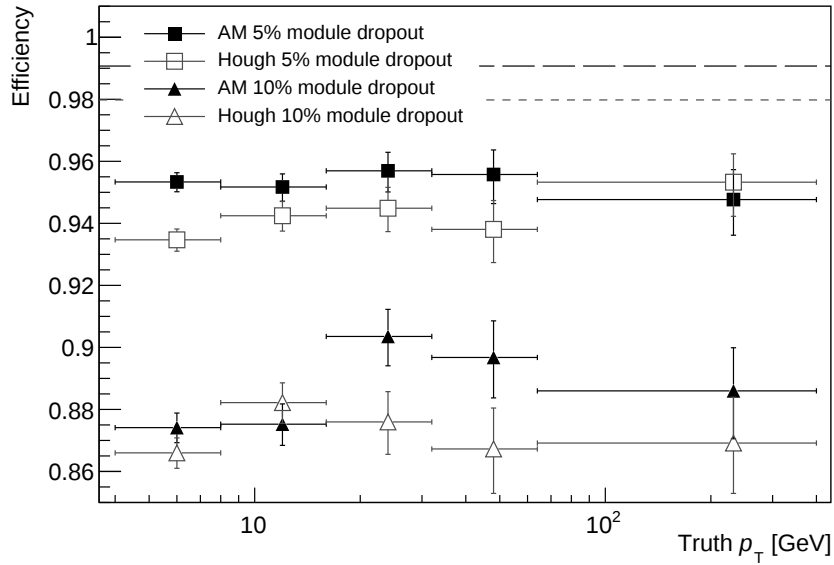


Figure 16. Muon track finding efficiency of the AM pattern matching and the Hough transform as a function of the true muon p_T for single muons embedded in 200 minimum bias events when randomly dropping 5 % and 10 % of the detector modules. For reference, the long and short dashed lines show the average nominal efficiency for the AM method and Hough transform respectively.

It should be noted that both methods could be made resilient for the event of severe detector failure, e.g. if a whole detector layer stops working, by keeping backup pattern banks and uploading them to the AM chips or simply switching the layer parameters in the Hough transform. However, it is more difficult to guard against modules dropping out for single events as studied here. If the beam parameters change or the pile-up conditions differ from the simulations, the Hough transform offers greater flexibility as it can run with an arbitrary number of layers which can be increased if need be, while the AM chips must be designed for a fixed number.

7 Summary

Both methods are able to find at least six of the muon hits in each event with an efficiency at the level of 98–99 %. This efficiency is not directly dependent on the amount of pile-up. Both systems can be tuned to have a nearly full efficiency, e.g. by using very large superstrips or accumulator bins. The optimization of the hit filtering must be done so that the efficiency is kept high while respecting the hardware constraints and rejecting enough hits to keep the latency of the fitter low. The number of hit combinations, as shown in figure 6, grows rapidly and non-linearly with more pile-up for both methods. With modern FPGAs capable of performing a track fit every nanosecond, both methods yield a mean number of hit combinations that could be fitted within a latency of a few microseconds. However, the tails of the distributions are long and both methods will likely produce some events with several thousand hit combinations, which might be truncated by the latency limit. The AM pattern matching shows a consistently lower mean and 75 % percentile of the number of hit combinations over the range of pile-up hits tested here. Thus, the AM option seems to give the best safety margin when it comes to fitting latency.

Another important factor for a track trigger is the quality of the tracks, e.g. the final trigger decision could be based on the χ^2 and the p_T measurement. Since in general the track fit quality will increase with the number of hits used, it is important for the hit filter to send as many of the true track hits as possible to the fitter within the same pattern or group. As shown in figure 12, the AM pattern matching finds more muon hits than the Hough transform and has a larger fraction of 8 out of 8 hits found.

Given these results, the AM pattern matching looks like the best option for a hardware track trigger. However, there are some drawbacks to this method. AM chips with specifications that meet the task are not commercially available and have to be custom-made, which is expensive. The Hough transform, on the other hand, can be implemented in commercially available FPGAs.

References

- [1] ATLAS collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, [2008 JINST 3 S08003](#).
- [2] CMS collaboration, *The CMS experiment at the CERN LHC*, [2008 JINST 3 S08004](#).
- [3] G. Apollinari et al., *High-Luminosity Large Hadron Collider (HL-LHC): preliminary design report*, CERN yellow reports Monographs [CERN-2015-005](#), CERN, Geneva Switzerland, (2016).
- [4] CMS collaboration, F. Ravera, *CMS tracker upgrade for HL-LHC: R&D plans, present status and perspectives*, [Nucl. Instrum. Meth. A 824 \(2016\) 455](#).

- [5] ATLAS collaboration, A. Cerri, *L1Track: a fast level 1 track trigger for the ATLAS high luminosity upgrade*, *Nucl. Instrum. Meth. A* **824** (2016) 263.
- [6] GEANT4 collaboration, S. Agostinelli et al., *GEANT4: a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.
- [7] ATLAS collaboration, *Technical design report for the ATLAS inner tracker strip detector*, [CERN-LHCC-2017-005](#), CERN, Geneva Switzerland, (2017) [ATLAS-TDR-025].
- [8] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [[arXiv:1410.3012](#)].
- [9] P. Hough, *Machine analysis of bubble chamber pictures*, *Proc. Int. Conf. High Energy Accel. Instrum.* **C 590914** (1959) 554.
- [10] R.O. Duda and P.E. Hart, *Use of the Hough transformation to detect lines and curves in pictures*, *Commun. ACM* **15** (1972) 11.
- [11] ALICE collaboration, G. Volpe, *Pattern recognition and PID procedure with the ALICE-HMPID*, *Nucl. Instrum. Meth. A* **766** (2014) 259.
- [12] V. Halyo, P. LeGresley, P. Lujan, V. Karpusenko and A. Vladimirov, *First evaluation of the CPU, GPGPU and MIC architectures for real time particle tracking based on Hough transform at the LHC*, *2014 JINST* **9** P04005 [[arXiv:1310.7556](#)].
- [13] ATLAS collaboration, *Technical design report for the ATLAS inner tracker strip detector*, [CERN-LHCC-2017-005](#), CERN, Geneva Switzerland, (2017).
- [14] R. Mankel, *Pattern recognition and event reconstruction in particle physics experiments*, *Rept. Prog. Phys.* **67** (2004) 553 [[physics/0402039](#)].
- [15] ATLAS collaboration, *Fast TracKer (FTK) technical design report*, [CERN-LHCC-2013-007](#), CERN, Geneva Switzerland, (2013) [ATLAS-TDR-021].
- [16] A. Annovi et al., *AM06: the associative memory chip for the Fast TracKer in the upgraded ATLAS detector*, *2017 JINST* **12** C04013.