Instrumentation of the upgraded ATLAS tracker with a double buffer front-end architecture for track triggering

D. Wardrope\textsuperscript{1,2}

University College London, London, U.K.

E-mail: David.Wardrope@cern.ch

Abstract: The Large Hadron Collider will be upgraded to provide instantaneous luminosity $L = 5 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$, leading to excessive rates from the ATLAS Level-1 trigger. A double buffer front-end architecture for the ATLAS tracker replacement is proposed, that will enable the use of track information in trigger decisions within 20 $\mu$s in order to reduce the high trigger rates. Analysis of ATLAS simulations have found that using track information will enable the use of single lepton triggers with transverse momentum thresholds of $p_T \sim 25$ GeV, which will be of great benefit to the future physics programme of ATLAS.

Keywords: Trigger concepts and systems (hardware and software); Front-end electronics for detector readout

\textsuperscript{1}Corresponding author.
\textsuperscript{2}On behalf of the ATLAS collaboration.
1 Introduction

The High Luminosity LHC (HL-LHC) upgrades to the Large Hadron Collider in the early 2020s will increase the instantaneous luminosity to \( L = 5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1} \), allowing ATLAS and CMS to collect large datasets corresponding to an integrated luminosity of order 3 ab\(^{-1}\) per experiment. This will enable high precision measurements — for example of Higgs boson couplings or WW scattering — as well as searches for extremely rare processes such as direct slepton production.

It is impossible to specify the physics programme of ATLAS ten years in advance, particularly so early in the LHC data-taking era. However, most putative programmes (e.g. [1]) require that ATLAS can trigger efficiently on single leptons with \( p_T \sim 25 \text{ GeV} \). This will be challenging, since the increased luminosity will increase event rates by a factor of five beyond the ATLAS design rates. Furthermore, the increase in instantaneous luminosity will lead to an increase in the number of proton-proton (pp) collisions per bunch crossing to \( \langle \mu \rangle \sim 120 \). This pile-up of pp collisions poses difficulties, particularly for the Level-1 trigger (L1) which is the first step in trigger and cannot use tracking in the current system.

So far, the LHC has delivered luminosity in excess of expectations and so it is prudent to plan for higher than expected luminosity and pile-up. For this reason, ATLAS is planning for HL-LHC conditions of \( L = 7 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1} \) or \( \langle \mu \rangle = 150 \).
Figure 1. A schematic of the current ATLAS Level-1 Trigger. The L1 decision is taken by the Central Trigger Processor based on calorimeter and muon trigger candidates. The Level-1 accept signal (L1A) is distributed to the detector front-ends through the Timing, Trigger and Control system. Geometrical Regions of Interest centred on the Level-1 candidates are passed to the Level-2 trigger for further examination.

2 ATLAS Level-1 trigger system

2.1 The current system

The ATLAS trigger system [2] consists of: a hardware-based Level-1 (L1) trigger that reduces from the bunch-crossing rate of 40 MHz to \( \sim 75 \) kHz; a software-based Level-2 trigger that examines in detail Regions of Interest defined by Level-1 and further reduces the rate to \( \sim 3 \) kHz; and the Event Filter that makes the final decision to record the event, with an output rate of \( \sim 300 \) Hz.

The Level-1 system (figure 1) is implemented in hardware since it must make decisions within a latency of \( < 2.5 \mu s \). This limit is imposed by the length of the on-detector pipeline memories. The combination of the short latency budget and high input rate precludes the use of tracking at Level-1 and limits the granularity and resolution of the calorimeter and muon data used.

The coarse data from the calorimeter and muon systems are used to identify objects such as muons, electrons and photons which satisfy programmable \( p_T \) thresholds. These objects are passed to the Central Trigger Processor which will accept the event if it meets the requirements specified in the trigger menu. These requirements are based solely on the multiplicity of the trigger objects.

2.2 Performance of the current system at higher luminosities

The Level-1 trigger is operating well in current data-taking [3], but rates become excessive at \( L = 5 \times 10^{34} \text{ cm}^{-2}\text{s} \) (a rate of \( O(20\text{kHz}) \) is deemed the maximum allowed for a single lepton trigger in this work). For example, figure 2a shows the L1A single EM rate for \( \langle \mu \rangle = 70 \) which corresponds to \( L = 3 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1} \). The rates can be moderated by increasing the \( E_T \) threshold, however for inclusive EM clusters the rate is above 20 kHz until \( E_T \geq 50 \text{GeV} \). Such a high threshold implies a greatly reduced acceptance for many physics processes, e.g. \( W \rightarrow e\nu \) as shown...
in figure 2b. The acceptance loss in $W \rightarrow l \nu$ is indicative, since much of the potential HL-LHC physics programme involves leptons at a similar $p_T$ scale.

Isolation criteria can be applied to EM clusters to reduce the L1A rate — figure 2a shows a rate of 20 kHz can be obtained for $E_T > 28$ GeV. However, the tight isolation demanded to obtain this rate is inefficient at $\langle \mu \rangle = 70$ and the even tighter isolation necessary to maintain this threshold at $\langle \mu \rangle = 150$ would be even worse - if indeed it were possible to control the rate to such a degree.

For single muon triggers, the L1A rates are lower than for electrons by virtue of the smaller background. However, they are still excessive at high luminosity — approximately 40 kHz for a threshold of $p_T > 20$ GeV at $\langle \mu \rangle = 150$, even with the foreseen muon trigger upgrades. The rate cannot be controlled by raising thresholds, since the $p_T$ resolution of the muon trigger chambers worsens with increasing $p_T$. Furthermore, calorimeter isolation cannot be used in conjunction with the muon trigger in the current L1.

2.3 Proposed upgrades to ATLAS

A programme of upgrades to ATLAS will be installed during two of the planned LHC long shutdowns. These will enable successful operation at $\langle \mu \rangle = 150$. Included in the upgrades for the 2022 shutdown is the complete replacement of the inner detector and almost all of the on-detector front-end electronics. These upgrades provide an opportunity to install a Level-1 track trigger, since the new inner detector can be designed with L1 trigger functionality and the new front-end electronics can have increased pipeline buffers and readout bandwidth.

The only front-ends not to be upgraded are a small fraction of the muon drift tube chambers which are inaccessible. At $\langle \mu \rangle = 150$ these limit the L1A rate to $\sim 200$ kHz and latency to $\sim 20$ $\mu$s.

2.4 Benefits of tracking information at Level-1

Tracking could greatly reduce L1A rates by improving the identification of objects. It could: ensure objects come from the primary vertex; measure pile-up robust, track-based isolation; and
make complementary measurements. For example, most EM trigger accepts are not due to signal electrons or photons but hadronic background. The electron trigger fake rate could be reduced by demanding that the energy measured by the EM cluster is consistent with the momentum measurement of a geometrically matched track, as shown in figure 3a. Simulations show this alone could reduce the L1A rate for triggering on single electrons by a factor of ten (figure 3b).

Improving the $p_T$ measurement of particles using the inner detector improves the application of $p_T$ thresholds. This is particularly beneficial for single muon triggers, where rate reductions of a factor three have been demonstrated in simulation.

Tracking also improves multiple object triggers by ensuring that candidates come from a common primary vertex.

3 Track trigger using a double buffer front-end architecture

Tracking at Level-1 would be beneficial, but it is unrealistic to read out the entire inner detector at 40 MHz, since the high output bandwidth would require many data links, a great deal of power and cooling and thus a great deal of material in the tracker. A trigger with two hardware-based levels in place of Level-1, the “Level-0 + Level-1” scheme can reduce the bandwidth needed.

In this scheme, Level-0 (L0) is the equivalent of the current Level-1 — coarse muon and calorimeter data are used to reduce the rate from 40 MHz to 500 kHz, within a latency $< 6.4 \mu$s. The Level-0 will also define geometrical Regions of Interest based on the Level-0 trigger candidates. In the new Level-1, the L1 Track system only uses data from the inner detector elements within the RoI, further reducing the bandwidth needed. The final Level-1 decision is taken with 20 $\mu$s, on the basis of the tracks reconstructed combined with refined calorimeter and muon candidates.

A benefit of this scheme is it does not require alterations to the tracker layout, which can be optimised for offline reconstruction. To illustrate the bandwidth reduction of this scheme, consider...
Figure 4. A schematic of the proposed two buffer scheme, showing how L0A is broadcast to all front-end chips, resulting in data being moved from the first pipeline into the second buffer. The L0 also defines geometrical Regions of Interest that are mapped to particular front-end chips. For L1 Tracking, data are requested only from the specified front-end elements.

the following parameters: rate of L0A, $r_{L0A} = 500$ kHz; rate of L1A $r_{L1A} = 100$ kHz; and the fraction of tracker data contained in an RoI, $f_{roi} = 10\%$. Then the total bandwidth is only $1.5 \times$ the bandwidth without the track trigger:

$$\text{Bandwidth} = (r_{L1A} + r_{L0A} \times f_{roi}) \times \text{event size} = 150\text{kHz} \times \text{event size} \quad (3.1)$$

3.1 Possible implementation of Level-0 + Level-1 scheme

A potential design of the L0+L1 trigger is shown in figure 4. The design of the ABC130 front-end chip for the new silicon tracker features a double buffer architecture as illustrated.

3.2 Front-end readout architecture

The Level-0 will operate synchronously with the bunch crossings. While it makes a decision, inner detector data will be stored in pipeline buffers on the ABC130s. These have 256 cells in the current design — corresponding to an L0 latency of $6.4\ \mu s$. When L0 accepts an event, an L0A signal is broadcast to all front-ends, triggering the transfer of the data into the second buffer. Since Level-1 is not bunch-crossing synchronous, the data in the second buffer must be tagged with an L0ID so the appropriate data can be read out for each event. With a L0A rate of 500 kHz, the latency of 20 $\mu s$ requires only ten cells in the second buffer, although the ABC130 is designed with 256.
Figure 5. A schematic of the front-end readout configuration, showing data flow from the ten ABC130s on the hybrid through to the L1 Track Finder.

After the L0A, RoIs are mapped to the corresponding ABC130s and Regional Readout Requests (R3s) targeting only those chips are issued. These trigger readout from the second buffer into the L1 Track Finder. An R3 is a self-describing sequence: a 3 bit header + 12 bit module ID + 12 bit L0ID + 1 bit trailer. This allows the R3 to get the data for the correct event from the correct chips. The L1A signal need contain only an L0ID since it is broadcast to all front-end elements.

3.3 Data flow

The planned silicon strip barrel detector consists of staves, each holding twelve strip modules which are readout by two hybrids. On each hybrid, there are two banks of five daisy-chained ABC130s which communicate with a Hybrid Chip Controller (HCC) (figure 5). Fixed length packets of data (59 bits) are generated by the ABC130 chips and passed via their neighbours at 160 Mb/s to the HCC. Due to the limited latency for L1 track finding, only one packet is transmitted per ABC130 in response to an R3, carrying the position of the first four hits registered by the chip. Data are transmitted at 160 Mb/s from the HCC along stave links to a gigabit transceiver and then on to a Readout Driver (ROD) via a virtual private link. The ROD then passes the data (with low latency) either to the DAQ system or the L1 Track Finder. Per-link bandwidth is the same along the entire HCC to ROD chain, so queueing only occurs on the hybrids.

3.4 Results on readout configuration

Discrete event simulations have been carried out to ascertain the optimal front-end readout configuration and whether this configuration is capable of reading out the R3 data within the 6 $\mu$s allocated from the total 20 $\mu$s latency budget. The parameters investigated were the bandwidth of links within the system; whether to have dedicated links for R3 and L1A data in order to prioritise R3 data or to share links; and whether to have separate buffers on the HCC for R3 and L1A data.

The results depicted in figure 6 were obtained for $\langle \mu \rangle = 200$; an L0A rate of 300 kHz; an L1 rate of 75 kHz; and an R3 rate of 3 kHz (an f roi of 1%). They show that a 160 Mb/s link shared between R3 and L1 data is the optimum, with an average latency 0.8 $\mu$s shorter than if the link were split into separate 80 Mb/s links. There are more outliers in the shared case, but 98.5% of R3 data is received at the end-of-stave within 5.5 $\mu$s.

As there are two banks of five ABC130s each linked at 160 Mb/s into the HCC, but only 160 Mb/s output bandwidth, the buffers on the HCCs will fill faster than they drain. It was found that it was beneficial to have separate buffers on the HCC for R3 and L1 data, since this allows the R3 data to be prioritised for transmission.
Figure 6. Distribution of time taken for R3 data to be transferred from the FE ABC chip to the End of Stave.

The bandwidth requirements were estimated for the expected parameters: $\langle \mu \rangle = 140$; an L0A rate of 500 kHz; an L1 rate of 200 kHz; and an R3 rate of 50 kHz (an $f_{\text{roi}}$ of 10%). The bandwidth required was found to be 125 Mb/s, within the envelope of the 160 Mb/s links under consideration.

4 Conclusions

Track Triggering at Level-1 can enable the ATLAS trigger system to cope with the challenges of HL-LHC and meet the likely physics goals of the experiment. A double-buffer front end architecture has been proposed to enable the use of “Level-0 + Level-1”. This reduces the bandwidth required for a Level-1 Track Trigger by both reducing the L1 input rate and by only using data from Regions of Interest defined by the Level-0.

The hardware for such an architecture is under development. Submission of the ABC130 and HCC for an engineering run is planned in Spring 2013. Physics studies are underway to determine the requirements for Level-1 track measurement quality, efficiency and fake rates.

References

[1] F. Gianotti et al., Physics potential and experimental challenges of the LHC luminosity upgrade, Eur. Phys. J. C 39 (2005) 293 [hep-ph/0204087].

[2] ATLAS collaboration, G. Aad et al., The ATLAS experiment at the CERN Large Hadron Collider, 2008 JINST 3 S08003.

[3] ATLAS collaboration, Performance of the ATLAS trigger system, ATL-DAQ-PROC-2012-004, CERN, Geneva Switzerland (2012).