PROCEEDINGS OF SCIENCE

Upgrades of the ATLAS trigger system

Francesca Pastore, on behalf of ATLAS Collaboration*
Royal Holloway, University of London
E-mail: francesca.pastore@cern.ch

In coming years the LHC is expected to undergo upgrades to increase both the energy of proton-proton collisions and the instantaneous luminosity. In order to cope with these more challenging LHC conditions, upgrades of the ATLAS trigger system will be required. This talk will focus on some of the key aspects of these upgrades. Firstly, the upgrade period between 2019-2021 will see an increase in instantaneous luminosity to $3 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$. Upgrades to the Level 1 trigger system during this time will include improvements for both the muon and calorimeter triggers. These include the upgrade of the first-level Endcap Muon trigger, the calorimeter trigger electronics and the addition of new calorimeter feature extractor hardware, such as the Global Feature Extractor (gFEX). An overview will be given on the design and development status of the aforementioned systems, along with the latest testing and validation results.
By 2026, the High Luminosity LHC will be able to deliver 14 TeV collisions with an order of magnitude larger instantaneous luminosity, expected to reach $7.5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$. ATLAS is planning a series of upgrades to prepare for this even more challenging environment. This paper will describe the baseline architecture for this upgrade, while also detailing on-going studies into new system components and their interconnections. The overall challenge here is to meet low latency and high data throughput requirements within the limits given by technological evolution. A discussion on the physics motivations and the expected performance based on simulation studies will be presented, together with the open issues and plans.

EPS-HEP 2017, European Physical Society conference on High Energy Physics
5-12 July 2017
Venice, Italy

*Speaker.

PoS(EPS-HEP2017)510

1. Introduction

The LHC [1] upgrade plans as shown in [2] will support the physics goals of the experiments with unprecedented amounts of collision data at the high-energy frontier. Starting from 2020, the instantaneous luminosity will go beyond the design specifications, growing first to $3 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ in the so-called LHC Run 3, and then to the ultimate target of $7.5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ during Run 4 which will start by 2026 (High-Luminosity LHC, HL-LHC). Being designed to explore the TeV energy scale and after discovering the Higgs boson, the ATLAS experiment [3] focuses its future physics programme on detailed measurements of the Higgs properties, allowing the study of the electroweak symmetry breaking as shown in Fig.1 (left), and on precision measurements of rare signatures and searches for new heavy states, probing different scenarios of New Physics beyond Standard Model.

To accomplish this broad physics programme, ATLAS is planning two upgrade phases that will allow to take advantage of the increased luminosity, with the target of collecting 3000 fb^{-1}. The upgrades will try to improve or at least maintain the performance of present detectors in the new harsh environment as well as to increase the trigger selectivity and the online processing capabilities with upgrades in the T/DAQ system.

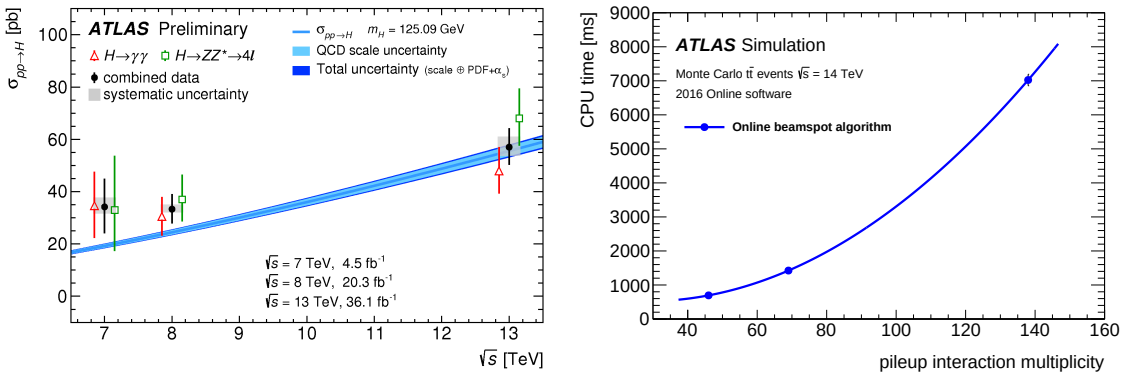

Figure 1: Left: Total pp→ $H + X$ cross section as function of sqrt(s) from H→ $\gamma\gamma$ and H→ ZZ→ $4ll$ [4]; Right: The time of online track reconstruction for a trigger algorithm that identifies the interaction region (beamspot), for 14 TeV $t\bar{t}$ Monte Carlo simulated with 46, 69 and 138 interactions per bunch crossing (pileup), measured on a 2.4 GHz Intel Xeon CPU with 2016 software release [5].

The current two-level T/DAQ system achieves approximately 10^4 rejection from the initial 40 MHz rate of LHC proton bunch-crossings, recording around 1 kHz of data for offline analysis. The first-level trigger (L1) is built from custom electronics, using reduced granularity calorimeter readout and a subset of the muon detectors, with a 2.5 μs fixed latency and a maximum of 100 kHz accept rate. Accepted events are read out and held by the DAQ system during High-Level Trigger (HLT) processing. The HLT selection is realised in software running on a commodity PC farm, through fast trigger algorithms with some code in common with the offline reconstruction. Regional readout and reconstruction using data in detector regions identified by the L1 trigger (Region-of-Interest, RoI) reduce processing time significantly and the readout network size by a factor of 10.

The current trigger strategy is based mainly on the identification of high momentum particles such as the decay products of the W and Z bosons, with 60% of the trigger bandwidth devoted to single and double leptons signatures. The trigger menu includes more than a thousand selections, mostly inclusive for a broad physics programme plus their supporting triggers. Selection algorithms are continuously monitored and optimised to maintain stable performance in spite of changes to luminosity, LHC and detector conditions, and to control rates and processing times.

The large increase in luminosity planned for the HL-LHC gives rise to many challenges for the T/DAQ system. The mean number of collisions per bunch-crossing (pileup) will increase from the current 40 to 80 in Run 3 and to 200 in Run 4. All trigger rates will increase correspondingly with the pileup level, some linearly and others more rapidly due to the changes of event shape. Ultimately more than ten thousand particles will overlap every 25 ns, compromising pattern recognition and detector resolution, which reduce the rejection power from algorithms. Maintaining the current low energy thresholds needed to preserve physics acceptance, for example 20-30 GeV for single muons/electrons, will cause the L1 accept rate to rise from 0.1 to 1-4 MHz. As a consequence of this and of the increased event size due to pileup, the DAQ system must handle an increased throughput from 1 to 50 Tbps. These issues motivate a major upgrade of the T/DAQ system that goes beyond a simple linear scaling up with luminosity of the current design, since event complexity is dominant: for example reconstruction time can grow quadratically with pileup, as shown in Fig.1(right).

A fundamental design principle of the upgrade is to increase the measurement resolution for better signal acceptance and background rejection, as early as possible in the trigger selection. Three different approaches are envisaged: first an increase of granularity at the L1 trigger, then an improved resolution and parallelism in the software trigger and lastly the inclusion of the silicon tracker in the hardware trigger decision. All of these are described one by one in the next sections.

2. Upgrades in the first-level trigger

The first approach tries to increase detector granularity at the first trigger level, with higher density electronics and faster data transmission links. Already in Run 3 this will lead to new L1 calorimeter Front-End electronics and new L1 muon detectors and trigger electronics, described in detail in [6].

For the muon L1 trigger, the majority of the rate is produced by either fake muons or muons below the p_T threshold in the forward regions ($|\eta| > 1.0$), because of the complex and reduced magnetic field in this region. To limit this effect, a completely new detector will be installed in this region for Run 3, as shown in Fig.2 (left), together with the design of a completely new front-end and trigger electronics. The New Small Wheel (NSW) [7], based on sTGC and MicroMegas technologies, will improve p_T resolution and robustness of the trigger, halving the L1 muon trigger rate in Run 3. At higher luminosity, in Run 4, the L1 muon trigger selection will be further improved with the p_T measurement of the precision muon chambers (MDTs), that will be fully readout at 40 MHz with a brand new readout electronics, now under design.

Exploiting finer-granularity digital inputs from electromagnetic LAr calorimeter is a powerful tool to suppress pileup by increasing energy resolution. Consequently for Run 3 the LAr front-end electronics will be replaced by a new dense and high-speed electronic system (40 MHz digitisation

with 12-bit precision), and the L1 trigger will make use of a completely new system of processors, called Feature Extractors (FEXes), that will implement dedicated object-finding algorithms at different granularities, even with access to event-based variables (Electron FEX, Jet FEX, Global FEX). An overview is shown in Fig.2 (right) while a similar upgrade of the Tile hadronic calorimeter is planned for Run 4.

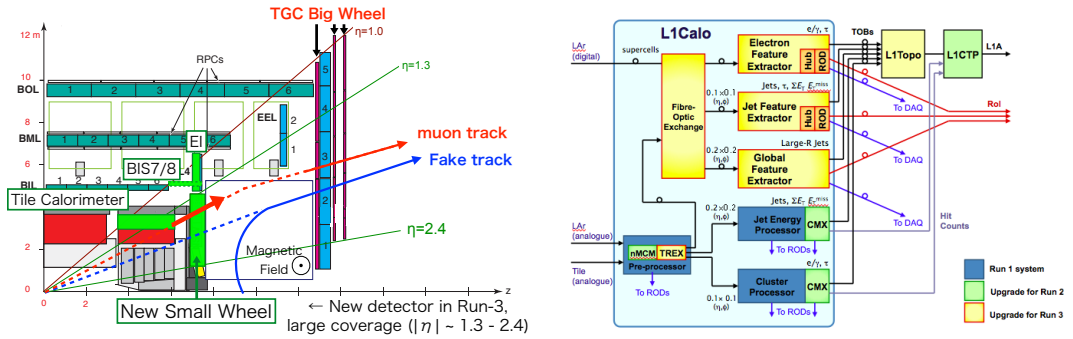

Figure 2: Left: Layout of the L1 muon trigger for Run 3, showing in green the region of the New Small Wheel and the new coincidence logic to fight fake muons from the beam-pipe [7]; Right: Overview of the planned Level-1 Calorimeter trigger system for Run 3, modified from [6]: in blue and green the legacy system, in yellow the new components added as part of the upgrade.

All the new Run 3 components of the L1 trigger are already in the prototyping and testing phase: both functional and integration tests on test-beams are ongoing, with the aim of starting the final production by the beginning of 2018 and installation one year later. QA/rad-tolerance tests on custom chips (LAr FE ASIC), development of complex algorithms in new generation FPGAs, test of high speed (up to 12.8 Gbps) links and data sharing technologies are currently under way. Details of these technologies and ongoing tests are described in [8] for muons and [9] for calorimeters.

3. Software trigger evolution

The second approach taken in these upgrades concerns the HLT software trigger, which needs to increase throughput and maintain efficiency in spite of increased pileup. Algorithms developed in the offline reconstruction to have stable performance with respect to pileup will be adapted for the trigger, with code speed-up and performance tuning. Due to the time needed to derive calibration and alignment (24-48h), ATLAS does not plan to merge the HLT and offline reconstruction like other experiments [12]. Technology trends combine multicore/multithread CPU processing with embedded hardware processors and accelerators, like GPUs and FPGAs, for extreme parallelism, to beat both frequency and memory walls while making code development more and more complex and linked to the hardware resources (vectorization, memory sharing). This is crucial in the trigger processing, mainly for data preparation, tracking and clustering algorithms that dominate CPU time (currently 50-70%) and grow exponentially with pileup. Tracking demonstrators [10] show large improvements in memory consumption when multi-threading paradigms replace current multi-processing, exploiting both multiple events in flight and sub-event parallelism (Fig.3

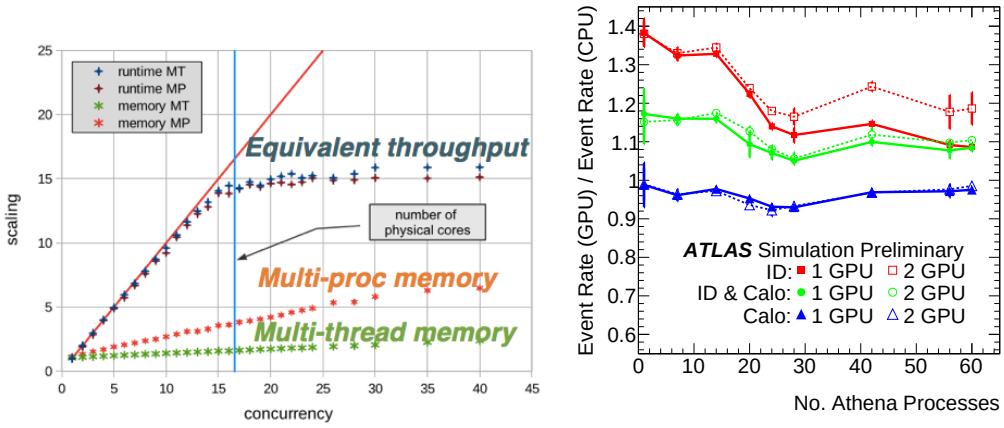

Figure 3: Left: Event throughput and memory scaling comparing AthenaMP (multi-process) and AthenaMT (multi-threaded) approaches for ATLAS simulation [10]; Right: The ratio of event throughput rates with GPU acceleration to the CPU-only rates as a function of the number of ATLAS trigger (Athena) processes running on the CPU [11].

left). Concurrently, tests with GPU prototype acceleration within the ATLAS software showed 20-40% gain in throughput for tracking algorithms with factor five acceleration, as show in Fig.3 (right) and the references there in.

4. Towards a hardware track trigger

The third approach is the inclusion in Run 4 of hardware-based tracking for early rejection, profiting from the extended coverage of the new planned silicon inner tracker (ITk [13]). The presence of a quasi-offline track at a given p_T matched with L1 leptons offers a new opportunity for reducing their rate, overwhelmed by muons from b-hadron decays and jets faking electrons. As shown in Fig.4 (left) even simple algorithms can provide a factor five rejection on these signatures. Tracking information can provide flexible tools to control hadronic signatures, for example identifying primary and secondary vertices and applying track-isolation. Moreover multi-object triggers can easily distinguish objects from the same interaction from random combinations by using the track position along the beam axis.

The hardware track-trigger will use dedicated electronics designed to apply fast algorithms combining hits on a reduced number of silicon layers. The system is an evolution of the current FTK [15] system, with the same two-step selection: track-filtering with pattern-recognition within Associative Memories (AM2020 chip, low power 28nm technology with 250 MHz clock and 0.5 Million patterns storage); secondly track-fitting, with linearised algorithms in new generation FPGAs. This system is massively parallel, with order of 500 boards, and 1-4 MHz input rate. Compared to previous similar projects (starting from CDF-SVX and then ATLAS FTK), technology advances in processing and data sharing make it possible to handle two orders of magnitude larger luminosities and input rates, in addition to almost three orders of magnitude larger number of readout channels. The big challenge in building a track-trigger system is represented by the reduced latency, which needs to include the readout of an extremely high granularity system (order

of 800 million channels) and to fight the large combinatorics expected at increasing pileup (10^4 hits per bunch-crossing expected at HL-LHC). Fig.4 (right) shows the readout latency as a function of rate as measured from a discrete-event simulation of the readout of the ITk-strip system.

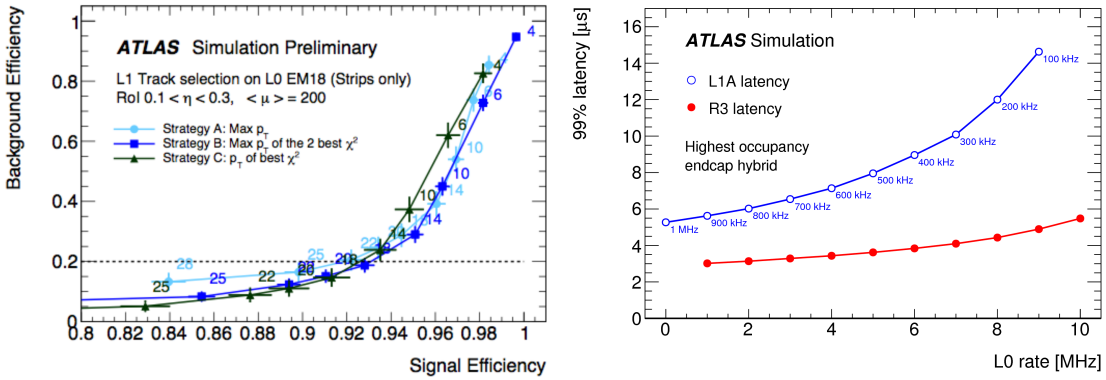

Figure 4: Left: Signal versus background efficiencies for simulated track-trigger electron selection, with three track selection strategies as functions of a track p_T threshold, in one region of interest [14]; Right: Read-out latencies of regional-readout data (R3, red) and L1 data (blue) as a function of the first-level (L0A) rate in the two-level trigger scheme, for the highest occupancy hybrid in the end-cap [13].

As shown in the figure, the system is designed to work in different scenarios, depending on luminosity and L1 rates (that can grow up to 4 MHz). It can help the HLT decision, reducing the processing time per event, as long as the HLT farm is able to process at the L1 accept rate. If this rate is too high, the track-trigger system can instead help to select events before HLT processing, adding a second hardware trigger level for what is known as the two-level trigger scheme. In this case the farm size can be limited, at the cost of reducing the latency of the track-trigger processing to within tens of microseconds. This can be obtained with more parallelism, for example with multiple copies of the same patterns, which require either more AM chips or reduced performance for limiting the pattern content. Depending on the L1 trigger selections, the track-trigger system can work either as regional tracking, seeded by single muon/electron RoIs, or as global tracking, for primary / secondary vertex reconstruction as needed by b-jets, taus and jets selections. The regional readout has the benefit of limiting the readout rate on the ITk front-end chips (only 10% of the full detector volume is expected to be processed in the RoIs) in the two-level architecture scenario. The overall T/DAQ system upgrade will utilise some or all of these options in a new architecture for Run 4.

5. Conclusions and plans

ATLAS is approaching an ambitious and exciting phase of upgrades that involve the T/DAQ system. For Run 3, the Technical Design Review [6] was approved by the LHCC in 2013. The new components are under final design review this year and will be ready for installation in 2019. For Run 4, the LHCC will review the T/DAQ Technical Design Review within the first half of 2018. Once approved, construction will take place until 2023 when installation and commissioning will be done. Different T/DAQ architecture schemes offer the flexibility to face unexpected new conditions and discoveries.

References

[1] L. Evans and P. Bryant (editors) 2008 JINST 3 S08001

[2] Frederick Bordry, "LHC roadmap"

[3] ATLAS Collaboration, 2008 JINST 3 S08003

[4] ATLAS Collaboration, ATLAS-CONF-2017-047

[5] ATLAS Collaboration, "Inner Detector trigger timing study"

[6] ATLAS Collaboration, "Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System", CERN-LHCC-2013-018

[7] ATLAS Collaboration, "New Small Wheel Technical Design Report", CERN-LHCC-2013-006

[8] S.Akatsuka on behalf of ATLAS Collaboration, "The Phase-1 Upgrade of the ATLAS Level-1 Endcap Muon Trigger", ATL-DAQ-SLIDE-2017-325 (TIPP 2017)

[9] A.Camplani on behalf of ATLAS Collaboration, "Phase-I Trigger Readout Electronics Upgrade for the ATLAS Liquid-Argon Calorimeters", ATL-LARG-SLIDE-2017-319 (TIPP 2017)

[10] G.A.Stewart et. al for the ATLAS Collaboration, "Multi-threaded software framework development for the ATLAS experiment", Journal of Physics: Conference Series 762 (2016) 012024

[11] ATLAS Collaboration, "ATLAS Trigger GPU Demonstrator Performance Plots"

[12] LHCb Collaboration, "The Trigger and Online TDR for the LHCb upgrade", CERN-LHCC-2014-016

[13] ATLAS Collaboration,"Technical Design Report for the ATLAS Inner Tracker Strip Detector", CERN-LHCC-2017-005

[14] ATLAS Collaboration, "Tracking for the ATLAS Level 1 Trigger for the HL-LHC"

[15] ATLAS Collaboration, "Fast TracKer (FTK) Technical Design Report", CERN-LHCC-2013-007

PoS(EPS-HEP2017)510