PROCEEDINGS OF SCIENCE

The Phase-1 Upgrade of the ATLAS Level-1 Calorimeter Trigger

Ralf Gugel[a,*] **on behalf of the ATLAS Collaboration**

[a] *Institut für Physik, Johannes Gutenberg-Universität, Staudingerweg 7, 55128 Mainz, Germany*

E-mail: rgugel@cern.ch

During the Long Shutdown 2 of the LHC, the ATLAS detector underwent its Phase-1 Upgrade in order to better handle the increased luminosity in LHC Run 3 and beyond. A significant part of this upgrade comprises new custom hardware for the first-level calorimeter-trigger path using state-of-the-art FPGAs. An overview of the different subsystems, their capabilities and roles in the overall ATLAS trigger and data acquisition architecture, and their relevance for the ATLAS physics program is presented. Finally, exemplary preliminary performance estimates are discussed.

The European Physical Society Conference on High Energy Physics (EPS-HEP2025)
7-11 July 2025
Marseille, France

*Speaker

 ISSN 1824-8039 . Published by SISSA Medialab.

PoS(EPS-HEP2025)564

1. Introduction

The formal transition of CERN's Large Hadron Collider (LHC) into the High-Luminosity (HL) LHC will occur in the Long Shutdown 3, between the ongoing Run 3 and the upcoming Run 4. In practice the increase in instantaneous luminosity provided by the LHC, however, is a gradual process. Already at the end of 2024 the integrated luminosity recorded by the ATLAS experiment [1] during Run 3 exceeds the combined values of Runs 1 and 2. Higher rates of processes of interest come not only with an equivalent increase in background rates but at the cost of a much larger number of simultaneous collisions (*pile-up*) per LHC bunch crossing, μ. The most probable value μ_{MPV} has more than doubled from Run 2 ($\mu_{\mathrm{MPV}} = 29$) to Run 3 ($\mu_{\mathrm{MPV}} = 62$) [2]. The Phase-1 Upgrades of and related to the Level-1 Calorimeter Trigger (L1Calo) were installed during the Long Shutdown 2 and have since gradually replaced the functionality provided by the legacy predecessor system. The newly installed trigger systems have not only been crucial to allow ATLAS to make full use of LHC's performance in Run 3 but also represent a first step towards ATLAS' readiness for the HL-LHC, during which most of the Phase-1 subsystems will continue to form an integral part of the ATLAS hardware trigger.

2. Overview of the ATLAS Trigger and Data Acquisition System

A scheme of the data flow in the current ATLAS trigger and data acquisition (TDAQ) system [3] is shown in Figure 1. Data resulting from collisions are initially buffered on the detector front ends. A reduced subset of detector signals from calorimeters and muon detectors are sent to the first-level trigger (Level-1 trigger) at the full LHC bunch crossing rate of 40 MHz. The Level-1 trigger system employs a variety of algorithms and criteria to select the most interesting events while maintaining an Level-1 trigger rate below 100 kHz limited by the readout capabilities of detector front-end electronics, some of which have been in operation since nearly two decades. Given limited front-end buffer sizes the decision to (not) issue a L1Accept (L1A) signal must be made within a total latency budget of only $\sim 2.5\,\mu$s imposing hard real-time constraints on the architecture of the Level-1 trigger system. Due to the challenging latency and throughput requirements the Level-1 trigger largely consists of a number of custom, FPGA-based hardware modules implementing a pipelined, fixed-latency processing chain. Three Feature EXtractors (FEXes) use calorimeter information to form candidate trigger objects (TOBs). Based on these and L1 muon candidates the topological processor (L1Topo) evaluates various event-level criteria before the Central Trigger Processor (CTP) forms the final L1A decision.

A second, high-level trigger (HLT) stage is implemented via software algorithms running on a farm of commercial off-the-shelf servers. The main requirements regarding the technical performance of the HLT emerge from the Level-1 trigger input rate, the size of off-detector readout buffers, the number of available processing slots (CPU cores) and the allowed bandwidth to CERN's Tier-0 data center for permanent storage and offline analysis. Some signatures such as electrons or photons require particularly CPU-intensive algorithms. To reduce the processing time these algorithms are restricted to detector regions of interest (RoIs) given by TOBs identified by the Level-1 trigger.

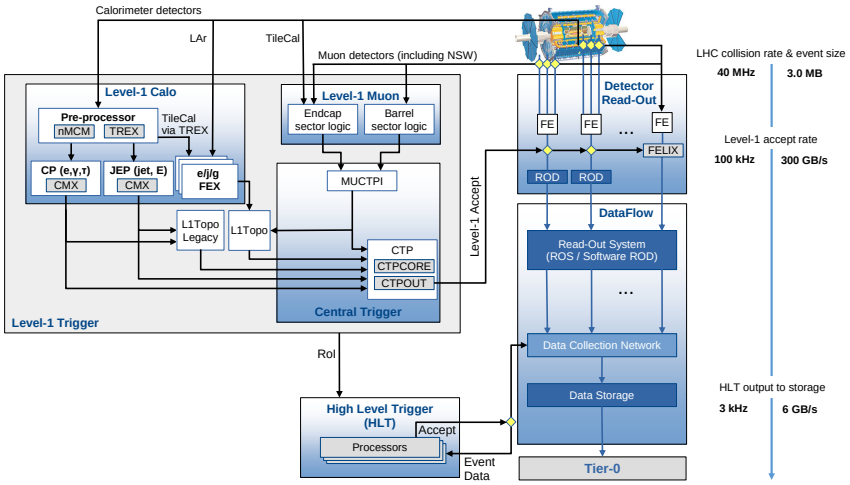


Figure 1: Overview of the ATLAS TDAQ system during Run 3 [3].

3. The Level-1 Calorimeter Trigger

The ATLAS calorimeter system employs two technologies. A scintilating-tile-based hadronic calorimeter covers the pseudorapidity range of $|\eta| \lesssim 1.6$ [4]. Hadronic calorimetry outside of this range as well as electromagnetic (EM) layers are realized by a Liquid Argon (LAr) based calorimeter [5] covering up to $|\eta| \leq 4.9$.

Prior to the Phase-1 Upgrade analog calorimeter signals were summed on-detector to form so-called *trigger towers* of sizes of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the central detector region and even larger towers towards the forward calorimeters. The resulting signals were digitized in the L1Calo PreProcessor system located in a service cavern next to the main detector cavern.

As part of the Phase-1 Upgrade, signals from the LAr calorimeter are digitized on-detector allowing for increased spatial and energetic granularity. Trigger towers are replaced by finer grained Super Cells (SC) increasing the number of corresponding channels by about an order of magnitude.

Digitized samples corresponding to about 25 Tbps are transmitted via optical fibers to backend electronics in the service cavern. There, estimates of transverse energies E_{T} per SC are obtained from the digitized samples using FIR (finite impulse response) filters and further corrections, e.g., those accounting for bunch-crossing dependent effects due to LHC fills consisting of bunch trains[1].

On-detector digitization of signals from the Tile calorimeter for the trigger path is only foreseen as part of the upcoming Phase-2 Upgrade [6]. Hence, newly developed Tile Rear EXtension (TREX) cards have been installed in the relevant subset of legacy PreProcessor crates. These extension cards provide additional optical interfaces to send real-time data to the FEX modules and interface with the new FELIX [7] based readout system which is also employed by various other Phase-1 Upgrade systems.

3.1 L1Calo Feature Extractors

The identification of various calorimeter based signatures is performed by three multi-FPGA based FEX systems. Each FPGA receives data from a fixed detector region at an initiation interval

[1] Sequences of multiple filled bunches colliding with their counterparts followed by several 25 ns time slices without colliding bunches due to the LHC filling pattern which may differ from fill to fill.

Table 1: Technical overview of FEX hardware, connectivity, and algorithmic boundary conditions. Main FPGAs are those performing most of the realtime processing, additional FPGAs and/or SoCs, e.g., for monitoring and slow control are omitted. Based on [1, 3].

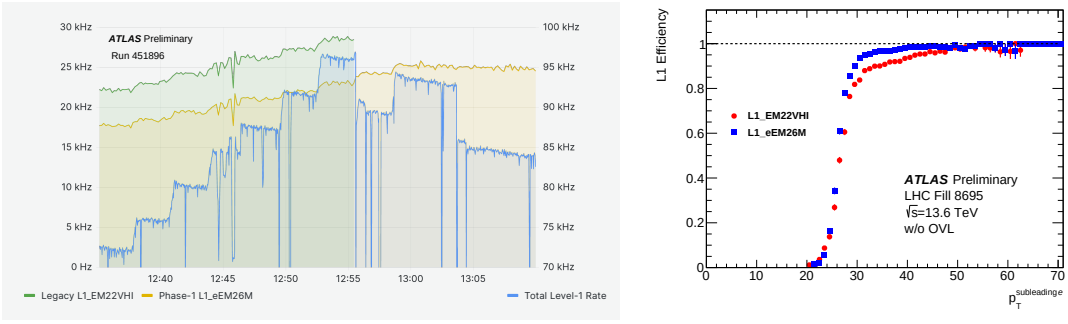
	eFEX	**jFEX**	**gFEX**
max. input granularity ($\Delta\eta \times \Delta\phi$, EM calorimeter)	0.025×0.1	0.1×0.1	0.2×0.2
algorithmic core region	$\lvert\eta\rvert \leq 2.5$	$\lvert\eta\rvert \leq 4.9$	$\lvert\eta\rvert \leq 4.9$
trigger objects (TOBs)	e/γ, τ_{had}	jets, τ_{had}, $\vec{E}_{\text{T}}^{\text{miss}}$, $\sum E_{\text{T}}$, $e/\gamma_{\lvert\eta\rvert>2.5}$	large jets, $\vec{E}_{\text{T}}^{\text{miss}}$, $\sum E_{\text{T}}$
main FPGAs (# per module)	XC7VX550T (×4)	XCVU9P (×4)	XCVU9P (×3)
# LUTs per FPGA	346 k [9]	1182 k [10]	1182 k [10]
# modules in full system	24	6	1
input bandwidth per module	~ 1.4 Tbps	~ 2.2 Tbps	~ 3.4 Tbps
unique bandwidth to L1Topo (number of copies)	537.6 Gbps (×5)	716.8 Gbps (×4)	89.6 Gbps (×3)

of 25 ns, i.e. at the full LHC bunch crossing rate. Besides information from the respective algorithmic core region upstream, an on-board duplication of data from neighboring detector regions is performed to allow for translation equivariant algorithms at the system level. Each FEX offers a different trade-off between (spatial) granularity of its inputs against the algorithmic window sizes available in each FPGA of a given FEX. A summary of technical figures of the FEX subsystems is given in Table 1. Excluding latency due to (de)serialization of data and other transmission related delays the algorithmic latency budget for FEXes in the Phase-1 system is $\lesssim 200$ ns.

The electromagnetic FEX (eFEX) receives E_{T} values at the highest granularity, i.e. individual SCs in case of inputs from the LAr calorimeter. It is therefore particularly suitable for the identification of narrow, high-precision signatures such as EM signatures (electrons or photons) as well as hadronically decaying τ leptons. In order to improve the energy response for EM signatures eFEX allows to apply corrections for varying amounts of detector material upstream of the calorimeters. Furthermore eFEX benefits from the increased granularity as the individual energies can be more accurately assigned to a core and isolation regions for each TOB candidate. As a result a significantly improved discrimination against misidentified EM signatures is achieved as shown in Figure 2.

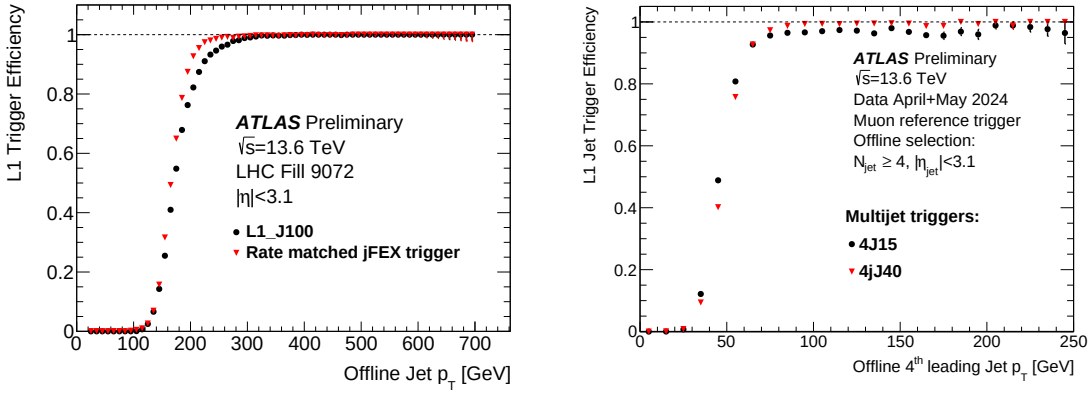
For discrimination of hadronically decaying τ leptons against hadronic jets from quark or gluons eFEX employs a boosted decision tree (BDT) based on sums of supercell energies in different layers and distances from the center of the τ candidate [8]. The BDT was trained using XGBoost [11] and converted for implementation in the FPGAs using Conifer [12].

The jet FEX (jFEX) makes use of intermediate granularity inputs ($\Delta\eta_{\text{min}} \times \Delta\phi_{\text{min}} = 0.1 \times 0.1$) to search for a large set of signatures including EM signatures outside the coverage provided by eFEX, auxiliary τ candidates, jets, and event-level quantities $\vec{E}_{\text{T}}^{\text{miss}}$ and $\sum E_{\text{T}}$. Crucial to the resolution and turn-on behavior especially of $\vec{E}_{\text{T}}^{\text{miss}}$ based triggers it is to account for contributions from pile-up collisions. To this end jFEX does not only allow to set lower thresholds (noise cuts) for each input energy but can also correct the received E_{T} values based on per-event estimates of the pile-up density in a given calorimeter region. The jFEX $\vec{E}_{\text{T}}^{\text{miss}}$ estimate (jXE) is taken as the vectorial



(a) **(b)**

Figure 2: **(a)** Comparison of legacy and Phase-1 L1 rates of the main single-electron triggers and overall ATLAS Level-1 trigger rate. Increase in these rates are due to ongoing trigger tuning and LHC ramp-up whereas the drop around 12:55 is due to the deactivation of the legacy EM trigger with the Phase-1 "eEM" trigger running at an approximately 20% lower rate. **(b)** Efficiencies of the legacy (EM22VHI) and Phase-1 (eEM26M) triggers as functions of the transverse momentum of offline reconstructed electrons passing tight isolation requirements and a tag-and-probe selection including a Z mass constraint [8].

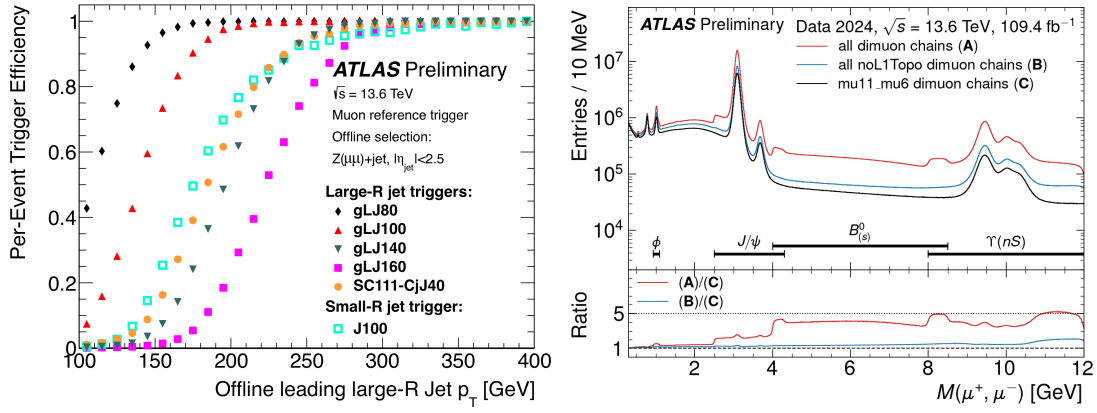


(a) **(b)**

Figure 3: **(a)** Efficiencies of legacy single-jet (J100) and corresponding jFEX triggers. Offline jets are matched to L1 objects requiring $\Delta R < 0.4$ [8]. **(b)** Efficiencies of legacy (4J15) and Phase-1 (4jJ40) four-jet triggers as functions of the transverse momentum of the fourth-leading offline jet [15]. Offline jets used as references are reconstructed using the anti-k_t algorithm with radius parameter $R = 0.4$ and particle flow based constituents [16]. The Phase-1 thresholds shown are tuned to yield the same L1A rate as the corresponding legacy trigger.

sum of E_T values after these corrections. Compared to the legacy $\vec{E}_T^{miss}$ trigger a slightly improved efficiency is achieved while saving about 20% in L1 rate [13].

The tight latency constraints are largely prohibitive of iterative algorithms such as anti-k_t [14]; jFEX jets (jJets) are therefore constructed as E_T sums in a fixed radius around a local-maximum based seed and radius of $\Delta R \lesssim 0.4$. Improvements in the single-jet trigger turn-on shown in Figure 3a are achieved through the capability to apply η dependent calibration factors to the jet E_T sum. The lower granularity in the legacy system gives rise to a plateau inefficiency for multi-jet triggers when two jets are in close proximity. Thanks to the quadrupled input granularity available to jFEX with respect to its predecessor system and a refined seeding algorithm this inefficiency has been vastly reduced as can be seen in Figure 3b.



(a) **(b)**

Figure 4: **(a)** Efficiencies of dedicated large-radius jet triggers compared to the legacy single-jet (J100) Level-1 trigger. The items employed for regular physics data taking are gLJ140 and SC111-CjJ40 [8]. **(b)** Enhancement of sample size across the low-mass di-muon spectrum due to the use of topological criteria (A) compared to using only multiplicity based di-muon triggers (B, C) [18].

The global FEX (gFEX) design is based on the goal to aggregate information from the entire calorimeters in a single module and extensive use of communication between its FPGAs. Its primary purpose is to produce large-radius jets (gLJ) and various global quantities including alternative $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ and $\sum E_{\mathrm{T}}$ estimates.

Exemplary turn-ons for L1 large-radius jet triggers are shown in Figure 4a. The main large-radius jet triggers gLJ140 (gFEX) and SC111-CjJ40 (jFEX small radius jets reclustered in L1Topo) show a more favorable behavior especially near the efficiency plateau. It should be noted that especially the comparison with small-radius jet triggers may differ significantly when evaluated on samples with different large-radius jet compositions. Both large-radius jet triggers incur only small additional L1A rate due to their large overlap with the single jet trigger. HLT trigger chains involving large-radius jets are simultaneously seeded by the single-jet trigger and both large-radius jet triggers.

3.2 The Level-1 Topological Processor

The Phase-1 L1Topo system receives muon candidates via the Muon-to-CTP-Interface (MUCTPI) as well as calorimeter based TOBs from all FEX systems via optical links at 11.2 Gbps per fiber. Each TOB contains information about its transverse energy E_T and direction (η, ϕ) as well as possibly type specific additions such as isolation criteria. L1Topo therefore allows for the most comprehensive view of every event produced in the center of the ATLAS detector including near arbitrary multi-object calculations. The total bandwidth of actively used links from FEXes and MUCTPI to L1Topo amounts to about 1.4 Tbps excluding upstream duplication. Each of the six XCVU9P FPGAs hosted across three identical boards receives only a subset of TOB types as the number of multi-gigabit transceivers in the FPGAs is slightly below the required number allowing for full connectivity. Instead the fiber mapping has been designed such that nearly every combination of TOBs is available in at least one of the FPGAs.

The algorithmic latency budget in the L1Topo system depends on the individual boards due to different outgoing connectivity to the CTP. The two FPGAs on the first board (Topo1) send

their results via two high-speed optical links each whereas FPGAs on the second and third board (Topo2/3) drive LVDS wire pairs in parallel electrical cables. While the optical links allow for a higher bandwidth of $2 \times 2 \times 6.4$ Gbps the required alignment buffers and (de)serialization for such links incur additional latency compared to the parallel, electrical transmission (4×1.28 Gbps).

Several instances of multiplicity algorithms for each calorimeter based TOB type are implemented on Topo1 with a maximum allowed latency of 50 ns. These determine the number of TOBs passing various, possibly object specific requirements such as E_{T} thresholds, restricted η ranges, and isolation or quality levels. Through coordinate matching of jTAU TOBs received from jFEX to eTAU candidates from eFEX so called *combined* cTAUs are formed. Identification levels and E_{T} estimates are inherited from the eTAU TOBs. The matched jTAU supplies a transverse energy sum from a larger surrounding area compared to what is accessible in a single eFEX FPGA, providing additional isolation based discrimination of hadronic decays of τ leptons against QCD jets.

So called *sort/select* algorithms on Topo2 and Topo3 are used to reduce the size of fixed length TOB arrays from up to 168 input TOBs to generally ten or less. Per-TOB criteria similar to those used in multiplicity algorithms are applied. TOBs failing these or empty TOBs are removed[2]. In case of *sort* algorithms the resulting TOB arrays are also sorted by E_{T}. Due to the large number of inputs sort/select algorithms are generally implemented with a latency of up to 50 ns.

Decision algorithms receive one or more lists of TOBs prepared by such sort/select algorithms. Most frequently used criteria involve requirements on the invariant mass and/or angular distances ($\Delta\phi, \Delta\eta, \Delta R$) of tuples of TOBs. Algorithms based on such criteria cover a wide range of applications such as tagging dijet topologies from vector-boson fusion or decays of light resonances. Gains due to such algorithms applied to low-mass di-muon events are shown in Figure 4b. The mapping of inputs to an algorithm's decision, however, is largely arbitrary as long as the required logic fits within limitations due to FPGA logic and routing resources as well as the latency budget of only 25 ns. At present about 80 different decision algorithm instances are defined in the ATLAS trigger menu [3] from which parts of the L1Topo firmware are automatically generated via a dedicated toolkit [19].

One example for a custom decision algorithm is GELATO[3], which in early 2025 became the first neural network deployed in ATLAS hardware. It is based on a variational auto-encoder architecture using 44 features ($E_{\mathrm{T}}, \eta, \phi$ of jJets, eTAUs, muons and $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$). The network is trained offline using Keras/Tensorflow [20, 21] and converted for deployment in the FPGA using HLS4ML [22, 23]. In order to stay within the allowed latency budget, discrimination during inference is solely based on information from the latent space allowing to drop the decoder part from the full auto encoder.

4. Conclusions

The ATLAS Level-1 Calorimeter Trigger has been upgraded and successfully commissioned. Its enhanced capabilities and performance are vital for efficient data taking during Run 3 and beyond. Largest benefits are found for electro-magnetic, as well as dense multi-jet signatures.

[2]Empty TOBs refer to parts of the FEX/MUCTPI→L1Topo data formats which are unused in a particular event. These formats are defined as to largely avoid truncations in case of highly localized activity in the core region of one FEX FPGA.

[3]Global Event Level Anomalous Trigger Option

References

[1] ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider: a description of the detector configuration for Run 3* , 2024, JINST 19 (2024) P05063

[2] ATLAS Collaboration, *Luminosity Public Results Run 3*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3

[3] ATLAS Collaboration, *The ATLAS Trigger System for LHC Run 3 and Trigger performance in 2022* , 2024, JINST 19 (2024) P06029

[4] ATLAS Collaboration, *Operation and performance of the ATLAS Tile Calorimeter in Run 1*, 2018, Eur. Phys. J. C **78** (2018) no.12, 987

[5] M. Aleksa *et al.*, *ATLAS Liquid Argon Calorimeter Phase-I Upgrade*, 2013, CERN-LHCC-2013-017

[6] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter*, 2017, CERN-LHCC-2017-019

[7] ATLAS TDAQ Collaboration, *The ATLAS FELIX Project*, 2022, doi:10.5281/zenodo.6347080

[8] ATLAS Collaboration, *Level-1 Calorimeter Trigger Public Results* , https://twiki.cern.ch/twiki/bin/view/AtlasPublic/L1CaloTriggerPublicResults

[9] AMD/Xilinx®, *7 Series FPGAs Data Sheet: Overview*, 2020, https://docs.amd.com/api/khub/documents/2LByHkO nSZXcei2D55fTg/content

[10] AMD/Xilinx®, *UltraScale™ Architecture and Product Data Sheet: Overview*, 2025, https://docs.amd.com/v/u/en-US/ds890-ultrascale-overview

[11] T. Chen, C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, 2016, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

[12] S. Summers *et al.*, *Fast inference of Boosted Decision Trees in FPGAs for particle physics*, 2020, JINST 15 P05026

[13] ATLAS Collaboration, *Missing Et Trigger Public Results*, 2024, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MissingEtTriggerPublicResults

[14] M. Cacciari et al, *The anti-k_t jet clustering algorithm*, 2008, JHEP04(2008)063

[15] ATLAS Collaboration, *Jet Trigger Public Results*, 2024, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetTriggerPublicResults

[16] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, 2017, Eur. Phys. J. C 77, 466

[17] D. Bertolini *et al.*, *Jet observables without jet algorithms*, 2014, J. High Energ. Phys. 2014, 13 (2014)

[18] ATLAS Collaboration, *BPhysics Trigger Public Results*, 2024, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BPhysicsTriggerPublicResults

[19] ATLAS TDAQ Collaboration, *TopoConverterPh1*, 2025, https://gitlab.cern.ch/atlas-l1calo/l1topo/topoconverterph1

[20] F. Chollet *et al.*, *Keras*, 2015, https://keras.io

[21] M. Abadi *et al.*, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015, https://www.tensorflow.org

[22] FastML Team, *fastmachinelearning/hls4ml*, 2025, 10.5281/zenodo.1201549

[23] J. Duarte *et al.*, *Fast inference of deep neural networks in FPGAs for particle physics*, 2018, JINST 13 P07027

[24] ATLAS Collaboration, *Triggering with Anomaly Detection in Run 3*, 2025, ATL-COM-DAQ-2025-039