

The ATLAS Hardware Track Trigger design towards first prototypes

Sebastian Dittmeier*, on behalf of the ATLAS TDAQ collaboration

Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg,

Im Neuenheimer Feld 226, 69120 Heidelberg, Germany

E-mail: sebastian.dittmeier@cern.ch

For the High-Luminosity LHC, planned to start in 2027, the ATLAS experiment will be equipped with the Hardware Tracking for the Trigger (HTT) system, a dedicated hardware system able to reconstruct tracks in the silicon detectors with low latency. The HTT will be composed of about 700 ATCA boards, based on new technologies available on the market, like high speed links and powerful FPGAs, as well as custom-designed Associative Memory ASICs, which are an evolution of those developed for the ATLAS Fast Tracker. The HTT is designed to cope with the expected extreme high luminosity in the so-called L0-only scenario, where the HTT will operate at the L0 rate (1 MHz). It will provide good quality tracks to the software High-Level-Trigger (HLT), operating as coprocessor to lighten the load of the software tracking. The implementation of the HTT allows the HLT farm size to be reduced by a factor of 10. All ATLAS sub-detector systems are designed also for an evolved, so-called "L0/L1", architecture, where part of the HTT is used in a low-latency mode (L1Track), providing tracks in regions of ATLAS at a rate of up to 4 MHz, with a latency of a few micro-seconds. This evolved architecture poses very stringent requirements on the latency budget and to the dataflow rates. All the requirements and the specifications of this system have been assessed. The design of all the components has been reviewed and validated with preliminary simulation studies. Soon, the development of the first prototypes will start. In this paper we describe the status of the HTT design, discuss the challenges and assessed specifications, towards the preparation of the first slice tests with real prototypes.

The 28th International Workshop on Vertex Detectors - Vertex2019

13-18 October, 2019

Lopud, Croatia

*Speaker.

1. Introduction

Starting operation in 2027, the High-Luminosity Large Hadron Collider (HL-LHC) [1] at CERN will exceed the LHC's nominal luminosity up to an ultimate peak value of $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. This enhancement in luminosity is accompanied by an increased number of inelastic proton-proton collisions per bunch-crossing (pile-up μ) of $\langle \mu \rangle \approx 200$. To cope with the higher event rates, increasing detector occupancies and radiation levels, the ATLAS [2] detector is planned to be upgraded, referred to as the Phase-II upgrade of the experiment. This upgrade involves the installation of the Inner Tracker (ITk) [3], a new all-silicon tracking detector, which will be able to withstand the high particle rates. Furthermore, the ITk enhances the tracking detector coverage with respect to the current ATLAS Inner Detector [4] up to pseudorapidities $|\eta| = 4$.

The ATLAS experiment aims to continue a broad physics programme at the HL-LHC, ranging from precision measurements of the Standard Model parameters including properties of the Higgs boson to flavour and heavy ion physics as well as more sensitive searches for signatures of physics beyond the Standard Model. To cover this large variety of physics, an inclusive trigger selection is used. The challenge for the ATLAS Trigger and Data Acquisition (TDAQ) system is to maintain low thresholds especially for single- and di-leptonic, but also hadronic signatures, while being able to handle the extreme rates and pile-up conditions at the HL-LHC. Therefore, the TDAQ system is also required to undergo a major upgrade [5]. This upgrade includes the addition of a hardware-based tracking system, referred to as Hardware Tracking for the Trigger (HTT), which allows trigger background rates to be significantly reduced by exploiting the excellent transverse momentum resolution of the ITk.

2. TDAQ Phase-II Architecture

The baseline architecture for the Phase-II TDAQ system, see Figure 1a, relies on a single Level-0 (L0) hardware trigger that processes data from the calorimeter and muon systems at 40 MHz. The processors deliver the L0 trigger decision at a rate of 1 MHz within a latency budget of $10 \mu\text{s}$. The triggered detector data is transferred to the Event Filter, where particle tracks are reconstructed using the ITk data. The Event Filter selects events according to the trigger menu and reduces the output rate of the data sent to permanent storage to 10 kHz.

The upgraded Event Filter system provides high-level trigger functionality. It consists of a CPU based processing farm complemented by the HTT acting as coprocessor. The addition of the HTT to the system reduces the CPU requirements significantly. Simulations show that the size of resources for tracking in the computing farm can be reduced by a factor of 10 [5].

The HTT is split to perform two different tasks, regional and global tracking. Regional tracking will be performed at the full L0 rate of 1 MHz in regions that are defined by the L0 trigger. These regions are limited to contain less than 10 % of the full ITk data. Within these regions of interest, all charged particles with $p_T > 2 \text{ GeV}$ are reconstructed. Regional tracking is meant for quick initial background rejection for single high- p_T lepton triggers as well as for multi-object triggers. A reduction factor of 5 can be achieved for muonic signatures at a signal efficiency of more than 98 %, see Figure 2a, without exploiting the full resolution of the ITk by only using a subset of the

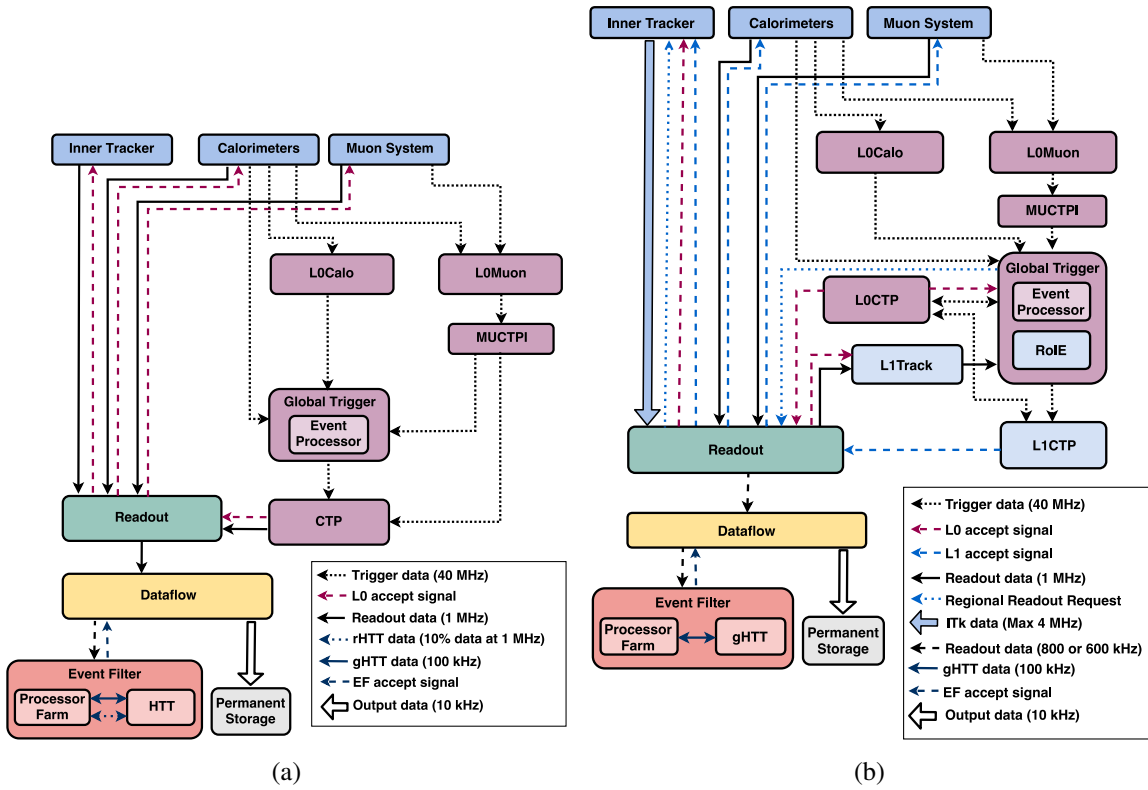


Figure 1: The two possible architectures of the TDAQ Phase-II upgrade: (a) the baseline architecture, where the HTT acts as coprocessor in the Event Filter [5]; (b) the evolved architecture, which includes low latency regional tracking (L1Track) contributing to the global trigger decision [5]. In the latter scenario, global hardware tracking is still performed in the Event Filter.

detector layers. More specifically, regional tracking only uses data from the outer ITk layers, which can also be equipped more easily with high-speed links.

Global tracking will be applied at a lower rate of 100 kHz, depending on the trigger menu. The full ITk event data will be processed and all tracks with $p_T > 1$ GeV will be reconstructed. Global HTT will provide tracks to the Event Filter with a quality close to offline track reconstruction. Global tracking will be requested mainly for hadronic signatures, to enhance primary vertex identification and to correct and mitigate for pile-up. Figure 2b displays that backgrounds contributing to the L0 trigger rate for missing transverse energy (MET) trigger signatures can be significantly reduced by means of global HTT.

Each ATLAS sub-detector system will be capable of evolving to a dual-level hardware trigger architecture, see Figure 1b. In this architecture L0 trigger rates can be up to 4 MHz. Trigger rates are subsequently reduced by a Level-1 (L1) trigger to rates of 600 kHz to 800 kHz before the data enters the Event Filter. This L1 trigger relies on regional hardware-based track reconstruction by reconfiguring part of the HTT, which is then called L1Track. The L1Track system reconstructs all tracks with a transverse momentum $p_T > 4$ GeV in regions specified by the L0 trigger. The tracks have to be provided to the Global Trigger within a latency budget of $6 \mu\text{s}$. In the Global Trigger, calorimeter and muon based trigger objects are combined with the reconstructed tracks, before the

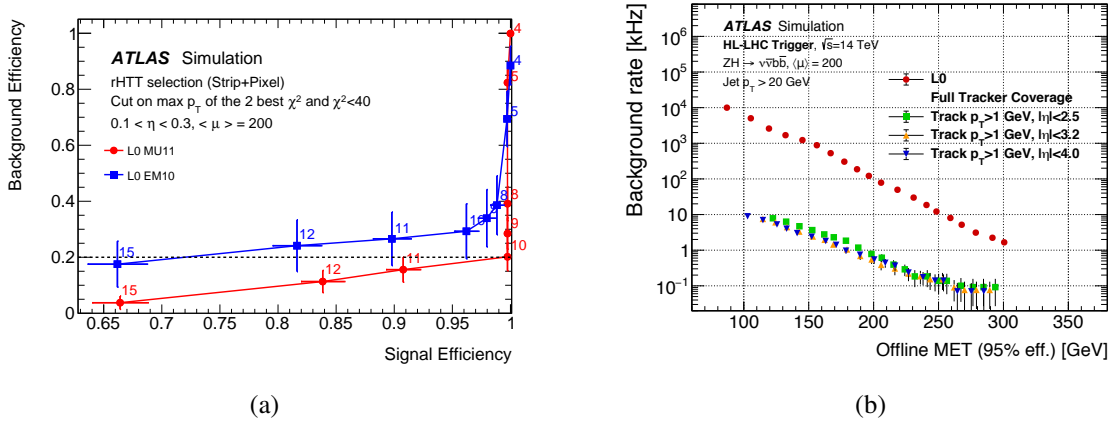


Figure 2: Simulated background reduction by applying regional and global hardware tracking in the Event Filter of the baseline TDAQ Phase-II architecture. (a) Signal versus background efficiency for L0 muon and electron signatures after regional tracking [5]. The data points correspond to different cuts on the transverse momentum (given in GeV). (b) Background rate versus missing transverse energy after offline reconstruction, after the L0 trigger, and after applying global hardware tracking in different η regions [5].

Central Trigger forms the L1 trigger decision.

On the one hand, this evolution scenario serves as a mitigation strategy if detector occupancies due to pile-up exceed expectations such that the detector readout capabilities are limited by the available bandwidth, or if the trigger rates for hadronic signatures have been significantly underestimated. On the other hand, the architectural evolution also provides opportunities to further reduce trigger thresholds and therefore improve the acceptance for physics signatures. While the evolved scenario is not the key driver for the system specifications, the HTT system is currently being developed to fulfill both the requirements for the baseline and the evolved architecture.

3. Hardware Track Trigger Architecture

There are three major data processing steps within the HTT. Firstly, Associative Memory (AM) ASICs are used for pattern matching of incoming detector data with pre-computed patterns of tracks derived from detector simulations. Only detector information of the eight outermost detector layers is stored in the patterns. The patterns do not make use of the full ITk detector resolution. Instead, several pixels or strips are combined to form so-called superstrips, which is a trade-off between momentum resolution and number of patterns that have to be stored. Secondly, for successfully matched patterns, a first-stage track fit is computed using the same clusters as used for pattern matching, but utilizing the full cluster resolution. These two tasks are performed for regional and global tracking requests, as well as for L1Track. Thirdly, for successfully fitted tracks, a second stage track fit is performed adding detector information from the remaining ITk detector layers, especially including all of the pixel layers, which enhances the track parameter resolution significantly. This processing step is only performed if global tracking is requested.

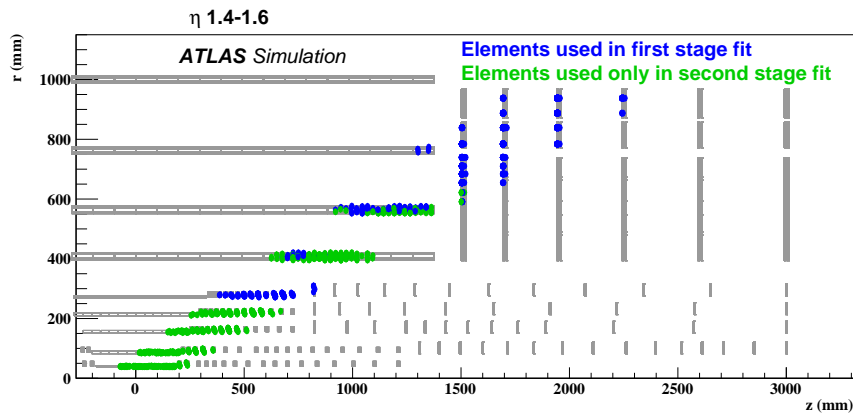


Figure 3: Elements of the ITk detector used for first and second stage track fits within a specific η region [5].

Figure 3 shows an illustration of the elements used in the first and second stage track fitting in a specific η region.

The architecture of the HTT is an evolution of the ATLAS Fast Tracker (FTK) [6] design. The HTT simplifies the hardware modularity by relying on a single ATCA main board, called Tracking Processor, that can be loaded with one of two different mezzanines depending on the application. For pattern recognition and first stage track fitting, one Pattern Recognition Mezzanine is mounted. The unity of main board and mezzanine is called Associative Memory Tracking Processor. To perform the second stage track fitting, two Track Fitting Mezzanines are mounted. The board is then referred to as Second Stage Tracking Processor. In the baseline design, see Figure 4, twelve Associative Memory Tracking Processors and two Second Stage Tracking Processors share one ATCA crate and form one independent unit of the HTT that processes data of a specific $\eta \times \Phi$

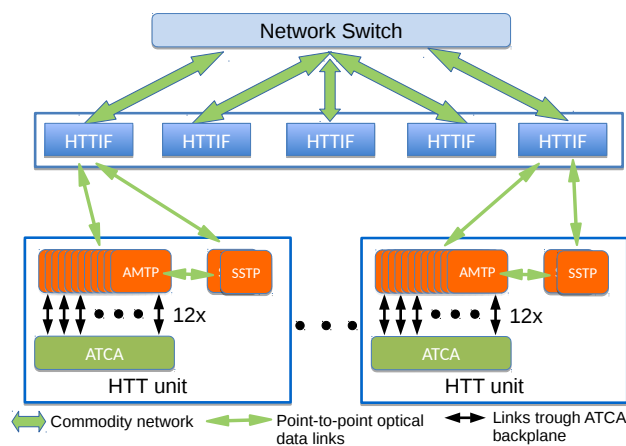


Figure 4: System diagram of the HTT in the baseline design [5]. The interconnections between the components within an HTT unit and between the units and the network are displayed. AMTP = Associative Memory Tracking Processor, SSTP = Second Stage Tracking Processor.

region. The numbers of Associative Memory Tracking Processors and Second Stage Tracking Processors are optimized for the expected rates at which hardware tracking will be requested from the trigger menu. The HTT comprises in total 48 of these units to cover the full ITk.

In contrast to the FTK, the HTT receives the data from the network via dedicated servers using the FELIX [7] cards, referred to as the HTT-Interface (HTTIF), rather than from the detectors directly. This allows commissioning to be initiated without beam before the start of Run 4. This also enables the HTT to be used as an offline coprocessor for Monte Carlo events. For the evolved scenario, however, dedicated low latency streams are foreseen to be used to transfer the data directly from the ITk to the regional tracking units that form L1Track.

4. Detailed Functional Description

The Tracking Processor will perform various tasks: It receives hit data via the Rear Transition Module (RTM), that is optically connected to the HTTIF servers. Furthermore, it shares and switches data between other Tracking Processors via the ATCA crate's backplane. Within the Tracking Processor's main FPGA, hits are clustered and provided to the mezzanines. From the mezzanines, the Tracking Processor receives the track output, removes duplicate tracks and sends them to the RTM for optical transmission downstream. A System-on-Chip is used for monitoring and control.

The Pattern Recognition Mezzanine hosts one powerful FPGA and four groups of five AM ASICs, which are the core component of the HTT design. The production version AM09 will offer an immense computational power of about 30 petabit comparisons per second. The AM09 will store $3 \times 128\text{k}$ patterns and will be able to process incoming data at a speed of 250 MHz. It will be produced in a 28 nm high performance computing process. The basic building block of the chip are the so-called KOXORAM+ cells, that are optimized for an even lower power consumption with respect to the KOXORAM cells tested in the latest prototype AM07 [8], which has been extensively characterized in the lab and found to be fully functional [9]. The next prototype version, AM08, will be submitted early 2020, featuring the new cell design and LVDS interface, whose design has already been silicon proved in the TIMESPOT submission [10].

The FPGA on the Pattern Recognition Mezzanine will handle communication with the AM ASICs and perform linearized χ^2 -fits relying heavily on the available digital signal processing blocks. The Pattern Recognition Mezzanine will be able to perform up to $4 \times 1\text{ Gfit/s}$, which is a specification driven by L1Track. For the computation of χ^2 and track parameters, numerous constants and parameters have to be accessed by the FPGA at very high rates. This need for high-bandwidth memory access is one of the main motivations for the choice of the FPGA, the candidate being an Intel Stratix 10 MX2100, which features several gigabytes of High-Bandwidth Memory (HBM) that is stacked on-top of the FPGA within the same package. The HBM removes the need to put an external RAM on-board which simplifies the board design significantly.

The Track Fitting Mezzanine will be used to extrapolate tracks from the Pattern Recognition Mezzanine to the remaining layers of the ITk not used in the first stage track fit, and to perform a linearized χ^2 -fit. For similar requirements on high-bandwidth memory access, the same FPGA has been chosen as the target device as for the Pattern Recognition Mezzanine. In contrast to the Pattern

Recognition Mezzanine, the mezzanine is only half the size, which means that two mezzanines will be mounted per Second-Stage Tracking Processor board.

5. System Aspects and Development Status

The HTT will be a massive parallel system. In the baseline design, it consists of 576 Associative Memory Tracking Processors, hosting 11 520 AM ASICs, and 96 Second Stage Tracking Processors. During 2019, power estimates have been significantly refined for all components, as the project has undergone and passed an internal system specification review. The total power budget for the HTT system adds up to 385 kW including contingency and margin. The HTT will process a system-wide hit data rate of 3.2 Tb/s, with about 1 Tb/s of output bandwidth.

At the time of writing, demonstrators for all boards are being developed. A mechanical demonstrator for the Tracking Processor will be available soon to perform mechanical and thermal studies as well as electrical studies of the Samtec Z-ray interposers [11], which are the candidate connectors for mounting the mezzanines. A fully functional Tracking Processor demonstrator will be submitted soon afterwards and is expected to be available at CERN from summer of 2020 to perform cooling and integration studies towards a full vertical slice. A demonstrator of the Pattern Recognition Mezzanine is expected to be ready in spring of 2020, while a Track Fitting Mezzanine demonstrator should arrive in early 2020.

6. Simulated Performance

Detailed simulation studies have been conducted to study the performance of the HTT and check that requirements can be fulfilled. Here, only few examples are discussed. More details can be found in [5]. Figure 5a shows simulated track finding efficiencies for high p_T muons using regional HTT, studied in different η regions. The requirement of an efficiency larger than 98 % for $p_T > 10$ GeV is fulfilled in all regions. Figure 5b shows the z_0 track parameter resolution for muons versus transverse momentum in different η regions using regional HTT. The requirements, varying with η , are summarized in Table 1 for the studied pseudorapidities. All requirements can be achieved.

η range	σ_{z_0} [mm]
$0.1 < \eta < 0.3$	0.8
$0.7 < \eta < 0.9$	0.8
$1.2 < \eta < 1.4$	3.8
$2.0 < \eta < 2.2$	7.1

Table 1: First-stage track fitting z_0 resolution requirements (σ_{z_0}) for muons ($p_T > 4$ GeV) in different pseudorapidity (η) regions.

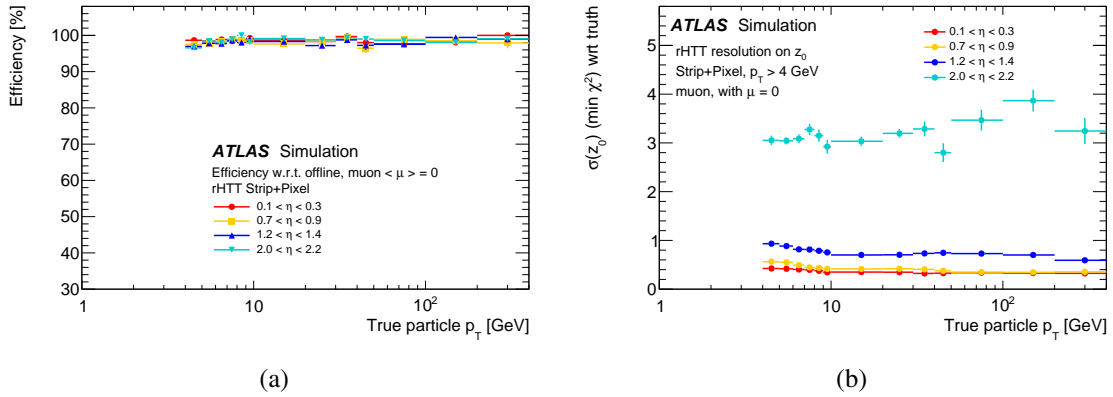


Figure 5: Simulated performance of the regional HTT, (a) track-finding efficiency [5], (b) resolution of the z_0 track parameter [5], as a function of the true muon p_T , for $p_T > 4$ GeV.

7. Summary and Outlook

The HTT is an important element of the ATLAS Phase-II TDAQ system at the HL-LHC. In the baseline architecture, the HTT is designed to be operated as a coprocessor to a CPU based Event Filter farm, significantly reducing the CPU requirements. Two types of hardware tracking will be performed: regional tracking at the full L0 trigger rate of 1 MHz using a subset of the detector data, and global tracking at a reduced rate of 100 kHz, using the full detector data. The HTT is also designed to be used in the so-called evolved TDAQ architecture in a low-latency mode referred to as L1Track. The L1Track provides tracks to the Global Trigger at rates up to 4 MHz within a latency of a few microseconds in regions defined by the L0 trigger before the data enters the Event Filter system.

The HTT design relies on pattern matching in AM ASICs and linearized χ^2 track fitting in FPGAs. A single ATCA main board, the Tracking Processor, is used that can be mounted with one of two different types of mezzanines depending on the purpose of the board: the Pattern Recognition Mezzanine, that hosts the AM ASICs and a powerful FPGA also used for first stage fitting, or the Track Fitting Mezzanine, that hosts another powerful FPGA for second stage fitting. Demonstrators for all components are currently being developed and will become available for integration studies within 2020.

References

- [1] G. Apollinari et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*. CERN-2017-007-M. CERN, 2017.
- [2] The ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider*. JINST **3** S08003. IOP Publishing, 2008.
- [3] The ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*. CERN-LHCC-2017-021. CERN, 2017.

- [4] The ATLAS Collaboration. *ATLAS inner detector: Technical Design Report, 1*. CERN-LHCC-97-016. CERN, 1997.
- [5] The ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*. CERN-LHCC-2017-020. CERN, 2017.
- [6] The ATLAS Collaboration. *Fast TracKer (FTK) Technical Design Report*. CERN-LHCC-2013-007. CERN, 2013.
- [7] on behalf of the ATLAS TDAQ Collaboration S. Ryu. *FELIX: The new detector readout system for the ATLAS experiment*. *J. Phys.: Conf. Ser.* 898 032057. 2017.
- [8] A. Annovi et al. “Design and Characterization of New Content Addressable Memory Cells”. In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. <https://ieeexplore.ieee.org/document/8351682>.
- [9] A. Annovi et al. “Characterization of an Associative Memory Chip in 28 nm CMOS Technology”. In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. <https://ieeexplore.ieee.org/document/8351801>.
- [10] A. Lai. “Recent results of the TIMESPOT project on sensors and electronics developments for future vertex detectors”. In: *The 28th International Workshop on Vertex Detectors (VERTEX2019)*. 2019 (publication pending).
- [11] Samtec. *Z-Ray® Ultra-Low Profile High-Density Arrays*.