Alma Mater Studiorum · University of Bologna

School of Science
Physics and Astronomy Department
Master Degree in Physics

Feasibility study and emulation of Hough transform algorithm on FPGA devices for ATLAS Phase-II trigger upgrade

Supervisor:
Prof. Alessandro Gabrielli

Co-Supervisor:
Fabrizio Alfonsi

Presented by:
Giacomo Levrini

Academic Year 2020/2021

Dedicated to all of you, who stayed with me until the end.

From the deepest part of me, sincerly, thank you.

Darker days seem to be
What will always live in me
But still I run
Its hard to walk this path alone
Hard to know which way to go
Will I ever save this day?
[Oper Your Eyes - Alter Bridge]

Abstract

In the next 10 years, a radical upgrade is expected for the Large Hadron Collider (LHC), mainly focused in achieving the highest possible values in the instantaneous luminosity (at maximum $7.5 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$) and the integrated one (over a factor of 10 with respect to the current one). As a consequence, both the subdetectrors of the main experiments at CERN and their data acquisition systems will need an upgrade, in order to manage a data flux considerably larger if compared with the one used until now.

For the Phase-II upgrade of the Trigger and Data Acquisition System (TDAQ) of the ATLAS experiment a common platform has been created to share the common firmware, software and tools that are ongoing and that will come in the next years within the ATLAS TDAQ collaboration. The environment includes a set of design procedures, a virtual machine as repository for the firmware and some automatic tools for the continuous integration and versioning. The platform is under testing, some aspects have to be consolidated and agreed within the TDAQ collaboration but they have been definitely required by ATLAS. As the firmware will be tested on the TDAQ upgraded boards in general, it will also be used for the prototype cards that will be produced as demonstrator for the ATLAS Hardware Tracking for the Trigger (HTT) system. For the HTT project in general a physical environment is being prepared, exploiting Peripheral Component Interconnect express (PCIe) and ACTA crates.

My personal work has been the design and implementation of a part of some track-fitting algorithms, in particular the one using the Hough Transform, to be integrated into the ATLAS Trigger Phase-II Upgrade. This implementation has been highly required by the ATLAS experiment as an alternative solution to the baseline proposal accepted and described in the TDAQ Upgrade Technical Design Report (TDR). Along with the study of this Hough Transform implementation I have developed and tested a set of pattern vectors used not only in the simulation and validation of the algorithm, but also in the hadware integration on a FPGA-based hardware accelerator. The used technology is based on high-performance Xilinx Ultrascale+ FPGA, implemented on the physical VCU1525 board.

This work is going to be validated by the ATLAS collaboration very soon, so to understand how we can proceed in the future upgrade. It is worth saying that Bologna is the only Italian institute which participates in the integration of a tracking algorithm in the ATLAS trigger upgrade, using high performance FPGA-based hardware.

Contents

Introduction

The first chapter is a brief overview of the Large Hadron Collider (LHC) with a summarizing description of the ATLAS multipurpose detector current setup, the major components and the final objectives of the latest years of operations.

The second chapter delivers a general presentation of the proposals for the Phase-II upgrade of the ATLAS experiment, giving the major changes and differences from the previous experiment's setups and the technological advancement of the whole experiment. In general it will be presented the development from the letter of intent of the experiment, and the further evolutions that the proposals have reached.

The third chapter focuses the on the upgrade of the trigger and data acquisition (TDAQ) system architecture of the ATLAS experiment, exploring the different proposals for the Phase-II upgrade. The two main designs are a baseline one, directly from the starting proposals, and evolved one, derived from the changes in the luminosity and pileup conditions. A particular interest is also posed in the differences proposed in the two designs for the Hardware Tracking for the Trigger (HTT), which provide the usage of Associative Memories in the baseline configuration, with respect to the implementation in hardware of the Hough transform in the evolved one.

The fourth chapter presents the heart of the task accomplished during the thesis working period. Briefly, a main introductory scope for the hardware design production is given, with a presentation of the Hough Transform (HT) algorithm implemented in hardware (for the HTT). The passage continues presenting the VCU1525 board and some of its technical details. Successively a paragraph is dedicated to the transceiver technology used in the board, eventually reaching the device emulation section, where the simulation and the actual hardware programming are presented.

The fifth chapter is a conclusive overview on the presented work and on the future developments.

The appendix A deepens some technical details regarding electronics, in particular a simple introduction to the FPGAs, the description of the JTAG system, the I^2C protocol and the Aurora protocol.

Chapter 1

LHC and ATLAS experiment

1.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is currently the biggest circular hadron collider in the world, with a circumference of 27 km; its purpose is to accelerate protons and heavy ions which will collide and consecutively allow the study of the high energy particle interactions. It is situated nearby the city of Geneva, between the French and the Swiss border, 100 m underground. It is located in the old Large Electron-Positron (LEP) collider location, and a collaboration of 22 nations known as CERN (Conseil Européen pour la Recherche Nucléaire) works together at high energy physic experiment. In Figure 1.1 is shown the LHC underground complex.

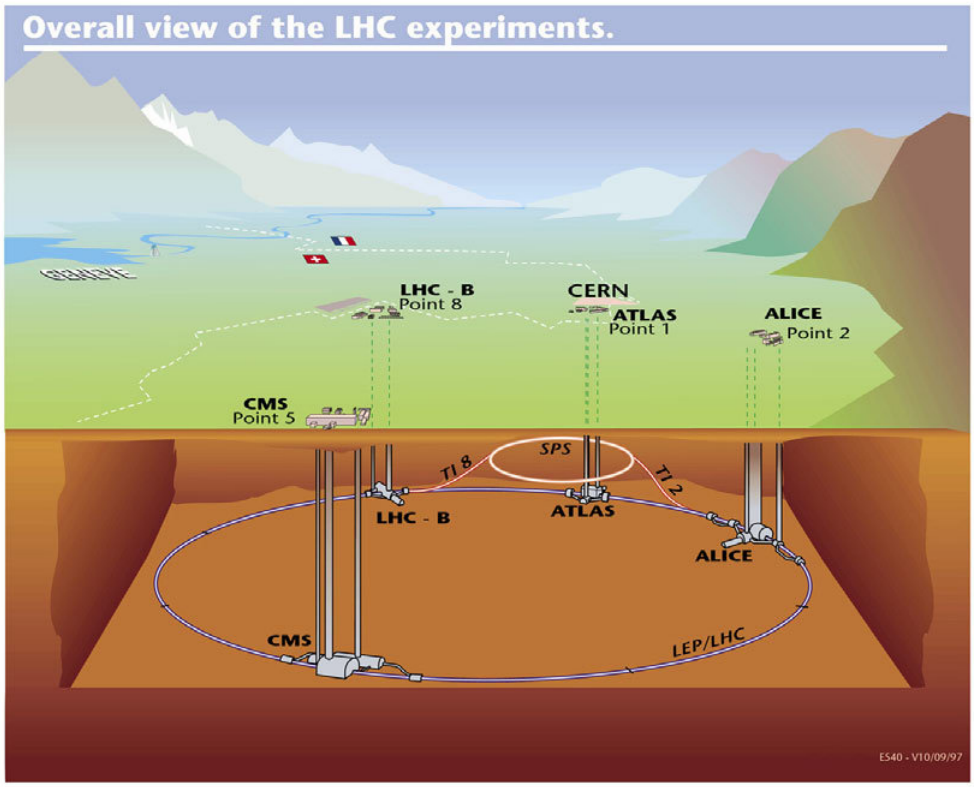

Figure 1.1: *The LHC underground complex below the French-Swiss border.*

Unlike previous particle-antiparticle colliders, in which both beams share the same phase space in a single ring, the LHC machine is based on a proton-proton collision. The two beams of protons travel in two counter-rotating rings, crossing in eight different locations (so called points) along the ring.

The two-ring superconducting collider is composed of a particle source and a series of accelerators which together should allow proton beams collisions with a total centre of mass energy of 14 TeV, that will be reached in the next Run upgrade. Currently the maximum energy reached is 13 TeV, 6.5 TeV for each proton. The proton source is an hydrogen container where the particles are split in their fundamental components, a proton and an electron; then the protons are collected and sent to the first stage of the acceleration of LHC, LINAC2, a linear accelerator where the protons reach the energy of 50 MeV. Then the protons are accelerated by 3 synchrotron accelerators before reaching the last accelerator (LHC): PSB (Proton Synchrotron Booster) where the protons reach 1.4 GeV, PS (Proton Synchrotron) where the protons reach 25 GeV, and SPS (Super Proton Synchrotron) which accelerates protons at 450 GeV. At the end, LHC, with radio frequency cavities working at 400 MHz, pushes the proton beams, at 6.5 TeV each, in the beam pipes.

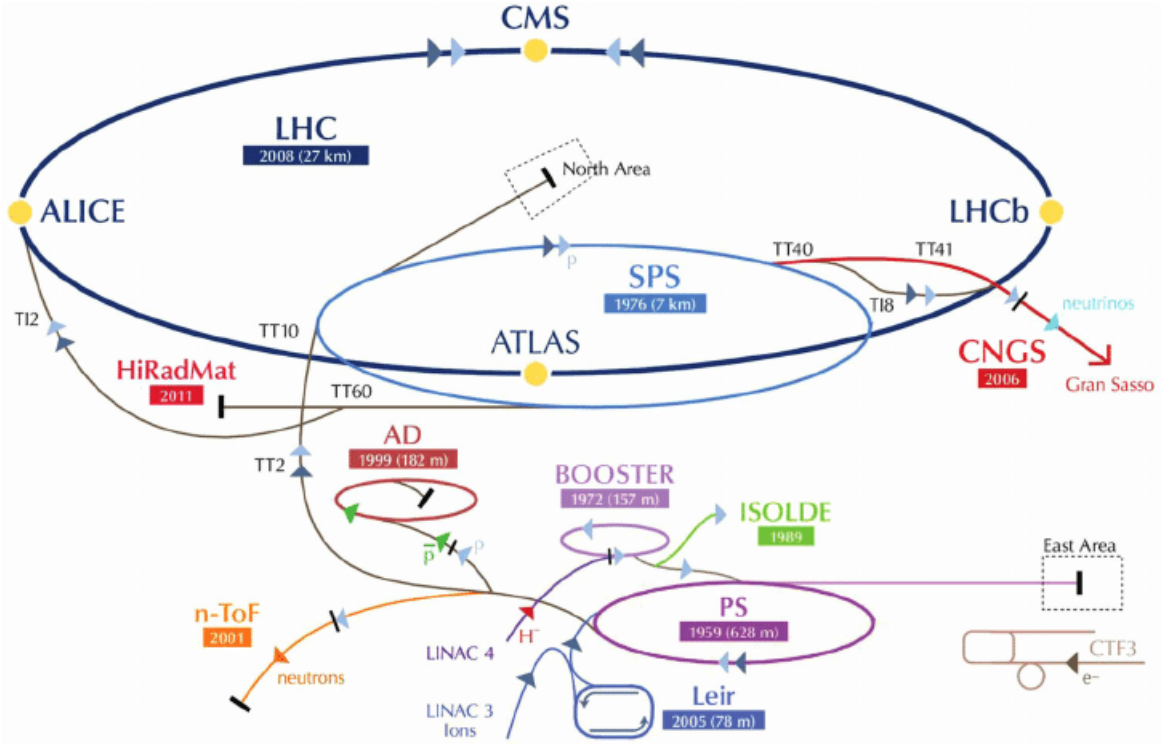

Figure 1.2: *The LHC accelerators complex. The figure also shows the injection chain of the beams: it starts with the linear accelerator LINAC, followed by the Proton Synchrotron Booster (PSB). The Proton Synchrotron (PS) further increases the beam energy and feeds the particle packets to the Super Proton Synchrotron (SPS). After the SPS the beams are injected directly into the LHC.*

In Figure 1.2 the acceleration process of LHC is shown. The proton beams are kept in the 2 beam pipes and far from its walls by a complex system of magnets,

precisely with 1232 superconducting dipole magnets, which maintain the beam in the beam pipes, and with 392 quadruple magnets for the beams focusing. After the acceleration, the beams collide in the 4 collision points where the 4 experiments now active at CERN are situated. The final beam consists of a 2808 bunches of protons, formed by the radio frequency cavities, with $\approx 1.2 \times 10^{11}$ protons per bunch. The collisions allows to reach an instantaneous luminosity with a peak of $1034\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

1.1.1 The LHC experiments

The bunches collisions occurs in correspondence of the 4 experiments currently active at LHC: ATLAS, CMS, ALICE, LHCb; each one of them has a specific motivation and task:

- ATLAS (*A Toroidal Large hadron collider ApparatuS*) is a multi-purpose detector. It is composed of a series of sub-detectors which surround the beam pipe: an Inner Detector (ID) for the particle tracking, a solenoid magnet to measure the momentum of the particles, an electromagnetic calorimeter to measure the energy of electromagnetic interactive particles, a hadronic calorimeter which measures the energy of particles which interact by strong interaction, a muon spectrometer to detect muons tracks and momenta;
- CMS (*Compact Muon Solenoid*) is another multi-purpose detector built with different technologies but with the same layout and purposes of ATLAS;
- LHCb (*Large Hadron Collider beauty*) is a specific apparatus for proton-proton collisions. Its purpose is to investigate the physics of the quark b, in particular the CP-violation of the B meson. Unlike the other experiments, it has a fixed target morphology: the apparatus is composed of a tracker around the region of proton interaction followed by a ring imaging Cherenkov detector (RICH), a series of other trackers, another RICH, an electromagnetic calorimeter, a hadronic calorimeter, and at the end a muon detector;
- ALICE (*A Large Ion Collider Experiment*) is an apparatus built to study Pb-Pb collision, where each couple of colliding nucleons reach an energy of 2.76 TeV. It studies concerns mainly QCD theory, in particular studying the condition of high temperature and high energy density. It is composed of 18 detectors surrounding the collision point, including: a time projection chamber (TPC), a transition radiation chamber, a "time of flight" detector, electromagnetic and hadronic calorimeters, and a muon spectrometer.

1.2 The ATLAS detector

The multi-purpose detector ATLAS is 46 m long and with a diameter of 26 m. The goals of this apparatus are the conformation and improvement of the values of the Standard Model (SM) and the study of the beyond SM theories. The experiment involves over 3000 physicists from over 175 institutes.

The general multi-purpose design of the ATLAS detector and its detecting features have put it on forefront in the research for the study of the Standard

Model. For instance, the detection of the Higgs boson ($m_{\mathrm{H}} = 124.98 \pm 0.98\,\mathrm{GeV}$) opened a new chapter in the particles physics research. The study of all its decay channels, in particular the ones involving the b quark, brought and improvement in the knowledge of the characteristics of this boson. Another relevant area of studies at ATLAS regards the top quark: the improvement of the knowledge about it leads to reach the Standard Model limits, finding new decay processes at lower cross sections and hints of new physics. At last but not least, the investigation of new physics, the Super Symmetric Model, is performed at the ATLAS experiment . The latter, at this point, has not been proved yet to be the natural extension of the Standard Model, together with other research like the extra-dimension and many other theories of new physics, that at the moment did not give any evidences.

1.2.1 ATLAS coordinate system

In the ATLAS experiment, the interaction point is considered to be in the coordinate $(0, 0, 0)$ in a 3-D Cartesian reference frame, with coordinates (x, y, z). The coordinate z is along the beam line, and the x-y plane is perpendicular to the beam line with the positive x-axis points to the centre of the LHC ring and the positive y-axis points upward to the sky (see Figure 1.3).

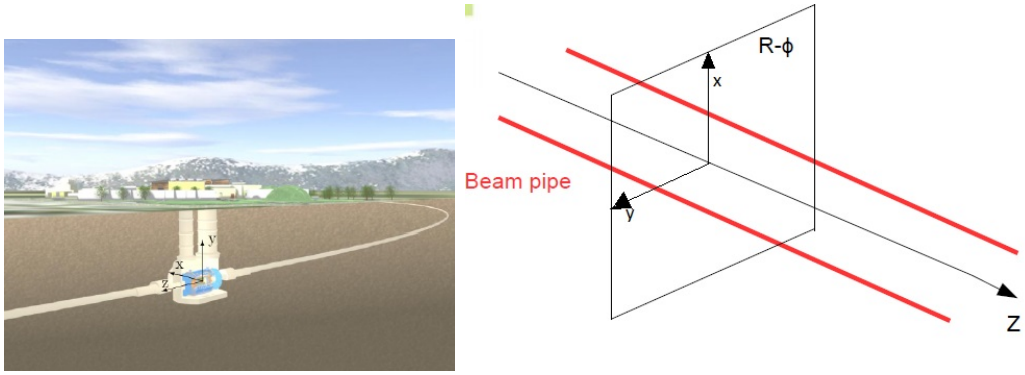

Figure 1.3: *The ATLAS coordinate "real" frame on the left, and the schematics of the frame and the transverse plane on the right.*

In the x-y transverse plane the position coordinates of a particle can be described in polar coordinates: R is the distance from the centre of the plane and ϕ is the azimuthal angle in the transverse plane, from the x-axis around the z-axis . The momentum measured in the transverse plane is called transverse momentum $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$. The polar angle θ is in the z-y plane from the positive z-axis. Using this coordinate representation it is possible to define Lorentz-invariant variables, such as

$$y = \frac{1}{2} \ln \left[\frac{E + p_z}{E - p_z} \right], \tag{1.1}$$

which is the rapidity, a Lorentz-invariant for transformations along the z-axis. For relativistic particles (i.e., for particles travelling close to the speed of light), this variable can be reduced to

$$\eta = -\ln \tan \left(\frac{\theta}{2} \right), \tag{1.2}$$

which is the pseudorapidity, a function of the angular position of the particle, not considering its nature and energy. It is therefore possible to measure the position of a particle in the new Lorentz-invariant coordinate system, composed by (η, ϕ, z), and it is possible to represent a difference in distance, ΔR, between two particle by the formulation

$$\Delta R = \sqrt{(\Delta\eta^2 + \Delta\phi^2)}, \tag{1.3}$$

where $\Delta\eta$ and $\Delta\phi$ are respectively difference in pseudorapidity and azimuthal angle of the particles.

The pseudorapidity ranges from 0, along side the y-axis, to infinity, alongside the z-axis. The high energies involved in the proton collisions makes the partons of the protons collide (each parton carries a fraction of the proton energy and momentum). For the analyses of the collisions many different quantities are used: the transverse momentum p_{T}, the transverse energy E_{T} and the transverse missing energy $E_{\mathrm{T}}^{\mathrm{miss}}$.

1.3 Detector composition

ATLAS apparatus is composed by different detectors, where each of them covers a η range and has a specific purpose. In Figure 1.4 is shown the ATLAS layout.

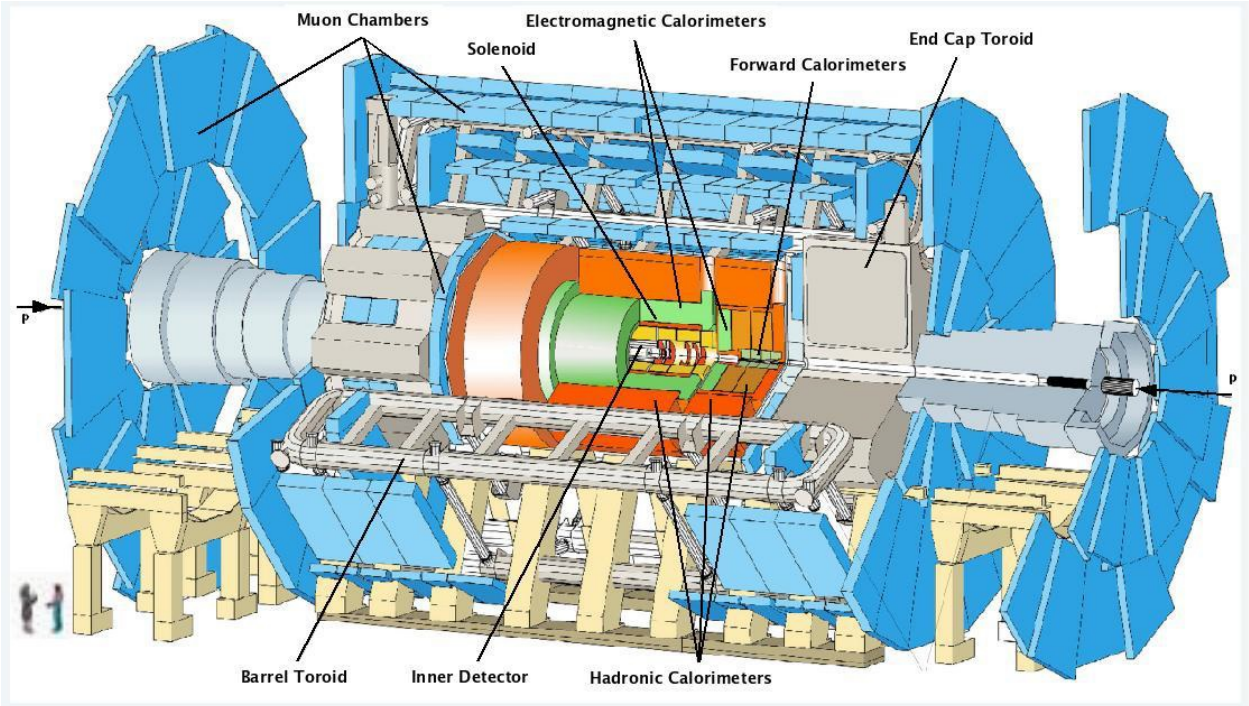

Figure 1.4: *The ATLAS experiment layout section, with all its components. The system is built with cylindrical symmetry with the beam pipe as the axis. From the innermost (closer to the beam pipe) to the outermost, all the sub-detectors of ATLAS are shown. Figure from [10].*

Immediately after the proton-proton collision, the first detector which the produced particles traverse is the Inner Detector (ID), a tracking apparatus formed

by 3 detectors surrounding the beam line and covering the region $|\eta| < 2.5$. Here, the particle tracking precision is very high, with an intrinsic accuracy varying approximately from 10 to 100 µm. Right after the ID, there is the first stage of the Magnet System, a central solenoid which provides a 2 T magnetic field and permits to perform a momentum measurements of p_{T} with a resolution of $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = (4.83 \pm 0.16) \times 10^{-4}\,\mathrm{GeV}^{-1} \times p_{\mathrm{T}}$.

In this detector section, only charged particles can be detected. After this stage, the following detector is the electromagnetic calorimeter, where electrons, positrons and photons are detected. Through electromagnetic interaction, electromagnetic showers are produced for each mentioned particle, inside of which other particles are produced and thereafter detected. The energy and the track of the particles in the shower can be measured. Furthermore, the more energetic particles survived from the previous steps encounter the hadronic calorimeter. Thanks to the strong interaction, the hadrons, formed in the proton-proton collisions or from secondary decays, develop hadronic showers. They are detected with an energy resolution which ranges from 0.13 to 0.06 when jet transverse momentum p_{T} increases. The calorimeter stage covers an angle up to $|\eta| = 4.9$.

Eventually, only the particles with a very low cross section survived, mainly muons and neutrinos. The latter can not be detected directly by ATLAS, thus they are studied with the missing energy technique. The muons instead can be detected in the muon spectrometer, a very large detector, composed of 2 tracking chambers and 2 trigger chambers. It has a coverage $|\eta| < 2.7$.

1.3.1 Inner Detector

The Inner Detector is a 6.2 m long apparatus with a diameter of 2.1 m, placed around the beam line, with a coverage of $|\eta| < 2.5$. Built for the early tracking stage of ATLAS, it is composed the Pixel Detector, the Semiconductor Tracking and the Transition Radiation Tracker. The technologies and components of all these detectors must be the most radiation hard possible, since they are the closest to the beam line.

The Pixel Detector (PD), is the second tracker encountered by a particle produced in the proton-proton interaction. It is a silicon based detector which uses the pixels technology; it has the highest granularity in all ATLAS and it consists of 4 barrels: Insertable B-Layer (see later), B-Layer, Layer1, Layer 2 and 3 disks for each side;

The Semi-Conductor Tracker (SCT) is 4 layers a Silicon microstrip detector. Each layer is formed by modules composed by two microstrip detector bounded together and glued with a 40 mrad angle of their planes, layout used to obtain a better z-measurement. The two microstrip detectors of a single module are glued with the angle between the two microstrip shifted of 90°. In the barrel region the plane of the microstrip detector is parallel to the beam line, while in the end-cap region is perpendicular.

The Transition Radiation Tracker (TRT) is the largest track detector of ID and surrounds the previous two. It consists of a large number, about $5{\times}10^4$, of straw tubes, that are cylindrical tubes with one positive wire in their inside and the internal wall at negative voltage. The straws all together contribute to the

measurement of the particle momentum thanks to the high number of hits, and each one is filled with a mixture of Xenon (70 %), CO_2(27 %), and O_2 (3 %). In the barrel region the tubes are parallel to the beam line, while in the end-cap region are perpendicular.

All of them are placed both in the barrel and in the end-cap region of ID. The Figure 1.5 shows the ID configuration, dimension and coverage, while the general characteristics of each sub-component are described in the Table 1.1.

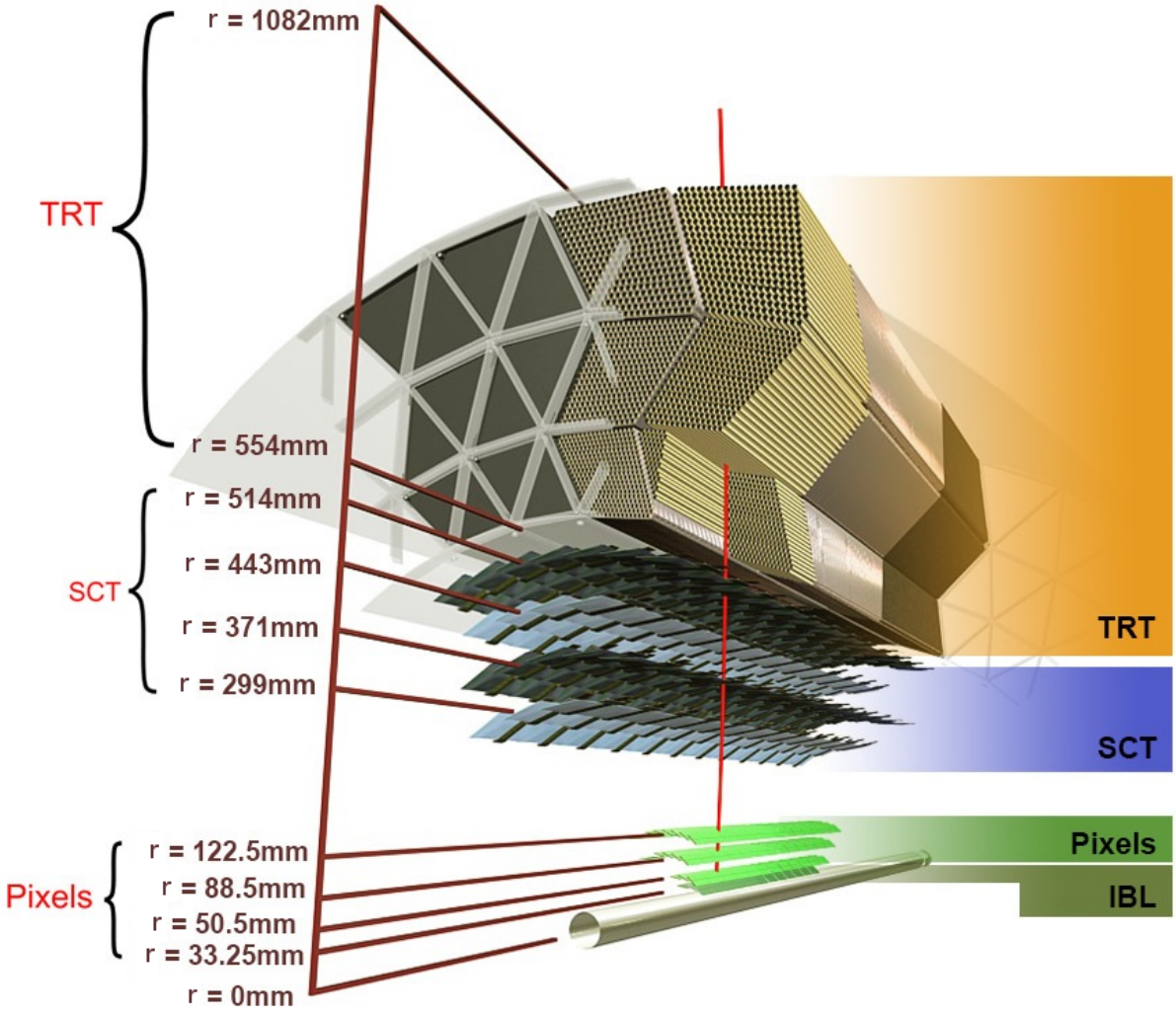

Figure 1.5: *Section of the ID subdivided detectors system, with the relative distances and distributions from the beam line. From the closest one to the further one the subsystems are: the Pixel Detector, the Semi-Conductor Tracker and the Transition Radiation Tracker. Figure from [10]*

All these detector's task is to provide a high precision measurement of the particles track, which may by reconstructed by two different algorithms:

- the inside-out algorithm, where primary tracks of charged particles born from primary interactions are reconstructed using 3 seeds in the silicon detectors (SCT and PD) and subsequently the following hits are added by a combinatorial Kalman-filter algorithm;

Table 1.1: *Main characteristics of the ID's detector.*

Detector	Hits Tracks	Element Size	Hits Resolution (µm)
PD, $\|\eta\| < 2.5$			
4 barrel layers	3	$50 \times 400\,\mu m^2$	10(R-ϕ)-115(z)
3×2 lateral disks	3	$50 \times 400\,\mu m^2$	10(R-ϕ)-115(R)
SCT, $\|\eta\| < 2.5$			
4 barrel layers	8	50 µm	17(R-ϕ)-580(z)
9×2 end-cap disks	8	50 µm	17(R-ϕ)-580(z)
TRD, $\|\eta\| < 2.0$			
73 barrel tubes	30	$d = 4$ mm, $l = 144$ mm	130/straw
9×2 end-cap disks	30	$d = 4$ mm, $l = 37$ cm	130/straw

- the outside-in algorithm, where the hits in TRD of secondary charged particles, formed by decays from primary particles or other secondary particles, are used to reconstruct the tracks adding the Silicon hits always with the combinatorial Kalman-fitter algorithm, if there are. The efficiency of track reconstruction is measured by simulated events, and it varies as a function of p_{T} and η.

Insertable B-Layer

In 2015, it was added a new innermost tracking detector, the Insertable B-Layer, in order to improve the ATLAS tracking performances during the Phase-I LHC operation. The Insertable B-Layer (IBL) is the first detector of the ATLAS setup and the latest upgrade of the pixel detector. The detector is based on 2 different technologies of the silicon sensors:

- the planar sensors, with a "typical" silicon layout, similar to the ones used in the B-layer, but slightly different. In fact, the inactive border passed from the previous 1mm to 450 µm, and from studies on the B-layer, the irradiated sensors increase the collected charge if the thickness is reduced;

- the 3D sensors, with a different geometrical arrangement of the silicon junctions. The signal is collected simultaneously by two different electrodes, since the number of produced charges is lower. Moreover, the active surface of these sensors extends much more, reducing the not sensible volume, and they are closer to each other in the arrangement of the detector, which means a reduction of the applied voltage and the leakage current contribution.

Thus, the new technology of IBL makes the detector itself more radiation hard and with a higher surface coverage, with the usage of the FE-I4 chip.

1.3.2 Calorimeter

The ATLAS calorimeter system is composed by sampling calorimeters whose scope is to perform the energy reconstruction of a crossing particle. The entire

sub-detector is divided into multiple parts, each of which dedicated to a different type of particle. Each calorimeter consists of four parts, to which correspond a region of coverage: a barrel part, an extended barrel part, a end-cap part and a forward part. The whole system covers a pseudorapidity up to η=4.9 and a complete ϕ coverage. The ATLAS calorimeter setup is presented in Figure 1.6.

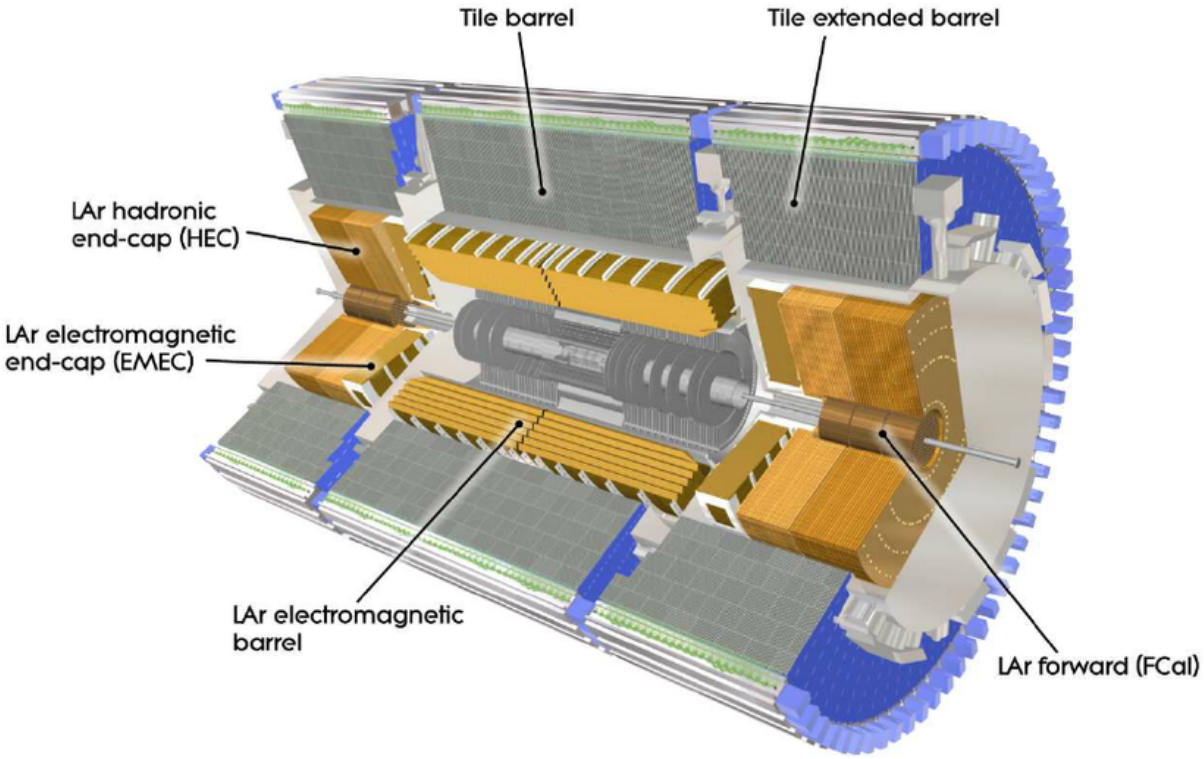

Figure 1.6: *The ATLAS calorimeters system setup. It consists of a Liquid Argon (LAr) electromagnetic calorimeter and an Hadronic Calorimeter. Interactions in the absorbers transform the energy into a "shower" of particles that are detected by the sensing elements. Figure from [10].*

The whole calorimetric system consists of :

- a Liquid Argon (LAr) electromagnetic calorimeter, focused in the measure of electron, positrons and photons in the pseudorapidity range $|\eta| < 3.2$, with a high granularity. It is divided into a barrel and two end-cap. These calorimeter are based on a lead-LAr detector technology, the lead is a good absorber, while the liquid argon is a good active medium, since its radiation hardness and its good energy resolution;

- a Hadronic Calorimeter (HCAL), focused in the energy measurement of jets originated from hadronic processes, as well as the determination of the missing transverse momentum. It is formed by the Hadronic Tile Calorimeters (HTC), a scintillator-tile calorimeter, by the Hadronic End-Caps Calorimeters (HEC) and the Forward Calorimeter (FCAL) which are both LAr calorimeters.

In terms of energy, the resolution of a sampling calorimeter can be written as

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{1.4}$$

where the term $\oplus$ refers to the sum in quadrature of the terms. Also, a is the stochastic term and comes from intrinsic fluctuations of the development of the shower due to their statistical behaviour; b is the term due to the electronic noise of the read-out channels; c is a constant that considers the temperature, the ageing of the detector, the radiation damage, and other factors.

The energy resolutions are different depending on the calorimeter composition: $\frac{\sigma_E}{E} = \frac{10\%}{E} + (1.2 \pm 0.1^{+0.5}_{-0.6}\%)$ for the electromagnetic calorimeter in the barrel region, from 0.13 to 0.06 (when the transverse momentum increase) for the hadronic one in the barrel and in the end-cap region, $\frac{\sigma_E}{E} = \frac{10\%}{E} + (2.5 \pm 0.4^{+1.0}_{-1.5}\%)$ for the forward electromagnetic calorimeter.

1.3.3 Muon Spectrometer

In Figure 1.7 is shown the scheme of the ATLAS Muon Spectrometer. High p_{T} muons provide signatures for many physics processes studied in ATLAS and that is the reason why the muon spectrometer has a key role in particle identification. It is designed in order to reach high precision and resolution in the measurement of high p_{T} muons and it also provides an independent muon trigger from the rest of the detector.

The measurement is based on the magnetic deflection of muon tracks in the large superconducting air-core toroid magnets (the magnetic system is presented in Section 1.3.4). Over the range $|\eta|<1.4$, magnetic bending is provided by the large barrel toroid, while for $1.6<|\eta|<2.7$,muon tracks are bent by two smaller end-cap magnets inserted into both endsof the barrel toroid. Over $1.4<|\eta|<1.6$ (transition region) a combination of barrel and end-cap fields provides magnetic deflection.

The detector is divided in barrel and end-cap region, in which are placed toroid magnets and the system is divided in two different groups of sub-detectors, composed by 4 different detector technologies: 2 of them are the Precision Chambers (Monitored Drift Tubes and Cathode Strip Chambers), which are focused on precision measurement of muon momentum and the other 2 are the Trigger Chambers (Thin Gap Chambers (TGC) and Resistive Plate Chambers (RPC)), which provide online trigger.

The Monitored Drift Tube (MDT) chambers are drift chambers of two multilayer drift tubes which are focused on precise measurement of the z coordinate in the barrel region, in the region $|\eta|<2$. By measuring the drift time in single tubes it is possible to recontruct the hit position of the particle. provide the measurement of momentum and track in the barrel and end-cap regions and

The Cathode Strip Chambers (CSC), are multi-wire chambers with strip cathodes for the measurement of muon momenta in the pseodurapidity region $1.0<|\eta|<2.7$. The CSC wires are composed of parallel anodes which are perpendicular to 1 mm large strips of opposite polarity. They are positioned close to the beam pipe in the innermost layer of the end-cap.

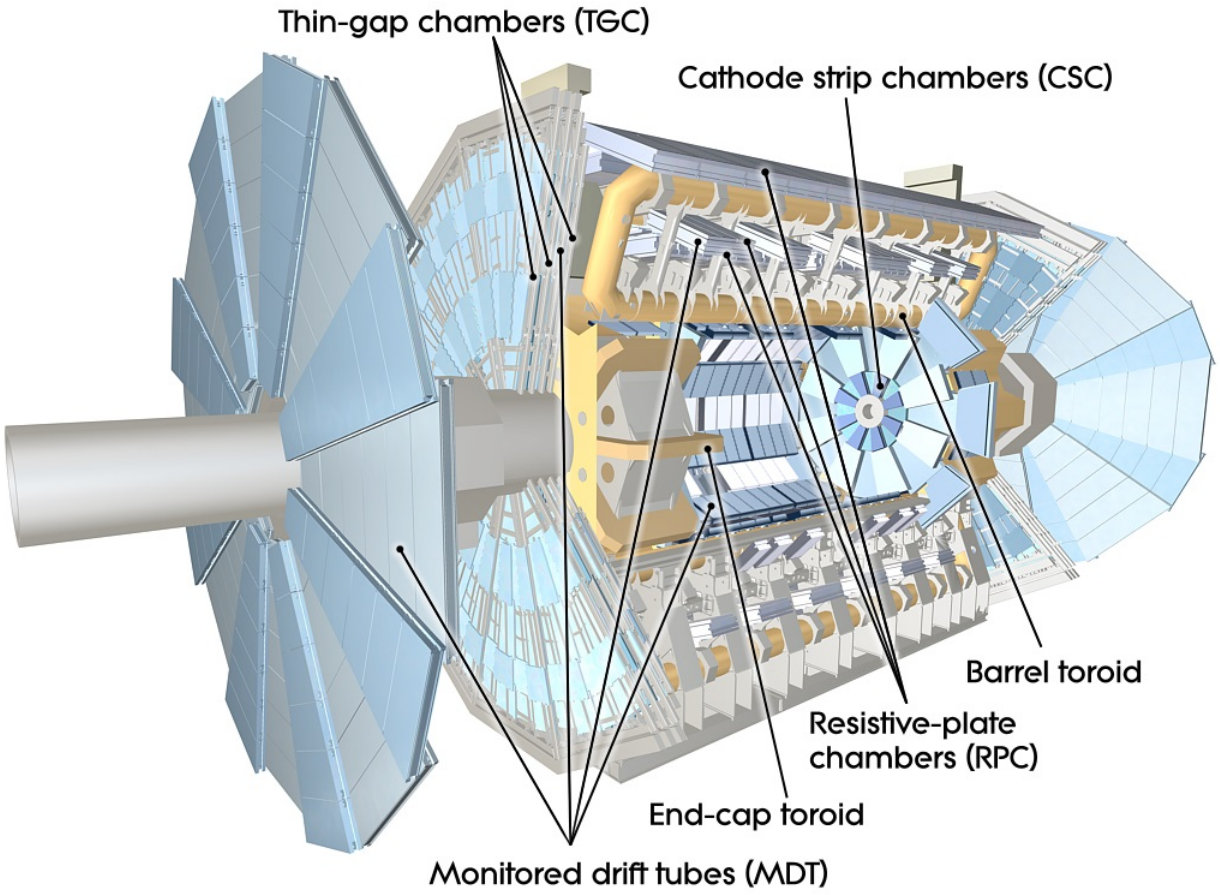

Figure 1.7: *The ATLAS muon spectrometer morphology. The system is divided in two different groups of sub-detectors, the Precision Chambers (Monitored Drift Tubes and Cathode Strip Chambers) for muon moment and Trigger Chambers (Thin Gap Chambers, Resistive Plate Chambers) for online data-acquisition triggering. Figure from [10].*

Thin Gap Chambers (TGC) in the end-cap region is a very thin multiwire chambers. The anode-cathode spacing is smaller than the anode-anode spacing, leading to a very short drift time ($< 20, \text{ns}$). The spatial resolution of these detectors is 4 mm in the radial direction and 5 mm in the ϕ coordinate. The TGC are also used to improve the measurements along the ϕ coordinate obtained from the precision chambers.

The Resistive Plate Chambers (RPC) in the barrel region are gaseous parallel electrode-plate detectors, with a spatial resolution of 1 mm in two coordinates and an excellent time resolution of $1, \text{ns}$. This sub-detector works in the avalanche regime: when a charged particle passes inside the chamber, the primary ionization electrons are multiplied into avalanches by a high electric field,

This whole system identifies muons for which $|\eta < 2.7|$ with a threshold of $p_{\text{T}} > 3\,\text{GeV}/c$, since muons with lower energy are completely absorbed (they lose

their whole energy) before reaching the muon spectrometer. The measurements resolution of p_{T} is about 20% at 1 TeV.

1.3.4 Magnetic System

ATLAS uses a system of superconducting magnets (shown in Figure 1.8) for the measurement of the charged particles momenta.

The system is composed by a Central Solenoid (CS) surrounding the Inner Detector, and by a system of 3 large air-core toroids (1 in barrel and 2 in end-cap) generating the magnetic field of the muon spectrometer, with a dimension of 26 m in length and 20 m in diameter. The CS, for the momentum measurement of ID, has a magnetic field of 2 T and it points in the positive z-axis direction, while the toroids magnet emits a magnetic field of 3.9 T (barrel) and 4.1 T (end-cap). The entire system work at 4.7 K of temperature. The most important parameters for momentum measurements are the field integrals over the track length inside the tracking volume:

$$I_1 = \frac{0.3}{p_{\mathrm{T}}} \int_0^l B \sin(\theta)_{(\mathrm{d}\vec{\mathbf{l}},\vec{\mathbf{B}})}\, \mathrm{d}l$$

and

$$I_2 = \frac{0.3}{p_{\mathrm{T}}} \int_0^{l\sin(\theta)} \int_0^{\frac{r}{\sin(\theta)}} B \sin(\theta)_{(\mathrm{d}\vec{\mathbf{l}},\vec{\mathbf{B}})}\, \mathrm{d}l\, \mathrm{d}r$$

where I_1 is the measurement of bending power field (p_{T}=q ×bending power=q $\times(B \times L)$) and I_2 represents the total transverse deflection of the particle from its initial path. θ is the longitudinal component of the angle between the track and the magnetic field and the integrals are calculated on the azimuthal direction of the particle ($l = r/\sin(\theta)$) and on its radial trajectory.

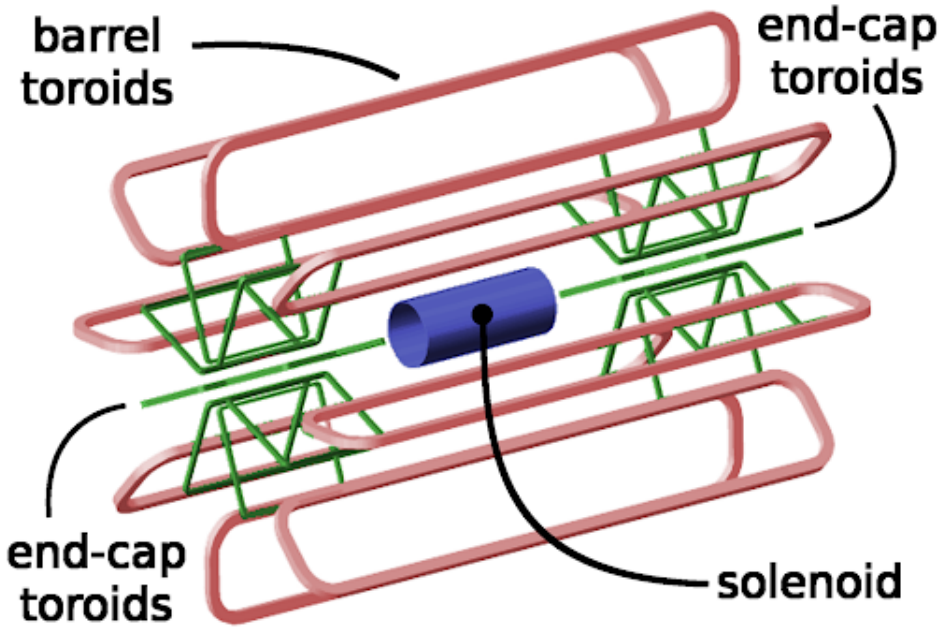

Figure 1.8: *The ATLAS Magnetic system schematics.*

1.3.5 Trigger and Data Acquisition system

Thanks to the high luminosity reached in the LHC ring, in 1 second are produced an order of magnitude of 10^8 proton-proton processes. Despite that, the read-out electronics can reach at maximum a rate for recording speed of 300 Hz and for this reason the ATLAS experiment has a system of multi-level trigger, composed by a series of three different trigger levels: Level-1 (L1), Level-2 (L2) and Event-Filter (EF). In Figure 1.9 a scheme of the current Trigger and Data Acquisition (TDAQ) system of ATLAS is presented.

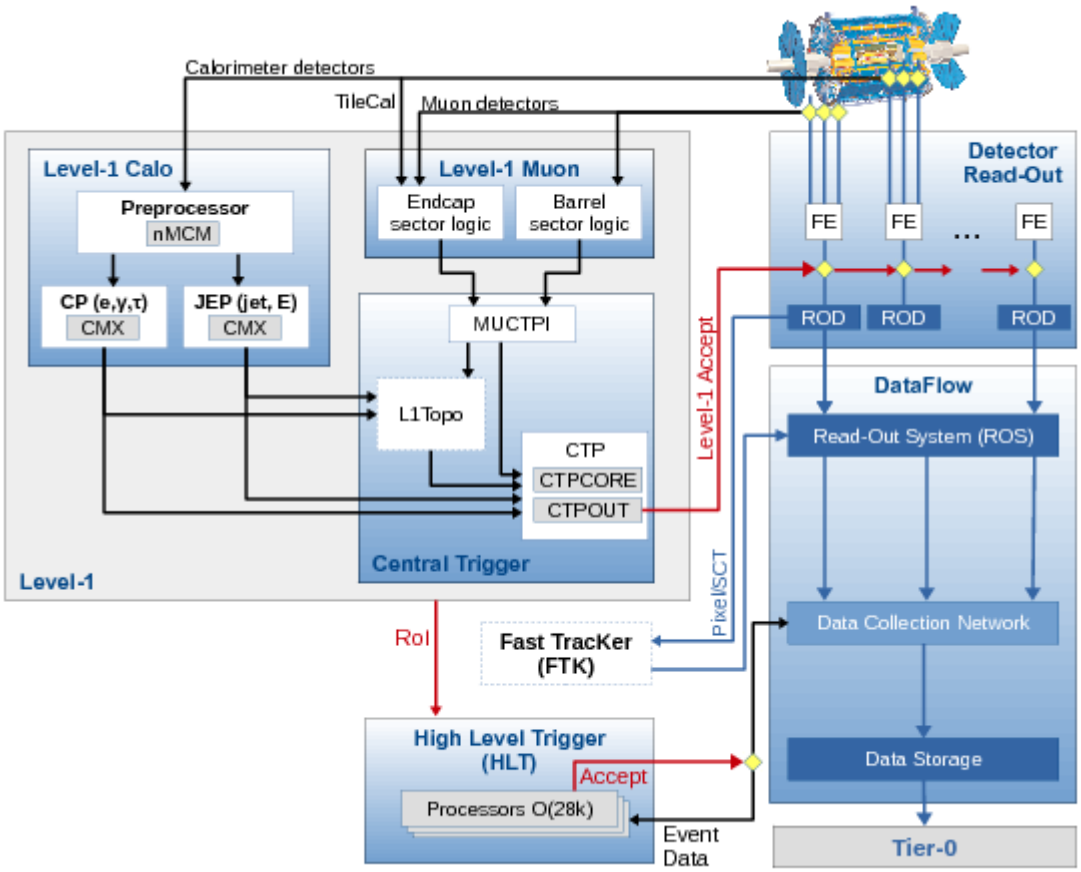

Figure 1.9: *The ATLAS current Trigger and Data Acquisition (TDAQ) system. The Level-1 trigger uses information from multiple subsystems to quickly identify events of interest, while the more refined High-Level Trigger exploits informations from all the subsystems. Accepted data is then sent to the Read-Out system to be stored. Figure from [10].*

The L1 trigger is a hardware-based trigger which splits data read by high p_{T} leptons, photons, jets, large missing transverse momentum and total transverse energy. It reduces the rate of the dataflow down to approximately 50 kHz, with a decision time for each collision of 2 µs from the collision itself. The data for the trigger comes from the calorimeters and from the Muon Spectrometer (MS), in particular from the RPC and the TGC chambers. The L1 trigger defines the regions in η and ϕ coordinates where the other subsequent triggers will have to start their

own works, so called Regions of Interest (RoI).

Furthermore, L1 muon trigger searches for coincidences of hits in different trigger stations within a road pointing to the interaction point, because the width of this road is correlated with the transverse momentum. There are six muon p_{T} threshold ruled by the hardware-programmable coincidence logic for this part of the trigger, three for the 6-9 GeV (low p_{T}) and three for the 9-35 GeV (high p_{T}).

The L2 trigger instead is a software based trigger which starts from RoIs defined in L1 and uses all the detector informations in these regions in its trigger algorithms. It allows to reach less then 5 kHz in less then 50 ms. Eventually, the Event Filter is the final stage of the trigger chain and it reaches the rate of approximately 30 Hz in less then 4 s. This time does not come by algorithms, instead it is the standard time of the off-line event reconstruction of ATLAS TDAQ.

A trigger menu is present, inside of which a list of characteristics (of an event like $E^{\mathrm{T}}_{\mathrm{miss}}$ and others) with a certain threshold, given by the luminosity, for each one. The whole set of events which passes the selection criteria of this menu are tagged and sorted into data streams. Additionally with the data streams are also present the streams used for the calibration data of the detector. The set of the two subsystem of the Trigger Level-2 and Event Filter trigger forms a High Level Trigger.

Chapter 2

ATLAS Phase-II upgrade

2.1 The Phase-II upgrade

The past decade of LHC running has been hugely successful. Even though the technological and experimental development challenge, major results were reached during the years of operation, such as the discovery of the Higgs boson or the high precision measurements of the physics at the electroweak scale. This led to a stronger confidence in the potential of the LHC ring, so that it emerged the idea to push the sensitivity and the precision of the detectors well beyond the the original proposals.

In fact, the ATLAS experiment as all the other LHC experiments, will need an upgrade, since it is expected a new operational phase of working: the High Luminosity upgrade of the Large Hadron Collider (HL-LHC). It was expected to be active for the operations in the second half of 2026, but because of some delays the timetable has been slightly postponed. The plan for the future years of the LHC ring are shown in Figure 2.1.

The ATLAS collaboration firstly gave a description of the initial plan for the Phase-II upgrade of the detector in the Letter of Intent (LoI) in 2012. Since then, the collaboration has been improving and refining the initial proposals, giving possible evolutions of the initial proposal.

In the first proposals of the HL-LHC, the nominal levelled instantaneous luminosity should have reached a value of $\mathcal{L} \simeq 5 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, corresponding to an average of roughly $\mu \simeq 140$ inelastic proton-proton collisions per each bunch-crossing. But in later times, a new scenario with a ultimate levelled luminosity was introduced, with a peak up to $\mathcal{L} \simeq 7.5 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, corresponding to $\mu \simeq 200$, delivering and integrated luminosity of around $300\,\mathrm{fb}^{-1}$ per year of operation. The latter, explores the technical limits of the LHC capabilities, leading to the study of performances through simulations to actually ensure that the Phase-II detector upgrade is able of taking advantage of the ultimate luminosity.

Since from the initial concepts the upgrade proposal has changed, the possible development of the upgrade will depend on the actual maximum luminosity reached and on the mean number of interactions per bunch-crossing. Thus, most of the evaluation of the performances have been done for the highest mean number of interaction, but some of them also at the lower value of $\mu \simeq 140$. Therefore, the

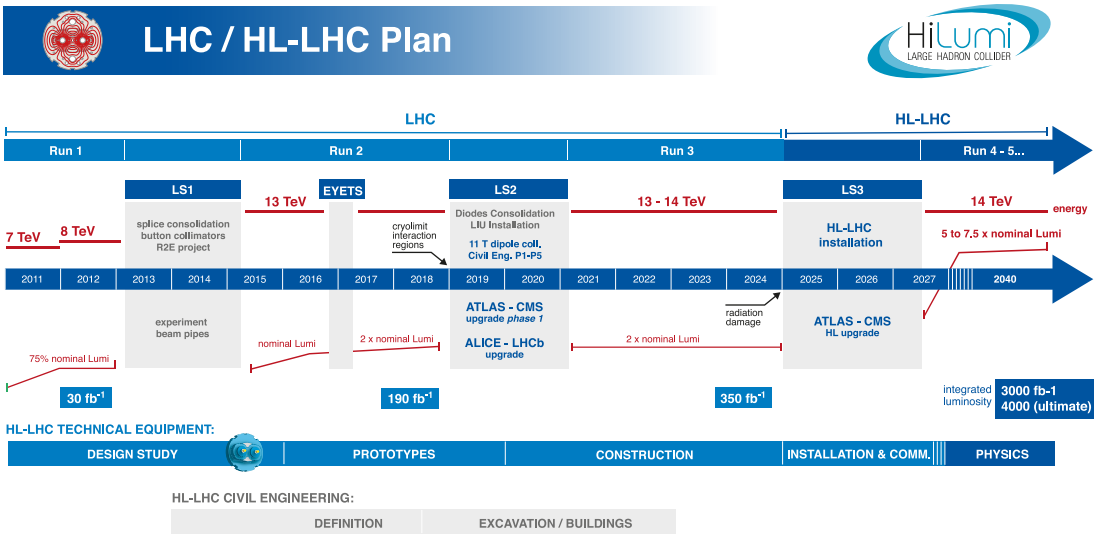

Figure 2.1: *Timetable plan for the future years in terms of different phases, respectively long shutdown phases (e.g. LS1) and run phases (e.g. Run 2). The proposals and the studies for the this upgrade started during the LS2, but it will be needed time to test and to decide which proposal will be implemented. In fact, the Phase-II upgrade will come not before 2027.*

possible detector configurations are currently divided into 3 different scenarios (per cost of implementation): the Reference, the Middle and the Low scenario. Of course they include different configurations for the upgrade for each of the major systems in the ATLAS experiment.

Eventually, the design and techniques used for the Phase-II upgrades represent an evolution from the new designs and technologies already introduced during the LS1 improvement program, which included the installation of the IBL pixel detector, and the Phase-I upgrades now being prepared for installation during LS2. The evolved approach leads to a relatively solid understanding of the technical challenges, even in the early stages of the upgrades activities.

2.2 Upgrade proposals

For the ATLAS experiment, the focus of interest is posed in the upgrade of the Trigger and Data Acquisition (TDAQ) and on the Inner Tracker (ITk) systems (which are also the more expensive components in the whole detector's system), the Calorimeters (both the Liquid Argon and the Tile one) and the Muon Spectrometer (MS). We will see that in some cases some of the sub-detectors are completely replaced with more modern ones, in other cases instead, particularly for the most new sub-detectors and for the ones which can sustain the high pileup, only the electronics for the readout are replaced.

In general, the focus in the following sections is posed in the whole possible elements and the possible scope of the ATLAS upgrades, giving only a statement of

the major changes with respect to the previous experiment set-up. More detailed information is in general not reported if not needed or postponed to further chapters.

2.2.1 Inner tracker

The Inner Tracker (ITk) for the Phase-II upgrade of ATLAS is an all-silicon tracker detector. Designed to measure the transverse momentum and direction of isolated particles (particularly muons and electrons), it allows to reconstruct the vertices of pile-up events and associate the vertex with the hard interaction. It is also able to reconstruct and identify secondary vertices in b-jets with a high efficiency and purity and also the decay of τ leptons, including impact parameter information.

The basis of the upgrade were posed in the initial proposals, and the current designs for the three possible upgrade scenarios are an ultimate evolution of the starting ideas. For instance, a major change in the ITk upgrade programme has been developed. It regards the improved understanding of the importance for the HL-LHC physics programme of the modification of the tracking coverage well beyond $|\eta| = 2.7$, reaching a maximum coverage of $|\eta| = 4$, matching with the improvements of all the other detectors.

In the Reference scenario proposal, the tracking capability of the tracker provides a full coverage up to $|\eta| = 4.0$. Its layout is shown in details in Figure 2.2. In the barrel region there are four pixel layers followed by five strip layers: the layout has been optimised for the coverage, with the preservation of small gaps between sub-detectors to allow supports, services and eventual insertions. Meanwhile, in the end-cap region there are 12 pixels discs on each side if the detector, to keep the number of space points on a track approximately constant in the prefixed range. This configuration is currently the ideal one, even though it still has to be optimized in terms of mechanical construction and maximum performance for a given silicon area.

In general, the new tracker has been designed to balance the tracking performance required for the upgrade programme against the cost of construction. Even though the coverage maximise the physics potential of detection, the extended regions coverage poses a challenging implementation design. The test studies make use of two different technologies, thus the division into two sub-parts of the sub-detector: silicon pixels modules used in the inner radii, surrounded by silicon micro-strip modules used at larger radii. For all the possible scenarios, the whole active length is 6 m long, including the space which is currently occupied by the straw tubes of the Transition Radiation Tubes.

2.2.2 Calorimeters

In the last decades of the LHC activity, the technologies used in the ATLAS Calorimeter system, designed to allow the measurement of the development of electromagnetic and hadronic showers, were two:

- The Liquid Argon (LAr) calorimeter technology, used for the electromagnetic calorimetry, in the barrel, in the end-cap and in the forward regions;

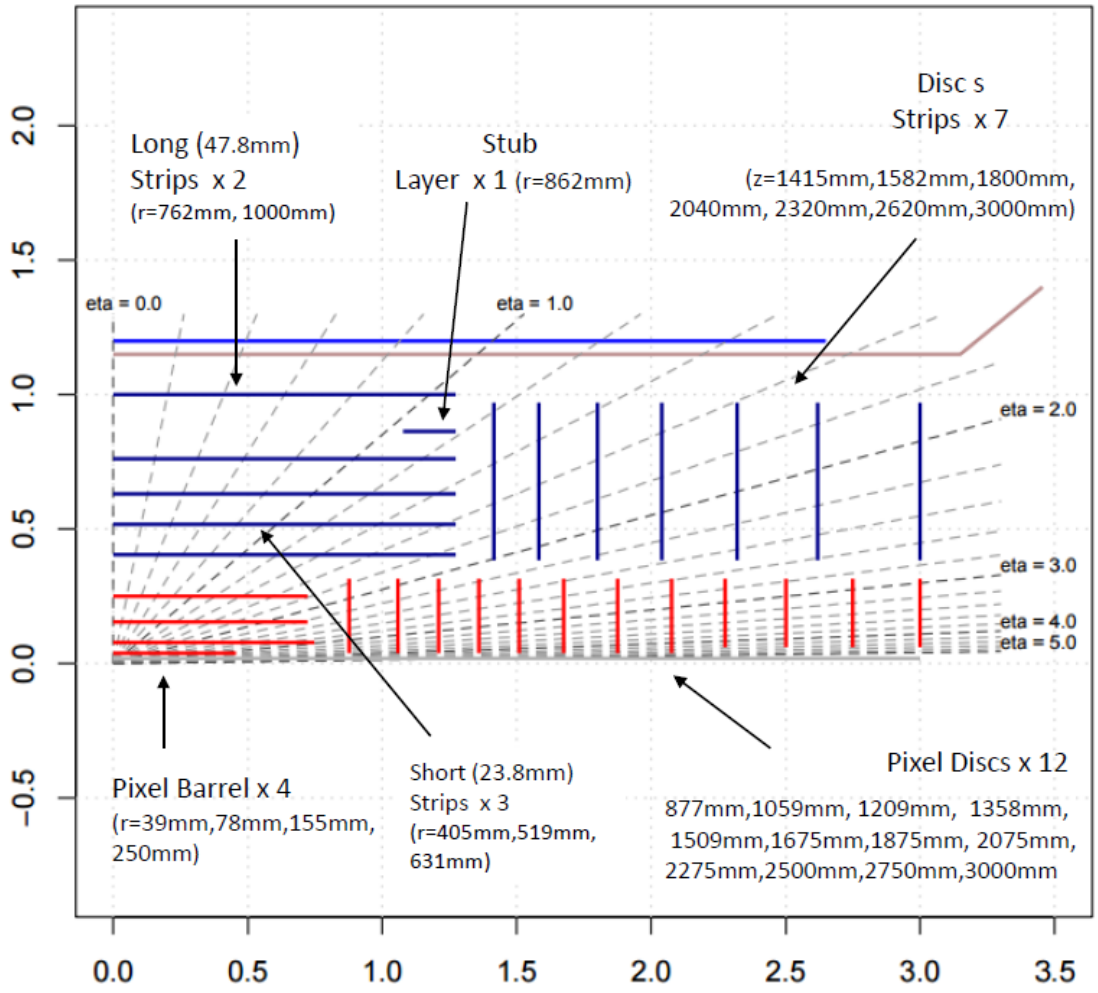

Figure 2.2: *Cross-section view of the ITk layout for the Reference scenario. It shows the coverage of the pixel detector in red and of the strip detector in blue. The pseudo-rapidity coverage extends up to $|\eta| = 4.0$. The blue and red lines represent respectively the strip and the pixel layers. Horizontal lines stands for barrel layers and vertical ones for end-cap layers. The blue line on the top of the view represents the coil of the solenoid magnet. Figure taken from [4].*

- The Scintillating Tile hadronic barrel (TileCal) and the Hadronic End-cap Calorimeter (HEC) made use of the scintillating technology for the energy measurement of hadronic showers developed from strong interacting particles.

Most of the components of the ATLAS Calorimeter system, in which are included the LAr of the barrel region, the TileCal and the HEC, keep their required performances during the HL-LHC conditions and they do not need any replacement. The only exception is for the LAr Forward Calorimeters (FCals), which will be degraded by the high energy and particle density of the HL-LHC upgrade and thus it will be replaced with a new forward detector with a higher granularity. Moreover, with the addition of a new High Granularity Timing Detector (HGTD), to be installed in front of the the LAr Calorimeter end-caps, it will be possible to lessen the pile-up

effects in the forward and end-cap regions.

What is going to be upgraded for sure for both the systems is the readout electronics, because the current one has a limited radiation tolerance, and for the upgrade the trigger will need higher performances in terms of rate and latencies. The LAr and the Tile electronics upgrade includes a replacement of the on-detector front-end electronics currently in use with a more radiation tolerant one and of the off-detector signal processing units. The upgrade provides a more sophisticated detector signal processing optimised for high pile-up conditions.

The increased event pile-up represent a challenge for precision calorimetry. As anticipated, in order to stand this increase it is replaced the current forward LAr detector with a higher granular one and it is added a new high time precision detector, the HGTD. The latter will be placed in the range $2.4 < |\eta| < 4.3$. For this region the segmentation of the electromagnetic calorimeter is thus significantly reduced with respect to the central region ($|\eta| < 2.5$) and the detector performance will be degraded because of the large pile-up. Its aim is to measure the arrival time of the charged particles in order to assign them to different collision vertices. This will provide and additional suppression in the pile-up.

The expected improvements in calorimeter performances are relevant particularly for physics events with jet formation in the end-cap and forward regions, and for physics signatures with missing transverse energy.

2.2.3 Muon spectrometers

The HL-LHC upgrade requires a substantial improvement in the performances of the muon spectrometer system and it must be true for precision tracking as well as the triggering system. In fact, the main scope of the muon spectrometer upgrade is the improvement of the performances of its trigger. Currently, the muon spectrometer provide a L1 hardware muon trigger based on hit coincidences within different detector layers. The high-level trigger performs a software confirmation of the L1 muon trigger, by using refined p_{T} measurements from the precision chambers. After the Phase-II upgrade:

- The muon spectrometer provides a finer granularity trigger based on Monitored Drift Tube (MDT), in order to improve the sharpness of the transverse momentum threshold at L1 or even at L0 if allowed by the L0 latency;
- The coverage and the redundancy of the L0 are increased, since new Resistive Plate Chamber (RPC) detectors in the barrel region ($|\eta| < 1$) are added;
- The original on-chambers electronics for the MDT may be partially or completely replaced.

The final configuration of the Muon spectrometer after the upgrade is shown in Figure 2.3. In all the three possible scenarios, the trigger electronics is upgraded, both in the barrel and the end-cap spectrometers. Moreover, a replacement of the trigger chambers in the forward region ($2.0 < |\eta| < 2.4$) is foreseen in all the scenarios, to largely improve the trigger selectivity.

Focusing only on the Reference scenario, the muon acceptance will be increased thanks to the addition of a very forward muon tagger, attached to the New Small

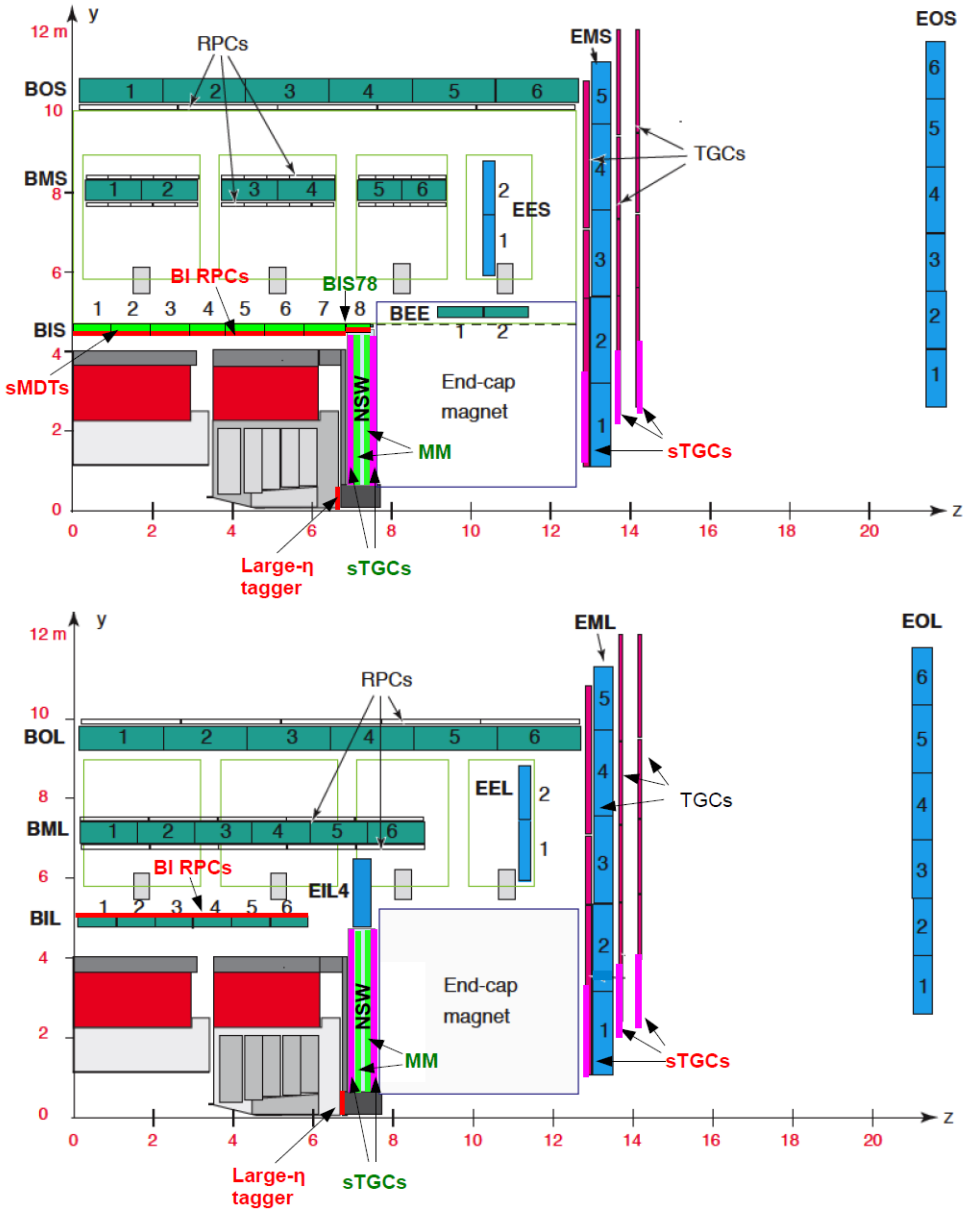

Figure 2.3: *Schematics of the ATLAS Muon Spectrometer with the new chambers proposed for installation in Phase-II upgrade (in red text), those to be installed during LS2 (in green text), and those that will be kept unchanged from the Run 1 configuration (in black text). The upper panel shows the R-Z view of the azimuthal sectors that contain the barrel toroid coils, while the lower panel shows a sector between the barrel toroid coils. Figure from [6].*

Wheel (NSW) shielding disk and covering the region of $2.6 < |\eta| < 4.0$. Furthermore, if we consider the barrel and end-cap detectors, it is possible to summarize the main changes in this scenario:

- In the barrel region, RPC and a small tube diameter MDT (sMDT) are installed in the small sectors of the Barrel Inner layer (BI), and the already existing on-detector electronics is entirely replaced.

- In the end-cap region, the upgrade plans to include the replacement of the MDT front-end readout and of the L0 trigger electronics of all the chambers, with the exclusion of the NSW, and also the installation of a very forward tagger to take advantage of the very-forward ITk tracking (which in the Reference scenario has the largest extension).

The main reasons for the electronics upgrade comes from the new ATLAS requirements of the L0/L1 trigger system, that are automatically applied to the readout electronics. The present MDT read-out electronics cannot cope with the expected trigger rate, thus the new trigger system will have to maintain a high level of efficiency so to find high-p_{T} tracks and keep the rate of fake triggers low. The present read-out electronics are not able to sustain the rate of 400 kHz, it must be replaced with an improved one.

Eventually, another important reason for the upgrade of the trigger electronics is the need for an improved selectivity for high p_{T} tracks, which calls for a better space resolution in the bending direction (η). The current read-out electronics of the trigger chamber systems does not supply information on the signals pulse height, therefore charge interpolation cannot be used to perform position measurement.

2.2.4 Trigger and Data Acquisition

The Phase-II upgrade of the ATLAS TDAQ system must satisfy the ATLAS physics programme planned for the operation decade of the HL-LHC. During the latter period, the expected amount of data to be collected is of the order of 4000fb^{-1}, and it will allow a deeper exploration of the electroweak symmetry breaking though the research of Higgs boson properties. Thus, both the necessity of a high efficient selection of Higgs events and of new physics events requires exceptional performances from the trigger and data acquisition system. Also, the need for a higher maximum rate and a longer latency of the trigger comes from the larger step increase in the luminosity.

As for most of the components of the ATLAS detector, the upgrade of the TDAQ system can be seen as an evolution from the older designs and technologies introduced during the LS1 programme, which are currently being prepared for the installation during the LS2. In the Reference scenario the front-end electronics of all the existing ATLAS detector system is replaced, with the exception of the systems upgraded during the LS1 for the Phase-I upgrade.

For the HL-LHC upgrade of the TDAQ system the adopted strategy forecast the development of three possible options: a Baseline Level-0 only design, a single level trigger, but with the possible addition of a L1track level, in the evolved design or the alternative design. The information is sent to a sophisticated FPGA technology to implement fast and powerful trigger algorithms (which will be treated with more details in the following chapter), which is a different technology with respect to the ASIC-basic systems used during the Run 1 operations.

A general architecture layout is given in Figure 2.4, where it can be distinguished three different sub-systems which will form the ATLAS TDAQ: the Level-0 Trigger, the DAQ (Readout and Dataflow subsystems) and the Event Filter. A deeper and more detailed description of the whole architecture development and also another

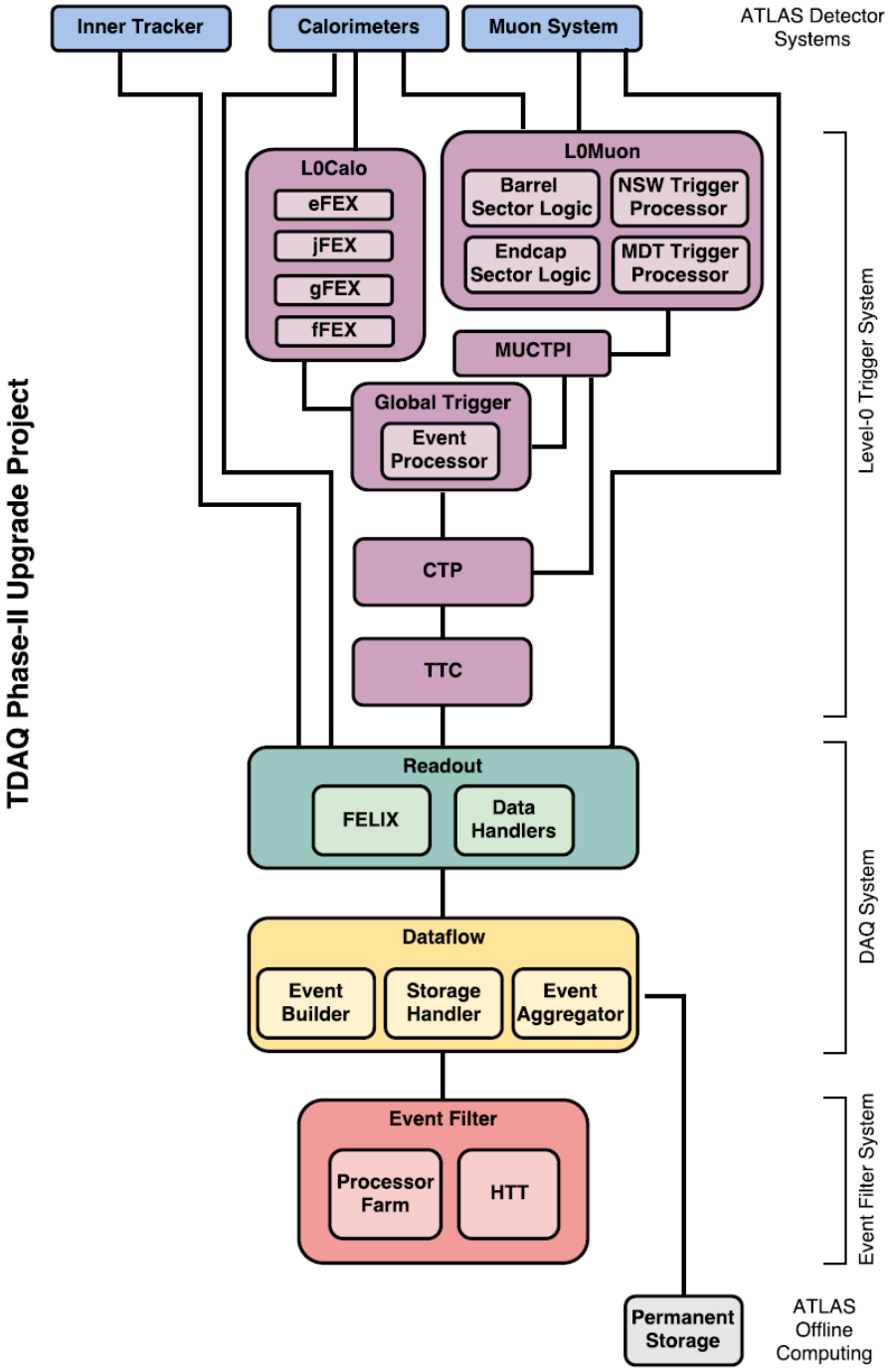

Figure 2.4: *Baseline L0-only design of the TDAQ Phase-II architecture. The upgrade project is divided into three main system: Level-0 Trigger, DAQ (Readout and Dataflow subsystems) and Event Filter. Figure taken from [6].*

possible evolution of the TDAQ system will be the main topic in Chapter 3.

Eventually, the Phase-II DAQ system is designed to make maximal use of networking and computing hardware. Data from detectors will be transferred from the detectors buffers to the software based event filter and event storage.

2.3 Upgrades summary

To summarize, the key features of the various scoping scenarios upgrades are the following:

- The Inner Tracker consists of two sections. The outer section uses silicon micro-strip technology with the layout as defined in the ATLAS Phase-II initial proposals. The inner section is based on silicon pixel technology, but extends the η-coverage from 2.7 to 4.0.

- The Calorimeter consists of two major systems. Liquid argon technology is used for the barrel electromagnetic calorimetry and all of the end-cap and forward calorimetry, while scintillating Tiles are used for the barrel hadronic calorimetry. Both the LAr and TileCal readout electronics will be upgraded to provide streaming of all data.

- The Muon upgrade focuses primarily on improved trigger performance, replacing almost all of the on-detector electronics, including the BI inner barrel region. In the BI region, the MDT precision chambers will be replaced by sMDT chambers to accommodate the installation of a layer of RPC trigger chambers. In addition, the MDT information will supplement the RPC information in the L0 trigger, providing an improved p_{T} resolution. Finally, a very-forward muon-tagger will be added in the region $2.6 < |\eta| < 4.0$.

- The Trigger and Data Acquisition system uses a single level hardware trigger (the L0 baseline has been approved) with enhanced specifications, but with the possible design option of an evolved design, which embeds an L1Track in addition to the L0.

Chapter 3

TDAQ system upgrade

3.1 Trigger and data acquisition

As anticipated in the previous chapter, in particular in Section 2.2.4, the main requirements imposed by the HL-LHC upgrade have posed significant challenges to the ATLAS Trigger and Data Acquisition System (TDAQ) to fully exploit the physics potential of the upgrade. Despite the challenge itself, the main fields which will be explored include for instance the exploration of electroweak symmetry breaking through the properties of the Higgs boson and the research of rare Standard model processes.

The initial plans of the collaboration were described both in the Letter of Intent in 2012 and in the Scoping document in 2015: two "custom-hardware" trigger levels were proposed, which allowed data streaming off-detector either after an initial trigger decision or, in some cases, at the full 40 MHz bunch crossing rate. Since then, the design of the upgraded architecture of the TDAQ has passed through a further evolution, resulting in a baseline architecture with a single-level hardware trigger that features a maximum rate of 1 MHz and 10 µs latency. This is the so called *baseline* system, but in recent times another possible configuration, with some changes, has been developed, and it is called *evolved* system.

3.1.1 Baseline system

As previously mentioned, the TDAQ Phase-II architecture is divided into three main systems: the Level-0 trigger, the DAQ (Readout and Dataflow) and the Event Filter. They form the so called *baseline* system.

The hardware-based L0 trigger system is composed of four different sub-systems, which are either evolution of their predecessors of the Phase-I operation at the LHC or new subsystems added for the new physics requirements of the upgrade. In details, the components of the L0 trigger are the following:

- Level-0 Calorimeter Trigger (L0Calo), which is based on the Phase-1 L1Calo system, with minor modifications applied and it uses coarse-granularity data in order to reconstruct electrons, tau leptons, jet candidate and to calculate the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$. Moreover, the Phase-I calorimeter feature extraction will be complemented with a new forward Feature EXtractor

(fFEX), to ensure an efficient electron identification in the region $3.2 < |\eta| < 4.0$;

- Level-0 Muon Trigger (L0Muon), which has been completely upgraded with respect to the Phase-1 system. The new sub-system will use the upgraded barrel and end-cap sector logic and the NSW trigger processors for the muon reconstruction in the barrel, forward and end-cap regions;

- Global Trigger, which is a new subsystem of the L0 trigger system. It will perform offline-like algorithms on full-granular calorimeter data. It will thus replace and extend the Run 2 and Phase-I Topological Processor, thanks to the full-granularity calorimeter information. L0Calo and L0Muon subsystems send their selected objects the Global Trigger, which are combine with the results of the latter calorimeter processing to refine electron, photon, tau, muon and jet selections;

- Central Trigger Processor (CTP) sub-systems, which have the task to perform the final trigger decision, by applying flexible pre-scales and vetoes to trigger items.

The sub-parts of the trigger itself and the baseline design for the TDAQ architecture are introduced in Figure 3.1. The data inputs can be categorised in the picture as Inner Tracker (i.e. the ITk pixel and strip detectors), Calorimeters (i.e. the LAr and Tile calorimeters) and Muon System (i.e. MDT, RPC, TGC and NSW subsystems of the muon detector).

After the L0 trigger in the transmission chain, the resulting data are transmitted to the Readout subsystem, whose scope is to sent it to all detectors and trigger processors. The connection between this subsystem and the other detectors is done by the Front-End Link eXchange (FELIX) subsystem. Data are received by Data Handlers, where detector specific processing can implemented before the data buffering in the Dataflow subsystem: the Readout subsystem receives data up to 1 MHz event rate. Both of these sub-systems are based on PC servers and standard networking infrastructures.

In order to compete with the 1 MHz input rate of data, a large Event Filter (EF) processor farm is needed. The main function of the EF system is to refine the trigger objects in order to get down to the final output rate. Events will be rejected as early as possible during their processing. The upgraded Event Filter (EF) system is able to provide a high-level trigger, and consist of a CPU-based processing farm, complemented by Hardware-based Tracking for the Trigger (HTT) co-processors, which are designed to allow a fast hardware track reconstruction. HTT consist of two subsystems with identical hardware but different tasks: the regional HTT (rHTT), which performs the tracking in Region of Interest (RoIs), and the global HTT (gHTT), performing tracking in the whole detector volume. More details will be given in Section 3.2.

Indeed, the regional tracking allows a fast initial rejection in the EF of single high-p_{T} lepton and multi-object triggers, reducing the incoming rate to approximately 400 kHz. This system should operate at 1 MHz, using up to 10% of the ITk data, by the selection of tracking modules in particular regions , depending on the results

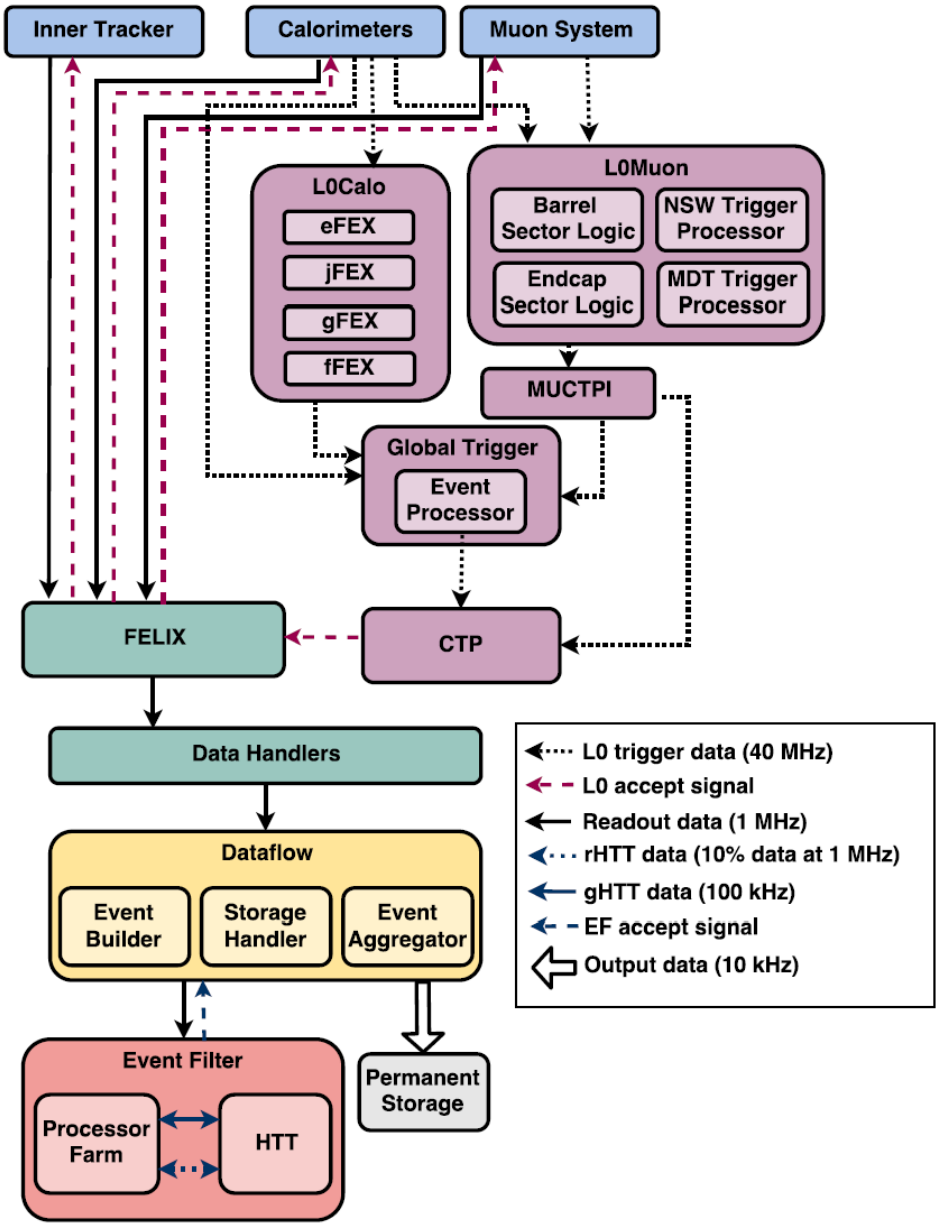

Figure 3.1: *Design of the TDAQ Phase-II baseline architecture. In purple it is shown the trigger: the single level hardware-trigger L0 is composed by the L0Calo and L0Muon sub-systems, it steps through the Global Trigger and eventually reach the CPT. The black dotted arrow represent the L0 trigger dataflow, while the full black lines represent the readout dataflow, which will start only if the detectors and the FELIX recive the L0 accept signal (dashed purple line). Figure taken from [6].*

of the L0 trigger. Moreover, software-based reconstruction will follow to achieve further rejection: this operation will be aided by global tracking around 100 kHz again thanks to HTT.

The EF trigger decision enables the transfer of data corresponding to selected events from the DAQ to permanent storage of the ATLAS offline computing system. The raw output event size is expected to be 6MB, and the total trigger output is

expected to be 10 kHz: the total bandwidth out of the system is 60 Gbit/s.

A management plan is in place to deliver the baseline TDAQ upgraded system fully commissioned by the end of the LS3 (i.e. for the start of Run 4 and of the HL-LHC phase).

3.1.2 Evolved system

The single hardware trigger discussed in the previous section is the selected trigger architecture for the TDAQ system, and it is the nominal design for the start of Run 4. Concerning the main impacts on the performance of the system itself, two risks have been identified. The first risk is the uncertainty in the projected trigger rates for hadronic objects at $\langle\mu\rangle = 200$. The second risk is the uncertainty on the occupancy in the inner pixel detector layers at the ultimate HL-LHC conditions. Because of these reasons, it has been taken into account a new possible architecture for the TDAQ system, so called *evolved* system.

Currently, the evolved design system, with the development of the L1track subsystem, is not supported by the ATLAS collaboration. Difficulties related to the needed time to build the system and the actual speed transmission of the Pixel Detector (at maximum it would be of 4 MHz) probably can not afford the requirements of the the evolved design. Anyhow, some of the proposals of the evolved design will be used in the Phase-II upgrade.

The main differences in the architecture design of the evolved system are shown in Figure 3.2. This two-level development architecture includes a L0 trigger rate up to 2-4 MHz and 10 µs latency, followed by a Level-1 (L1) trigger rate of 600-800 kHz and latency up to 35 µs. In this possible upgrade, the hardware-based track reconstruction is implemented in the L1 trigger system, through the reconfiguration of a part of the HTT.

Unlike the single-level trigger, the hardware trigger is in fact split into a two-level hardware trigger system, in which the HTT performs the primary reduction of the L0 rate for an affordable EF farm size. The basic functionalities of the L0 trigger are unchanged.

However, the Region of Interest Engine (RoIE) is added to the Global Trigger, to calculate RoIs. The data from the ITk strip and ITk outer pixel layers are now used in the Readout system to select the relevant data for L1Track. This subsystem is practically the rHTT of the baseline system: it is reconfigured to be able to reconstruct tracks within a latency time of 6 µs, but only for tracks with $p_{\mathrm{T}} > 4$ GeV. It reconstructs tracks in the RoIs, which is composed of the same hardware and firmware as the HTT components, but with an additional latency of 10 µs.

Right after the track reconstruction, the resulting trigger is combined with calorimeter and muon-based trigger objects in the Global Trigger, after which the CTP forms the L1 decision. Then the main features of the trigger chain are almost the same, with the exception that in the evolved system, right after the L1, the whole data rate of 1 MHz is shared between the full and regional detector readout. So, the regional readout gets priority over the full readout, with respectively a rate of 200-400 kHz and 600-800 kHz detector readout.

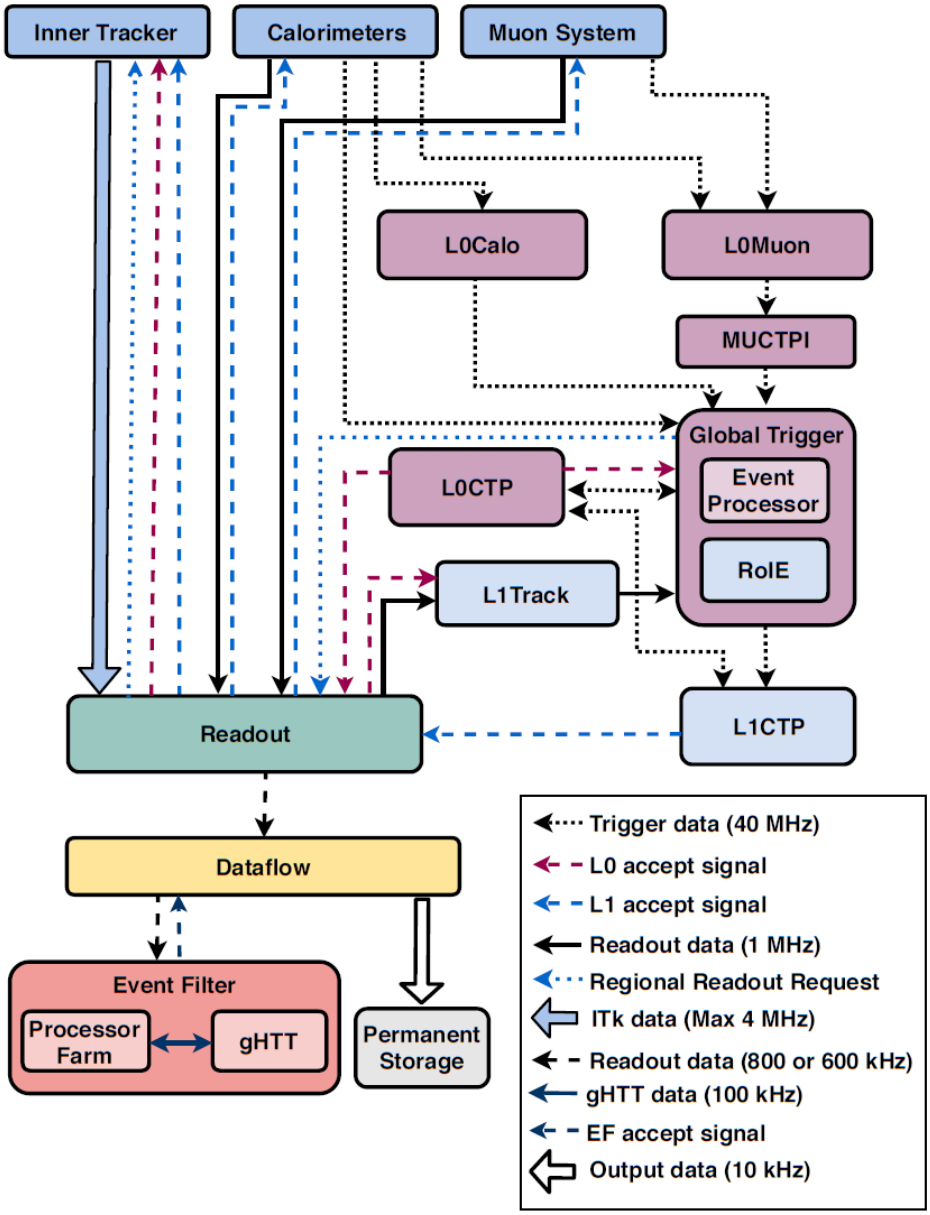

Figure 3.2: *Design of the TDAQ Phase-II evolved architecture. In purple is shown the L0 trigger, composed by the L0Calo and L0Muon sub-systems, the Global Trigger and the L0CPT. In light blue is shown the L1 trigger, composed of L1Track and L1CTP. As for the baseline configuration, the black dotted lines represent the L0 and also L1 dataflow. The full black lines represent the readout dataflow at 1*MHz*, while the dashed black line represent the readout dataflow at 800 or 600*kHz*. Figure taken from [6].*

3.2 Hardware tracking for the trigger

The reconstruction of charged particle tracks in the high pileup conditions is a challenge for the trigger system of the ATLAS experiment in the HL-LHC conditions, mostly because the information from the ITk must be use by the trigger as early as possible in the trigger selection. The task of the reconstruction will be handled

by the Hardware Tracking for the Trigger (HTT) system, a hardware-based track reconstruction system. Briefly, for the baseline scenario, the HTT is made of two main systems: gHTT, which provides global tracking for tracks with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$, and rHTT, which provides regional tracking for tracking with $p_{\mathrm{T}} > 2\,\mathrm{GeV}$. In the evolved one instead, the so called L1Track provides regional tracking for tracks with $p_{\mathrm{T}} > 4\,\mathrm{GeV}$.

The choice on the type of system to be used in the tracking system in not unique. Different technologies may be taken into account to do so:

- a hardware-based system (i.e., HTT), made of custom-designed Associative Memory (AM) ASICs for pattern recognition and Field Programmable Gate Arrays (FPGAs) for track reconstruction and fitting;
- a software system, mostly CPU-based servers with or without accelerators (e.g., GPGPUs).

In the baseline option, the HTT is used in the EF as a tracking co-processor for several reasons, which include considerable experience in the AM technology, the short latency time, a less demanding space requirements and a lower power budget. Depending on the trigger scenario, the EF may transmit two types of requests: finding tracks in regions of interest identified by the previous selection stage (regional tracking, rHTT), and reconstructing tracks in the entire ITk coverage (global tracking, gHTT). Both the regional and the full-scan track reconstruction are provided by the same hardware system.

The HTT system performs the track reconstruction in two separated steps. During the first one, which is the same for both the regional and global reconstruction, the *hits* from the eight ITk layers are clustered into consecutive ITk strip or pixel channels (so-called "super strips"). These super strips are compared to a large bank of pre-computed and simulated patterns, through the AM ASICs. The pre-configured patterns are called *roads*, and they are physical regions of the detector defined by the physic events of interest. For instance, if in these regions one hit occurs with a specific pattern (more details will be given in Section 4.4.1), then all the hits inside the same region must be studied and processed in the track fitting.

Since the inputs of the HTT are hits from the ITk, the incoming data need to be organized. In fact, the data from the pixel is made into clusters using clustering algorithms, which are subsequently converted into the so called super strips, groups of consecutive silicon strip or pixel channels (shown in Figure 3.3). Each super strip is associated to a unique SuperStrip Identifiers (SSIDs). Super strip width can be any integer value: their dimension is linked to the choice of the number of layers to be used in the AM step and the choice of the actual layers.

In the second step, for each matched hit pattern, the track parameters and quality are computed from the corresponding full-resolution hits in a FPGA. At this point, rHTT tracks obtained from the eight ITk layers are complete, but for gHTT, each track candidate goes through a second stage of processing, where its track fit is extrapolated to remaining layers of the gHTT. It is eventually produced a full track fit also on gHTT after this further stage. The produced output consists of candidate tracks and the associated χ^2 of the track-fit.

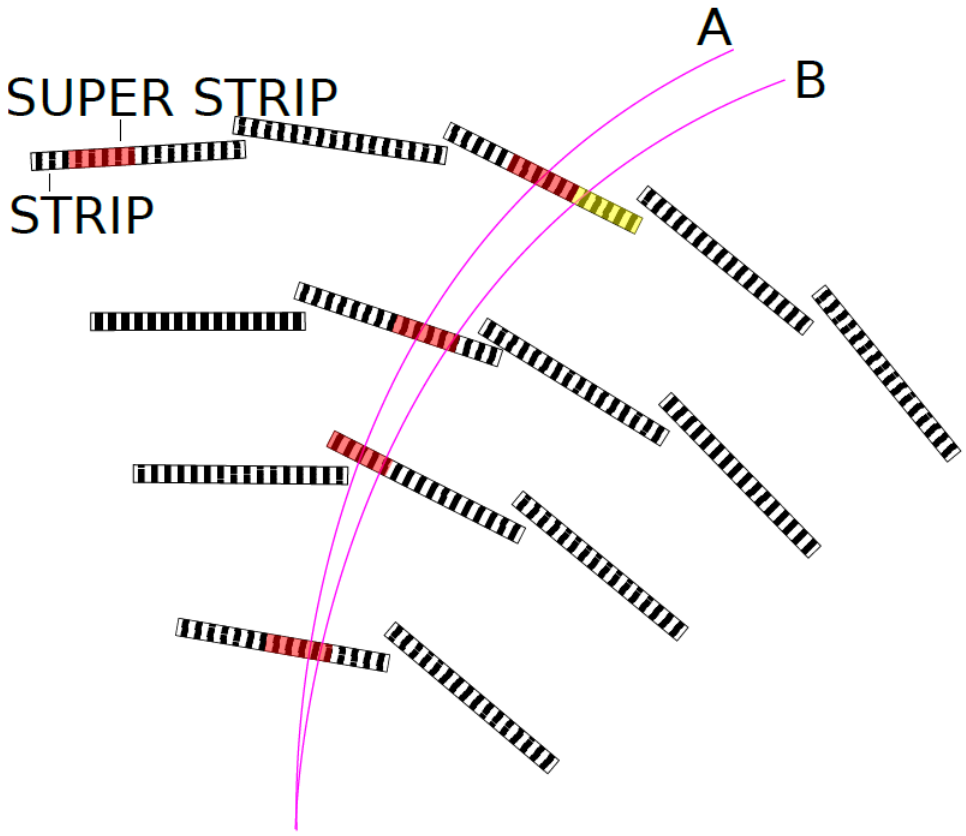

Figure 3.3: *Graphical illustration of the tracks, crossing layers of strips or pixels organized into superstrips. The figure is taken from [12].*

Another plausible option for the tracking, instead of using the AM as comparator of possible track candidates, in the evolved scenario another possibility has been developed. Instead of the AM technology, in this scenario the track reconstruction will be performed by a clustering filtering method based on the mathematical Hough transform. Shortly, the transform allows to identify points belonging to straight or curved lines, and its implementation will be done on FPGAs boards. Further details will be provided in Section 3.3.

3.2.1 Associative memories

The HTT pattern recognition method is based on the Fast TRacker (FTR) method, the currently used one for the ATLAS Inner Detector, but applied to the ITk layout and the HL-LHC pileup levels. The main idea of the method relies in the usage of template patterns, obtained form single-muon tracks simulated training events with SSIDs from a given number of layers in the ITk. These patterns are eventually stored into pattern banks which, in the hardware implementation, are stored in AM ASICs. Notice that a single patter describes a sequence of eight SSIDs in different layers of the detector.

In the possibility of implementation of this type of pattern matching, the technology on which the AM are based is required to accomplish the task. Indeed, each bank inserted into the AM is stocked with an address in the memory (the

used technology is called content-addressable memory). Such a system compares the input data which receives to the data stored in its own memory, and if it finds a match, it sends as output the address of the data that best matches the input. This function is thus utilized by the AM to compare sets of ITk real clusters to predefined pattern from tracks.

Content-addressable memories often have the so called *don't care* bits, which match independently from the input. Those bits are used in the ASICs to combine two or more similar patterns to one, which frees up space to include more patterns and increase the track finding efficiency. In the ATLAS baseline configuration, the AM ASICs use a subset of eight layers of the ITk to construct a bank of patterns, hence with eight super strips in each pattern.

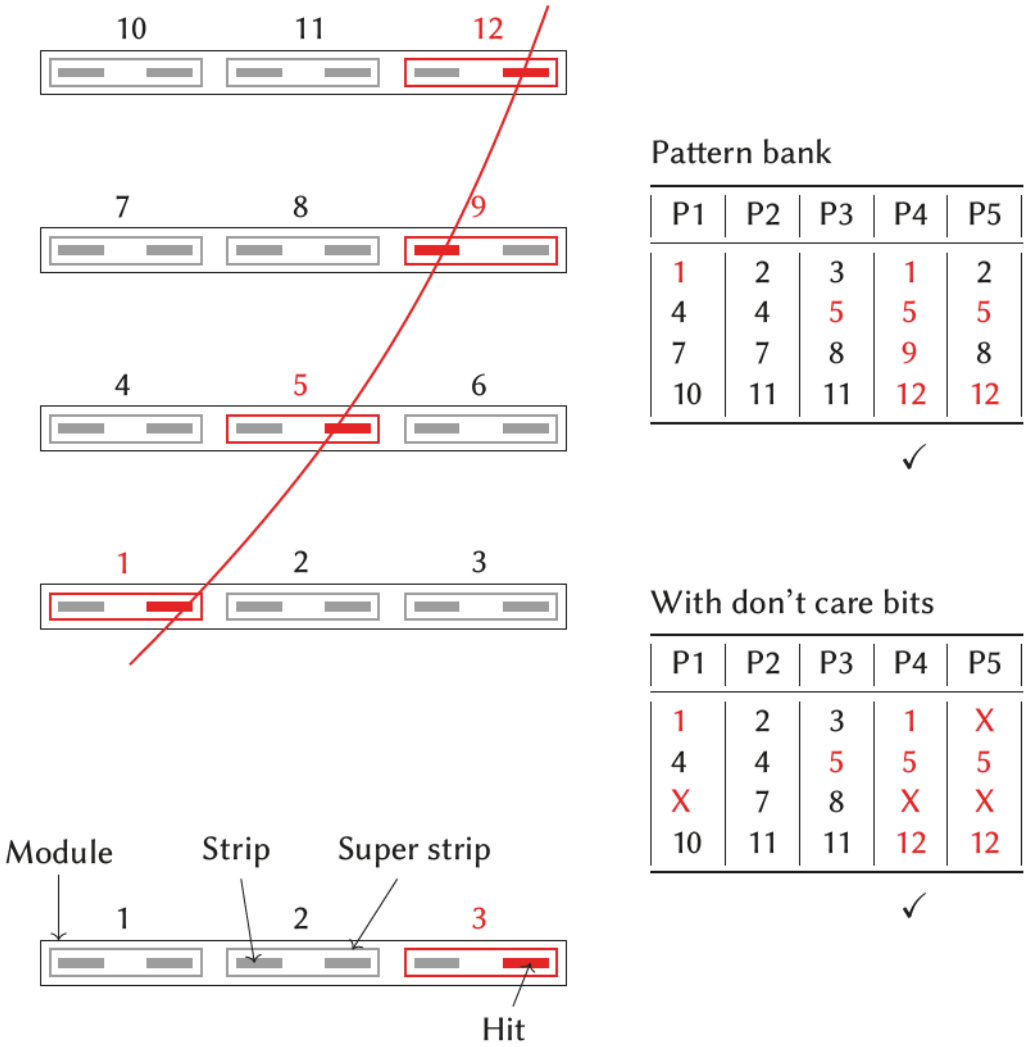

Figure 3.4: *Simplified example of the pattern matching done with AM ASICs in just four detector layers. For instance, each layer has a single module with six strips combined into 3 super strips, and the pattern banks contains 5 possible patterns. The "X" symbols represent the* don't care *bits. The figure is taken from [12].*

An example on how the pattern matching method works is shown in Figure 3.4.

In the figure, with only four different layers, adjacent strips or pixels are combined into the super strips. For a track traversing the detector will trace out a pattern consisting of one super strip in each layer of the detector. Consequently, using a large set of simulated tracks allow to build a bank of patterns can be created. The AM chips will try to match an input pattern and it will output the address which best suits with the input pattern.

3.2.2 Track fitting

The clusters sent to the AM and which passes the match comparisons are given to a track-filter implemented in FPGAs, but the track fitting is performed in two different stages.

Firstly, the FPGA in the Pattern Recognition Magazine (PRM) takes the full-resolution hits from the roads passed by the pattern matching and extrapolates the track parameters and the χ^2 of the fit. The track parameters p_i (i.e., $p_{\mathrm{T}}, \phi_0, \eta, d_0$ and z_0) are calculated through a linear interpolation:

$$p_i = \sum_{j=1}^{N} C_{ij} x_j + q_i, \tag{3.1}$$

where x_j are the full-resolution local cluster coordinates and (C_{ij}, q_j) are constants unique for each fit. These constants are different for each so called *sector*, which consists of a unique combination of eight modules from different ITk layers. Their values are determined from a large sample of simulated muon tracks, having the same parameter ranges and distributions as the ones used in generating the patterns. The quality of the fit is evaluated using a linearised χ^2 method which is fast to compute with FPGAs:

$$\chi^2 = \sum_{i=1}^{N-5} \Big(\sum_{j=1}^{N} A_{ij} x_j + k_i \Big)^2, \tag{3.2}$$

where the constants A_{ij} and k_i are additional constants needed per each sector, obtained similarly as the previous ones. For instance, several thousand sectors are needed for fitting a $\eta \times \phi = 0.2 \times 0.2$ region (about forty million coefficients need thus to be stored in external memories of the PRM or in the internal FPGA memory).

In the second stage, another FPGA installed on the Track-Fitting Magazine (TFM) and for each track it still calculate the 5 helix parameters and the χ^2 still using the equations 3.1 and 3.2, but it receives the eight layer tracks form six PRM cards in addition to all the hits from the detector layers not used by the PRM. The TFM carries out two functions: the *extrapolator*, which finds the hits on the additional silicon layers close to the PRM track, and the *track fitter*, which fits the hits on the PRM track with each combination of hits on the other layers, applying a χ^2 cut.

After both the phases are terminated, the track candidates are to the final stage where duplicates of same track obtained are removed, and eventually the true track candidates are sent to the Event Filter.

3.3 Hough transform for particle tracking

The alternative proposed in the evolved design of the TDAQ system still needs a fast and efficient HTT able to sustain the high rate of events in the future pileup conditions, but also at a relatively affordable costs. Since the stock of the pattern banks would require a huge amount of local memory, another eventual design has been tested in the efficiency and in the expense: this design will include the hardware implementation of the Hough transform in one of the subsystems of the TDAQ.

Recalling briefly what has already been presented, the proposal for the evolved design in the Phase-II upgrade is to introduce a two-level hardware trigger. In the second level of these two, the so called L1Track, it will occur the read out of the inner tracker in the RoIs and search for tracks of charged particles. Despite that, the different baseline approach for the pattern recognition is to use custom designed AM ASICs to perform pattern matching and eventually track fitting would be performed in an FPGA. A track trigger of this kind has been shown to maintain reasonable trigger rates while still maintaining low energy and momentum thresholds. Instead, the evolved approach choice for the pattern recognition, which will be indeed done in the L1track step, is the hardware implementation of the Hough transform.

In its original form, the Hough transform is an image processing algorithm which searches for straight lines, but the method itself can be used to analyse and search any type of shape (as we can see from the example in Figure 3.5). Its first usage in particle physics was in the photographic analysis of bubble chambers plates. In the recent years its usage has developed in the high energy physics field, mostly because of the advancements in the computational capacities of Graphics Processing Units (GPUs) and FPGAs. The transform aims to detect a fixed type of feature in a

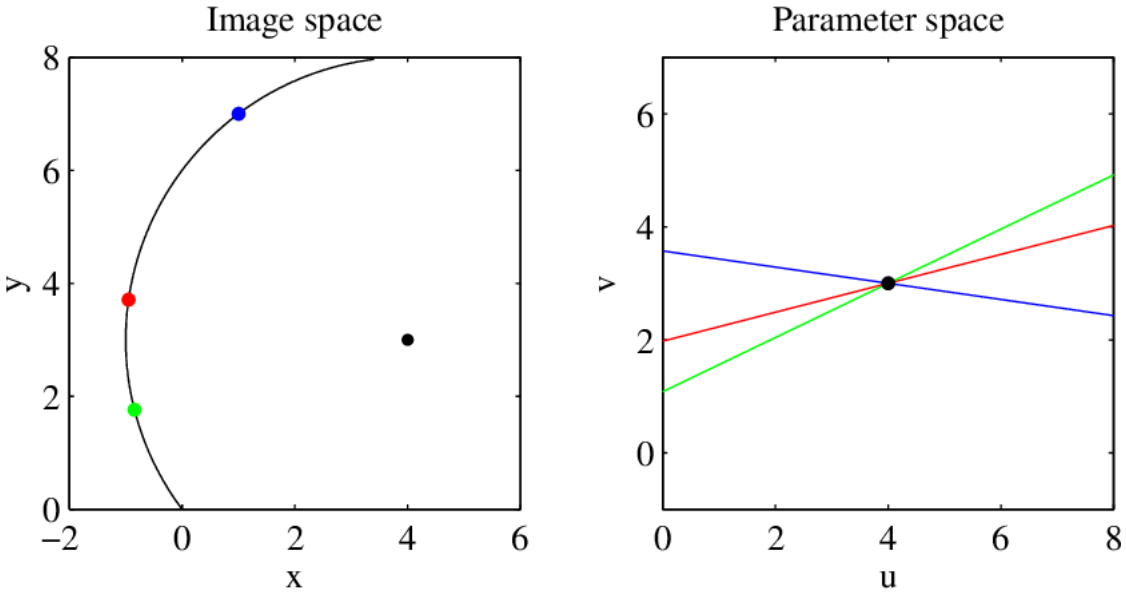

Figure 3.5: *Example of Hough transform for the identification of circles. The green, red and blue dot, in the left plot, are transformed into the corresponding coloured lines in the parameter space, in the right plot. The intersection of these lines gives the coordinates of the circle centre in the image.*

binary image by parametrising that feature and mapping each point in an image to the set of all parameters compatible with that point. Then, votes are cast in a discrete parameter space, called an accumulator (see section Section 3.3.2), for each point. If two points belong to the same feature, their votes in the accumulator will overlap at the specific parameter values describing that point.

For instance, if the feature is a line, it would be parametrised by a slope and an offset. The accumulator would then be a two-dimensional histogram with the slope on one axis and the offset on the other. All combinations of slopes and offsets would then be calculated for each point and the corresponding bins in the accumulator incremented. Bins with contributions from several points would then be considered line candidates with slope and offset given by the bin coordinates.

In the physics contest of interest, the possibility of implementing such algorithm in hardware devices with high performances is more than concrete. Likely, in a track trigger, the feature one would like to detect is high-momentum tracks. The track of a charged particle in the transverse plane (x-y plane) of the ATLAS tracker has the shape of a circular arc which can be described, as in the following Section 3.3.1, by the transverse momentum p_{T} and its initial angular direction ϕ_0.

One clear advantage of this type of algorithm is that its latency time will grow linearly with respect to the number of hits, instead of a combinatorial algorithm, whose latency time grows much more rapidly with the number of hits. Moreover, the Hough transform is much more tolerant to "missing" hits or hits which does not perfectly match for instance with a given default pattern (because of limited resolution). Lastly, the costs of the Hough transform implementation would be much more for a limited number of input hits, but its performances are significantly improved with a large set of hits, another reason in favour of the introduction of this algorithm.

3.3.1 Hough transform for circles

In this section, we will derive the Hough transform useful for the physics scenario of our interest, or rather the motion of charged particles inside a magnetic field. In presence of any electromagnetic field, charged particles are subject to the Lorentz force

$$\boldsymbol{F} = q\boldsymbol{E} + q\boldsymbol{v} \times \boldsymbol{B}, \tag{3.3}$$

where q is the electric charge of the particle, $\boldsymbol{E}$ is the electric field, $\boldsymbol{v}$ is the velocity of the particle and $\boldsymbol{B}$ is the magnetic field. We can assume a negligible electric field and a uniform magnetic field along the z direction, i.e. $\boldsymbol{B} = B\hat{\boldsymbol{z}}$. Choosing cylindrical coordinates in our frame with and origin so that $\boldsymbol{v} = v\hat{\boldsymbol{\phi}}$ the Lorentz force becomes

$$\boldsymbol{F} = qvB\hat{\boldsymbol{r}}. \tag{3.4}$$

Thus, if the particle's momentum remains constant, the particle's trajectory will be circular and the force can be described through a radial acceleration:

$$\boldsymbol{F} = \frac{p_{\mathrm{T}} v}{r}\hat{\boldsymbol{r}}, \tag{3.5}$$

where r is the circle's radius and p_{T} is the transverse component of the relativistic momentum of the particle. By substituting $\boldsymbol{F}$ definition in 3.4 with the one in 3.5

and dropping the vector notation it can be written:

$$p_{\mathrm{T}} = qBr. \tag{3.6}$$

equation 3.6 holds for SI-units, i.e. the momentum p_{T} expressed in $\mathrm{kg\,m\,s^{-1}}$ and the charge q in C. But in high energy physics it is much more convenient to express it in units of elementary charge e and GeV/c:

$$p_{\mathrm{T}} = \frac{cqeBr}{e} \cdot 10^{-9} = cqBr \cdot 10^{-9} \approx 0.3\, qBr, \tag{3.7}$$

with $c \approx 3 \times 10^{8}\mathrm{m\,s^{-1}}$. Now the transverse momentum is in units of GeV/c, q in e, B in T and r in m. equation 3.7 determines the relationship between the transverse momentum of a particle and the radius of the circle it traces in a uniform magnetic field.

Now, since the purpose is the reconstruction of particle's track, it is necessary to determine all the possible radii of the circles which pass through at minimum two points (i.e. the hits that the tracker has registered). In Cartesian coordinates the equation of a circle is

$$(x-a)^2 + (y-b)^2 = r^2, \tag{3.8}$$

and by substituting the centre's coordinates (a, b) with the usual polar coordinates $(r\cos\theta, r\sin\theta)$, evaluating the equation for a pair of points (x_1, y_1) and (x_2, y_2) and subtracting the two results it turns

$$r(\theta) = \frac{1}{2}\frac{(y_1^2 - y_2^2) + (x_1^2 - x_2^2)}{(y_1^2 - y_2^2)\sin\theta + (x_1^2 - x_2^2)\cos\theta}. \tag{3.9}$$

Furthermore, we can simplify the equation by breaking out $1/r$, using polar coordinates for the points $((x_i, y_i) \to (r_i\cos\theta_i, r_i\sin\theta_i))$ and by substituting $\theta = \phi_0 + \frac{3\pi}{2}$, with ϕ azimuthal angle at the track's closest approach to the beam line (i.e. the angle in the x-y plane of ATLAS with which the particle enters in the tracker). Using trigonometric identities, it turns

$$\frac{1}{2r}(\phi_0) = \frac{r_1\sin(\phi_0 - \phi_1) - r_2\sin(\phi - \phi_2)}{r_1^2 - r_2^2}, \tag{3.10}$$

Substituting then r with the p_{T} definition in equation 3.7, we obtain the equation used to parametrize p_{t} as a function of ϕ in the Hough transform:

$$0.15\frac{qB}{p_{\mathrm{T}}} = \frac{r_1\sin(\phi_0 - \phi_1) - r_2\sin(\phi_0 - \phi_2)}{r_1^2 - r_2^2}. \tag{3.11}$$

equation 3.11 applies to pairs of generic points, it is not constrained to any vertex. However, it is possible to apply a vertex constraint, for instance by fixing one of the two points in the origin (i.e. setting the point 2 in the origin means to impose $r_2 = 0$), and keeping the remaining point coordinates:

$$0.15\frac{qB}{p_{\mathrm{T}}} = \frac{\sin(\phi_0 - \phi_1)}{r_1}. \tag{3.12}$$

In the ATLAS experiment, the magnetic field modulus is $B = 2\,\mathrm{T}$ and the radius is commonly measured in mm instead of m. Knowing that, we can write equation 3.12 as

$$\frac{qA}{p_\mathrm{T}} = \frac{\sin(\phi_0 - \phi)}{r} \tag{3.13}$$

where $A = 3 \times 10^{-4}\mathrm{GeV\,mm^{-1}}c^{-1}e^{-1}$ and the 1 index have been dropped. For small values of ϕ the equation can be further simplified through the usage of the first order Taylor expansion of $\sin(\phi_0 - \phi_1) \approx \phi_0 - \phi_1$.

3.3.2 Implementation

As an example to the applicability of the algorithm to the tracking system, the following paragraphs will present the studies developed for the actual feasibility on the hardware implementation, carried out by Mikeal Mårtensson in his PhD thesis [12].

The main idea behind the Hough transform obtained in equation 3.13 is the transformation of spatial points into curves in a new defined parameter space: a pair (r, ϕ) is turned into a space spanned by the parameters qA/p_T and ϕ_0. This parameter space is referred to as the *accumulator*, as anticipated in Section 3.3.

The starting point of the algorithm development is shown in Figure 3.6. In practice, each arc of circumference in grey represent layer of the tracker, on each of which it is possible to distinguish the so called *clusters*. Notice that these ones does not correspond to the *hits* in the tracking detector: one *hit* is registered by several next to next pixels, while *clusters* are reconstruction of the track hits through the charged distribution in the "hit" pixel, which indeed have higher spatial precision.

In the presented example, clusters are shown as black dots if aligned along a signal track, or as black crosses if not associated to any particular track. Once the Hough transform operates, the clusters are transformed into lines: their intersection in the parameter space represent a possible match for a particle track.

In fact, when this type of transformation is applied to several clusters coming from the same track, in the parameter's space the corresponding curves drawn cross at the point corresponding to the parameters of the track. While clusters not coming from the any track or from tracks outside the considered parameters, form randomly crossing curves in the accumulator.

This procedure has to include the whole set of operating tracking layers (i.e., for each set of tracking layers a parameters space is done), which means that the track candidates have to be found with a match within the different layers. To better explain this concept, the description on how the accumulator works will be more clarifying.

The accumulator

The accumulator has a key role in the Hough transform and its implementation. The whole process can be reduced to the creation of this object, on which many operations are performed, such as filling it and selecting tracking candidates from it. In fact, the accumulator has been implemented as a two dimensional histogram. This histogram contains two variables : a boolean value for each detector layer,

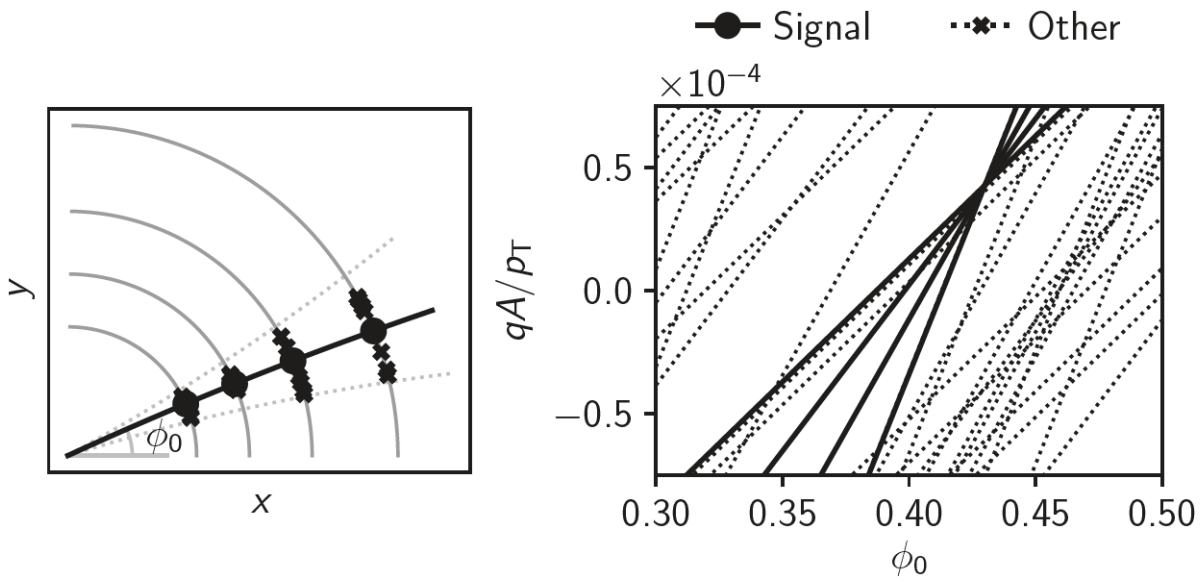

Figure 3.6: *On the left, a quadrant of the transverse plane of the tracker with clusters along a signal track (black dots) and clusters not associated with a track (black crosses) in a given angular range. On the right instead is shown the accumulator, where full lines are the transformation of the black dots, while dotted lines correspond to the black crosses. The figure is taken from [12].*

which is true in the case if one or more clusters in that detector layer goes through this particular bin in the accumulator or false otherwise, and a list of all clusters going through this particular bin. An example of an accumulator is shown in Figure 3.7.

It shows a set of clusters which are registered in the whole set of used tracking layers, but only for a single muon event. With a given single track, finding a good candidate simply means to get the intersection point of the parametrized curves, which in the plot is represented as the small region coloured in yellow: this means that in this region of the parameter space a higher number of layers have registered the same hypothetical cluster. Thus, if we apply a simple threshold on the number of layers which match a given cluster would lead to find a track candidate pretty easily.

However, adding the pile-up, it is indeed harder to discriminate the muon signal from the background. As we can see in Figure 3.8, the same muon signal is shown with the minimum bias signal, which is produced by 200 proton-proton interactions. It is now almost impossible to apply a threshold decision to try to find matches in the possible track candidates.

Although the RoI considered is small in $\Delta\eta_0 \times \Delta\phi_0$, it is wide in z, due to the 300 mm longitudinal spread of the interacting proton beams and this case a huge amount of overlapped clusters which indeed makes harder the pattern recognition to the algorithm. Despite that, the situation can be improved: it is possible to split the RoI into smaller regions along the z axis, shaped like the figure Figure 3.9. Applying this shapes to the RoIs, then the occupancy can be reduced by paying the cost of having more accumulators to deal with (for each added slice).

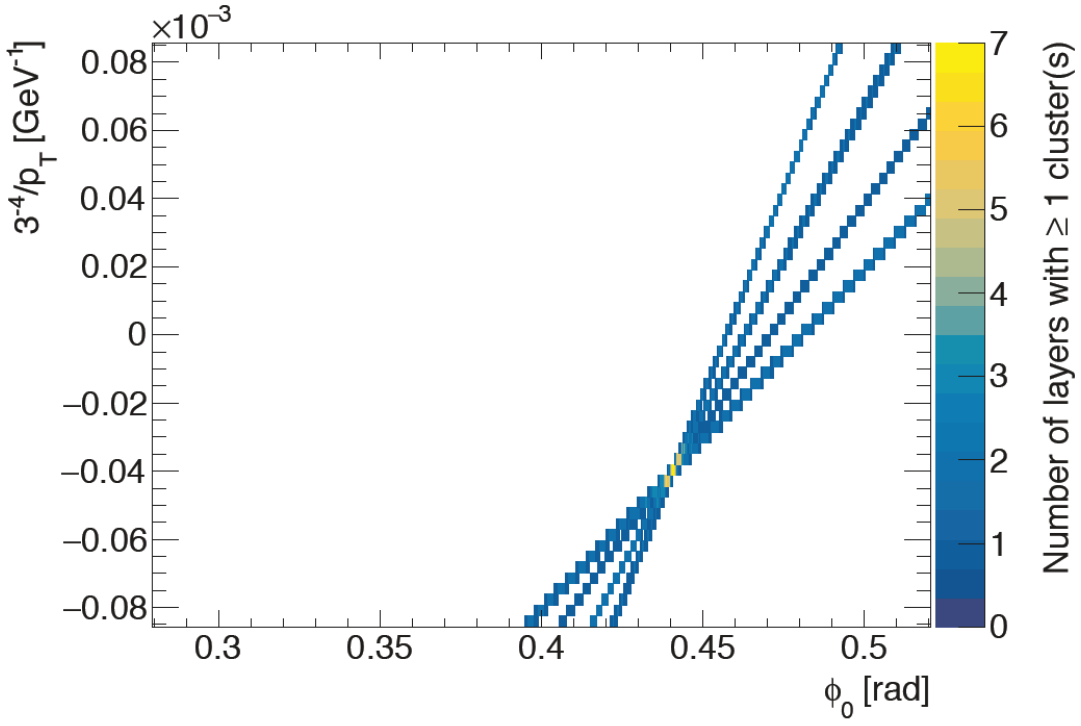

Figure 3.7: *Accumulator filled with clusters from a single muon track. The intensity of the colours represent in how many layers the same cluster have been registered. This figure is taken from [12].*

The splitting boundaries in the z-r plane are identified by

$$\begin{aligned} z_{\min,n}(r) &= z_0 + n\Delta z + r\sinh\eta_{\min}, \\ z_{\max,n}(r) &= z_0 + (n+1)\Delta z + r\sinh\eta_{\max}, \end{aligned} \tag{3.14}$$

where $n = 0, 1, ..., N$ is the split index, $[\eta_{\min}, \eta_{\max}]$ is the η range, $[z_0, z_0 + N\Delta z]$ is the z range and r is the radial coordinate of the detector layer. This splitting technique reduces the occupancy in the accumulators, improving the track finding efficiency and the rejection of unwanted clusters and, nevertheless, the technique suits for a multi.threaded environment, such as an FPGA. It should be taken into account that nearby splits are overlapping and the same hit can show up in multiple accumulators.

Eventually, in the accumulator, a selection of bins can be seen as a track candidate with a set of clusters associated with it. The simplest way to choose them is to select single bins from where the number of clusters in unique layers is above a given threshold. But another criterion as simple as the threshold could be to apply another threshold to neighbouring bins, in order to cross-check for the presence of clusters in the close bins, but selecting eventually the central bin. Of course, more elaborate selection could have potentially a better signal efficiency and background suppression, but the system may still be changed before the ultimate implementation, since the several possibilities are still under testing and it has not been decided if the baseline or the evolved design will be installed for the Phase-II upgrade.

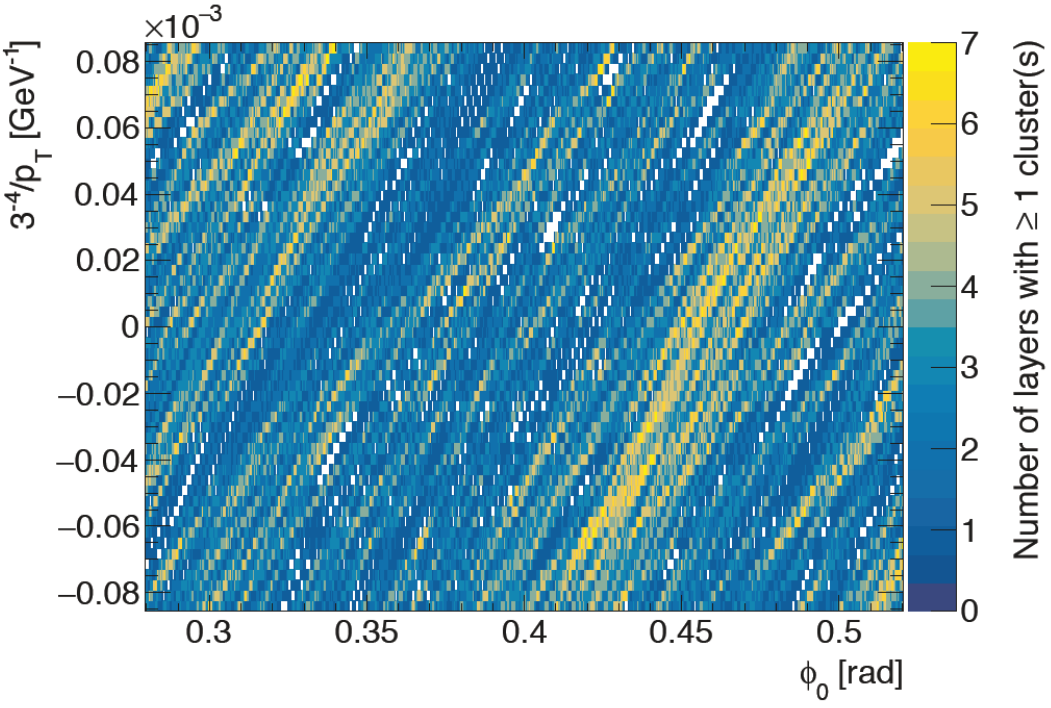

Figure 3.8: *Accumulator filled with clusters from a single muon track and minimum bias event, which correspond to 200 proton-proton interactions. It is indeed much harder to discriminate track signals within different layers. This figure is taken from [12].*

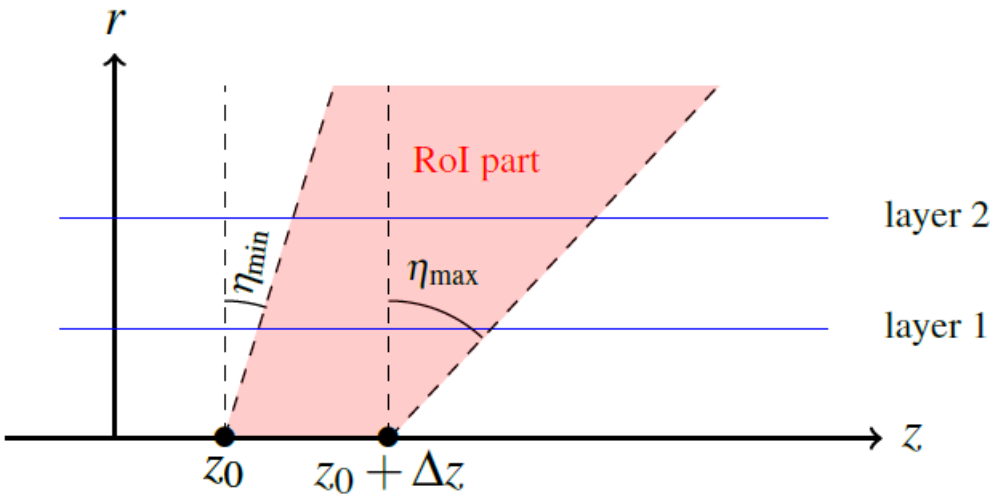

Figure 3.9: *Shape of the region in which the RoI is split along z, to reduce occupancy in the Hough accumulators. The boundaries are defined by equations 3.14. Of course, nearby splits are overlapping and the same hit can show up in multiple accumulators. This figure is taken from [12].*

The results provided by this ultimate selection are subsequentially transmitted to the global trigger processor in the DAQ chain, end eventually through the L1 accept signal it will trigger the dataflow transmission to the readout sub-system.

3.4 TDAQ upgrade summary

A definitive decision about how the tracking reconstruction is going to be accomplished by the TDAQ system has not been made. Currently, the different designs are under several tests, in terms of the performances through simulated events of the Phase-II upgrade, and in terms of feasibility of the implementation for both the AM design and the Hough transform FPGAs one.

The AM ASICs design implementation would be cheaper in terms of costs and since it will include a software-based track recontruction processing, the latency time of these processes can not be fully controlled and synchronized with the experiment clock. Instead, the evolved design, through the implementation of the Hough transform in hardware-based system, would be completely synchronized with the experiment clock, although its implementation costs would be more expensive than the baseline design ones.

Whatever the final decision will be, either the Baseline or the Evolved design will be implemented and installed, at last, during the Long Shutdown 3 (LS3).

Chapter 4

Demonstrator development for HT algorithm

4.1 Bologna's hardware framework

The frame presented in the previous chapters shows an escalation from the most general plans for the HL-LHC Phase-II upgrade, eventually reaching the possible options which can be implemented in the new TDAQ particle tracking subsystem. Major details were posed in the possible outlines of the future TDAQ system, also because of the main scope of the presented work regards the preparation of some of the hardware parts that will be implemented physically added in the hardware tracking subsystem.

The evolution path followed has led to the ultimate purpose of this paper: the implementation of a hardware design for the data transmission, compatible with the ATLAS TDAQ software. As already mentioned, the proposals for the HTT actual implementation and design are still under evaluation of costs and efficiencies, but since in either possibilities the HTT firmware has to be designed, an architecture for transmission will be useful. In practice, such data transmitter would be the interface between the front-end electronics of the detectors and the physical boards of the TDAQ system.

In details, my personal work has regarded the construction of a firmware architecture which allowed the physical test of the data transmitted , in order to verify the feasibility of high speed transmission, by maintaining the capability to communicate with the main customized board which will embed the HTT hardware design. Such structure, will be installed on a physical board proposed to be part of the HTT whole structure. Its main scope is to read data from a file (either generated by a simulation software or a simulated physics data file from the ATLAS "official" software) and transfer it to the hardware board programmed with the Hough Transform (HT) algorithm.

The HTT designed block has been taken under development by the electronic group in Bologna: studies and simulations for the implementation of an hardware algorithm to be embedded in the Hardware Tracking for the Trigger are currently conducted. More precisely, the firmware design for the HT is being developed by the Bologna group.

The main purpose is to produce a hardware algorithm able to perform track reconstruction from a data cluster externally received, reaching valuable results to demonstrate the real potential of such hardware architecture. In order to verify the correctness of the algorithm behaviour, once the preliminary works will come to an end, the system developed in Bologna should be eventually compared with the software's results provided by the ATLAS collaboration. The framework of the project is shown in a synthetic scheme in Figure 4.1. The work of the group

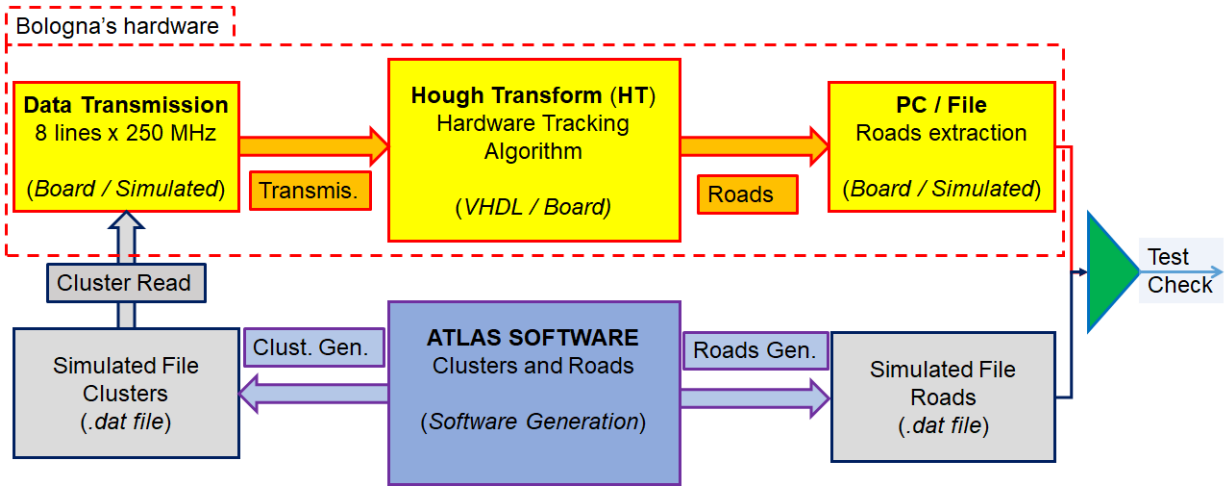

Figure 4.1: *The scheme shows the framework of the Bologna group, inside the red dashed line, and the whole chain for the algorithm functioning test. The Hough Transform (HT) hardware results must be compared with the one of the simulations of the ATLAS software for the roads generation.*

is focused in the data transmission and the hardware HT algorithm for track reconstruction. The whole structure foresee the presence of clusters and roads provided by the ATLAS collaboration, in order to validate the current status of the work and allowing to proceed with further steps. Ideally, the simulated *clusters* file will be used to test the correct functioning of the data transmission and the hardware tracking algorithm and the resulting extracted *roads* will be compared with the simulated *roads* file (based on theoretical predictions).

As mentioned, my work was focused in the development of the data transmission structure, but before entering in the details a major overview on the developed work of the hardware HT will be given, in its features and characteristics and providing the main idea behind the algorithm.

4.1.1 Hardware HT algorithm

As already presented in Section 3.3.2, the HT converts lines into points and vice versa. For the case of particle's track reconstruction in the ATLAS experiment, hits are converted into lines and roads are converted into points. According to this task, the possibility to implement the HT algorithm on FPGAs is concrete. This option can be optimized to reach the same performances of the AM ASIC structure implementation for the ATLAS tracking system. The electronic group in Bologna

has actually developed an implementation for the Hough Transform algorithm, on commercial high-end FPGA, along with simulations.

In details, the hits part of the roads in the ATLAS system have to be extracted, and to achieve correctly the result, the interception points of the several lines must be extracted from the HT parameter space, following a specific "bin clustering" (for instance, the Figure 3.6 shows a schematic example of the idea, and it will be taken as a reference in the further arguments). In the ATLAS case, in relation to the AM ASIC implementation, the physical constrains are of 1200 ϕ_0 bins, where each bin contains a given range of ϕ_0, 64 qA/p_T bins for 8 different layers of the tracker. The system have to keep a latency of 175 ns for up to 16 clusters concurrently in output, at 250 MHz rate clock. All this studies were validated in the past by ATLAS, but until now no-one has been able to provide a concrete proof of its feasibility in terms of digital implementation on commercial programmable devices, such as FPGAs.

The accumulator of the HT is a two dimensional space, which will represent singularly every layer of the tracker. This means that each plane of the detector will have its own parameter space, from which the hits have to be extracted.

The road extraction is not trivial. In the parameter space, roads are counted in a given window, they are processed (n roads to be processed requires n clock cycles) and the process is shifted to another window. Since the FPGA technology can be optimized, probably a multiple window analysis can be performed. Currently, the aim is to provide per each clock cycle an output road.

The main block diagram of the HT firmware structure is shown in Figure 4.2, with the macro components embedded inside the architecture. The diagram flow is not in its final configuration, but it is the requested one for the present requirements for the HTT features and performances.

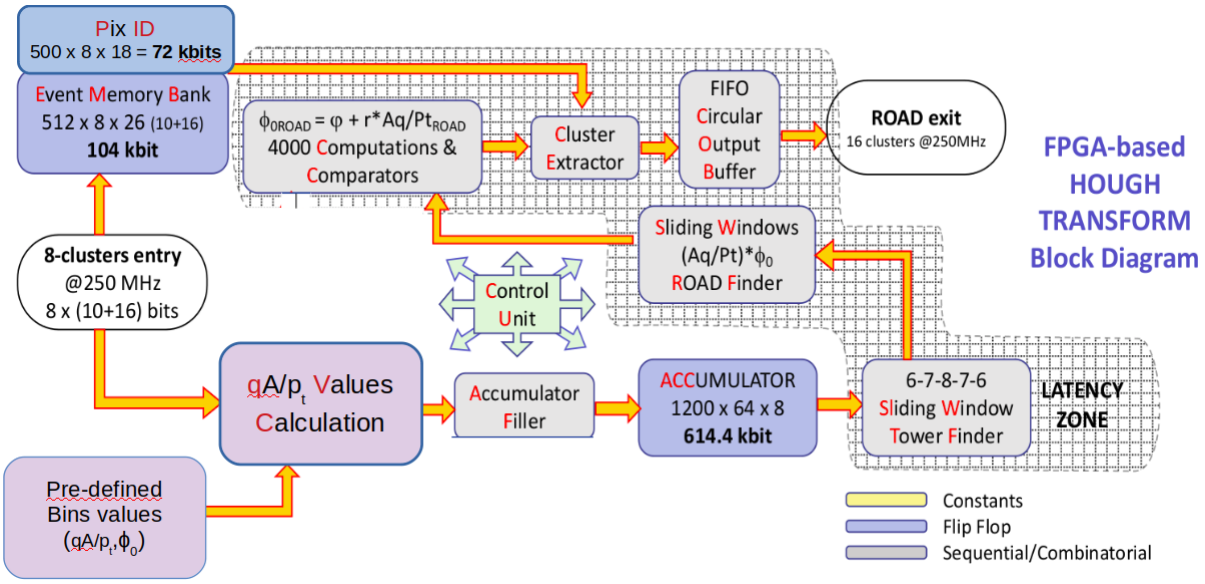

Figure 4.2: *The block diagram of the Hough Transform (HT) algorithm, to be implemented on a FPGA. The 8-cluster entries are stored and processed inside the board. From those values the extraction of qA/p_T is performed, to generate an accumulator. Eventually, the extracted roads from the input clusters are set as an output.*

In details, the input of the firmware is of 8 clusters (r_i, φ_i with $i = 1, 2, ..., 8$), in the figure the block **[8-cluster entry @250 MHz 8x(10+16(bits)]**. The input is sent to two different blocks: one is the **[Event Memory Bank 512 × 8 × 26 104 kbit]**, where the data is temporary stored. The other is the block **[qA/$\mathbf{p_T}$ Values Calculation]**, where from each of the 8 inputs are calculated 1200 qA/p_T values, through the equation 3.13. Thus, a total of 9600 qA/p_T values activate specific bins in the accumulator, "drawing" the transformed clusters as lines, for each of the 8 layers (done by the block **[Accumulator Filler]** and stored in **[ACCUMULATOR 1200 × 64 × 8 614.4 kbit]**. Eventually, a specific pattern is searched to identify the real roads (in two blocks **[6-7-8-7-6 Sliding Window Tower Finder]** and **[Sliding Windows (qA/$\mathbf{p_T}$*ϕ_0) ROAD Finder]**. The searched pattern is of 5 adjacent bins with respectively the values $(6, 7, 8, 7, 6)$, where the values represent the matching pattern of crossed layers, or the number of clusters found in the same bin but in different layers. The input clusters that cross in the Hough space are extracted in the **[Cluster Extractor]**, right after the computation and comparations of the block **[4000 Computations & Comparators]**, using the Hough transform back-propagated. Eventually, the roads are buffered into **[FIFO Circular Output Buffer]**, befor is sent as an output of 16 different clusters (**[ROAD exit 16 clusters @250 MHz]**.

The main logic core of the architecture has been completed in its design, now the architects is under several tests whose purpose is to get a positive feedback in terms of violations from the Xilinx (vendor) synthesis tool. Optimization tests are performed with different set of parameters to prove the real possibilities of implementation and of computing of the hardware algorithm, thus the ultimate version of the HT algorithm in firmware in not currently available, but it will surely be for the time of the LS3.

The presented firmware will be installed on a board which will have to communicate with the front end electronics indirectly, through the usage of another board whose scope is to transfer data from the detector's readout to the HT board. The work presented in the following sections regards the structure of the architecture built for the data transmission, presenting its behavioural features and its simulation and emulation tests.

4.2 Hardware device for data transmission

The main purpose of my work in the electronic group in Bologna was to develop a functioning architecture for data transmission, with specific features able to suit the requirements of the data input of the HT algorithm.

In order to accomplish the task, the design has been studied and tested on the commercial hardware board VCU1525. This new board, produced by Xilinx, features the implementation of the Virtex UltraScale+ FPGA, one of the latest evolution FPGAs available on the market in the high speed hardware devices. As a matter of fact, the ultimate generation hardware technologies are almost always required for high performances detectors readouts or triggers, because of their fast response and the high precision in time for the response (i.e., the hardware-device operations can be controlled at the order of the clock period needed). Nevertheless,

new type of devices are produced practically at each improvement in the silicon-based technology advancement. This entails a constant upgrade on the knowledge of such devices, which means that the necessity of spending time in the understanding of the boards features and testing their actual capability becomes crucial in hardware design master.

The reasons for which the tests have been performed on a commercial device are mostly two. On the one hand, the typical hardware boards installed in the HEP detectors are practically always customised before the actual installation: it is needed a preparation in steps of the board, starting from the paper proposal for the design of the board, ending in the programmed and functionally tested board. Such process requires a longer preparation, deeper technical studies which will start only when a final decision on the TDAQ system will be made. On the other hand, using a commercial device has practical advantages. Once the board is powered and connected it should be ready to be used in all its components, since it is a brand new device. Moreover, the devices' features are fully displaced in their user guide manuals, which allows an easier approach to the hardware boards programming, particularly for beginner firmware developers.

At last, the main scope of the research was to produce a compatible hardware design able to communicate with the future hardware components of the ATLAS TDAQ system, in charge for the hardware tracking reconstruction. In order to accomplish the goal, the designed firmware exploits the transceiver devices for transmission, embedded in the FPGA technology. Its purpose is to perform a data generation of possible track-like events, which would be transmitted by a similar type of device in the actual TDAQ system. The test has been done through the data *loop-back*: data have been internally generated by the board itself, transmitted as output from the board and taken back eventually as an input, checking the correctness of the communication.

4.2.1 The VCU1525 board

The VCU1525 reconfigurable acceleration platform is a a peripheral component interconnect express (PCIe) board and it features the Virtex Ultrascale+ XCVU9P-L2FSGD2104E FPGA. Produced by Xilinx, this FPGA-based PCIe accelerator board is designed to accelerate compute-intensive applications (like machine learning and data analytics). The following section aims to show the main features of the board, enhancing the benefits of the main components, also explaining why the choice fell upon the device.

In its versatile technology, the VCU1525 feature an ultimate generation FPGA which provides the optimal balance between the best required system performances and the smallest power supply envelope. These devices are a optimal option for applications ranging in several applications, from 1+Tbit/s networking transmissions to small network interface controllers.

The board provides a Quad Serial Peripheral Interface (SPI) flash memory for FPGA bitstream storage. The Quad provides up to 1 GB of non-volatile memory, and through it external devices can communicate with the FPGA. Moreover, another access "bridge" to the FPGA is granted via the Joint Test Action Group (JTAG) interface: the configuration is available through the *Vivado* tool hardware device

programmer (i.e. the software provided by Xilinx to programme and build designs for FPGAs), which accesses the onboard USB-to-JTAG bridge device. Also, the Micro-AB USB connector available on the VCU1525 PCIe panel/bracket provides external device programming access.

The board is available with both passing and active cooling. Of course, depending on the type of cooling system, the working conditions and type of installation for the two devices are different, but their operational characteristics are essentially the same. The operating temperature of the board ranges from, 0 °C to 45 °C, while the storage temperature is from −25 °C to +60 °C. Indeed, these ranges must be kept to guarantee the optimal work condition of the board and to prevent an over-temperature shut-down and possible damages to the board.

The board used for the hardware design implementation is the active cooling one and it is shown in Figure 4.3.

Figure 4.3: *The VCU1525 Reconfigurable Acceleration Platform design with active cooling system.*

The VCU1525 embeds also a SI570 programmable low jitter (50 ppm) differential oscillator. When the power-up is on, the oscillator's user clock defaults to an output frequency of 156.250 MHz, but it can be modified by the user in the frequency range from 10 MHz to 810 MHz. To communicate the desired frequencies to be set, the instructions can be sent the device through an internal I^2C interface. In fact, inside the board, an I^2C bus network is implemented, allowing the interconnection within the components, which can be thus programmed and controlled from the external.

Another important feature of the board is the presence of 24 GTY Transceiver, of which 8 are available to the usage. Data transmission is ruled by those units, and

the board uses 16-lane PCIe edge connector CN1, which performs a data transfer at four different rates. The FPGA of the VCU1525 can support the transfer at 8.0 Gbit/s for Gen3 applications and 16.0 Gbit/s for Gen4 applications. The cables for the connection, used in the loop-back, are made of optical fibers. The actual connection is done through a Small Form-factor Pluggable 28G (SFP28) cable, able to reach a maximum data rate of 25 Gbit/s, and potentially up to 28 Gbit/s .

The presented components of the board do not represent the whole set, but just the ones needed to reach the final implementation of the studied firmware. The first task was to understand how to communicate with these sub-parts and to activate them through the usage of the hardware description language. Once the task was accomplished, the focus shifted to the data transmission with the activation of the integrated transceiver components.

4.3 Transceiver activation and programming

In its most general definition, a *transceiver* is a device able to both transmit and receive information via a transmission medium (its name is in fact a combination of the words *transmitter* and *receiver*). Their usage in high performance electronics is crucial to provide a data transmission, capable to sustain the needed rates of operation of high energy physics experiments and it strongly depends on the technological level reached by the board producers. For the specific case of the VCU1525 board, the primitives implemented are GTY transceivers and such technology is indeed very performing in its purposes.

Inside the FPGA, the transceiver are located inside so called *banks* of the FPGA, and their physical location is signed in the user guide of the product. It is important notice that only few of the whole set of 24 transceivers can be activated by external devices, since some of them are used internally by the FPGA and are not programmable by the user, while some of them allow the communication with the outside world. For instance, the 8 available transceivers used for the purpose are showed in their relative banks in Figure 4.4.

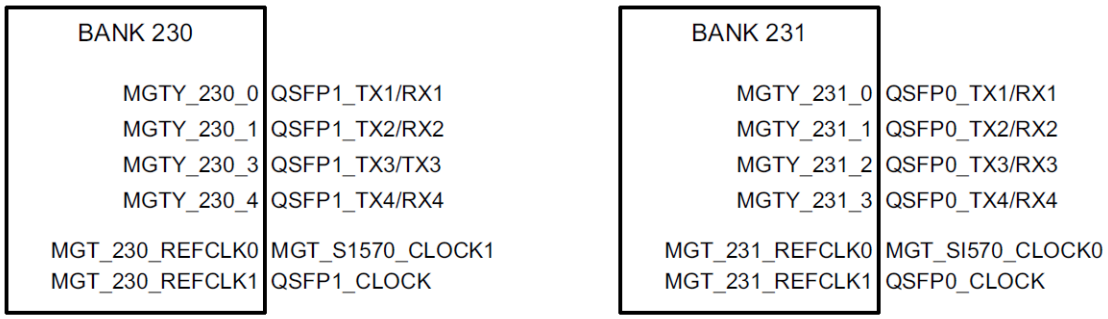

Figure 4.4: *The banks scheme 230 and 231, where the used transceivers are located. Only 8 of the total 24 are shown.* MGTY_230_*i* *and* MGTY_231_*i* *sets have been activated to allow data transmission for the device emulation.*

To be activated, the transceiver must receive an external clock. First of all, before programming the actual components, the first step is to set the operational

reference clock, starting from the default differential clock generator inside the board, which generates a clock signal with a frequency of 156.250 MHz. After the activation to the SI570 oscillator through the I^2C communication, the need is to produce a new clock: for the requirements of the emulation the frequency has been set to 250 MHz, of course for each of the 8 activated channels. The choice on the reference frequency is not trivial. The associative memories option presented in the baseline design works in Section 3.2 at such frequency, thus also the other candidate option has to satisfy such rate. In general, the data input and output rate planned to be used in the HTT algorithm has to be of 250 MHz.

4.3.1 Data transmission

The data protocol used for the transmission is the Aurora Protocol developed by Xilinx, at 64/66 bit encoding. The features of such protocol, like the scalability and light-weighting, make it an optimal link-layer protocol for high-speed serial communication. According to that, having such characteristics for the encoding with the set clock frequency, allows for each active line of the transceivers a data transmission of 16 Gbit/s per line.

The structure of the data transmission is actually simple: the transceiver generates the a sample of data which is sent to a randomiser. This part of the circuitry re-distribute randomly the logical values of "0" and "1", and send the aleatory data to another receiving component. The data is eventually de-randomised with the same logical principle of the randomiser, and it is controlled to be the same generated data cluster initially sent by the first transceiver. The "randomisation" is often exploited in data transmission protocols for a high speed transmissions. In fact, like in most of the cases, the two activated devices receives the clock source which may be out of phase or the edges may not be aligned, or simply from different physical clock sources. Thanks to the randomise procedure, and the usage of the *clock and data recovery* methodology, the receiver is thus able to reconstruct both the clock and the data which is receiving, and it is eventually able to directly synchronize its clock with the one of the data generator. The synchronization is possible thanks to the presence of the Phase-Locked Loop (PLL), an electric circuit present in the FPGA, which allows the phase matching of two different signals.

Besides, the Xilinx's Aurora protocol is available in the 8/10 bit and the 64/66 bit encodings. These type of protocols for data transmissions includes in the serialized lines also another information: the instruction which allow the communication between different components of the transmission chain. Such information is inserted within the data itself, and in the case of the Xilinx's protocol, is identified through the usage of a *gearbox* structure. This circuit adds to the words transmitted two bits, which allows the devices to identify if the incoming word is a numerical data or if it is an internal instruction to be processed. In fact, in both the encoding, the transmission occurs via words of n, which correspond to the real data or an instruction, and words with $n + 2$ bits, where the additive bits identify a starting sequence of data or a starting sequence of instructions (more details about the protocol can be found in Appendix A.4).

4.4 Device emulation

The architecture development occurs in several progressive steps. In the most general view, the process starts from the *Register Transfer Level* (RTL), where the circuits are "built" in software and it proceeds to the *Simulation* step, in which the input and outputs of the built architecture are set to test its correct behaviour. Successively, if the previous in-software parts are correctly completed, the actual in-hardware development starts with the *synthesis*, or the logic scheme creation and the instantiation of the needed component for the architecture (i.e., the *netlist*. At last, the physical *implementation* of the architecture, where the type and the number of components are counted to verify if the FPGA is able to physically generate the connections, while the timing is kept (i.e., all the signals in the flip-flops do not switch their values in the internal transmission within components).

For the specific task of the VCU1525 emulation, the first step regards preparation of the Transceiver Wizard. It is an Intellectual Property (IC) core which provides an industrial standard configuration set for the transceivers cores, to which is possible to apply modifications to give another set of features to the device. For instance, the clock generation is not included in the default codes, but it has to be attached externally, with a block like the one presented in Section 4.3 whose task is the activation of the SI570 clock source.

The further simulation and emulation steps are presented in the following sections, underling the main features and operations that allowed to reach the correct functioning of the transceiver technology on the latest generation FPGAs.

4.4.1 Simulation

Once all the modifications have been inserted in the default code, the architecture is posed under *simulation*. The behaviour of the logic built in circuits is tested in software simulation. In specifics, the transceiver architecture is programmed to *calibrate* the transmitter and the receiver communication. For such technology, the calibration procedure is automatized: the incoming data is a 64 serial word randomised, to which is added a 2 bit word (*gearbox*) after the randomisation. To achieve calibration, firstly the transceiver searches with a bit-slip of the 66 serial bit a sequence like "01" or "10", which can be associated only to the gearbox. This frame searching procedure is possible because the randomiser generate random sequences avoiding the generation of the gearbox sequences inside the 64 transmitted word. Once the gearboxes are identified, the transceiver starts the scan of the 64 in order to align the transmission with the reception.

Simultaneously to the calibration, the data word packs are aligned with the system clock. In order to avoid wrong data transmissions which may be caused by jitter effects of the system clock, its rising ad falling edges are internally delayed in the transceivers, so to be "temporary" located at the centre of a single data, during a transmission. Such operation is a default set operation for transceivers, allowing to correctly operate during the transmission. The alignment procedure work either in single data rate or double data rate.

The architecture's calibration is shown in Figure 4.5. In all the figures of the simulation only the **channel0** of the transceiver is shown, but all the other 7

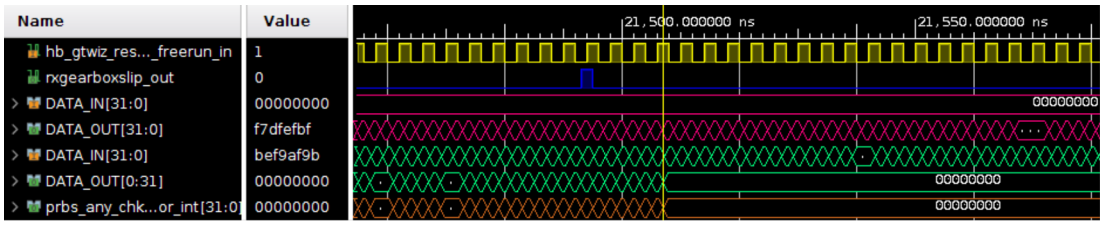

Figure 4.5: *Signals of the channel-0 transceiver, in the moment of the calibration. In yellow the system clock, in blue the bit-slip check signal, in purple the data input and output of the transceiver in green the data input and output of the receiver, and in orange the error checker of the transmission.*

behaves in the same way. The word used to calibrate is a 32-bit "0" word. The serial word is read bit per bit in the word transmitted and it is compared with the one received, and if the words do not match, the received word is shifted by one bit in the reading, searching for a possible match. The procedure is governed by the bit-slip checker, which restarts the calibration each time the data input and output do not match.

As the devices reaches the calibration, the data input of the transceiver and the data output of the transmitter show the same values, and the bit-slip check is not any more triggered to restart the calibration.

While the architecture remains in this idle state, another subpart of the architecture is triggered to be activated: the data transmission of the data set by the user. With the usage of a counter as a cross check of the actual calibration, the data transmission starts after the counter reaches a fixed value (e.g., in the architecture is 64). When the value is reached, the data internally generated in the transmitter (but it can be replaced with an external data source) starts to flow from the transmitter itself. For instance, the start of the actual data transmission is shown in the upper part of Figure 4.6.

In the simulation the data transmitter is a simple counter, generated internally in the VCU1525 board, to actually see if a recognizable data is transmitted. Of course, that does not exclude the possibility to connect this data source with external devices (like other FPGAs) or memories (like associative memories), in order to receive a real data cluster to be transmitted in other parts of a larger architecture.

Once the transmission starts, the data takes a little time (≈ 50 ns) to reach the receiver block, because of the randomization and de-randomization processes the data undergoes before reaching receiver. In the bottom part of Figure 4.6 is shown the moment at which the receiver starts to read the data sent by the transmitter. Even if delayed, the data output is equal to the one internally generated in the FPGA.

After the verification of the correct behaviour of the simulations, the firmware design was synthesized and eventually implemented to eventually program the VCU1525 board.

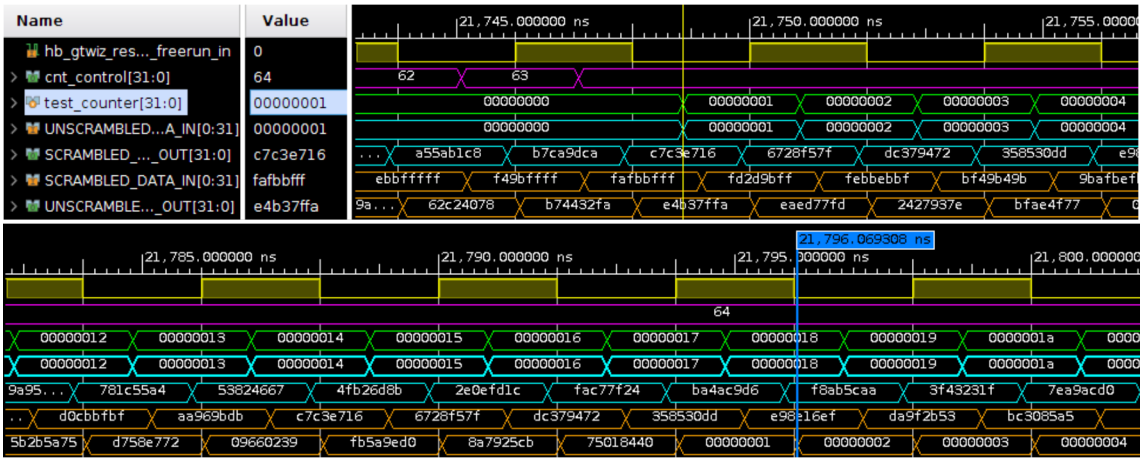

Figure 4.6: *In the top figure: the signals of the channel0 when the data transmission starts from the transmitter (yellow vertical line). In the bottom figure: the signals of the channel0 when the receiver starts to read the previously emitted data (blue vertical line). The signals are: system clock (yellow), control counter (purple), test counter (green), data input/output of transmitter (cyan) and data input/output of receiver (orange).*

4.4.2 Hardware emulation

The VCU1525 has been inserted in a tower case structure, it has been powered with a specific power supply cable and it has been connected to the *Vivado* tool software though a USB connection. The setup for the hardware emulation is shown in figure Figure 4.7.

Figure 4.7: *The case structure where the VCU1525 was inserted to be emulated. On the left the the upper view of the setup. On the right loop-back cables and the USB connection.*

The possibility to check the correct behaviour of the implemented logic is guaranteed by the Integrated Logic Analyzer (ILA), an IP core provided by Xilinx. The logic analyser can monitor internal signals of a design, showing in function of the time the chosen signals, with the possibility to insert some internal triggers

for the analysis. Moreover, it is possible to trigger particular input-output signals through the usage another logic core, the Virtual Input Output (VIO), to manually control or activate some processes of the architecture.

Thus, after the FPGA was correctly programmed and after the general reset of the board components, the logic was triggered with the VIO to start to operate. The first operational step that the architecture undergoes is the data calibration, which is shown in Figure 4.8.

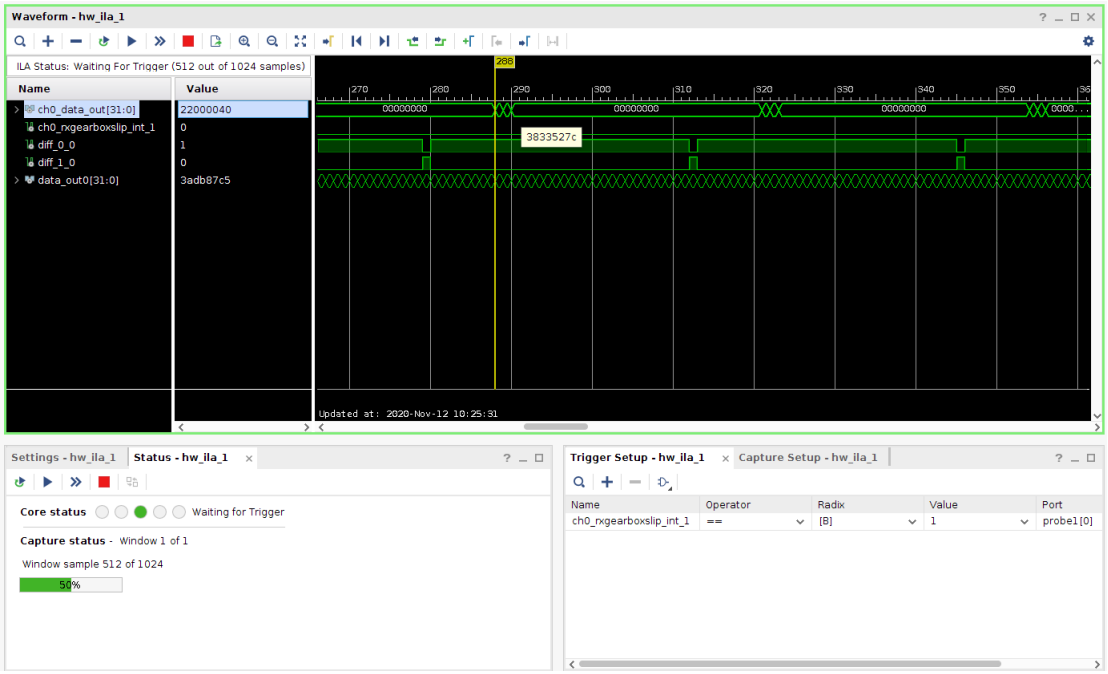

Figure 4.8: *View of the waveforms in the instantiated ILA, when the architecture reaches the calibration. Notice in fact that the trigger is set to the bit-slip controller, but it is never triggered (i.e. the system is calibrated). The signals shown are:* ***ch0_data_out****, the output data of the calibration system,* ***ch0_rxgearboxslip_int_1****, the bit-slip controller,* ***diff_0_0*** *and* ***diff_1_0*** *which are simple checker of the correct transmission and the* ***data_out0*** *is the actual data output of the receiver.*

As mentioned in Section 4.4.1, the focus was posed only in the **channel0** of the architecture, since all the other channels are set to the same behaviour. To verify that the architecture works properly, inside the ILA the trigger was set to the **ch0_rxgearboxslip_int_1** signal, which is a controller of the bit-slip process, and it raises every time the calibration fails, but it keeps its state "0" when the transmission is eventually calibrated. The ILA is set to search a "1" logic state for that signal, but as it is possible to see in the bottom left part of the figure, the IP core is not able to find the trigger state. Thus, the calibration in hardware emulation is reached.

The data transmitted when the architecture is calibrated and in "idle" state is a constant "0". It is actually possible to see that two signals in the transmission are different from the zero. Their timing distance is constant, and it coincides with the internal enables of the randomizer and de-randomizer. The time dedicated to fix

this issue was not sufficient to correct eventually the bugs, but since the problem relies only on the fact the the Virtex Ultrascale+ has recently entered in the global market. The internal architectures provided by the vendor requires a deep and long period of work, and almost the same time is required to understand to how correctly exploit the complete functionalities of the newer release of the firmware.

The 1-bit signals **diff_0_0** and **diff_1_0** have been generated just a control of the correct counter transmission. The are respectively checking if the difference between consecutive data is either the logic state "0" or "1". Thus, it represent a double control on the calibration (difference "0") and the counter transmission (difference "1"). The counter transmission in the architecture is shown in the ILA in Figure 4.9. In the figure, even if in hexadecimal base, a counter is shown in

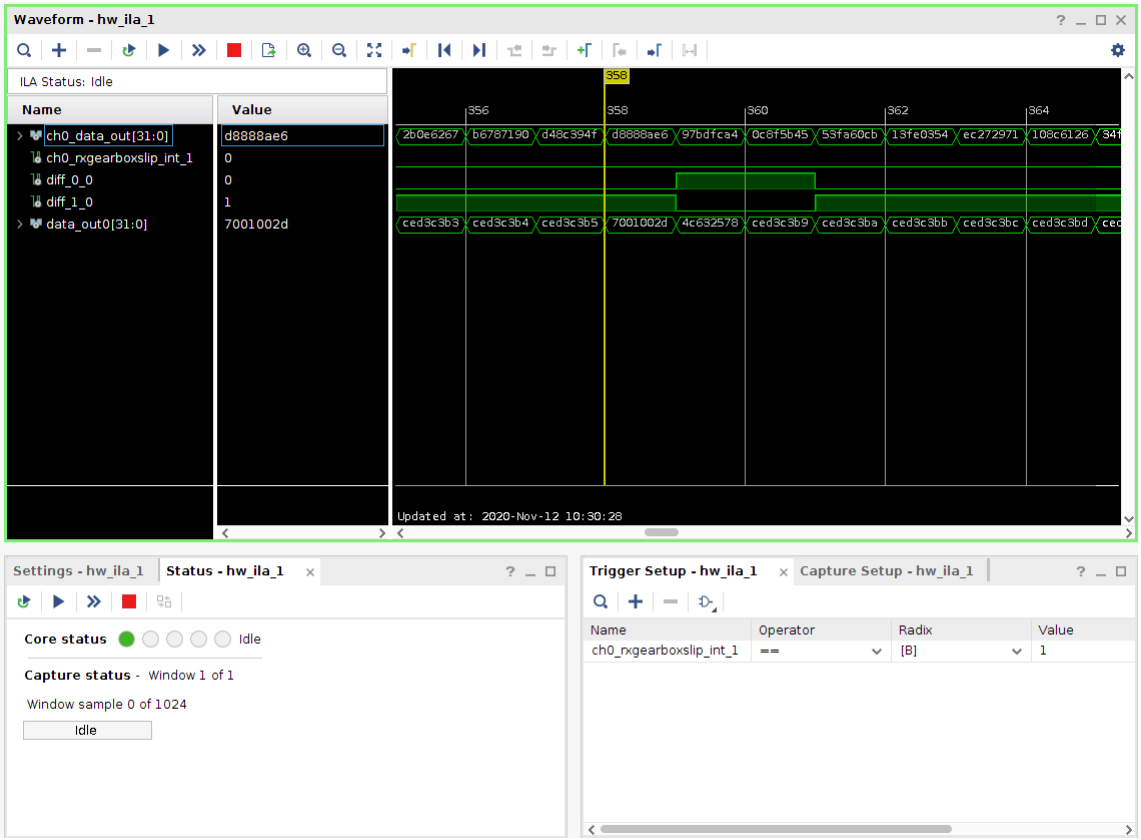

Figure 4.9: *View of the waveforms in the instantiated ILA, when the architecture starts to transmit data. It is possible to see the counter structure in the waveform of the* ***data_out0****. The signals shown are:* ***ch0_data_out****, the output data of the receiver,* ***ch0_rxgearboxslip_int_1****, the bit-slip controller,* ***diff_0_0*** *and* ***diff_1_0*** *which are simple checker of the correct transmission the* ***data_out0*** *is the actual data output of the receiver.*

the **data_out0** signal (including also the two wrong data previously mentioned). The counter behaviour is clear also because of the signals of the **diff_0_0** and **diff_1_0** "difference" checking signals, which behaves exactly as expected.

In conclusion, the whole architecture has been correctly simulated and emulated, even with minor internal bugs that will be surely fixed for the future tests of the HT hardware algorithm.

Chapter 5

Conclusion

5.1 Future developments

In the future years, the research and the innovation in silicon based electronic devices will receive a large acceleration process. The main reasons rely on the huge technological development which the research field has encountered in its process of evolution. Starting from the 1970s, the silicon metal-oxide-semiconductor (MOS) technology worked with junctions of the order of tens of µm, reaching the order of less than tens of nm in the latest years. All the integrated circuits in the high speed devices are built upon this growing technology, and year per year, new devices with better performances are launched in the global market.

The usage of FPGAs as hardware accelerators in high energy physics experiments has become over time a necessity in the whole setup of the build-up detector systems. The reasons why these devices are chosen as high speed processors are clear: they are fully programmable and controllable in their operations and most importantly, once their operational task is over, they can be re-programmed a re-used in different experiments with completely different needs. Although, FPGAs evolution goes hand in hand with the silicon technology, reaching new generations of devices at each new generation of semiconductor junction. It is indeed a good thing for experimental physicist, because in terms of performances new generation devices are able to work in borderline conditions, but at the same time a specific knowledge and capability to fully take advantage of such technologies is surely hard to achieve in a short time.

As an example, the proposals for the TDAQ system of the ATLAS experiment includes the development of a track reconstructor with several hardware components. The heart of the tracking reconstruction algorithm is being developed by the electronic research group in Bologna, and it is still currently being under several simulations and timing tests. My work inside this framework was the development of some track fitting algorithms for the Hough Transform (HT) hardware implementation, beside the development and testing of a set of pattern vectors used both in the simulation and validation of the HT algorithm itself and in the hardware integration on a FPGA-based hardware accelerator.

Once the work of the group would be completed, the Hough Transform (HT) algorithm designed in Bologna will undergo the validation through external func-

tional tests provided by the ATLAS collaboration and eventually the approval of the collaboration itself in case of satisfactory results. Surely, such researches and developments are a certain proof of the performance the technology provided by the FPGAs, not just in the ATLAS experiment, but in all the future experiments that will be proposed in the future years. Nowadays, it is almost completely indispensable for particular subsystems of the high energy physics detectors.

In conclusion, the presented work frame focused in the emulation and the functional testing of the high speed hardware device VCU1525 board (which embeds a Virtex UltraScale+ FPGA), is in perfect alignment with the future experiment's needs. The improved knowledge in development of firmware architectures has in fact become a task of major interest in high physics experiments. As long as large experiments would need a trigger system and a data acquisition system as basic requirements to handle the huge amount of data generated in a reasonable amount of time, the need for improvements in the hardware accelerators designs will be probably taken into account as of the optimal solution to work on.

Appendix A

Technical Appendix

A.1 FPGA

Field Programmable Gate Arrays (FPGAs) are reprogrammable integrated circuits, internally made of an array of reconfigurable logic states, subdivided in blocks. They are semicondutor flexible devices, with a high hardware-timed speed and reliability and by thei own nature they work in parallel way: different processing operations do not have to compete for the same resources, thus the performances of a single part are not affected by the addition of another .

FPGAs embed a two dimensional array of programmable logic blocks, and a hierarchy of reconfigurable interconnections which allow the "wiring" of the component blocks. An FPGA contains, in its integrated circuit (IC), millions of gates. The main components are the Configurable Logic Blocks (CLB), which includes the digital logic, inputs and outputs, the Programmable Interconnects (PI) which provide direction between the logic blocks to implement the use logic, and the input-output Blocks (I/OB), which are used to communicate with the FPGA from the outside world. A general scheme view is shown in Figure A.1.

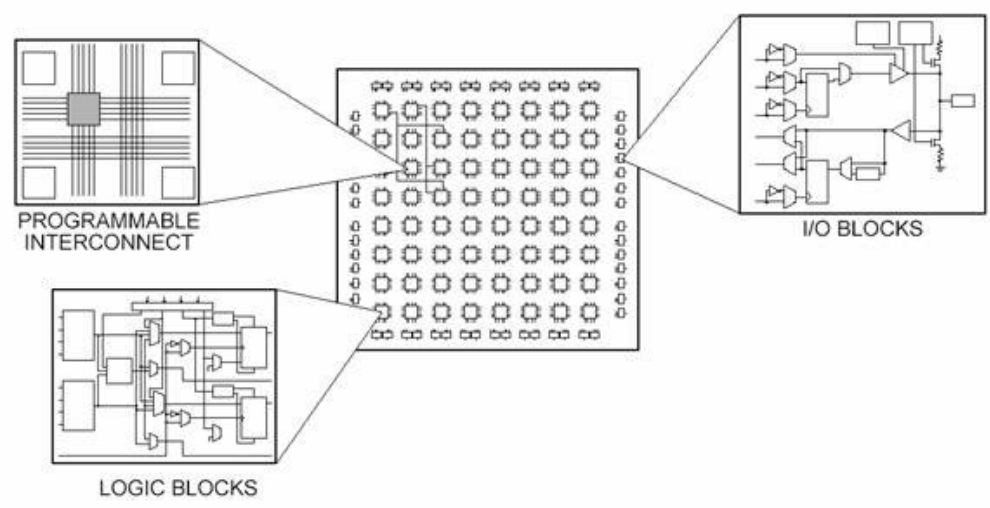

Figure A.1: *An FPGA structure design, with the internal component blocks: the Configurable Logic Blocks (CLB), the Programmable Interconnects (PI) and the input-output Blocks (I/OB).*

The idea of "reprogramming" an FPGA is different from the common concept of software programming: depending on the market vendor, the programming is done either by blowing interconnections fuses or by establishing links with anti-fuses, or by using static-RAM technology. Logic blocks are implemented using multiple level low fan-in gates, which can be configured even to emulate a microprocessor. Their implementation can be done by:

- transistor pairs;
- combinational gates like basic NAND or NOR gates;
- n-input Lookup Tables (LUT);
- multiplexers (MUX);
- wide fan-in AND-OR structure.

In general, a logic block consist of few logical cells, as in the example shown in Figure A.2. Inside an FPGA, when a design circuit have to be mapped with the internal components, these cells and most importantly the connections within these cells are considered as resources. In order to generate an efficient design, the internal resources should not be stressed to the limits of the FPGA resources.

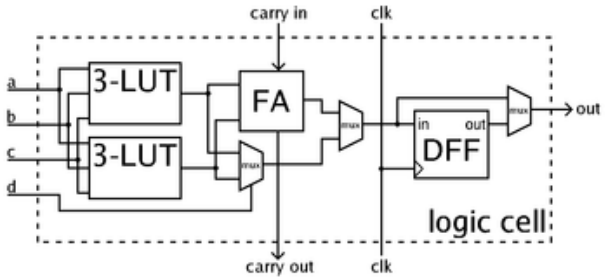

Figure A.2: *An example of a FPGA unit cell: the main sub-parts are the Lookup Tables (LUT), Full Adder (FA) and D-type flip-flop (DFF).*

Eventually, routing in FPGA consists of wire segments of varying lengths which can be interconnected via electrically programmable switches. The main configuration of an FPGA's arrays and components can be done in three possible ways:

- *Symmetrical arrays*, where CBLs are arranged in rows and columns of a matrix and interconnected by PIs. In the "periphery" of the FPGA the I/OBs provide the communications with the electronics right outside the array structure. Each CLB consists of an n-input LUT and a pair of programmable flip-flops. I/O blocks also control functions such as tristate control, output transition speed. Interconnects provide routing path, and obviously direct interconnects within adjacent logic blocks have smaller delays with respect to general purpose interconnect.
- *Row based Architecture*, where now the structure of logic modules is of alternating rows of logic modules and PI tracks. The I/O are at the end of

each row, and these last are connect each other vertically by the interconnect. Combinatorial modules contain only combinational elements while sequential modules contain both combinational elements along with flip-flops. Routing tracks are divided into smaller segments connected by anti-fuse elements between them.

- *Hierarchical Programmable Logic Device*, where the architecture is structured with a top level containing only logic blocks and interconnects. Each logic block contains number of logic modules and each logic module has combinatorial as well as sequential functional elements. Each of these functional elements is controlled by the programmed memory. Communication between logic blocks is achieved by Programmable Interconnect arrays. I/OB surround this scheme of logic blocks and interconnects.

A.2 JTAG

The hardware devices used in high energy physics researches, or in general for devices which works at very high rates, need a debug and testing structure safe and performing. During the 1980s, in order to achieve such debugging system, the Joint Test Action Group (JTAG) was formed, in order to provide a pinout access from the external world and verify the electronic components behaviour. The system is nowadays the standard option for electronic debugging, in particular the Standard Test Access Port (TAP) and the Boundary Scan (BS) structure.

The Boundary Scan

The main advantage offered by the BS technology is the capability to set and read the values on the pins of the FPGA without direct physical access. In fact, all the signals between the device's core logic and the pins are intercepted by a serial scan path known as the Boundary Scan Register (BSR), which consists of a number of boundary scan "cells". In normal operations these cells are invisible, however, in test mode the cells can be used to set and/or read values from the device pins (or in "internal" mode from values of the core logic).

The collection of boundary scan cells is configured into a parallel-in, parallel-out shift register. A parallel load operation, called a "capture" operation, causes signal values on device input pins to be loaded into input cells, and signal values passing from the core logic to device output pins to be loaded into output cells. Instead, a parallel unload operation, called an "update" operation, causes signal values already present in the output scan cells to be passed out through the device output pins.

Data can also be shifted around the shift register in serial mode. To activate this operation, the BS needs four different signals:

- the Test Clock (TCK) signal allows to synchronize the internal state machine operations;

- tha Test Mode Select (TMS) is a sampled signal at the rising edge of the TCK for the determination of the successive state;

- the Test Data In (TDI) is a signal which represents the data shifted *into* the device's test or programming logic (i.e, it is connected to an input pin). As the TMS, it is sampled at the rising edge of the TCK, while the internal state machine is in a correct state;

- Test Data Out (TDO) is a signal which represents the data shifted *out* the device's test or programming logic (i.e, it is connected to an output pin). Differently from the previous signals, it is sampled at the falling edge of the TCK, while the state machine is in a correct state.

In Figure A.3 a scheme of the BS and the TAP structure is shown. Thanks to

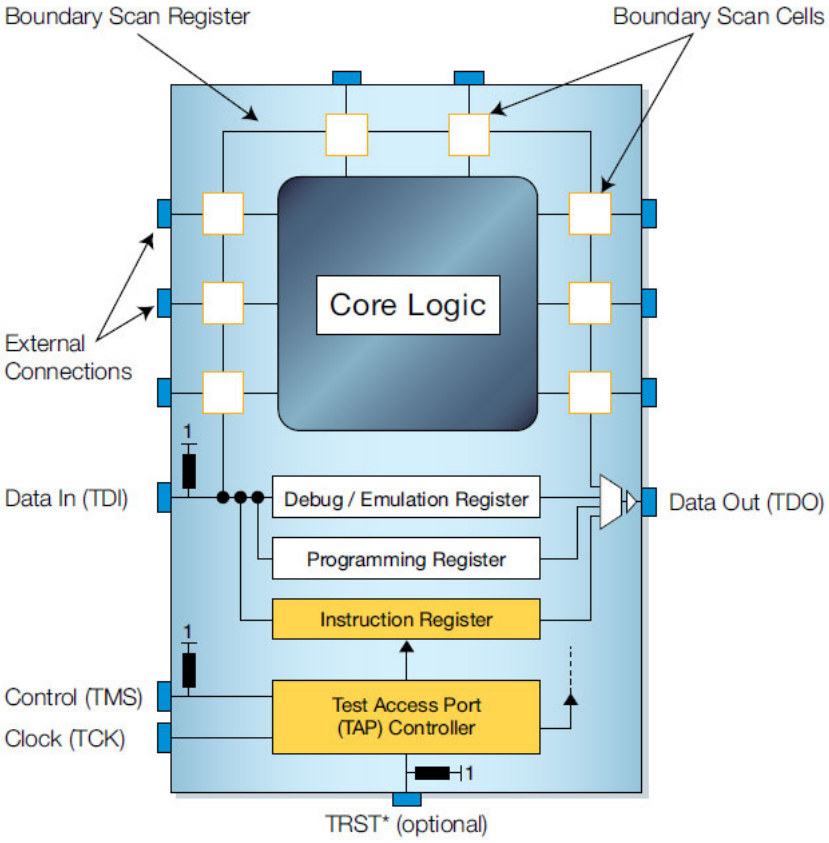

Figure A.3: *Scheme of the Boundary Scan structure. The main sub-structures are the BS cells, the various operational registers, the Test Access Port (TAP) and the connections with the external.*

the installed components inside the device, specific tests can be applied to the latter after the connection through the global scan path: loading the stimulus values into the appropriate device-output cells via the edge connector TDI (*shift-in operation*), applying the stimulus (*update operation*), capturing the responses at device-input scan cells (*capture operation*) and eventually shifting the response values out to the edge connector TDO (*shift-out operation*).

In fact, a single boundary scan cell has four modes of operation:

- Normal mode, in which DataIn is passed directly to Dataout;

- Update mode, in which the contend of the output register is passed through to DataOut;

- Capture mode, in which DataIn signal is routed to the shift register and the value is captured by the next ClockDR;

- Shift mode, in which the ScanOut of one register flip-flop is passed to the ScanIn of the next via a hard-wired path.

Notice that both capture and shift mode do not interfere with the normal passing of data from the parallel-in terminal to the parallel-out terminal. This allows the capture of operational values "on the fly" and the movement of these values for inspection without interference with functional modes of operation.

At last, the usage of the Boundary Scan cells to test the core functionality of a device is known as "internal test", while the test of the interconnections between two devices is called "external test".

Registers

The registers associated with the Boundary Scan and its functions are of two types: *instruction registers* and *data registers*. The standard devices embed in the BS one instruction register and at least two data registers, but possibly more of these ones.

The instruction register holds the "current" instruction that the BS receives, its content is used by the Test Access Port (TAP) controller to decide what to do with signals that are received. Most commonly, the content of the instruction register will define to which of the data registers signals should be passed. Instead, for what regards the data registers, the main one is the Boundary Scan and it is used to move data from the input-output pins of a device. Another one is a single-bit register which passes information from TDI to TDO, named BYPASS, and it allows the test of other devices with minimal overhead. Moreover, a register is used for the identification of the code and the revision number for the device, it is called IDCODES. The information inside this register allows the device link to the Boundary Scan Description Language (BSDL), and it also contains the details on the BS configuration for the device.

Test Access Port Controller

The Test Access Port (TAP) controller, a state machine whose transitions are controlled by the TMS signal, controls the behaviour of the JTAG system. In Figure A.4 is shown the state-transition diagram. All states have two exits, so all transitions can be controlled by the single TMS signal sampled on TCK. The two main paths allow to set or retrieve information from either a data register or the instruction register of the device. The data register operated on (e.g. BSR, IDCODES, BYPASS) depends on the value loaded into the instruction register.

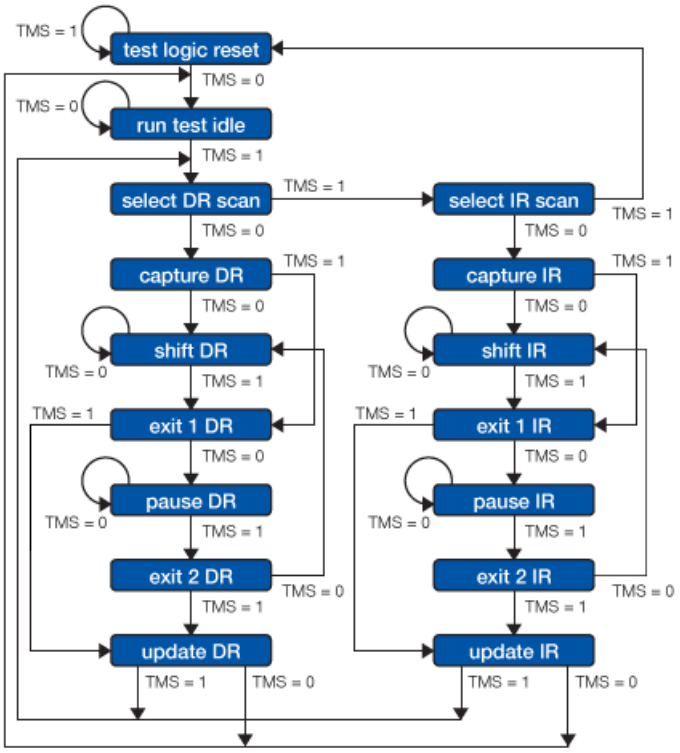

Figure A.4: *Scheme of the state transition of the Test Access Port (TAP) controller.*

Boundary Scan instructions

The IEEE 1149.1 standard defines a set of instructions that must be available for a device to be considered compliant:

- BYPASS: instruction which causes the connection of TDI and TDO lines via a single bit pass-through register (the BYPASS register). This instruction allows the testing of other devices in the JTAG chain without any unnecessary overhead;

- EXTEST: this instruction causes the connection of TDI and TDO to the Boundary Scan Register (BSR). The device's pin states are sampled with the "capture dr" JTAG state and new values are shifted into the BSR with the "shift dr" state. These values are successively applied to the pins of the device using the "update dr" state;

- SAMPLE/PRELOAD: this instruction connects the TDI and TDO to the BSR. However, the device is left in its normal functional mode. With the instruction, the BSR can be accessed by a data scan operation to take a sample of the functional data entering and leaving the device. The instruction is also used to preload test data into the BSR prior to load an EXTEST instruction.

A.3 I²C protocol

The Inter-Integrated Circuit (I²C) is a synchronous serial protocol for two wire interface, which connect through a master-slave hierarchy low-speed electronic devices embedded in larger systems. It is indeed widely used to connect peripheral Integrated Circuits (ICs) to processors and microcontrollers in a relatively "short" distance intra-board.

The protocol uses only two signals for the connection: the *Serial CLock* (SCL), an output signal of the master devices, providing the other devices the functioning clock, and the *Serial DAta* (SDA), an input-output wire which provides the actual data exchange within the devices. An example of the structure of a typical I²C connection is given in Figure A.5.

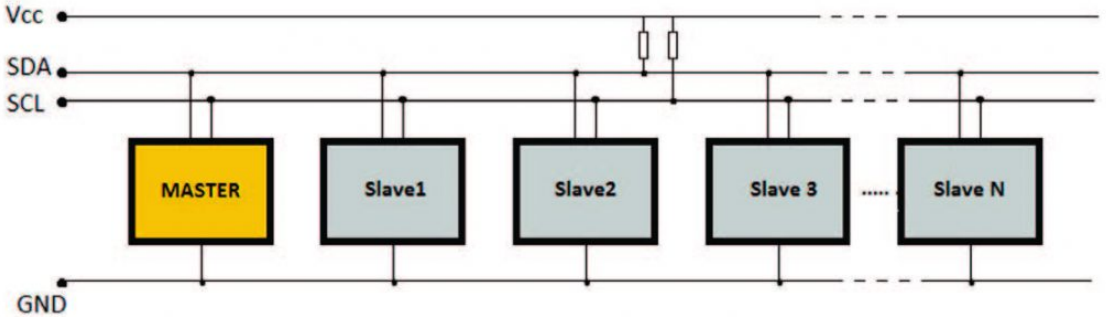

Figure A.5: *A simplified scheme on how the I²C communication structure and components. A master (or even more than one) can control an arbitrary number of slaves with different address (limited of course by the device resources). The communication between the components is established only through the wires Serial DAta (SDA) and Serial CLock (SCL).*

The drivers which uses the I²C bus are "open drain", which means that SDA can be changed only from a high state to a low one. Thus, the contention between the devices which try to drive high and the ones which try to pull low is eliminated, reducing potential dames to the drivers or the excessive power dissipation. Besides, I²C allows to connect devices with different input-output voltages: in general, in a system where one device is at a higher voltage than another, it may be possible to connect the two devices via I²C without any level shifting circuitry in between them.

The initial I²C specifications defined a a maximum clock frequency of 100 kHz, but it was later increased up to 400 kHz as a fast mode. For the latest implementations of the protocol, also high speed mode (up to 3.4 MHz) and a ultra-fast mode (up to 5 MHz) have been developed.

The master-slave structure in its simplicity has granted a large success for this data protocol. In fact, each slave device has a unique address, typically identified with a sequence of 7-bit. The information transfer from and to the master device is serial and it is split into 8-bit sequences. These are the only requirements to initialize such a data protocol, but paying attention in particular sequences that have a key role in the data transfer.

The actual data transmission protocol is shown Figure A.6. In details, the

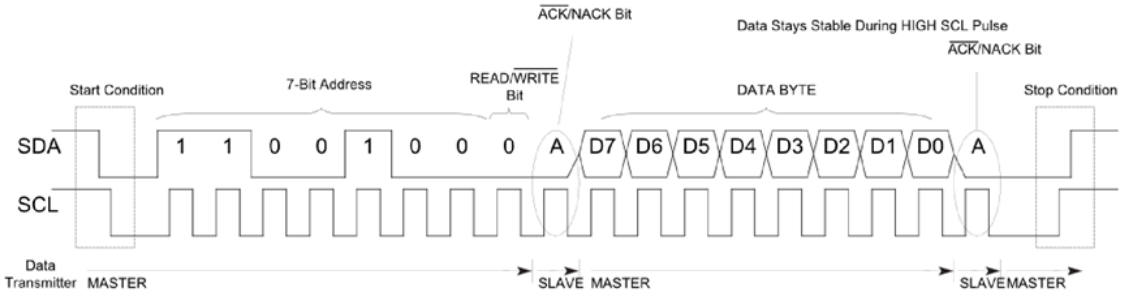

Figure A.6: *A typical I^2C data transmission in its main phases.*

transmission may be divided into several phases of operation:

- *Idle* state: it is the "no-operation" currently running in the bus of the I^2C, in particular neither *start* nor *stop* conditions have been identified by the device, and both SDA and SCL are in a high logic state;

- *Start* condition: it corresponds to a falling edge (i.e., a high-to-low state transmission) of the SDA signal, while the SCL is in a high state. In the multiple master case, the master device which makes the Start condition first, wins the possibilities to control the slave or slaves;

- *Address slave* frame: the master sends an 8-bit word to the slave, containing in the first 7 bits (starting from the most significant bit, MSB) the physical *address* of the target slave and a last bit, the least significant bit LSB, which defines the operation to be done in the successive phase of the transmission. If the LSB is a "0", the master will send a word to be written inside the slave, otherwise the master will read from the slave a word in a given address;

- *Address register* frame: in this phase, the master sends an 8 bits frame which represents the location of the register where the data either is written from the master or is read from the slave. The transmission is always from the MSB to the LSB;

- *Data* frame: this is the 8 bits data written from master to slave or read from slave to master, in function of the value of the LSB of the *address slave* 8-bit word;

- *Acknowledge* bit: after every transmission frame from master to slave, there is a bit sent from the slave (so the slave takes the SDA control) which determines if the transmission has succeeded, telling it by sending a "0" (acknowledge) or otherwise a "1" (not acknowledge);

- *Stop* condition: after all the previous operations, there is a *stop* condition, implemented by a transition from a low state to a high one of SDA while SCL is in a high state. In case of a read command, an *acknowledge* bit is not sent from the slave before the stop condition, but from the master it is sent a "not acknowledge" value to the slave.

Finally, despite the I^2C is an "aged" communication system, thanks to its simplicity, the protocol is still used and implemented in hardware devices, including the high performance ones.

A.4 Aurora protocol

Aurora 64B/66B is a link-layer protocol that can be used to move data point-to-point across one or more high-speed lane connectors. The protocol describes how to transfer user data through serial mode connection using Aurora channels, consisting of one or more Aurora lanes. The transfer can be full-duplex or simplex. A scheme of the functional protocol is shown in Figure A.7.

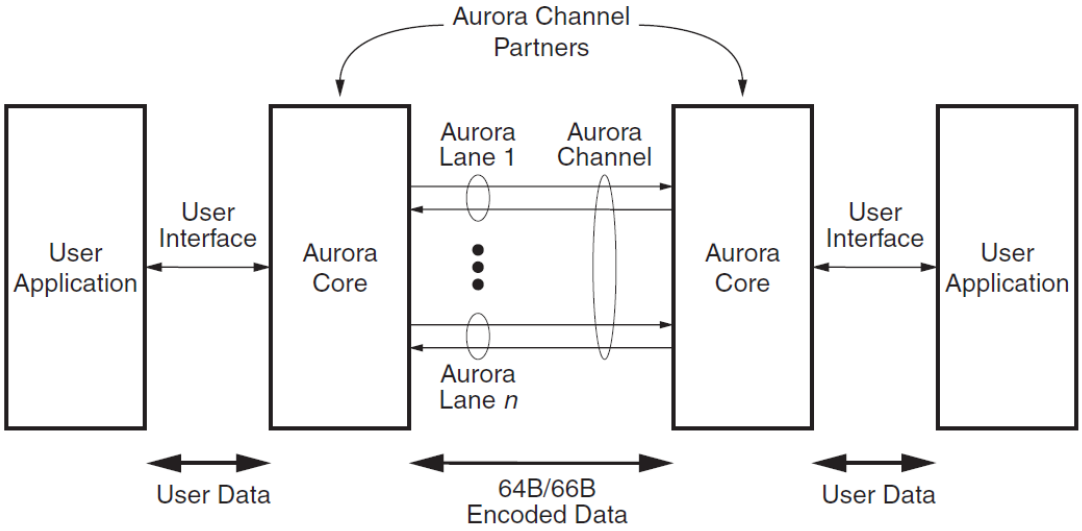

Figure A.7: *Overview of the Aurora 64B/66B data transfer protocol.*

In the Aurora channels, data are transferred in frames, which share the same channel with control information as flow control messages, clock-compensation sequences and idles. Also, frames can be of any possible length an can have any possible format, and they can be interrupted at any time by the flow control messages or idles.

Data transmission and reception

The data transmission in the Aurora protocol is performed with 64-bit codes called *blocks*, where each of these blocks can be transmitted through an Aurora channel. For this reason, block are prioritized, to resolve possible conflicts which can occur if two or more blocks are sent in the same lane. The main blocks are:

- Clock Compensation: it is an idle block which can be used to prevent data corruption due to small differences between recovered clock and local reference clock. This problem could occur if different clock sources are used to drive TX and RX clocks;

- Not Ready: it is an idle block sent while attempting to align data from the channel and performing channel bonding;
- Channel Bonding: it is an idle block ,used to bond the channel, sent to every lane in the channel simultaneously, and the receiver lanes receive this block if their channel bonding FIFOs are adjusted;
- Native Flow Control: it requests native flow control from the Aurora interface on the other side of the channel;
- User Flow Control: the same as the previous one, but the flow control messages are customized by the user;
- User K-Block: it is a not decoded block which is passed directly to the user and can be implemented with specific control application functions;
- Data: this block, with the successive two, creates frames carrying user data. In particular this block carries eight octets of data;
- Separator: this block indicates the end of the current frame; the next frame begins in the next block. Separator blocks carry from 0 to 6 octets of data;
- Separator-7: it does the same of Separator, but always carries 7 octets of data;
- Idle: it is transmitted when no other higher priority blocks can be transmitted. It is the lowest priority block.

The transmission and reception procedure of a frame is shown in Figure A.8. Through the user application, the procedure of a frame transmission to an initialized

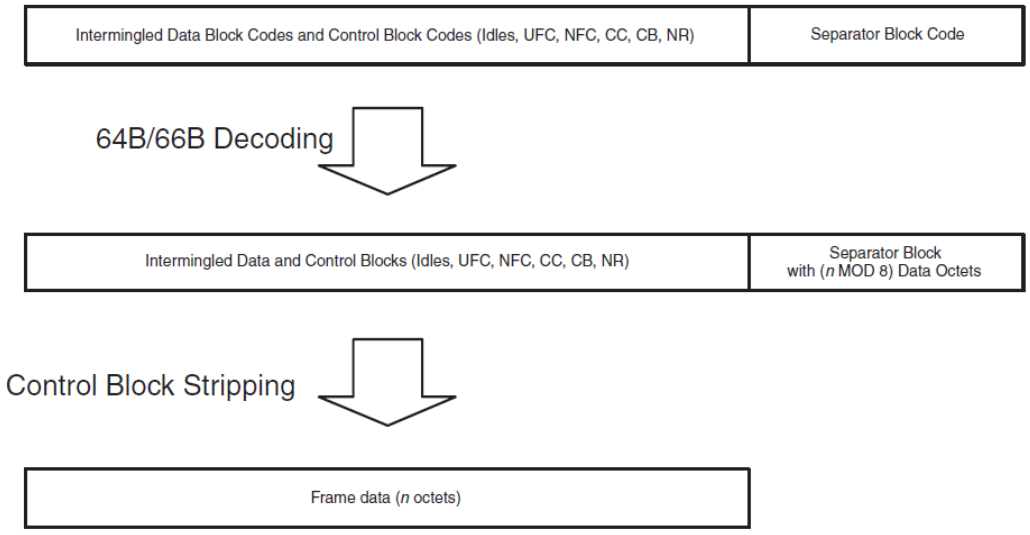

Figure A.8: *Scheme of the transmission and reception procedure of a frame of the Aurora 64B/66B protocol.*

Aurora channel is the following:

- frames are delineated at their end using Separator and Separator-7 blocks, fact that allows the channel partners to distinguish between different frames;
- data and separator block are 64b/66b encoded by a Physical Coding Sub-layer (PCS) prior the transmission, transforming 64-bit blocks to 66-bit blocks;
- the encoded blocks, comprising the frame, are serialized for transmission, using a differential non-return-to-zero (NRZ) format.

A.4.1 PCS Layer and PMA Layer

The physical layers of the Aurora protocol are shown in Figure A.9. They are the PCS and the PMA layers.

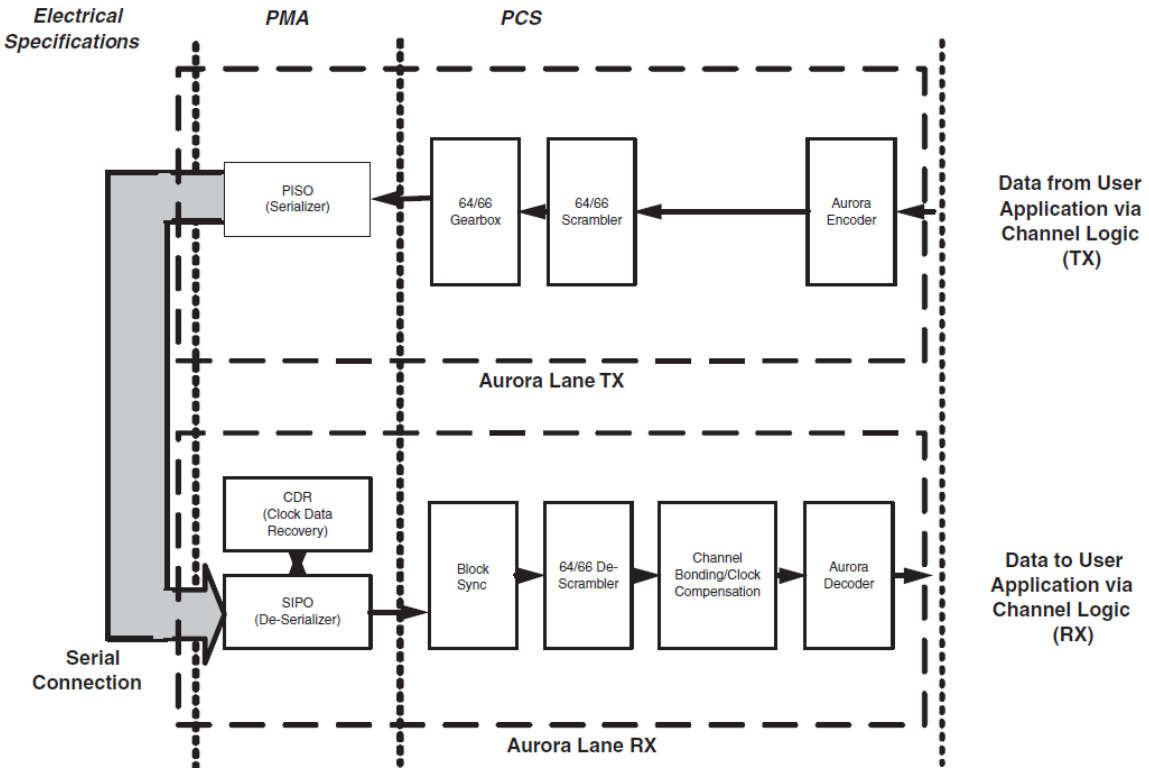

Figure A.9: *Scheme of Aurora 64B/66B logic blocks and physical layers.*

PCS

In the *PCS layer* is present the Aurora Encoding: it specifies how data and control information must be encoded before transmission through an Aurora channel, and decoded upon reception. In fact, all data and control information in Aurora 64B/66B are encoded in 64-bit blocks and each 64-bits block is marked with a sync header value, indicating whether or not it is either Data or Control block.

In addition, it embeds the 64B/66B scrambling, which operates as it follows: whenever a lane transmits a block code, the 8 octets, following 2 bits sync header, must be scrambled using a self-synchronizing scrambler with the polynomial $G(x) =$

$1 + x^3 9 + x^5 8$. When a lock code is received, the 8 octets, following 2 bits sync header, must be de-scrambled using a self-synchronising descrambler with the same polynomial. The 2 bits sync header don't undergo the scrambling or the de-scrambling.

Data at 64B/66B are transmitted with a gearbox: after exits from the scrambler, the gearbox combines the 64 bits block and the 2 bits sync header to present a 66 bits block to the PMA layer. Otherwise, before entering the de-scrambler, the gearbox separates the 66 bits block to send at the de-scrambler the 64 bits block and the 2 bits sync header.

Eventually, the layer has Clock Compensation (CC) idle blocks can be used to prevent data corruption due to small differences between the recovered clock and local reference clock. These differences occur when independent clock sources are used to drive the clocks on the TX and RX side of a connection. If a shared clock is used, Clock Compensation blocks are not required.

PMA

On the other hand, the PMA layer includes a *Bit and Byte Ordering Convention*: the leftmost bits of the encoded block code are the sync header bits. These are the most-significant bits of the block code. The leftmost byte of the block is the most significant byte. The rightmost byte is the least-significant byte.

In fact, each active lane must transmit the most-significant bit of each block code first, followed in order by the remaining sync bit and the bits of the block code (*serialization*, Figure A.10). Each lane should expect the data that it de-serializes

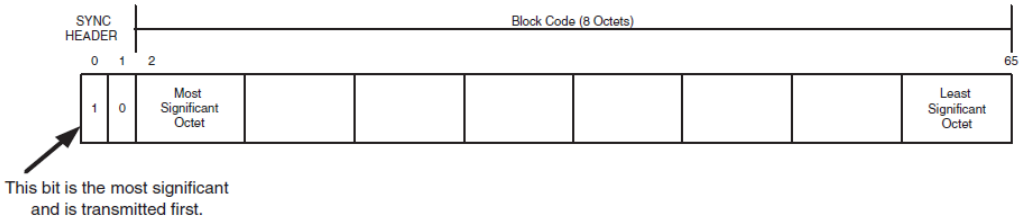

Figure A.10: *Ordering of Aurora 64B/66B data serialization.*

to be in the serialization order presented in "Bit and Byte Ordering Convention".

In this layer, is present the clock data recovery: from incoming data transmission streams a high speed serial clock is recovered.

Bibliography

[1] K. A. Mohammad Kamal Azmi, W. A. T. Wan Abdullah, and Zainol Abidin Ibrahim. "Hough transform method for track finding in center drift chamber". In: *AIP Conference Proceedings* 1704.1 (2016), p. 030014. DOI: 10.1063/1.4940083. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.4940083. URL: https://aip.scitation.org/doi/abs/10.1063/1.4940083.

[2] M. Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. Sept. 2010. URL: http://cds.cern.ch/record/1291633.

[3] ATLAS Collaboration. *ATLAS Liquid Argon Calorimeter Phase-II Upgrade Technical Design Report*. Tech. rep. CERN-LHCC-2017-018. ATLAS-TDR-027. Geneva: CERN, June 2018. URL: https://cds.cern.ch/record/2285582.

[4] ATLAS Collaboration. *Phase-II upgrade Scoping Document*. Tech. rep. CERN-LHCC-2015-020. LHCC-G-166. Geneva: CERN, Sept. 2015. URL: https://cds.cern.ch/record/2257755.

[5] ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*. Tech. rep. CERN-LHCC-2017-021. ATLAS-TDR-030. Geneva: CERN, Sept. 2017. URL: https://cds.cern.ch/record/2285585.

[6] ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer*. Tech. rep. CERN-LHCC-2017-017. ATLAS-TDR-026. Geneva: CERN, Dec. 2017. URL: https://cds.cern.ch/record/2285580.

[7] ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter*. Tech. rep. CERN-LHCC-2017-019. ATLAS-TDR-028. Geneva: CERN, June 2018. URL: https://cds.cern.ch/record/2285583.

[8] ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS Trigger and Data Acquisition System*. Tech. rep. CERN-LHCC-2017-020. ATLAS-TDR-029. Geneva: CERN, June 2018. URL: https://cds.cern.ch/record/2285584.

[9] ATLAS Collaboration. *Technical Design Report: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*. Tech. rep. CERN-LHCC-2020-007. ATLAS-TDR-031. Geneva: CERN, July 2020. URL: https://cds.cern.ch/record/2719855.

[10] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.8 (Aug. 2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[11] V Halyo V Halyo et al. "First evaluation of the CPU, GPGPU and MIC architectures for real time particle tracking based on Hough transform at the LHC". In: *Journal of Instrumentation* 9.04 (Apr. 2014), P04005. DOI: 10.1088/1748-0221/9/04/p04005. URL: https://doi.org/10.1088%5C%2F1748-0221%5C%2F9%5C%2F04%5C%2Fp04005.

[12] M. Mårtensson. "A search for leptoquarks with the ATLAS detector and hardware tracking at the High-Luminosity LHC". PhD thesis. Uppsala U., 2019.

[13] M. Mårtensson. "Fast pattern recognition with the ATLAS L1Track trigger for the HL-LHC". In: *Proceedings of Science* (Jan. 2017), p. C12036. DOI: MÃërtensson. URL: ATL-DAQ-PROC-2016-034.

[14] Nicola Pozzobon, Fabio Montecassiano, and Pierluigi Zotto. "A novel approach to Hough Transform for implementation in fast triggers". In: *Nuclear Instruments and Methods in Physics Research A* 834 (Oct. 2016), p. 81. DOI: 10.1016/j.nima.2016.07.020.

[15] L. Rinaldi et al. "GPGPU for track finding in High Energy Physics". In: *arXiv: Instrumentation and Detectors* (2015), p. 17.

[16] A. La Rosa. "The ATLAS Insertable B-Layer: from construction to operation". In: *Journal of Instrumentation* 11.12 (Dec. 2016), p. C12036. ISSN: 1748-0221. DOI: 10.1088/1748-0221/11/12/c12036. URL: http://dx.doi.org/10.1088/1748-0221/11/12/C12036.

[17] XILINX. *Aurora 64B/66B Protocol Specification*. Tech. rep. San Jose: XILINX, Oct. 2014. URL: https://www.xilinx.com/support/documentation/user_guides/ug578-ultrascale-gty-transceivers.pdf.

[18] XILINX. *UltraScale Architecture GTY Transceivers* User Guide. Tech. rep. San Jose: XILINX, Sept. 2017. URL: https://www.xilinx.com/support/documentation/user_guides/ug578-ultrascale-gty-transceivers.pdf.

[19] XILINX. *VCU1525 Reconfigurable Acceleration Platform* User Guide. Tech. rep. San Jose: XILINX, Mar. 2019. URL: https://www.xilinx.com/support/documentation/boards_and_kits/vcu1525/ug1268-vcu1525-reconfig-accel-platform.pdf.

[20] XILINX. *Virtex UltraScale+ FPGA Data Sheet: DC and AC Switching Characteristics*. Tech. rep. San Jose: XILINX, Aug. 2020. URL: https://www.xilinx.com/support/documentation/data_sheets/ds923-virtex-ultrascale-plus.pdf.