Design and characterisation of monolithic CMOS detectors for high energy particle physics and SEU radiation tests for ATLAS Inner Tracker Upgrade readout chip

Leyre Flores Sanz de Acedo

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Physics and Astronomy
College of Science and Engineering
University of Glasgow

July 2023

Abstract

This thesis covers the characterisation results and the design of monolithic CMOS detectors designed in TowerJazz 180nm CMOS technology for High Energy Particle Physics applications. Three different detectors have been studied the MALTA, the Mini-MALTA and the MALTA2. The MALTA sensor showed some efficiency losses at the corners of the pixels after irradiation, which meant that it was not suitable for the radiation environments in which it was supposed to be installed. Therefore, the front-end electronics and the fabrication process were modified to overcome this issue. The Mini-MALTA prototype was designed including the above mentioned improvements, fabricated and fully characterised. Finally taking into account all the knowledge acquired during these years of developments another large scale sensor the MALTA2 has been produced which should be radiation tolerant and have very good time resolution. The description and studies of the different architectures used in this family of detectors are covered and a simulation to estimate the bandwidth capabilities have been reported.

Furthermore, this work will present characterisation of single event effects in the ITkPixV1, the prototype version of the ATLAS Inner Tracker Upgrade chip for the High Luminosity LHC. Measurements were made in testbeam campaigns with high energy ions and protons to evaluate the level of single event effects in the chip.

Contents

List of Tables

List of Figures

Acknowledgements

Fist, I would like to thank my supervisor at the University of Glasgow, Professor Craig Buttar for his support during the whole PhD and for letting me pursue the topics I was interested in. I would also like to thank my supervisor at CERN, Dr. Heinz Pernegger for his help and guidance throughout this PhD project. Working with you has been an honour and an unforgettable experience. Thanks to Carlos Solans Sanchez, Valerio Dao, Walter Snoeys, Xavi LLopart, and everyone else in the EP-ADE-TK and EP-ESE-ME groups who had made this work possible by leading the main activities. Thanks also to my office mates, Ivan Berdalovic, Roberto Cardella, Francesco Piro and Dominik Dobrijevic for endless discussions and to my lab companions Ignacio Asensi, Enrico Junior, Milou Van Rijinbach, Patrick Freeman, Andrea Gabrielli and Florian Dachs for all the testing instructions, code development, explanations and graphs provided. Finally, thanks to my family and partner, who have supported me on every step of the way, even from the distance and during pandemic times.

Declaration

Chapters 1 and 2 are introductory material to CERN and to detector physics. Chapter 3 introduces the detector FE electronics and the sensor design that were done by the CERN/STREAM collaboration and it shows characterisation results. The cluster analysis for the Mini-MALTA and the CZ MALTA and the efficiency plots are my work and were the basis for further analysis by the group for publications and conferences. The data was taken during several test-beam campaigns in which I participated. I performed the laboratory measurements of occupancy and average cluster sizes on the mini-MALTA and analysed the data. Concerning chapter 4 the explanation of the readout architecture is based on the original design as well as on new findings done to understand it better. The architecture simulation was carried out by myself based on data-sets provided by ATLAS experiment with the goal of understanding the bandwidth limitations of the devices. Chapter 5 presents SEU characterisation results for ITkPixv1, the ATLAS detector foreseen for the HL Upgrade of the Inner Tracker. All this work was undertaken in collaboration with CPPM institute in Marseille and another colleague from CERN. The preparation of apparatus, test routines and analysis are the work done for this thesis. Finally chapter 6 introduces the new design that I did for the MALTA2 slow control and how it was verified and tested after fabrication.

Chapter 1

CERN, the Large Hadron Collider and ATLAS Upgrade

This is an introductory chapter that puts into context where the work of this thesis took place. It introduces CERN (Conseil Européen pour la Recherche Nucléaire) as an institution, the LHC (Large Hadron Collider) and gives an overview of the ATLAS detectors as the work done was intended to be in this experiment.

1.1 Introduction to CERN

The theoretical model that describes elementary particles and their interactions is known as the Standard Model (SM). This model has been widely tested for more than 30 years and its predictions closely match most experimental observations. Nevertheless, some observed phenomena have yet to be explained by the SM with the aim of looking for currently unknown physics. Accelerators are being used all over the world and the most powerful currently available can be found at CERN.

The European Organization for Nuclear Research, most commonly known by its French acronym CERN (Conseil Européen pour la Recherche Nucléaire), is a research organisation that operates the largest particle physics laboratory in the world. It was established in 1954 and it is located in the suburbs of Geneva (Switzerland) on the French border. It has 23 member states and is also supported by many other countries and organisations via cooperation agreements.

CERN's primary research concerns fundamental particle physics to uncover what constitutes the universe and how it works. Its programme covers topics ranging from the basic structure of matter to cosmic rays and from the SM to super-symmetry. However, as a consequence of all the developments made during experiments, the laboratory also plays a vital role in developing technologies for the future.

A list with some of the main milestones and discoveries of CERN during its existence are in chronological order: the weak neutral currents that helped unify electromagnetism and the weak

force (1973), the measurement of W and Z bosons, elementary particles that mediate the weak force (1983), antimatter hydrogen which is made with an antiproton and positron was created at CERN (1995) and recently the detection of the Higgs boson, which is a manifestation of the Higgs field that generates mass through its interaction with other particles (2012).

1.2 The standard model

The Standard Model of particle physics is an accurate framework that predicts the behaviour of fundamental particles and their interactions. Figure 1.1 shows the list of the fundamental particles in the Standard Model. These particles can be classified firstly as particles with half-integer spin or fermions that follow Fermi-Dirac statistics and bosons which are exchange particles enabling the four fundamental forces. The matter particles can be further subdivided into quarks and leptons. The gauge carriers of the forces are: the photon, the $W^{\pm}$ and Z^0, and the gluons for the electromagnetic, weak and strong interaction respectively. It is important to mention that the Standard Model does not include a description of gravity. Similarly, it presents shortfalls in the description of the matter-anti-matter asymmetry in the universe and it does not provide a dark matter candidate.

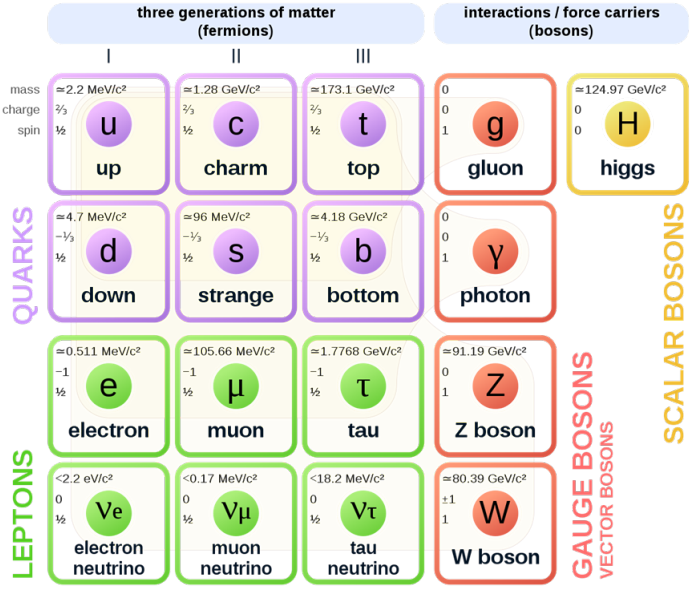


Figure 1.1: The Standard Model of Particle Physics.

1.3 LHC: the accelerator complex, the experiments and main parameters

The Large Hadron Collider (LHC) [1] is the largest and most powerful particle accelerator in the world. It was built by CERN and it consists of a 27km ring of superconducting magnets with several acceleration stages that boost the energy of the particles that travel along the accelerator. It is situated 100 m beneath the Franco-Swiss border. Protons are accelerated through a complex of accelerators shown in figure 1.2, in two counter rotating beams that reach an energy of 6.5 TeV in the LHC. The protons are produced by the ionisation of hydrogen gas inside an electric field. In the linear accelerator (LINAC), the kinetic energy of the proton is increased from 100 keV to 1.4 GeV, using radio-frequency cavities and the Proton Synchrotron Booster (PSB), before arriving to the Proton Synchrotron (PS). The PS and the Super Proton Synchrotron (SPS) facilities increase the energy of the protons to 25 GeV and 450 GeV respectively. Finally, the beams are injected into the LHC for further acceleration, from 450 GeV to 6.5 TeV. The superconducting magnets in the tunnel are cooled to temperatures below 2 K and are operated at magnetic field strengths above 8 T, to bend the particles in the beam around the circular collider.

Beam collisions take place at four different places around the LHC, where the four biggest experiments are located ATLAS [4], CMS [6], ALICE [3] and LHCb [5]. ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) are multi-purpose detectors designed to have a wide sensitivity to many physics processes. They are situated at opposite points in the LHC ring. Designing both experiments independently is of primary importance because it allows for cross-confirmation of measurements and possible new discoveries. The ALICE (A Large Ion Collider Experiment) detector was built to study heavy ions collisions in order to improve the knowledge of the dynamics of quarks and gluons at high energy density. LHCb (Large Hadron Collider beauty), sits facing ALICE , and it is a one arm spectrometer designed to cover forward region to perform detailed studies in the field of B-physics. B-physics is a category within particle physics that studies the properties of B-hadrons which are the ones that contain at least one bottom quark. There are two smaller experiments on the LHC, TOTEM and LHCf, that focus on forward physics.

The number of events generated per second in the LHC collisions is given by:

$$N = L\sigma \tag{1.1}$$

where σ is the cross-section of the event in question and L is the machine luminosity. L is entirely dependent on the beam parameters and can be expressed as:

$$L = \frac{N_b^2 n_b f_{rev} \lambda_r}{4\pi\varepsilon_n \beta^*} F \tag{1.2}$$

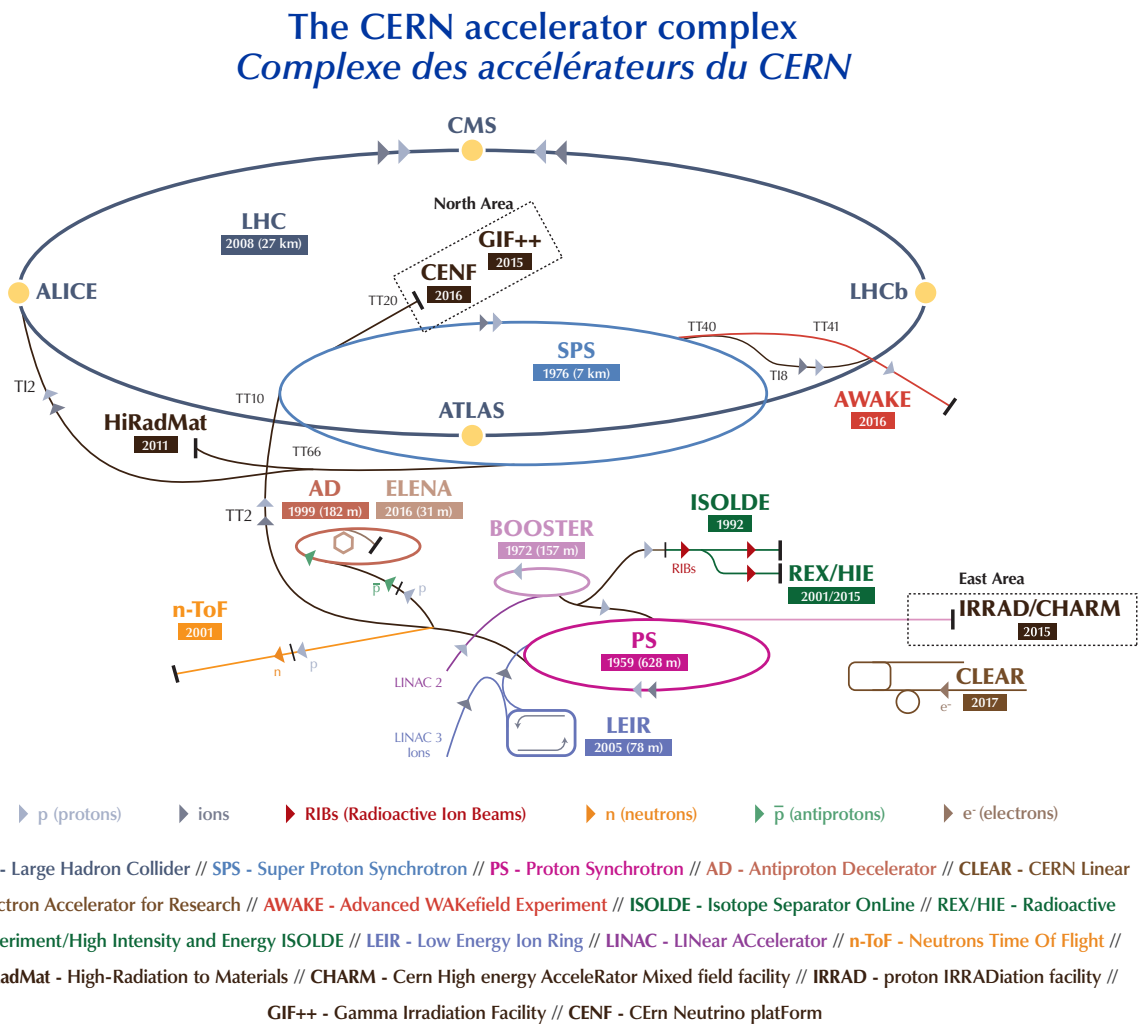


Figure 1.2: The CERN accelerator complex [9].

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, λ_r the relativistic gamma factor, ε_n the normalised transverse beam emittance, β^* the beta function at the collision point and F the geometric luminosity reduction factor due to the crossing angle at the interaction point [7]. The majority of the physics events that are studied in the LHC, such as decays or particles like the Higgs, occur very rarely. Therefore having both high beam energies and high beam intensities are crucial. For the two largest experiments, ATLAS and CMS the peak operational luminosity is around $L = 10^{34}cm^{-2}s^{-1}$ for proton collisions while in the case of lead-lead ion operation for ALICE the luminosity is $L = 10^{27}cm^{-2}s^{-1}$. The beams are not continuous, the protons are bunched together so the interactions between the two beams take place at discrete intervals, every 25 ns, meaning that the bunch collision rate is 40 MHz. The particles created in these collisions, as well as the decay products of particles with a short lifetime, pass through a variety of detectors within the experiments. This provides information about the charge, momentum and energy of the particles created.

1.4 ATLAS detector

The ATLAS detector [4] is one of the two general purpose detectors. It was designed and optimised to maximise the discovery potential for new physics beyond the SM as well as to perform high precision measurements of established physics. ATLAS is the largest detector at the LHC with a diameter of 25 m, a length of 46 m and a weight of roughly 7000 tonnes. It consists of several sub-detectors that provide the data for the event reconstruction. Those subsystems are: The Inner Detector, Electromagnetic Calorimeter, Hadronic Calorimeter and the Muon Chambers. Momentum measurements are possible due to a large magnetic field permeating through the experiment. This magnet system is divided into a solenoid magnet providing a 2 T field at 2.4 m radii and 3 toroidal magnets. One of these is in the barrel region, made out of 8 superconducting air-core toroidal magnets that give ATLAS its characteristic shape averaging 0.8 T, and two are in the end-cap (forward) regions averaging 1.3 T. The tracking system provides efficient track reconstruction up to high luminosities to enable the measurement of high transverse momentum leptons, identification of electron, τ-lepton and charged particles. In order to achieve the best performance the tracking system should cover a large region in pseudorapidity (η) and the full azimuthal angle ϕ. The coordinate system used in ATLAS is a right-handed system using cylindrical coordinates r and ϕ (azimuthal angle around the z-axis) in the transverse plane, its centre point being located at the interaction point in the middle of the detector and the z-axis along the beam pipe. For the x and y axis the LHC is taken as frame of reference, where the x-axis points towards the centre of the LHC and the y-axis points from the ring towards the surface. The pseudorapidity is defined as $\eta = -ln(tan\frac{\theta}{2})$ with θ being the polar angle with respect to the beam direction. The calorimeter system is optimised for electron and photon identification as well as for precise measurements of jets and missing transverse energy, while the muon system provides high-precision measurements of muon tracks. Figure 1.3 shows the detector and its parts.

1.4.1 ATLAS Inner Detector

The ATLAS Inner Detector is designed to reconstruct tracks and decay vertices in any event with high efficiency and measure the momentum of charged particles. It forms a cylinder with a length of 7 m and a radius of 1.15 m and consists of three different systems, each optimised for specific requirements according to their distance from the interaction region. There are two silicon-based systems closest to the interaction point, the Pixel Detector and the Semiconductor Tracker (SCT). The original pixel detector was made out of 4 barrel layers between 50.5 and 120.5 mm in radius, and 3 disks in each of the two end-cap regions. The barrel has 1744 modules, defined as structures that contain the sensor and chip, with 46080 read-out channels each, and a pixel size of 50×400 μm, with a total of 80 million channels, covering an area of 1.7 m^2. Every endcap disk is formed of 288 modules. The power consumption is 15 kW subject

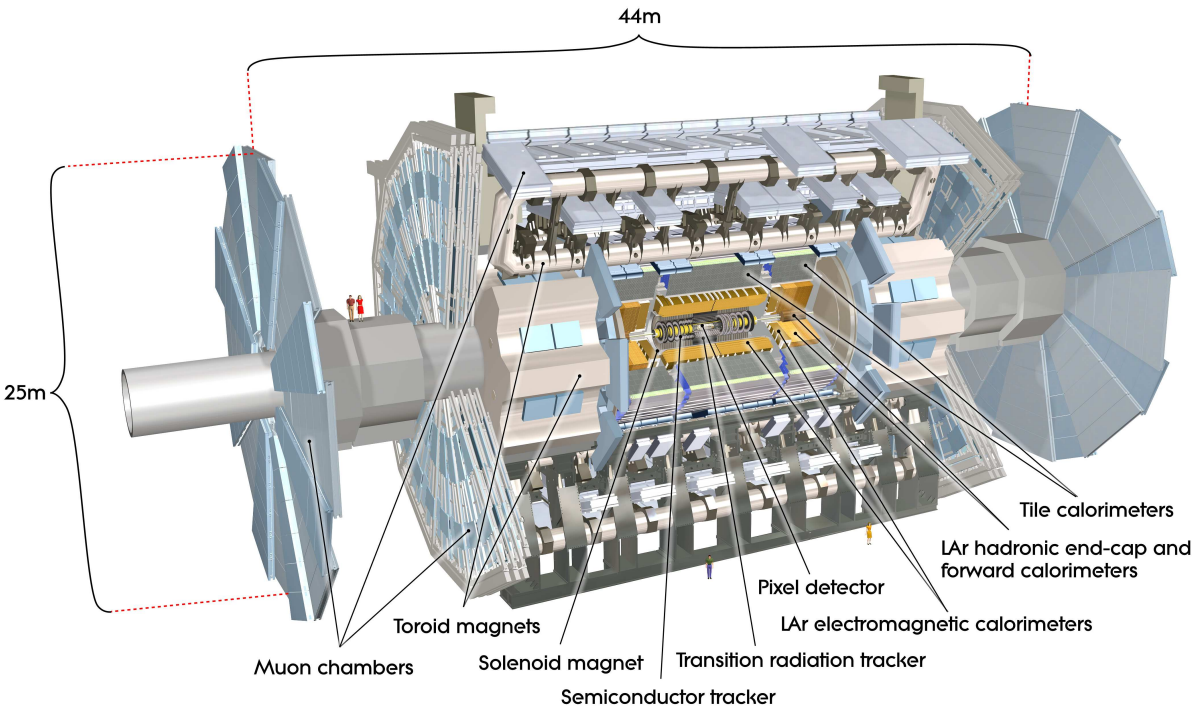


Figure 1.3: Cut-away view of the ATLAS detector highlighting its main sub-parts [10].

to -25°C by an evaporative cooling plant that uses C_3F_8. In 2015 an additional pixel barrel layer, the IBL (Insertable b-layer) [19], was added at a radius of 33.25 mm. It is composed by 224 modules with 50×250 µm pixel size having a mix of planar, and 3D sensors with respective thicknesses of $150 - 250$ µm and $230 - 150$ µm. This new layer uses the FEI4 [20] front-end chip.

The SCT is an assembly of 4 barrels placed between 299 and 514 mm from the beam line with a length of 1492 mm and is made of 2112 modules cooled to -7°C as well as 18 disks [21]. It is operated with a high voltage between 350 and 450 V to compensate for inefficiencies due to radiation damage. It consists of silicon strip modules with 80 µm pitch where each module contains two sensor layers rotated with respect to each other by a 40 mrad stereo angle.

Outside the SCT a continuous straw-tube detector called TRT is located as the outermost part of the tracking system. It consists of straw tubes with a diameter of 4 mm, holding a thin wire in their centre. The tubes are filled with Xenon gas that gets ionised when traversed by charged particles. The TRT can be used to distinguish between electrons and hadrons by measuring their transition radiation. The total number of read-out channels in the SCT is around 6.3 million. The achieved spatial resolution is of 17 µm in the direction of the strip pitch, and of 580 µm in the direction determined by the strip crossing.

The Transition Tracker (TRT) is the outermost component of the Inner Detector and it is composed of drift tubes (straws), each 4 mm in diameter and up to 144 cm long, that hold a thin wire in their centre. The tubes are filled with Xenon gas that gets ionised when traversed by charged particles. The TRT can be used to distinguish between electrons and hadrons by

measuring their transition radiation The TRT only provides information about the radial position, for which it has an intrinsic accuracy of 130 μm per straw.

The tracking system is surrounded by a solenoid magnet with a field intensity of 2 Tthat bends the trajectories of charged particles and provides the measurement of particles momentum. Schematics of all three subsystems are shown in Figure 1.4

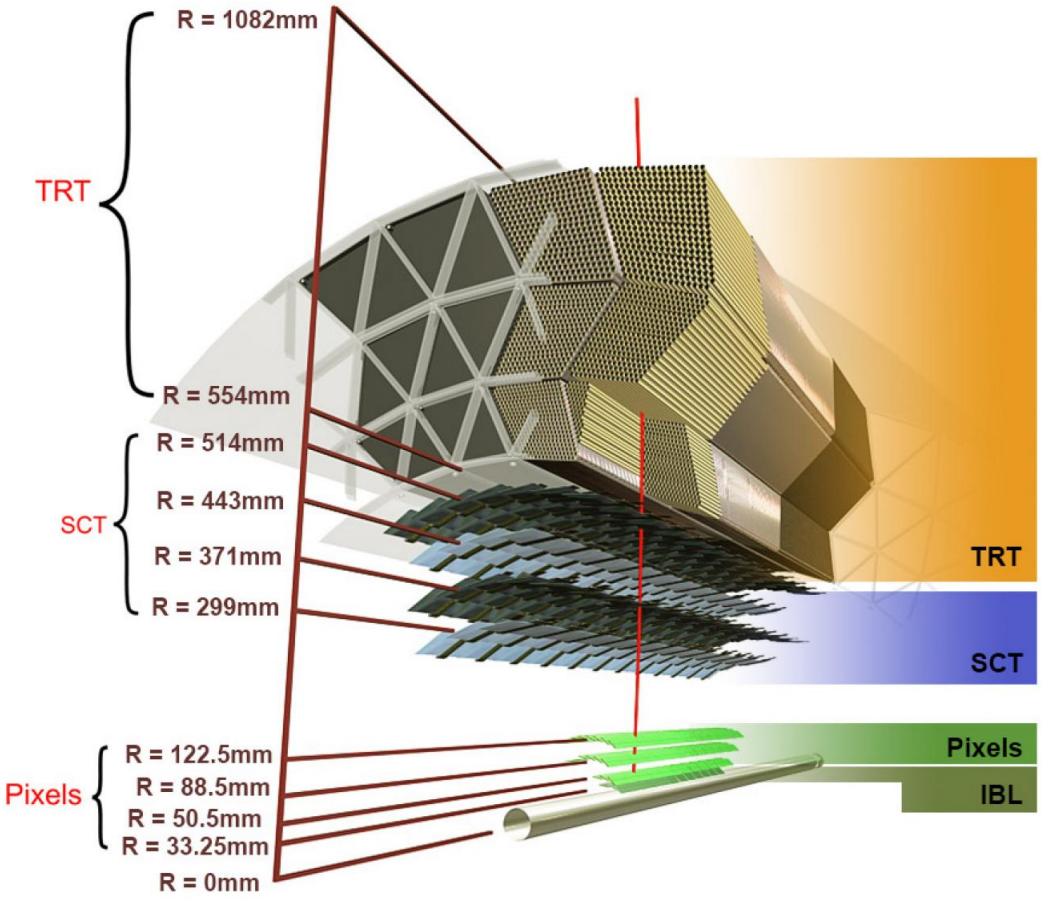


Figure 1.4: Overview of the ATLAS Inner Detector layers [17].

1.4.2 Calorimeters

The calorimeters are the first detectors placed outside the solenoid magnet that surrounds the Inner Detector, as can be seen in Figure 1.5. They cover the range of $|\eta| < 4.9$.

The electronic calorimeter (ECAL) works using liquid argon (LAr) as active material and 1.8mm thick lead plates as the absorber. The liquid argon solution was adopted for its intrinsic linear behaviour, high ionisation yield, stability and resistance to radiation. The lead plates have a characteristic accordion shape and are oriented in the radial direction. This allows a complete symmetric coverage without cracks in the azimuthal direction. High voltage is applied between absorber plates to collect the ionisation electrons from the interaction in the liquid argon as

well as to produce the signal amplification. The electric signal is read from shaped cathodes in the plates through capacitive coupling. The ECAL is composed of two identical half-barrels separated by a 4 mm gap at z=0. It also has two end-caps mechanically divided into two co-axial cylinders. In the central region a pre-sampler layer is located in close contact with the cryostat wall. The information from this layer is exploited in the calibration to estimate the energy lost by the electron or photon in the passive material of the solenoid.

The Hadronic calorimeter (Tile calorimeter) is composed of different independent sampling calorimeters, each with its own particular technology and choice of material. The choice was dictated by the different conditions in term of radiation flux and performance requirements encountered as a function of the pseudo-rapidity of the particle. In the central region there is the Tile Calorimeter that consists of a sampling calorimeter employing iron as passive material and plastic scintillators read by wavelength shifting fibres as the sensitive material. The central region is covered by the Hadronic End-Cap Calorimeter (HEC), which uses copper plates as absorbers and liquid argon as active material for its superior radiation resistance. The electric signal in this case is still collected by the cathodes of the plates. Finally, the Forward Calorimeter (FCAL), is assembled with tungsten rod absorber embedded in a copper matrix. Between the two, a thin gap filled with liquid argon, which provides the active material.

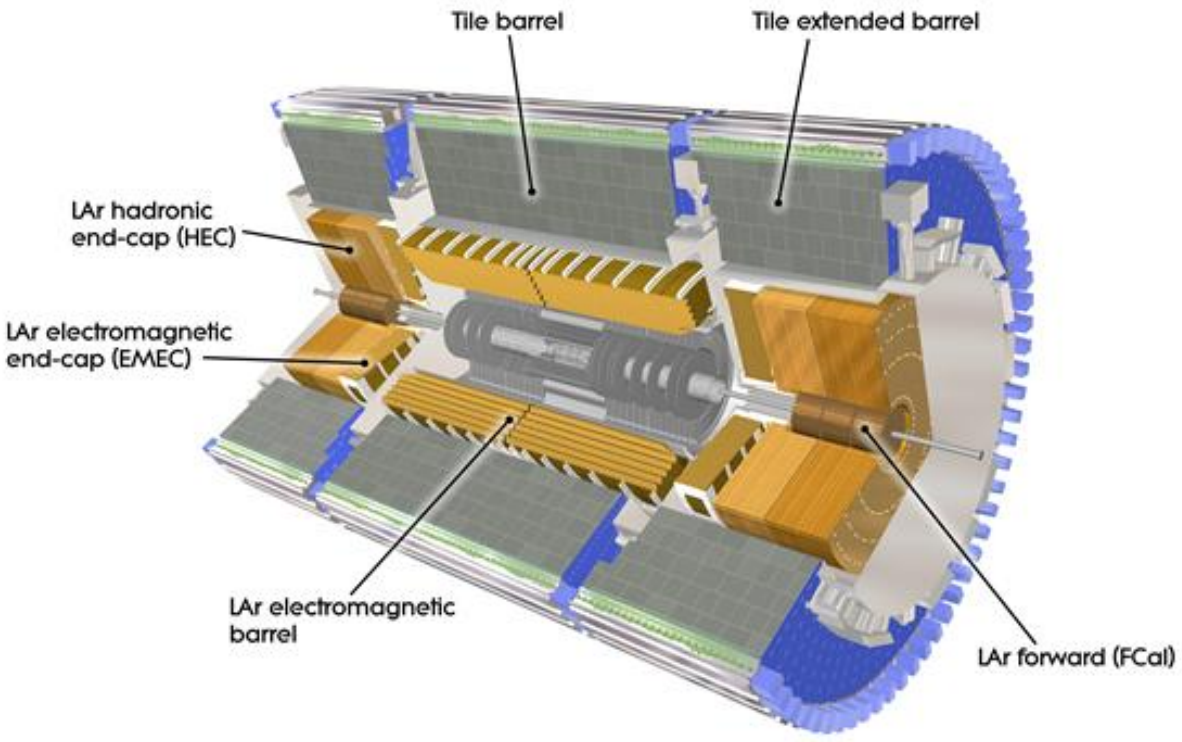


Figure 1.5: Layout of the ATLAS calorimeters.

1.4.3 ATLAS Magnet System

The ATLAS magnetic field configuration is optimised for particle bending around the various detectors in a light and open structure which minimises scattering effects. The experiment magnet system arrangement consists of a central solenoid servicing the Inner Detector trackers with an axial field of 2 T, and a barrel toroid and two end cap toroids that generate a tangential magnetic field of approximately 0.5 T and 1 T for the muon spectrometer in the barrel and end cap regions respectively. The central solenoid is designed to provide a 2 T strong magnetic field in the central tracking volume made out of a single layer coil. It shares the cryostat with the Liquid Argon calorimeter and the flux is returned by the steel of the Hadronic calorimeter.

1.4.4 ATLAS Muon Spectrometer

The ATLAS muon spectrometer has been designed to fulfil the following requirements: efficient use of the magnet bending power, pseudo-rapidity coverage of $|\eta| < 3$, and practical chamber dimensions for production, transport and installation. The spectrometer is divided into three re- gions, barrel region ($|\eta| < 0.5$), transition region ($1.05 < |\eta| < 1.4$) and end-cap region ($|\eta| < 1.4$). Four different technologies depending on the spatial and timing resolution, resistance to radiation and engineering considerations have been used: Monitored Drift Tube chambers (MDT), Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). Figure 1.6 shows the position of the muon chambers.

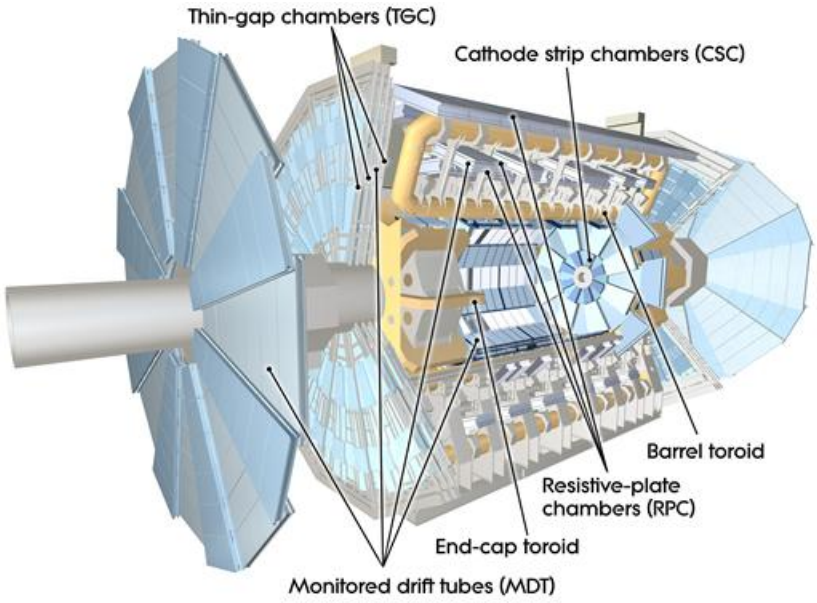


Figure 1.6: Muon system in ATLAS.

The MDT chambers are composed by multi layers of high-pressure drift tubes. Each multi layer is mounted on each side of the support structure. The drift tubes are made of aluminium,

30 mm of diameter, with a central wire of W-Re. They work at 3 bar absolute pressure with a non-flammable mixture of Ar-CO_2.

The CSCs are multi wire proportional chambers operated with a mixture of Ar-CO_2- CF_4. The distance between anode wires (2.5 mm) equals the distance to the cathode. The cathode readout is segmented into strips (5.08 mm) orthogonal to the anode wires. The precision coordinate is obtained by measuring the induced avalanche in the segmented cathode, achieving space resolutions better than 60 μm.

The RPC is a gaseous parallel-plate detector with a typical space-time resolution of $1cm \times 1ns$ with digital readout. It is composed by two parallel resistive plates made of Bakelite. The plates are separated by spacers that define the size of the gas gaps. The gas is a mixture of $C_2H_2F_4$. A uniform electric field of a few kV/mm produces the avalanche multiplication of ionisation electrons. The signal is readout via capacitive coupling to metal strips placed at both sides of the detector and grounded.

A TGC is built with 50 μm wires separated 2 mm. The wires are placed between two graphite cathodes at a distance of 1.6 mm. Behind the graphite cathodes, strips or pads are located to perform a capacitive readout in any desired geometry. Some advantages of these chambers are a fast signal, typical rise time 10 ns and low sensitivity to mechanical deformations.

Finally there is the New Small Wheel (NSW) under construction, installation and commissioning during LS2. It is composed of 16 sectors equipped with a Micro-megas for tracking and a sTGC chamber for triggering. The NSW will replace the current small wheel composed of CSC and TGC chambers.

1.4.5 ATLAS performance

ATLAS showed an exceptional performance throughout Run 2 where it recorded more than 90% of the delivered luminosity. The evolution of the luminosity over the year and the data recorded by ATLAS is shown in Figure 1.7.

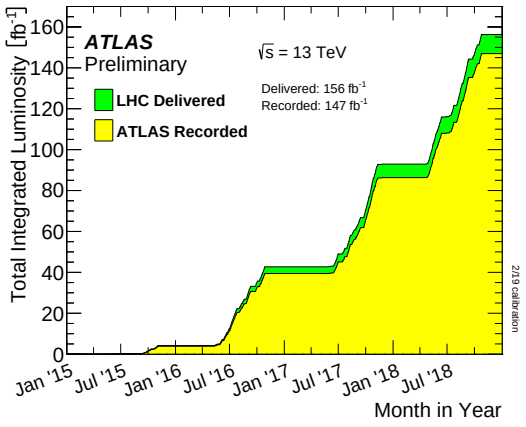


Figure 1.7: Integrated delivered and recorded luminosity by the ATLAS experiment during Run 2 [22].

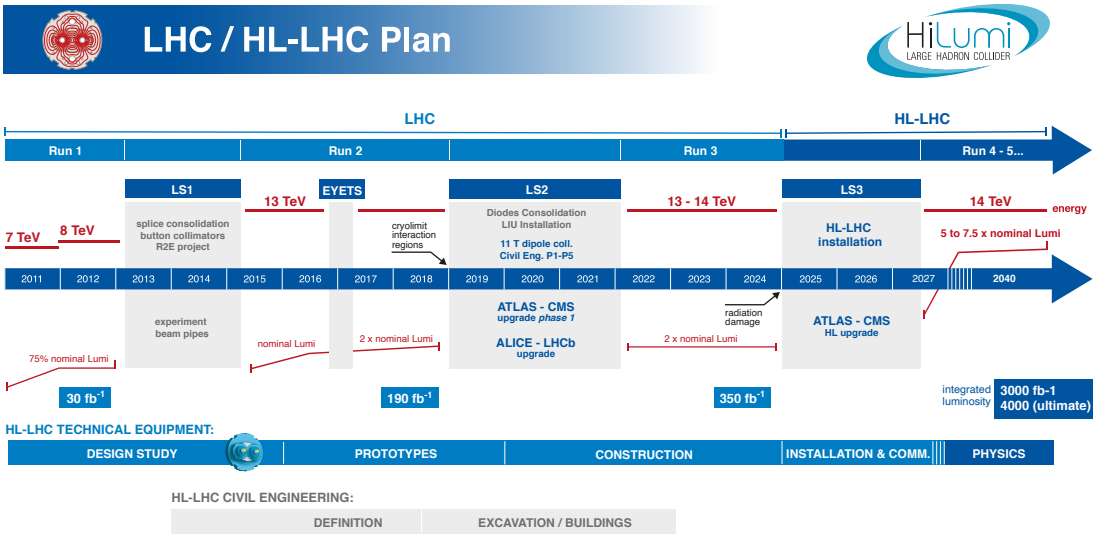


Figure 1.8: LHC timeline of the upgrade, showing energy of the collisions and luminosity [24].

Many physics searches at the LHC experiments require precise information about hadron-jets containing b-quarks, so called b-jets. Tracking does not only give information about the trajectories of particles produced in collisions and their subsequent decay products but also information about a particle's charge and momentum. Information on impact parameter resolution and distinguishing primary and secondary vertices is used as inputs to so called b-tagging algorithms which are able to identify b-jets. Identifying events that include high momenta p_T b-jets in their final state is essential for analyses that aim for measurements of the topquark, studies of the Higgs boson decays involving b-quarks and also searches for new physics therefore directly beneficial for those analyses. Equation 1.3 [23] details the effect of material thickness on tracking momentum resolution where β is v/c of the particle, B the magnetic field, L the distance to the detector perpendicular to the beam axis, X_0 the radiation length of the material being defined as the mean length (in cm) to reduce the energy of an electron by the factor 1/e and d is the total thickness. A CMOS based pixel detector can be extremely thin and yield an improved momentum resolution.

$$\frac{\Delta p_t}{p_t} \approx \frac{0.0136 GeV/c}{0.3 \beta B[T] L[m]} \sqrt{\frac{d}{X_0 \sin\theta}} \tag{1.3}$$

1.5 Towards High Luminosity Upgrade

The LHC started operating in 2008 and the initial data taking spanned from 2010 to 2012. Note that the beginning is not covered by the planning shown in Figure 1.8. 30 fb_{-1} were collected by ATLAS during this period.

Phase 0. The first long shutdown (LS1) of the LHC occurred from February 2013 to June 2015 to reinforce the LHC's magnet interconnects. This was necessary to operate the LHC at its design luminosity of 10^{34} $cm^{-2}s^{-1}$ with each beam having an energy of 7 TeV. During, LS1 the addition of the ATLAS pixel detector IBL (Insertable Barrel Layer) also took place. This shutdown was followed by the Run 2, which lasted from 2015 to 2018 collecting up to 190 fb^{-1} of data at a centre of mass energy of 13 TeV.

EYETS. The Extended Year End Technical Stop in 2016 saw the replacement of the laser drivers by new service quarter panels of the rest of the pixel detector and the installation and commissioning of the DBM (Diamond Beam Monitor).

Phase 1. While this thesis is being written, the LHC is undergoing a major upgrade, Phase 1, to sustain its luminosity at approximately 2.2 x 10^{34} $cm^{-2}s^{-1}$ during a long shutdown from 2019 to 2022 (LS2), during which period two upgrades are planned. The injector system for the Proton Synchrotron Booster (PSB) will be the LINAC4, replacing the LINAC2. The PSB beam intensity will be doubled enabling a greater LHC luminosity. The luminosity will be further increased through upgrades to the LHC collimation system which will focus the beams at the interaction points. A delivery of 350 fb^{-1} is expected from the LHC over the course of Run 3. In order to ensure the robust performance of the ATLAS detector during this period of greater luminosity, significant upgrades specifically targeting the ATLAS Trigger system will take place. ATLAS will experience 55 to 80 pile-up events per bunch crossing at an instantaneous luminosity of 1.9 x 10^{34} $cm^{-2}s^{-1}$. Reducing the event rate by altering the threshold on the pT would result in a loss of physics signal efficiency. In order to avoid this, background events, namely, jets misidentified as electrons, will need to be reduced. This error originates in the calorimeter with fake muons in the spectrometer. An increase in the granularity of the calorimeter readout will therefore also take place. New calorimeter readout boards for the electromagnetic and forward calorimeters are needed to achieve this goal. The fake muon background is addressed by introducing a new small wheel device and new MicroMegas tracking detector technology in the forward muon spectrometer.

Phase 2. In order to sustain and extend its discovery potential, the LHC will yet require a major upgrade. This upgrade will take place between 2025 and mid 2027 in what is known as the long shut down 3 (LS3), see Figure 1.8. The reason for this is that the statistical gain in running the accelerator without a considerable luminosity increase beyond its design value will become marginal. After 2027, the peak luminosity will be increased to more than 5 x 10^{34} $cm^{-2}s^{-1}$, which is a factor of 5 to 7.5 higher than the current design value. If the luminosity over the full operation time of the machine is integrated it gives a value of 3000 fb^{-1} to be reached by 2040. This integrated luminosity is around ten times higher than the expected luminosity after the long shut down 2 (LS2). In order to achieve this operational target, the relevant technologies are undergoing intensive R & D. These will encompass upgrades to the injector chain with crab cavities as well as to niobium and tin(Nb_3Sn) alloy focus quadrupole magnets.

The 5 to 7.5 fold increase in luminosity with respect to the LHC design is expected to lead to around 140-200 instantaneous interactions per bunch crossing. The ensuing higher radiation damage and occupancy are aspects the next generation detector will have to take into consideration. Parts of the present tracker system would likely experience close to 100% occupancy along with radiation induced catastrophic failure in such an environment. The higher luminosity will also equate to reaching the lifetime of the current pixel and strips tracker systems much sooner. Manifestation of detector ageing will appear in the form of increased depletion voltages, exceeding the operational 600 V high voltage limit [25]. A comprehensive upgrade of the inner tracking system will therefore take place through LS3 in 2025 to achieve improved vertex resolution and track reconstruction performance in ATLAS over the Phase 2 period.

The new inner tracker design under consideration will be entirely constituted of silicon, comprising five inner layers forming the pixel detector at close proximity to the beam pipe to ensure improved pattern recognition as well as vertex measurements. Beyond the pixel detector will be four strip layers. The different geometrical layouts being considered are aiming to allow improved tracking performance for pile-up events of up to 200. Subsequently,an upgrade of the readout system will also occur to account for the much greater data rates expected in conjunction with the higher occupancy.

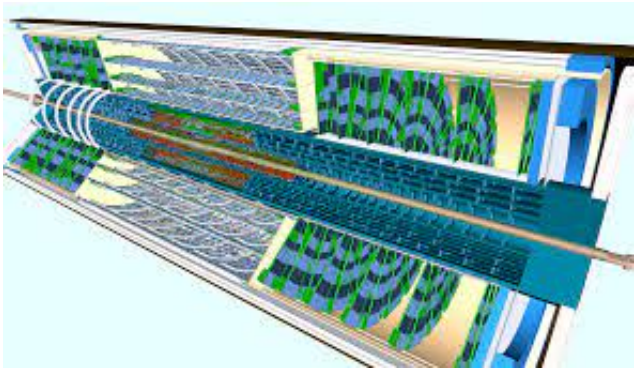


Figure 1.9: Display of the ATLAS Phase-II Inner Tracker ITk with the Inclined Duals detector layout in which barrel and encaps can be observed for both strip and pixel detectors. Figure from [2]

3000 fb^{-1} is expected to be collected during HL-LHC running which, for the pixel layer closest to the interaction point, will correspond to a total ionising dose (TID) of 7.7 MGy as well as a 1 MeV neutron equivalent fluence of 1.4 x $10^{16}cm^{-2}$. This will be significantly greater than the levels the current detector can withstand. On the opposite end, the outermost pixel barrel layer will be exposed to a TID of 0.9 MGy and a 1 MeV neutron equivalent fluence of 1.7 x $10^{15}cm^{-2}$. After the expected fluence the minimum sensor hit efficiency requirement is 97% [26], which allows for good track reconstruction. As a result of the order of magnitude reduction in radiation environment from the inner to outer layers, the technologies considered for the different layers target very different designs. This thesis evaluates the performance of a new

technology foreseen to be placed in the outermost layer of the Inner Detector. Figure 1.10 shows the expected high radiation environment the extended ITk region would be exposed to during the HL-LHC phase. The expected fluence at different z locations don't always decay along r in an $1/r^2$ fashion but instead display minor variations. These can be attributed to neutrons produced from nuclear collisions in the inner layers.

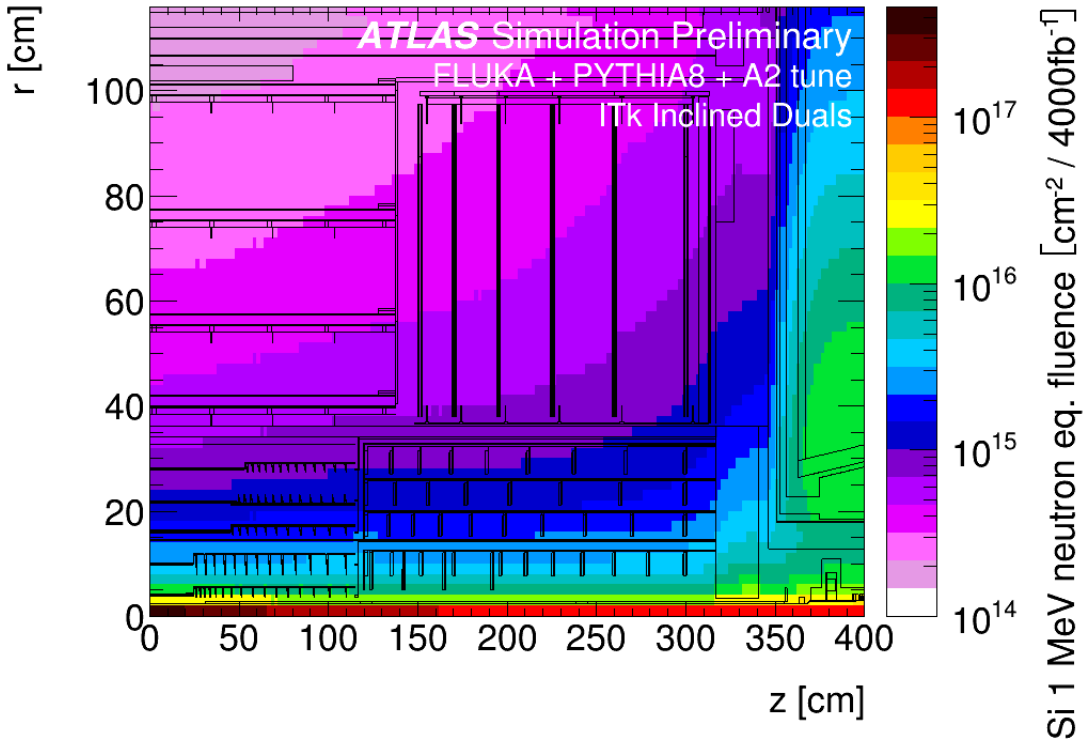


Figure 1.10: Radiation environment within the extended ITk region [30].

The ITk pixel detector readout electronics is designed to provide adequate bandwidth for a 1 MHz new Level-0 trigger, with the option to upgrade to 4 MHz for the track based trigger system. The trigger algorithm exploits regions of interest to estimate the pT of the tracks, which is then used in the Level-1 trigger algorithm following the new proposed trigger rate, with the only restriction of a minimum bandwidth. Resolving multiple pile-up vertices in the high luminosity environment and assigning the high pT jets, tracks and secondary vertices to the vertex of particular interest is the main task of the new pixel detector. Table 1.1 has a summary of the electronics requirements for the Inner Tracker.

Table 1.1: Readout chip specification for the HL-LHC Inner Tracker.

Technology	65nm CMOS
Pixel size	**50x50 um²**
Pixels	192x400 = 76800 (50% of production chip)
Detector capacitance	< 100fF (200fF for edge pixels)
Detector leakage	< 10nA (20nA for edge pixels)
Detection threshold	<600e-
In -time threshold	<1200e-
Noise hits	$< 10^{-6}$
Hit rate	**< 3GHz/cm² (75 kHz avg. pixel hit rate)**
Trigger rate	Max 1MHz
Digital buffer	12.5 us
Hit loss at max hit rate (in-pixel pile-up)	≤1%
Charge resolution	≥ 4 bits ToT (Time over Threshold)
Readout data rate	**1-4 links @ 1.28Gbits/s = max 5.12 Gbits/s**
Radiation tolerance	**500Mrad at -15°C**
SEU affecting whole chip	< 0.05 /hr/chip at 1.5GHz/cm² particle flux
Power consumption at max hit/trigger rate	**< 1W/cm² including SLDO losses**
Pixel analog/digital current	4uA/4uA
Temperature range	-40°C ÷ 40°C

In turn, this requirement dictates the granularity of the detector at different radii. Commensurate to the above requirements, a pixel detector layout is proposed, by splitting the pixel detector in two parts. Each of these parts uses distinctive technologies known as planar or 3D sensors. For the innermost layers 3D sensors have been shown to provide an exceptionally high radiation tolerance due to their stronger electric fields across a shorter distance coupled with their reduced carrier drift distances as shown on Figure 1.11. For a detailed overview of 3D sensors, the reader is referred to [32]. Planar sensors will be used for equipping the outer layers, as shown in the proposed layout for the ITk in Figure 1.11.

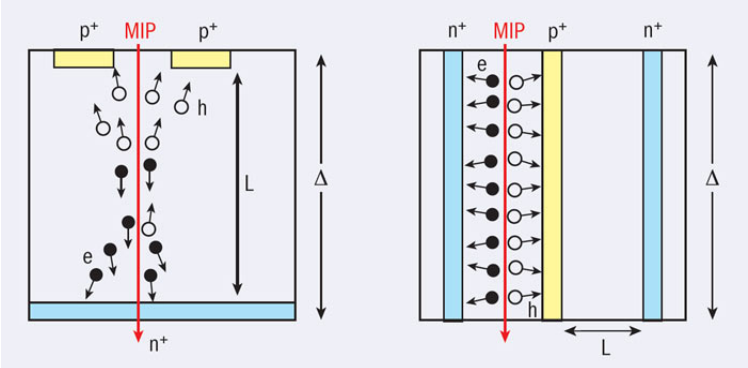


Figure 1.11: Schematic cross-sections of (left) a planar sensor design and (right) a 3D sensor [31].

The front-end chip of the hybrid will be $\approx 20 \times 20\ mm^2$ large, for a total silicon area of $\approx 40 \times 40\ mm^2$, in the case of a quad module. The front-end chip is being developed in the TSMC 65 nm CMOS technology by a joint collaboration between the ATLAS and CMS experiments, the RD53 Collaboration [27] [28]. The first prototype developed, the RD53A chip, has been

implemented and tested with prototype sensors [29]. The ATLAS front-end chip, ITkPix, will consist of a pixel matrix of 400×384 pixels and a 2 mm wide chip periphery. The pixel size will be either 50×50 μm^2 or 25×100 μm^2 according to the different detector regions.

The process that forms the hybrid pixel detector exploits a strategy involving the sensing element and the readout to be designed in two distinct silicon parts. These are then connected to each other through a process called bump bonding and a hybrid pixel detector is thus formed. This is a very expensive process, it reduces the granularity of the sensor because of the required added gaps for the bump pads as well as increases cumulative thickness. Therefore other approaches have been considered to replace the hybrid detectors in the outermost layers of the tracker. The monolithic sensor approach, detailed in Chapters 3, 4 and 5 is an interesting possibility. Here both the amplification and the additional readout logic are in the same silicon wafer. This approach gives the chance of exploiting industrial grade silicon manufacture for LHC and future HEP experiments. As mentioned, the pixel detectors are required to cope with very high events pile-up. Table 1.2 summarises the estimated hit occupancy for the modules in each area of the ITk. An average hit rate of 80 Mhz/mm^2 can be considered for the outermost pixel layer. The sensors and the readout architecture have to be sufficiently fast to match the hit rate requirement within the 25 ns of the bunch crossing. Modules will have to transmit data at high speed with an expected maximum bandwidth of 5.12 Gb/s. The hit loss at maximum hit rate should be kept smaller than 1%.

Table 1.2: Average hits per chip per event for 50×50 μm^2 pixels, using tt events with 200 pile-up, for different layers and regions of the ITk pixel detector [26].

Layer/Ring	Flat Barrel	Inclined Barrel	End-cap
Layer 0	223	136.7	80.9
Layer 1	26.6	27.8	37.7
Layer 2	18.3	20.1	21.0
Layer 3	12.9	12.7	13.3
Layer 4	9.9	9.1	9.3

Together with the high granularity of the detection system, a low material budget is needed, limiting the multiple scattering to achieve the targeted track reconstruction efficiency. A particle passing a material of thickness L will leave it with some displacement on the plane and a deflection angle θ, with respect to the entry trajectory. In good approximation, the θ angle follows a Gaussian distribution centred in zero, with a width corresponding to:

$$\theta_0 \propto \frac{1}{p}\sqrt{\frac{L}{X_0}} \tag{1.4}$$

where p is the particle momenta and X_0 the radiation length. To achieve the expected track reconstruction, the θ_0 has to be limited, according to its simulated maximum acceptable value. The design parameter to control θ_0 is the material budget, usually indicated in units of radiation

length X_0. The expected material budget of the ITk pixel modules is 0.67% of the radiation length X_0, see Table 1.3. The large number of pixel modules of the ITk imposes to minimise material of the local support and cooling to limit the material budget of the detector system. Therefore low mass carbon fibre local supports with thin cooling pipes will be used. For the same reason, a serial powering scheme [33] will be implemented to power the readout chips. A full parallel powering scheme would increase the material budget considerably, because of the large density of modules and the consequently large number of power cables needed. Parallel power is currently foreseen within a single module, while groups up to 14 modules will be powered in series [49]. An additional chip, the Pixel Serial Powering Protection (PSPP) [34], should be used parallel to every module to mitigate the risk of losing a full serial powering chain in the case of an open module failure.

Table 1.3: Material budget for the main components of the ITk pixel modules in units of the radiation length $\%X_0$, which has been estimated based on experience with the ATLAS IBL [26].

	$\%X_0$
Front-end chip, 150 µm thick	0.19
Sensor, 150 µm thick	0.17
Bumps	0.03
Module incl. SMD components	0.27
Flex adhesive	0.01
Total	0.67

Chapter 2

Silicon Pixel Sensors

This chapter covers the fundamentals of how particles are detected in matter and applies this to how silicon detectors detect the passage of particles. It goes through the most important parameters that determine the performance of silicon pixel detectors and the distinct types that are used in the LHC experiments. It describes in detail how the different types of radiation affect the detection capabilities. Finally, it introduces Depleted Monolithic Active Pixel Sensors, which are the focus of this thesis. These concepts will be used in the following chapters.

2.1 Interaction of Particles with matter

The detection of particles happens via the energy loss in the material that they traverse. The amount of deposited energy depends on the particle itself, on its energy and on the traversed material. This section describes the passage of different particles through matter, with a special focus on what happens in silicon detectors.

2.1.1 Electronic energy loss by charged particles

There are two main mechanisms in which charged particles lose energy: ionisation and excitation. The mean rate of energy loss for relativistic heavy charged particles (M»m_e) is described by the Bethe-Bloch equation.

$$-\frac{dE}{dx} = Kz^2\frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_ec^2\beta^2\gamma^2W_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right] \tag{2.1}$$

where K is a constant defined as $K = 4\pi N_A r_e^2 m_e c^2$, β is the velocity of the traversing particle $\beta = \frac{v}{c}$ and γ is the Lorentz factor $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$.

- N_A is the Avogadro's number or number of atoms per cm^3
- m_e is the classical mass of the electron

- r_e is the classical radius of the electron
- c is the speed of light 299792458 m/s
- Z is the atomic number of absorbing material
- A is the atomic mass of the detector material
- W_{max} is the maximum kinetic energy which can be given to a free electron in a single collision
- I is the mean excitation energy
- λ is the density effect correction factor

The Bethe-Bloch describes the mean rate of energy loss in the region of $0.1 < \beta\gamma < 1000$ for intermediate Z materials with an accuracy of few %. At the lower limit the particle velocity becomes comparable to atomic electron velocities and at the upper limit radiative effects are relevant. The minimum of the curve is known as the Minimum Ionising Energy and most relativistic particles seen in particle physics experiments have a mean energy loss close to this value. Consequently, the concept of Minimum Ionising Particles (MIPS) is defined as charged particles, which embody the minimum ionising losses in substances. Both values usually constitute a specification when designing a sensor.

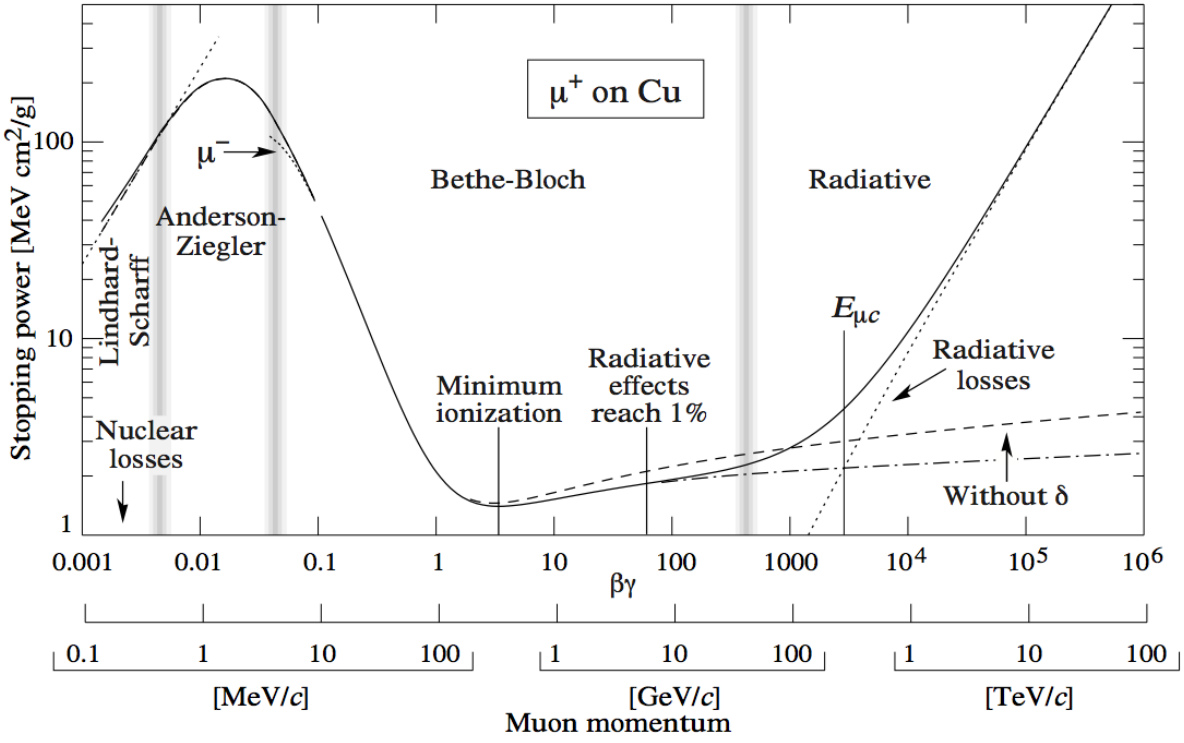
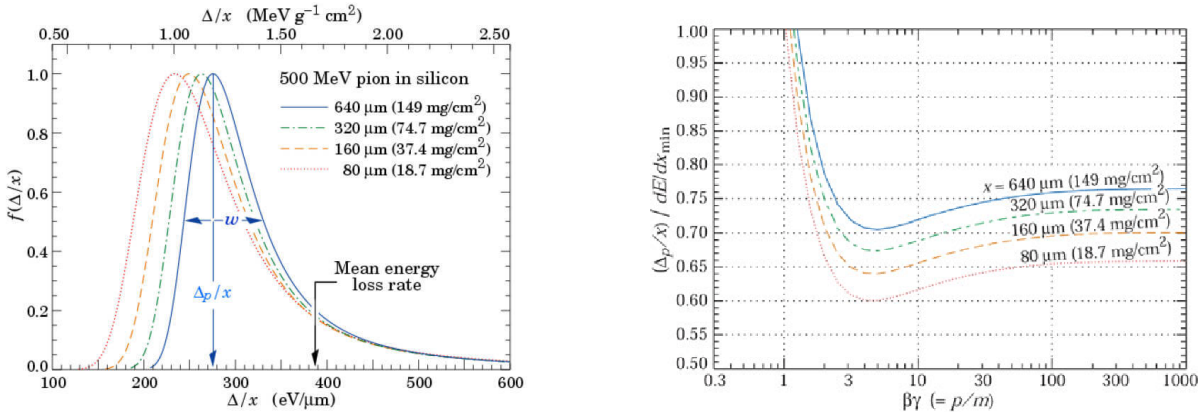


Figure 2.1: Stopping power $-\frac{dE}{dx}$ for muons in copper as a function of its momentum [35].

However, the statistical nature of the ionising process during the passage of a fast charged particle through matter results in large fluctuations in the energy loss Δ in absorbers that are thin in comparison with the particle range. The probability density function that describes the distribution of energy loss Δ for an absorber thickness x was calculated by Landau and Vavilov. Further corrections were later incorporated. For thick absorbers, this distribution has a Gaussian shape.



(a) *(b)*

Figure 2.2: (a) Straggling functions in silicon for 500 MeV pions, normalised to unity at the most probable value Δ_p/x. The width w is the full width at half maximum. (b) Most probable energy loss in silicon scaled to the mean loss of a minimum ionising particle [35].

As shown in Figure 2.2, it can be observed that the most probable energy loss for pions in silicon is below the mean predicted by the Bethe-Bloch equation, mainly because of the tail in the distribution. The distribution is asymmetrical due to energetic delta rays. Furthermore the most probable value decreases for thinner silicon thicknesses.

2.1.2 Energy loss of electrons

The energy loss of the electrons happens via two processes.

- **Ionisation**: Similar qualitative behaviour as the one described by the Bethe-Bloch equation but with some differences due to the electron's lower mass, it is more important at low energies.

- **Bremsstrahlung**: This is the electromagnetic radiation produced by the deceleration of a charged particle when deflected by another charged particle, typically an electron by an atomic nucleus.The moving particle loses kinetic energy, which is converted into a photon, thus satisfying the law of conservation of energy. The contribution of this process is more relevant for high energies.

2.1.3 Photon interactions with matter

Photons interact with a material primarily via three different processes. They can interact with matter and be completely absorbed by depositing its energy. They can interact and and be deflected from its original direction depositing part of its energy. Finally, photons that have more energy than twice the rest mass of an electron can give rise to the production of a pair of particles. These processes are described below:

- **Photo-electric effect**: In the photoelectric (photon-electron) interaction a photon transfers all its energy to an electron located in one of the atomic shells. The electron is ejected from the atom by this energy and begins to pass through the surrounding matter. The electron rapidly loses its energy and moves only a relatively short distance from its original location. The energy transfer is a two-step process. The photoelectric interaction in which the photon transfers its energy to the electron is the first step. The depositing of the energy in the surrounding matter by the electron is the second step. The photoelectric effect dominates for rather low photon energies. A portion of the energy is used to overcome the electron's binding energy and to free it from the atom. The remaining energy is transferred to the electron as kinetic energy and is deposited near the interaction site. Since the interaction creates a vacancy in one of the electron shells, typically the K or L, an electron moves down to fill in. The drop in energy of the filling electron often produces a characteristic x-ray photon. The energy of the characteristic radiation depends on the binding energy of the electrons involved. Characteristic radiation initiated by an incoming photon is referred to as fluorescent radiation. Fluorescence, in general, is a process in which some of the energy of a photon is used to create a second photon of less energy. This process sometimes converts x-rays into light photons. Whether the fluorescent radiation is in the form of light or x-rays depends on the binding energy levels in the absorbing material.

- **Compton scattering**: A Compton interaction is one in which only a portion of the energy is absorbed by an electron and a photon is produced with reduced energy. The intensity of the scattered beam has two peaks showing this. One peak appears at the wavelength λ of the incident radiation and the second peak appears at wavelength λ'. This photon leaves the site of the interaction in a direction of travel different from that of the original photon. Because of the change in photon direction, this type of interaction is classified as a scattering process. The separation between the peaks depends on the scattering angle. In effect, a portion of the incident radiation "bounces off' or is scattered by the material. Therefore, this process can result in particles with different energies. Figure 2.3 illustrates the intensity peaks for a setup in which Monochromatic x-rays with wavelength λ are incident on a sample of graphite, where they interact with atoms inside the sample. They later emerge as scattered x-rays with wavelength λ' . A detector placed behind the target can measure the intensity of radiation scattered in any direction θ with respect to the

direction of the incident x-ray beam.

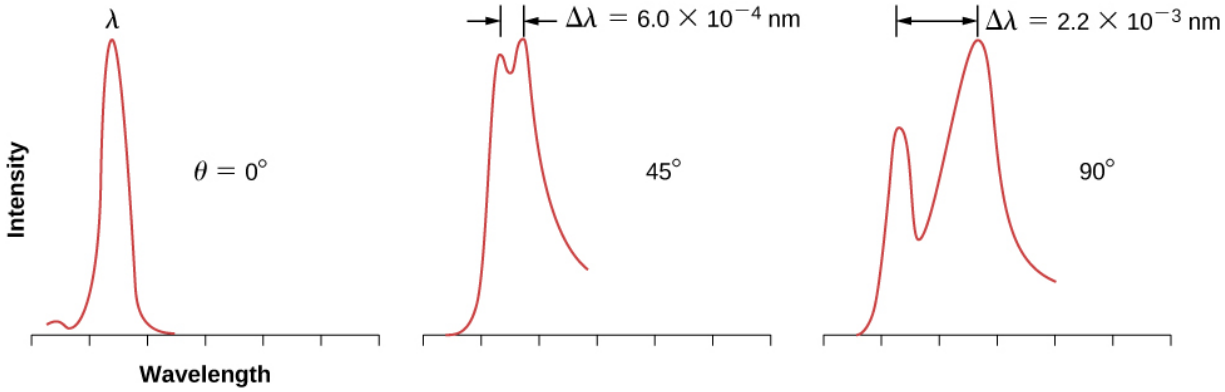
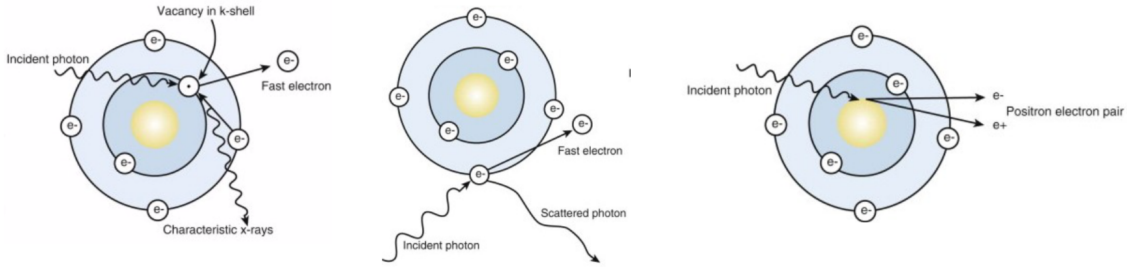


Figure 2.3: Experimental data for the described experiment. The data are plotted in arbitrary units so that the height of the profile reflects the intensity of the scattered beam above background noise [37].

- **Pair Production**: Pair production is a photon-matter interaction that occurs in photons with energies in excess of 1.02 MeV. In this process, the photon interacts with the nucleus in such a manner that its energy is converted into matter. The interaction produces a pair of particles, an electron and a positively charged positron. These two particles have the same mass, each equivalent to a rest mass energy of 0.51 MeV.

Finally a graphic representation of the phenomena described above is shown in Figure 2.4.



(a) Photoelectric effect. (b) Compton effect. (c) Pair production.

Figure 2.4: In the image the three process in which photons interact with matter are sketched [36], in (a) the Photoelectric effect, in (b) the Compton effect and in (c) the Pair production

2.2 Silicon detectors

2.2.1 Material properties

Looking at an isolated atom, the electrons have only discrete energy levels. However, in solid state materials the atomic levels merge into energy bands. These bands determine the properties

of the material. For metals the conduction and the valence bands overlap while for insulators or semiconductors these levels are separated by an energy gap. In Figure 2.5 a graphical representation of different types of materials is shown according to its energy bands.

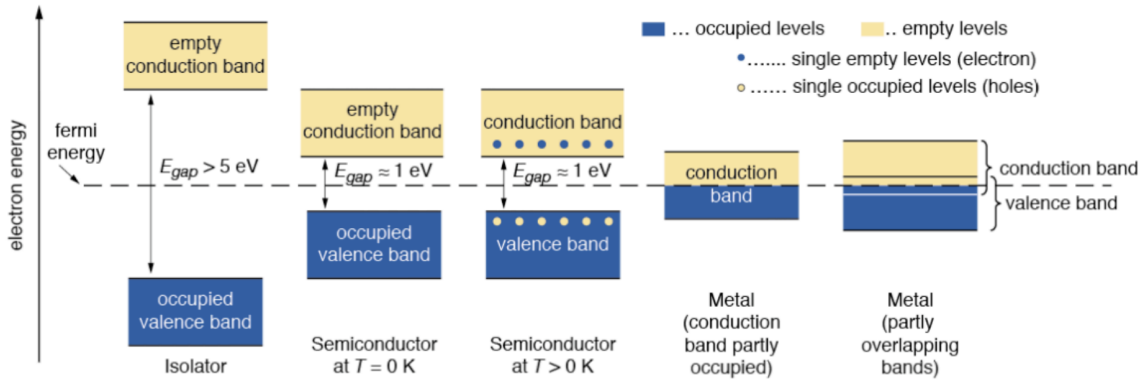


Figure 2.5: The image shows that for insulator materials the energy that is needed to go from the valence to the conduction band is almost impossible to overcome. At the other end of the spectrum, metals have both bands overlapping and semiconductors materials fall in the middle, so their properties can be modified to fulfil detection purposes. Furthermore, the properties of the material are dependant on temperature, becoming more conductive for higher temperatures [38].

For semiconductors, at room temperature, some electrons, represented as n, are in the conduction band due to the small energy gap. These electrons are likely to recombine with the holes, named p, present as well in the conduction band. Once the equilibrium between excitation and recombination is reached, the semiconductor is said to have the intrinsic carrier concentration, ie $n = p = n_i$. The equation for the intrinsic carrier concentration is:

$$n_i = N_s e^{\left(-\frac{E_g}{2K_b T}\right)} \tag{2.2}$$

where,

- n_i is the intrinsic carrier concentration, the number of electrons in the conduction band (and also the number of holes in the valence band) per unit volume in a semiconductor that is completely free of impurities and defects.
- N_s is the number per unit volume of effectively available states; its precise value depends on the material.
- k_B is Boltzmann's constant, $k_B = 1.381 \times 10^{-23}$ Joules/Kelvin
- T is the absolute temperature in Kelvin

The band gap energy for silicon is 1.12 eV at room temperature (300K) [39]. However, when calculating the mean ionisation energy for this semiconductor the value is more than three

times higher the band-gap energy, 3.62 eV, because all the energy supplied is not only used for ionisation purposes but there are also vibrations and other excitation that result in phonon creations.

The electrons and holes, move because they are subject to an electrical field. This process is known as drift.

$$\overrightarrow{v} = -\mu_{n/p}\overrightarrow{E} \tag{2.3}$$

and where $\mu_{\mathrm{n/p}}$ is the mobility of the holes and the electrons which depends on the mean time between collisions. Electrons move faster than holes. For silicon, the electron mobility at room temperature is $\mu_{\mathrm{n}} = 1450 cm^2/Vs$ and the holes mobility is $\mu_{\mathrm{p}} = 450 cm^2/Vs$.

2.2.2 P-N junction as a detector for High Energy Physics

The principle of a detector is that the particles that penetrate or traverse it create electron-hole pairs. Applying an external electric field, these pairs are drifted to the electrodes. This signal is what will be read out. One of the most important parameters of the detector is the signal to noise ratio.

A p-n junction consists of n and p doped substrates. Doping is the replacement of a small number of atoms in the lattice by atoms of neighbouring columns from the atomic table (with one valence electron more or less compared to the basic material). These doping atoms create different energy levels within the band gap altering the conductivity of the material. A doped semiconductor is called an extrinsic semiconductor and contrary to the intrinsic ones, there is a surplus of electrons or holes depending on the doping. At the transition between the n-type and p-type materials majority carriers from one side diffuse to the other side and recombine with the majority carriers producing a region depleted of free carriers [40]. The space charge in the depletion region, creates an electrical field that stops further diffusion. See Figure 2.6 for a better understanding of the p-n junction. The potential built is known as the built-in voltage $\mathrm{V_{bi}}$ (represented in Figure 2.6 as α) and for a p-n junction in thermal equilibrium can be calculated with the following equation:

$$Vbi = \frac{kT}{e}\ln\left(\frac{n_{0n}p_{0p}}{n_i^2}\right) \approx \frac{kT}{e}\ln\left(\frac{N_D N_A}{n_i^2}\right) \tag{2.4}$$

where $\mathrm{n_{0n}}$ and $\mathrm{p_{0p}}$ are the majority carrier concentrations on both sides, and as complete ionisation can be assumed they can be approximated by the donor and acceptor concentrations, $\mathrm{N_D}$ and $\mathrm{N_A}$, respectively. k is the Boltzmann constant, T the temperature and e is the electron charge, while $\mathrm{n_i}$ is the intrinsic carrier concentration in silicon.

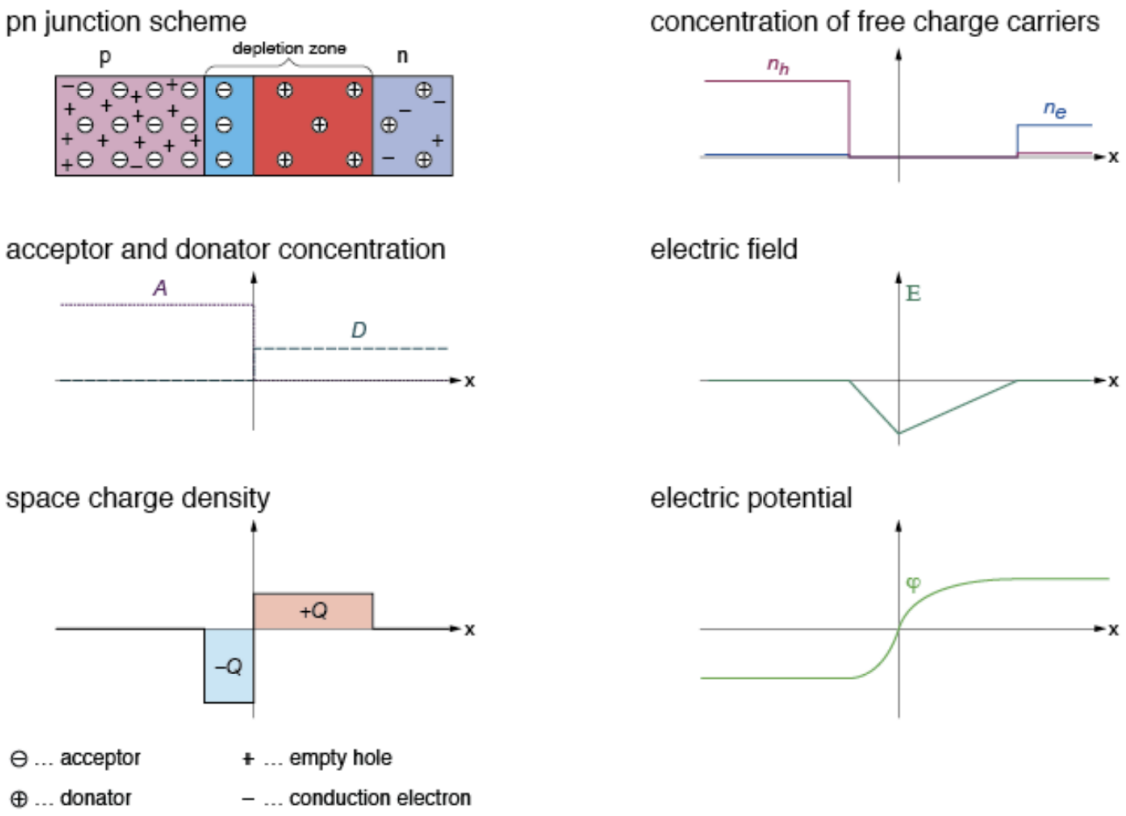


Figure 2.6: Example of a p-n junction in thermal equilibrium, with no bias voltage applied.Plots for the charge density, the electric field, and the voltage are also reported [38].

If an external voltage, V, is applied such that +V is to the n-semiconductor and -V to the p-semiconductor then electrons and holes are removed from the doped semiconductor resulting in an extension of the depletion region. In addition, the potential barrier will also increase and the diffusion of carriers across the junction is suppressed. In this process, an excess of carriers is generated thermally in the extended depletion region, which results in a small current known as leakage current. Detectors are operated in this mode. The width of the depletion region for a planar junction of silicon is obtained as the sum of the depletion region width on both sides as given in the following equation:

$$W = x_n + x_p = \sqrt{\frac{2\varepsilon_0\varepsilon_{Si}}{e} + \left(\frac{1}{N_D} + \frac{1}{N_A}\right)(V + V_b i)} \tag{2.5}$$

Assuming that the reverse bias voltage applied is substantially higher than the potential across the junction and considering that the semiconductor doping is higher on the n-side than on the p-side, the above equation can be simplified to:

$$W = \sqrt{\frac{2\varepsilon_0\varepsilon_{Si}}{eN_A}V} \tag{2.6}$$

This is a very important equation that is extensively used to calculate the depletion depth of

sensors. The maximum electrical field can be obtained as follows:

$$E_{max} = \frac{2V}{W} = \sqrt{\frac{2eN_A}{\varepsilon_0 \varepsilon_S i} V} \tag{2.7}$$

All these equations described above show evident that applying a higher reverse bias enlarges the depletion region and increases the electrical field. Furthermore, having a low doped junction or high resistivity helps to increase the depletion region.

A p-n junction also has a capacitance given as:

$$C = \varepsilon_0 \varepsilon_S i \frac{A}{W} \tag{2.8}$$

where A is the area of the junction and W is the width of the depletion region. From this last equation it is important to realise that capacitance is minimised for a small junction area and a wide depletion region. This capacitance has a strong role from the sensor point of view as it is a key parameter to determine the signal to noise ratio as well as the timing capabilities of the detector.

2.2.3 Transport of charged carriers

In a semiconductor there are two different mechanisms of creating current due to transport of charged carriers: diffusion and drift. These are crucial for generating the signal in the detector that will then be read-out. Carriers can also move because of their thermal energy. In this case, they behave as a particle with a kinetic energy of $3/2kT$. However the travelled distance should be averaged over many charge carriers in equilibrium condition so no current will be produced.

If there is a gradient in the concentration of carriers, random movements of free charges cause a diffusion current density J_{diff}.

$$J_{n,diff} = -qD_n \nabla n \tag{2.9}$$

$$J_{p,diff} = -qD_p \nabla p \tag{2.10}$$

where D_n and D_p are the diffusion constants [40]. Diffusion transport does not depend on the electric field. The charge velocity is low and the carrier path from one point to another is usually long because it is originated from random movement. The direction of the current is determined by the concentration gradient.

The second mechanism occurs if an external electrical field is applied to a semiconductor, the free charge carriers are accelerated between random collisions with the semiconductor lattice. The drift current densities $J_{n,drift}$ and $J_{p,drift}$, for electrons and holes respectively, can be found as:

$$J_{n,drift} = -en\mu_n E \tag{2.11}$$

$$J_{p,drift} = ep\mu_p E \tag{2.12}$$

where μ is the mobility of the free charges. Figure 2.7 shows the mobility of electrons and holes as a function of the applied electric field E. Increasing the electric field above a certain value, around $10^5 V/cm$, does not lead to an increase of the drift current. The mobility decreases and, therefore, the carriers drift velocity $vd = \mu E$ saturates at about $1x10^7 cm/s$, in the case of silicon. The drift velocity no longer obeys Ohm's law, and becomes non-linear in the applied field with a clear tendency to saturation due to the appearance of a new dissipation mechanism involving optical phonon generation. The proportionality factor between the current density and the electric field is the conductivity σ or, equivalently, the specific resistance ρ of the material:

$$\sigma = \frac{1}{\rho} = en\mu_n + ep\mu_p \tag{2.13}$$

The transport of free charges is a critical aspect of silicon particle sensors. When a particle traverses the detector it ionises charge carriers that should be sensed before they get trapped or recombine with others. Drift is preferred over diffusion due to the higher velocity reached by the carriers. The carrier movement in the semiconductor is used to induce a signal in the sensor. The shorter the path for the carriers to drift the smaller the risk of losing charge before the collection electrode. The risk of charge loss in a path is higher after irradiation, as the induced damages increases the trapping probability. The basis of the monolithic detector that this thesis focuses on is operation at full depletion and charge collection via drift.

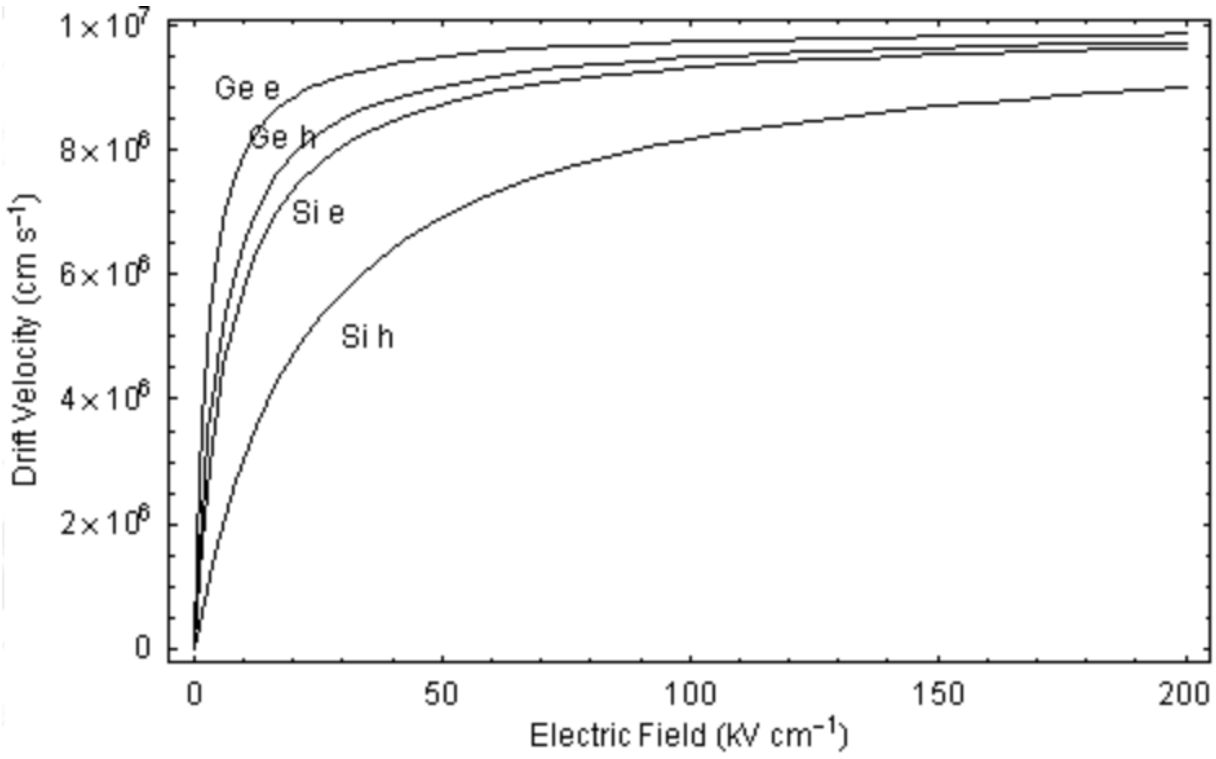


Figure 2.7: Dependence of carrier drift velocity on electric field for Ge and Si [52].

2.2.4 Signal formation on silicon detectors

Electron-hole pairs generated by a traversing particle induce a signal in the electrode. This phenomenon is known as charge collection. The free charge generated in a depleted p-n junction will move by drift under the effect of an electric field, with an average velocity v_{drift}. Applying the Shockley-Ramo theorem, the induced current in the detector can be calculated as:

$$i(t) = q v_{drift} E_w \tag{2.14}$$

where E_w is the so-called weighting field and it is different from the actual electric field in the sensor. It can be obtained by applying a unit potential to the electrode under consideration and zero potential to all other electrodes. The charged induced on the electrode, Q, by the movement of carriers from position x_1 to x_2 is the integral of the current over the collection time.

$$Q = \int_{t_1}^{t_2} i(t) = e[\Phi_w(x_1) - \Phi_w(x_2)] \tag{2.15}$$

Φ_w is the weighting potential, also obtained by raising the electrode under consideration to unit potential, setting all others to zero, and solving the Poisson equation.

From these equations, there are two important concepts to retain. First is that the smaller the electrode, the larger the area where the weighting potential comes near to zero, meaning that drift of carriers in this area will induce very little signal on the collection electrode. Second is that the closer the charge is to the electrode, the more signal it will induce, as the gradient of the weighting potential increases, and for small electrodes, most of the signal is induced in this last part of the drift path.

In order to have a better understanding Figure 2.8 shows the weighting potential, Φ_w, of different collection electrodes sizes. It is assumed that the electrode is located at 0 in the ordinate axis. The dimension of the opposite electrode, placed at 1 in the ordinate axis is considered as infinite in the abscissa. Consequently it can be observed that if the electrode size is smaller or similar to the detector thickness the weighting potential decreases along with the sensor depth and the charge will need to travel to the collection electrode to induce current. This is an important concept to take into account when designing small collection electrode detectors for which the electrode is smaller than the pixel size, they will be introduced in Subsection 2.4.1.

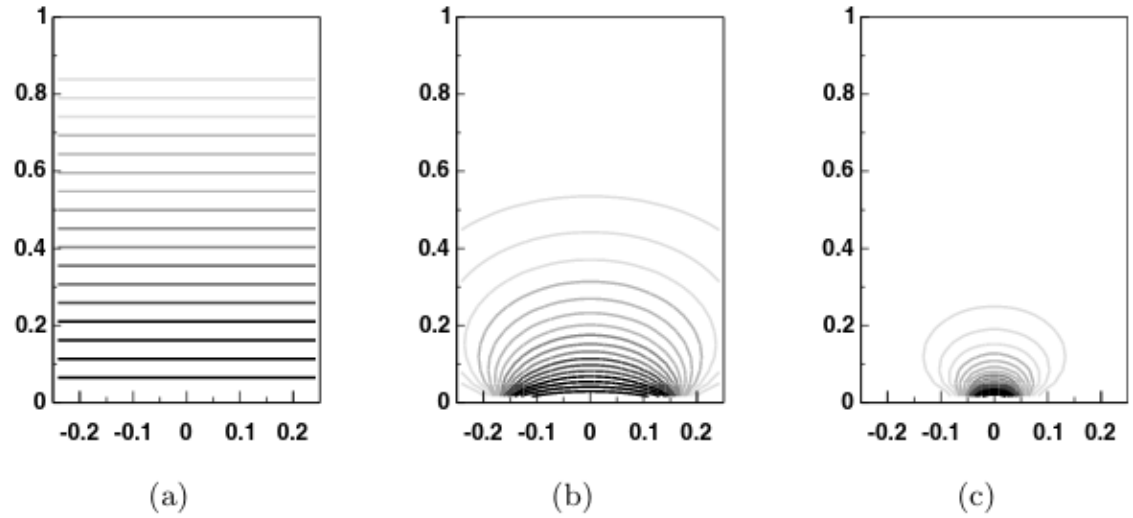


Figure 2.8: Weighting potential for two infinite parallel plates in a), for a collection electrode of 1/3 of the wafer thickness in b) and 1/10 of the wafer thickness in c) [51].

2.3 Hybrid pixel detectors

The word pixel is directly linked to image processing applications and it describes the smallest controllable element of a picture represented on the screen [11]. Consequently, a pixel detector is a device that detects an image with a granularity corresponding to the pixel size. Photo and video cameras as well as x-ray films are just examples of devices used in daily life based on sensing elements, pixels, that interact with photons of different energies and generate an intensity distribution known as an image. In the case of HEP applications, images and patterns are not generated by visible light but by charged particles or photons in the energy range of keV to MeV which experience an ionising interaction with the detector. Most of HEP experiments, like the ones in the LHC, have been using the so-called Hybrid Pixel Detectors (HPD) [51] [11] because they are fast and able to detect high-energy particles and electromagnetic radiation. The detection mechanism is performed via different devices that have specific functions:

(i) A sensor that converts the energy of the radiation into an electrical signal.

(ii) The front-end electronics that processes the signal and sends it to the digital readout circuitry.

(iii) Eventual processing and storage that allows for later inspection and data analysis

The notion of "hybrid" detector comes from the fact that the sensor and readout electronics, are fabricated separately then joined together through a process named bump-bonding, which can be seen in Figure 2.9. Because each pixel is visible in the radiation image, a high bump bonding yield is required to avoid visible defects. Solder or indium bumps have been typically

used as connecting the chips electrically and mechanically together. Bump-bonding process has a rather high cost although it has been fully characterised and it is suitable for withstanding high-radiation levels, gives a good resolution and copes with high data rates. Another important characteristic of HPDs is the high density connectivity between the sensing elements and the readout electronics.

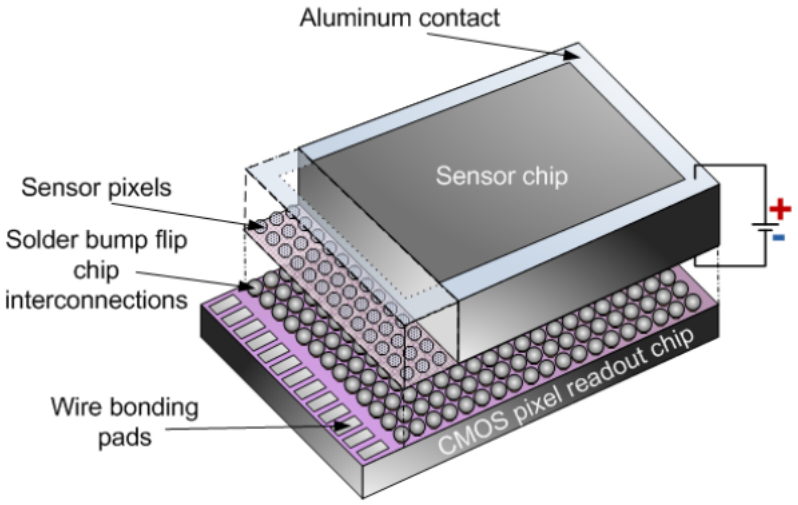


Figure 2.9: Image of a hybrid pixel detector taken from [50].

In order to get an idea of how the pixel detectors have evolved the ATLAS Pixel Detector used in Run 1 was made of 1744 modules that constitute 47232 pixels with a typical size of 50×400 μm^2. In the IBL, the pixel size of the sensors is 50×250 μm^2 which is 60% of the pixel size used for the Pixel Detector. The IBL has about 12 million pixels in total [16]. The ATLAS ITk Pixel Detector at the HL-LHC Upgrade will have pixels of different sizes region-dependent: in the outer layers the pixels will be 50×50 μm^2 but in the inner system will be 25×100 μm^2 and they will also be of different technologies 3D in the innermost layers and planer in the outer ones. There will be a total of 9000 modules and around 5 GPixels [12].

2.3.1 Silicon radiation sensor

Up until now, a wide variety of radiation sensors based on different types of materials have been developed, like CCDs, silicon strips, gas electron multipliers, vacuum tube photo-multipliers, avalanche photo-diodes, etc [47]. Nonetheless, planar silicon sensors have been extensively used in previous generations of experiments, see [48], and are the baseline for the future Upgrades of the tracking systems of both ATLAS and CMS. Nevertheless for IBL and the innermost layer of the Atlas Upgrade ITK 3D sensors optimised to perform in high radiation environments [49] have been designed, characterised and used.

In Figure 2.10 shows the cross section of a single sided p-in-n silicon planar sensor. A large area p+ implantation is placed in a n-bulk and a positive bias is applied to the back side through

an ohmic n+ contact and metal layer. The ionising particles create charge carriers (electrons and holes) that are then collected due to the electrical field in the depletion region. The sensor is acting in this case as a reversed bias p-n junction as described in section 2.2.2. The charge is then fed to the analogue front-end in the ASIC readout chip through the bump bond connection by DC-coupling. This type of sensor has been extensively used due to their simplicity [51]. Other types of planar sensors have also been studied and optimised in order to achieve maximum charge collection, spatial resolution or radiation hardness.

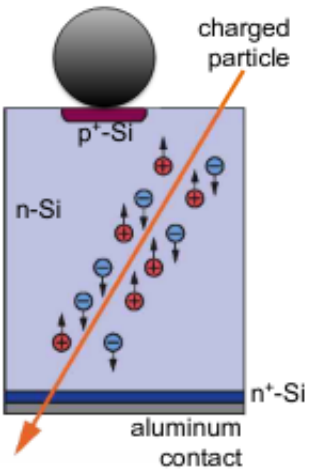


Figure 2.10: Cross section of a single-sided p-in-n silicon planar sensor, with n-bulk and p+ implant [69].

2.3.2 ASIC Readout chip

Readout chips for pixel detectors feature different geometries, readout approaches and analogue front-ends, but the main properties and the hierarchy are common to most of them. They are composed of an active area that has a repetitive matrix of elementary pixels directly interfacing to the sensor via the bump-bonds and of a chip periphery that manages the global control, data buffering, readout and global configuration. In HEP, the active area is known as the pixel matrix or pixel array and it is formed of basic units called Pixel Unit Cells (PUCs). In the small pixel size, a PUC integrates an analogue front-end, required to perform analogue-to-digital conversion of the charge collected in the sensor, and digital processing, most likely including some data storage.

An analogue front end is commonly designed with a cascade of amplifying stages. The first stage is the preamplifier and the following stages are bandwidth-limited determining the frequency spectrum of the output pulse and its shape, they are the filter or pulse shaper. The filtering is necessary because detector signals are very fast and their shape cannot be preserved with limited power. The most typical front-end architecture for HEP applications is the one in which the output of the analogue circuitry is fed to a discriminator, that outputs a digital

signal. This can be either considered a binary hit [62] or a further amplitude measurement can be performed. In the analogue front end, an inverting amplifier with feedback capacitance converts the input charge to a voltage. The preamplifier is a critical part of the circuit and it is carefully designed bearing in mind several metrics (e.g gain, bandwidth, power, noise, etc.). In circuits where the preamplifier output is directly interfaced with the discriminator (without a separate filter), the discharge must be completed before the next signal arrives, to avoid overlap. Analogue to digital conversion of the collected charge is done through a Time over Threshold (ToT) measurement, where the ToT represents the number of clock cycles for which the signal is higher than the discriminator threshold. The pulse width should be proportional to the input charge. The digitisation of the signal into a defined number of bits can be done with simple approaches, for example via a clock signal and a digital counter for each channel. Otherwise, a clock counter can be centralised and its output is latched into local registers when the leading and trailing edge transitions are detected by the single channels. The ToT is then computed as the difference. An example of a generic in-pixel front-end circuit can be found in Figure 2.11.

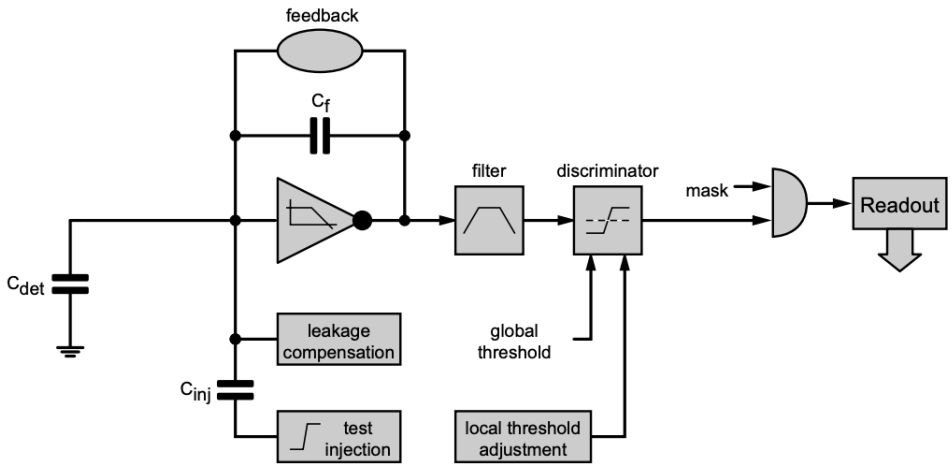


Figure 2.11: Block diagram of a Pixel Unit Cell [51].

The readout architecture of the ASICs of HPDs is closely related to the targeted application. Position, time and the corresponding pulse amplitude of all hits belonging to an interaction must usually be provided in HEP. When dealing with high-data rates like in the LHC, on-detector data reduction is needed in order to obtain a feasible data rate towards the Data Acquisition System (DAQ). For this reason, usually a trigger signal is used for selection of hits of interest. The generation of the trigger signal is based on the analysis of many sub-detectors of the experiment. This analysis has to happen within a fixed latency after the particle has been detected, for the trigger to be correctly produced. Storage logic is required to maintain the data until the trigger latency has expired. As far as triggered architectures are concerned, data buffering can be implemented following diverse approaches and storage elements can be located in different parts of the pixel chip (End of Column (EOC), single PUC, region of certain number of PUCs),

implementing different readout schemes. Moreover, limited buffering constrained by the area available, can be a substantial source of hit loss unless this issue is properly addressed at design time. An overview of triggered architectures can be found in [63].

2.4 Monolithic Active Pixel Detectors

Monolithic Active Pixel Sensors are fabricated in standard CMOS technology, often with minor process adjustments, allowing the chip to include both sensor and front-end electronics, unlike the HPDs [67]. A typical monolithic cross-section, like the one shown in Figure 2.12, consists of the n-well collection electrode situated in a p-type epitaxial layer that collects the charge generated by the traversal of an ionising particle. The p-wells around the collection electrode have the in-pixel electronics that in this case are the NMOS transistors. A small depletion area is created around the collection electrode in which the generated charge is collected via drift mechanism. Nevertheless, most of the epitaxial layer remains undepleted so the majority of the charge is eventually collected by diffusion. In this example, because the collection electrode takes a small portion of the total pixel area it is referred to as a small collection electrode. In order to extend the depletion region, achieve faster charge collection, mainly dominated by drift and enhance the radiation tolerance of the detector, the front-end and readout electronics can be placed inside the collection electrode. For this approach, the collection electrode occupies the majority of the pixel area so it is known as large collection electrode.

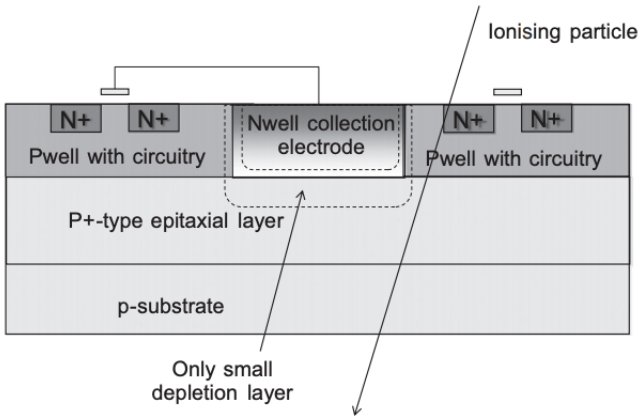


Figure 2.12: Cross section of a DMAPS detector, taken from [64].

2.4.1 Small collection electrode introduction

As has just been introduced, for the small collection electrode designs, the free carriers are collected at an electrode that is a small fraction of the area of the pixel, a few μm^2. The in-pixel circuits are isolated using a deep p-well, which prevents electrons from being collected by the

PMOS n-well. The main advantage of the small collection electrode approach is having a low sensor capacitance, in the order of a few femtofarad, as it allows for a low power consumption and noise [66]. The voltage signal seen by the sensor can be calculated with the following relationship:

$$V = \frac{Q}{C} \tag{2.16}$$

where Q is the charge deposited by the incident particle and C is the sensor capacitance. After collecting the charge in the sensor there is an amplification. In the following subsection the evolution of this type of detectors for CERN designs will be introduced to show the state of the art.

History of DMAPS at CERN

In 2014, the ALICE experiment in the LHC adopted the ALPIDE CMOS MAPS sensor implemented in the TowerJazz 180 nm technology CMOS imaging sensor process for an upgrade replacement for their Pixel Inner Tracking System [70]. This fabrication process will be referred to in this thesis as the standard process. It is the first experiment in the LHC making use of MAPS instead of a hybrid pixel solution. A schematic drawing of the cross section of this sensor is shown in Figure 2.13.

This technology was initially chosen because of the possibility of using a deep p-well to shield the n-well of the PMOS transistors allowing full CMOS technology and complex readout circuitry inside the pixel and different starting materials, epitaxial substrates with high resistivity, so the signal generation could be optimised. In this design only the n-well collection electrode is not shielded from the epitaxial layer by the deep p-well and therefore it collects the signal charge from the epitaxial layer. The electrode is separated from the in-pixel electronics by several microns to reduce the lateral capacitance to the wells. The low capacitance of the small collection electrode made it possible to only consume 40 nW in the front-end which was then combined with a zero suppressed readout for low overall power consumption. To further increase the depletion zone and further reduce the sensor capacitance, a reverse bias of up to 6 V is applied to the p-type substrate of the sensor. Since the bodies of the NMOS transistors see the same reverse voltage applied to the substrate, this bias is limited by the breakdown of the source/drain junctions of NMOS transistors, which takes place at $\approx 8V$.

In terms of radiation hardness of the sensor, for ionising radiation it follows the observed trend in sub-micron CMOS technologies that as the thickness of the oxide of the gates decreases, the tolerance increases. In particular this technology uses a 3 nm gate oxide thickness. In the case of NIEL irradiation, this is a harder requirement to achieve as most of the charge in the sensor is collected by diffusion which is a slower mechanism than drift allowing for more time for the signal charges to be trapped. In the standard process it is difficult to deplete the epitaxial layer over its full width unless the p-well area is kept very small with respect to the total pixel

area. For the ALICE experiment this is not an issue as the NIEL radiation levels needed are in the order of $10^{13}\ MeV n_{eq}/cm^2$ and timing information is not as critical as in the other experiments.

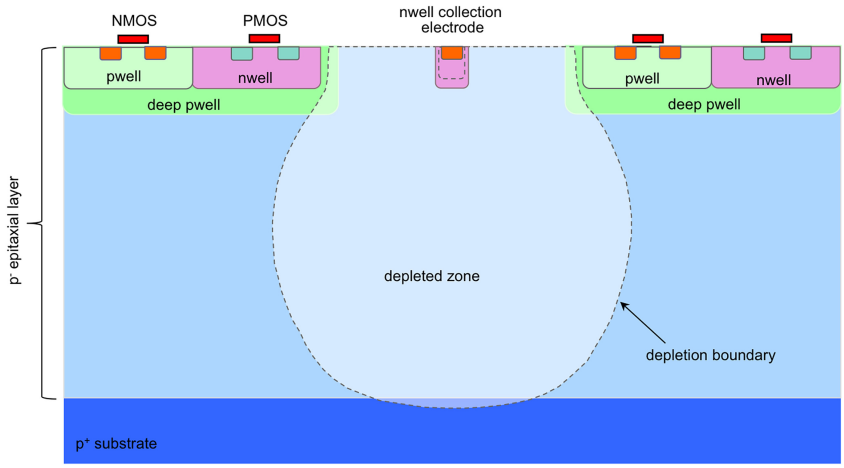


Figure 2.13: Cross section of the ALPIDE detector, featuring the TowerJazz 180 nm standard process [82].

The promising measurement results of the ALPIDE, received interest from the ATLAS ITk community and a group inside this experiment was created in order to develop sensors that could meet specifications for the outer layer of the ATLAS pixel detector. The specifications that were the hardest to achieve were the radiation tolerance of 80-100 MRad for ionising doses and $10^{15}\ MeV n_{eq}/cm^2$ fluence for silicon bulk defects. This effort resulted in the submission of an engineering run in July 2017 with the MALTA and Monopix sensors. These two sensors, of $2 \times 2\ cm^2$ and $1 \times 2\ cm^2$ respectively, have very similar front-end electronics but different readout architectures. The readout architectures will be studied in detail in Chapter 4 while the front-end and the sensor will be covered in Chapter 3, focusing on the MALTA results as they are part of the work of this thesis. Prior to the fabrication of these two large scale sensors there were several studies done to optimise parameters like pixel size, electrode size, electrode to deep p-well distance in the TowerJazz process that are not object of this work but whose results were taken into account in the design of these DMAPS, see [73].

In order to improve the radiation tolerance by two orders of magnitude, from $10^{13}\ MeV n_{eq}/cm^2$ to $10^{15}\ MeV n_{eq}/cm^2$, the depletion region needs to be vastly extended with respect to the ALPIDE and the charge should be collected by drift mechanism instead of by diffusion. In the ALPIDE it is difficult to extend laterally the depletion region far into the epitaxial layer in between the high resistivity substrate and the deep p-well, as this requires a potential gradient or an electric field in between two equipotentials. Therefore a process modification was done for the MALTA and Monopix sensors named the standard modified process [83]. This modification consists of adding a low dose n-implant layer in order to move the junction of the sensor away from the small collection electrode and create a fully planar junction. The cross section of the standard modified process can be seen in Figure 2.14.

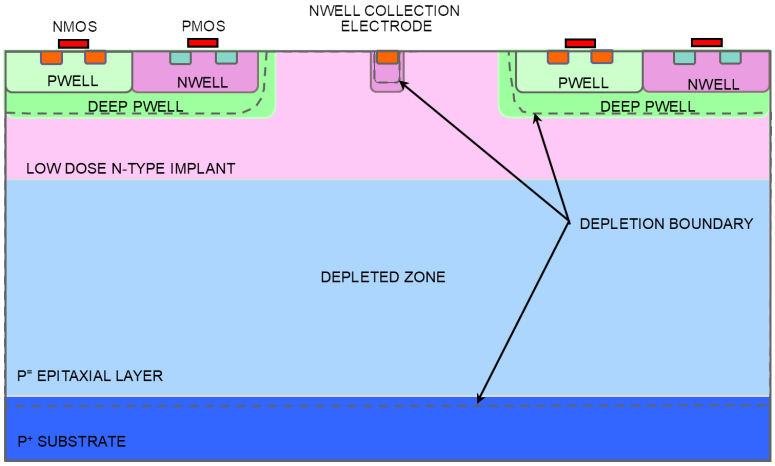


Figure 2.14: TowerJazz 180 nm standard modified process cross section [83].

As for the standard process, the electronics are in a separate well outside the collection electrode. The depletion layer also separates and isolates the p-wells containing the electronics from the substrate, which implies that they can be biased independently. For this process, the sensor junction is between the low dose n-type implant and the p-epitaxial layer and it is planar. The depletion starts at the junction and with a reverse substrate bias of few volts, the depletion extends to the n-well collection electrode and over the full pixel width. In general, the higher the reverse bias is, the higher the electrical fields are in the sensitive region and consequently the charge is collected more quickly. The limit of when applying higher reverse bias stops having a beneficial effect for the different sensors will be shown in the characterisation results in Section 3.4.6. All of this, allows the combination of the benefits of full depletion with a small sensing node capacitance ($\approx$ 2.5 fF) that is essential for a low power pixel design and a lower material budget. The pixel size is 36.4×36.4 μm^2 and the electrode constitutes only 10% of this area.

Another positive point of this approach is that, apart from defining the region of the n-implant over the pixel matrix, the process modification does not require any other layout changes in the design of the sensor or the circuitry. Therefore, the same design can be done in both the standard and the modified process, permitting a direct comparison between the two. To the contrary, in order to achieve the high collection efficiency specified by the LHC experiments, very fine tuning of design and processing parameters is needed and extensive simulations have to be run.

Measurements on this detector confirmed that the addition of the low dose n-type implant maintains the low sensor capacitance. The full depletion improves the timing performance and increases the non-ionising radiation dose tolerance [74].

2.4.2 Large collection electrode

A large collection electrode design [65] refers to the one seen in Figure 2.15 in which the deep n-well constitutes the collection electrode, that extends underneath the pixel area. The front-

end and readout electronics are situated inside the collection electrode. The NMOS transistors are isolated from the collection electrode with a p-well layer. Since the transistor junctions are now isolated from the p-type substrate, a high reverse bias voltage in the order of 50-100 V can be applied to the collection diode. These characteristics improve the radiation hardness of the detectors, because full depletion can be achieved and the electrons have to travel a shorter path to generate a current reducing the risk of being trapped. The charge will be mostly collected by drift mechanism so it is fast. The main disadvantage of this type of DMAPS is the large input capacitance seen by the collection electrode, in order of a few 100 femtofarads, due to the large junction between the p-substrate and the deep n-well. This capacitance is comparable to the one of the HPD detectors. In this case, to achieve a high Q/C there are two possibilities, the first is to limit the area and complexity of the in-pixel electronics or to increase the collected charge by using thick high-resistivity substrates for the sensitive layer. A large signal is needed to achieve a good S/N ratio because, as was seen in the previous section, it is dependent on the capacitance.

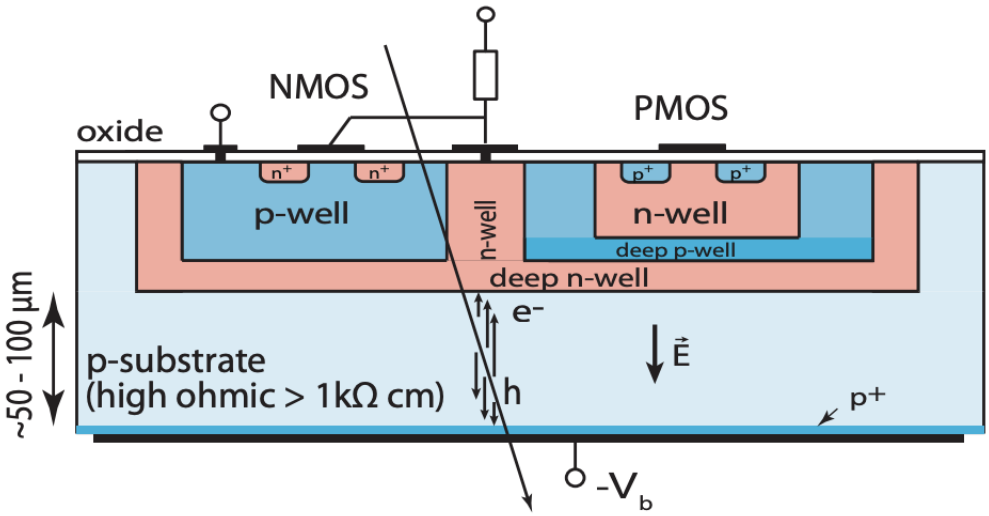


Figure 2.15: Cross-section of a large collection electrode design taken from [68].

2.5 Radiation damage in silicon detectors

The hostile radiation environment where silicon detectors operate means that its performance is degraded due to radiation damage. Cumulative radiation effects are gradual effects that occur during the whole lifetime of the electronic component in a radiation environment. The effects of radiation on electronic devices and materials depend on the type of radiation, the rate of interaction and the detector characteristics. This section explores the main sources of this damage.

2.5.1 Silicon bulk defects

Bulk damage is caused by the interaction of the incident particles with the nuclei of the lattice atoms. In order to remove a silicon atom from its lattice position a minimum recoil energy of around 25 eV is required. Electrons need an energy of at least 260 keV in order to provide such a recoil energy in a collision, while protons and neutrons, because of their higher mass, require only 190 eV. When only a small amount of energy is delivered to the recoil Si atom isolated displacements are created, leaving a single vacancy-interstitial pair. They are created mostly by electromagnetic radiation of low energy electron and x-ray photons.

If the recoiling silicon atom gets enough energy through the collision, it can cause further defects. A recoil Si atom needs about 5 keV of energy to displace other Si atoms in the crystal, creating a dense agglomeration of defects. Since these displacements are formed in a small volume, there is rapid annealing and only 2% of all generated defects form electrically active states. These disordered regions are referred to as defect clusters. Defect clusters have high local defect density and can be tens of nanometers wide.

To be able to compare the damage caused by the different types of particles with different energies, radiation damage is scaled with non-ionising energy loss (NIEL). This quantity summarises all energy deposited in the crystal that does not cause ionisation. The NIEL corresponding to a total fluence of particles is quoted in terms of the equivalent fluence of 1 MeV neutrons to provide a common reference for comparing the damage from different particles.

Cluster defects severely affect the performance of the detector, depending on their concentration, energy level and the respective electron and hole capture cross-sections.

Radiation-induced defect levels close to the middle of the bandgap are very efficient charge carrier generation centres that lead to an increase of the leakage current of silicon devices. This current is also called generation current or dark current. The experimental calculation of the leakage current must be done with great care as it depends on several parameters.

The leakage current increases considerably after irradiation because of the new centres of charge carriers generated inside the space charge region. The increase in current, ΔI_{vol}, is proportional to the particle fluence, Φ_{eq}, and independent of the type, resistivity, and impurity content of the used silicon material [53]. It can be expressed as shown in Equation 2.17:

$$\Delta I_{vol} = \alpha \Phi_{eq} \tag{2.17}$$

where α is the current related damage rate. Figure 2.16 shows data obtained from various silicon detectors irradiated in a neutron field with 5.2 MeV mean energy and measured at room temperature after a dedicated annealing of 80 min at 60° C which is used as a common reference for measurements [59]. An annealing process consists of a heat treatment process that alters the structure of the material.

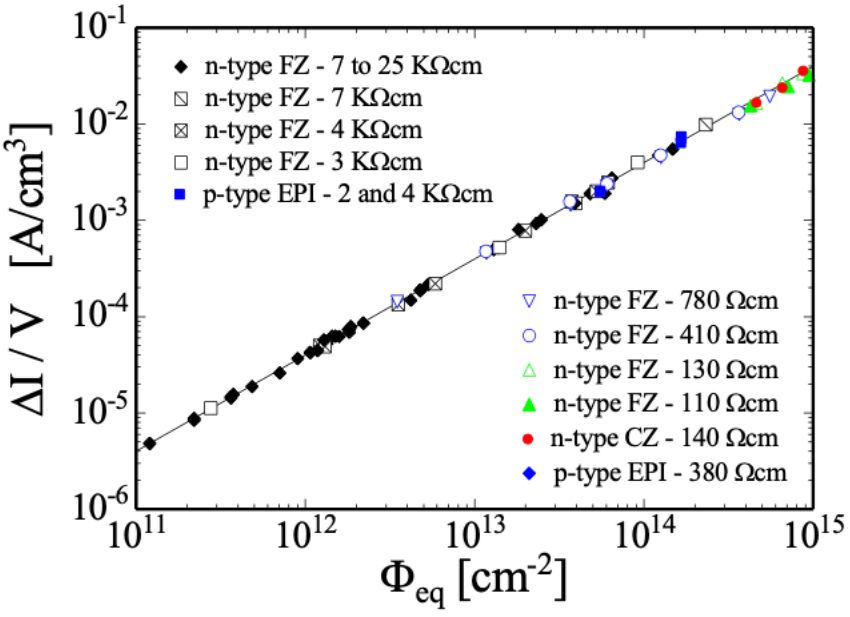


Figure 2.16: Radiation induced leakage current increase as function of particle fluence for various silicon detectors made from silicon materials produced by various process technologies with different resistivities and conduction type. The current was measured after a heat treatment of 80 min at 60° C and is normalised to the current measured at 20° C. Figure taken from [54].

Leakage current is also dependent on temperature. This dependence is dominated by the position of the energy levels in the band gap, their cross sections, their concentrations, and the temperature dependence of the bandgap itself. The most efficient generation centres are the ones at the intrinsic energy level. In this case, the leakage current temperature dependence will follow one of the intrinsic carrier concentrations n_i. Chilingarov compared experimental results obtained on several different irradiated silicon particle detectors using the parameterisation $I(T) \propto T^2 exp(-E_{eff}/2K_BT)$, where $E_{eff} = 1.214 \pm 0.014eV$ [55]. The leakage current generated in the volume has a strong temperature dependence and therefore the current can be reduced by lowering the temperature. This is important as the sensor power generated by the leakage current can raise the temperature, causing more leakage current and leading to thermal runaway.

Finally, the current related damage rate, α factor has an annealing behaviour. Figure 2.17 shows the change in α with time for different annealing temperatures. The annealing temperature is the temperature at which the samples are stored or heated to accelerate the defect reactions in the silicon bulk. This temperature shall not be confused with the measurement temperature of the leakage current which in the given example is 20°C. The α value is continuously decreasing with increasing annealing time.

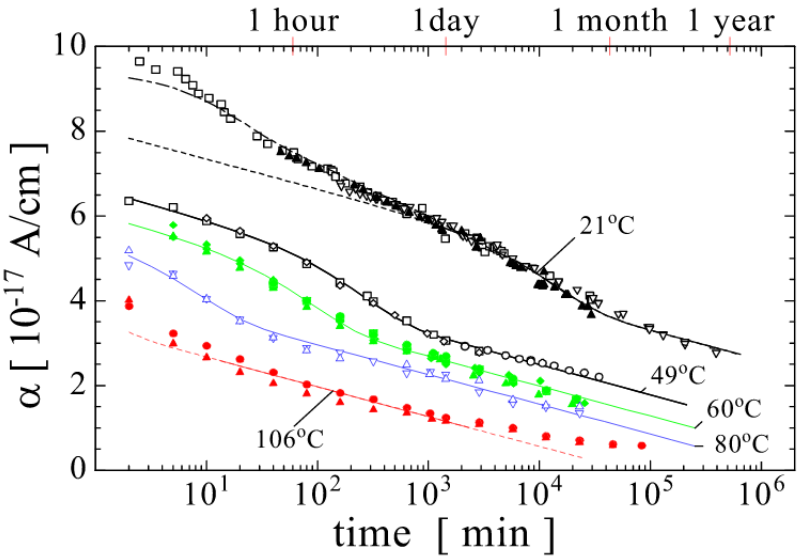


Figure 2.17: Current-related damage rate α as a function of cumulated annealing time at different temperatures. Solid lines: fits to the data. Figure taken from [56].

The radiation-induced defects lead to a change in the effective space charge N_{eff} that is reflected in a change of the depletion voltage V_{dep} of silicon detectors. The depletion voltage V_{dep} is given as:

$$V_{dep} = \frac{q|N_{eff}|d^2}{2\varepsilon\varepsilon_0} \tag{2.18}$$

where d is the thickness of the device, q is the elementary charge, ε is the relative permittivity of silicon and ε_0 is the vacuum permittivity. It shall be noted that Equation 2.18 is assuming a constant space charge over the volume of the damaged detector, which is not always the case [60]. However, the most common way of calculating depletion voltage is with a capacitance versus voltage measurement. Figure 2.18 shows an example of the evolution of the effective space charge or depletion voltage for an n-type sensor with particle fluence. Before irradiation, the sensor was of high-resistivity n-type (phosphorus-doped) base material resulting in a positive space charge of some 10^{11} cm^{-3}. Irradiation of the sensor results in the formation of negative space charge which compensates the initial positive space charge. With increasing particle fluence, the net space charge decreases and reaches very low values corresponding to almost intrinsic silicon. This point is called type inversion or space charge sign inversion (SCSI) as the space charge sign changes from positive to negative. Increasing the particle fluence beyond the SCSI point leads to more and more negative space charge values. Furthermore, this type inversion leads to the bulk Si changing from n-type material to p-type material which can affect the charge collection in underdepleted sensors. The depletion voltage rises accordingly and eventually reaches values that cannot be applied to the detector any more without resulting in breakdown. High-resistivity p-type sensors do not suffer from type inversion because the ini-

tial space charge is already negative before irradiation. The effective doping concentration after irradiation does change with time and can be accelerated at elevated temperatures and decelerated or frozen at lower temperatures. This effect is known as reverse annealing and it is one of the main reasons why measurements of irradiated detectors are performed at cold temperatures. One last note to mention is that the procedure of leaving samples for 80 minutes at 60°C is designed to get to the N_{eff} min. This reduces the depletion voltage and acts as a common reference point for comparing some measurements.

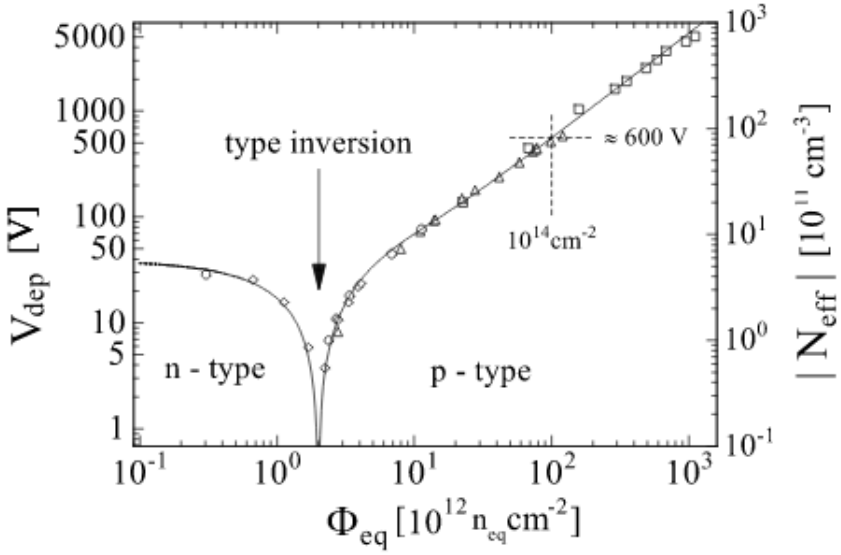


Figure 2.18: Depletion voltage V_{dep} and effective doping concentration N_{eff} of a 300 μm thick silicon detector up to a fluence $\Phi_{eq} = 10^{15} MeV n_{eq}/cm^2$.The n-type substrate is inverted to p-type for a $\Phi_{eq} = 10^{12} MeV n_{eq}/cm^2$. Image taken from [59].

Finally, defects associated to lattice interactions, could behave as trapping centres. The charge carriers generated by ionising particles or photons in the depleted bulk of the silicon sensor are travelling toward the electrodes and constitute the sensor signal. If a charge carrier is trapped into a defect level and not released within the signal collection time of the sensor, the charge is lost and the corresponding sensor signal is reduced. With increasing particle fluence, more and more charge carriers get trapped leading to signal reduction. The trapping is due to the different time constants of the electron capture and emission processes. Traps are mostly unoccupied in the depletion region due to the lack of free charge carriers. Trapping in irradiated devices is described by the trapping time constant τ_{trap}, which is the average time for a charge to be trapped after irradiation. This parameter is inversely proportional to the fluence Φ and defined by the following equation:

$$\frac{1}{\tau_{trap}} = \frac{1}{\tau_{trap0}} + \gamma\Phi \tag{2.19}$$

where τ_{trap0} is the trapping constant before irradiation and γ the proportional factor mod-

elling the dependence of the trapping time constant on the fluence. For most of the tracking devices used in particle physics this effect is not as relevant as the leakage current and depletion . Above a fluence of $10^{14} n_{eq}/cm^2$ about 90% of the generated charge in a 300 µm thick detector can be collected. However for the applications that this thesis targets, where silicon sensors are placed in radiation environments with fluences of at least $10^{15} n_{eq}/cm^2$, the amount of charge collected decreases to 50% due to trapping. As for the leakage current and the depletion voltage, the effective trapping damage constant depends on the annealing status of the sensor after irradiation. It has also been shown that the trapping times are only weakly temperature dependent [61].

As it has been presented in this section, after irradiation, the leakage current increases and the charge collected decreases. In consequence, the minimum threshold of the front-end defined as the minimum charge for which the electronics will react, needs to be adjusted, balancing the maximum acceptable noise hit rate and the minimum requirement of charge collection efficiency.

2.5.2 Total Ionising Dose Effects in ASICs

Ionising radiation generates electron–hole (e–h) pairs in insulators, such as the gate, field and spacer oxides (SiO_2) used in integrated circuits (ICs) that can lead to component degradation or failure via total ionising dose (TID). Therefore, the electronics (ASICs) in the detectors for High Energy Physics need to be designed bearing in mind the TID dose that they will be exposed to in order to avoid premature failures. Radiation sources typically used for TID testing are Cobalt-60 (^{60}Co) and Cesium-137 (^{137}Cs) radioisotopes and low-energy ($\approx$ 10-keV photons) x-ray generators. Radioisotopes are high-energy photon sources that are mainly used for device characterisation. Low-energy x-ray generators are instead commonly employed for transistors characterisation because of their relatively low cost, the possibility to reach elevated doses within short testing time and their high safety standards [43]. The total ionising dose is measured in Gray although it is more common to express it in rad or radiation absorbed dose: $1\ Gray = 1\ J/Kg$ (SI unit) and $1\ Gray = 100\ rad$.

When ionising radiation is passing through the oxide (SiO_2) electron-hole pairs are generated by the radiation energy deposited. The electron's mobility in the oxide silicon is $\mu_{n,oxide} \approx 20\ cm^2/V\ s$ which is almost ten times higher the mobility of the holes in the oxide $\mu_{p,oxide} \approx 2\ cm^2/V\ s$. This means that the electrons are swept out of the oxide in picoseconds while the holes remain trapped in the presence of an electric field. The accumulation of holes in the oxide results in a charge build up in this region that modifies the electrical behaviour of the detector [44].

A similar effect occurs in the shallow trench isolation (STI) or field oxides used to isolate transistors. The damage of these structures will cause an increase of the leakage current as the trapped holes can create a leakage path between source and drain. The effect is modulated by the distance of the oxide from the channel of the MOSFET and the length of the parasitic path,

which implies that the proper architecture for a CMOS device can vastly mitigate field oxide or STI leakage. Figure 2.19 shows a 3-dimensional (3D) schematic of an irradiated nMOSFET illustrating the formation of two lateral parasitic devices. The main nMOSFET is surrounded by the STI structure, as shown in light green. The front face of the STI structure is represented by the light-green frame to make the channel doping profile and the STI-related trapped-charge (+ markers) distribution visible.

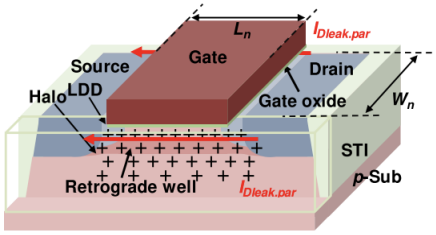


Figure 2.19: 3D schematic of an irradiated transistor showing the STI issue with the leakage current [41].

In order to reduce the effects of TID there are several mitigation techniques that can be applied while designing the electronics [42]. One of these techniques is the used of enclosed transistors for which the source-to-drain current flows underneath the gate, therefore eliminating any leakage path underneath the field oxide or along the active area edge. Another way of reducing radiation damage is the use of p+ guard rings to isolate different silicon areas.

2.5.3 Single Event Effects, SEE

Single Event effects (SEEs) are the third type of radiation damage on electronics and they are caused by the interaction of a single particle with the component's sensitive region [119]. In the context of a high-energy accelerator radiation field, these effects are typically caused by high energy hadrons (HEH, defined as hadrons above 20 MeV), intermediate energy neutrons (in the 0.2-20 MeV range) or thermal neutrons (with energies around 0.025 eV). The physical mechanism through which SEEs are induced is via indirect energy deposition events, meaning that a discrete interaction between the hadron and a nucleus in (or near) the device's sensitive region needs to occur to trigger the event which causes a permanent or temporary change in the operation of the ASIC. These events affect the electronics of the detectors and can lead to "hard" or permanent errors when single event burnouts, gate ruptures or latch-ups take place. However, they can also generate what are known as "soft" errors, which can can be split into two classes: transients created by Single Event Transients or glitches in the clock also named Single Event Effects [45]. The typical way these errors manifest themselves is when an ionising particle traversing a device depositing sufficient charge to create an event such as a memory bit flip. In consequence, SEEs can be classified in two different classes: destructive or hard effects

and non-destructive or soft effects. The most relevant damages concerning readout electronics for silicon detectors are given below [13].

- Destructive (hard) SEE's
  - Single Event Latchup (SEL): potentially destructive creation of a parasitic pnpn thyristor structure in a device. When a latchup occurs, the current increases and if the power supply is maintained, the device can be destroyed by thermal effects. The use of a current monitoring and a power control circuit allows the power to be shut down quickly after the latchup is detected in order to protect the device against thermal destruction [46].
- Non-destructive (soft) SEE's
  - Single Event Upsets (SEUs): single bit flip induced in a digital element either by direct ionisation from a traversing particle or by a recoiling nucleus emitted from a nuclear reaction. This event induces no damage to the basic element which can be rewritten with the right value.
  - Multiple Cell Upset (MCU) and Multiple Bit Upset (MBU): MCUs occur when two or more bits (physically adjacent or not) become corrupted by a single particle, or its secondary particles from nuclear interactions. If the two or more bits corrupted by the same event are in the same logical word, the effect is known as a Multiple Bit Upset (MBU). This effect is rarely observed in the devices studied for this thesis.
  - Single Event Transient (SET): momentary voltage excursion (voltage spike) at a node in semiconductors, originally formed by the electric field separation of the charge generated by an ion passing through or near a circuit junction. The signal perturbation is very short but can in some cases propagate through the whole system depending on where and when it occurs.

There are several techniques that can be applied to mitigate SEUs.

The scale down of technology plays an important role in the influence of SEE. From the technology level two scenarios can be differentiated. On one hand, with the technology scaling, the supply voltages are lower as well as the capacitance which means that less charge is needed to change a state. On the other hand, as the physical dimensions are smaller, it is less likely that a particle will hit the sensitive area. The overall behaviour is determined by the circuit topology and radiation environment.

At cell level, if the node capacitance is increased, the charge needed to be deposited by a heavy ion to induce errors is higher. It will also help to have the information stored in multiple nodes.

Finally at a system level redundancy can be introduced in the form of data encoding using an error-correcting codes such as a Hamming protocol [114] or with Triple Modular Redundancy

(TMR). This last technique is based on a majority voter cell that has an odd number of inputs, 3 being the most common number, and 1 output equal to the input value that is repeated the most. In the case of 3 inputs, an upset in one of them would be masked by the voter and will not be seen in the output. TMR is widely used in HEP.

Finally, as to what regards the general SEE test and rate estimation approaches and tools, the description of the sensitivity of a component is typically expressed as the probability per unit fluence, better known as the cross section, that an SEE occurs for a given incident beam. The different types of beam are characterised by their energy in the case of (hadron-induced) indirect energy deposition and by its Linear Energy Transfer (LET) for heavy ions. This quantity it is the stopping power that was described in Section 2.1.1 or dE/dX. The LET of a particle in a certain material is defined as its energy loss per unit length and density and is tabulated for a wide range of ions and materials.

In Chapter 6 a full analysis of SEEs for the detector that is foreseen to be used in the ATLAS ITk Upgrade will be presented.

Estimation of upset rates

To estimate the upset rate, one must consider the mechanism by which radiation particles cause the anomaly. Heavy ions have a high $\frac{dE}{dx}$ and hence low range so they will deposit their energy in a small region causing directly a SEU. Protons on the other hand tend to ionise the material in the interaction with electrons and have a lower $\frac{dE}{dx}$. However occasionally they will collide with a nucleus and have a nuclear reaction. This is basically a non ionising energy loss event which will result in a recoil of the nucleus and generate a lot of charge in a small volume. Figure 2.20 shows the two ways in which heavy ions and protons cause SEUs.

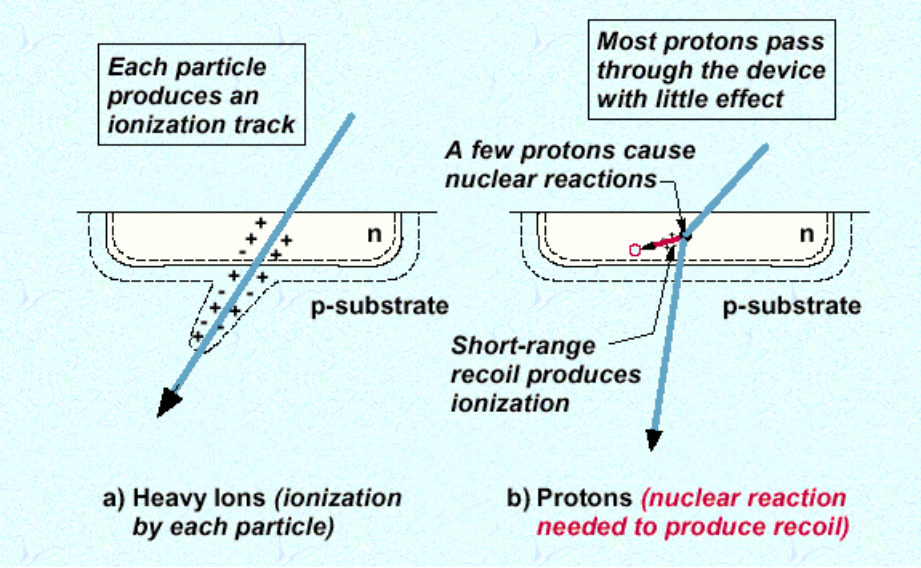


Figure 2.20: a) Generation of SEU via ionisation caused by heavy ions. b) Creation of SEU via nuclear reaction by protons. Images from [93].

Device immunity is determined by its Linear Energy Transfer Threshold (LET_{th}). The LET_{th} is defined as the minimum LET that will cause a single-event effect at a certain particle fluence. Devices foreseen to be working in radiation environments should be analysed for SEU rates and effects during testbeam campaigns, as the rate of this events during operation is low you need to go to dedicated facilities to be able to qualify the technology. In the following section results of the response of the chip to high energy ions, with different LET, measured in Louvain, Belgium, and to 480 MeV protons (HEH) at TRIMUF in Vancouver will be presented.

The LET_{th} usually reduces as a device accumulates large TID. The present trends (e.g., device size and power reduction, increased memory and speed) will only heighten the SEU susceptibility. This is easily seen when one considers the device as a simple capacitor (C) upon which the ionised particle deposits sufficient charge (Q) to result in a voltage (i.e., logic state) change. SEU occurs when LET > Q_{crit}.

Since the LET_{th} is equivalent to the LET required to produce a voltage change (ΔV) sufficient for an SEU, then mathematically:

$$LET_{th} \propto \Delta V = Q/C \tag{2.20}$$

As the size of these active devices decreases, the capacitance will decrease and so will the charge necessary to induce the SEU. If a square device of feature size L x L is considered, the critical charge for state change is proportional to the feature size squared ($Q_{crit} \propto L^2$). This critical charge is the charge necessary to flip a binary "1" to a "0" or vice-versa, but is less than the total stored charge. Specifically, Q_{crit} is then the difference between the storage node charge and the minimum charge required for the sensing amplifier to read correctly. For example, in SRAM circuits, Q_{crit} depends not just on the charge collected but also the temporal shape of the current pulse [94].

The energy deposited per unit path length as an energetic particle travels through a material is the linear energy transfer (LET). Note that the LET is normally defined by dE/dx. However, the LET used in SEU studies is actually the mass stopping power defined by $(dE/dx)/\rho$ where ρ is the material density. This results in an LET unit of $MeVcm^2/mg$ of material, which is the energy loss per density thickness. Density thickness (t_d) is the product of the material density and its thickness (t), i.e., $t_d = \rho \times t$. Therefore, the density thickness describes the area density of electrons (electrons/cm^2). The LET is dependent on the particle, its energy, and the material traversed. The upset rate is usually reported as errors per day per chip, or errors per day per bit (errors/bit-day). The practical way of calculating the SEU rate is by measuring the cross section (σ) versus LET for example using accelerator testing. The device cross section is defined as the ratio of the number of upsets to the particle fluence. The experimentally determined cross section is a function of particle energy (LET).

2.6 Sensor characterisation parameters

In order to determine if a sensor is suitable for a HEP experiment there are certain parameters that need to be evaluated. They will be explained in this section adapted to the depleted monolithic active sensors with small collection electrode designs.

2.6.1 Basic sensor electronics

The in-pixel electronics used in the detectors for High-Energy physics can be described by the block diagram seen in Figure 2.21. They consist of the sensor that captures the incoming particles charge and it is represented by the diode. This diode is connected to an amplifier whose output is digitised by discrimination stage that can be tuned via different DAC values to suit different tests. Finally the generated pulse will be processed by the digital logic. Having this picture in mind will help for a better understanding of the characterisation of these devices.

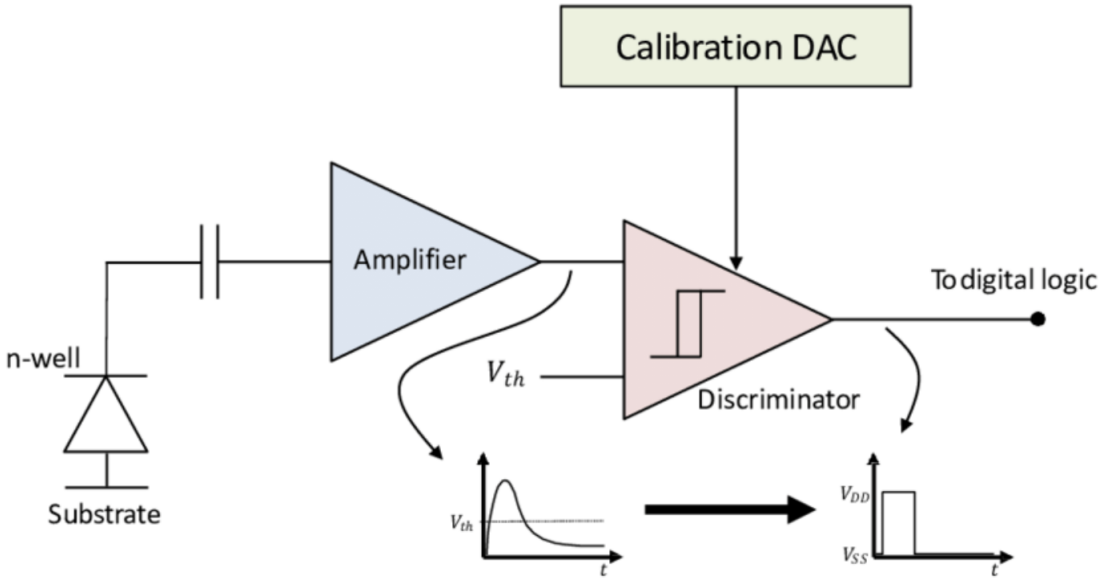


Figure 2.21: Block diagram of the pixel electronics that shows how the detected charge becomes a pulse at the output of the discriminator. Figure taken from [14]

2.6.2 Gain and charge collection efficiency

A certain amount of charge is deposited in the sensor by a traversing particle. This charge is then used to generate a signal by the front-end. The charge collection efficiency of a sensor, ε_{coll}, is defined as the ratio between the collected charge, Q_{coll}, and the generated charge, Q_{gen}.

$$\varepsilon_{coll} = \frac{Q_{coll}}{Q_{gen}} \tag{2.21}$$

The gain of an amplification circuit is defined by the ratio between the output voltage, V_{OUT}, and the input voltage V_{IN}. For the pixel detectors studied in this thesis $V_{OUT} = \frac{Q_{det}}{C_{IN}}$ where Q_{det} is the detected charge and C_{IN} is the input capacitance of the detector.

$$Gain = \frac{V_{OUT}}{V_{IN}} \tag{2.22}$$

This parameter is given by the front-end design which is explained in Section 3.1.2. In the MALTA sensor, due to its small input capacitance, for a collected charge of 1600 e- the voltage step created is around 50 mV. The sensor thickness is around 25-30 μm. To calculate the expected ionisation charge for this sensor a mean ionisation charge of 63 electron-hole pairs per μm path length [78] and a total electrode capacitance of 5 fF is assumed. In order to calibrate the sensor, fluorescence sources such as ^{55}Fe can be used as they produce point-like charge deposits of fix value in the sensor via the photo-electric effect. Measuring the signal amplitude at the front-end output for several Fluorescence lines therefore allows an accurate calibration and calculation of the gain of the sensor. The amount of collected charge can then be computed from the signal amplitude and compared to the expected deposited charge based on energy loss models as discussed in Section 2.1.1.

2.6.3 Noise

Having a large Signal-to-Noise ratio, SNR, is another important parameter of every detector. The goal of these devices is to generate the largest signal possible from the incident charge while keeping the lowest possible noise levels. For pixel detectors the most common way of expressing the noise is with the equivalent noise charge, ENC, which is the ratio of produced noise,V_{NOISE_OUT}, over the signal generated from a single collected electron,$V_{OUT\_1e^-}$.

$$ENC = \frac{V_{NOISE\_OUT}}{V_{OUT\_1e^-}} \left[\frac{V}{V/e^-}\right] \tag{2.23}$$

Another way of expressing this noise is:

$$ENC^2 = a_{shot}\tau + a_{1/f}C_D^2 + a_{therm}\frac{C_D^2}{\tau} \tag{2.24}$$

where a_{shot} is the shot noise contribution which can be modelled with a Poisson distribution and is due to the discrete nature of charge, τ is the filter time of the in-pixel circuitry, $a_{1/f}$ is the 1/f noise contribution and C_D refers to the detector input capacitance that depends on the capacitance between the collection diode and the deep p-well as well as the sensor back side and a_{therm} is the thermal noise contribution.

For MOSFET transistors the shot noise term can be neglected as it is linearly proportional to the leakage current of the sensor. τ is the filter time of the front-end amplifier and it usually is $<= 1$ μs making this term also negligible.

On the other hand, the thermal noise contribution is more important and it derives from the Brownian motion of charge carriers in the channel of the MOSFET [79]. a_{therm} can be defined for strong inversion i.e when the voltage between the gate and source is larger than the threshold voltage, Equation 2.25, and for weak inversion i.e when the gate-source voltage is below the threshold and there is a small current that depends on V_{GS} voltage, Equation 2.26:

$$a_{therm} = \frac{4K_B T\gamma}{g_m} \tag{2.25}$$

$$a_{therm} = \frac{2K_B T n}{g_m} \tag{2.26}$$

where K_B is the Boltzman's constant, g_m the transconductance defined as dV_{GS}/I. The factor γ can be approximated by 0.67-1.0 for CMOS processes and n is the weak inversion slope. The factor n is a bias dependent parameter.

Finally, the 1/f contribution, which is also known as Random-Telegraph-Noise is also important [15]. According to literature, the term $a_{1/f}$ can be defined in several ways but within the context of this work the definition taken is:

$$a_{1/f} = \frac{K_f}{C_{ox}^{'2} WL}\frac{1}{f} \tag{2.27}$$

where $C_{ox}^{'}$ is the oxide capacitance per unit area in a MOSFET, W and L are the gate width and length of the transistor and K_f is a technology dependent constant for the 1/f noise. In other words it can be said that this noise is caused by a change in the charge state of a trap or defect under or near the transistor gate. This abruptly changes the threshold voltage of the transistor and hence its current. In most hybrid pixel detectors the $C_{ox}^{'}$ can be neglected due to the high input capacitance of the sensor. This is not the case for DMAPS as the sensor input capacitance C_D is in the same order of magnitude as $C_{ox}^{'}$. Consequently, the thermal noise component because of its frequency independence, dominates for very fast charge deposits and the total SNR of a detector can be expressed as:

$$\frac{S}{N} \propto \frac{Q}{C_D}\sqrt{g_m} \tag{2.28}$$

To sum up, the recipe to design a low noise DMAPS is based on having the smallest possible sensor input capacitance, C_D and the largest possible $C_{ox}^{'}$ to reduce RTS noise. For the sensors described in this thesis, the latter two contributions: random telegraph noise and thermal noise are the ones that have most influence and certain parameters have been optimised to reduce their effect.

2.6.4 Efficiency, threshold, time over threshold (ToT) and time-walk

The efficiency of a detector is possibly its most important parameter and can be expressed as:

$$\varepsilon = \frac{N_{det}}{N_{total}} \tag{2.29}$$

where N_{det} is the number of detected particles and N_{total} is the total number of incident particles. This is the broadest definition possible of detector efficiency. Nevertheless other parameters can be specified, like in the case of the ATLAS detector, all the particles should be detected within a 25 ns window. Furthermore, some pixels might be too noisy or not working so they will be removed or masked, minimising the efficiency over a certain area. Consequently, more specific definitions of efficiency exist such as in-time efficiency, global efficiency and in-pixel efficiency for which different cuts to the data are applied during the analysis stage.

The most influential parameter in the efficiency of the detector is the threshold, which can be defined as the minimum incident charge for which the discriminator, seen in the block diagram of Figure 2.21, fires. The lower the threshold the better the efficiency value is providing that the noise hits do not compromise the operation. For pixel detectors, the global threshold which is applied across the pixel matrix typically ranges from a few hundred to a few thousand electrons, depending on the sensor design. Pixel designs with large collection diodes usually operate at a higher threshold than small electrode designs due to the higher input capacitance and larger signal. Irrespective of the detector design (monolithic or hybrid), per-pixel thresholds tend to vary with respect to the global threshold and this phenomena is known as threshold dispersion. Even when operating at high thresholds some residual noise is observed if the dispersion is large. As a result, many front-ends support a per-pixel threshold fine tuning to minimise this effect and to allow a lower overall threshold, which is not the case for MALTA as the area penalty was too high. The way of measuring the threshold is by injecting a fixed charge into the front-end electronics. This injection is repeated multiple times, and the percentage of injections that result in a hit being readout is recorded. The value of the injected charge is then increased by a discrete voltage step, and the process is repeated until a specified charge range has been covered. Ideally, this process would produce a step function, with zero injections resulting in a hit for any charge below threshold and all injections resulting in hits for any charge above threshold. However, due to the electronic noise of each front-end channel, this step function is smeared into an S-curve shape. The S-curve is then fit with a Gaussian error function. The mean value of this fit is recorded as the threshold for each pixel. The σ of the fit defined as the difference between the charge values that correspond to 16.5% and 83.5% hit efficiency, is regarded as the noise of the pixel [126]. Figure 2.22 shows in a) the ideal s-curve or step function. If our detectors were perfect, every charge injected below threshold will produce zero detected hits and every injection above threshold will make the pixels fire, however in reality, the behaviour is as shown in Figure 2.22 b), where injected charges close to threshold will fire some times but neither zero

nor all the times.

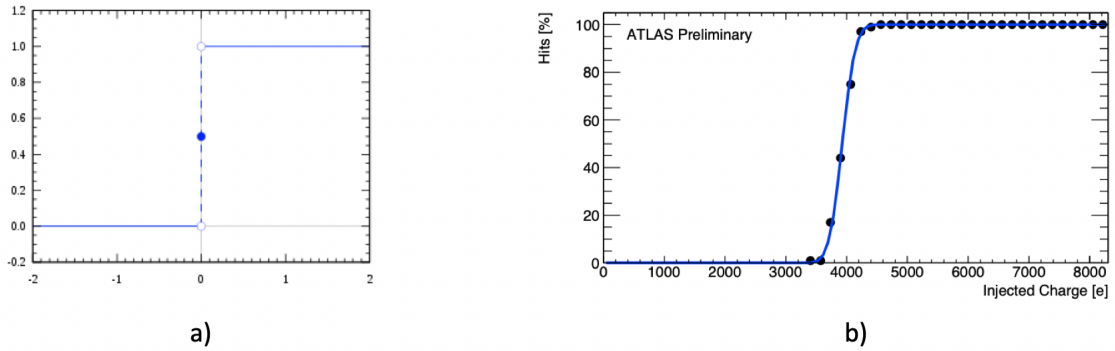


Figure 2.22: (a) Heaviside step function or unit step function with image taken from Wikipedia (b) ATLAS threshold curve distribution [84]

Other parameters that have a large dispersion and are relevant to particle physics measurements are time-walk and time over threshold or ToT. The first one is defined as the delay between the deposition of the charge in the sensor and the pixel firing. Here is where the 25 ns time-window, which corresponds to a bunch crossing in the LHC, comes into play as the time-walk should be kept below this number so the particles can be reliably identified as belonging to a particular bunch-crossing. The ToT is the time that a signal stays above the discriminator threshold. With this information, the amount of deposited charge or the size of the signal can be estimated. Having a well calibrated detector means that the time resolution will be as accurate as possible [80].

Figure 2.23 shows how these parameters will be computed for two different signals, one created after a high-charge deposition and the other with a lower-charge impact. The signal of high charge rises steeply and therefore does not need a lot of time to reach threshold. The ToT counter clock starts with only a small delay and thus time-walk is short. However, the signal with low charge takes more time to reach threshold and thus its time-walk is longer. Since ToT counts are a measurement of the signal charge, determining the relation between ToT counts and time-walk is of great importance to determine the Time of Arrival (ToA) of a hit.

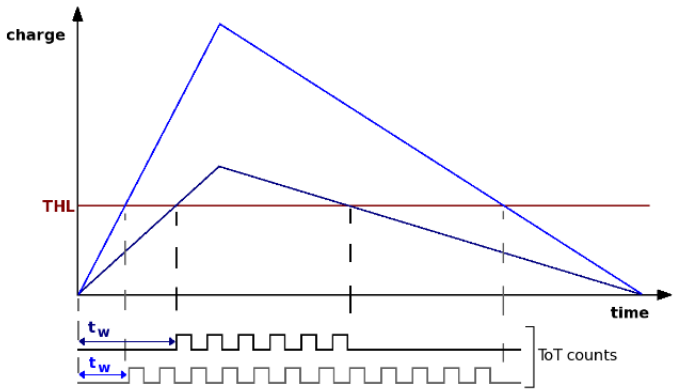


Figure 2.23: Signal created by two hits of different charge. Low charge corresponds to little ToT counts and long time-walk t_w. High charge leads to many ToT counts and shorter time-walk [80].

Figure 2.24 is a measurement performed with the MALTA detector that shows the relationship between the amplitude of the signal created by the front-end and the time-walk. It was taken by exposing the chip to a ^{90}Sr radioactive source that emits electrons that generate a signal similar to the response to minimum ionising particles. Therefore, the expected most probable value of charge deposition for the 25µm thick epitaxial layer of the chip is around 1500 e^-. The charge threshold setting of the front-ends with diode reset was around 200 e^-. The most probable amplitude value obtained by the most probable charge deposition of the MIP is around 500 mV. It puts into perspective the importance of being able to operate at low thresholds as the time-walk delay will be smaller and the ToT measurement can be more precise. In this case signals above 130 mV arrive with a delay smaller than the cited 25 ns which corresponds to 300 e^- for the operating front-end measurements. These numbers correspond to $\approx$ 95% in-time threshold detection which means that when operating with a 300 e^- threshold 95% of the hits will be processed within the 25 ns window of the LHC allowing for good particle identification. It is important to note that the MALTA detector has an asynchronous architecture explained in detail in Chapter 4, so measuring ToT is not possible as there is no clock that goes to the matrix. In consequence, this time-walk estimation is a key measurement to have an estimation of the charge deposited by the incident particles. The time shown corresponds to when the hits are available in the periphery which is "immediately" after the discriminator fires.

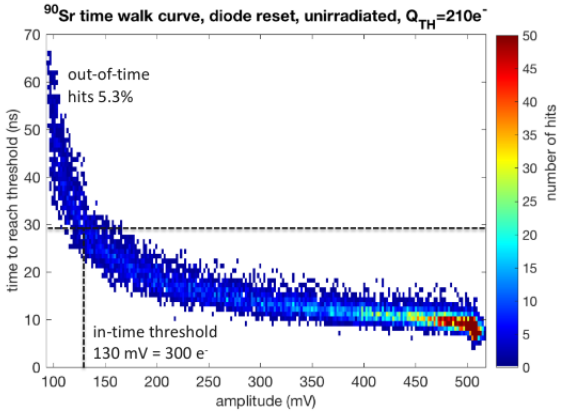


Figure 2.24: Signal amplitude versus time-walk for MALTA detectors with ^{90}Sr source zoomed in the time axis to focus on in-time efficiency measurement (not all ^{90}Sr energy spectrum covered). Curve provided by [77].

2.6.5 Cluster size

The cluster size of a detector is a very important quantity that defines the spatial resolution. It is defined as the number of adjacent pixels that fired simultaneously. The cluster size depends on parameters like the pixel pitch, operating threshold, bias voltage of the detector or thickness of the depletion area because the thicker the sensitive regions, they give more lateral diffusion and therefore the charge spread over more pixels. A detailed study of the cluster size will be presented in Chapter 3.

All the parameters discussed are influenced by the radiation effects explained in Section 2.5. It is important to do the measurements for both devices irradiated and unirradiated and then compare the degradation.

Chapter 3

Particle detection with Monolithic Active Sensors

This chapter will describe the different small collection electrode DMAPS that have been developed within the context of the ATLAS Upgrade. It explains how the sensor and front-end electronics have evolved in order to comply with the specifications of the outermost layer of the Inner Tracker detector of the ATLAS Upgrade. It focuses on how to achieve a good charge collection and the modifications that have been done to improve the radiation tolerance of the initial MALTA detector. It contains experimental results on cluster analysis and efficiency for the Mini-MALTA and MALTA Cz sensors that have been obtained in both laboratory and testbeam campaigns for this thesis work.

3.1 Description of the MALTA detector

The MALTA chip whose name is taken from "Monolithic from ALICE To ATLAS" is the largest monolithic CMOS sensor designed to meet the ATLAS requirements [90]. It has been designed in TowerJazz 180 nm with the standard modified cross-section. Some wafers were fabricated with an epitaxial layer of 25 μm and others with 30 μm with the idea of studying the charge collection differences. The substrate layer thickness has also been thinned to different values 50 μm to 100 μm and 300 μm mainly to reduce material and scattering. It has a matrix of 512×512 pixels with a size of 36.4×36.4 μm^2. The physical chip size is $\approx 2 \times 2$ cm^2. Its layout is illustrated in Figure 3.1.

The main part of the chip is the large pixel matrix. The digital periphery contains parts of the readout logic as well as the function control registers used to defined the best operational mode according to the application of the detector. In the case of bias currents and voltages needed by the front-end pre-amplifier, the register values are converted to the desired current or voltage values using digital-to-analogue converters (DACs).

The digital address and timing information of the pixels that have been hit is transmitted

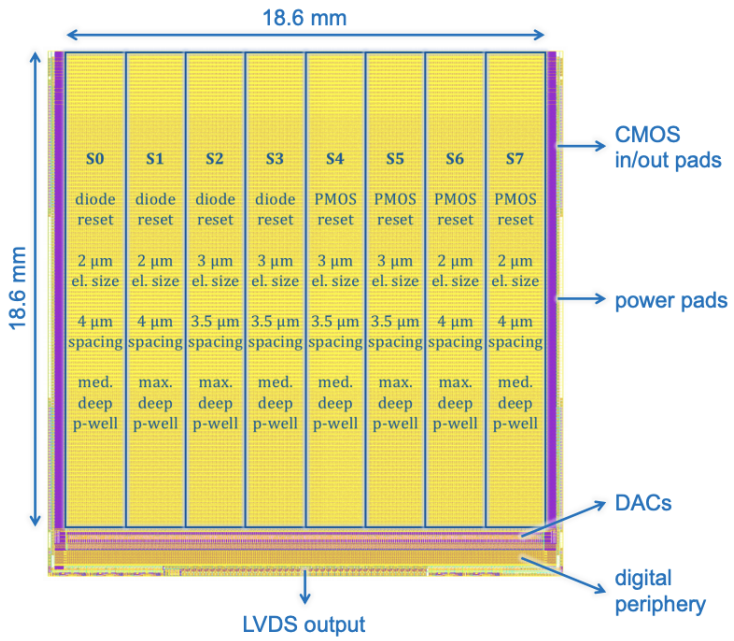


Figure 3.1: Layout of MALTA [77].

off-chip in a 40-bit parallel output, that utilises either a low-voltage differential signal (LVDS) standard or a full-swing 1.8 V CMOS standard. The LVDS drivers [75] are designed to operate up to 5 Gb/s and ensure robust data transmission to off-chip data acquisition systems, while the CMOS input/output pads offer the possibility of chaining multiple chips together and transmitting the hit data between chips.

The pixel matrix is divided into eight sectors with minor changes in the sensor and front-end preamplifier designs. Sectors 0-1 and 6-7 have a circular collection electrode of 2 µm diameter and there is a 4 µm spacing between the electrode and the deep p-well that surrounds the electronics. In sectors 2-5 the circular electrode diameter is 3 µm and the separation is 3.5 µm. Another significant distinction between the sectors is the deep p-well coverage inside the pixel. The deep p-well is only needed to shield the n-wells of the PMOS transistors and therefore half of the sectors implement what it is shown in Figure 3.1 as med. deep p-well or medium deep p-well where this layer has not been extended below the p-wells of the NMOS transistors. The sectors with max. deep p-well or maximum deep p-well have this layer extended over both types of transistors in the same way as shown in Figure 2.14.

Concerning the front-end electronics, that will be discussed in the following section, there are no differences between sectors except for the mechanism used to reset the voltage of the collection electrode, which can be either done with a PMOS transistor or with a diode.

Finally the in-pixel circuitry, as shown in Figure 3.2 has two main parts, the analogue that includes the front-end circuitry and the digital with all the logic needed to readout the matrix. It follows a similar implementation to the diagram shown in Figure 2.11. From the layout it can be observed that the collection electrode is situated in the middle of the pixel and that the

electronics are separated from it either by 3.5 or 4 µm. There is as well shielding between the analogue and digital parts in order to avoid any possible cross-talk and two separated power domains are in placed for the same reason.

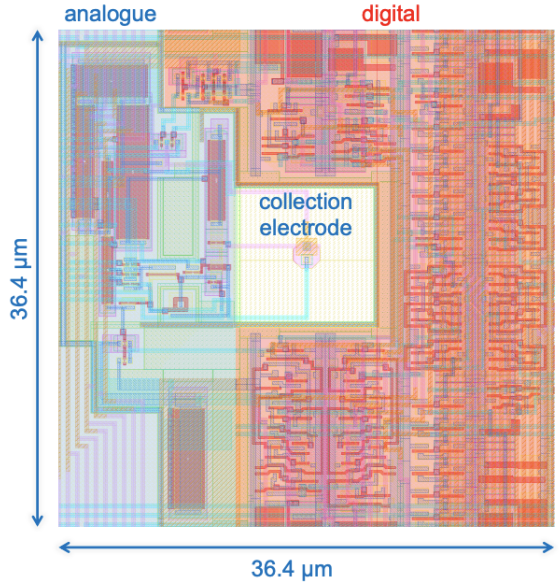


Figure 3.2: Layout of a pixel from the MALTA sensor [77].

3.1.1 MALTA main features

MALTA detector main design characteristics are specified in Table 3.1 and the specifications that it should fulfil to be compatible with the outermost layer of ATLAS Inner-Tracker Upgrade [81] are given in Table 3.2.

Table 3.1: Main characteristics of MALTA and Monopix sensors.

	MALTA
Sensor Size	2/3 μm collection diode 3.5/4 μm spacing to electronics
Pixel Size	36.4 x 36.4 μm^2
Readout Architecture	Asynchronous
Fabrication process	TowerJazz 180 nm standard modified process with continuous n-blanket
Time Resolution	Hits available in the periphery according to timewalk

Table 3.2: Main specifications for the sensors in the outermost layer of the ATLAS Inner Tracker Upgrade.

Requirement	Unit	Value
Detection efficiency	%	> 97
Time resolution	ns	25
Particle rate	MHz/mm^2	1
Non-ionising radiation fluence	1 MeV n_{eq}/cm^2	10^{15}
Ionising radiation dose	Mrad	50
Power consumption	mW/cm^2	< 500
Material budget	% of x/X_0	< 2

3.1.2 MALTA front-end electronics

The analogue front-end in MALTA consists of the following circuits: the reset of the voltage of the collection electrode after a hit, an amplifier and a discriminator or comparator.

The mechanism of the front-end behaviour, the equivalent circuit of which is shown in Figure 3.3 and relies on the transfer of charge from a large capacitance, the one of the sensor, to a small capacitor in order to generate voltage gain. The charge deposited by the incoming particles is transformed into a negative voltage step $\Delta V_{PIX\_IN} = Q_{IN}/C_{IN}$ at the input terminal. The transistor M1 with current source IBIAS from VDDA acts a source follower, meaning that the input voltage will be almost the same at the source. This results in the following charge transfer $Q_{source} = C_{source} \times \Delta V_{PIX\_IN}$ from C_{source} to C_{OUT_A} in case the current sink to GNDA is IBIAS. So ideally ΔV_{OUT_A} would be:

$$\Delta V_{OUT\_A} \approx \frac{Q_{source}}{C_{OUT_A}} = \frac{C_{source}\Delta V_{PIX\_IN}}{C_{OUT\_A}} = \frac{C_{source}}{C_{OUT_A}}\Delta V_{PIX\_IN} = \frac{C_{source}}{C_{OUT\_A}}\frac{Q_{IN}}{C_{IN}} \tag{3.1}$$

In consequence, the voltage at the output becomes much larger due mainly to the fact that $C_{Source} \gg C_{OUT\_A}$ or expressed in a different way $C_{OUT} \ll C_{IN}$.

The circuit practical implementation, with the amplifier and discriminator, can be seen in Figure 3.4. Transistor M0 behaves as the current source that supplies the IBIAS current to the source follower M1. M1 and its capacitance CS are placed in their own well and have a gain close to 1. Transistors M5 and M6 shape the voltage that will be seen in OUT_A, setting the correct baseline and the return to it after a hit. The bias voltage VCASN and the gate-source voltage of M6 that conducts the current ITHR determine the DC voltage on OUT_A. M2 is a cascode transistor and therefore it is used to negate the Miller effect, preventing capacitive coupling between the input, IN, and the analogue output, OUT_A. Transistor M3, takes the

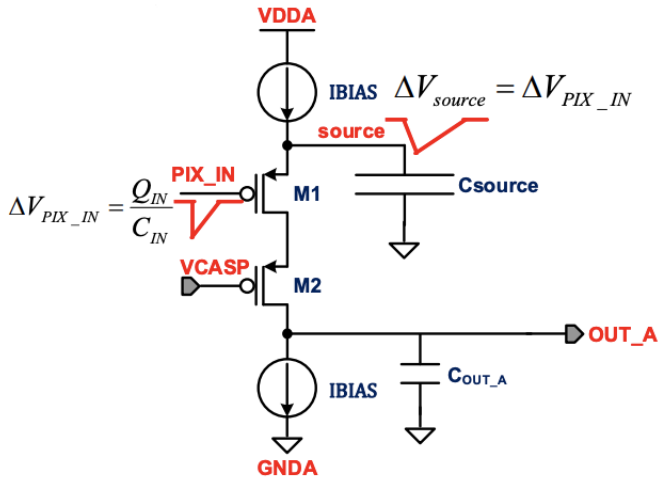


Figure 3.3: Operating principle of MALTA front-end [76].

sum of the currents IBIAS and ITHR. A traversing particle will increase the voltage at OUT_A provoking transistor M6 to go out of saturation so ITHR will charge the GN node. The current seen by M3 becomes larger and it will eventually discharge OUT_A node bringing it back to its initial state. It is important to underline the role of the capacitor CS that determines the voltage gain and it is also controls the return to baseline. The final shape of the signal depends on the biasing currents IBIAS and ITHR and on the CS value. The functon of M4 in this circuit is to avoid having too long signals when the input charge is large, therefore it only conducts when OUT_A voltage is above a certain value which can be controlled using the VCLIP bias. Finally transistors M7 and M8 constitute a simple discriminator. In steady-state conditions, the baseline voltage on OUT_A sets the DC current of M9 and the whole discriminator branch. M8 is a cascode transistor that decouples in this case OUT_D from OUT_A. M7 is biased in a way that provides a current IDB higher than the DC current of the branch. However it is pushed out of saturation while OUT_D is close to the supply voltage of 1.8 V. During the transient or hit, when the OUT_A voltage increases, the current drawn by M9 also increases until it becomes larger than IDB and then it starts to discharge the OUT_D node. While the signal in OUT_A returns to its baseline value, OUT_D will be charged up again by the IDB current. Consequently the threshold of the discriminator is controlled by the combination of the DC current setting in the branch (indirectly by VCASN) and the IDB current setting. When the OUT_D voltage drops below the threshold voltage of the following inverter during the transient, a digital pulse is produced at the output of the front-end [77].

The front-end uses what it is called a continuous reset of the collection electrode and has two different implementations. The first one, the reset diode, realised through a p+ implant in the collection n-well is very compact. However it has the drawback that the reset current

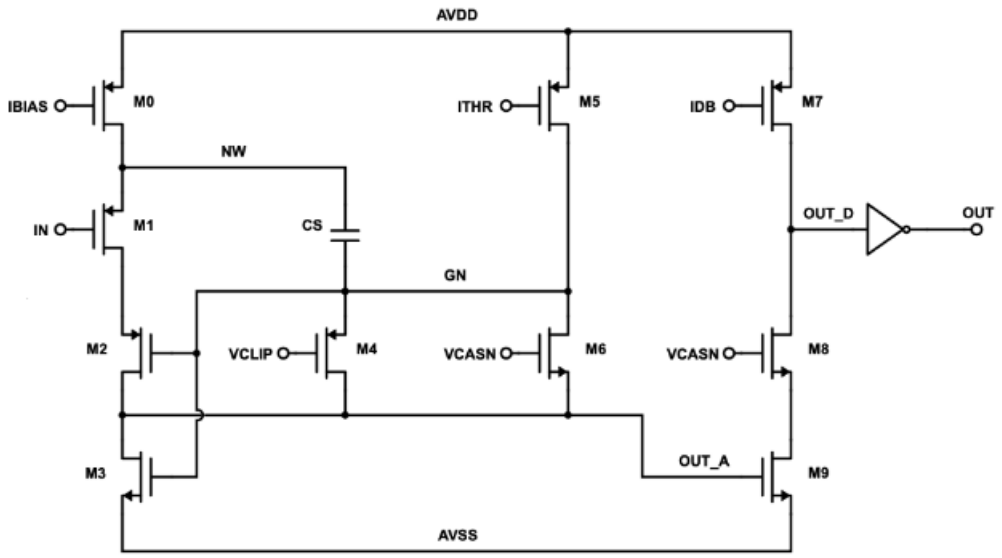


Figure 3.4: MALTA front-end schematic [89].

depends on the sensor leakage current and on the signal amplitude. The second one, the PMOS reset, needs a larger area and its contribution to the input capacitance of the sensor is larger. It has the advantage that the PMOS conductance is controlled by the bias current IRESET that is larger than sensor leakage current. Using these two ways of resetting the collection electrode means that the operation point of the different parts of the matrix will need slightly different bias settings. The circuit representation of both resets can be found in Figure 3.5.

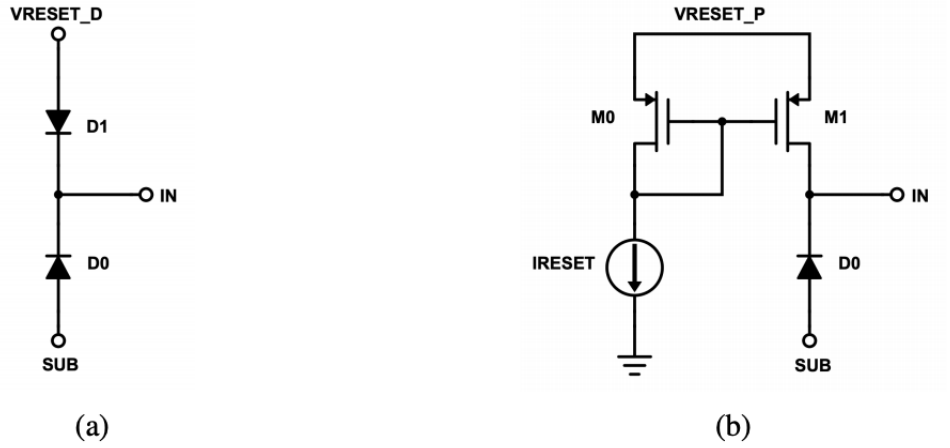


Figure 3.5: MALTA's electrode reset mechanisms [77], in a) the diode reset and in b) the PMOS reset.

The last features of the the analogue front-end include the masking and pulsing circuitry. In case there are noisy or malfunctioning pixels, the masking mechanism allows to disable them so results are not perturbed. The most common way of implementing this functionality is with

a latch that will enable or disable the output. For the MALTA design in order to save some area necessary to keep the pixel pitch small, it has been done with three parallel NMOS transistors situated between the source of transistor M9 in the discriminator stage and the analogue ground AVSS. The gates of these transistors are connected to the digital signals used for addressing every pixel by its row, column and diagonal. If the three lines are set to "1" or whenever these three signals intersect, the pixel will be disabled. A disadvantage of this scheme is that if more than one pixel is masked other unintended pixels will be disabled because of the intersection created by the three coordinates grid. For the analysis of the results these pixels are referred to as "ghost" pixels. Using three coordinates to implement the masking scheme is a compromise between saving some area in the front-end and being able to minimise the number of ghost pixels.

Being able to input test pulses is a very useful functionality, as it allows the response of the pixels to a fixed injected charge to be studied. A V_{PULSE} signal is sent to a specific pixel selected by its row and column number. Two PMOS switches force the output of the pulsing circuit to one of the two preset DC voltage levels: V_{HIGH} or V_{LOW} that determine the value of the charge injected creating a signal that is a voltage step with an amplitude of $V_{HIGH} - V_{LOW}$. This signal is coupled to the input node of the front-end by a 230 aF metal-to-metal capacitor. This capacitance value has been obtained via the parasitic extraction engine of Cadence tool Virtuoso. Nonetheless, its real value depends on other "settings" like bias voltage so it becomes complicated to have a very systematic measurement of the threshold or a good calibration. This injection circuit will be addressed in upcoming designs. As an approximation, the charge injection can be calculated as follows:

$$Q_{IN} = C_{IN}\Delta V_{IN} = C_C(V_{HIGH} - V_{LOW}) \tag{3.2}$$

The pulsing circuitry described that is connected to the input node can be seen in Figure 3.6.

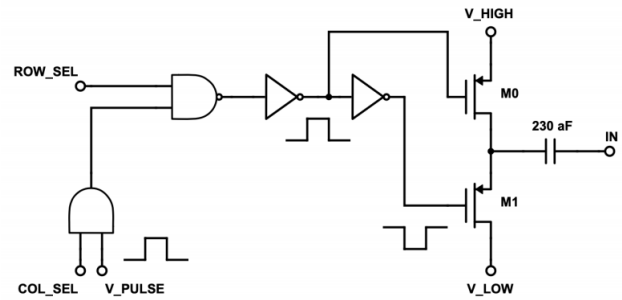


Figure 3.6: MALTA pulsing circuitry schematics used to capacitively inject test pulses to the input of the front-end [77].

The readout circuitry will be explained in Chapter 4 but it follows an asynchronous architecture so hits are transmitted "immediately" after the discriminator fires to the periphery via

LVDS signals. The DACs are common for the full matrix and have 7 bits to act on the bias of the different transistors in the FE electronics. In the characterisation results the two DACs settings that have the biggest influence on the threshold are specified. These are ITHR because it has a direct influence on the gain of the front-end and IDB as it influences the threshold level of the front-end comparator.

3.1.3 MALTA sensor characterisation results

In this section, the main results of the MALTA characterisation will be discussed as it shows what needs to be corrected in order to successfully meet the requirements of the experiment. The cross-section of the MALTA sensor had been changed with respect to the ALPIDE to achieve full depletion and therefore charge collection by drift rather than by diffusion mainly because of the higher radiation levels in ATLAS. Nevertheless, after testing sensors irradiated to $10^{15}\ MeV n_{eq}/cm^2$, it was found that the charge was not collected when deposited in the corner of the pixels [88]. All the sensors referred to as irradiated in this thesis, have been exposed to neutrons at the NIEL fluence specified at the TRIGA reactor in Ljubljana [92]. Figure 3.7 shows the efficiency maps of a 2 x 2 pixel region of the MALTA chip. Before irradiation, the efficiency is around 98% but after it drops due to the charge loss at the corner of the pixels because the lateral electric field near the pixel borders is too low for charge to be collected by drifting. The threshold for the irradiated sensor is as low as the RTS noise allows. This means that the sensor is still not fit for purpose and further improvements are required.

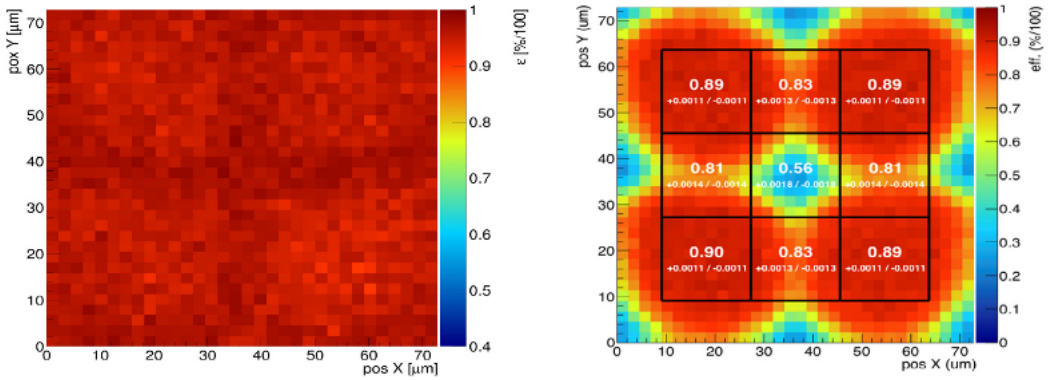


Figure 3.7: Efficiency map of a 2 x 2 pixel region of the MALTA chip, on the left before irradiation for 250 electrons threshold and on the right after $10^{15}\ MeV n_{eq}/cm^2$ *fluence for a 350 electrons threshold [88].*

Another indication of this charge loss can be seen with the cluster analysis [101]. In Figure 3.8 it can be seen how the cluster sizes evolve according to the threshold and irradiation levels. For the unirradiated pixel maps, the centre of the pixels have a cluster size of ≈ 1, while if we lower the threshold it is clear that more charges deposited at the corners of the pixels is collected. For the 600 e^- threshold the cluster size average at the corners is around 1.2 while for the lowest

operational threshold this number increases to 2.4. It is normal to have higher cluster sizes in the corners as the charge can be collected from the adjacent electrodes. For the irradiated, the centre of the pixels still has an average cluster size of ≈ 1 but the corners have an average of ≈ 1.1, even with a lower threshold applied. This is further proof that the charge deposited in the corners is not reaching the electrode.

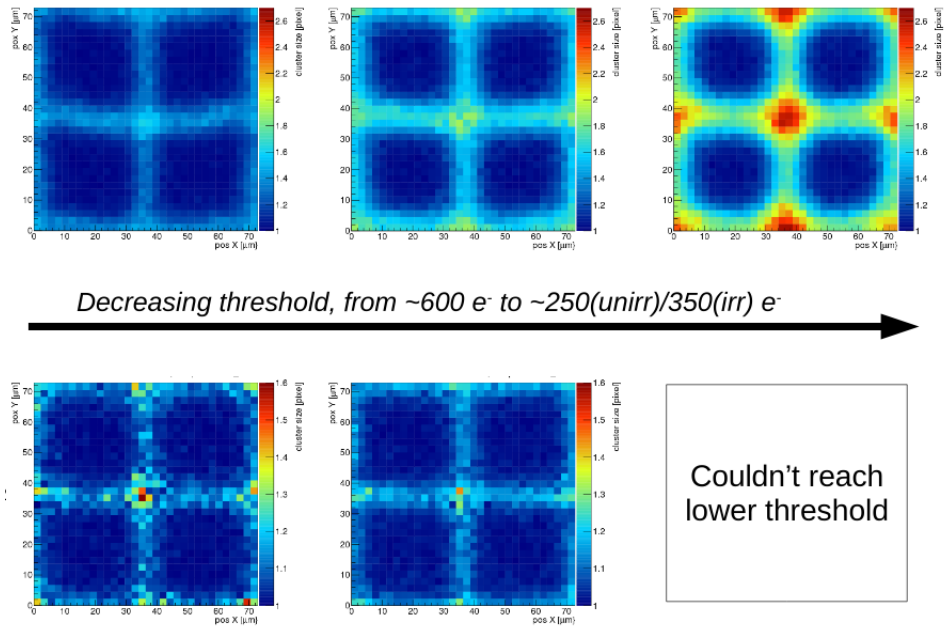
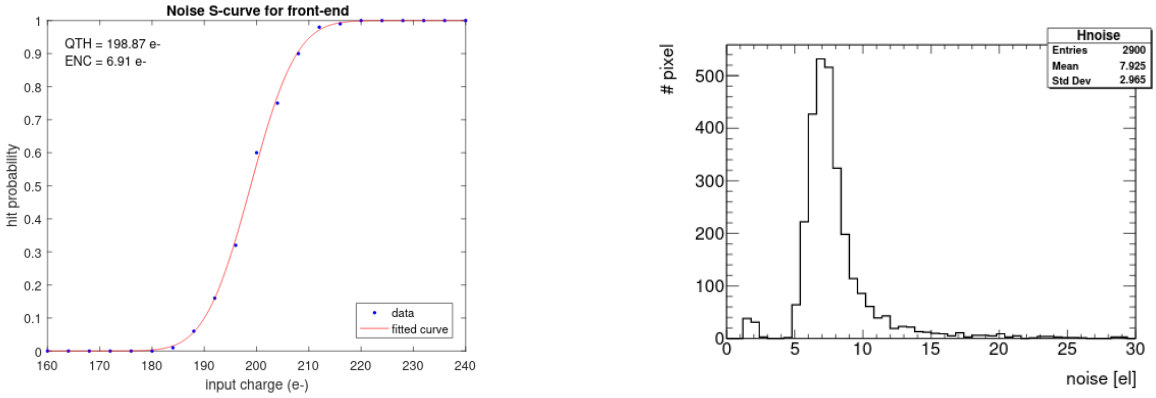


Figure 3.8: Cluster analysis for a 2 x 2 pixel region of the MALTA sensor unirradiated (upper set of plots) and irradiated to 10^{15} $MeVn_{eq}/cm^2$ (bottom set of plots) for 600, 400 and 250 electrons threshold respectively from left to right. The lowest threshold for the irradiated sensor is not achievable so it is not represented.

There was another unexpected finding during MALTA testing that needed to be addressed. Figure 3.9 shows both the simulated S-curve for the front-end and the measured distribution of the noise, done with the in-pixel injection mechanism. The simulated S-curve shows that the expected threshold value is around 200 e^- and the ENC noise is 7 e^- giving a S/N ratio of 28.6. Usually, with a threshold larger than 10σ a lower operational threshold could be achievable. However, the distribution of the noise shown in Figure 3.9 b) shows a non-Gaussian distribution with a tail due to a small number of pixels having noise value larger than 10 e^-.

The larger values of noise were believed to be due Random Telegraph Noise. This was attributed to "M3" transistor being much smaller than on previous circuits. To verify this assumption, the Mini-MALTA sensor includes sectors with the same "M3" size as on MALTA and sectors with a larger M3 transistor. Measurements on the Mini-MALTA sensor confirmed that the larger transistor sizes significantly decreased the RTS noise, both before and after irradiation. In addition, the larger NMOS size also yielded a significantly larger gain and lower charge threshold for the same settings. Further measurements confirmed that the "M3" output conductance was higher than expected, which caused gain degradation for the front-end. It was

also found that for lower threshold settings, where the influence of this output conductance is larger, the spread on the gain and hence the threshold spread increased.



(a) *(b)*

Figure 3.9: (a) Simulated S-curve of MALTA front-end (b) Noise distribution of MALTA [77].

Finally, the slow control block in the MALTA detector was implemented following an Ethernet like protocol but didn't function as foreseen. The clock tree was not generated properly which meant that to send the commands the power supply had to be toggled from 1.8 V to 1.2 V. This reduced the speed of the clock and helped with loading all the registers. Once the chip is configured, the supply had to be returned to 1.8 V to properly bias the transistors. The chip could be configured by doing the toggling of the supply but scans such as measuring the threshold of the detector were impossible to perform for the full matrix as they would take a very long amount of time. Therefore, the threshold cited for MALTA usually corresponds to the pixels in the region being studied. This issue was corrected for Mini-MALTA, in which a shift register was used for configuring the detector.

3.2 Mini-MALTA

The efficiency loss in the corner of the pixel was studied with detailed TCAD simulations. The goal was to enlarge the lateral electrical field and hence to accelerate the charge collection at the edge of the pixels especially for the irradiated sensors [86]. The CMOS circuitry for all the sensors would still be located inside the deep-p-well and the charge collection would continue to be done at the small n-type electrode. In September 2018, a new prototype, the Mini-MALTA, was designed with a number of process variations to address issues identified in the MALTA chip. The Mini-MALTA chip matrix contains 64 x 16 square pixels with a pitch size of 36.4 µm, the same as in MALTA. The full chip size is 1.7 x 5 mm^2 including both the matrix and all the periphery with the digital circuitry. Inside the Mini-MALTA sensor there are 8 different pixel flavours, differing in analogue front-end designs, reset mechanism and electrode/well geometries

as detailed in Figure 3.10.

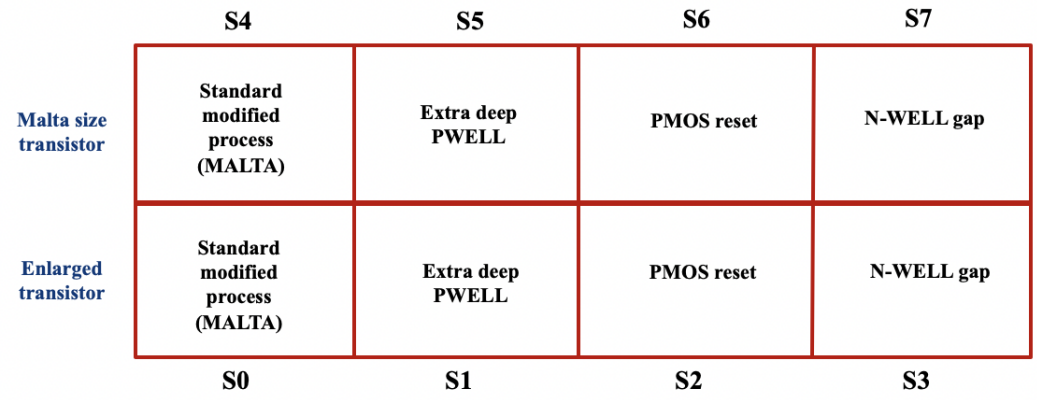


Figure 3.10: Different sectors of Mini-MALTA.

The matrix is divided in two halves, at the top, the transistor size has been kept at the same dimensions as was used in the MALTA design. For the bottom half of the matrix critical transistors for the front-end performance have been enlarged. This is to mitigate the significant Random Telegraph Noise measured on MALTA, which prevented low threshold operation. By enlarging the transistor size, fluctuations due to trapped charge will be smaller. The enlarged transistors are the input, M1, and the NMOS transistor of the current source, M3, that absorbs the current going to the input see Figure 3.4. Another important effect of this change in size of the critical transistors is that the gain of the amplifier for this half of the matrix becomes higher due to a lower output conductance of the enlarged transistor. The combination of higher gain and the lower noise allows operation at lower thresholds, which will improve the detection of charge in the corners.

In Mini-MALTA, these process variations are emulated using mask variations for different sectors in the matrix. The extra deep p-well modification is shown on the right image in Figure 3.11. In this, the standard modified process from TowerJazz has been further modified with an additional deep p-type implant under the deep p-well. The middle cross-section represents the n-well gap sector of Mini-MALTA for which the low dose n-blanket mask has been modified to include a gap. The left image is the cross-section corresponding to the Standard Modified Process, which is the same as in MALTA. The purpose of both changes is to improve charge collection for charges generated near the pixel edges by creating a stronger lateral field which focuses the generated charge (electrons) to the collection electrode.

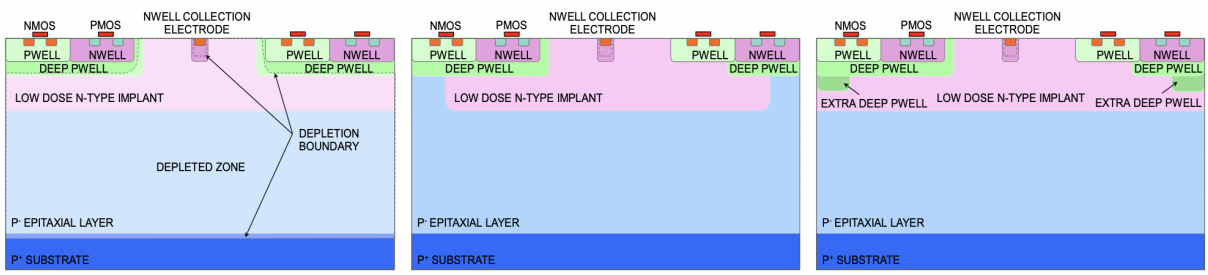


Figure 3.11: From left to right: standard modified process, process modification and new mask with gap in the n-well and extra deep p-well cross section [90].

For the standard modified process the lateral electrical field at the corner of the pixel is so low that some of the charge deposited there is not registered by the electrode after irradiation, as shown in the left image of Figure 3.12. The colour scale represents the electrostatic potential, red are the strongest potential regions, then orange, green, light blues and being the dark blue the weakest potential areas. The arrow lines show the electric field and the star-symbol is the minimum electrical field. The white lines show the edge of the depletion region. What can be extracted from this simulation is that the gap in the n-well and the additional p-implant, middle and right images of Figure 3.12, bend the drift path away from the minimum to the collection electrodes, increasing the lateral electrical field. Thus the drift path is reduced and charges will not end up trapped in the minimum. Finally the charge collection will be faster. Consequently, the gap in the n-implant and the extra deep p-well reduce the charge sharing and increase the single pixel signal, making the sensor more radiation hard.

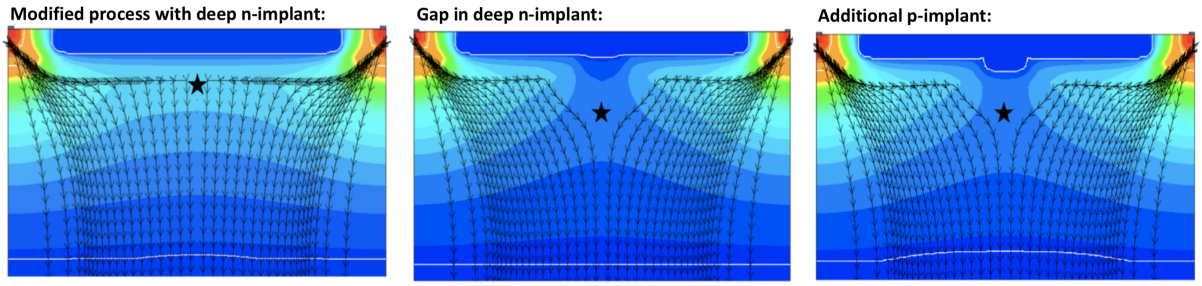


Figure 3.12: Results of electrostatic simulation for different process variations. The black arrows mark the electric field lines, the star symbol marks the electric field minimum and the white lines mark the edges of the depleted regions [86].

The front-end design remains the same as in the MALTA, Figure 3.4, except for the increased size of the transistors M1 and M3. The layout of the full Mini-MALTA chip is shown in Figure 3.13. The readout architecture, which will be discussed in Chapter 4, is done in a similar way to MALTA but with an additional logic at the end-of-column for synchronisation purposes: the synchronisation FIFO and memory. The DACs have been modified to utilise 8 bits and to be more modular and independent of the matrix size (the previous design implemented in MALTA

was a single block occupying the full width of 512 pixels). The slow-control is a different implementation as well due to its unreliable behaviour in MALTA. Finally, the data transmission unit (DTU) used in the ALPIDE chip was added to send the serialised hit data off-chip using a single LVDS output transmitting the data at 1.2 Gb/s with double data rate (DDR). The 600 MHz needed for this transmission is also generated inside the DTU using a phase-locked loop (PLL).

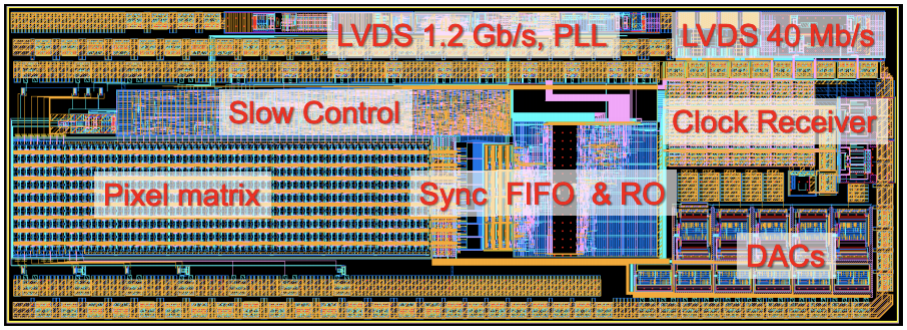


Figure 3.13: Layout of Mini-MALTA chip [87].

3.3 MALTA Czochralski

Some further ideas to tackle the charge loss at the corner of the pixels were developed. Some samples from the original MALTA design were produced on Czochralski silicon wafers and in combination with the new process variations developed for the Mini-MALTA prototype and presented in Section 3.2. The signal generated in a sensor is proportional to the thickness of the sensitive layer. MALTA and Mini-MALTA were fabricated using a p-type substrate with a high resistivity epitaxial layer of 25-30 µm thickness, which constitutes the sensitive layer and can be fully depleted at low voltages of around 6 V. An actual value of the resistivity cannot be quoted due to the complexity of the doping profiles that are kept confidential by the foundry. However, by moving to high resistivity Czochralski material for the fabrication of the MALTA sensor the sensitive volume can be considerably increased for higher reverse biases compensating at least in part the higher noise and higher charge threshold that the MALTA sensor has in comparison to the Mini-MALTA. Czochralski is the standard method to grow single crystalline silicon and it is only recently that it has become available with higher resistivity, 3-4 KΩ per cm. Some of the wafers fabricated with the new starting material were processed with the same variations as discussed above (Figure 3.11) but for all of the wafers the circuitry and consequently the critical transistors were kept the same as in the original MALTA design. Therefore, the RTS noise tail observed in Figure 3.9 will still be a feature in Czochralski devices and there is a limit to the low thresholds that can be reached. Finally, Czochralski single crystal growth technique results

in a lower quality silicon than Float-Zone. It is possible that there will be larger variations in threshold dispersion or noise.

Figure 3.14 shows the cross-sections for MALTA in the standard modified process fabricated in epitaxial and Czochralski respectively. From now on, all devices that have been fabricated with Czochralski starting material will be referred as MALTA CZ.

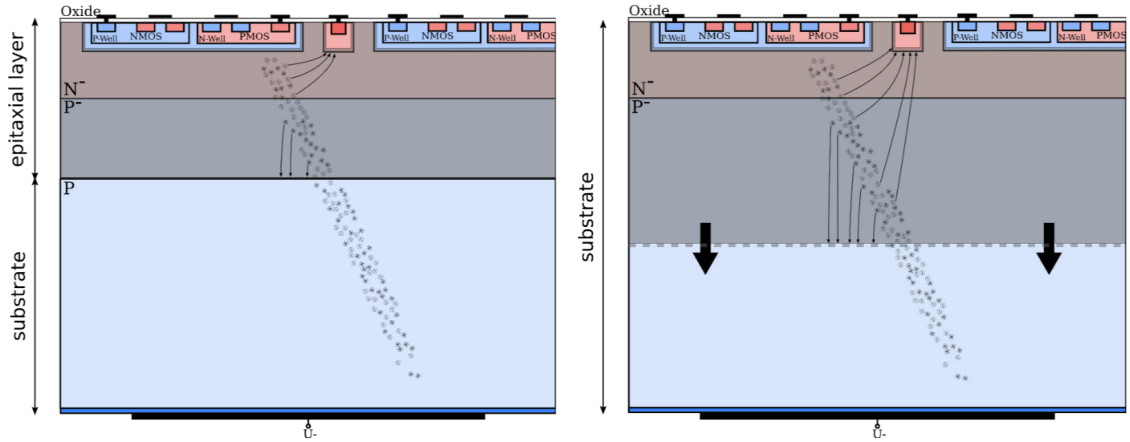


Figure 3.14: Cross section of the original MALTA pixel on epitaxial silicon (left) and the new MALTA pixel on Czochralski silicon (right).

3.4 Characterisation results

In this section the description of the setups that were used to characterise the detectors in both the laboratory at CERN and during testbeam campaigns will be explained.

3.4.1 Mini-MALTA laboratory test description

Mini-MALTA chips arrived in the lab in January 2019. As soon as they were received some of them were sent to Ljubljana for neutron irradiation and others were diced and wire-bonded onto printed circuit boards (pcb) in order to be tested. Basic tests were performed: establishing communication with the chip, exposure to sources to make sure that particles can be detected, noise scans etc. The main goal of this prototype was to prove that the fabrication and front-end changes led to an improvement in sensor performance, particularly after irradiation.

An in-depth analysis of the clusters measured in the different sectors of Mini-MALTA was made to determine the effect of the different design changes on the performance of the sensor. Clustering of hits will improve the spatial resolution and typically reduce the number of noise hits as it is unlikely that they are part of the cluster. If the average cluster size for Mini-MALTA is higher than for MALTA it would be a clear indication that more charge is detected. It is important to know the typical cluster size of the specific detector you are working with to then

make sure that during beam operation the data taken looks as expected. A custom clustering algorithm that computes the number of adjacent pixels that fire within a time window of 50 ns was developed to perform this analysis. Having a small time window guarantees that all the hits in the cluster belong to the same physics event. Because of the asynchronous trigger-less architecture every hit that goes above threshold will be recorded and used in the analysis.

The methodology and most relevant information applied to perform these tests is described here:

- The detector is placed inside a climate chamber operated at a temperature of -20° degrees for all the samples but one unirradiated chip that was put on a table at room temperature. The real sensor temperature was not measured.
- The detector has a protection for the wire bonds, which leaves the matrix exposed to light but the rest is covered in black tape to avoid noise and current due to stray light.
- The radiation sources used are ^{90}Sr and ^{55}Fe due to their different properties.
- The radioactive source is placed manually on top of the matrix.
- Time of exposure of the detector to the radioactive source was not accurately tracked for all the tests and can lead to different count numbers.
- The opening of the source is not collimated.
- Reverse bias is applied and threshold and noise scans are made to determine the operation point of the device.
- Once the source analysis is done, the data collected is passed through the clusteriser algorithm whose results will be presented.

Table 3.3 summarises the conditions for all the chips that were analysed in the laboratory under the conditions described above. The chip name makes reference to the wafer number and the reticle number. As an example W4R6 corresponds to wafer number 4 and 6th reticle. Furthermore, all the results presented have no data for the PMOS reset sectors because they require a different bias to be operated under similar conditions to those that have the diode reset. The p-well voltage is always kept at -2 V for these tests and for all the Mini-MALTA's characterisation results due to a design bug that prevents going to lower values. The threshold is computed for both halves of the matrix as the modifications done to front-end especially the enlargement of some transistors should have an effect on this parameter. The aim is to bias them so they have a similar threshold. One chip where the response of the two halves was studied at two different thresholds so that the effect of operating the sensors at high or low thresholds could be understood. Most noisy pixels were masked so they do not influence the results. Finally, one

of the Mini-MALTA's was analysed for different biases voltages as this parameter can have an effect on the total depleted area and consequently on the charge collection efficiency.

Two radioactive sources were used to perform these measurements ^{90}Sr, and ^{55}Fe, due to their different properties. The ^{55}Fe source emits photons that interact with the silicon via the photoelectric effect. These photons knock electrons out of their shell and those electrons get all the energy from the photons. The photons are fully absorbed. The different electrons have a lot of energy but inside the detector they will collide with other electrons immediately and spread the energy around. This creates a cloud of ionised electrons (and holes) in an almost point-like volume in the silicon so all the charge will be then collected by the sensor from a small volume. This means that if the photon is absorbed in a part of the sensor that is depleted, all the charge will be collected but if the photon is absorbed outside of the depletion region, then the charge will move slowly by diffusion and will be usually lost through recombination or trapping. As a summary it can be said that basically the ^{55}Fe source induces a "binary" signal in the sense that either the sensor gets all of the charge or virtually none of it.

^{90}Sr source decays to Yttrium, which is a beta source that emits electrons with an endpoint energy of 2.2 MeV. The higher energy electrons behave like minimum ionising particles see Section 2.1, similarly to the particles in testbeam. The low energy electrons are absorbed in the material surrounding the sensor. The electrons pass through the sensor depositing energy and creating charge along its path. This results in a signal that is proportional to the sensitive volume, which depends on the depletion voltage.

In conclusion, the ^{55}Fe produces a peak in the energy spectrum and the number of counts in the peak will depend on the size of the depletion region. The ^{90}Sr electrons will produce a signal that is proportional to the size of the depletion region.

Table 3.3: Operating conditions of the sensors used for the Mini-MALTA laboratory cluster analysis. The Mini-MALTA sectors and names are defined in Figure 3.10. Note that when two values of threshold or subtrate voltage are specified is to make comparisons.

CHIP NAME	IRRADIATION DOSE ($MeVn_{eq}/cm^2$)	MEASURED THRESHOLD S0-S3	MEASURED THRESHOLD S4-S7	SUBSTRATE VALUE	SOURCE TYPE ANALYSIS	TEMPERATURE
W4R6	Non irradiated	188 e^-	371 e^-	-6 V	Sr	25C
W1R2	1 x10^{15}	280 e^- | 174 e^-	507 e^- | 310 e^-	-6 V	Sr	- 20C
W5R3	2 x 10^{15}	150 e^-	277 e^-	-6 V	Sr	- 20C
W4R3	2 x 10^{15}	173 e^-	327 e^-	-6 V | -9 V	Sr	-20C
W1R10	Non irradiated	253 e^-	428 e^-	-6 V	Fe	-20C
W1R2	1 x10^{15}	200 e^-	438 e^-	-6 V	Fe	-20C
W1R3	2 x 10^{15}	200 e^-	420 e^-	-50 V	Fe	-20C

3.4.2 Mini-MALTA laboratory cluster analysis

The results presented first discuss the occupancy or total number of hits seen by the different sectors in Mini-MALTA and then the average cluster size for the sectors with the enlarged transistor sizes. This last plot is presented as a histogram with the number of single clusters to several hit clusters.

The first study is a comparison between an unirradiated sensor, W4R6, and one irradiated to a dose of $10^{15}\ MeVn_{eq}/cm^2$, W1R2. The source used for the tests is ^{90}Sr. The thresholds of W1R2 are the higher ones according to the Table 3.3. Note that the temperature column refers to the one of the climate chamber and not to the one of the sensor and that in the two threshold columns S0-S3 and S4-S7 refer to the Mini-MALTA sectors shown in Figure 3.10. The main conclusion of this test is that both before and after irradiation the sectors with the enlarged transistor see a larger count of hits than the smaller size transistors sectors. The reason for this is that both halves of the matrix have a different operational threshold, being lower for the enlarged transistors regions, so more charges are registered. Figure 3.15 shows the measured maps. The average cluster size shown in Figure 3.16 decreases with irradiation from ≈ 1.6 to ≈ 1.1 which indicates that some charge might be still trapped. The unirradiated standard modified sector has more charge sharing than the sectors with modified processes where the lateral electric field has been improved. After irradiation the cluster size becomes larger for sectors S1 and S3, this can be explained because it is known that MALTA loses 30% hit efficiency after irradiation and the lateral electrical field is not high enough to direct them to the corner of the pixels. New cross-sections allow to detect more charge, see Figure 3.16. This also confirms the TCAD simulations presented in Section 3.12 where it is expected that the average cluster size will be smaller for the new modified processes because more charge is collected by the electrode without being trapped.

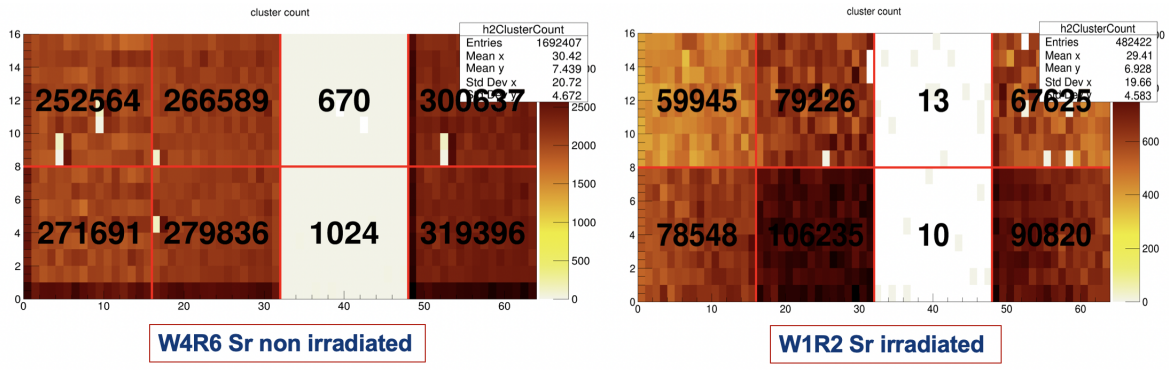


Figure 3.15: Occupancy maps of Mini-MALTA for unirradiated and irradiated to $10^{15}\ MeVn_{eq}/cm^2$ sensors done with ^{90}Sr source.

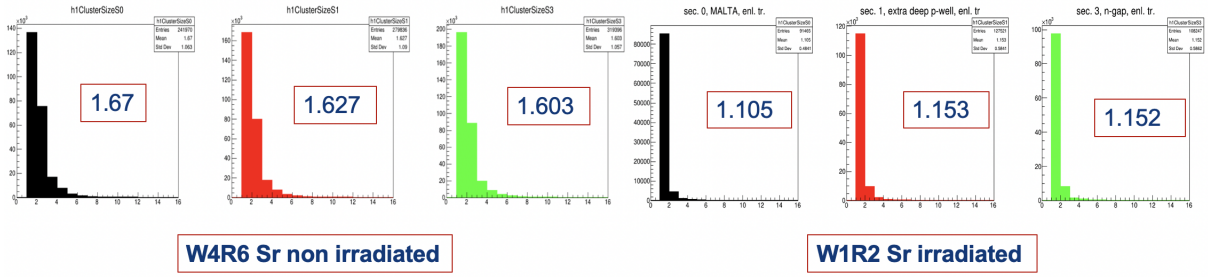


Figure 3.16: Average cluster size histogram of Mini-MALTA for unirradiated and irradiated to 10^{15} $MeVn_{eq}/cm^2$ *sensors done with* ^{90}*Sr source.*

Another comparison was done between an unirradiated Mini-MALTA, W1R10 and W1R2 irradiated to 10^{15} $MeVn_{eq}/cm^2$. The radioactive source used in this case was ^{55}Fe. The same trend as with the ^{90}Sr source is observed. Due to the different operational thresholds of the MALTA like sectors and the enlarged transistors regions, the first ones have a much lower count of hits than the second ones, see Figure 3.17. In terms of average cluster size, the results are interesting as they show what it is expected by the theory, that either all the charge is collected or not seen by the sensor. As it has been explained, ^{55}Fe will deposit the charge in a small area (smaller than the pixel pitch) that can either be in the depleted region or outside it. The average cluster size is then very close to 1 for all the sectors. For the irradiated device the amount of double clusters gets even smaller than for the unirradiated one following the observed trend of smaller cluster sizes after irradiation. The most positive observation is that after irradiation, the average cluster size is still bigger for the new cross-sections meaning that the lateral electrical field is more effective at guiding charge towards the collection electrode leading to reduced charge losses, see Figure 3.18.

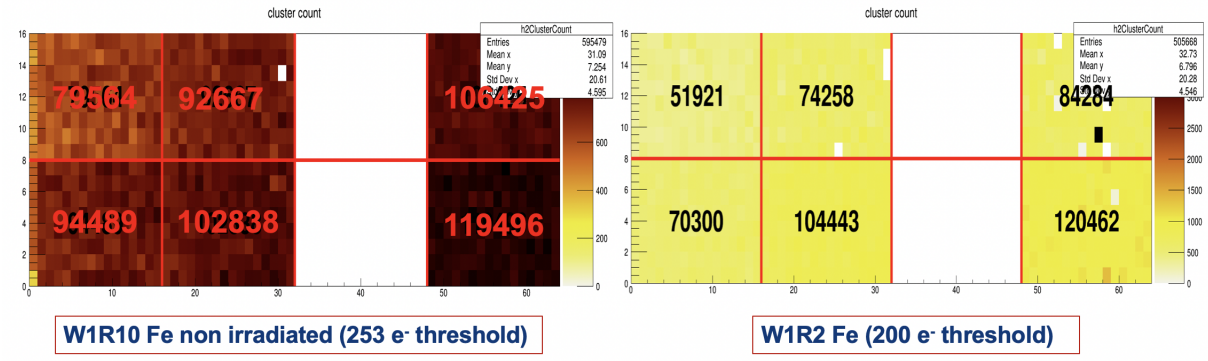


Figure 3.17: Occupancy maps of Mini-MALTA for unirradiated and irradiated to 10^{15} $MeVn_{eq}/cm^2$ *sensors done with* ^{55}*Fe source.*

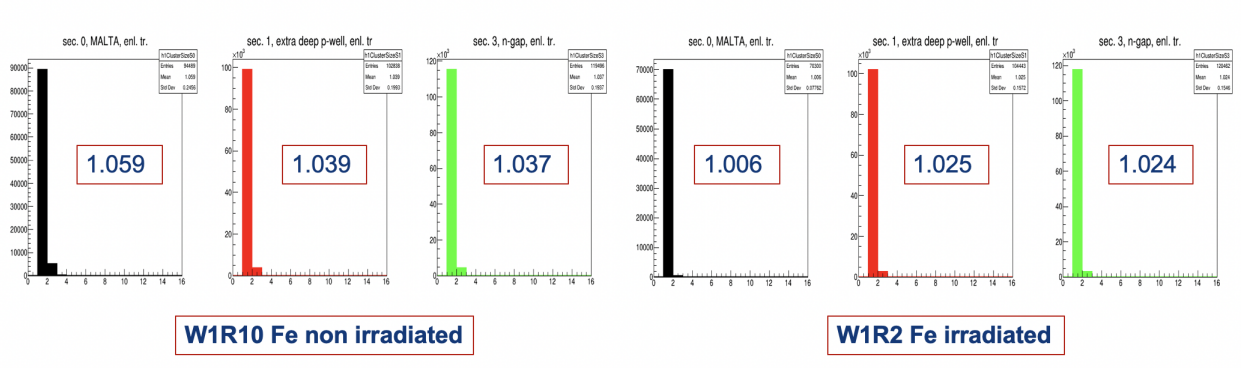


Figure 3.18: Average cluster size histogram of Mini-MALTA for unirradiated and irradiated to 10^{15} $MeVn_{eq}/cm^2$ *sensors done with* ^{55}Fe *source.*

The third comparison is between the behaviour of the same irradiated sensor sample to 1×10^{15} $MeVn_{eq}/cm^2$, W1R2, operated at different thresholds. As expected, for a similar time of exposure to the ^{90}Sr source the device that was configured to run at lower threshold has a higher occupancy than the one set to operate at a high threshold, this is represented in Figure 3.19. With regards to the average cluster size, for a lower threshold it is slightly bigger. This can be explained because the pixels are sensitive to smaller charges and despite having the same number of hits in the corners more pixels will fire due to the low threshold setting, see Figure 3.20.

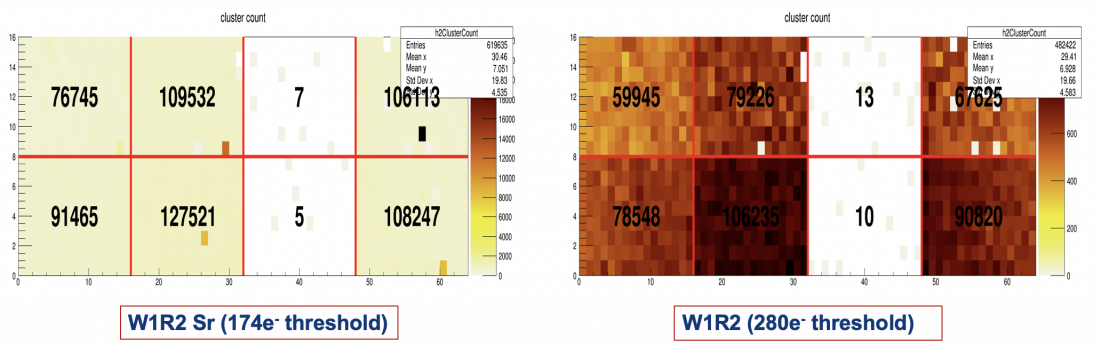


Figure 3.19: Occupancy maps of Mini-MALTA for irradiated to 10^{15} $MeVn_{eq}/cm^2$ *sensor set to operate at different thresholds done with* ^{90}Sr *source.*

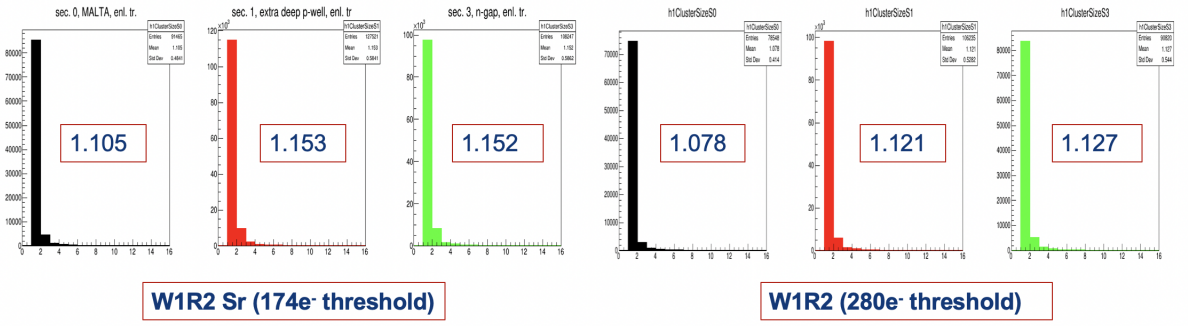


Figure 3.20: Average cluster size histogram of Mini-MALTA for irradiated to 10^{15} $MeVn_{eq}/cm^2$ *sensor set to operate at different thresholds* ^{90}Sr *source.*

The last of this set of comparisons looks at the different response of Mini-MALTA chip irradiated to 2×10^{15} $MeVn_{eq}/cm^2$ for different substrate bias applied. The theory indicates that there should not be a major difference observed in the charge collection because at -6 V the sensor should already be fully depleted. However, the electric field will increase for higher reverse bias and the charge will move faster, reducing the possibilities of the charge to get trapped. The plots observed in Figures 3.21 and 3.22 show that for a similar exposure time a higher absolute value of the substrate voltage makes the MALTA alike sectors see a larger count of hits while the modified sectors number of hits remain very similar. The average cluster size does not change for the different substrate bias as predicted. One last thing can be extracted from Figure 3.22 , if it is compared to the histogram shown in Figure 3.20 for low threshold operation it is clear that higher irradiation levels are linked to further charge losses as the average cluster size decreases. These initial lab studies will be complemented with data from testbeam campaigns in which the performance will be compared against the different monolithic active sensor developments that have been presented in this chapter.

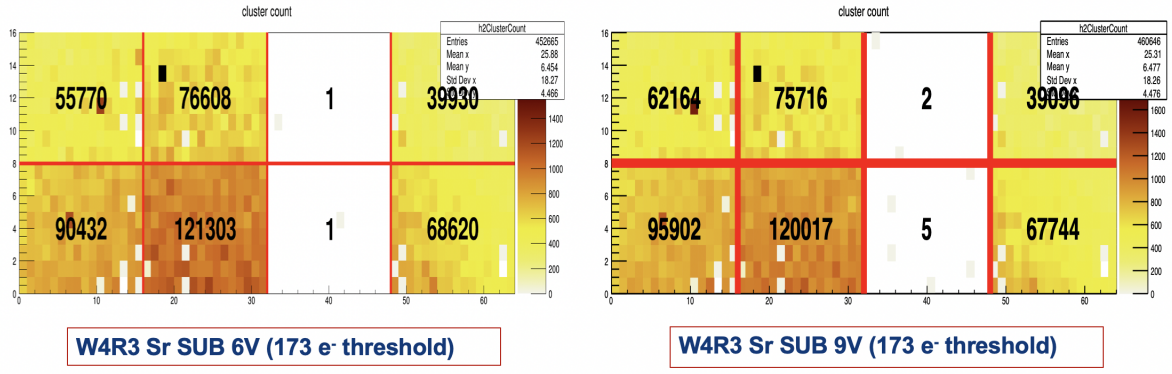


Figure 3.21: Occupancy maps of Mini-MALTA for irradiated to 2×10^{15} $MeVn_{eq}/cm^2$ *sensor, set to operate at a "high" and "low" substrate voltage done with* ^{90}Sr *source.*

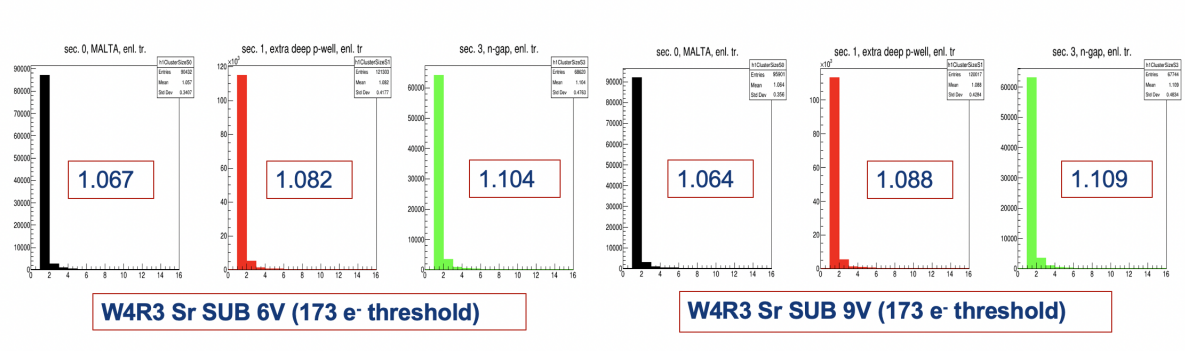


Figure 3.22: Average cluster size histogram of Mini-MALTA for unirradiated and irradiated to 2×10^{15} $MeV n_{eq}/cm^2$ *sensors done with* ^{55}Fe *source.*

3.4.3 Testbeam facilities

Detection efficiency is a parameter that is measured mainly during testbeam campaigns. All the data that will be shown from the original MALTA design was taken in CERN SPS facilities with a 120 GeV pion beam [100] by the ATLAS CMOS collaboration. The data acquisition for Mini-MALTA and MALTA CZ, was carried out either in the beam facilities in DESY accelerator in Hamburg, or in the ELSA accelerator operated by the University of Bonn during the timeline of this thesis. For both facilities they use an electron beam with an energy that can be set up to 6 GeV in DESY and of 2.5 GeV in ELSA. This difference in the energy of the beam will influence the way particles interact with the detectors.

3.4.4 Testbeam data acquisition setup

The device under test (DUT) that is going to be measured, is situated in the middle of a beam telescope consisting of a number of detector planes for defining tracks, which are located on either side of the DUT. For all the SPS measurements, the telescope used was made up with MIMOSA 26 sensors and had six reference planes [102]. A telescope with MALTA detectors was built for more recent measurements of the MALTA Cz and Mini-MALTA. Figures 3.23 and 3.24, show the layout of the telescope and how it looked in reality during the 2019 testbeam campaign in DESY. It has 7 tracking planes made of thin MALTA sensor of 100 µm thickness in order to minimise multiple scattering. Each plane is readout with a Xilinx Virtex-7 FPGA.

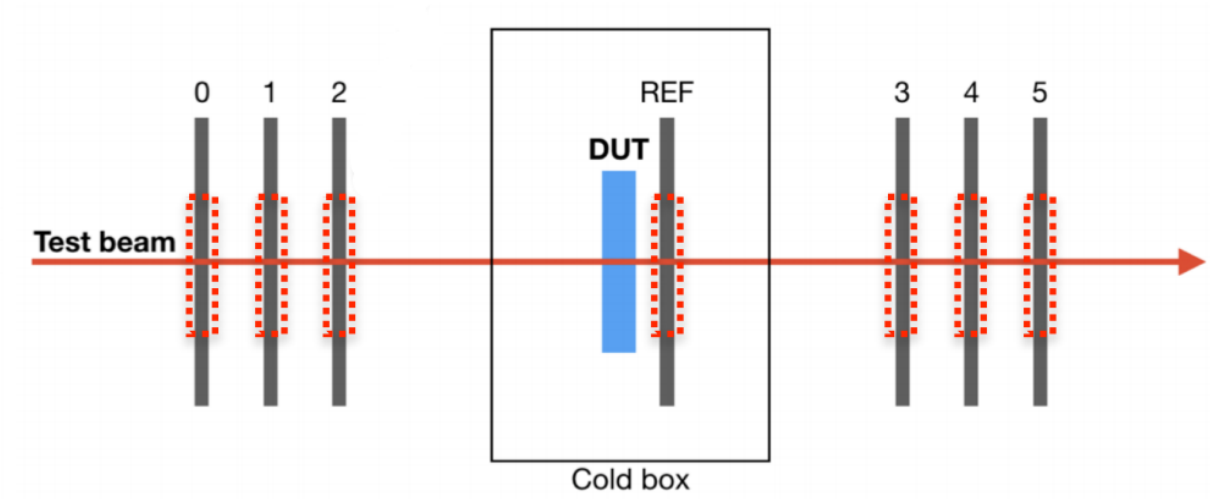


Figure 3.23: Layout of MALTA telescope used during beam campaign in DESY 2019.

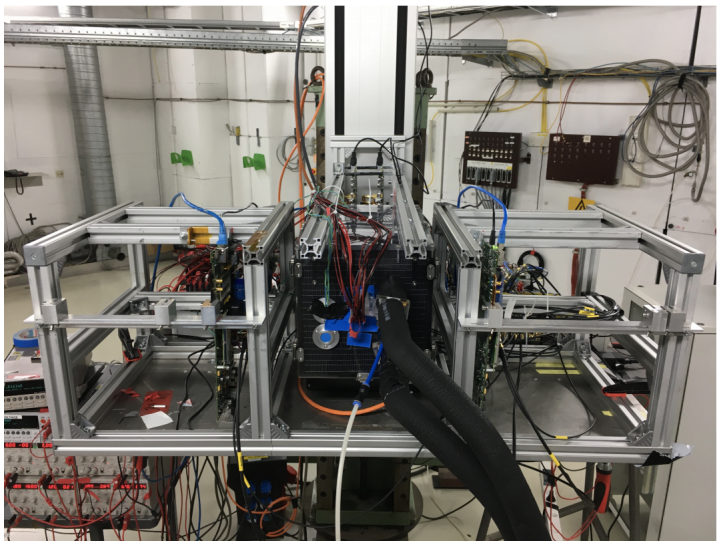


Figure 3.24: Picture taken during the 2019 DESY beam campaign of the MALTA based telescope.

In the case of the MALTA telescope, if a particle crosses both the reference plane and the DUT in the same location within a few micrometers and in a time-window of 75 ns it will be used to perform the tracking. If the hit is not observed in both the reference plane and the DUT the event is discarded. The rest of the 6 planes are used to provide the tracking resolution.

The main algorithm used for reconstruction is General Broken Lines or "GBL" [103], as it accounts for the multiple scattering of the electron beams. The detection efficiency is calculated as the ratio of the number of events with telescope tracks and a corresponding hit on the DUT over the total number of events with a telescope track within the given time acquisition window. The clusters are identified when adjacent pixels in both the reference plane and DUT get hit within the time window. The telescope and the DUT are aligned in such a way that the beam

intensity is the highest over the region of interest of the DUT. If needed the rest of the pixel matrix can be masked. Since the telescope provides such a precise track resolution in relation to the pixel size of the DUT, the detection efficiency can be analysed with sub-pixel precision to check whether a uniform high efficiency can be achieved over the full area of a pixel. This has been key for all the developments presented in this chapter.

3.4.5 Testbeam efficiency analysis

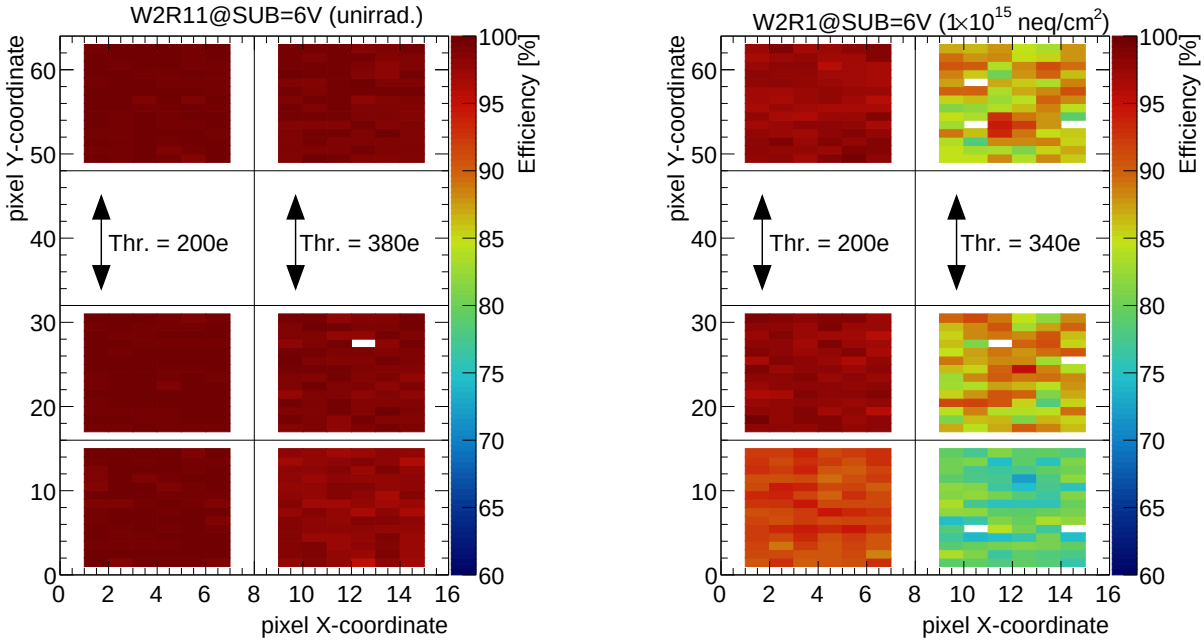
First results to be reviewed are if the positive indications with respect to charge collection that were found during the laboratory cluster analysis tests could be confirmed with the efficiency plots obtained during testbeam campaigns.

In Section 3.1.3 it has been explained that the difficulty of operating the detector at low thresholds due to RTS noise plus a sensor not fully optimised meant a drop in efficiency at the corners of the pixel after irradiation at a fluence of $10^{15}\ MeV n_{eq}/cm^2$. In this case the measured chip efficiency of MALTA was $\approx 70\%$.

The first measurement to do is to measure the efficiency of all all the regions in Mini-MALTA for an unirradiated device and verify that they meet the required 98% detection efficiency. Figure 3.25a shows a 2D efficiency map for an unirradiated Mini-MALTA sample measured with a 2.5 GeV electron beam at ELSA in 2019. Different sensor regions are visible: standard MALTA-like (bottom part of each chip), modified with extra deep p-well (middle part) and modified with extra n-gap (top part). Results are also shown for sensor regions with standard (right side of each chip) and enlarged (left side) transistors. The chips were operated at 6 V SUB voltage, -2 V P-WELL voltage and -20° C, and were tuned for low threshold operation. Note the difference in threshold for the two halves of the matrix after the transistor size modifications. The expectations were met as the efficiency is above 98% for all the regions.

The second image, see Figure 3.25b, is the same 2D efficiency map taken during the same beam campaign but this time for an irradiated sensor to $1 \times 10^{15}\ MeV n_{eq}/cm^2$. The operating conditions were 6 V SUB voltage, -2 V P-WELL and -20° C, and were tuned for low threshold operation. In this case, the MALTA sector should see an efficiency of around 70-80% while for the other sectors, the charge collection should have increased. The efficiency should be higher for the enlarged transistors regions than for the ones with the small size devices as the operating thresholds as the use of larger transistors reduces the RTS noise. These results confirmed the measurements taken in SPS for MALTA. A hit efficiency of 75% was observed for the regions of the irradiated sensors with standard transistors, as was observed in previous measurements at the SPS. The regions with an extra deep p-well and the gap in the n-layer see a small improvement, having an efficiency of 85%. The efficiency is not as high as required because the operational threshold remains too high due to noise. The left half of the matrix with the larger transistors show that even for the MALTA sector the capability of reaching a low threshold brings a 10% enhancement in the charge collection. The other two sectors that combine the sensor optimisation

plus the RTS noise reduction are fully efficient after irradiation.

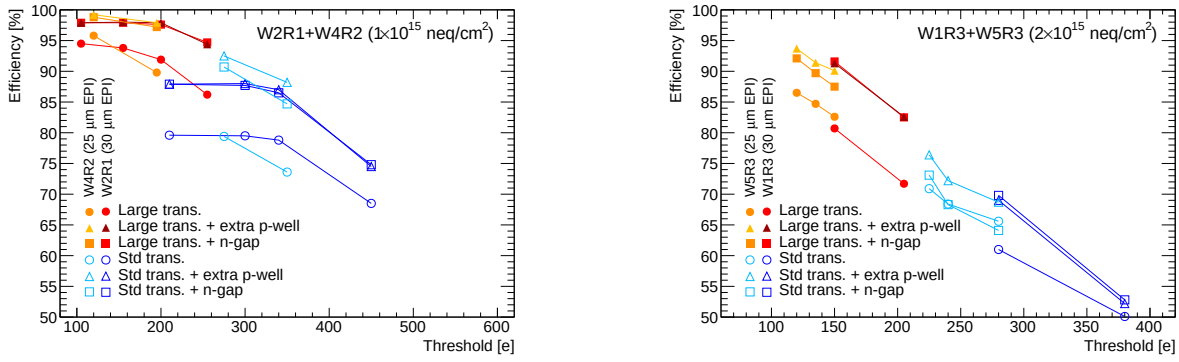


(a) *(b)*

Figure 3.25: (a) Unirradiated Mini-MALTA efficiency map. (b) Irradiated to 1×10^{15} $MeVn_{eq}/cm^2$ *Mini-MALTA efficiency map [87].*

To sum up the Mini-MALTA efficiency measurements Figure 3.26 shows the efficiency versus threshold mean for two neutron irradiated Mini-MALTA samples at 1×10^{15} $MeVn_{eq}/cm^2$ and 2×10^{15} $MeVn_{eq}/cm^2$ respectively measured with a 2.5 GeV electron beam at ELSA in 2019. The chips were operated at -6 V SUB voltage, -2 V P-WELL voltage and -20° C. Different sensor regions are presented: standard MALTA-like (circles), modified with extra deep p-well (triangles) and modified with extra n-gap (rectangles). Results are also shown for sensor regions with standard (open markers) and enlarged (full markers) transistors, as well as for sensors with different epitaxial layer thicknesses: 25 µm (orange or light blue symbols) and 30 µm (red or dark blue symbols). The conclusions from these graphs are that the data taken are consistent for all the samples as the results are very close for the different samples operated under the same conditions. In addition, enlarged transistors regions have a higher efficiency than the standard transistor size ones and it decreases for higher thresholds. The sensor optimisations solve the charge loss at the corner of the pixels for the radiation levels of the outermost layer of ATLAS. Both extra deep p-well and n-gap efficiency numbers are almost identical. If the device is irradiated to 2×10^{15} $MeVn_{eq}/cm^2$ similar trends are observed. The main differences are that the decline in the efficiency with the threshold happens faster and that a slight charge loss is observed even for the modified regions. The maximum efficiency drops to $\approx 95\%$. One remark to these plots is that the efficiency at the borders is a bit lower than in the centre. The reason for this is that when a particle hits at the borders of the DUT it can be influenced by the scattering

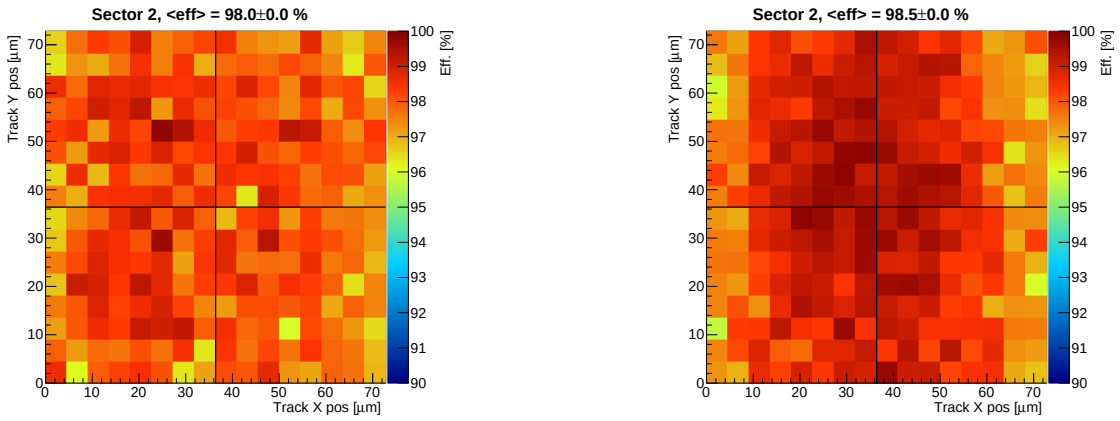
or noise and fall outside the boundaries of the reference plane.



(a) *(b)*

Figure 3.26: (a) Efficiency vs threshold for all regions of Mini-MALTA irradiated sensor at 1×10^{15} $MeV n_{eq}/cm^2$. *(b) Efficiency vs threshold for all regions of Mini-MALTA irradiated sensor at* 2×10^{15} $MeV n_{eq}/cm^2$. *Graphics from [87].*

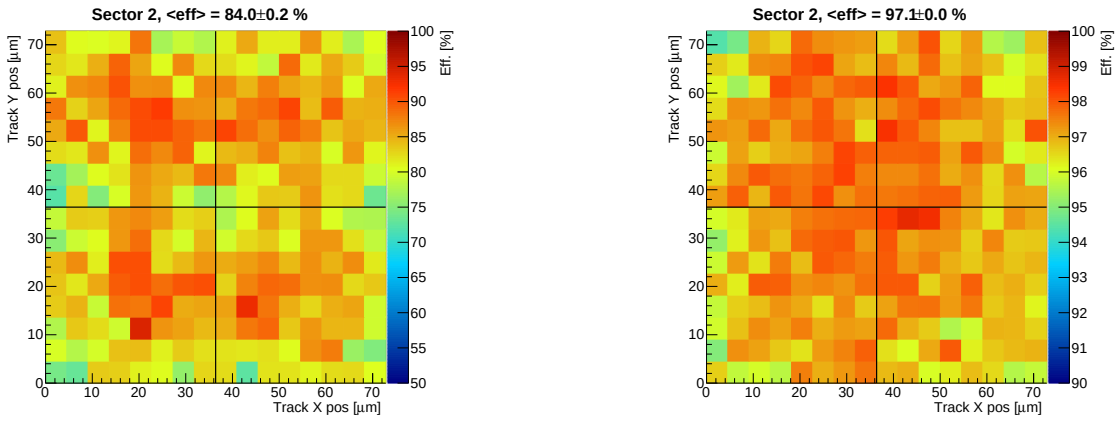
Finally the efficiency for the new MALTA CZ samples will be analysed. The increase of the sensitive volume for these detectors means that more charge will be deposited and the generated signal will be larger. Consequently, the hit efficiency will improve for a fixed threshold raising the efficiency in the corners. The only difference between the original MALTA and the MALTA CZ is the substrate type, one has epitaxial and the other high resistivity Czochralski, but the electronics design remains completely unchanged. Several wafers with all the different process modifications were fabricated. To have a direct comparison Figure 3.27a shows the in-pixel 2D efficiency map for Sector 2 of an unirradiated MALTA epitaxial sample W7R4 (standard modified process) projected over a 2 x 2 mini-matrix measured with a 3 GeV electron beam at DESY in October 2019. The binning corresponds to 8 x 8 entries per single pixel. The sample was operated at -20° C in a climate chamber with the bias voltages set to P-WELL -6 V and SUB -10 V. Effective threshold was measured to be 485 electrons with a noise of 11 electrons. On the other hand, in Figure 3.27b can be seen the same plot measured in the same facility and testbeam for an unirradiated sample of Czochralski MALTA, W7R12 (Czochralski standard modified process). The bias settings in this case were P-WELL -6 V and SUB -30 V. The idea was to have a similar effective threshold and it was measured to be 427 electrons with a noise of 9 electrons. The efficiency for both cases is almost 100% but it is already noticeable the higher efficiency for the Czochralski sample.



(a) *(b)*

Figure 3.27: (a) 2D efficiency map for Sector 2 of unirradiated MALTA epitaxial sample W7R4 projected over a 2×2 *mini-matrix (b) 2D efficiency map for Sector 2 of unirradiated MALTA CZ W7R12 projected over a* 2×2 *mini-matrix. Note: error is 0% because it is smaller than the one decimal digit.*

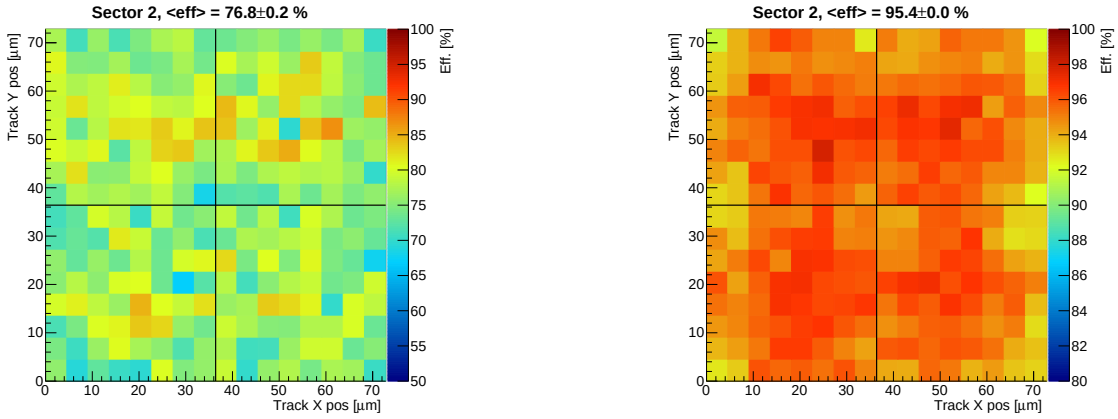
The next question to consider is if the Czochralski efficiency after irradiation is as good as the results obtained for Mini-MALTA after enlarging the transistors and with the new process modifications. During the experimental campaign there were no MALTA detectors with the standard modified process that had been irradiated so as to be able to compare the data, Figure 3.28 compares the efficiency between MALTA detectors with the n-gap fabricated in epitaxial and high resistivity Czochralski as starting material, both irradiated to $1 \times 10^{15}\ MeV n_{eq}/cm^2$. The operating conditions for both chips were -20° C with P-WELL -6 V, and SUB -10 V and -20° C with P-WELL -6 V and SUB -50 V respectively, trying to set them at a low threshold. The specific values of the threshold for these chips had not yet been measured due to lack of time before the testbeam and the MALTA slow control issue. Despite the similar colours in the plots, the scales are very different. The results shown in Figure 3.28 show that the epitaxial MALTA has an average efficiency across the 2x2 mini-matrix projection of sector 2 of 84% that is similar to the one observed in the Mini-MALTA regions with n-gap and without enlarged transistors. The corners also show lower efficiency compared to the the centre, as was observed previously. In contrast, the efficiency of the MALTA CZ sensor is 97% proving that increasing the sensitive depth is a solution to the loss of charge in the corners of the pixels. There is still a drop in efficiency in the boundaries of the pixels but it reduces to 95% rather than 75% that was observed for the epitaxial designs.



(a) *(b)*

Figure 3.28: (a) 2D efficiency map for Sector 2 of an epitaxial MALTA with n-gap (sample W4R1) irradiated to to $1 \times 10^{15}\ MeV n_{eq}/cm^2$ *projected over a* 2×2 *mini-matrix. (b) 2D efficiency map for Sector 2 of a MALTA CZ with n-gap (sample W7R12) irradiated to* $1 \times 10^{15}\ MeV n_{eq}/cm^2$ *projected over a* 2×2 *mini-matrix. Note: error is 0% because it is smaller than the one decimal digit.*

The last comparison is between a MALTA epitaxial n-gap and a MALTA CZ n-gap irradiated to $2 \times 10^{15}\ MeV n_{eq}/cm^2$. The sensors were operated under the following conditions: at -20° C with P-WELL -6 V and SUB -10 V and -20° C P-WELL -6 V, and SUB -55 V respectively. The efficiency of the CZ sample is 95%, which is only a small decrease compared to efficiency of the CZ sample irradiated to $1 \times 10^{15}\ MeV n_{eq}/cm^2$. There is only a small drop in efficiency in the corners. In the case of the epitaxial sensor the average efficiency drops from 84% to 76%.



(a) *(b)*

Figure 3.29: (a) 2D efficiency map for Sector 2 of an epitaxial MALTA with n-gap (sample W4R4) irradiated to $2 \times 10^{15}\ MeV n_{eq}/cm^2$ *projected over a* 2×2 *mini-matrix. (b) 2D efficiency map for Sector 2 of a MALTA CZ with n-gap (sample W9R4) irradiated to* $2 \times 10^{15}\ MeV n_{eq}/cm^2$ *projected over a* 2×2 *mini-matrix.*

3.4.6 Testbeam efficiency vs substrate bias analysis

It is important to understand the optimal reverse bias settings for both sensor types. For the epitaxial samples the sensor is already fully depleted for a SUB setting of -6 V. A further increase of this value results in a drop of efficiency because the electrical field profile changes and the charges deposited in the corners are less likely to be detected. The opposite behaviour is observed for the CZ samples. Their efficiency increases with increasing reverse bias as it directly influences the volume of the depleted region. The efficiency as a function of bias voltage is shown in Figure 3.30. Measurements were done with a 4 GeV electron beam at the DESY facility in 2019 on the second and third sector of the MALTA matrix. The deep p-well was biased at -6 V and the samples were kept at -20° C. The measured thresholds are specified in the legend of the graph. As it has been explained for the epitaxial samples a substrate bias higher than 10V decreases the performance of the sensor while for the CZ samples, the higher the reverse bias the higher the efficiency. This trend is applicable for both fluences.

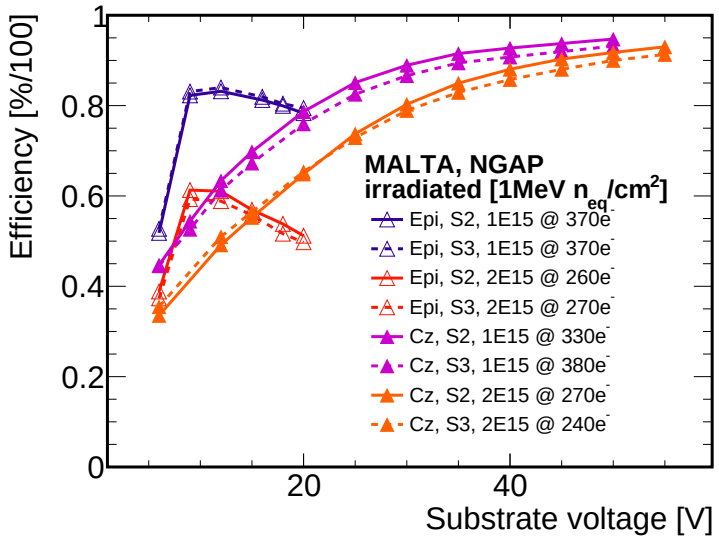


Figure 3.30: Efficiency versus substrate voltage for irradiated to $1 \times 10^{15}\ MeV n_{eq}/cm^2$ *and* $2 \times 10^{15}\ MeV n_{eq}/cm^2$ *MALTA samples fabricated in epitaxial and Czochralski starting materials.*

3.4.7 Testbeam efficiency vs threshold analysis

A key for the success of DMAPS to be more radiation tolerant is to be able to operate at low thresholds. Figure 3.31 shows the efficiency versus the threshold for the same irradiated MALTA epitaxial and CZ n-gap samples that were presented in the previous section. It is evident that for all devices a lower threshold results in a higher detection efficiency.

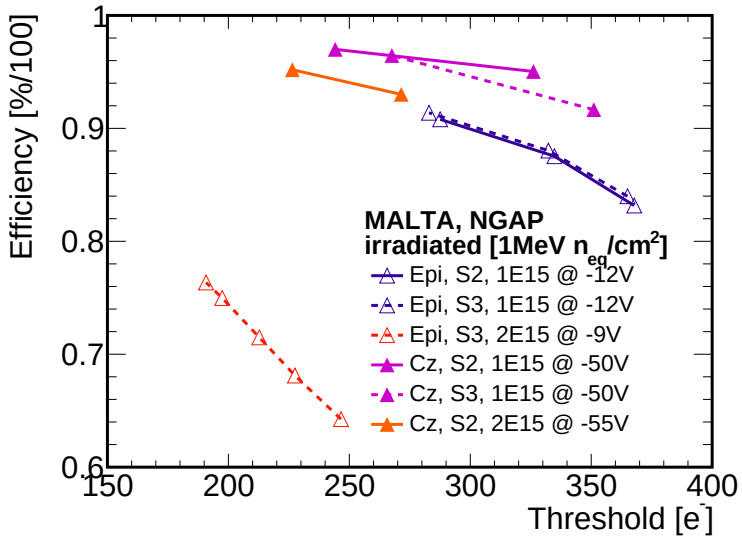
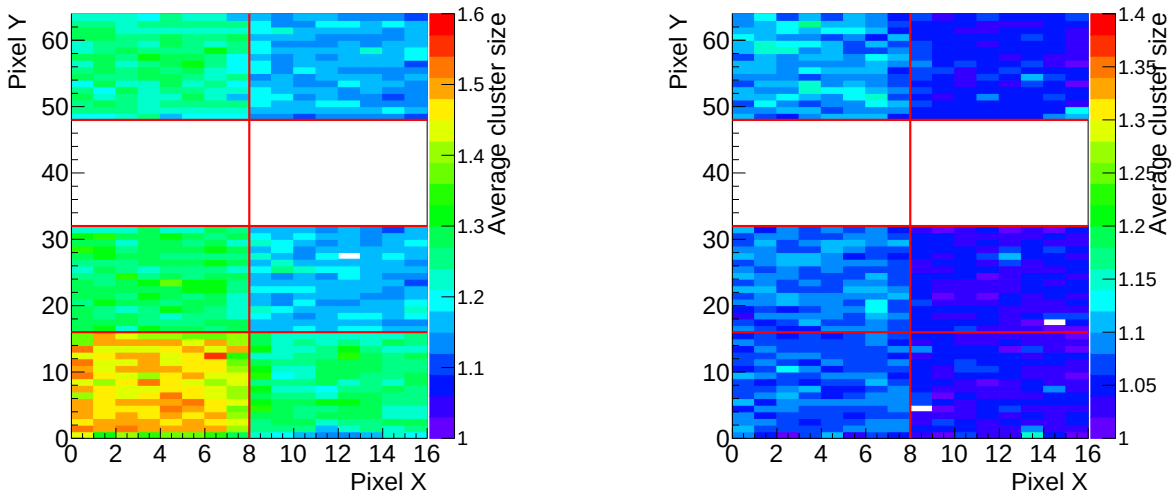


Figure 3.31: Efficiency versus threshold for epitaxial and CZ MALTA sensors irradiated to $1 \times 10^{15}\ MeV n_{eq}/cm^2$.

3.4.8 Testbeam cluster size analysis

Testbeam studies of mini-MALTA sensors were made to determine the cluster size. The cluster size in the testbeam was compared to that measured in the lab. Figure 3.32 shows the average cluster size per pixel for two Mini-MALTA samples one unirradiated and another irradiated to a fluence of $1 \times 10^{15}\ MeV n_{eq}/cm^2$. The data for these plots was taken with a 2.5 GeV electron beam at ELSA in 2019. The operating conditions of the samples were bias settings of P-WELL -2 V and SUB -6 V and at a temperature of -20° C. The threshold values were optimised to be comparable and the blank regions correspond to the PMOS reset sector that was not studied due to different biasing. Results show that the sectors with enlarged transistors have larger cluster sizes. Furthermore, the n-gap and extra deep p-well process modifications regions have a smaller cluster size in average than the standard modified regions due to their increased lateral electrical field. These results are in agreement with the Mini-MALTA laboratory measurements previously presented.



(a) *(b)*

Figure 3.32: (a) Average cluster size per pixel for an unirradiated Mini-MALTA sensor. (b) Average cluster size per pixel for a Mini-MALTA sensor irradiated to $1 \times 10^{15}\ MeV n_{eq}/cm^2$.

Similarly to what has been shown for the efficiency results, the response in terms of cluster analysis for the MALTA fabricated in epitaxial material and the MALTA CZ has been studied. Figure 3.33 is a graph of the average cluster size of unirradiated MALTA samples on 25 µm epitaxial silicon and CZ silicon respectively with no process modifications (standard modified process) versus substrate bias. Measurements were done with a 4 GeV electron beam at the DESY facility in 2019 on the second and third sector of the MALTA matrix. The deep p-well was biased at -6 V and the samples were kept at room temperature. The most noticeable outcome of this plot is that CZ samples have a much larger average cluster size than the epitaxial devices, for example at - 6 V of substrate bias the CZ samples pixels see $\approx$ 1.5 hits per cluster while the epitaxial pixels get $\approx$ 1.2 hits per cluster in average and that this parameter increases with higher substrate voltages. MALTA epitaxial average cluster size is constant, independent of the substrate voltage, which is due to the sensor being fully depleted at -6V. However, the MALTA CZ samples see an increase on the average cluster size to $\approx$ 2.2 hits per cluster when biased at -30 V due to the increase of the depleted volume.

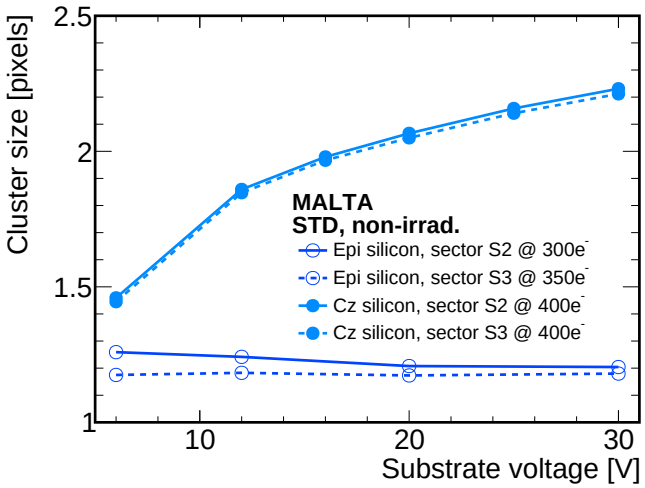


Figure 3.33: Average cluster size for unirradiated MALTA with standard modified process fabricated in epitaxial and Czochralski for different substrate bias and thresholds.

Figure 3.34, shows an equivalent graph but in this case for unirradiated MALTA devices fabricated in 30 µm of epitaxial and CZ with the gap in the n-layer process. P-well bias was set to -6 V and they were measured at room temperature. The behaviour has changed with respect to what was observed in the previous plot. The MALTA CZ samples have still a larger average cluster size but the difference with respect to the epitaxial sensor is not as large, it goes from 1.2 hits per cluster to 1.3 hits per cluster on average and the number is smaller than for the standard modified process. The change with respect to the substrate bias is much smaller as well, even though, the scale range in the substrate voltage is reduced to half. The reason for this is that the sensor optimisation modifies the electric field to increase the charge collection at the electrode.

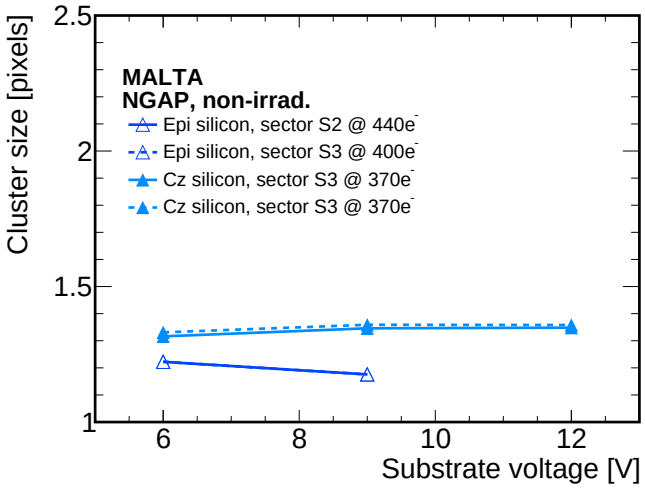


Figure 3.34: Average cluster size per pixel for unirradiated MALTA with gap in the n-well process fabricated in epitaxial and Czochralski for different substrate bias and thresholds.

There was no MALTA epitaxial sensor with the standard modified process irradiated so only devices with the gap in the n-well were analysed. Measurements were done with a 4 GeV

electron beam at the DESY facility in 2019 on the second and third sector of the MALTA matrix see Figure 3.1. The deep p-well was biased to -6 V and the samples were kept at -20° C. Figure 3.35 shows that the average cluster size after irradiation decreases considerably independently of the substrate material. The higher the reverse bias applied for CZ samples the larger the average cluster size as the depleted volume increases while for the epitaxial this parameter stays constant as the sensor is already depleted at -6 V. The differences in threshold are not big enough to observe any difference and both S2 and S3 sector behave very similarly.

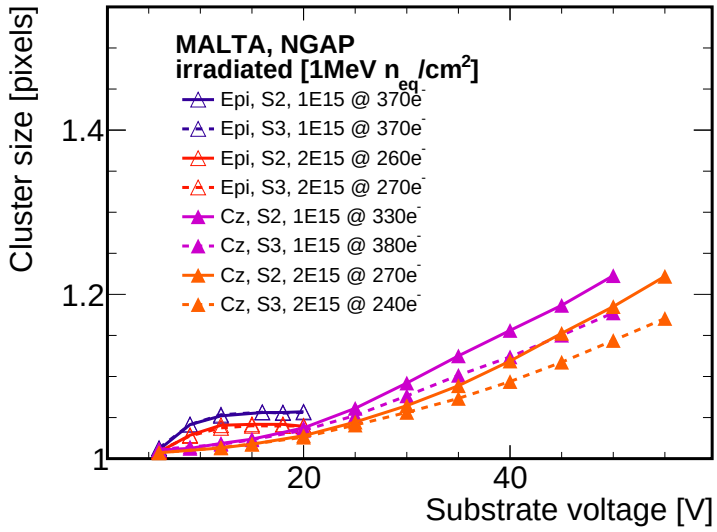


Figure 3.35: Average cluster size per pixel for irradiated MALTA samples with gap in the n-well process fabricated in epitaxial and Czochralski for different substrate bias and thresholds.

To complete this section, the influence of the threshold and bias voltage on the average cluster size for the same irradiated MALTA n-gap samples with both CZ and epitaxial material is studied. The facility, temperature of operation and p-well bias are the ones previously specified in Section 3.4.6. The general trend observed in Figure 3.36 is that the lower the threshold the larger the average cluster size. As more hits are registered this effect is much more noticeable for the Czochralski samples.

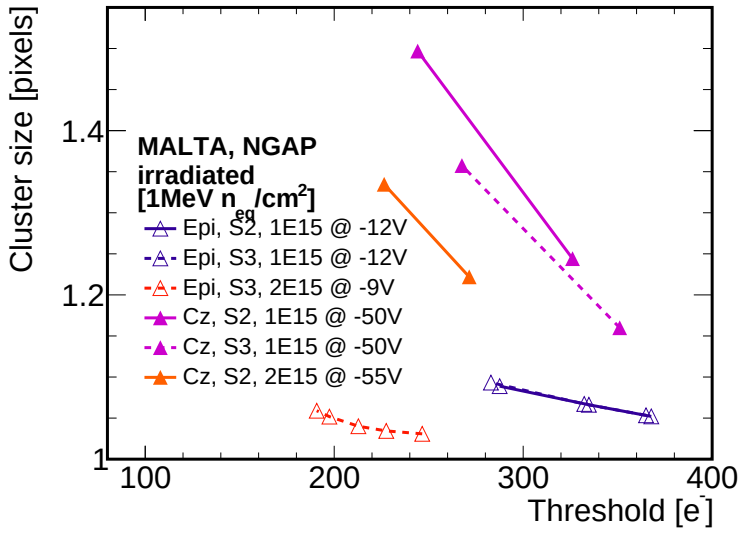


Figure 3.36: Average cluster size per pixel for irradiated MALTA samples with gap in the n-well process fabricated in epitaxial and Czochralski for different threshold settings.

3.5 Conclusions

This chapter has described the front-end schematics and the sensor designs for the MALTA and Mini-MALTA DMAPS. It has also results on cluster size and efficiency plotted against different operational parameters. All the improvements done to increase the radiation tolerance and to overcome the initial charge collection inefficiency after irradiation in the corner of the pixels for these detectors have been proven in silicon and understood. Both of the new cross-sections: the gap in the n-blanket layer and the extra deep p-well, have a very similar behaviour in terms of cluster size and efficiency after irradiation and meet the ATLAS Inner Tracker Upgrade requirements for the outer barrel. MALTA detectors fabricated with Czochralski material can be operated at higher reverse bias, increasing the depletion region to around 100 µm and generating more signal. The cluster size increases considerably with respect to the original MALTA detector and they have a good efficiency up to 98% when irradiated at a fluence of $10^{15}\ MeV n_{eq}/cm^2$.

Chapter 4

Digital Readout Architecture

So far this thesis has covered the different sensor designs, front-end electronics and characterisation results for the small collection electrode DMAPS devices of the MALTA family. This Chapter will talk about the digital readout architectures present in the MALTA and Mini-MALTA emphasising on how to synchronise in the periphery and asynchronous matrix.

4.1 Main concepts on digital readout architectures for HEP

As mentioned in Section 2.3.2 for HEP architectures, the main parameters that need to be provided for the physicists to analyse the data are hit position, event number and pulse amplitude. The individual implementation of the architecture then depends on the chip technology, in the case of this thesis, TowerJazz 180 nm, on the acceptable data losses and on the precision of the stored data. In the case of ASIC designs for the High-Luminosity LHC, different architectures usually process only the hits that are above a specified threshold. This is to reduce the amount of memory required to store all the data as the noise hits due to to electronic noise are not recorded and therefore to optimise the use of bandwidth. This type of readout approach where not all the matrix is processed is known as zero suppression [105]. Experiments that have lower input hit-rates can afford to readout every single hit seen by the matrix, this type of architecture is known as trigger-less. Rolling shutter architecture schemes in which different parts (usually rows or columns) of the pixel matrix are read at different times have been widely used in HEP. However when the data rates are too high some data reduction is needed so the Data Acquisition System (DAQ) can cope. This is known as a triggered architecture. In this case a triggering signal based on system level measurements selects the events of interest [106]. The analysis of the data has to be done within a fixed latency after the particle has been detected for the trigger to be correctly pointing out to the right event. In the case of ATLAS Upgrade the trigger latency is in the order of 500 bunch-crossings or 12.5 μs and the trigger rate has been set to 1 MHz maximum. Consequently, for triggered architectures hits need to be stored for the total trigger latency. The storage needed for this purpose in order to fulfil the data losses requirements is determined with

simulations like the one that will be described in the coming sections. In the case of MALTA, the architecture is trigger-less and its compatibility with the triggered system is done in the DAQ side. The bunch crossing identifier (BCID) of the hits is stored so the triggered hits can then be filtered.

4.2 MALTA architecture

The MALTA chip has a novel asynchronous architecture, which means that there is no clock distribution over the pixel matrix. The main reason to choose such an architecture is to reduce the digital power consumption and to be able to handle high data rates with good timing precision. For the readout, the 512×512 pixels are organised in double columns. Every double column is subdivided in groups of 16 pixels with 2 columns and 8 rows each having 64 of these 16 pixel groups in total per double column. These groups are further distinguished by two alternate colours red and blue depending on the output bus that they are connected to. This is depicted in Figure 4.1.

When a pixel within a group has its discriminator fired, it generates a reference or hit signal in the corresponding line of the pixel bus and the 5 bit group address corresponding to the hit pixel is propagated as well. A pulse with a programmable length of 0.5, 0.75, 1 or 2 ns is transmitted in parallel on every line that needs to transmit a logic one. Due to the different positions of the pixels with respect to the end of column, the propagation time of the signals in the bus will be different. This issue was addressed and the delay in the transmission is matched for all the different lines in each bus. The total number of lines for each bus is 22, 16 for each pixel in the group, 5 for the address of the 32 groups of the same colour in the double column and 1 for the reference programmable length pulse. Assuming that the charge is shared between two or more pixels in the same group, a hit arbitration is performed within the group to ensure that the hit data of all fired pixels is transmitted correctly. If the discriminators of two pixels fire within a 1.6 ns time window, two corresponding lines are activated simultaneously and will be readout with the same reference pulse. If the pixels receive a different enough amount of charge and the discriminators react successively, two words are transmitted sequentially over the bus, making sure that there is sufficient separation of the signals on the bus for proper transmission. Alternating the groups of 16 pixels in the two different buses is done in order to avoid collisions in case a particle hits in the boundary of two groups. Data is transmitted instantaneously after the hit is detected so the latest it will be present at the periphery is within a few nanoseconds of the hit arrival. Every double column has its own readout so they can all be transmitting simultaneously. This parallelism is done to be able to cope with high-data rate applications.

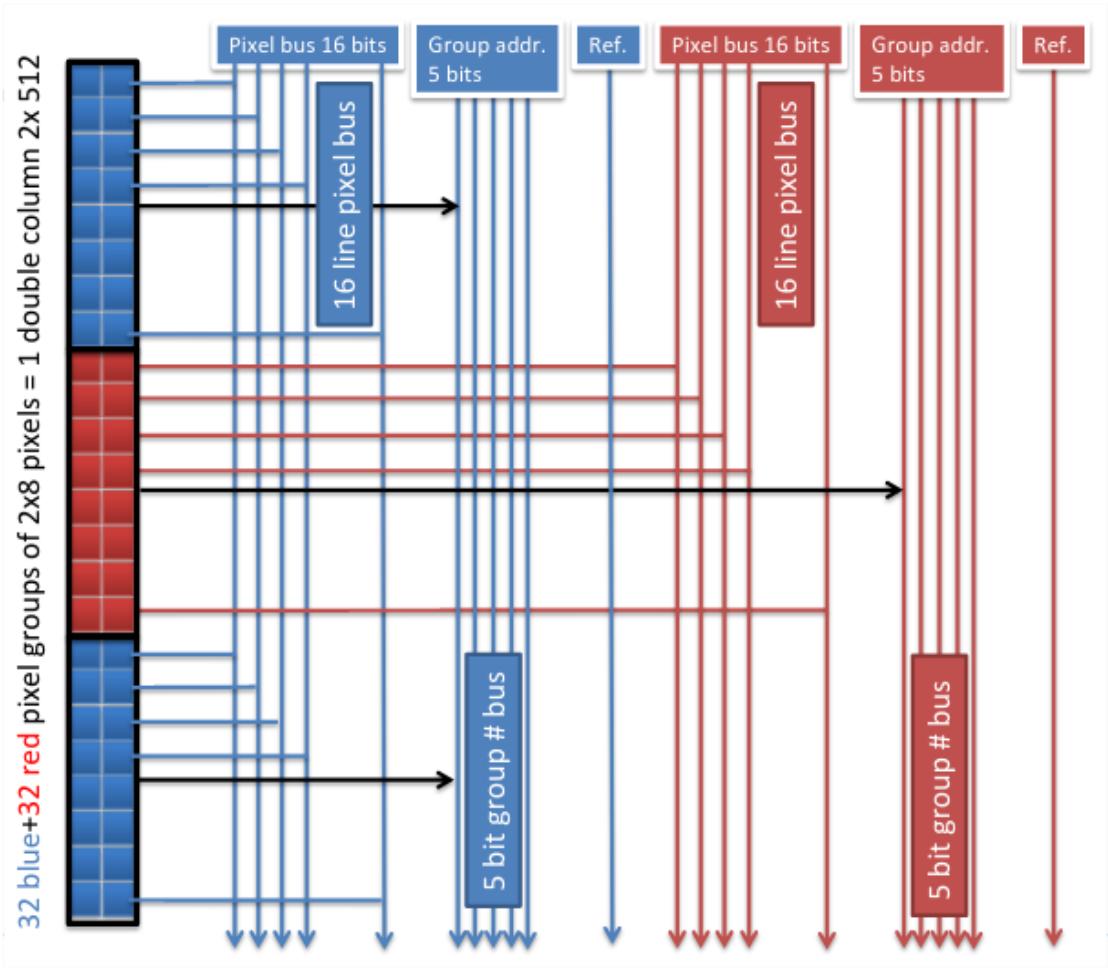


Figure 4.1: Double column readout organisation in MALTA architecture [91].

4.2.1 In-pixel digital circuitry

The in-pixel logic circuitry used for the generation of the transmitted pulses after the output of the comparator or discriminator in the front-end amplifier and for hit arbitration when multiple hits in the same group occurs is shown in Figure 4.2.

The assumed initial value of the input signals IN, RESET, MIRROR and VALID is 0. Once the front-end has seen a hit, the output of the discriminator produces an output signal (node IN for the digital readout logic) and toggles the STATE signal of the corresponding pixel to 1. The fact that the STATE signal goes to a high level means that a logic one is latched by the left flip-flop, provoking MIRROR to become high. The VALID signal is in common for all the pixels within the group, and goes to a high level whenever any of the STATE signals for the 16 pixels are high, making impossible the latching of further hits until the first hit has been read out. At the same time, it starts the generation of the reference pulse in the circuit shown in Figure 4.2. The VALID signal is inverted, delayed and then combined with the original signal to get a short pulse on REF_B, that is the inverted version of the reference pulse that is transmitted down the

column. The procedure is repeated again to obtain a similar pulse on the RESET immediately after the reference pulse. The duration of the reference and RESET pulses can be chosen using 4 control bits that act on the number of delay stages used within the delay cells, resulting in a pulse width between 500 ps and 2 ns. The reference pulse is combined with the STATE signals from all 16 pixels, generating the 16-bit pixel address, as well as 5 hard-wired group bits fixed for each of the 32 groups, which generates the 5-bit group address. Once the RESET pulse is generated, the and/or logic in Figure 4.2 will cause the MIRROR_B signal, now a logic zero, to be latched by the flip-flop on the right, effectively resetting the STATE signal and allowing the circuit to read the next incoming hit. The MIRROR signal will be reset as soon as the input signal goes down, and at this point the in-pixel logic will be ready to detect any new incoming hit.

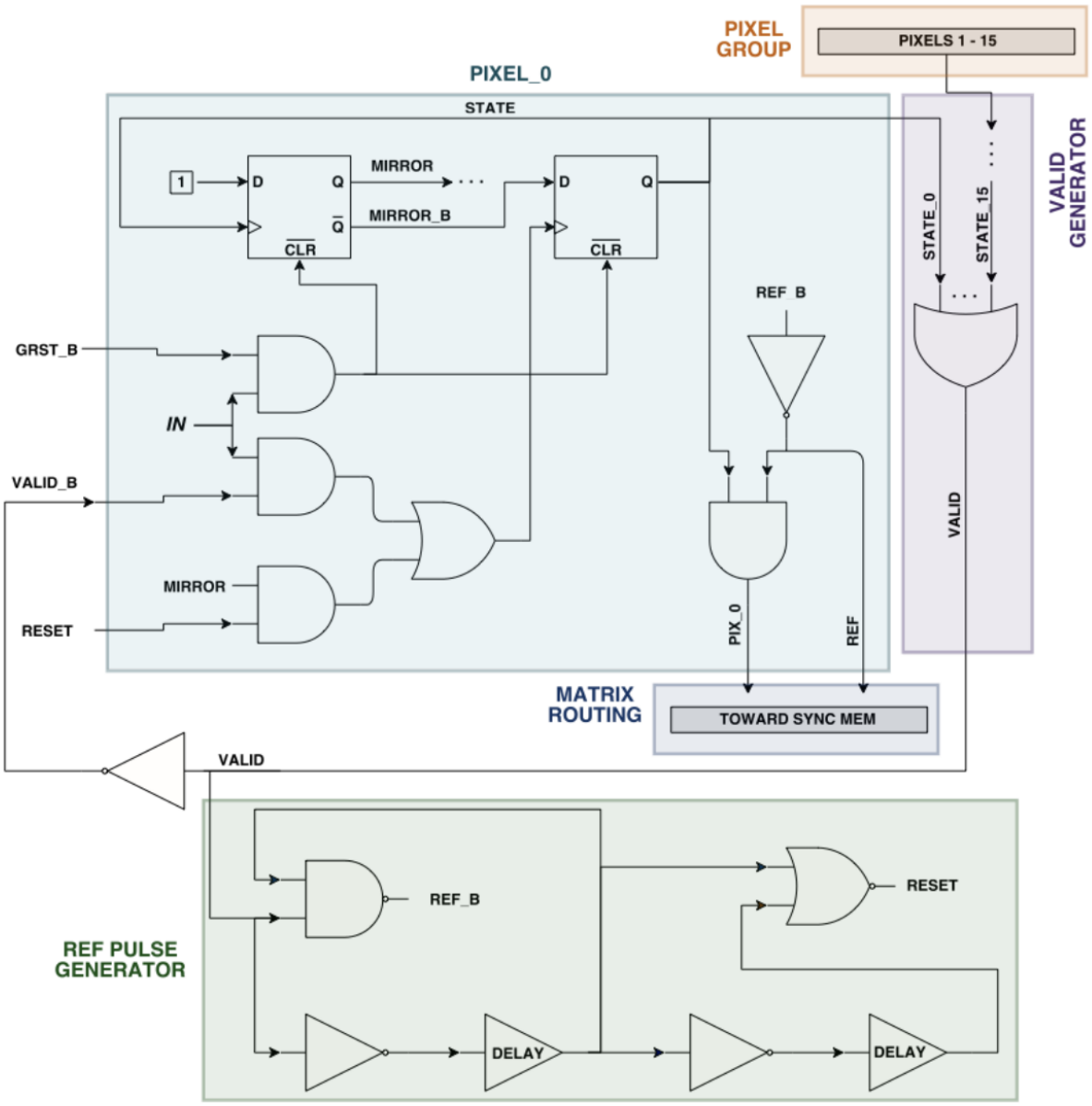


Figure 4.2: Schematic diagram of the in-pixel readout digital logic circuitry that includes the hit arbitration mechanism made out of a double flip-flop structure and the reference pulse generator common for all the pixels in the group.

As mentioned above, there is only one possibility of getting signals from several pixels readout at the same time and it is when two hits belonging to the same group arrive within a 1.6 ns time window. This will happen if both of the pixels collect a similar amount of charge and their discriminators react simultaneously. In this case the STATE signal for both will become high and the two pixel addresses pulses will be sent with the same reference signal. If the delay between the input signals is larger than the specified 1.6 ns, the VALID signal of the first hit will prevent the second pixel STATE to go high, so the hits will be read one after the other, after the generation of a second reference pulse.

Figure 4.3 is a simulation of the in-pixel circuitry just described. Nine pixels of the same

group have been injected with different charges of 3000 e^-, 2500 e^-, 2000 e^-, 1500 e^-, 1000 e^-, 800 e^-, 600 e^-, 400 e^-, 300 e^- and 200 e^- respectively. The waveforms in the blue rectangle are the discriminator outputs according to the received charge with the correct timing. It can be observed that if the signals are very close together only one VALID pulse is generated and several hits are read out at the same time. In the highlighted rectangle four of them will transmit simultaneously.

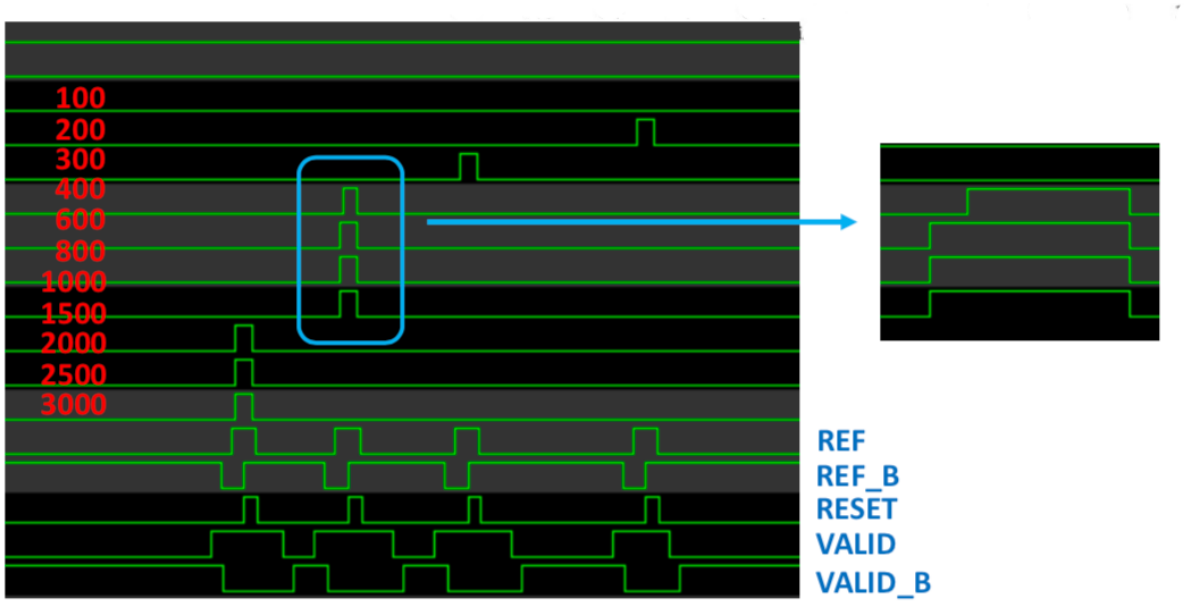


Figure 4.3: Simulation waveforms of the in pixel logic circuitry.

The pulses generated by the pixel logic digital circuitry are transmitted down the column using a special buffering structure that can be seen in Figure 4.4. The input in this case could be any one of the 22 reference/address bits from the pixels (the buffering structure is the same for all the bits). There are two separate 22-bit buses for alternating groups of pixels: the "blue" groups (A) or the "red" groups (B). In each "blue" group, the pulse is injected into one input of a NAND gate, while the other input is used to propagate pulses coming from higher up within the matrix. The output of this gate is then inverted in the next "red" group to achieve the correct polarity of the signals when they reach the next "blue" group. The procedure is repeated over the full height of the column, and is similar for pulses generated in the "red" groups.

The most critical part of the design just described is to preserve the pulse width and the alignment of the different pulses in the bus when they travel down the column. In order to achieve this a very careful routing was performed in which the the capacitance at the output of the NAND gates was "perfectly" balanced. A minor difference in this capacitance will cause a delay between the bits which accumulated over the length of the column that could lead to a bad miss-alignment of the 22 bits at the end of column. Another important consideration is to shield adjacent lines that can toggle at the same time by adding an extra line connected to ground to ensure that the capacitive load of each line is perfectly balanced . Eventually, to avoid

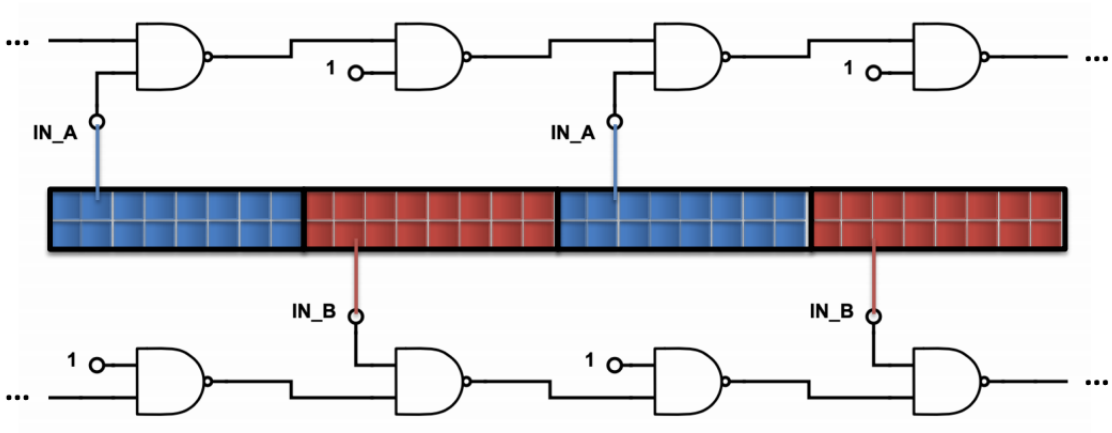


Figure 4.4: Circuitry to buffer the signals from the pixel to the end of column [77].

deformation of the pulses the rising and falling edges have to be carefully looked at as this can lead to a significant stretch or to the disappearance of the pulse depending on which edge has the smaller propagation delay. Parasitic extraction of the capacitance shows a maximal capacitance difference between lines of less than 3% resulting in a simulated maximal delay difference of $\approx$ 150 ps over the full column height with a worst case scenario of 30 ps for the total deformation of the pulses. The propagation time from top to bottom of the matrix ranges from 7.5 ns to 10 ns depending on the bias setting of the p-well and it has been confirmed by measurements in silicon.

4.2.2 Digital periphery and end-of-column readout

It is at the end of column logic that signals coming from both red and blue busses are merged onto a single bus. In the case of having pulses in the two buses at the same time an arbitration mechanism implemented in a similar manner to the one described in the pixel logic is in place. It gives priority to the signals on one bus and delays the ones on the other. In addition, a delay counter bit gets added to the address bits to show that there are simultaneous pulses and to be able to retrieve the timing information later on. The MALTA architecture is binary meaning that it does not store the time over threshold of the hit. However, if the time of arrival is known, via the time-walk, relative information about the charge can be retrieved. If a hit has a large amount of charge it will be present in the periphery before a hit with a charge level close to the threshold. Furthermore, there are two more bits added at this stage to identify the bunch crossing in which the hit took place. They are generated with a counter running at 40 MHz, so the hits will be timestamped and resolved within 4 bunch crossings or 100 ns. Eventually a group identifier bit is added to the data word that specifies whether the hits come from a red or a blue group.

The process of merging and delaying pulses is repeated in a tree like structure until all the pixels in the full matrix have been combined into a single bus as seen in Figure 4.5. In each level a double-column identifier is added so the final data word will contain the full pixel address. In

the second level of the tree, the initial 1 bit delay counter is expanded to 3 bits, increasing every time a pulse is delayed, which means that a pulse can be delayed by a maximum of 8 times in the total 9 levels of merging. There is an additional 10th level for combining signals coming from different chips, as there is an option on the chip for transmitting data from chip to chip via some lateral pads. After the data merging, 4 bits are added to identify the chip. This allows up to 16 MALTAs to be readout together. This leads to a final MALTA word that has a total of 40 bits as summarised in Table 4.1. The 40 bits are transmitted off-chip with 40 parallel LVDS drivers, LAPA, designed to operate at up to 5 Gb/s with a peak-to-peak jitter below 30 ps, and are capable of transmitting the shortest 500 ps wide pulses [104].

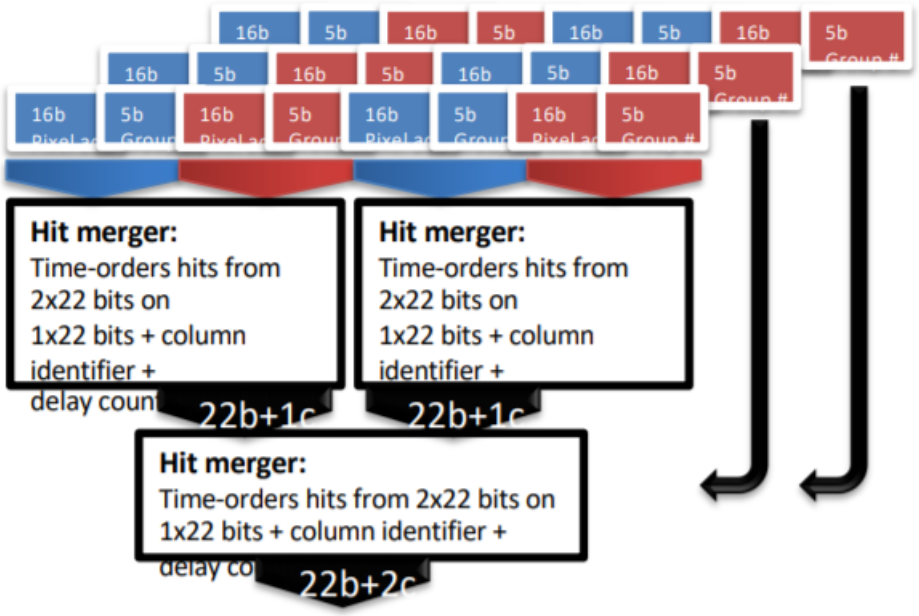


Figure 4.5: MALTA merging structure to append the data word.

Table 4.1: MALTA word bits information.

Bits	Content
0	Reference pulse
1-16	Pixel address
17-21	Group address
22	Group identifier
23-25	Delay counter
26-33	Double column address
34-35	BCID counter
36-39	Chip identifier

As a backup feature a simplified way of merging the hits from the matrix is also included in the readout logic. It consists of a binary tree of OR gates for each bit, which basically merge all

the pulses coming from the full pixel matrix, but without any arbitration in case of simultaneous hits. For low data rates, this provides a similar functionality to the more complex merger logic. However, in the case of multiple hits happening at the same time on multiple buses, combining their address bits with an OR gate will cause a corruption in the address data. Consequently, for very high hit rates this mode of operation could cause data loss and hence a loss in detection efficiency. All the chips have been tested using this mechanism instead of the merger as it did not work.

4.3 Mini-MALTA architecture

The asynchronous readout has proven to have very positive features like the low power needed to operate the matrix and the excellent time resolution. Nonetheless, some other features should be improved, this is the case of the tree like merger structure. Therefore the Mini-MALTA prototype was not only made to study sensor and detection efficiency improvements but also to synchronise the data at the end-of column.

The basic idea of this new circuitry is to use the reference signal generated by the pixels to store the asynchronous pixel and group address signals into a random-access memory (RAM). As described in Section 4.2, the pixel and group addresses are aligned with the reference pulse, so this last one can be used as the write enable signal (WRITE) for the memories. Each of the two 22 bit buses in a double column, corresponding to the red and blue groups, have their own synchronisation block, making a total of 512. Besides the pixel and group address information, the value of two counters will be latched with respect to the reference signal, a 3 bit BCID counter running at 40 MHz and a 4-bit fine time counter working at 640 MHz, provided with external clocks. This leads to a precision on the hit time of arrival of 1.5 ns which can give a good indication of the charge deposited. If there are multiple consecutive pulses, up to four data words can be stored in four different rows of the RAM memory. Figure 4.6 depicts the 28 bits x 4 RAM cells present in the Mini-MALTA.

The basic building block of the RAM cells is the standard dual port RAM of TowerJazz 180 nm. It consists of 8 transistors, two cross coupling inverters and and two sets of two NMOS transistors used for writing and reading. The main advantage of using a dual-port cell is that it enables the simultaneous writing and reading of different rows within the memory. The routing structure connecting the matrix signals to the inputs of the RAM, following the MALTA procedure, has a balanced capacitive load, and additional buffers and delay gates are added to the address lines to ensure that a maximum timing misalignment of 400 ps with respect to the reference signal still results in the correct writing of the RAM cells.

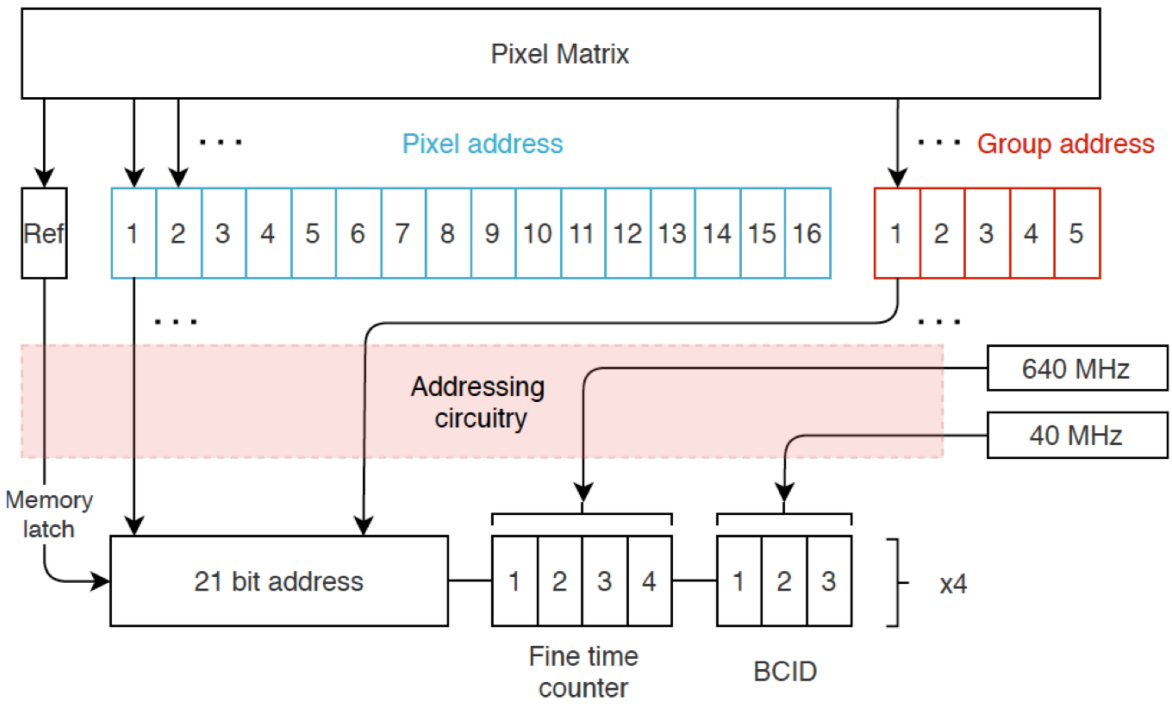


Figure 4.6: Mini-MALTA synchronisation block composed of 4 RAM cells that store 28 bits.

4.3.1 Readout logic

The WRITE signal of the RAM cells is the reference pulse and the READ is generated by the synchronous peripheral readout logic. This logic works with the 640 MHz clock used for the fine timing tagging. When a hit is stored in a memory a signal is generated and sent to a priority encoder announcing that an event has been stored. Then, the priority encoder sends the READ signal to the corresponding memory. After that, the data word will by transferred into a 64 row first-in-first-out (FIFO) memory. This second memory is a standard clocked RAM memory provided directly by the foundry. There is one 64 row FIFO for every 16 synchronisation blocks. In the FIFO, 4 more bits corresponding to the column number and 16 bits to increase the BCID identifier are appended to the data word, making up for a final length of 48 bits. The extra BCID bits will allow to store the word for longer without losing the timing information. The priority encoder simply gives preference to the leftmost synchronisation memories when several hits have been stored in them. The time to decide which memory is read first and then transfer the data from the synchronisation memory to the FIFO takes 3 clock cycles of the 640 MHz clock. The buffer depth has been chosen so the data losses are negligible. The choice of having three bits for assigning the BCID timestamp during synchronisation also was taken considering the readout rates. Eight 25 ns clock cycles are enough to read out the synchronised data in all cases. Once the data has been stored in the FIFO there are two different possibilities to send it out to the

DAQ system. The first one is after two 40 MHz clock cycles the word is sent to a 8b/10b encoder so a DC balanced output is guaranteed, with a similar number of ones and zeros. Furthermore, having this signal will allow the chip output to be AC coupled to the readout systems, which is the usual way in the LHC experiments. In order to be able to encode the 48 bits word six 8b/10b encoders are used to provide the final 60 bits. The parallel 60 bits are then transmitted to the ALPIDE data transmission unit (DTU), which serialises the bits and transmits them off-chip through a single LVDS data output at 1.2 Gb/s using double data rate, sampling at both edges of the clock, and a 600 MHz clock generated by an internal PLL [107]. The second readout option, known as the "slow readout " was introduced as a back up, the output of the FIFO is directly serialised using a shift register with a clock of 40 MHz. Finally, the data is transmitted off-chip using the LAPA LVDS driver, already implemented and tested in MALTA. It takes 25 ns x 48 bits = 1.2 μs to take the data out instead of the 0.8333 ns x 60 bits = 50 ns that will take with the fast readout option. This mode of operation limits the maximum output bandwidth and readout rate of the chip, but preserves all the address and timing information, and provides a functionality equal to the "fast" readout for sufficiently low particle hit rates. A sketch of the readout is shown in Figure 4.7.

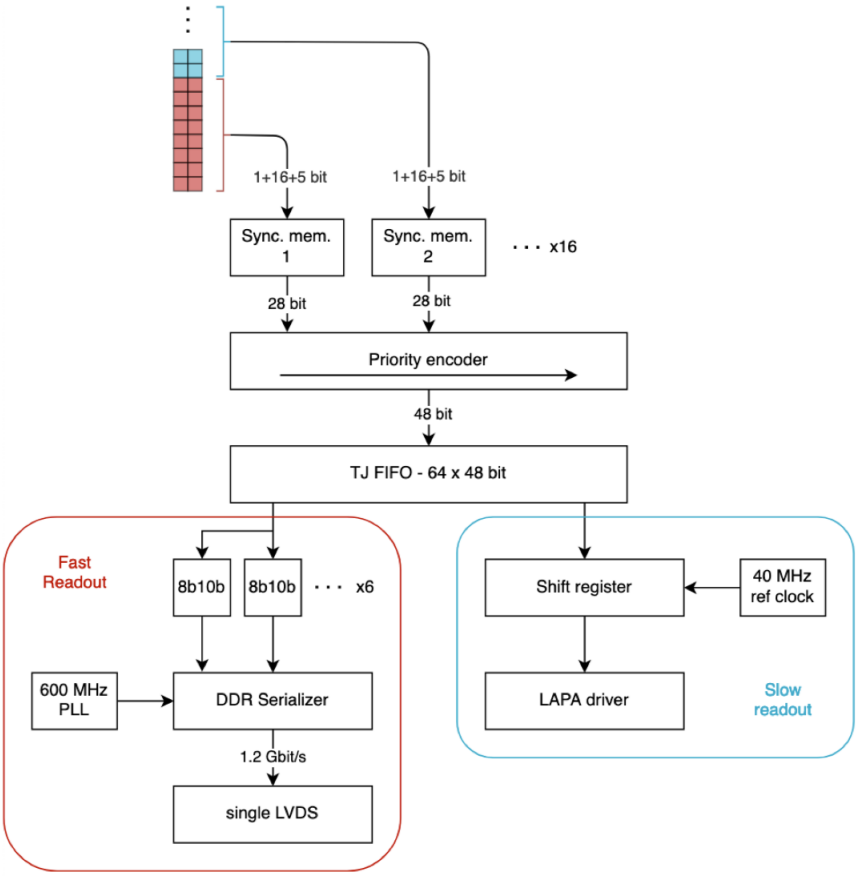


Figure 4.7: Mini-MALTA readout block diagram with the two different output options.

4.4 Architecture simulations

After having tested the MALTA and Mini-MALTA devices in detail, it was determined that some further studies and simulations were required to evaluate the architecture performance of the Mini-MALTA devices and properly dimension the memories if the same architecture will be implemented in full size sensor. The main goals of this study are to determine the number of synchronisation memories needed per column, to dimension the size of the FIFO and to determine the necessary readout rate. Having the BCID information stored will allow to discard all events that have not been selected by the experiment's trigger system. When these simulations were carried out, ITk specifications foresaw two different trigger levels with a maximum rate of 4MHz[1] that will determine which data is readout and then analysed. These triggers will have associated to them a bunch crossing identification (BCID). The BCIDs are related to the LHC collision rate and there is one BCID generated every 25 ns.

The simulation was carried out in a python framework using standard libraries and taking into account the following considerations. The triggered events were generated according to a Poisson distribution that selects on average one every 10 BCIDs to be read out. In this way, considering that the frequency of the BCID is 40 MHz a trigger rate of 4 MHz is emulated. The detector expected hits are MonteCarlo data-sets generated and provided by the experiment in their own simulation framework called Athena [111] in this case for the outer barrel of the ITk detector. The simulations correspond to average conditions in the detector for the luminosity scenario of HL-LHC, ttbar events with 200 minimum bias pile-up events. In these data-sets the geometry of the detector and the correct pixel dimensions of 36.4×36.4 μm^2 for a 2×2 cm^2 sensor are taken into account as well as its thickness of 150 μm . The data-sets give X-Y positions and the charge of the particles traversing the detector. In every file there are all the events corresponding to 5000 BCID and in order to be as realistic as possible a noise level of average 10 electrons following a Poisson distribution has been applied to the given charge. In addition, all the hits with a charge smaller than 200 electrons have been removed as this is the minimum operational threshold measured in the MALTA sensors. Taking into account all of these considerations the occupancy or total number of hits expected per module for the outer barrel layer of the ITk is shown in Figure 4.8. The line connecting the different modules hit rate has been included to be able to better understand the variations. However it is important to note that every module is independent from the others.

[1]The 2-level trigger design was later dropped in favour of a single level 1 MHz trigger

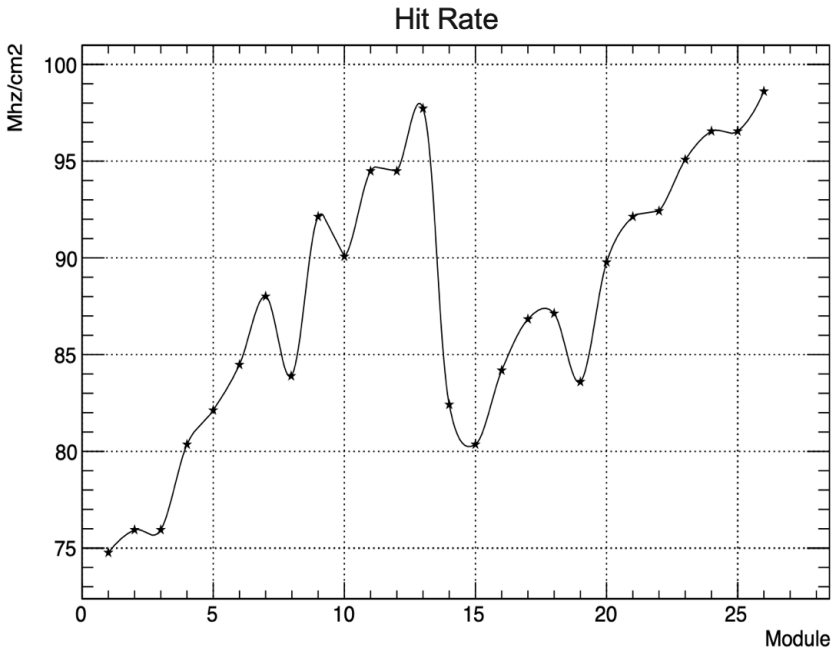


Figure 4.8: Simulated occupancy rate in MHz/cm^2 per module number for the outer barrel of the ITk according to MALTA performance. The module number corresponds to the z-position.

From this plot it can be extracted that the average number of hits expected per module is higher than 80MHz/cm^2. The dip that can be observed in Module 15 is due to transition from the flat barrel to the inclined barrel. ATLAS Pixel Detector, in order to reduce material and therefore improve tracking, consists of five barrel layers with inclined modules starting from $|\eta| > 1.4$ as shown in Figure 4.9. By knowing the rate in MHz / cm^2 per module and how many 2×2 cm^2 sensors are in each module, the total number of expected hits per BCID per chip can be calculated and it varies from 6.4 to 8.4 hits depending on the module z-position.

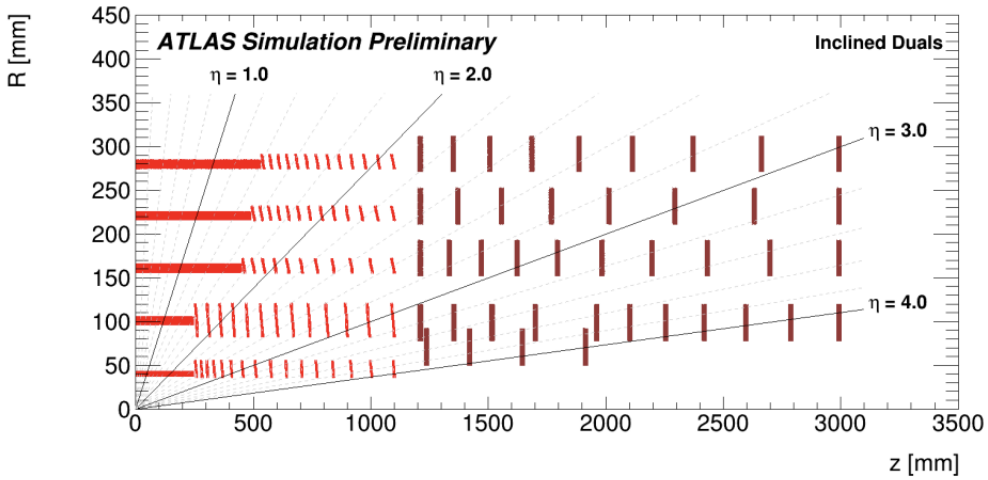


Figure 4.9: ITk layout for different radius and heights. Note that this plot does not agree with Figure 4.8 in the total number of modules due to the simulation of different layouts.

A deeper analysis of the occupancy can be done and in Figure 4.10 the pixel hit map for 5000 BCIDs is shown for chip 0 in module number 22 which will be in the inclined region. As mentioned above, the hits in a module have been re-mapped to their corresponding X-Y coordinates in the pixel matrix for each chip, therefore the axes are 512×512.

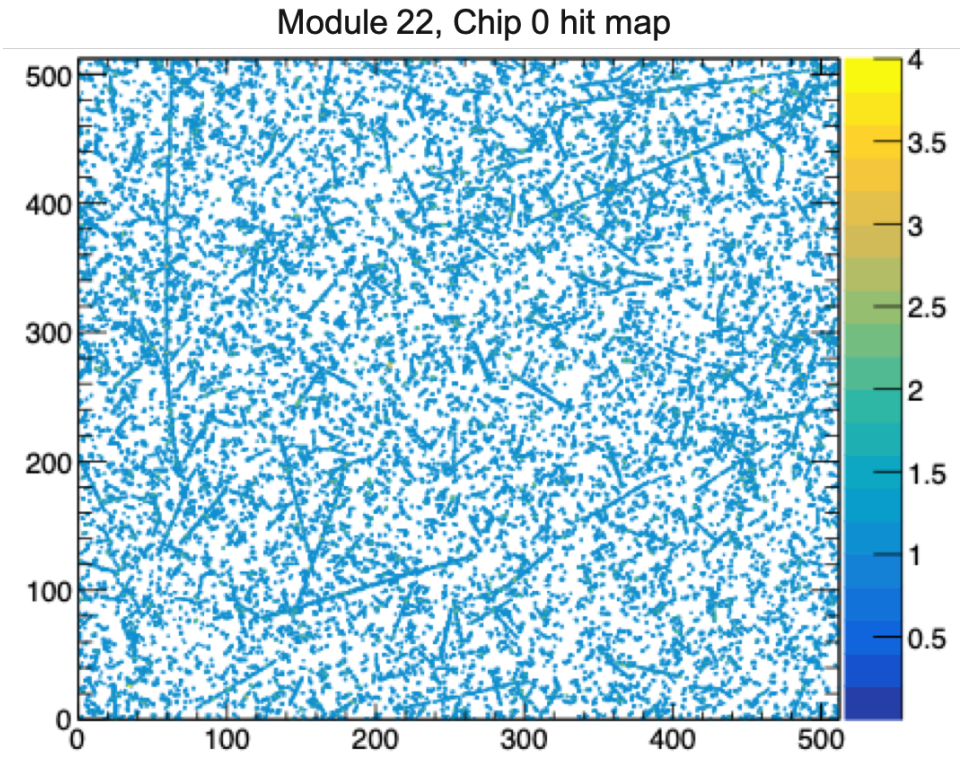


Figure 4.10: Hit-map for chip 0 in module 22 of the outer barrel of the ITk according to the MALTA sensor performance integrated over 5000 BCIDs.

A remark from this hit-map is the long delta electron tracks that can be observed in all directions. Delta ray or delta electron is a secondary electron with enough energy to move a significant distance away from the hit position, producing a long trail of ionisation. These events can leave hundreds of hits and will definitely be problematic if they have to be stored, specially as they are not interesting for the physics analysis. Therefore the readout should have mechanisms in place to stop the reading of such phenomena.

Figure 4.11 shows a histogram with the total number of hits per BCID for module number 1 chip 0, that has a low hit rate. This information of the number of hits that will need to be read per triggered event is key to defining the dimension of the memories.

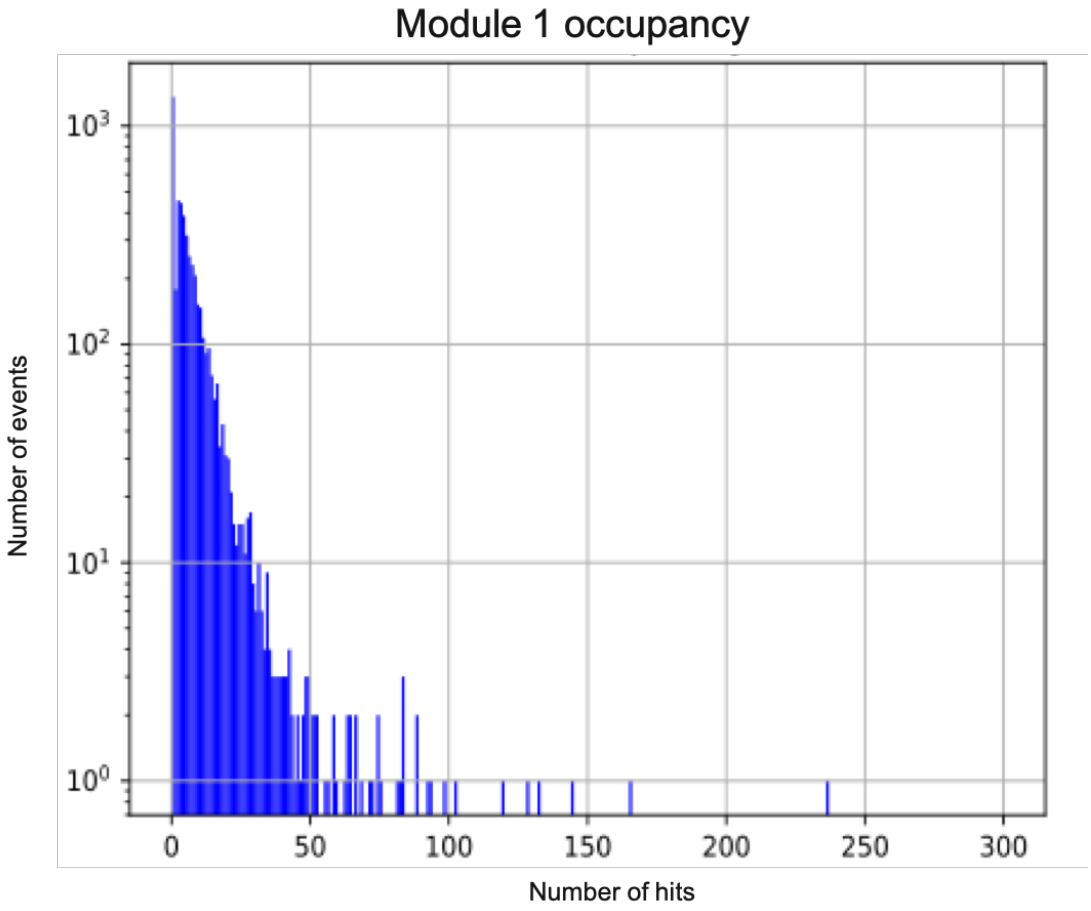


Figure 4.11: Simulated histogram of the number of hits per BCID for module 1 chip 0 for 5000 BCID, the average is 6.9 hits per BCID.

In most of the events it is observed that there are no hits or a few but there is a tail with high occupancy, and there is a non-negligible probability that for every module, events with over 50 hits will be triggered, agreeing with the observation from the previous Figure 4.10.

After having studied the occupancy values of MALTA, the in-time efficiency of the chip was determined. To calculate it, the charge value given in the simulation is converted into time according to the time-walk curve that has been measured in a MALTA chip for a threshold of 200 electrons, see Figure 2.24. For this time extrapolation, not only the charge value has been taken into account but also the propagation delay of the hits until they are avalilable at the end of column logic. This propagation delay varies from 0 ns to 7 ns. To give more details on how the charge conversion into timing is achieved, in the MALTA chip there is a pulsing mechanism, described in Section 3.1.2 in which different fixed charges are injected into the circuit via a capacitor to specific pixels along the matrix and the time that takes the information to get to the periphery is measured. Gathering together all this information similar curves to the one presented in Figure 4.12 can be obtained for all the pixel positions. The plateau seen at low charges is due to the fact that the sensor could not be operated at such low thresholds so it was modelled in this way.

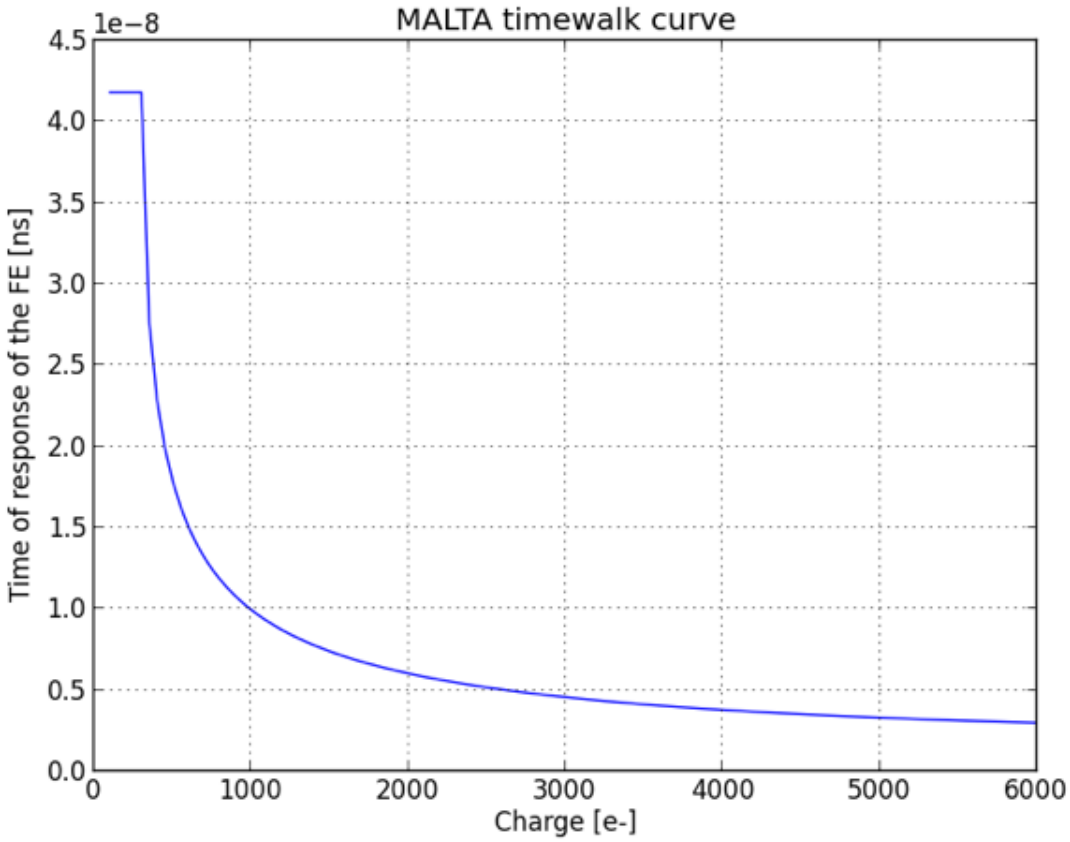


Figure 4.12: Time-walk of one MALTA pixel measured via pulsing and used to extrapolate charge values given in data-set into time.

The analysis of these curves shows that for small charges, the time-walk is very long. Actually for charges smaller than ≈300 electrons the hit will be in the periphery after more than 25 ns. On the other hand, for large charges, this time is short and the hit will be readout in a few ns.

The data-sets where modified so the charge information is translated into time. In Figure 4.13, a histogram that represents how long the hits take once the discriminator fires until they get to the periphery for chip 0 in module 1 is shown. There is a non negligible amount of hits that arrive after the 25 ns time window that determines the in-time efficiency. It is good to remember that this time takes into account two different effects, the time-walk and the propagation delay in the bus. According to the plot, on average any charge smaller than 300 electrons will be tagged in the following BCID. If the worst combination of a hit with little charge comes at the top of the matrix then it could be seen in the periphery after 50 ns. It is worth pointing out that this plot is a very pessimistic model of the reality as it assumes that the BCID goes from 0 to 25 ns when it should be considered from 5 to 30 ns because the fastest hits take 5 ns to reach the end of column. Furthermore, it is known that a version of the MALTA chip with the FE improvements demonstrated in Mini-MALTA could be operated at lower thresholds, improving the in-time efficiency. The gap observed in the plot from 43 ns to 65 ns is an artefact derived

from the modelling of the time-walk of small charges. Finally, after scrutinising the data of this plot it can be concluded that over 99% of the hits that will have associated a wrong time reference are members of a cluster for which the seed is always detected on time. This will have an implication for position resolution but in order to make any conclusions it would require a much more detailed study.

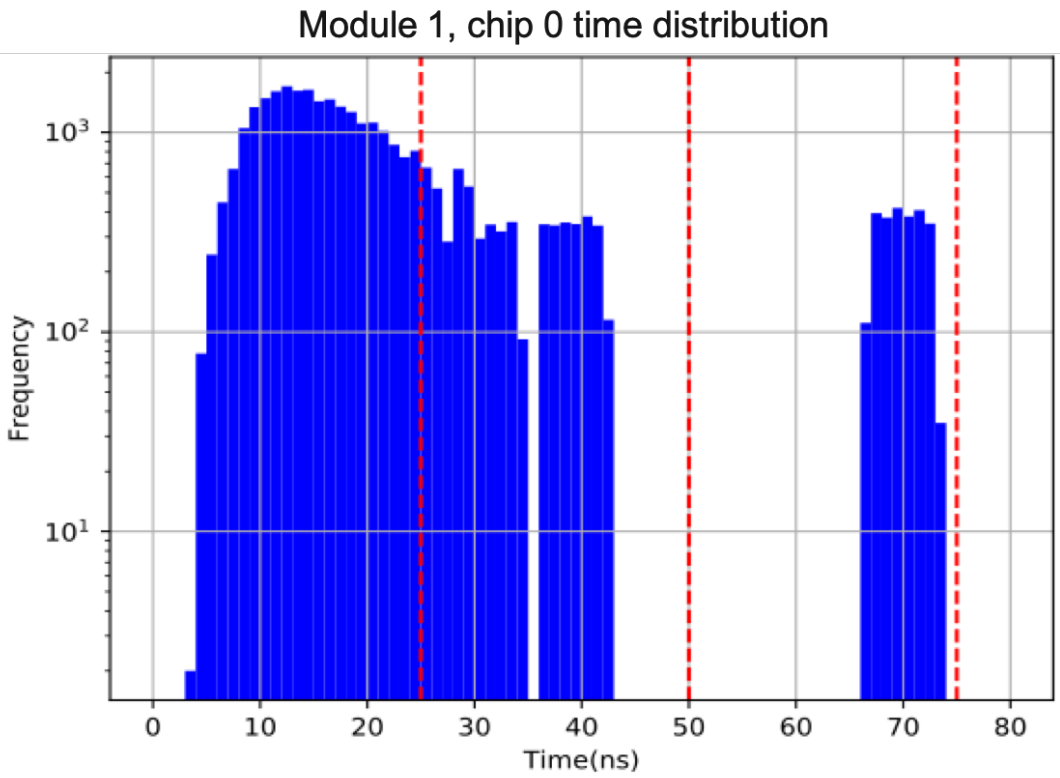


Figure 4.13: Time distribution of the hits arriving at the end-of column logic. Red dotted lines denote the different BCIDs.

The next step was to simulate the data losses in the synchronisation block or short term memory, similar to the one presented in the Mini-MALTA architecture, assuming one memory with 4 rows per column. Every hit in the column will be written in this memory in order of arrival. The read-out rate capability for this memory has been fixed to 8 hits per BCID, which is equivalent to having a readout rate of 880 Mbps, see Equation 4.1 where 22 is the number of bits coming from the matrix, consisting of the reference pulse plus 16 pixels plus 5 for the group identifier, and 25 ns for the BCID period.

$$Readout\ rate = \frac{22\ bits\ \times\ 8\ hits}{25\ \times\ 10^{-9}\ s} = 880\ Mbps \tag{4.1}$$

The percentage of data losses as a function of the module positon assuming that each column has 4 memories can be seen in Figure 4.14. The graph predicts that the data losses in this circuit are almost zero as they vary from 0.3% to 0.9%. In fact, the events that are lost are those with a large number of hits like the delta electrons that produce hits over many columns and cannot

be written and read-out in time so the memory positions will be overwritten. For the future, one of the proposed solutions to tackle these losses is to implement a triggered FIFO. It will work in a way that if it takes more than a certain amount of time to read a single event, the readout of that event will be stopped and discarded and the chip will proceed with the readout of the next trigger. These mechanisms are already in place in the current ATLAS detector, the FEI4 and in the RD53 chips. In the light of these results it can be concluded that 4 memory positions per column is enough as the observed losses are minimal. Nevertheless, the same calculation was repeated for the assumption of having 8 memories per column and an identical readout rate of 880 Mbps. In this case the losses were zero for most of the modules and 0.2% in the worst case. The delta rays type events are the primary generators of these losses.

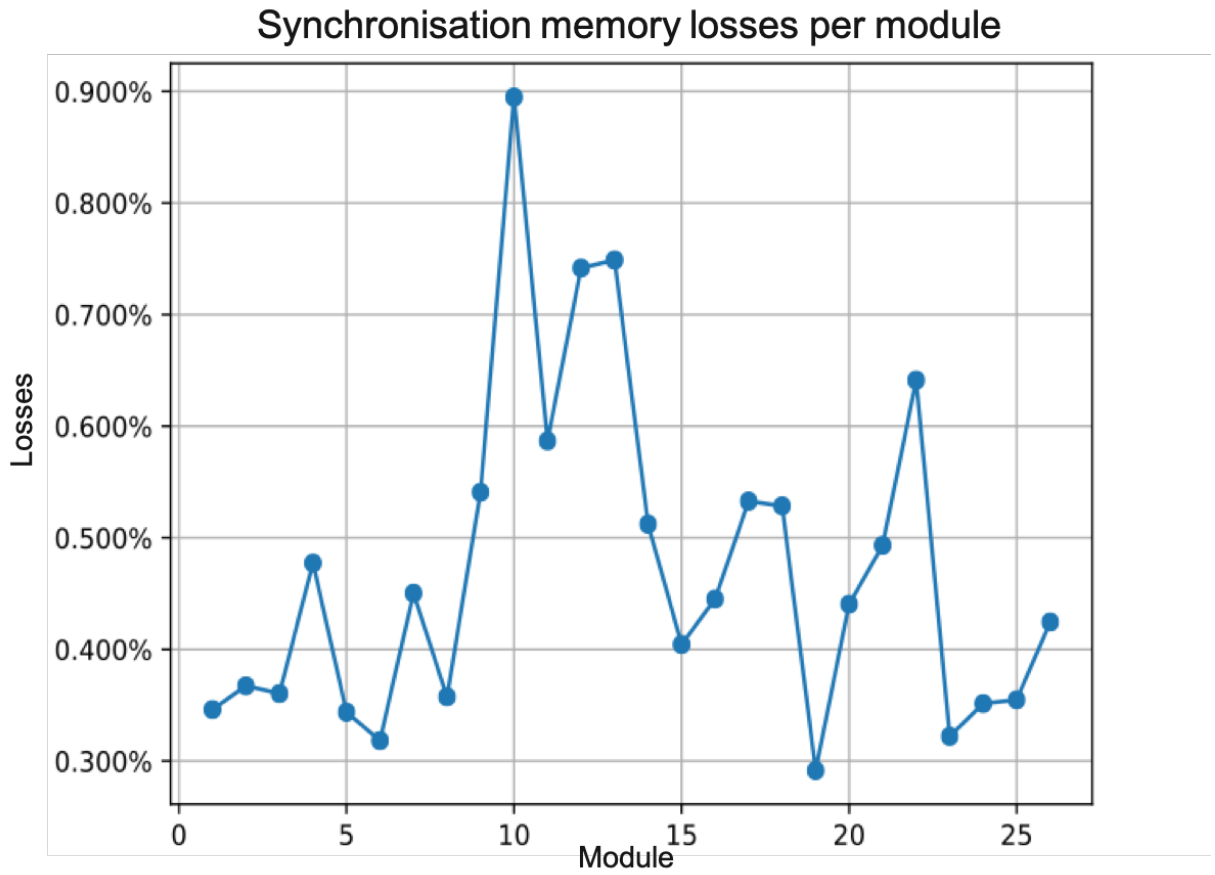


Figure 4.14: Simulation of the percentage of expected data losses in the synchronisation memories for every module assuming there are 4 memories per column based on the physic simulations of 5000 BCIDs.

A final simulation was done, to calculate the number of positions needed in the long-term memory or FIFO to be able to write and read the triggered events for a given output data-rate. To provide a realistic description of the experiment, as stated before, a trigger signal was simulated with a Poisson distribution of average 10, as this represents a 4MHz trigger rate because you trigger on average one out of 10 BCIDs because they come with a frequency

of 40MHz. The Poisson distribution gives a data set as follows [2,9,16,28,37, ...] and these numbers are used to highlight the BCID events triggered by the experiment and in consequence that contain interesting data to be analysed.

This time the simulation was done in a different way, instead of estimating the percentage of data losses the idea is to see what would be the minimum readout rate needed to be able to empty the FIFO and simultaneously estimate the number of memory rows needed in it so the specifications of data losses by ATLAS ITk are achieved. It is assumed that there will be one FIFO for every 16 columns, so a total number of 32 to cover the 512 columns. The data rate will still be expressed in number of hits emptied per BCID but the conversion to Mbps is straight forward and it only depends on the number of bits written into the FIFO as shown in Equation 4.1. The data can be thrown away from the FIFO once the latency has expired, so after 12.5 µs, and only if there is no data to be read. Approaching the simulation in this way is essential from the design point of view in order to determine the specifications of the serialiser and to estimate the physical area needed in the periphery for the memories. In addition, the number of bits stored in the FIFO and how the number of columns grouped per FIFO are also parameters to optimise. With reference to Figure 4.15, if the readout rate is fast enough to cope with the expected hit-rate, the number of memory rows occupied should remain constant on average regardless of the number of events that have been triggered. On the contrary, if the readout rate of the chip is not high enough, the number of memory positions required will linearly increase with the number of triggered events. This plot corresponds to the FIFO number 17 of module 24, so it groups columns that go from 320 to 336 of sensor 0.

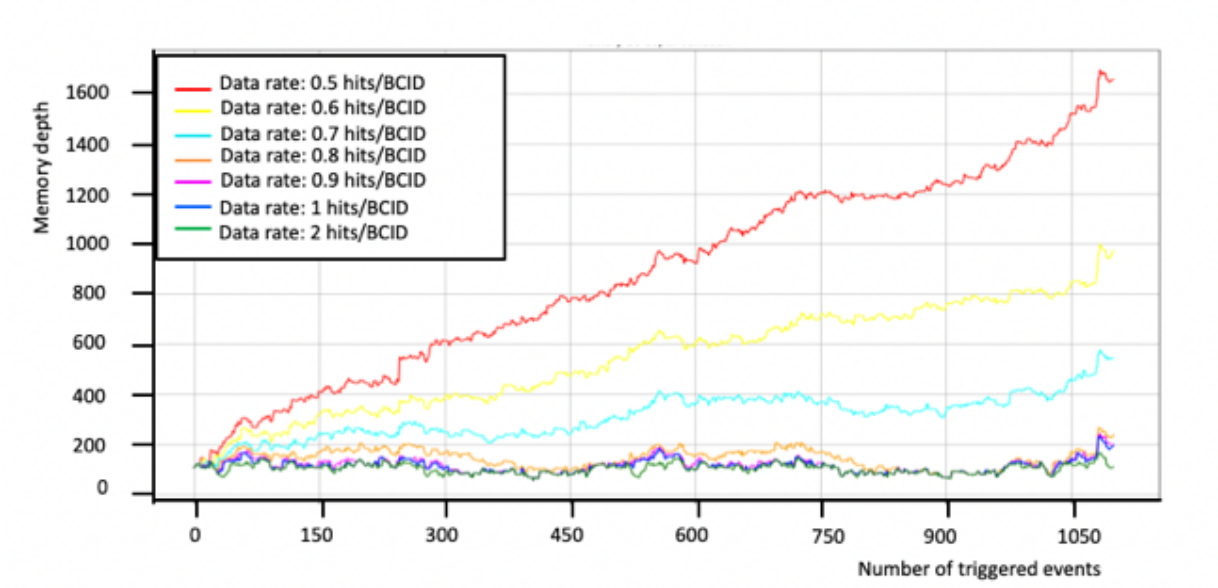


Figure 4.15: Memory depth required for a given number of events that have been triggered for different readout rates. This is for FIFO number 17 of module 27 corresponding to columns that go from 320 to 336 of sensor 0.

Figure 4.15 shows the two possible scenarios described above. For readout rates higher than 0.8 hits per BCID, the depth in number of memory rows of the FIFO remains constant as the number of triggered events increases. The number of memory positions needed is ≈ 160. However, for readout rates smaller than 0.8 hits per BCID, the total depth of the memory needed increases as more events are triggered because the previous events have not been read. To give a bit more of context, if 32 bits were to be stored in every memory position of the FIFO and the design of the future monolithic sensor would have a readout rate of 1 hit per BCID the serialiser in the chip should work at a rate of:

$$Readout\ rate = \frac{32\ bits}{25\ x\ 10^{-9}\ s} = 1.28\ Gbps \tag{4.2}$$

This rate corresponds to the one required by the ATLAS ITk Upgrade chip, the ITkPixV1, and the current LAPA driver present in the MALTA and Mini-MALTA devices, which was introduced in Section 3.1 is capable of reading out at 1.28 Gbps. If instead of storing 32 bits 40 bits were saved, the same readout rate will be sufficient to empty 0.8 hits per BCID without losing information.

Figure 4.16 shows the maximum number of rows needed in the FIFO for each module. As in the previous example the assumption is that there is one FIFO for every 16 columns or 8 double columns. The points represent the maximum number of rows necessary not to lose any data for a certain readout rate. This number has been taken for the worst device in the module and worst FIFO out of the 16 present in every chip.

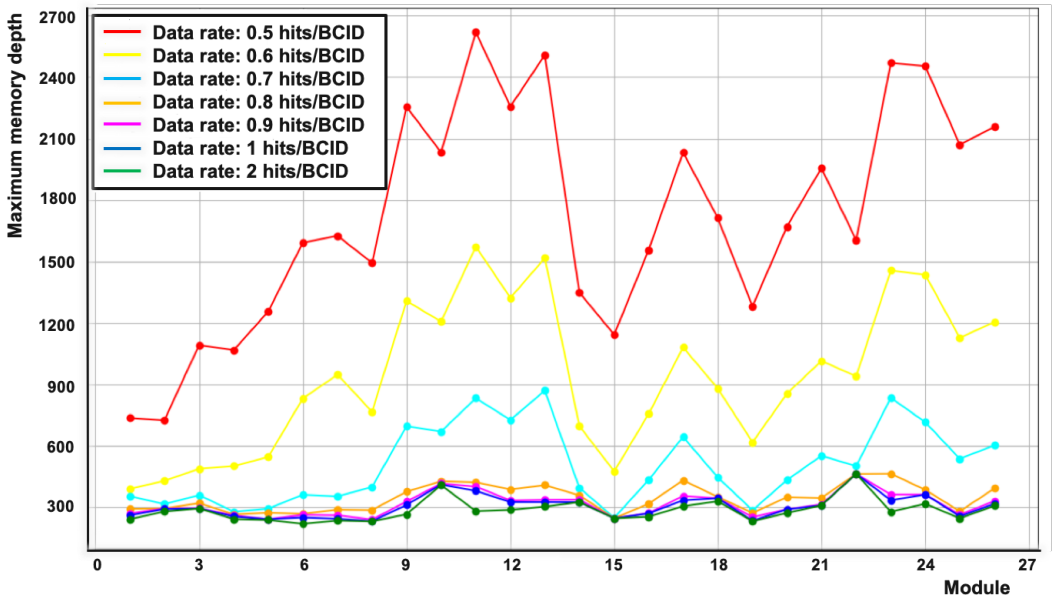


Figure 4.16: Maximum number of rows needed per FIFO for each module.

In agreement with the previous plot, the maximum number of memory positions is very similar for readout rates greater than 0.7 hits per BCID and it increases considerably for lower

rates. The maximum number of rows is around 400 which is double the number of the example shown in Figure 4.15. The 400 rows can then be considered as a safety factor of 2, and will have a good margin to accommodate the largest events. When the readout rate is not sufficient, the maximum number of rows has a similar shape to the simulated occupancy plotted in Figure 4.8 which is to be expected as more rows are needed for higher occupancy levels.

Lastly more simulations were performed to study what the benefits of doing in-chip clustering would be. For all the work presented so far the assumption is that every hit is associated to an specific pixel. If instead of doing this, hits were associated to hard coded addresses for every group of 1×4 or 2×2 pixels, it would be possible to read these 4 pixels simultaneously, in a single word, for the common case of a cluster of pixels being associated to a hit. In order not to lose information the word assembled at the end of column should have timing or charge information for each of the pixels in the cluster. This means that the total number of bits stored per memory will be larger than in the current scenario where only the coordinates of one pixel and its charge related information are stored. Another effect of having the addresses hard coded is that some clusters will still be split in different groups. Figure 4.17 is a representation of the different cluster possibilities explored and has a proposal of how the clustered word could look like. For the case of the MALTA family sensors as there is not TOT information it could be Time of Arrival or TOA that it is stored. In the picture for both clustered options, there are 4 hits recorded, represented with the red dots that get split between two hard coded groups. For the case of single pixel reading this will mean transmitting 4 words of 29 bits while for the clustered options it will be 2 words of 39 bits. Note that the length of the addresses and BCID are random just for the purpose of this example. What will determine the efficiency of the different options studied is the geometry of the experiment and the cluster shapes and distributions.

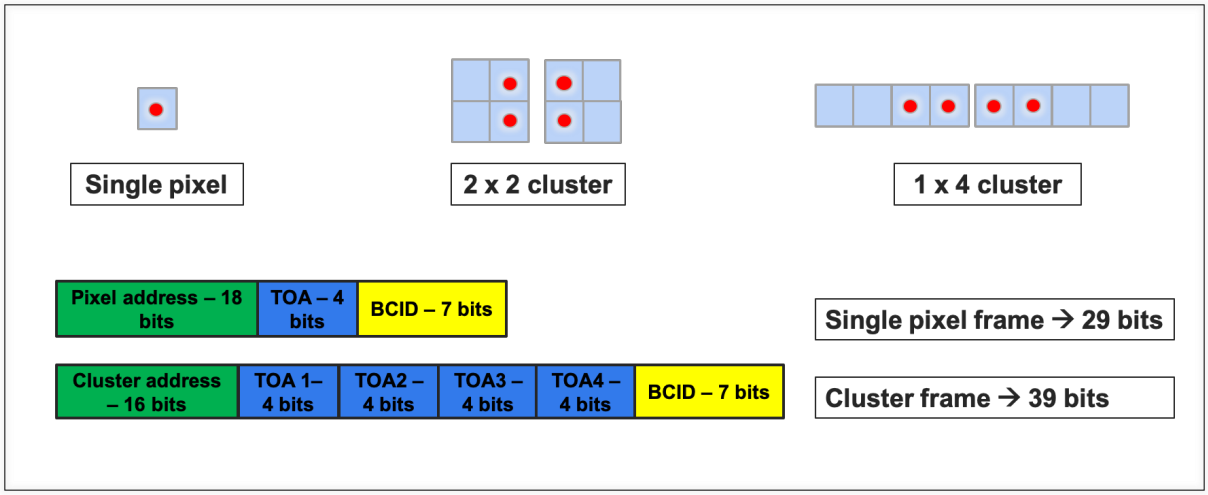


Figure 4.17: Different proposals to do in-chip clustering and difference in number of bits of the different words.

The idea is to see if the amount of memory positions needed per FIFO will be smaller and therefore if the bandwidth requirements are less strict when on-chip clustering is implemented.

Figure 4.17 shows the maximum memory depth required to store all the hits per module with remapping of the addresses and clustering the hits according to their new hard coded groups to which they belong. The result is that the equivalent readout rate from the non-clustering data of 1 hit per BCID is 0.7 clusters per BCID due to the difference in the bit length of the word as shown in Figure 4.17.

Figure 4.18 shows the maximum number of rows needed per module with on-chip clustering grouped in 2×2 pixels.

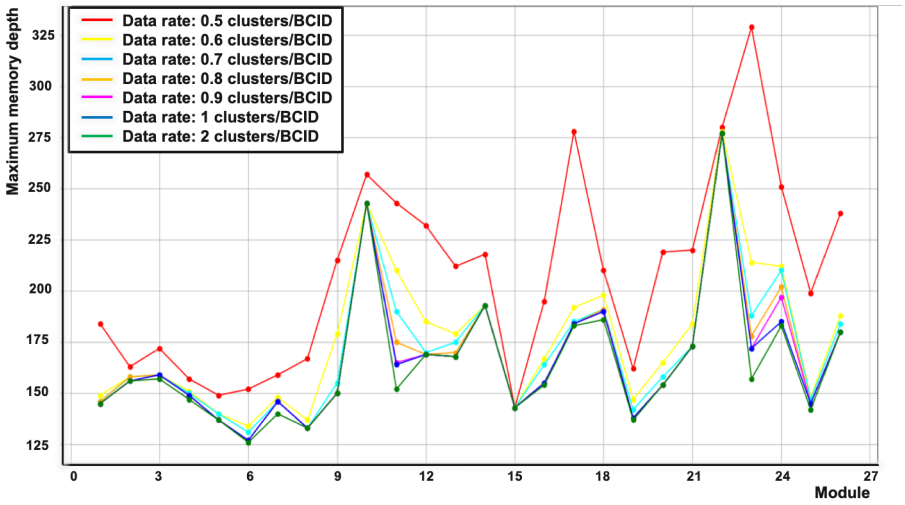


Figure 4.18: Maximum number of rows needed per FIFO for every module considering on-chip clustering with a 2×2 geometry.

Figure 4.19 depicts a similar graph for clusters of 1 x 4 pixels.

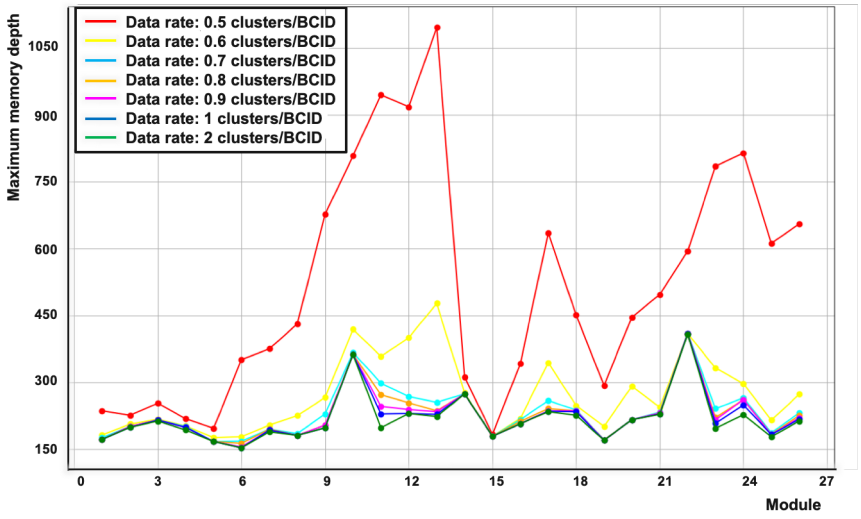


Figure 4.19: Maximum number of rows needed per FIFO for every module considering on-chip clustering with a 1×4 geometry.

There are several conclusions that can be extracted from these two plots. The most notable

one is that for both clustered options the maximum number of rows is significantly smaller than in the case of single pixel readout schemes by more than a factor of two. With regards to the most suitable orientation for the clusters, it seems that the 2×2 geometry is more appropriate to the way the particles pass through the sensor. The worst number of memory positions is 275 and even for lower readout rates this number does not increase rapidly as it does for single pixel readout and 1×4 clusters.

4.4.1 Conclusions

Architecture simulations for the ATLAS outer pixel layer have shown the need for a minimum readout rate of 1.28 Gbps. These simulations demonstrate the advantages of being able to operate at low thresholds and of minimising the propagation delays down the matrix in order to fulfil the in-time efficiency specifications given by the experiment. On-chip clustering can save some area in the design by reducing size of memories required to buffer the data. Finally, there should be protection in place in order to avoid large events. To summarise, it can be said that all the ingredients needed to design a system that will not lose data are known. Having a synchronisation memory per column with 4 memory positions that is emptied at a frequency of 880 MHz followed by a FIFO that groups 16 columns with 350 positions that is readout at 1.28 Gbps should be able to achieve this goal.

Chapter 5

The MALTA2 Monolithic Active Pixel Sensor

This Chapter will present the design of the MALTA2 sensor, a continuation of the MALTA family monolithic active pixel devices designed in TowerJazz 180 nm technology. It is a large prototype including all the improvements in radiation hardness obtained with the new processing and mask changes in Mini-MALTA, the gap in the n-well and the extra deep p-well introduced in Section 3.2, a new front-end design with a higher gain and reduced noise and dispersion, good timing resolution of a few ns and a new slow control that targeted overcoming the operational issues identified with communication in previous MALTA chips. This last part will be discussed in detail as it is part of the work of the thesis. Finally it will give a glimpse of the initial testing results.

5.1 MALTA2 design overview

MALTA2 is a 20 x 10 mm^2 sensor half the size of the original MALTA. It has a pixel matrix of 512 rows and 224 columns with a pixel pitch and cell-size of 36.4 x 36.4 μm^2. The front-end electronics are very similar to Figure 3.4. The new FE design is shown in Figure 5.1, where the main differences with respect to the previous front-end design are highlighted. Unlike in the MALTA, all the pixels in the matrix have the same FE so there are no sectors.

The main difference between the old and this FE is that it has been cascoded with transistor M3. The reasons for such a change is to have a higher gain while preserving the band-width. Furthermore, the size of the transistor M4 has been further increased. In Mini-MALTA its size was 1.22 μm x 0.36 μm, while in the new implementation it is 2.72 μm x 0.36 μm to be less sensitive to RTS noise. Finally the capacitance CS has also been significantly increased by a factor of 2.7 for the purpose of reducing noise and having a lower threshold dispersion per pixel. There is only one difference between one half of the pixel matrix and the other which is that if required a mask change could be done to remove the cascode from the FE and revert to the

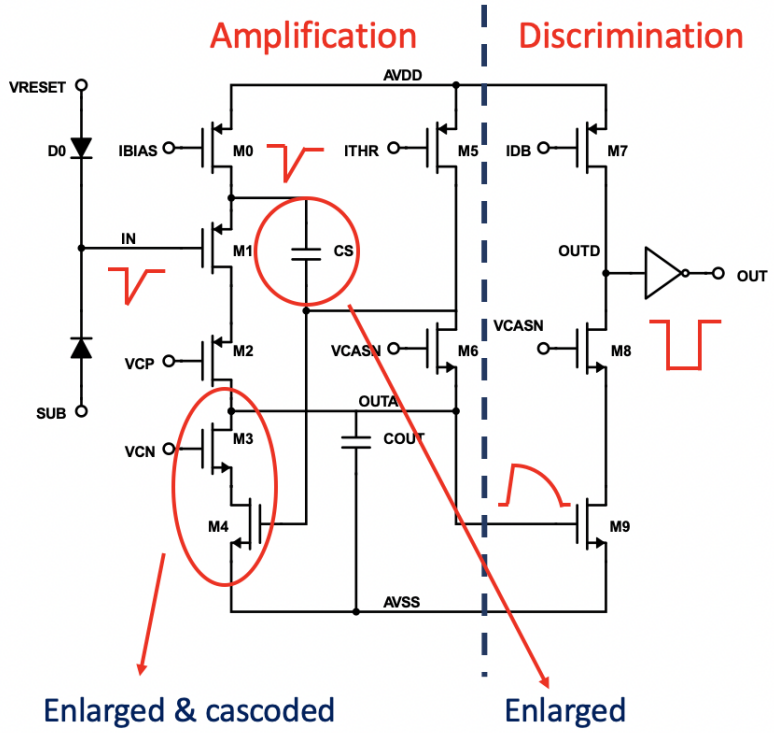


Figure 5.1: MALTA 2 front-end schematic with changes with respect to the old front-end design highlighted.

previous FE. The collection electrode has a diameter of 2 μm and the separation from it to the rest of the electronics is 4 μm.

The new simulated gain and expected time-walk curves compared to the previous front-end can be seen in Figure 5.2 to have a better feeling of the effects of the changes done. The gain has been increased by a factor two. The new sensor will have a faster charge collection time because of the process modifications to improve the electrical field were implemented.

The digital circuitry of MALTA2 has kept the same asynchronous architecture as in MALTA however it has a new slow control protocol. This change was made to overcome the difficulties of the previous control block that because of the absence of a clock tree made its operation difficult as it has been explained at the end of Section 3.1.3.

5.2 MALTA2 function control

MALTA2 function control which will be referred to as slow control is a full digital block in charge of providing the configuration to the device. It has been implemented as a shift register using only standard cells from the 180 nm technology of TowerJazz. It is a sequential protocol consisting of D-type flip-flops connected one after the other that move or shift the data from the input to the output of the flip-flops in every clock cycle. Once the load signal is asserted

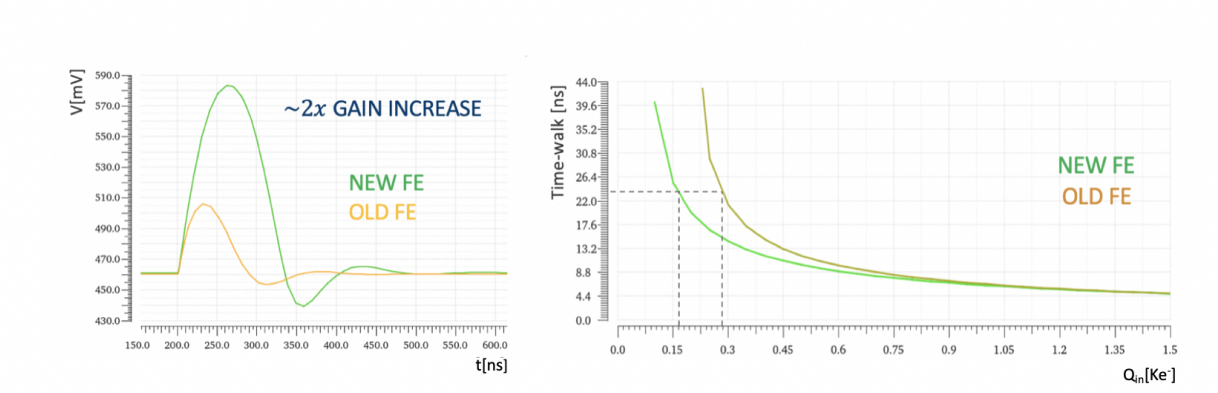


Figure 5.2: MALTA 2 vs Mini-MALTA simulated gain and time-walk curves. The input to this simulation was a delta function.

for one clock cycle, the total number of bits needed to configure the chip are stored in their corresponding registers in the matrix.

The main functionalities provided by the slow control are the following:

- Loading the default configuration of the chip. After power-up, a global reset-low signal is sent to the chip and a default configuration is provided to the matrix. These values, have been selected according to the results obtained from measurements on the MALTA chip and adapted for the new cascoded front-end.

- Pulsing. There is a pulsing mechanism in which a voltage is sent through a capacitor that allows a defined input charge to be injected into the FE as described at the end of Section 3.1.2. Both the voltage difference and the 'pulse_In', a signal needed to enable the propagation of the charge in the pulsing circuitry, are controlled via the slow control.

- Masking. Noisy pixels are masked by deactivating individual pixels or and double columns if there are many noisy pixels in that column.

- Biasing of the matrix. All the DACs are controlled by the slow control. There are 6 voltage dacs and 5 current ones that set the operation point of the FE, mainly its threshold and noise sensitivity. The first ones are one-hot encoded and the latter ones thermometer encoded. The encoding is done in software. A one-hot is a group of bits among which the legal combinations of values are only those with a single high '1' bit and all the others low '0' [127]. The main advantage of using this encoding scheme is that any change only accesses two flip-flops, it is easy to design and modify, it is simple to detect any illegal state and it runs with a fast clock. Thermometer encoding is done in a way that every number, n, is represented with with n ones followed by zeros to fill the length of your word. It is also known as unary code [128]. In a similar way to the one-hot encoded it is an easy encoding scheme that allows for high accuracy in the current settings and easy detection of invalid codes.

- Reset of the arbiters and mergers. The slow control is responsible for resetting the arbiter and merger mechanisms that ensure that data is sent to the periphery properly.
- It determines the width of the reference pulse which influences the throughput capabilities of the detector.

At any moment the configuration of the chip can be modified. In total 4322 bits need to be sent every time something needs to be configured via a serial input line. For the new changes to take effect the load signal has to be set to '1' for one clock cycle exactly after 4322 clock cycles, the necessary time to shift all the bits to their correct position. As the clock is running at 10 MHz, writing a new configuration takes ≈ 450 μs. The configuration loaded into the matrix cannot be readout. However, there is a serial output line that replicates what the chip has received before it has been stored in the registers to make sure that what has been sent is what the user has written and it should correspond to the latched data. Finally for the pulsing to take place, there is another signal "pulse_in" that has to be set to '1' to propagate the charge to the pixels selected to be injected.

5.2.1 Design steps

This section describes how the design of the slow control was carried out.

1. Register Transfer Level. This is commonly known as RTL. It is a design abstraction which models a synchronous digital circuit, in this case the shift register, in terms of the flow of digital signals (data) between hardware registers, and the logical operations performed on those signals [129]. Therefore it is constituted of two types of elements: sequential logic or registers and combinational logic in which the output only depends on the input. It is used to create high-level representations of a circuit, from which lower-level representations and ultimately actual wiring will be derived. A test-bench in order to check the design is implemented corresponding to the circuit at this stage and it will be used in all the following steps. The tools used were text editors to write the code in System Verilog [108] language and the Cadence Xcellium simulator [109] to see the wave-forms.
2. Synthesis. Once the RTL code has been verified and it is working as expected, it is translated into standard cells of the technology by the Cadence tool Genus [110]. This is a very important step in the design phase as the timing constraints that will determine the performance of your block have to be described. In the slow control, the main constraints are the frequency of the clock that will shift the data which is set to 10 MHz with a 50% duty cycle. The setup time is defined as the minimum amount of time that the data must be stable before the clock's active edge and was set to 500 ps, and the hold time which is the minimum amount of time after the clock's active edge during which data must be stable was set to 100 ps. In addition, the capacitance of certain outputs were specified as

they were extracted from the circuits to which they are connected. During synthesis and place and route these constraints will determine if the design is ready for manufacture or not. During this phase of the design a Logic Equivalence Check (LEC) is performed to guarantee that the functionality of the block in RTL is preserved after its translation to logic gates.

3. Place and Route (P&R). This phase consists of different steps that lead to the final circuit layout:

   - Floorplan. In this step, the physical area of the block that is being designed is defined. Pins and ports are assigned a rough location, which can further be refined depending on the Place and Route results. For this design it has a rectangular shape.
   - Powerplan. A robust power grid that addresses static and dynamic voltage IR drops is specified. In the TowerJazz 180 nm technology used for the MALTA2 design, the M6 and M5 metal layers are used for this purpose, one in vertical and the other in horizontal direction. In the original MALTA design most of the power stripes had minimal width and for the shift register of MALTA2 they were enlarged and more lines added.
   - Placement. During placement, all standard cells are placed in permitted locations on site rows. The aim of this step is to minimise the wire length, while ensuring optimal placement that will help faster timing convergence. No real routes are laid but a first estimation of timing is performed.
   - Clock Tree Synthesis. During clock tree synthesis, clocks are propagated and the clock tree is synthesised using clock buffers and based on the constraints initially generated for the synthesis. The major goal of this step is to achieve optimal clock latency while minimising clock skew. As the clock is the signal with highest toggling frequency in the design, the clock buffer tree accounts for over 75% of the dynamic power dissipated in an ASIC. This step was missing for the MALTA design as there was a line in the constraints setting a false path between the slow control clock and rest of its logic, which led to issues in configuring the block. The new design has been made to overcome the operational difficulties on the MALTA device.
   - Detail routing. With all instances placed and clocks routed, the signal nets are eventually routed. The aim of detail routing is to ensure minimum detours because these may have implications on timing, and to ensure Design Rule Check (DRC) violations like opens, shorts etc are minimised. This step performs multiple search and repair loops to keep the overall DRC violation count low or non-existent.
   - Sign-off. Once the block is finalised the full design is checked again to make sure it can be sent for manufacturing. Logic verification ensures correct functionality, phys-

ical verification ensures correct layout. During this step the following physical verification takes place: DRC (Design Rule Checks), LVS (Layout versus Schematic), electromigration which is the phenomenon in which an electrical current passing through a conductor causes the atomic-scale erosion of the material eventually resulting in device failure [130], electrostatic discharge violations (ESD), Antenna violations, Pattern Match (PM) violations, Shorts, Opens, Floating nets and many more. Timing verification verifies that the chip runs at the specified frequency by ensuring setup and hold is met for all timing paths in the design.

The MALTA2 slow control was designed following the steps mentioned above. Nevertheless there is one more step missing in the list because it has to be done continuously throughout the design, which is verification. Current ASIC designs are becoming so complex that in industry for every design engineer there are three people verifying the circuit.

5.2.2 Functional control design verification

The verification of the MALTA2 slow control was done with system verilog test-benches [112] for all the different design stages and with the development of an emulator with two different FPGAs.

Functional testbench

The sytem verilog test-bench was prepared so that all the functionalities of the block: masking, pulsing, biasing, resetting etc. were checked. It allowed the optimisation of the timing of external signals like the reset and the load. Once the design was clean according to the sign-off step, the final simulation with all the routing delays and parasitic capacitances extracted proved that the capabilities of the new configuration block were maintained with the correct timing. Three different corners were checked, meaning that the process variations, temperature changes and voltage supply of the cells are taken into account as well as the different responses from NMOS and PMOS transistors. These corners, are a set of libraries, known as liberty files, provided by the foundry that include the timing information for all standard cells and interconnects in the layout. The corners chosen were the typical corner in which the power value for the transistors is its nominal one of 1.8 V and the operating temperature is 300 K, the maximum corner also referred to as "fast" corner for which the transistor models are set to work at 1.92 V and the temperature to 233,15 K and the minimum corner or "slow" corner that specifies a power of 1.62 V and a working temperature of 398.15 K. The reason for analysing these three corners is because they cover the most likely operation conditions and the most extreme cases so if the circuit works for the three of them it should as well work well when fabricated. The foundry may also consider the temperature sensitivity of the library cells, which can vary for different processes and add extra corners to cover for those cases.

The following plots are simulations to verify that the design responds as expected. Figure 5.3 shows how two different configurations, visible in the serial_input signal, are sent to the chip. After waiting 4322 clock cycles in which the serial input bits are sent the load signal is asserted for a clock cycle. This procedure is repeated for the second configuration. If the value of the registers in the matrix has been changed, the new value is loaded, this can be seen in the signals SET_IDB and SET_VRESET_P that toggle their value after the load is set to '1', SET_VRESET_P register takes a new value after the first time the load is set to '1' and SET_IDB register after the second time the load is asserted .

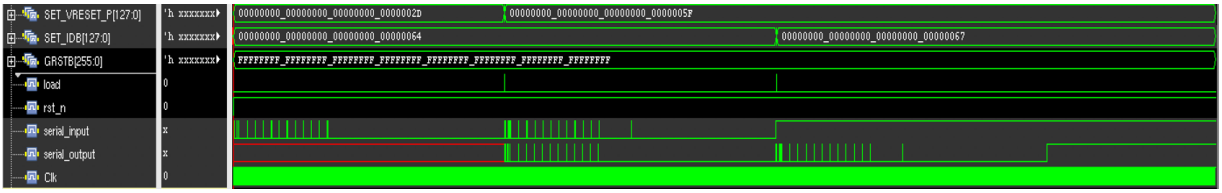


Figure 5.3: MALTA2 write and read configuration operations.

Figure 5.4, is an example of the pulsing operation simulation of MALTA2. The pixels that are going to be injected should be defined by their row and column positions, for this example columns 0, 1 and 2 are enabled in the matrix and the other input of the block, denoted as PULSE_I, has to be set to '1'. Then the specified columns will receive the pulse.

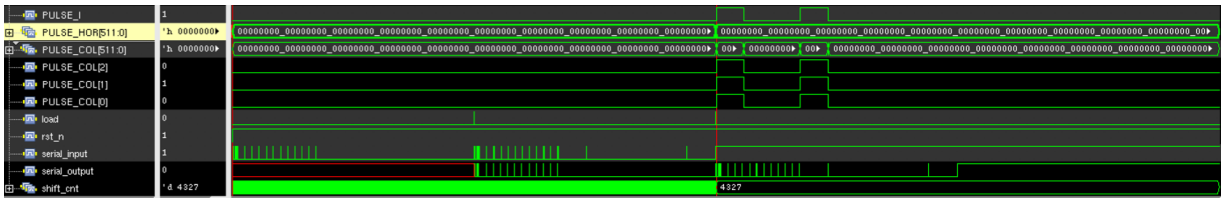


Figure 5.4: MALTA2 pulsing operation.

Emulator design

The development of the slow control emulator aimed to ensure that the timing in MALTA2 in response to external signals is correct, that the results of the test-bench simulation can be repeated with hardware and that the read back in the serial output will happen as expected. In addition, to make this setup function correctly all the software and firmware had to be developed which allows the chip to be tested immediately after it is fabricated. Eventually, the timing differences between the scans performed with the MALTA slow control and the new implementation could be evaluated.

The emulator setup consists of two Kintex 7 FPGAs connected together via a FMC cable. One of them is connected to a computer from where the serial input is sent. Figure 5.5 shows the boards in the laboratory.

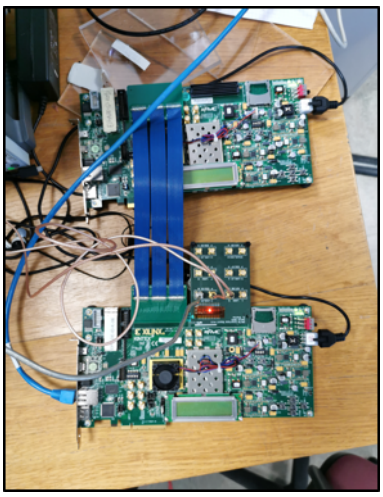


Figure 5.5: MALTA2 emulator setup in the lab with the two Kintex 7 FPGAs.

The communication protocol used to send the data from the computer to the FPGA is IPbus [125] which is a suite of software and firmware that implements reliable high-performance control links for particle physics electronics and is widely used at CERN. All the 4322 required to configure the MALTA2 bits are stored either in registers or FIFO's according to their size and functionality. On top of the IPbus own software a MALTA2 slow control class was defined to implement the control of the device. Operations like changing individually every DAC, setting default values, defining the pixels to mask and pulse are examples of the methods included.

Concerning the firmware development, on one side the code from the chip was ported into VHDL [108] to be compiled in Vivado without any modification and keeping similar timing constraints. Vivado is a software produced by Xilinx that allows for synthesis and analysis of hardware description language designs targeting development of embedded firmware for Xilinx FPGA and CPLD integrated circuit (IC) product families [131]. On the other side in the second FPGA, the IPbus registers were defined and a serialiser and a deserialiser are set up in order to send and receive the serial data, respectively. Figure 5.6 is a diagram that shows all the modules defined in the code.

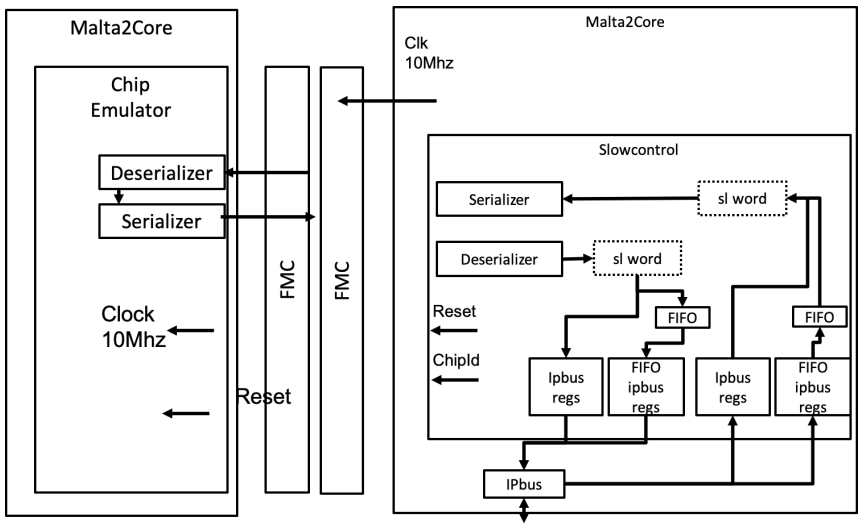


Figure 5.6: MALTA2 firmware implementation diagram in each of the FPGAs that constitute the setup.

Finally, Figure 5.7 is an image taken from the oscilloscope. The blue signal is the word sent from the user in the computer and the green waveform corresponds to the response by the emulator. This picture is equivalent to the simulation presented in Figure 5.3 and therefore shows that the chip configuration and read-back of the registers are working. It also means that the firmware and software are ready to be used once the chip is fabricated.

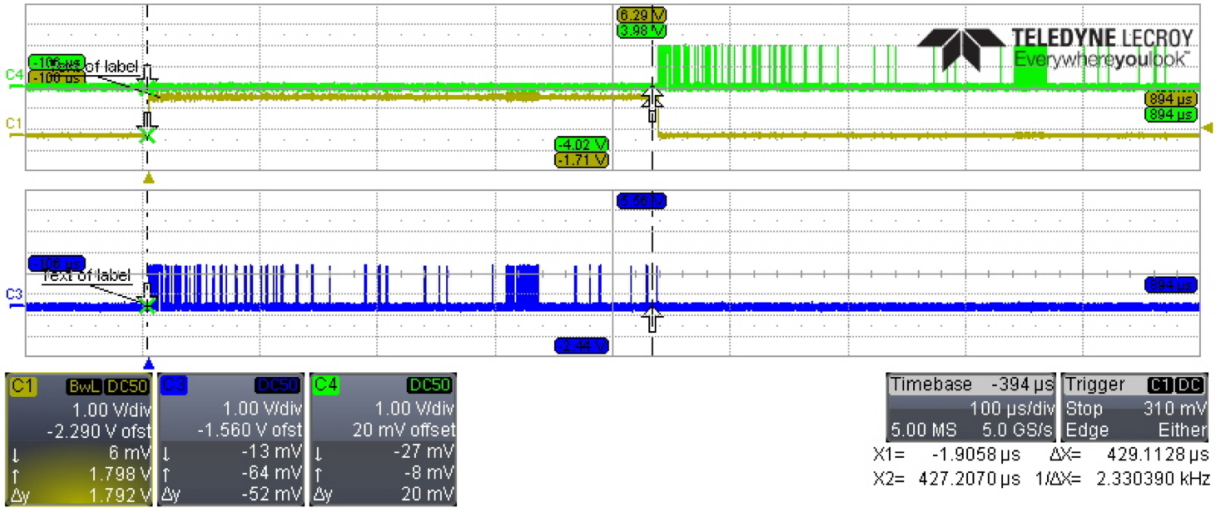


Figure 5.7: Serial input and serial output signals from MALTA2 emulator measured in the oscilloscope.

5.3 MALTA2 initial characterisation results

MALTA2 was sent for fabrication at the end of summer 2020 and received in the lab and bonded to the single chip card in the last week of March 2021. First tests targeted communication with

the device to make sure that the device can be configured as expected and without the need to toggle the digital power supply as had to be done for the MALTA. The hardest part was finding the correct alignment of the load signal so that the sampling of the 4322 bits occurs in the right place. The slow control was a fully functional block.

Initial tests were done to check the DAC linearity, source scans, analogue scans and to find the correct operational point of the front-end in order to perform threshold scans.

Figure 5.8 is an example of a DAC scan. It corresponds to the DAC called VLOW responsible for setting the voltage that will then be input into the injection capacitor. The actual injection voltage is controlled by the subtraction of two DACs VHIGH and VLOW. Therefore it is widely used in the threshold scans and for pulsing. The reason why it saturates is because it is buffered. As expected it goes all the way to 1.8 V covering all the voltage range from 0 to the supply level. The slope corresponds to 13 mV per DAC bit. This DAC scan shows that the DAC works as expected.

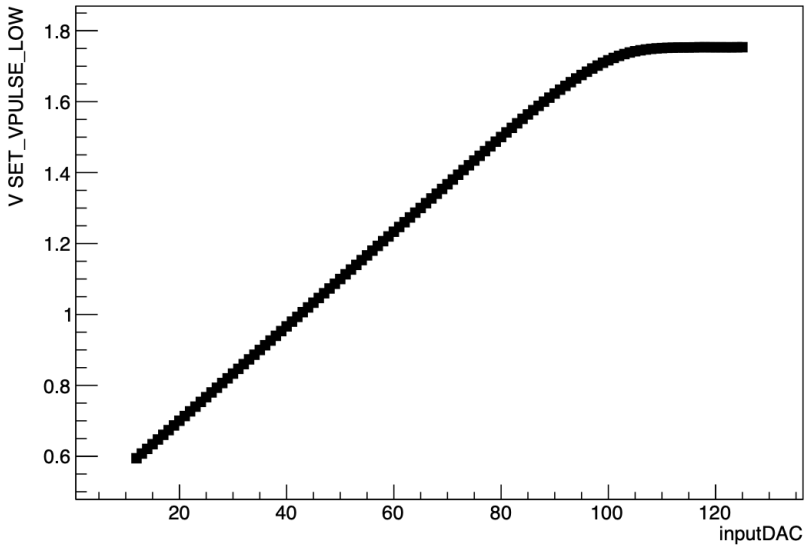


Figure 5.8: VLOW dac linearity scan for a MALTA2 device.

In Figure 5.9 is a plot of an analogue scan. The procedure is accomplished by injecting 2000 electrons, which is a very high charge with respect to the minimum detectable threshold by the detector, 50 times into each pixel and reading if the FE responds to it. By sending such a high charge, you ensure that there will be no noise. If a pixel does not see any injections then it can be declared dead. In the picture shown the full matrix responded as expected to all the injections. Note that the rows start at 288 as the matrix has been reduced from the old MALTA one.

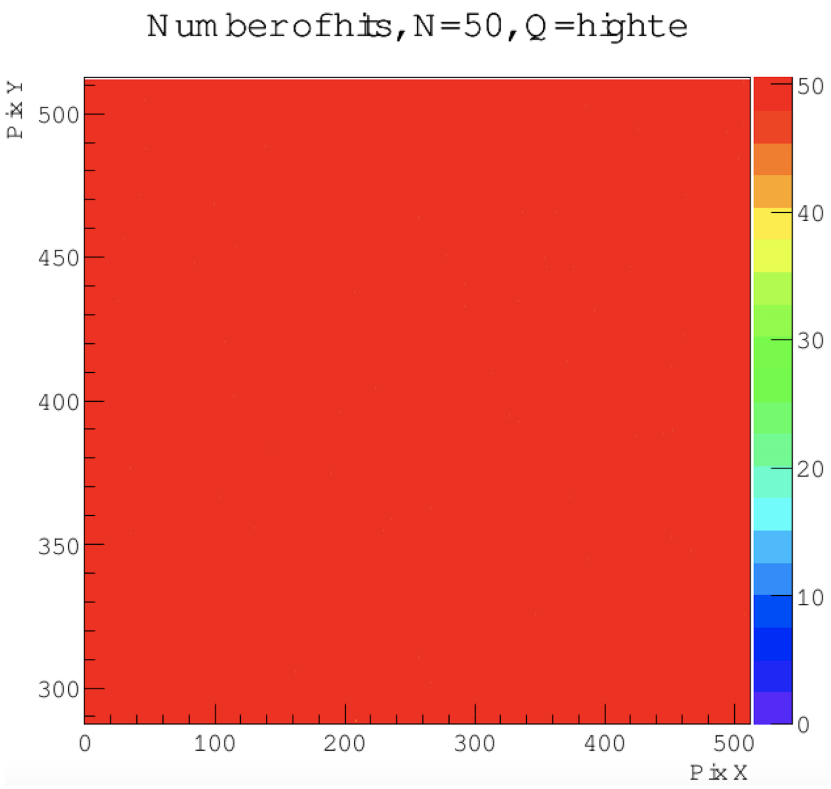
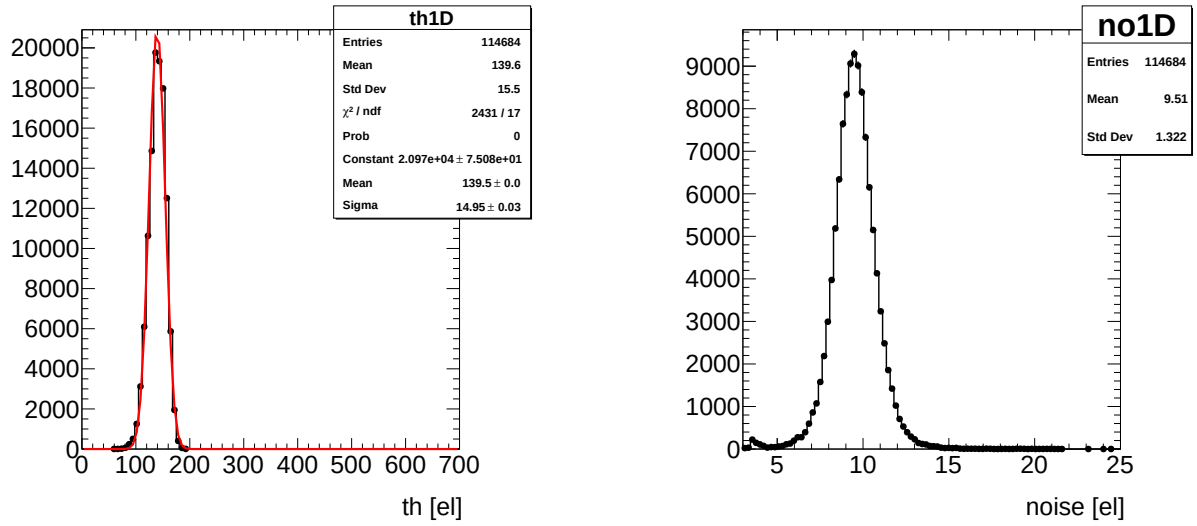


Figure 5.9: Analogue scan for a MALTA2 device. The injection charge is 2000 electrons and each pixels is injected 50 times.

Finally Figure 5.10 is a threshold scan comparison between an epitaxial with gap in the n-well process modification MALTA and MALTA2 devices. The DACs have been set to

threshold scan, carried out in the way described in Section 2.6.4 for a device fabricated with an epitaxial substrate. The graphs show that the MALTA2 matrix responds to any charge larger than 140 electrons and it has a dispersion of 16 electrons. The noise is around 10 electrons with a deviation of 1. Certainly the biasing of the matrix should be optimised for the different applications but this result is a big step forward in the MALTA devices as it is the first time that a threshold scan of a full matrix has been run. Previously due to the clock tree issue the time needed to change configuration was too long to be able to scan every pixel. It should be noted that 4 pixels were masked in this threshold scan and that the small peak observer in the noise plot around 0 electrons is due to the fact that the s-curve is sharper than the charge injection step so the fitting gives a small noise value. Statistically the probability of not getting the s-curve increases as the noise decreases.



(a) *(b)*

Figure 5.10: (a) Threshold scan for a MALTA2 device with the following dac settings IBIAS 43, IDB 50, ICASN 10, VCASN 110, VRESET_P 29, VRESET_D 65 and ITHRESHOLD 15 (b) Noise estimation from the threshold scan for the same settings.

5.4 Conclusions and future of MALTA family monolithic active pixels detectors

MALTA2 modifications, specially the new slow control design which was big part of the work of this thesis, constitute a vast improvement with respect to the original MALTA device. New possibilities for further characterisation of the upgraded cascoded front-end arise with the availability of performing threshold scans of the full matrix. The device has been fabricated in both epitaxial and Czochralski materials with the extra deep p-well and gap in the well processes described in Chapter 3. The new front-end has proven to be more robust against noise and therefore it can be operated at lower thresholds. The next step consists of an in-depth characterisation of the sensor to determine its radiation hardness towards NIEL and TID and its timing behaviour. Some wafers have already been sent for irradiation and a similar analysis of efficiency, in-time efficiency and cluster size will be performed in MALTA2.

Chapter 6

ITkPixV1 Single Event Effects characterisation

As it has been introduced in Section 2.5.3, Single Event Effects (SEE) are a type of radiation induced effect caused by single particles of three different types: high energy hadrons, intermediate energy neutrons and thermal neutrons traversing the readout electronics of a detector. Due to the fact that their effects become more damaging with the scale down of technologies and they are independent of the other types of radiation like TID and silicon bulk defects, it is crucial to determine the resistance of the chips to these effects. Careful designs including SEE mitigation techniques are made and several testbeam campaigns are needed in order to fully characterise how well the mitigation works. This chapter will describe the SEE protection implemented in the ITkPixV1, which is the prototype of the chip that will go in the ATLAS pixel tracker Upgrade, in order to minimise SEE and the results obtained during testbeams with heavy ions.

6.1 ITkPixV1 mitigation techniques

ITkPixV1 is the full scale prototype readout chip for the ATLAS ITK Upgrade. It is approximately 2 x 2 cm^2, and has a pixel matrix of 400 columns and 384 rows. It has been fabricated by Taiwan Semiconductor Manufacturing Company (TSMC) in their 65 nm technology. The sensor will be bonded to the chip with bump bonds, making a hybrid pixel detector as described in Section 2.3. It has been designed to minimise the impact of SEEs and this section will present all the mitigation mechanisms in place.

Due to power and area limitations not every single block that constitutes the chip can be protected in a similar manner. Therefore, different choices of protection are implemented in different areas of the chip.

In ITkPixV1 there are different techniques or levels of triplication protection in place to mitigate SEE:

- Triplicated registers with triplicated skewed clocks. This is the highest level of protection available. It has been used for the Global Configuration Registers, as their functionality is critical. They define the operational condition of the chip and must remain reliable for a long time. Furthermore, it is not possible to recover their normal state with a simple reset. This type of mitigation has also been done for all the state machines in the readout chain that control the output data. Figure 6.1 shows a schematic of how this technique is implemented. The output of this logic block is voted which means that if the three outputs of the corresponding logic blocks are not the same, the voter output will be the two that are equal. The voter output is then fed to the registers that will latch the signal with a clock of a similar frequency but different phase. The clock skew allows for correction in the case of SET glitches shorter than ΔT which is set to 250 ps according to the SET prototype test results [115]. The triplicated clock skew is made by clock tree synthesis and the timing closure becomes difficult to obtain since the delay ΔT depends on the process, supply voltage, temperature and TID. It should be mentioned that TMR is inserted during synthesis and single cells are replaced with the triplicated version during the synthesis step. The redundant cells are constrained to be placed with 15 μm^2 distance so that one particle cannot affect multiple cells storing the same redundant information. The output of the registers will be then compared with the feedback mechanism against the input to check for potential errors in the path and correct them. Thus this is the maximum level of protection achievable by design. It is not perfect, because if the voter suffers from SEE the errors will be propagated but the conditions to get an SEE when this type of mechanism is in place are statistically very small.

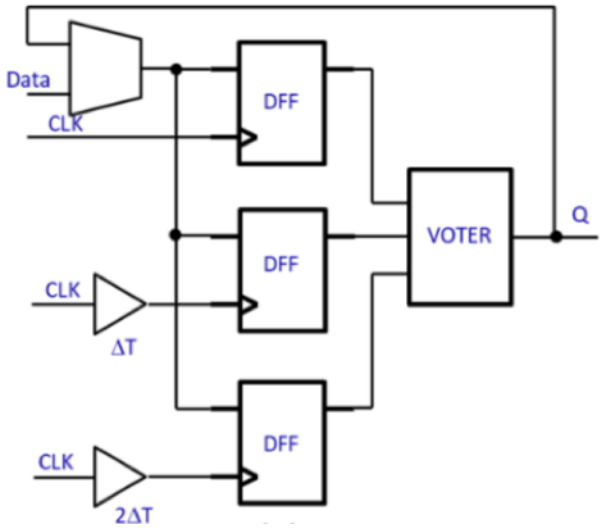


Figure 6.1: Schematic of a full triplication protection with clock skews TMR technique [118].

- Triplicated registers without auto-correction. This is a level of protection lower than the

previous one, as it does not account for SET protection. Each pixel has 8 bits that can be configured and this type of protection has been implemented in 6 of those bits, the most critical ones corresponding to enabling the pixel and adjusting the threshold due to routing congestion and area limitation inside the matrix. The schematic of this TMR mechanism is shown in Figure 6.2. For this architecture any error in the voter will be seen at the output.

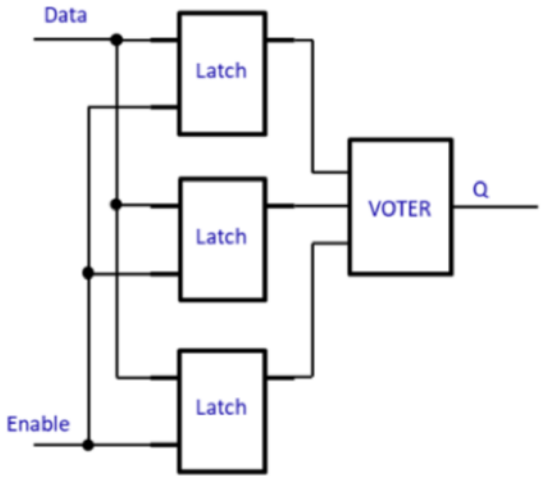


Figure 6.2: Schematic of a simple triplication protection [118].

- Other protections. The previous mechanisms presented target the digital logic of the chip but other custom blocks have been designed to avoid SEE. A key SEU protection technique is to avoid the use of asynchronous reset as an error in this line would imply reloading all the configuration. Custom designs for the PLL [116] as well as the Shunt LDO [117] that controls serial powering have been optimised for SEU/SET immunity by including bulk contacts of both NMOS and PMOS transistor close to the drain-source or using oxide trenches to isolate the bulks. The resistance of this specific technology, TSMC 65 nm, had been previously tested against latch-up showing no sign of it. Finally the non critical hit or event data was decided not to have any protection after extensive SEU injection simulations.

The implementation of TMR measurements has consequences. Depending on the level of protection, design resources like power, area or speed are affected. The maximum triplication level or full TMR implies triplication of the following structures: flip-flops, combinatorial logic, voters and clocks and therefore the area and power consumption will significantly increase. On top of that, the voters add a delay to the clock tree that needs to be taken into account in order to meet timing constraints. In the case of triplicated registers without auto-correction only the flip-flops are triplicated and the voter delay will have to be considered but the level of protection is several orders of magnitude worse than for the full TMR. At the end, it becomes a compromise

between the requirements and the design resources.

In consequence there is another outstanding factor that will determine the behaviour of the chip to SEE which is its recovery capabilities. To start with, the architecture should be robust enough to ensure that in the case of getting SEU the events will be removed quickly and will not affect other hits. This translates to having a Data Acquisition System (DAQ) that ensures continuous reconfiguration fast buffer clear and dedicated resets, like a power reset that will not imply a full power cycling of the chip.

6.2 ITkPixV1 SEU cross-section measurements

In order to estimate the cross section of the ITkPixV1 three different testbeam campaigns have been done: two of them with high energy ions and one with a proton beam. The dates of the campaigns were the 18th of October, 17th of November and 19th and 20th of December of 2020. The two first ones took place in the Centre de Ressources du Cyclotron (CRC) in Louvain-la-Neuve [97], Belgium, and the last one was run remotely with the help of local people from TRIUMF facility in Canada [98]. All the testbeam work presented is part of a collaboration between CPPM (Centre de Physique des Particules de Marseille) and CERN.

The expected High Energy (> 10 MeV) Hadrons (HEH) rates during the HL-LHC in the pixel detector for the inner layers, situated at a radius of $\approx$ 30 mm from the interaction point, go from a 100 MHz / cm^2 to 1 GHz / cm^2. For the further layers of the detector the estimated rate scales with $\approx$ 1 / r^2. In consequence, for the middle layers the expected rates go from 10 MHz / cm^2 to 100 MHz / cm^2. All the numbers quoted above have an uncertainty factor from 10 to 100.

The typical 65 nm latch, flip-flop or memory cross-section has been measured with 400 MeV protons and heavy ions to $\approx$ 1 x 10^{-14} cm ^{2} / bit [99]. Therefore taking into account the numbers from the paragraph above, the SEU rate for a unprotected memory structure is:

$$1\ GHz/cm^2\ \times\ 1\ \times\ 10^{-14}\ cm^2/bit = \ 1\ \times\ 10^{-5}\ Hz/bit \tag{6.1}$$

This is equivalent to having every latch flipping once a day. For the outer layers, they will have SEUs once every 100 days.

The results obtained from high energy ion measurements in Louvain are then scaled to the High Energy Hadron spectrum typical of the LHC. The results from the testbeam at TRIUMF done with 400 MeV protons and a flux of 1.5 x 10^9 particles / cm^2 s = 1.5 x 10^9 particles / cm^2 s are similar to the conditions expected in the inner layer of the pixel detector. The conclusions from the three campaigns will determine if more effort needs to be put into protection in the final design of the chip or if the current mitigation strategy is good enough.

Before starting with the characterisation results it is important to highlight that the majority

of the measurements have been done in bypass mode, which is not the expected operating mode of the chip. In normal operation conditions, inside the chip there is a PLL [95] that receives commands from the firmware at 160 Mbps and it is in charge of generating two clocks from them, the command one that allows for communication and configuration of the chip and works at 160 MHz and the serialiaser clock that operates at 1.28 GHz. Previous work has shown that the PLL and in general the clock and data recovery block are extremely sensitive to SEU [116], especially for high LET ions, and because the testing time is limited, in order to maximise data taking these clocks will be supplied from an external source. This is what is referred to as bypass mode and it was done in this way because the priority is to check how the memories in the global configuration registers and in the pixels respond to SEE and not the PLL. When time allowed some additional tests were as well done in order to determine the performance of the Clock and Data Recovery but, as will be shown, the results are not conclusive and more measurements will be necessary to determine if they are robust enough against SEU.

There are also several available possibilities to power the chip. For all cases the currents and voltages were monitored all the time during the beam measurements and the chip was kept at room temperature. For the two tests in Louvain the chip was powered in "direct mode" providing directly the required 1.2 V for the analogue and digital domains, instead of using the sLDO. For the tests in TRIUMF the setup was run in low-dropout regulator (LDO) mode that will provide the 1.2 V for the two different domains from a 0.6-0.8V input supply.

An issue was found in the design of the ToT memory cells that uses custom multi-bit latches. If a certain ToT value is registered it results in a short circuit between the VDD and ground leading to high digital currents of 2A [96]. For the SEU tests the chip was run with configurations that ensured that the digital current was kept to the correct value and the current was monitored.

6.2.1 Tests performed

The tests that were performed during the three testbeam campaigns will be explained in this section. As discussed before, the basic idea is to monitor how many SEU are observed in the different memories of the chip. The measurements performed were:

- Global Configuration Register Test. Global configuration registers define the operational condition of the chip. They have been implemented with full TMR protection. In total there are 209 with 2894 configurable bits. The test routine consists of writing the "default values" for the chip outside the beam and reading them back to make sure that the desired value has been correctly stored. Then the chip is placed inside the beam until the desired fluence has been reached. After that, the beam is switched off and the global configuration registers values are read and compared to check that they have the same value that was written outside the beam. A further offline analysis indicates which registers have suffered an SEU and how many bits were flipped. The effect of several parameters like: power

supply, angle of the chip with respect to the beam and the presence of clock will be studied.

- Pixel Configuration Register Test. In the ITkPixV1 each pixel is configured with 8 bits, 6 of which are protected using triplication without auto-correction and 2 of them are standard without SEU protection as a bit flip on them will not be critical for the overall chip behaviour. Figure 6.3 shows a table with the functionality of the 8 in pixel registers and it specifies the ones that have TMR protection. The test procedure consists of writing the configuration outside the beam of the 400 x 384 x 8 = 1228200 bits. Usually the chosen bit pattern is 01010101 so it can be studied if a flip from 0 to 1 or from 1 to 0 is more likely. Then the beam is switched ON until the desired total fluence is reached and no communication is made with the chip. After, the beam is turned OFF, communication with the chip is reestablished and the pixel configuration is read back. Finally the offline analysis will produce an SEU pixel map, the distribution of SEUs among the 8 bits, the differences between SEUs that produce a bit flip from 0 to 1 and 1 to 0 and also the number of bit flips in the TMR bits versus errors in the standard ones.

Bits	Name	Description
[0]	Enable	Include the pixel in the DAQ data path
[1]	Cal Enable	Turn on charge injection (*)
[2]	HitOr Enable	Add the pixel to its wired OR core col. hit line
[3:6]	TDAC value	Value for in-pixel theshold trim DAC
[7]	TDAC sign	Selects differential branch set to TDAC value

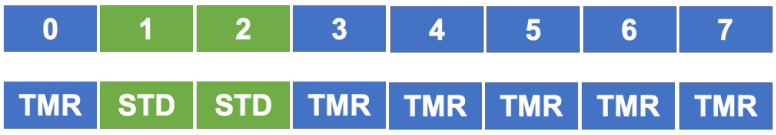


Figure 6.3: Pixel register bits functionality [121] and type of SEU mitigation in place for the pixel registers. Note that the TMR is of the type without auto-correction.

- Digital scan test. This is run continuously with the beam ON until the desired fluence is reached. The procedure consists of injecting every single pixel with a digital pulse 100 times. If the scan has an ideal result, the analysis should show that all the pixels have received 100 injections. However if there were SEUs, noise, etc. some pixels will store a different number. This test is performed in both bypass and PLL mode to have an idea of the robustness of the clock and data recovery block of the chip and of the DAQ. It is important to note that the scan is a digital scan, because the injection pulse is sent via a digital pulser circuit bypassing the analogue front end of the chip. This ensures that all the pixels receive the same signal.

- Noise scan test. The threshold of the chip is set very high, 6000 electrons, so in theory it should not be affected by noise. The chip is kept in the beam during the desired total

fluence while a loop of a scans without injection is carried out. After every scan, triggers are sent so for each iteration the number of pixels (and their locations) that have received a hit is recorded. The analysis will give us the noise rate in the testbeam environment.

The goal of all these tests is to compute the SEU cross section of the different memory elements in the chip. For all the plots that will be shown in this chapter, the cross sections have been calculated with the following formula:

$$Cross\_Section = \frac{Tot\_errors}{Num\_config\_bits \times \Phi} \tag{6.2}$$

where Tot_errors is the total number of bits that have flipped during each specific test, Num_config_bits is the total number of bits that were configured to a specific value, this number is over a million for the pixels registers and ≈ 2000 for the global configuration registers, and Φ is the fluence expressed in ions per cm^2.

6.2.2 Louvain-la-Neuve testbeams

Facility and setup description

These testbeam campaigns took place in the Centre de Ressources du Cyclotron (CRC) [124]. The facility provides a beam with a diameter of 2.5 cm ($\pm$ 10% uniform inside the circle). The chip was placed to completely cover the global configuration registers and most of the pixel matrix. The test set-up consists of the single chip card with the ITkPixV1 and the DAQ system BDAQ53 [113] that were connected with 2 meter cable of Display Port (DP) to SMA in the control room and 60 cm cable SMA to DP in beam area. The break in the cabling is needed because there is a patch panel in the facility that has SMA connections. Figure 6.4 is a picture of the single chip card in the beamline during the first campaign situated in the stage where the DP connections and the power cables of the chip can be seen.

The facility allows the use of several ions with different LET. Table 6.1 collects all the information with respect to the ions used, the LET, the fluence reached for both the pixel configuration tests and the global configuration tests, as well as the power settings and orientation of the chip during the first campaign.

The information for the second testbeam campaign can be seen in Table 6.2. For this campaign there is an extra column, the clock in GC that stands for global configuration, because as it has been explained the global configuration registers have the clock skewed and auto-correction for SEU protection while this is not implemented in the pixel registers. Therefore, if the clock is switched off during the global configuration register test the level of protection becomes the same as for the pixel registers with triplication and the SEU cross sections should be similar instead of differing by a couple of orders of magnitude while the clock is running.

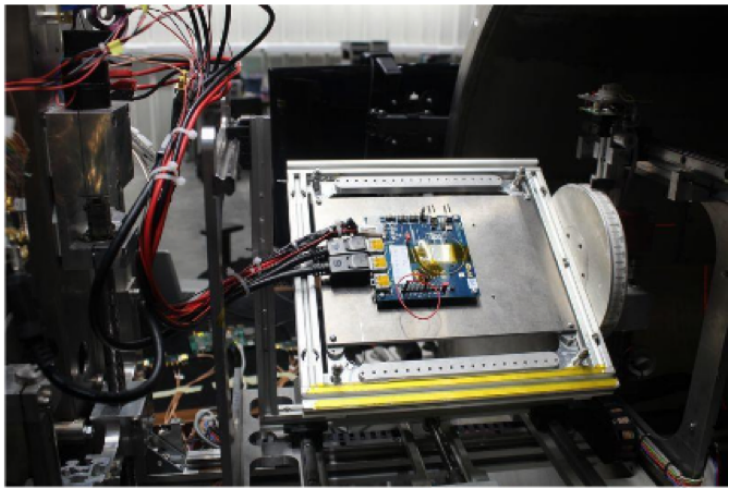


Figure 6.4: ITkPixV1 testing setup placed in the beam facility in Louvain-la-Neuve CRC.

Table 6.1: Ions, LET, fluence reached and chip power and orientation with respect to the beam table for the tests performed during the first testbeam campaign in Louvain.

Ion	LET[MeV/(mg/cm²)]	Pixel Configuration Fluence [ions/cm²]	Global Configuration Fluence [ions/cm²]	VDD [V]	Tilt [degrees]
Carbon	1.3	5×10^5	Not tested	1.2	0
Neon	3.3	5×10^5	1×10^6	1.2	0
Aluminum	5.73	5×10^5	1×10^7	1.2	0
Argon	9.9	2×10^5	2×10^7	1.2	0
Chromium	16.13	2×10^5	5×10^6	1.2	80
Krypton	32.4	5×10^5	Not tested	1.2	0
Xenon	62.5	1×10^5	5×10^6	1.2	0

Table 6.2: Ions, LET, fluence reached, and chip power and orientation with respect to the beam table for the tests performed during the second testbeam campaign in Louvain.

Ion	LET[MeV/(mg/cm²)]	Pixel Configuration Fluence [ions/cm²]	Global Configuration Fluence [ions/cm²]	Clock in GC	VDD [V]	Tilt [degrees]
Carbon	1.3	1×10^7	Not tested	No	1.2	0
Neon	3.3	1×10^7	Not tested	Yes	1.2	0
Aluminum	5.73	5×10^7	Not tested	Yes	1.2	0
Argon	9.9	3.02×10^7	Not tested	Yes	1.2	0
Chromium	16.13	3×10^7	Not tested	Yes	1.2	0
Krypton	32.4	1×10^7	Not tested	Yes	1.2	0
Krypton	32.4	5×10^5	1×10^7	Yes	0.9	0
Krypton	32.4	Not tested	1×10^7	No	0.9	0
Krypton	32.4	5×10^5	5×10^6	Yes	0.9	80

Power consumption

To begin with an analysis of the power consumption of the chip will be done. Figure 6.5 is a plot of the digital current drawn by the chip during the irradiation. As noted in section 6.2, there is a design issue that means when a specific value is set in the TOT memory a short circuit between GND and the power supply is created and the digital current becomes very high and has to be constantly decreased. This value is avoided as much as possible. What can be observed in the graph is that during the configuration of the chip in the digital scans, the current goes high. Therefore a chip operating script called "decrease current" was created to set the value of the corrupted ToT cell to the one that avoids the short circuit. Once invoked the 700 mA baseline is recovered.

The global configuration register scan has the longest duration to accumulate more statistics because the number of memory bits available is much smaller than for the pixel register, therefore the increase in the current is much larger, going above 2.5 A. The cause of this increase is that ions are hitting the pixel matrix while running the global configuration scan and the TOT cells record different values so eventually there will be a short circuit. For the pixel register test, as for the digital test, once the script "decrease current" is run the current baseline is reset to 700 mA. The pixel configuration test does not have any influence on the TOT memories, as they are configured to a working value and remain unchanged, or on the matrix activity so the current is low and it even goes below 500 mA before increasing back to the 700 mA baseline after running the script. Finally during the noise scans there is another increase in the current with time because as noise accumulates the TOT memories can end up in the states which result in the short circuit. This proves that despite the extra difficulties due to the design issue the digital current can be kept under control during testbeam campaigns.

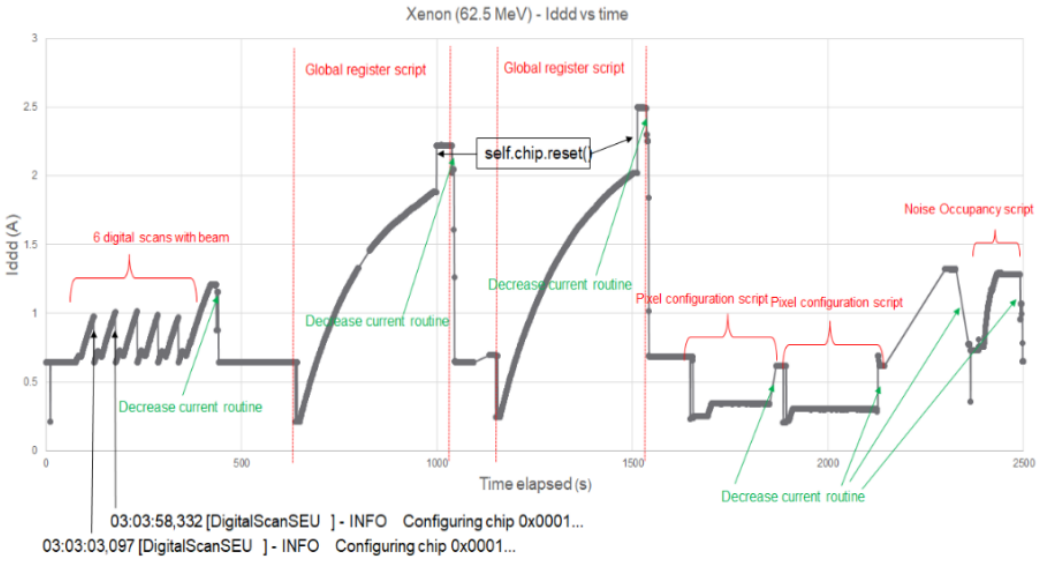


Figure 6.5: Digital current monitoring while performing SEU tests with Xenon

Pixel configuration characterisation results

A compilation of the pixel configuration register test results for both campaigns is shown in Figure 6.6 . It is a series of histograms per ion which include, the total number of SEU that provoke a bit flip from '1' to '0' and vice-versa. The total number of SEUs recorded and how many happened in the two standard pixel registers without protection and how many in the triplicated ones. The first observation is that the number of errors depends on the LET of the ion, the higher the LET the more errors observed. Secondly, for all the ions the number of SEU that result in a bit flip from 1 to 0 and from 0 to 1 is very similar so it can be concluded that there is no difference between flips from 0 to 1 and 1 to 0. For the case of the carbon, the plot can be a bit misleading but the number of errors is very low to draw any different conclusion. The more noticeable aspect of this plot is that the TMR registers have considerably less flips than the unprotected ones, demonstrating the impact of the SEU mitigation.

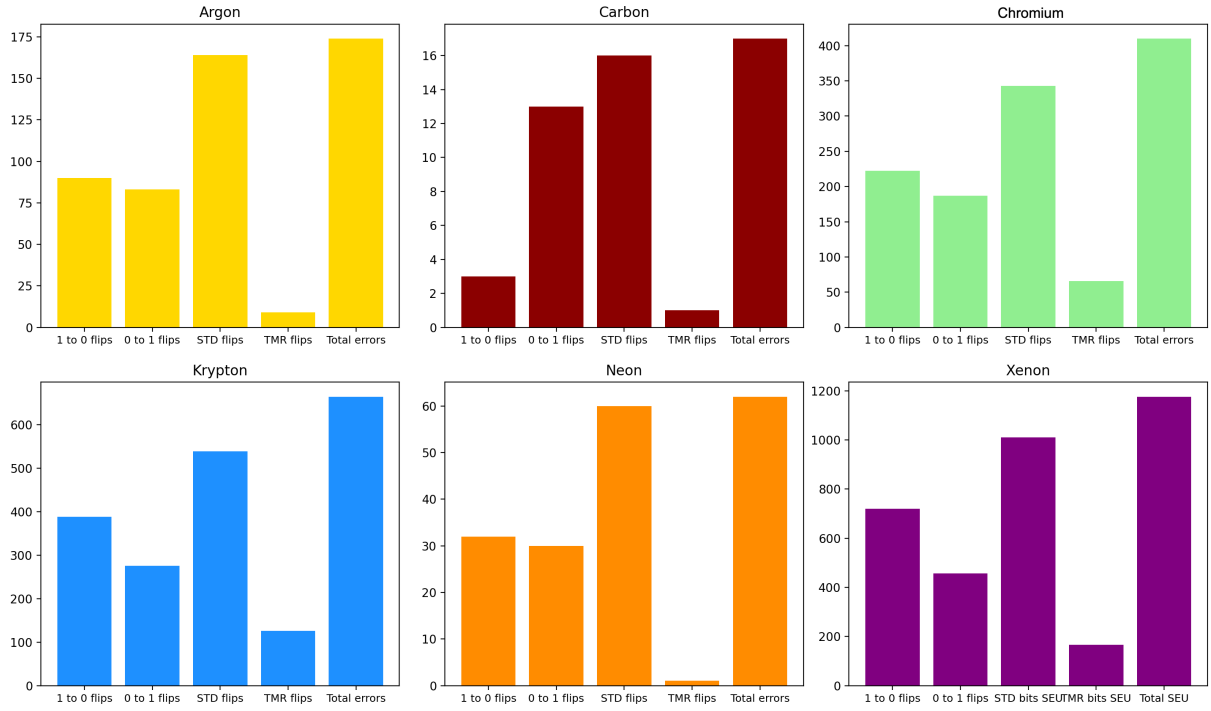


Figure 6.6: SEUs measured in the pixel registers for each ion. Each plot shows the number of SEU that induce a '0' to '1' flip, a '1' to '0' flip, total number of SEU, number of SEUs in non protected registers and in triplicated registers.

Another observation, after analysing the data for both campaigns, is that for the pixel configuration registers that have the triplication show an increasing number of SEUs with increasing fluence. This is expected because these bits have no clock refreshing and so the errors accumulate with fluence. Figure 6.7 shows the number of errors for Neon for two different fluences, 5×10^5 $ions/cm^2$ and 1×10^7 $ions/cm^2$. The number of errors measured in the triplicated bits of the pixel configuration is 30% higher for the 1×10^7 $ions/cm^2$ run than for the 5×10^5 $ions/cm^2$.

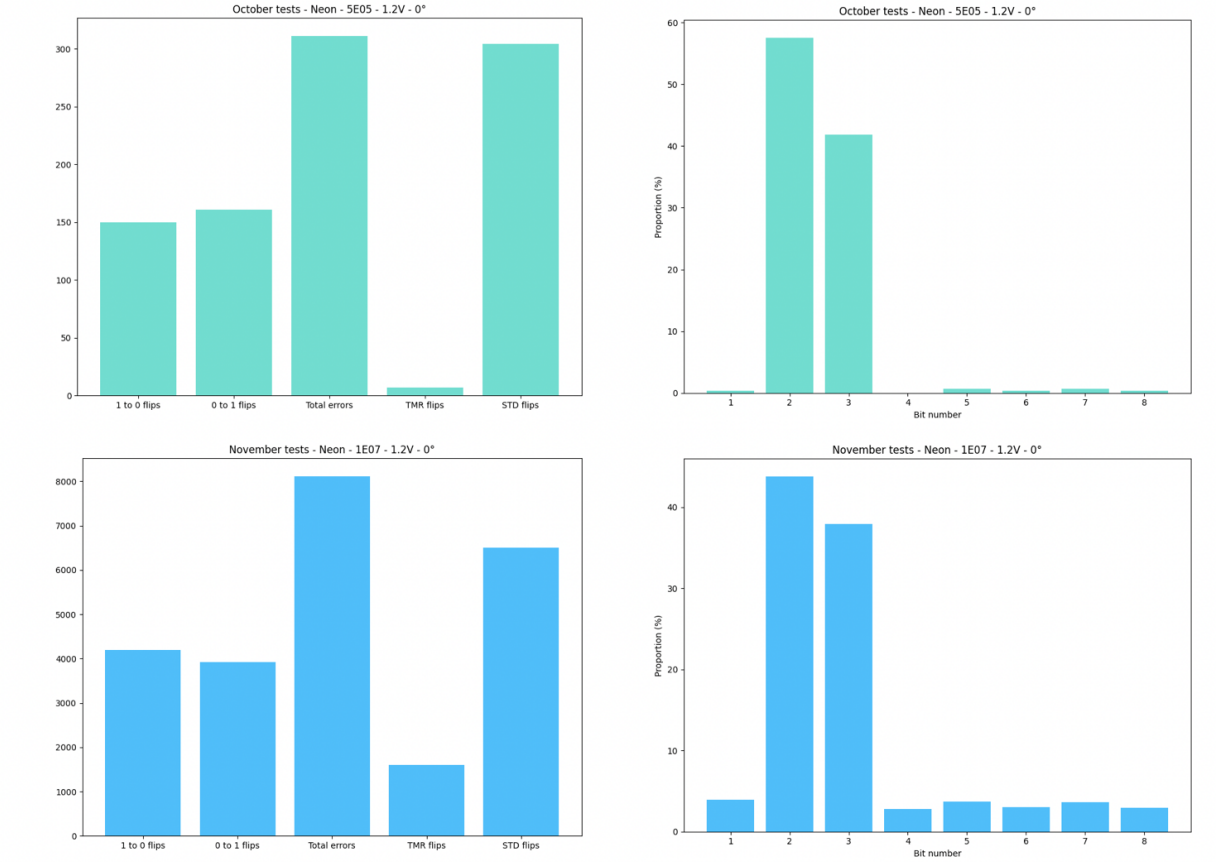


Figure 6.7: On the left, histogram for Neon ion that covers the number of SEU that induce a '0' to '1' flip, a '1' to '0' flip, total number of SEU, number of SEUs in non protected registers and in triplicated registers. On the right the fraction of bit flips per register. The top two histograms show the results for total fluence of 5 x 10^5 ions/cm^2 and the bottom histogram shows the results for a total fluence of 1 x 10^7 ions/cm^2 fluence.

Finally, the estimated cross sections for both campaigns were calculated using equation 6.2. The results are shown in Figure 6.8. There is at least one order of magnitude between the curves corresponding to the SEU cross section of the triplicated pixel registers (in blue) and the standard pixel registers (in yellow). This demonstrates that the TMR techniques have at least some some mitigating effect even without the clock refresh. The measured cross sections go from $\approx 10^{-11}$ to $\approx 10^{-9}$ for the TMR bits and from 8 x 10^{-10} to 7 x 10^{-8} for the standard bits. The error bars have been estimated with a chi-square distribution, $\tilde{\chi}^2$, with a confidence level of 95% and the a total number of degrees of freedom equal to 2 times the total number of SEUs per test, following the methods of [120].

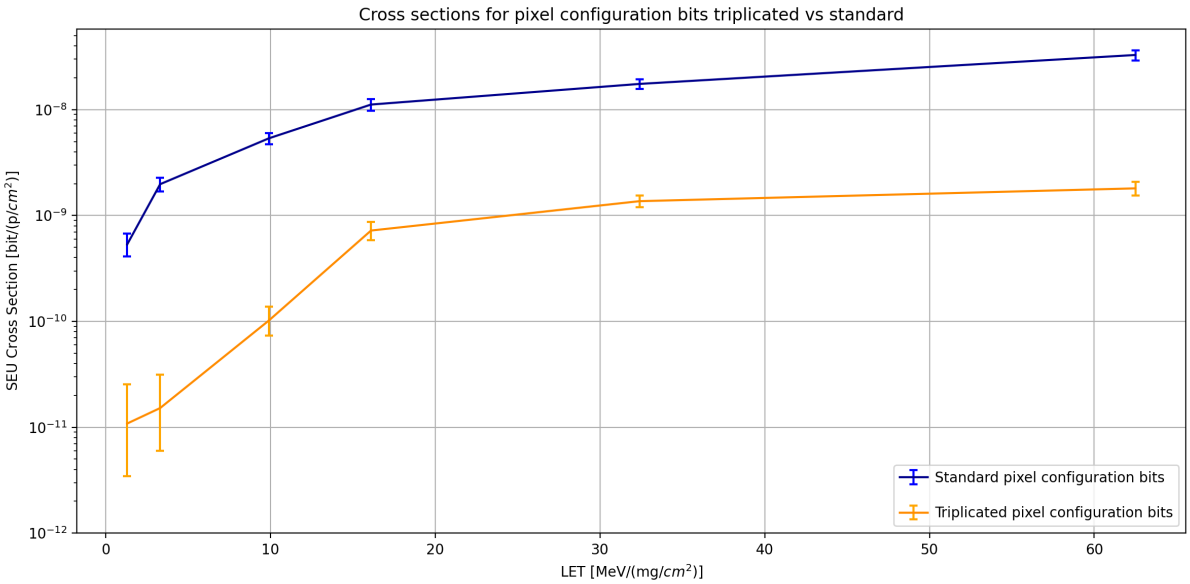


Figure 6.8: Cross section of pixel configuration registers obtained in both testbeam campaigns in Louvain for the different ions, normalised to its fluence. The blue points are the cross-sections measured in the standard registers and orange points are the cross-sections for the registers with TMR.

Global configuration characterisation results

The next results will examine the global configuration register tests and results. The main difference with respect to the pixel register tests is that these registers receive the clock and have auto-correction if a bit flip is identified. They are 16 bits wide and when SEUs are identified in a register, all the bits are checked for possible flips. In total, there are $\approx$ 2000 global configuration bits. Due to the much smaller number of bits for the global configuration tests than for the pixel registers tests and their higher protection level, the duration of the tests is considerably longer. Figure 6.9 is a histogram that shows how many registers were upset and how many flips happened per register. The observed trend shows that in $\approx$ 95% of the cases the registers errors are due only to a single bit flip. In addition, the registers with upsets are completely random and vary from scan to scan which is the expected behaviour for these tests. The fluence for each ion is the one specified in Table 6.2. It can be pointed out that for the global configurations registers the number of errors seems to be less dependent on the fluence than for the pixel registers due to the different protection mechanism.

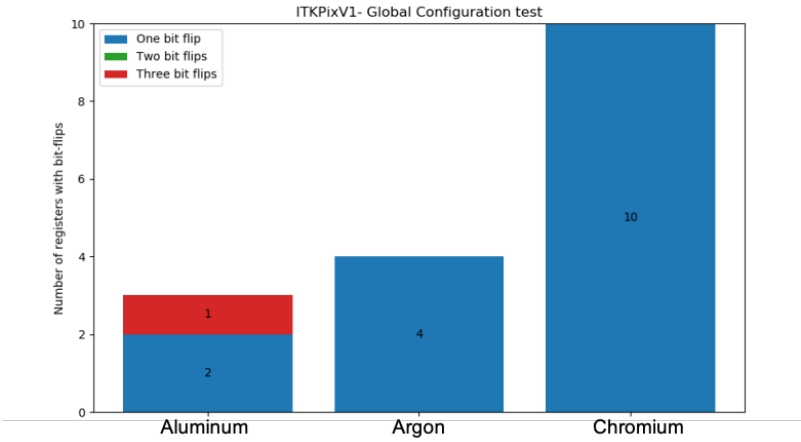


Figure 6.9: Total number of register with errors and number of bit-flips per register for medium LET ions.

The cross-section of the global configuration bits has also been calculated according to the equation 6.2. Figure 6.10 is a plot with the measured cross section for all the ions with the exception of Carbon as the error rate is too low. The error bars have been calculated in the same way as for the pixel configuration registers, but the statistics are much smaller and therefore they are larger. Note that in the case of the Neon, no errors were observed during the scan so a cross-section could not be determined but the range is given by the error bar. The cross-section increases with the LET, and it goes from $\approx 10^{-11}$ for Neon to ≈ 4 x 10^{-10} for Xenon. In general, this is an order of magnitude smaller than for the triplicated bits of the pixel configuration registers that do not have a clock refresh.

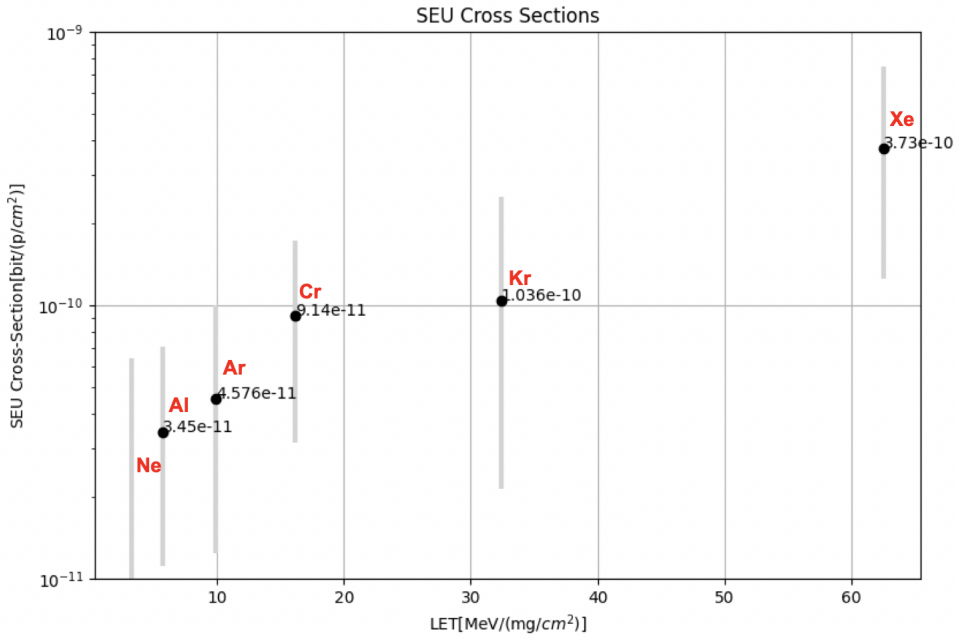


Figure 6.10: Cross section of global configuration registers obtained in both testbeam campaigns in Louvain for the different ions, normalised to its fluence.

The next tests seek to explain the influence of other parameters on the chip. They were done with the Krypton ions. The idea is to calculate the different cross-sections of the global configuration registers while powering the chip at its nominal value of 1.2 V and at its lower operating limit of 0.9 V. Check that when the clock is not provided to the global configuration registers that have auto-correction, they behave like the pixel registers with triplication as this is the main difference between both mitigation strategies. Finally the influence of tilting the chip 80 degrees with respect to the beam axis was analysed. This is motivated by the fact that the angle of incidence of the particles on the chip varies across the detector. Figure 6.11 is a representation of how the chip is placed when straight or tilted. In Figure 6.12 the estimated cross section can be seen for all the tests performed in which these parameters are studied. Firstly, the influence of the powering settings will be analysed and it can be seen that the cross-section increases as the voltage decreases by a factor 3, this is marked in the plot as "Influence of VDDD". This is expected as by operating at lower power, transistors get slower having a reduced current capability which makes more likely that a deposited charge will induce an SEU. Secondly, the effect of the tilt is analysed and even if the difference is not high, the ratio is 1.7, the chip suffers less SEUs when placed perpendicular to the beam than when it is tilted 80 degrees, it is represented as "Influence of the tilt" in Figure 6.12. This is because even though the registers have a smaller effective cross section area the track length of an ion in a register is increased. Lastly, the verification that if the clock is not provided to the global configuration registers they should behave like the in-pixel registers was done. It is important to note that the chip was powered with 0.9 V in the digital but the result shows a very similar cross section to the one obtained in the pixel registers for the same ion, $1.37e{-}91$ for the pixel registers versus $1.19e{-}9$ for the configuration ones without clock. There is a factor of 4 gain in protection when the clock is triplicated, shifted with respect to each other and enabled. These results prove that the understanding of the different levels of protection is good and the behaviour with respect to the different parameters follows the expected trend. The gains of the different triplication protections have been quantified and now it is a matter of deciding if the numbers are good enough for when the chip will be operating in the ATLAS Upgrade experiment after the LS3.

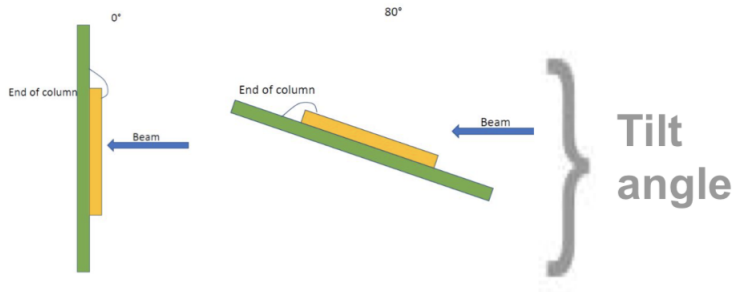


Figure 6.11: Graphical representation of the tilt of the chip with respect to the beam as it was done during testbeam campaigns in Louvain.

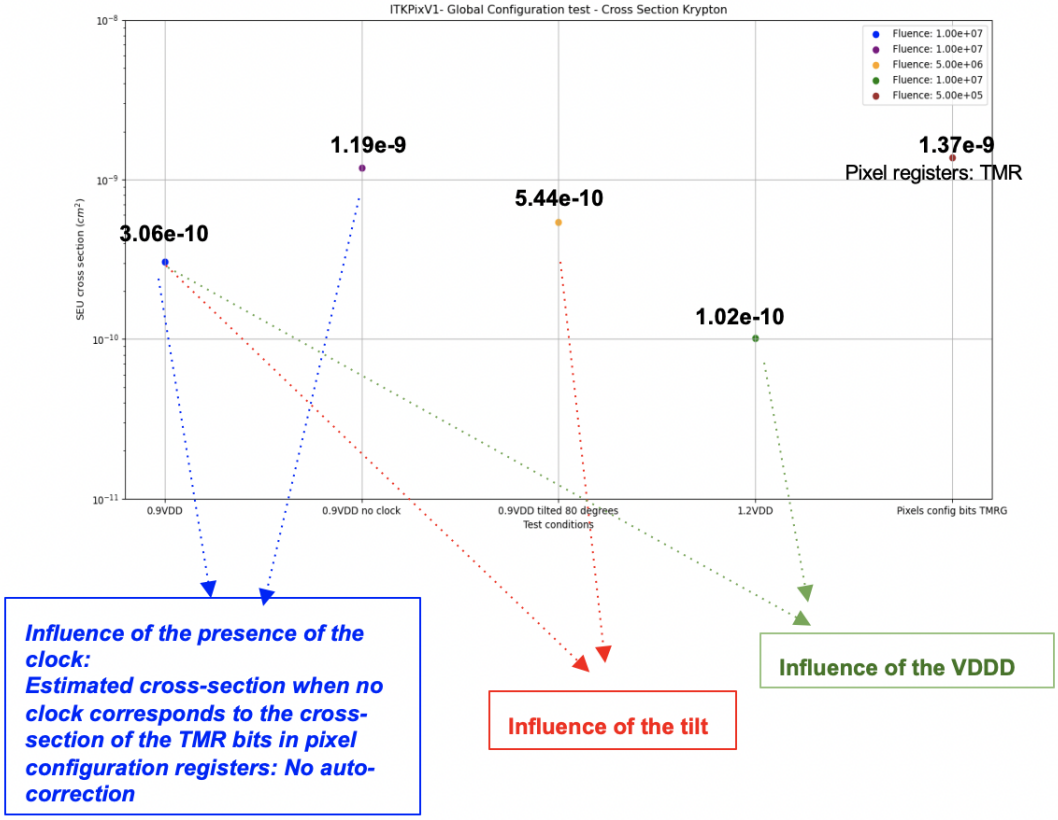


Figure 6.12: Calculated cross section of ITkPixV1 for Krypton ion under the influence of different parameters: tilt, power and clock. Error bars are not shown because the ratios are not significant.

Digital tests characterisation results

Finally the results of the digital tests will be analysed. As a brief reminder the test procedure consists of enabling two neighbouring pixels in every second core column (belonging to the same pixel region) and performing 100 digital injections per pixel. The analogue FE is bypassed and the TOT is set to code 0 in order to avoid the large current issue. This scan is continuously running until the desired fluence is reached. For this test it is very important to distinguish between the results obtained while operating the chip in bypass mode and the behaviour of the setup while running it in PLL mode. In all cases the chip was operated at 1.2 V supply for both analogue and digital power. Focusing on the scans done in bypass mode the results can be classified in three different categories: out of the 100 injections per pixel some of them will receive more or less hits than this number, when there is an "extremely" noisy pixel that shows more than 10000 injections or there are some dead pixels that don't register any hits at all. If we consider the dimension of the chip the total expected number of injections in a clean scan digital scan is:

$$Tototalinjections = 384 \times 400 \times 100 = 1536000 \tag{6.3}$$

The digital scans were performed for four different ions: Carbon, Neon, Argon and Chromium. The ones with lower LET were chosen for this scan with the idea of trying some runs in PLL mode. Previous testing had shown difficulties on running the chip in this mode with very high LET ions [116]. The histogram seen in Figure 6.13 has the percentage of the total scans that fall into each of the categories. The analysis shows that the most likely output is that the majority of the pixels get the correct number of injections and a small fraction of them record extra or fewer injections than expected, this is shown in the plot as Category 1. This is something normal considering the large amount of pixels in the matrix and factors like noise and readout could be the reason for this to happen. The other two categories appear for higher LET ions than carbon. It should also be noted that the noisy pixels and dead pixels seem to be random as they are not the same ones for the different scans.

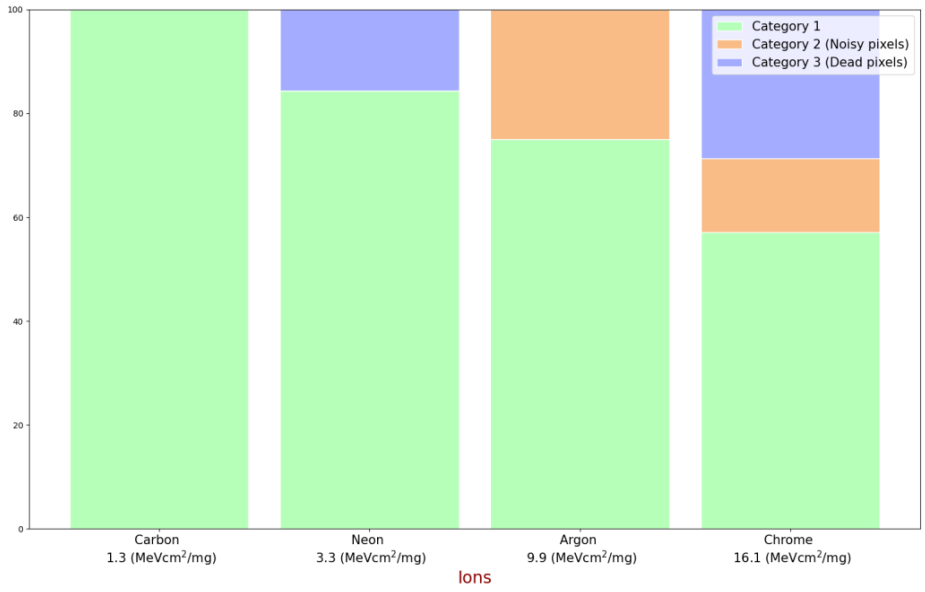


Figure 6.13: Histogram with the percentage of digital scans that fall into each category per ion when running in bypass mode.

If the total number of upsets for all the runs and ions is examined, there is a clear region that is more affected by SEUs than the rest of the pixel matrix. It corresponds to the 8 last rows of the first core columns, indicated by the blue rectangle in Figure 6.13. The matrix is divided into cores of 64 pixels, within every core the pixel hits are processed in pixel regions of 1 x 4 pixels. The number of pixels registering an SEU inside every region has been analysed as this can be easily related to the chip architecture.

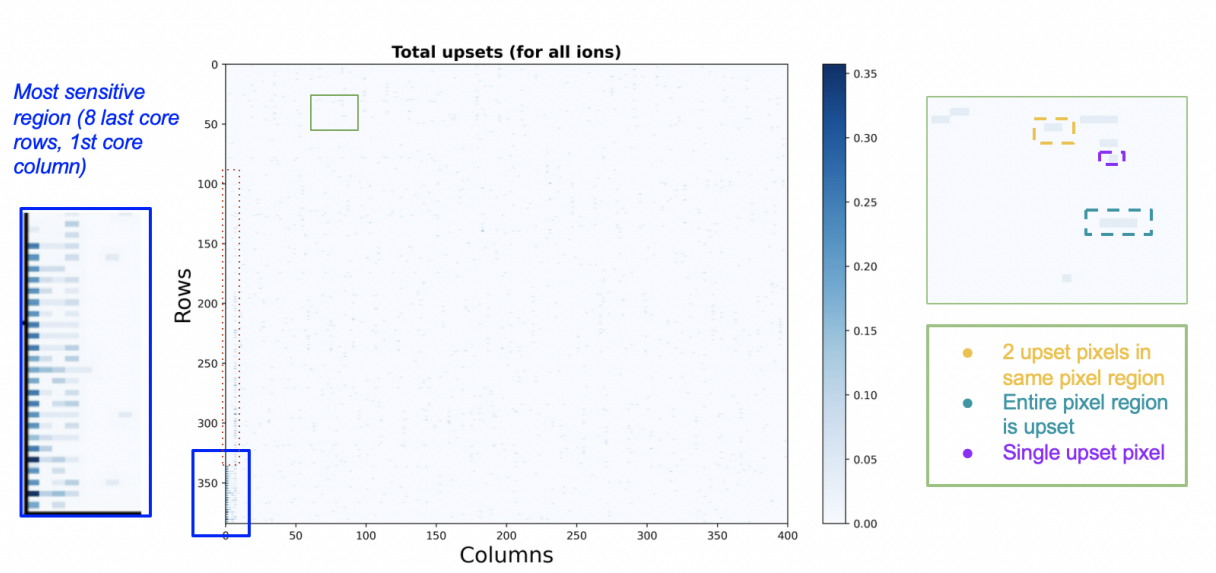


Figure 6.14: Pixel matrix including total number of upsets in all the digital scans performed for all ions. Zoom-in to the most affected region. Example of how many pixels per region have an error.

This study allows to go a bit further when counting the total number of SEUs that affected every scan as well as identifying the specific group in the readout chain. Based on ITkPixV1 chip manual [121], the pixel region diagram is as depicted in Figure 6.15.

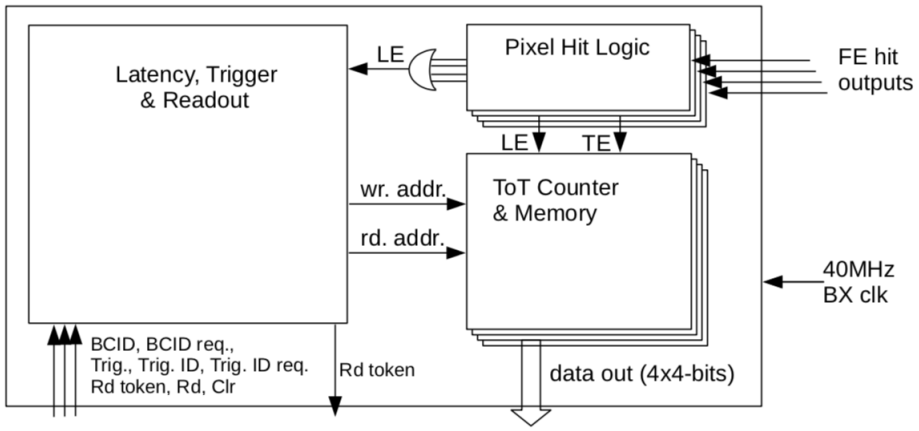


Figure 6.15: 4-Pixel region block diagram. LE is leading edge, TE is trailing edge, BCID is bunch crossing counter value. Image taken from [121].

In order to be able to count the total number of SEUs it is important to relate them to the pixel region diagram. Whenever there is a single pixel upset it means that it has taken place in the Pixel Hit Logic block as it is the only time in which the hits are stored individually. On the other hand, if two neighbouring pixels belonging to the same region did not have the right number of injections this should be interpreted as a single upset that occurred in the Latency, Trigger &

Readout block because in the test procedure always 2 neighbouring pixels in one pixel region are injected and read out at the same time. The last case is when the 4 pixels have incorrect data this is regarded as two independent SEU that happened in the Latency, Trigger & Readout for the same reason as the double pixel upset. By doing this analysis it can be pointed out which logic in the chip is the most sensitive to getting SEUs and study if changes should be made to make it more robust. In the case of the ITkPixV1 the majority of time more than one pixel is affected by SEU during the digital scans.

The last point to mention is the digital scans performed with the chip configured in PLL mode. During the testbeam in October it was observed that after a few injections the error "No Rx sync" will appear. This implies that Aurora channel in the FPGA lost its link to the chip making the communication impossible. For more information on the Aurora protocol used to transmit and receive the data in ITkPixV1, refer to [122]. Consequently, no data was collected to analyse from these scans. For the testbeam in November, the scan was modified to have error handling so the communication between the DAQ system and the chip can be reestablished if a similar error is observed. The outcome showed that the communication can be recovered although the data saved is damaged and cannot be decoded. In the future, dedicated SEU tests will be made just to address this issue which seems to happen every time there is a phase jump in the Clock and Data recovery unit of the chip.

6.2.3 TRIUMF testbeam

Facility and setup description

On the 19th and 20th December 2020 a third SEU testbeam campaign took place in TRIUMF, the Canadian national centre of particle acceleration, located in Vancouver. The facility provided a 480 MeV proton beam with a flux of $15 \times 10^8\ particles\ /\ cm^2\ /\ s$. The beam has a diameter of 25mm and it was aligned so it covered the full digital chip bottom. Giving health circumstances of the time, this campaign was run remotely with support from a local physicist, who helped with the installation of the setup and the operation of the beam.

The setup installation consisted of a BDAQ board connected with 2m display port cables to the single chip card inside the beam area. The power supplies, 3 single port TTis, were in the same space attached with 2m banana plug to molex power cables. The laptop from where all the scans were performed was connected to the global network in the control room and could be remotely accessed. A description of the set-up is shown in Figure 6.16.

In this case and differently from the testbeams in Louvain the chip was powered in LDO mode. While the setup was running VDDD and VDDA voltage values were monitored with the on chip ADC as well as the temperature with the 3 temperature sensors present on chip. The outcome of this monitoring showed that despite having seen an increase of VDDD due to the TOT issue and having reached the compliance current of 3A after an hour, a reset of the chip

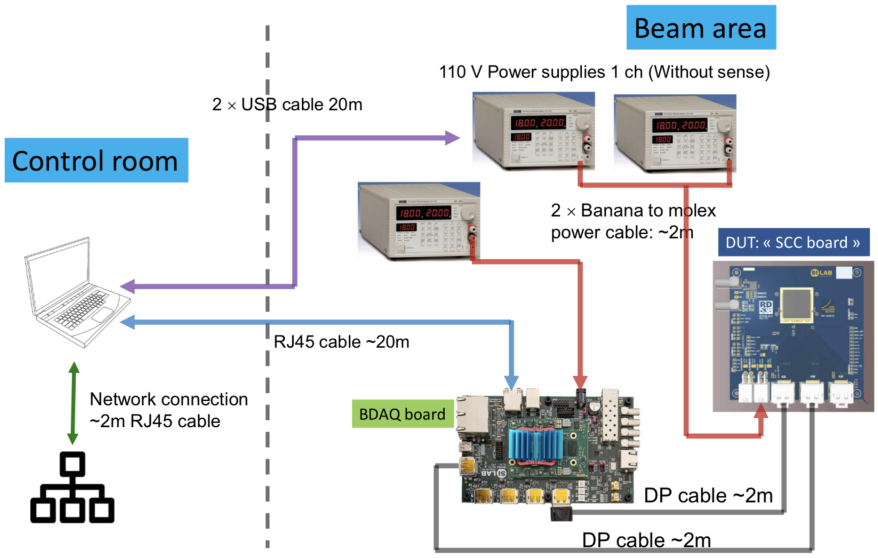


Figure 6.16: Schematic of the connections of the set-up installed for the SEU testbeam in TRIUMF.

allows the LDO to recover nominal power values. The chip certainly heated up several degrees during operation but kept functioning as expected delivering the correct regulated voltage.

The main goals of this campaign were to have some long acquisition runs for the global configuration registers and calculate their cross-section, to verify the in-pixel cross-section and to run all scans in PLL mode to see how these type of particles affect the Clock and Data Recovery.

Initially, testing was dedicated to perform the tests in bypass operation mode and once enough data has been recorded to switch to PLL operation mode. The procedure for the different scans remain the same as described in Section 6.2.1 and used in Louvain-la-Neuve so the results are comparable. Table 6.3 covers all the pixel configuration scans performed in bypass mode. The dosimetry foil test is done to check the correct alignment of the beam with respect to the digital chip bottom.

Table 6.3: The pixel configuration tests performed in TRIUMF in bypass mode.

Test	Duration (min)	Fluence	Comments	Bit Flips
Pixel Configuration	9	8.85×10^{11}	Write and read outside beam	3802
Pixel Configuration	10	9.38×10^{11}	Write and read outside beam	5007
Pixel Configuration	40	4.32×10^{12}	Write inside and read outside beam	27056
Pixel Configuration	5	4.44×10^{11}	Write inside and read outside beam	1768
Dosimetry foil	2	5.37×10^{10}	Write and read outside beam	206

Pixel configuration characterisation results

The results of the analysis are very similar to the results observed with the heavy ions in Louvain. It allows to verify that for low fluences the two bits that don't have triplication are considerably more affected by SEUs than the ones that have protection. The ratio becomes more equal when the fluence is higher due to the accumulation effect previously explained. For the set of scans shown in Figure 6.17 it is more likely to get a flip from zero to one than from one to zero although the direction of the flip should be completely random but the amount of data is not big enough to be conclusive.

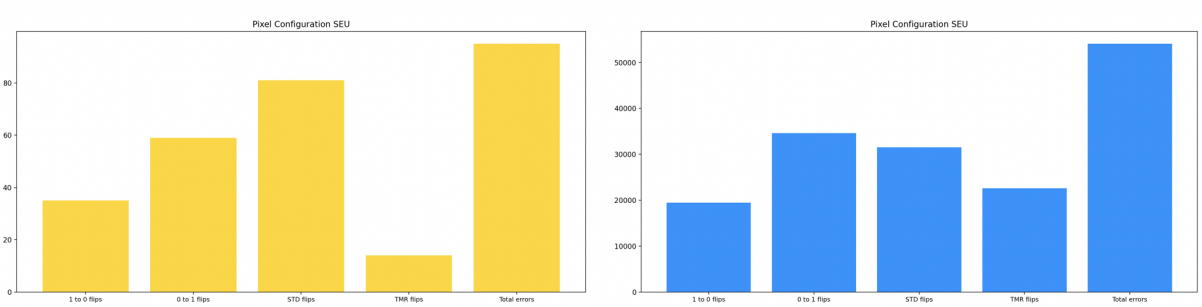


Figure 6.17: Histograms for two of the pixel configuration tests performed in TRIUMF that covers the number of SEU that induce a '0' to '1' flip, a '1' to '0' flip, total number of SEU, number of SEUs in non protected registers(left plot) and in triplicated registers (right plot).

The final plot in Figure 6.18 proves that the unprotected bit cross section remains the same regardless of the fluence reached and the operational mode while the TMR bit cross-section increases with fluence but seems coherent between modes.

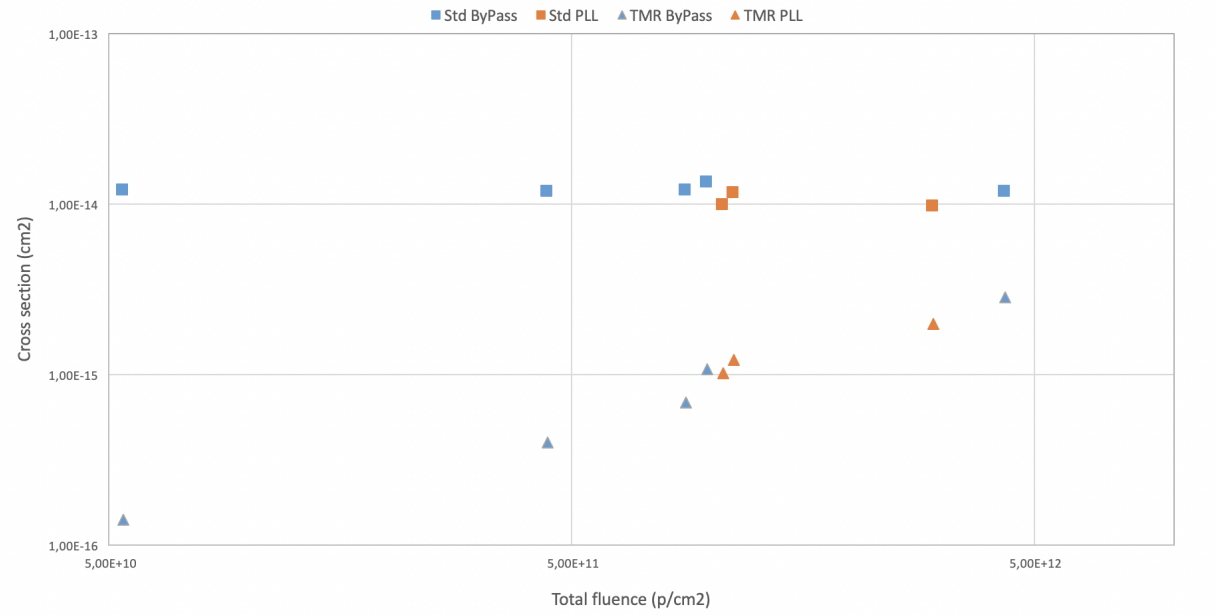


Figure 6.18: Cross section of the pixel configuration tests performed in TRIUMF according to the operation mode and fluence reached.

Global configuration characterisation results

Looking at the global configuration register scans three of them were done.

- Long duration, 10:30 hours, global register run in bypass mode. A total fluence of $7.21 \times 10^{13}\ ions/cm^2$ was reached and the number of SEUs observed was 8.
- Long duration, 2 hours, global register run in PLL mode. A total fluence of $1.39 \times 10^{13}\ ions/cm^2$ was reached and the number of SEUs observed was 4.

Based on the measurements, the calculated cross-section of the global registers with protons is in the order of 10^{-17} per cm^2.

Finally, from Table 6.4, it can be concluded that the results are compatible with what is the expected prediction using the results of the heavy ion tests for the doses expected for the HL-LHC. There is an improvement of ≈ 2 orders of magnitude between the cross-section of unprotected latches and the full TMR registers of the global configuration registers. These estimations were done by CERN to predict the number of SEUs that could be expected in the electronics when exposed to different hadrons like protons, pions or neutrons [123].

Table 6.4: Estimated cross-section of global registers and normal latches for a 200 MeV proton based on heavy ion testing and measured cross-section in the TRIUMF with a 480 MeV proton beam [123]. Note that the single latch cross-section is equivalent to the non triplicated pixel registers.

	Single latch cross section (cm^2)	**Global register cross-section (cm^2)**
Calculation for 200 MeV proton based on heavy ion testing	8.63×10^{-15}	7.86×10^{-17}
Experimental value 480 MeV proton test	STD: ~ 1.2×10^{-14} TMR: 1-5 x 10^{-16}	3.57×10^{-17}

Digital tests characterisation results

The last type of tests done were the digital scans. A total number of 22 of them were measured. For the first 15 iterations, the fluence reached was 6.02 x 10^{12} particles / cm^2 and for the last 7 the flux was decreased 50% with respect to the initial value so the total fluence was 9.57 x 10^{11} particles / cm^2. All these tests were done in PLL mode, with LDO powering and the software modified with respect to the one run in Louvain to have error handling and to be able to see what happens to the data when SYNC errors appear. For the high flux tests the analysis of the data taken reveals that, for 8 iterations the scan successfully finished due to the error handling modifications but the analysis of the data could not be done. The raw data files which were also

stored turned out to be corrupted. The other 7 iterations generated all the expected plots. For the low flux runs, 4 of them gave corrupted outputs and 3 of them successful ones.

The Aurora protocol is structured in a way that has data blocks and control blocks. Both blocks have their own specific header bits, 01 and 10 respectively, that allow for synchronisation of the serial data. The SYNC errors observed in the previous campaign are related to the following two scenarios:

- Identification of Aurora protocol soft errors. Headers 00 and 11 are illegal and are reported. This type of error is transient and also expected in the normal course of operation. In bypass mode during ion tests performed in Louvain soft errors were also reported but without creating sync issues with the receiver link. All data was captured and interpreted
- Sync lost in the Aurora protocol. This error is caused by a burst of soft errors, signal integrity issues in the set-up or clock jitter.

In the testbeams so far while running in bypass mode the digital scans, soft errors had been reported but did not cause sync issues. However, while the scans are run in PLL mode, both types of errors appear. The average time between soft errors has been calculated and it is directly proportional to the fluence reached. For the high fluence scans, the average soft error rate is 6.59 per second and the total number of sync errors is 5 which can be extrapolated to one every 22.8 s. For the lower fluence tests, the average soft error rate is 3.20 per second and the sync error rate 5 or one every 42.06 s.

The data that could be analysed showed two different types of occupancy maps: ones that have some noisy pixels and others where the injection pattern is visible. The first type of plots had been observed in the previous testbeam and sometimes they can even appear on normal laboratory scans but the second type was new. The most interesting fact is that if the noisy pixels are masked, the injection pattern becomes clear for the first type of maps. Once the injection pattern is visible, the amount of corrupted pixels depends on the number of sync errors received during the scan duration.

6.3 Conclusions from the SEU testbeam campaigns

The three SEU campaigns have demonstrated that the protection techniques in place on the chip lead to a lower count in the number of SEU with respect to normal latches. Furthermore, the full triplication protection for the global configuration registers has proven to be around two orders of magnitude better than the pixel configuration triplication without clock refreshment. Furthermore, if the clock is removed from the global configuration registers the cross-section values calculated are very similar to the triplicated pixel configuration registers confirming that the schemes behave as expected.

Nevertheless, there will need to be a lot of effort put on monitoring the PLL and CDR to be able to understand the phase and clock variations that deal to the synchronisation issues during the digital scans run in PLL mode.

New specific testbeam campaigns targeting this last point are being scheduled and the outcome of them will determine if some modifications will need to be done in the final version of the chip to be able to overcome the difficulties explained.

It is still not fully clear if the level of protection present in the current design is enough to stand the doses that the chip will receive during operation in the HL-LHC. Once the upcoming campaigns are finished, the experiment will look at all the results in combination with the refreshing capacities provided by the DAQ system and take a decision consequently.

Chapter 7

Conclusions

This thesis has presented the different designs, characterisation results and simulations done for the depleted monolithic active sensors of the MALTA family. It has also shown SEU results for the ATLAS prototype FE-chip that will be installed in the Inner Tracker detector during the HL-LHC Upgrade.

The MALTA DMAPS that was fabricated in the TowerJazz 180nm CMOS process had shown a loss in efficiency in the corners of the pixel after NIEL irradiation. The charge did not reach the electrode when deposited in the pixel corners. Simulations to understand this behaviour showed that the lateral electrical field was not strong enough to collect the charge. Two different process modifications, adding an additional mask in the fabrication process to have an extra deep p-well and modifying the n-blanket mask to include a gap, were proposed. These changes were implemented in the Mini-MALTA sensor that was fabricated as a demonstrator for the radiation hardness, to try a new synchronisation scheme for the asynchronous architecture and to address other identified issues like RTS noise that prevented operation at low thresholds. Instead of being a 512×512 pixel matrix as in the MALTA sensor, it was 16×64. Measurements of the charge collection using sources in the lab on unirradiated devices and devices irradiated to $10^{15}\ MeV n_{eq}/cm^2$ were made. The outcome of the analysis demonstrated that the process modifications improved the charge collection and this was verified by testbeam results. These results demonstrated that the mini-MALTA with the process modifications was radiation tolerant up to $10^{15}\ MeV n_{eq}/cm^2$.

In parallel, a completely different idea was explored to see if the charge collection issue in the MALTA could be overcome using an alternative substrate material. MALTA had always been fabricated using a p-type substrate with a high resistivity epitaxial layer of 25-30 μm thickness, which constitutes the sensitive layer and can be fully depleted at low voltages. By using high resistivity Czochralski material for the substrate in the fabrication of the MALTA sensor, the sensitive volume can be considerably increased, which provides a larger signal that compensates, at least in part, the higher noise and charge threshold. This development also showed a very positive result as the new MALTA batch fabricated with this new material in

combination with the mini-MALTA process modifications was efficient even after irradiation. Laboratory and testbeam measurements show that it operates in a similar way to the standard MALTA but that the increased bias voltage that can be applied results in an increased sensitive volume, which gives a larger signal, increased cluster size and improved charge collection.

A simulation of the asynchronous readout for ATLAS data rates was made. This demonstrated that the readout could operate at data rates corresponding to the outer radius of the ITk Pixel system. This was also used to define the dimension of the different memories used in the Mini-MALTA end-of-column readout.

Eventually, incorporating all the knowledge acquired with the previous version of MALTA, the MALTA2 sensor was designed and fabricated. The front-end electronics were changed to include a cascoded transistor that provides higher gain, reducing noise and consequently allowing operation at lower thresholds. The readout kept the original asynchronous architecture. Furthermore, the difficulties with the slow control operation were tackled and the block was completely redesigned, which was a significant part of the work of this thesis. A design based on a shift register was developed and then implemented into the chip including RTL, synthesis and place and route. This modification enabled threshold scans to be made for the full matrix, which had limited the operation and characterisation of the previous versions of the chip. Testbeams are already scheduled which will determine if the radiation tolerance has been further increased. In the meanwhile the MALTA3 sensor is being designed, a new large sensor that will provide excellent time resolution <2 ns and tracking capabilities with an improved synchronisation of the data.

On a completely different topic, this thesis presents the characterisation results of the Single Event Upset testing of the ITkPixV1 chip that is the prototype foreseen to be used by the ATLAS ITK Upgrade for the HL-LHC. It describes the mitigation techniques used for the design and the measurements made to determine the SEU cross-sections of the different memories in the chip. The results were compiled from several testbeam campaigns. The outcome is that the protection techniques lead to a lower SEU count compared to normal latches. The full triplication protection has proven to be around two orders of magnitude better than the pixel configuration that is implemented without clock refreshing. More testing is needed to assess the PLL robustness against SEE, and DAQ experts should determine if the chip is robust enough to ensure good operation during HL-LHC.

Bibliography

[1] Lyndon Evans and Philip Bryant (editors) *LHC Machine* JINST, 2008, pp. S08001 164 pgs

[2] The ATLAS Collaboration *Technical Design Report for the ATLAS Inner Tracker Pixel Detector* 15th of June 2018 with DOI:10.17181/CERN.FOZZ.ZP3Q and available at https://cds.cern.ch/record/2285585?ln=fr

[3] The ALICE collaboration *"The ALICE experiment at the CERN LHC. A Large Ion Collider Experiment* JINST, Vol. 3, 2008, pp. S08002. 259 p, also published by CERN Geneva in 2010, available at: https://cds.cern.ch/record/1129812

[4] The ATLAS collaboration *The ATLAS experiment at the CERN large hadron collider* Journal of Instrumentation, Vol. 3, No. 08, aug 2008, available at: https://doi.org/10.1088%2F1748-0221%2F3% 2F08%2Fs08003

[5] LHCb collaboration *The LHCb Detector at the LHC* JINST, Vol. 3, No. LHCbDP-2008-001. CERN-LHCb-DP-2008-001, 2008, pp. S08005, also published by CERN Geneva in 2010, available at: https://cds.cern.ch/record/1129809

[6] The CMS Collaboration *The CMS experiment at the CERN LHC* Journal of Instrumentation, Vol. 3, No. 08, aug 2008, available at: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08004

[7] Herr, W., Muratori, B. *Concept of luminosity* 2006, available at: https://cds.cern.ch/record/941318

[8] Aad, G. et al. *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC* Lett. B, Vol. 716, No. arXiv:1207.7214. CERN-PH-EP-2012-218, Aug 2012, pp. 1-29. 29 p, available at: https://cds.cern.ch/record/1471031

[9] De Melis, C. *The CERN accelerator complex* Jan 2016, general photo, available at: https://cds.cern.ch/record/2119882

[10] The ATLAS Collaboration *The ATLAS experiment at the CERN large hadron collider* Journal of Instrumentation, Vol. 3, No. 08, aug 2008, available at: https://cds.cern.ch/record/1129811/files/jinst8_08_s08003.pdf

[11] Maurice Garcia-Sciveres and Norbert Wermes *A review of advances in pixel detectors for experiments with high rate and radiation* Reports on Progress in Physics, Volume 81, Number 6, may 2018, doi: 10.1088/1361-6633/aab064

[12] Giovanni Calderini, on behalf of the ATLAS Collaboration *The ATLAS ITk detector for High Luminosity LHC Upgrade* Elsevier, Volume 1040, 1 October 2022 with doi: https://doi.org/10.1016/j.nima.2022.167048

[13] Federico Faccio *Radiation effects in the electronics for CMS* http://lhcb-elec.web.cern.ch/papers/radiation_tutorial.pdf

[14] P. Valerio et al. *A monolithic ASIC demonstrator for the Thin Time-of-Flight PET scanner* 2019 JINST 14 P07013 DOI 10.1088/1748-0221/14/07/P07013

[15] Hung, K.K. and Ko, P.K. and hu, Chenming and Cheng, Yiu *Random telegraph noise of deep-submicrometer MOSFETs* IEEE Transactions on Electron Devices, 1990, DOI: 10.1109/55.46938

[16] B. Mandelli on behalf of the ATLAS Collaboration *The Pixel Detector of the ATLAS Experiment for the Run 2 at the Large Hadron Collider* Nuclear and Particle Physics Proceedings, April 2016, 273-275:1166-1172 with DOI: 10.1016/j.nuclphysbps.2015.09.183

[17] Potamianos, K. *The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider* CERN, Geneva, Tech. Rep. ATL-PHYS-PROC-2016-104, Aug 2016, 15 pages, EPS-HEP 2015, Available at: https://cds.cern.ch/record/2209070

[18] Seiden A. *Characteristics of the ATLAS and CMS Detectors* Available at: https://doi.org/10.1098/rsta.2011.0461

[19] Capeans, M., Darbo, G., Einsweiller, K., Elsing, M., Flick, T., Garcia-Sciveres, M., Gemme, C., Pernegger, H., Rohne, O., Vuillermet, R., *ATLAS Insertable B-Layer Technical Design Report* Tech. Rep. CERN-LHCC-2010-013. ATLAS-TDR-19, Sep 2010, available at: https://cds.cern.ch/record/1291633

[20] M. Garcia-Sciveres et al. *The FE-I4 pixel readout integrated circuit* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 636, 1, Supplement (2011) S155 – S159. 7th International Hiroshima Symposium on the Development and Application of Semiconductor Tracking Detectors.

[21] A. Abdesselam et al. *The barrel modules of the ATLAS semiconductor tracker* NIMA Volume 568, Issue 2, 1 December 2006, Pages 642-671

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

[23] Zbynek Drasal et al. Nucl. Instrum. Meth., A910 (2018) 127–132.

[24] Apollinari G., Alonso I. B., Bruning O., Fessia P., Lamont M., Rossi L., Tavian L. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report* V. 0.1, ser. CERN Yellow Reports: Monographs. Geneva: CERN, 2017, available at: http://cds.cern.ch/record/2284929

[25] Apollinari G. et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1* CERN Yellow Reports: Monographs. CERN, Geneva (2017).http://cds.cern.ch/record/2284929

[26] *Technical Design Report for the ATLAS Inner Tracker Pixel Detector* CERN, Geneva, Tech. Rep.CERN-LHCC-2017-021. ATLAS-TDR-030, Sep 2017, available at: https://cds.cern.ch/record/2285585

[27] Christiansen, J. C., Garcia-Sciveres, M. L *RD Collaboration Proposal: Development of pixel readout integrated circuits for extreme rate and radiation* CERN, Geneva, Tech. Rep. CERN-LHCC-2013-008. LHCC-P-006, Jun 2013, available at: https://cds.cern.ch/record/1553467

[28] Arteche Gonzalez, F. et al. *Extension of RD53* CERN, Geneva,Tech. Rep. CERN-LHCC-2018-028. LHCC-SR-008, Sep 2018, available at: http://cds.cern.ch/record/2637453

[29] Garcia-Sciveres, M. L *The RD53A Integrated Circuit* CERN, Geneva, Tech. Rep. CERN-RD53-PUB-17-001, Oct 2017, available at: https://cds.cern.ch/record/2287593

[30] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/RadiationSimulationPublicResults

[31] https://cerncourier.com/a/silicon-sensors-go-3d/

[32] P. Grenier. Physics Procedia, 37 (2012) 874–881.ISSN1875-3892. Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011) http://www.sciencedirect.com/science/article/pii/S1875389212017865

[33] Ta, D. et al. *Serial powering: Proof of principle demonstration of a scheme for the operation of a large pixel detector at the lhc.* Nucl. Instrum. Methods Phys. Res., A, Vol. 557, 2006, pp. 445-59, available at: https://cds.cern.ch/record/1020524

[34] Lehmann, N. et al. *Prototype Chip for a Control System in a Serial Powered Pixel Detector at the ATLAS Phase II Upgrade* PoS, Vol.TWEPP17, 2017, pp. 026.

[35] M. Tanabashi et al. *Review of Particle Physics* Phys. Rev. D, 2018, Volume 98, issue 3, doi: 10.1103/PhysRevD.98.030001

[36] *Photons interactions with matter* https://clinicalgate.com/basics-of-radiation-therapy-2

[37] *The Compton Effect* Anon, 2022. Available at: https://phys.libretexts.org/@go/page/4522

[38] D. Bortoletto. *Detectors for Particle Physics* Available at: https://indico.cern.ch/event/190055/attachments/269161/376713/daniela_l4.pdf

[39] S. M. Sze *The physics of semiconductors Devices* Wiley, NewYork, 1969 pp. 12-20

[40] Adel S. Sedra, Kenneth C. Smith , Tony Chan Carusone, Vincent Gaudet *Microelectronics circuits* Oxford University Press; 8th edition (November 15, 2019)

[41] Chun-Min Zhang et al. *Characterization and Modeling of Gigarad-TID-Induced Drain Leakage Current of 28-nm Bulk MOSFETs* IEEE Transactions on nuclear science, Vol. 66, No. 1, January 2019

[42] Anelli G., et al. *Radiation tolérant VLSI circuits in standard deep submicron CMOS technologies for the LHC experiments: Practical design aspects* IEEE Transactions on Nuclear Science Volume 46, Issue 6 PART 1, 1999, Pages 1690-1696

[43] Federico Ravotti *Dosimetry Techniques and Radiation Test Facilities for Total Ionizing Dose Testing* IEEE transactions on Nuclear Science, Vol. 65, No. 8m August 2018

[44] Henri Koch *Effects of Ultra-High Total Ionizing Dose in Nanoscale Bulk CMOS Technologies* Cern thesis available in: https://cds.cern.ch/record/2641496/files/CERN-THESIS-2018-184.pdf

[45] Miryala S. et al, *Characterization of soft error rate against memory elements spacing and clock skew in a logic with triple modular redundancy in a 65nm process* PoS TWEPP 2018

[46] Federico Faccio et al. *Total dose and Single Event Effects (SEE)in a 0.25m CMOS technology* 4th Workshop on Electronics for LHC Experiments (LEB 98), 105-113 Available at: https://www.researchgate.net/publication/2386557

[47] A.Rivetti *CMOS: front-end electronics for radiation sensors (Devices, Circuits, and Systems)* Boca Raton, FL: CRC Press, 2015.

[48] Y. Allkofer, C. Amsler, D. Bortoletto, V. Chiochia, L. Cremaldi, S. Cuc-ciarelliet et al. *Design and performance of the silicon sensors for the cms barrel pixel detector* Nuclear Instruments and Methods in Physics Re-search Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 584, no. 1, pp. 25 – 41, 2008

[49] Matthias Hamer on behalf of the ATLAS ITk Community *Phase-II Upgrade of the ATLAS Pixel Detector* March, 2018. Available: http://cds.cern.ch/record/2665116

[50] *Flip chip technology* Available at: http://x-ray.camera/technology/flip-chip-bonding/

[51] L. Rossi, P. Fischer, T. Rohe, and N. Wermes *Pixel detectors: From fundamentals to applications* Springer, 2006

[52] Edited by N.Balkan *Hot Electrons in Semiconductors. Physisc and devices* Oxford Science Publications

[53] J. R. Srour and J. W. Palko, *Displacement damage effects in irradiated semiconductor devices* IEEE Trans. Nucl. Sci., vol. 60, no. 3, pp.1740–1766, Jun.2013.

[54] *Radiation damage in silicon particle detectors: Microscopic defects and macroscopic properties* Ph.D. dissertation, Dept. Phys., Univ. Hamburg, Hamburg, Germany, 1999

[55] A. Chilingarov *Temperature dependence of the current generated in Sibulk* J. Instrum., vol. 8, no. 10, p. P10003, 2013

[56] Moll M. et al. *Relation between microscopic defects and macroscopic changes in silicon detector properties after hadron irradiation* Nucl. Instrum. Methods Phys. Res. B, Beam Interact. Mater. At., vol. 186, pp. 100–110, Jan. 2002

[57] Michael Moll *Displacement Damage in Silicon Detectors for High Energy Physics* IEEE transactions on nuclear science, Vol. 65, No. 8, August 2018

[58] A. Chilingarov *Generation current temperature scaling* Available at: http://cds.cern.ch/record/1511886

[59] Michael Moll *Radiation damage in silicon particle detectors: Microscopic defects and macroscopic properties* Thesis: PhD Hamburg U. (1999) Report number:DESY-THESIS-1999-040

[60] V. Eremin, E. Verbitskaya, and Z. Li. *The origin of double peak electric field distribution in heavily irradiated silicon detectors* Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip., vol. 476, no. 3, pp. 556–564, 2002.

[61] G. Kramberger, V. Cindro, I. Mandic, M. Miku, and M. Zavrtanik *Effective trapping time of electrons and holes in different silicon materials irradiated with neutrons, protons and pions* Nucl. Instrum. Methods Phys.Res. A, Accel. Spectrom. Detect. Assoc. Equip., vol. 481,pp. 297–305, Apr. 2002

[62] J. Kaplon and W. Dabrowski *Fast CMOS binary front end for silicon strip detectors at LHC experiments* IEEE Transactions on Nuclear Science, vol. 52, no. 6, pp. 2713–2720, Dec 2005

[63] R. Horisberger *Readout architectures for pixel detectors* Nucl. Instrum. Meth., vol. A465, pp. 148–152, 2000

[64] Snoeys, W *CMOS monolithic active pixel sensors for high energy physics* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Vol. 765, 2014, pp. 167-171, Proceedings of the 9th International "Hiroshima" Symposium on Development and Application of Semiconductor Tracking Detectors (HSTD-9 2013)

[65] I.Peric *A novel monolithic pixelated particle detector implemented in high-voltage CMOS technology* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Vol. 582, No. 3, 2007, pp. 876-885, VERTEX 2006

[66] Snoeys, W. *Monolithic pixel detectors for high energy physics* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Vol. 731, 2013, pp. 125-130, PIXEL 2012

[67] N.Wermes on behalf of the ATLAS CMOS Collaboration *Depleted CMOS pixels for LHC proton–proton experiments* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,Volume 824, 11 July 2016, Pages 483-486

[68] Kolanoski, H., Wermes, N. *Teilchendetektoren: Grundlagen und Anwendungen* Berlin: Springer Spektrum, 2016

[69] Sara Marconi *Design and optimisation of low power hybrid pixel array logic for the extreme hit and trigger rates of the Large Hadron Collider Upgrade* Available at: http://cds.cern.ch/record/2665116/files/Fulltext.pdf

[70] G. Aglieri Rinella et al. *The ALPIDE pixel sensor chip for the upgrade of the ALICE Inner Tracking System* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Vol. 845, 2017, pp. 583-587

[71] Caicedo, et al. *The Monopix chips: Depleted monolithic active pixel sensors with a column-drain read-out architecture for the ATLAS Inner Tracker upgrade* http://cds.cern.ch/record/2665944/files/1902.03679.pdf

[72] Enrico Junior Schioppa, et al. *First tests of a novel radiation hard CMOS sensor process for Depleted Monolithic Active Pixel Sensors* https://iopscience.iop.org/article/10.1088/1748-0221/12/06/P06008

[73] van Hoorne, J. et al. *The Investigator: an efficient tool to optimize design parameters of a CMOS pixel sensor* IEEE Nuclear Science Symposium and Medical Imaging Conference (2016 NSS/MIC), November 2016

[74] H. Pernegger et al,. *First tests of a novel radiation hard CMOS sensor process for Depleted Monolithic Active Pixel Sensors* 2017 JINST12 P06008

[75] Cardella, R. et al. *"LAPA, a 5 Gb/s modular pseudo-LVDS driver in 180 nm CMOS with capacitively coupled pre-emphasis* Proceedings of Science, Vol. TWEPP17, 2017, pp. 038.5

[76] Kim, D. et al. *Front end optimization for the monolithic active pixel sensor of the ALICE Inner Tracking System upgrade* Journal of Instrumentation, Vol. 11, No. 02, February2016, pp. C02042

[77] Ivan Berdalović *Design of radiation-hard CMOS sensors for particle detection applications* Available in https://cds.cern.ch/record/2702884/files/CERN-THESIS-2019-221.pdf

[78] Meroli S., Passeri D. and Servoli L. *Energy loss measurement for charged particles in very thin silicon layers* 2011 JINST 6 P06013.

[79] C. Z. Yuan and W. Sansen *Low-noise wide-band amplifiers in bipolar and CMOS technologies* Kluwer, 1991. ISBN 0792390962, 9780792390961. doi:10.1007/978-1-4757-2126-3

[80] Mijke Schut *Characterisation of the Timepix3 chip using a gaseous detector* Available in https://wiki.nikhef.nl/detector/pub/Main/ArticlesAndTalks/thesis_mijke_schut.pdf

[81] J. Christiansen and M. Garcia-Sciveres *RD Collaboration proposal:Development of pixel readout integrated circuits for extreme rate and radiation.* Available in: http://cds.cern.ch/record/1553467?ln=en, CERN, July 2013

[82] Kugathasan T. et al., *Review on depleted CMOS* Available in: https://cds.cern.ch/record/2696397?ln=fr

[83] Snoeys W. et al., *A process modification for CMOS monolithic active pixel sensors for enhanced depletion, timing performance and radiation tolerance* Nucl. Instrum. Methods Phys. Res., A, Vol. 871, 2017, pp. 90-96. 7 p, available at: https://cds.cern.ch/record/2280552

[84] P. Behara, G. Gaycken, C. Horn, A. Khanov, D. Lopez Mateos *Threshold T eshold Tuning of the A uning of the ATLAS Pixel Detect el Detector* The ATLAS Collaboration, ATL-INDET-PUB-2010-001, CERN(2010).

[85] The ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector* CERN, Geneva, Tech. Rep. CERN-LHCC-2017-021. ATLAS-TDR-030, Sep 2017, available at: https://cds.cern.ch/record/2285585

[86] Munker M. et al. *Simulations of CMOS pixel sensors with a small collection electrode, improved for a faster charge collection and increased radiation tolerance* J. Instrum., 14 (2019), Article C05013 [1903.10190]

[87] Dyndal M. et al., *Mini-MALTA: radiation hard pixel designs for small-electrode monolithic CMOS sensors for the High Luminosity LHC* 2020 JINST 15 P02005

[88] Schioppa E.J. et al., *Measurement results of the MALTA monolithic pixel detector* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment Volume 958, 1 April 2020, 162404

[89] D. Kim et al., *Front end optimization for the monolithic active pixel sensor of the ALICE Inner Tracking System upgrade* 2016 JINST 11 C02042

[90] Cardella R. et al., *MALTA: an asynchronous readout CMOS monolithic pixel detector for the ATLAS High-Luminosity upgrade* 2019 JINST 14 C06019

[91] Berdalovic I. et al., *MALTA: a CMOS pixel sensor with asynchronous readout for the ATLAS High-Luminosity upgrade* DOI 10.1109/NSSMIC.2018.8824349

[92] Snoj L., Zerovnik G., Trkov A. *Computational analysis of irradiation facilities at the JSI TRIGA reactor* Applied Radiation and Isotopes, Vol. 70, No. 3, 2012, pp. 483-488

[93] Aashish Agrawal *Radiation effects in Microelectronics* Available at: https://slidetodoc.com/radiation-effects-in-microelectronics-ee698-a-course-seminar/

[94] *Single Event Effects* Available at: http://holbert.faculty.asu.edu/eee560/see.html

[95] Wang T., et al. *A high speed transmitter circuit for the ATLAS/CMS HL-LHC pixel readout chip* PoS TWEPP 2018, 098

[96] Timon Heim *First test results from the ITkPixV1 pixel readout chip* Vertex 2020 presentation available in: https://indico.cern.ch/event/895924/contributions/3968862/

[97] *https://uclouvain.be/en/index.html*

[98] *https://www.triumf.ca/*

[99] S. Bonacini et al. *Characterization of a commercial 65 nm CMOS technology for SLHC applications* 2012 JINST 7 P01015 available at: https://cds.cern.ch/record/1428524/files/1748-0221_7_01_P01015.pdf

[100] Hiti, B. et al. *Development of the monolithic MALTA CMOS sensor for the ATLAS ITK outer pixel layer* Proceedings of Science, Vol. TWEPP-18, 2018, pp. 155.

[101] Sharma A. et al. *The Malta CMOS pixel detector prototype for the ATLAS Pixel ITK* PoS 2019, doi = 10.22323/1.348.0014

[102] H. Jansen, et al. *Performance of the EUDET-type beam telescopes* https://arxiv.org/pdf/1603.09669.pdf

[103] C. Kleinwort. *General broken lines as advanced track fitting method* NIM:A, 673:107–110, 2012. doi:10.1016/j.nima.2012.01.024

[104] Roberto Cardella *CMOS Detector and System Developments for LHC Detector Upgrades* Available in https://cds.cern.ch/record/2702969/files/CERN-THESIS-2019-224.pdf

[105] L. Rossi *Pixel detectors hybridisation* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detecors and Associated Equipment, vol. 501, no. 1, pp. 239 – 244, 2003,proceedings of the 10th International Workshop on Vertex Detectors

[106] R Gonçalo *The ATLAS trigger: high-level trigger commissioning and operation during early data taking* Journal of Physics: Conference Series 110 092013 DOI 10.1088/1742-6596/110/9/092013

[107] G. Mazza, et al. *A 1.2 Gb/s Data Transmission Unit in CMOS 0.18 µm technology for the ALICE Inner Tracking System front-end ASIC* February, 2017, Journal of Instrumentation, DOI: 10.1088/1748-0221/12/02/C02009

[108] Sarah L. Harris and David Money Harris *Digital Design and Computer Architecture* Morgan Kaufmann, chapter 4-Hardware Description Languages, pages 172-237, DOI: https://doi.org/10.1016/B978-0-12-800056-4.00004-2

[109] *Fastest Simulator to Achieve Verification Closure for IP and SoC Designs* https://www.cadence.com/ko_KR/home/tools/system-design-and-verification/simulation-and-testbench-verification/xcelium-simulator.html

[110] *Genus Synthesis Solution* https://www.cadence.com/ko_KR/home/tools/digital-design-and-signoff/synthesis/genus-synthesis-solution.html

[111] Calafiura, P. and Lavrijsen, W. and Leggett, C. and Marino, M. and Quarrie, D. *The Athena control framework in production, new developments and lessons learned* 14th International Conference on Computing in High-Energy and Nuclear Physics, pages 456-458, 2005

[112] Harris, David and Harris, Sarah *Digital design and computer architecture* Morgan Kaufmann, 2010, Paperback ISBN: 978012394424

[113] Michael Daas, et al. *BDAQ53, a versatile pixel detector readout and test system for the ATLAS and CMS HL-LHC upgrades* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detecors and Associated Equipment, vol. 986, DOI: 10.1016/j.nima.2020.164721

[114] *Hamming Code* Available at: https://en.wikipedia.org/wiki/Hamming_code

[115] Miryala S, Hemperek T and Menouni M *Characterization of Soft Error Rate Against Memory Elements Spacing and Clock Skew in a Logic with Triple Modular Redundancy in a 65nm Process* PoS TWEPP-2018

[116] K. Moustakas et al. *A Clock and Data Recovery Circuit for the ATLAS/CMS HL-LHC Pixel Front End Chip in 65 nm CMOS Technology* PoS TWEPP 2019, 046.

[117] Florian Winkler *Verification of the Shunt-Low-Dropout voltage regulator for the current based supply of the serially connected pixel detector modules of the ATLAS- and CMS-experiments at the High-Luminosity Large Hadron Collider* Master Thesis Dortmund University

[118] Menouni M. et al *Single event effects testing of the RD53B chip* Submitted to TIPP 2021

[119] F. Faccio, G. Anelli, M. Campbell, M. Delmastro, P. Jarron, K. Klouldnas, A. Marchioro, P. Moreira, K. K.loukinas, E. Noah and W. Snoeys *TOTAL DOSE AND SINGLE EVENT EFFECTS (SEE) IN A* $0.25\mu m$ *CMOS TECHNOLOGY* Proc. 4th Workshop Electronics for LHC Experiments

[120] European Space Components Coordination *Single event effects methods and guidelines* Available at https://escies.org/download/webDocumentFile?id=62690

[121] Garcia-Sciveres, Maurice and Loddo, Flavio and Christiansen, Jorgen *RD53B Manual* Available at https://cds.cern.ch/record/2665301

[122] Xilinx *Aurora 64B/66B Protocol Specification* Xilinx, Aurora 64B/66B Protocol Specification, SP011 (v1.3) October 1, 2014

[123] M. Huhtinen and F. Faccio *Computational method to estimate Single Event Upset rates in an accelerator environment* Nuclear Instruments and Methods in Physics Research A 450 (2000) 155-172

[124] Url:https://uclouvain.be/fr/instituts-recherche/irmp/crc

[125] C. Ghabrous Larrea, K. Harder, D. Newbold, D. Sankey, A. Rose, A. Thea and T. Williams *IPbus: a flexible Ethernet-based control system for xTCA hardware* February 2015, Journal of Instrumentation, DOI: 10.1088/1748-0221/10/02/c02019

[126] Prafulla Behara, Goetz Gaycken, Claus Horn, Alexandre Khanov, David Lopez Mateos, Lucia Masetti, Klemens Muller, Kerstin Perez *Threshold Tuning of the ATLAS Pixel Detector* Available at http://cds.cern.ch/record/1283928

[127] Harris, D. and Harris S. *Advanced VL Codes*

[128] Salomon D., Motta G. *Advanced VL Codes* Springer, London. https://doi.org/10.1007/978-1-84882-903-9_3

[129] *Register Transfer Level definition* Available at: https://en.wikipedia.org/wiki/Register-transfer_level

[130] *Electromigration definition* Available at: hhttps://www.usf.edu/engineering/news-room/electromitigation.aspx

[131] *Vivado software* https://www.xilinx.com/products/design-tools/vivado.html