Characterization of radiation-hard monolithic CMOS sensors

Lluís Simón Argemí

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

April 2021

Abstract

The work presented in this thesis consists of the characterisation of monolithic CMOS sensors targeting the requirements of the outer-most layer of the ATLAS Inner Tracker after the High Luminosity upgrade of the Large Hadron Collider. Three detectors are investigated: an investigator chip and two large scale demonstrators (MALTA and mini-MALTA). The investigator chip is designed in the standard TowerJazz 180 nm technology and served as a tool to investigate the geometric parameters that affect the pixel capacitance. The MALTA chip is designed in the modified TowerJazz 180 nm technology and implements a novel asynchronous readout to minimise power consumption. The sensor is irradiated with X-rays up to 1.25 MRad to test the resistance of the front-end circuit to ionising radiation effects. The mini-MALTA chip is designed following the results obtained on MALTA and implements an improved front-end and pixel layout to enhance the radiation hardness of MALTA. A similar X-ray irradiation campaign is done for this chip showing good radiation hardness after 80 MRad of TID. Aside from the characterisation work, FPGA-based readouts for the MALTA and mini-MALTA chips were developed in collaboration with the CMOS development group at CERN.

Dedico aquest escrit a l'avi Pepe i l'àvia Teresina, perquè l'exemple i la confiança que m'heu donat m'acompanyaran sempre, vagi on vagi, sigueu on sigueu.

Acknowledgements

The work presented in this thesis would have never been possible without the help and support of many people. To begin with, I would like to thank my supervisors Prof. Craig Buttar and Dr. Richard Bates for their help and guidance throughout my time as a PhD student. I really appreciate every conversation we had, which always added value to the ongoing research, and the freedom you have given me to pursue the projects I was interested in. I would also like to address a big thanks to Dima Maneusky for your help and kindness inside and outside the lab.

My most sincere gratitude to the CMOS development group at CERN for allowing me to work with the group and providing invaluable expertise and motivation. Carlos Solans, Valerio Dao, Abhishek Sharma, Enrico Junior, Thanushan Kugathasan, Walter Snoeys, Roberto Cardella and Ivan Berdalovic, thank you for everything you have taught me and for taking the time to explain me all these complicated concepts of electronic circuits.

During my time as a PhD, I had the privilege of participating in the STREAM programme. A big thanks to Heinz Pernigger for organising this fantastic network and to Sonia Allegratti for making sure that we all did our paperwork. A special thanks to my STREAM fellows for every night out around Europe.

I would also like to express my gratitude to the people from the Detector Development group of the University of Glasgow. To Leyre, Calum, Joe, Sneha, Kenny, Fred, Tom, Liam, Andy, Lars and Val, thank you for every lunch together and every chat during coffee breaks. A special mention to Gordon Stewart for helping me out with every IT-related issue.

I would certainly have not survived this adventure without the good mood (and the caffeine) provided by the members of the Beans on Roast Coffee Club. Neil Moffat, Dwayne Spiteri, Laurynas Mince, Adam Rennie, Dima Manueski, Lojius Bin Lombigit and Conor Harkin, thank you for making more pleasant the several office hours we shared.

M'agradaria dedicar un últim reconeixement a totes les persones que formen part del meu àmbit personal, sense les quals no seria qui soc avui. A la meva família: la Clara, en Pep i l'Elena, que tot i la distància sempre m'heu sabut enviar un gest d'ànim. A tots aquells que m'heu vingut a veure, fins i tot més d'un cop, i els que, incansablement i sempre amb la mateixa il·lusió, m'heu rebut i acomiadat cada cop que tornava a casa. Al Marçal, que m'ha acompanyat en un bon tros d'aquesta aventura i al Marcio amb qui n'he viscut moltes més. Y a María, por ser mi compañera en este viaje y siempre animarme a seguir.

This research project has been supported by a Marie Skłodowska-Curie Innovative Training Network Fellowship of the European Commission's Horizon 2020 Programme under contract number 675587 STREAM.

Declaration

With the exception of chapters 2 and 3, which contain introductory material, all work in this thesis was carried out by myself unless otherwise explicitly stated. The chips characterised in this work and presented in chapter 4 were designed by my STREAM collaborators. My main contribution to the STREAM programme was the characterisation of the Investigator1 chip (chapter 5), development of a readout firmware for the MALTA and mini-MALTA chips in collaboration with the CERN CMOS group (chapter 6) and studies of the response of these chips to ionising radiation (chapter 7).

Contents

List of Tables

List of Figures

List of acronyms

AC Alternating Current. 74

ALICE A Large Ion Collider Experiment. 2, 5, 47, 48, 54, 55, 70, 86

ALPIDE ALICE pixel detector. 55, 65

AO Asynchronous Oversampling. xv, 95, 97–100, 105

ASIC Application Specific Integrated Circuit. 2, 86, 105

ATLAS A Toroidal LHC ApparatuS. ix, xi, 1, 2, 6–14, 16, 41, 46, 48, 54, 70, 78, 114, 118, 137, 138

CERN European Organization for Nuclear Research. 4, 115, 116

CLB Configurable Logic Blocks. xiv, 86–88

CMOS Complementary Metal-Oxide Semiconductor. xii, 1, 12–14, 16, 27, 31, 32, 35–37, 41, 43, 45–47, 49, 55, 65, 69, 70, 87, 138

CMS Compact Muon Solenoid. 1, 5

CMT Clock Management Tile. 89

COMPASS Common Muon and Proton Apparatus for Structure and Spectroscopy. 86

CP Charge Pump. 89

CSC Cathode Strip Chambers. 11

DAC Digital to Analog Converter. 65, 67, 69, 109, 110, 115, 116, 123, 127, 128

DAQ Data Acquisition. xvi, 2, 68, 85, 86, 93, 95, 101, 102, 105, 116, 117

DC Direct Current. 57, 58

DDR Double Data Rate. 65, 92

DMAPS Depleted MAPS. 2

DRU Data Recovery Unit. xv, 95, 98–100

DUT Device Under Test. 73, 113, 116

ELT Enclosed Layout Transistor. xii, 43, 44

EMEC ElectroMagnetic End-Cap. 10

FCAL Forward Calorimeter. 10

FET Field-Effect Transistor. 32

FET Metal Oxide Semiconductor Field-Effect Transistor. 32, 41, 129

FIFO First In First Out. 65, 95, 97

FMC FPGA Mezzanine Connector. xiv, 93–95, 98–100, 105, 116

FPGA Field Programmable Gate Array. ix, x, xiv, xv, 2, 64, 65, 67–69, 75, 86–89, 91–99, 101, 102, 105, 106, 116, 127, 137

GPAC General Purpose Analog Card. xiii, 73–75

HDL Hardware Description Language. 92

HEC Hadronic End-Cap. 10, 11

HEP High Energy Physics. 4, 9, 15, 45

HL-LHC High Luminosity LHC. xi, 1, 6, 9, 11, 12, 14, 16, 54, 71, 122

HPC High Pin Count. 93, 94

IBL Insertable B-Layer. 9

ID Inner Detector. 7–9, 12, 13

IO Input/Output. 86, 90, 91, 93

ITk Inner TracKer. ix, xi, 1, 2, 12–14, 16, 41, 46, 48, 54, 70, 78, 118, 137, 138

ITS Inner Tracking System. 47, 55, 70

LAPA pseudo-LVDS for ATLAS Pixel Apparatus. 65

LAr Liquid Argon calorimeter. 10

LED Light Emitting Diode. 95

LEP Large Electron-Positron Collider. 4

LF Loop Filter. 90

LHC Large Hadron Collider. xi, 1, 4–7, 11, 13, 15, 28, 31, 57, 86, 115, 138

LHCb Large Hadron Collider Beauty. 5

LINAC 2 Linear accelerator 2. 4

LPC Low Pin Count. 93

LS Long Shutdown. 6, 9, 11

LUT Look-Up Table. 88, 100

LVDS Low-Voltage Differential Signaling. 60, 65, 89, 94, 95, 97, 104, 105

MALTA Monolithic from ALICE to ATLAS. ix, x, xii–xvii, 2, 48, 54–57, 59–65, 67–71, 83–88, 91–109, 111, 114–117, 119–124, 126–129, 131–138

MAPS Monolithic Active Pixel Sensors. 1, 2

MDT Monitored Drift-Tube. 11

MIO Multi-I/O board. xiii, 73–75

MIP Minimum Ionizing Particle. 15, 28, 37, 78

MM Mini-Matrix. 78, 84

MMCM Mixed-Mode Clock Manager. xiv, 89, 90, 92, 94, 97

MPV Most Probable Value. 29, 134, 136

NIEL Non Ionizing Energy Loss. 138

NMOS N-type Metal Oxide Semiconductor Field-Effect Transistor. xii, 32–36, 40–42, 44–47, 50, 58, 68, 129

PC Personal Computer. xv, xvi, 68, 85, 86, 95, 97, 101, 102, 105, 113, 116, 117

PCB Printed Circuit Board. 68, 73, 99, 116, 117, 120

PCIe Peripheral Component Interconnect express. 74

PFD Phase-Frequency Detector. 89

PLL Phase-Locked Loop. 65

PMOS P-type Metal Oxide Semiconductor Field-Effect Transistor. 32, 36, 37, 40–42, 45–47, 50, 53–56, 58, 64, 68, 70, 71, 78

PS Proton Synchroton. 4

PSU Power Supply Unit. xiii, 74, 75, 109, 116, 127

R&D Research and Development. 1, 2, 16, 70

Rad Radiation absorbed dose. 41

RAM Random Access Memory. 64, 88

RF Radio Frequency. 4

RHIC Relativistic Heavy Ion Collider. 2

RINCE Radiation-Induced Narrow Channel Effect. 43, 44

RMS Root Mean Square. xvii, 38, 77, 110, 132, 133, 135

RPC Resistive Plate Chambers. 11

RTN Random Telegraph Noise. 40

RTS Random Telegraph Signal. xvii, 40, 41, 63, 71, 124, 126, 127, 133–136, 138

SC Slow Control. ix, xv, 60–62, 65, 67, 68, 71, 85, 86, 94, 101–103, 105, 106, 121, 122

SCT SemiConductor Tracker. 8, 9

SDR Single Data Rate. 91, 92

SEE Single Event Effects. 86

SEL Single Event Latchup. 86

SEU Single Event Upset. 65, 86

SGMII Serial Gigabit Media-Independent Interface. 94

SM Standard Model. 1, 3, 4, 6, 13, 15

SMU source measure unit. xvi, 117

SNR Signal to Noise Ratio. 37, 38, 55, 120

SPS Super Proton Synchroton. 4

SRAM Static Random Access Memory. 86, 87

STAR Solenoidal Tracker at RHIC. 2

STI Shallow Trench Isolation. 36, 42, 43

TCAD Technology Computer Aided Design. 62

TGC Thin Gas Chamber. 11

TID Total Ionising Dose. xii, xvi, xvii, 41–46, 55, 63, 71, 107, 110, 112, 114–116, 118–132, 135, 138

TileCal Tile Hadronic Calorimeter. 10

TJ TowerJazz. ix, xii, xiii, 2, 46, 54, 69–71, 73, 74, 88, 129

TRT Transition Radiation Tracker. 8, 9, 12

UK United Kingdom. 115

VCO Voltage Controlled Oscillator. 90

Chapter 1

Introduction

The *Standard Model* (SM) of particle physics is the most successful theory explaining the matter content of our Universe and its interactions. Particle accelerators play a critical role in the experimental study of this model. An example is the discovery of the Higgs boson at the *Large Hadron Collider* (LHC) by the *A Toroidal LHC ApparatuS* (ATLAS) and *Compact Muon Solenoid* (CMS) experiments in 2012 [1] [2]. The LHC will be upgraded to the *High Luminosity LHC* (HL-LHC) in 2025-2027. The HL-LHC upgrade will allow the LHC experiments to continue their physics programs at a higher luminosity and to collect a significantly higher amount of data. An introduction to the SM, the LHC, the ATLAS detector and its requirements for the HL-LHC upgrade is given in chapter 2.

Silicon-based detector technologies are employed in particle accelerator experiments for track reconstruction and particle identification. These detectors have to withstand the damage produced by highly energetic particles. Moreover, they have to be precise enough to distinguish the tracks of different particles. The increased luminosity of the HL-LHC exceeds the limits of the sensing technologies currently installed in the ATLAS experiment. ATLAS has started an R&D program to develop a detector technology that fulfils the requirements of the upgraded experiment. These requirements are especially demanding at the regions closest to the interaction point, in which the particle fluence is higher. Detectors have to maintain a good performance in terms of detection efficiency and readout time after several years of being exposed to radiation damage. Moreover, high spatial resolution is needed in order to distinguish individual particle tracks in a more dense environment. *Monolithic Active Pixel Sensors* (MAPS) are being considered as an option for the outer-most layer of the ATLAS Inner Tracker (ITk). This technology introduces the possibility to embed the pixel sensor and its readout electronics within the same Silicon bulk. This arrangement offers advantages in terms of material budget and power consumption. Moreover, MAPS are fabricated in commercial Complementary Metal-Oxide Semiconductor (CMOS) imaging processes lowering significantly the cost of this technology. However, traditional MAPS are susceptible to radiation damage as their charge collection mechanism is through diffusion. Thus, R&D is focused on developing a *Depleted MAPS*

(DMAPS) that collects charge through drift and can be commercially produced. A general introduction to Silicon detectors and DMAPS is given in chapter 3.

MAPS are already operated in a collider environment at the *Solenoidal Tracker at RHIC* (STAR) experiment of the *Relativistic Heavy Ion Collider* (RHIC). The *A Large Ion Collider Experiment* (ALICE) collaboration in the LHC has also developed a successful prototype in the TowerJazz (TJ) 180 nm technology which will be installed in the Inner Tracker upgrade of the ALICE detector. A test chip, named TowerJazz Investigator, was produced during the ALICE R&D process. This sensor was designed in the standard TJ 180 nm process, presented in chapter 4. The TJ Investigator includes several pixels with varying geometries dedicated to investigate the effect of the different characteristics on the signal. In this thesis, the Investigator was used to study the effect of the pixel geometry on the signal amplitude. This is presented in chapter 5.

Although the TJ 180 nm standard technology showed good performance for the requirements of the ALICE experiment, previous work has relegated its viability for the outer layers of the ATLAS ITk. In order to increase the resistance to radiation damage of the device, a new chip design including a process modification was implemented in a full ATLAS size prototype named MALTA. A description of the process modification and of the MALTA chip is given in chapter 4. One of the tests required to validate this technology for the new ATLAS ITk environment is to measure its performance after high doses of ionizing radiation. Tests were done at the University of Glasgow and are presented in chapter 7. The design presented issues in terms of electronic noise and detection efficiency, especially after receiving radiation damage. Following these results a smaller version of MALTA, named mini-MALTA, was designed implementing an improved front-end and different solutions to the efficiency loss issue. The Mini-MALTA chip is presented in chapter 4. The resistance to ionizing radiation of this design was validated at the University of Glasgow with successful results. These are presented in chapter 7.

The MALTA and mini-MALTA chips require the design of a hardware interface to connect the sensor to the Data Acquisition (DAQ) system. Field-Programmable Gate Arrays (FPGAs) are becoming a popular choice to implement these designs because of their reduced price and development effort as compared to Application Specific Integrated Circuits (ASICs). Xilinx FPGAs were chosen to develop the readout circuits of the MALTA and mini-MALTA chips. A description of both designs is given in chapter 6.

Chapter 2

Study of the Standard Model of particle physics

2.1 The Standard Model of particle physics

Particle physics aims to explain the Universe in terms of its most fundamental building blocks and their interactions. The most successful theory nowadays that is able to explain most of these phenomena is the Standard Model (SM). All the particles considered in this model are sketched in figure 2.1. One set of particles, known as *fermions*, are responsible for the matter content of

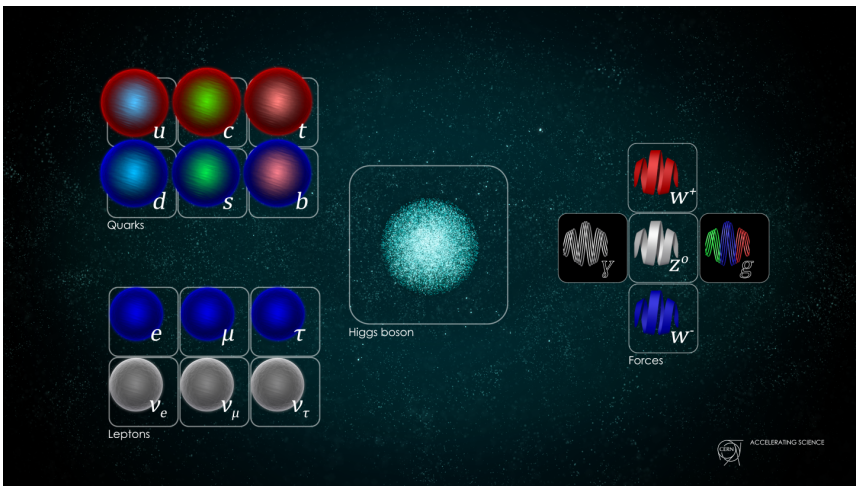

Figure 2.1: Particles included in the Standard Model of particle physics.

the universe. The other set, called *bosons*, are the mediators of their interactions.

There exist four known forces in nature that account for the interaction of fermions: the gravitational, the electromagnetic, the weak and the strong. Three out of the four forces in nature can be explained by boson interactions with the exception of gravity[1]. In this picture,

[1] It is speculated that there is also a mediating boson for gravity, called the graviton, in extensions of the SM.

photons (γ) are mediators of the electromagnetic force, eight different gluons (g) for the strong force and the $W^{\pm}$ and Z^0 for the weak force.

Fermions are subdivided in two groups: quarks and leptons. These groups, in turn, are subdivided in three families. The lepton group is formed by the electron (e), the muon (μ) and the tau (τ), together with their associated neutrinos ($\nu_{e,\mu,\tau}$). The e, μ and τ have an electrical charge of -1 while the neutrinos are chargeless. The family of quarks are the up (u) and down (d), the charm (c) and strange (s) and the top (t) and bottom (b). Quarks have a charge of either +2/3 or -1/3. Each charged lepton has its own antiparticle. These have the same quantum properties of their relative particle but with opposite charge.

Finally, the Higgs boson is a scalar boson predicted by Peter Higgs in 1964 [3] [4] [5] [6]. Vector bosons acquire their mass via spontaneous symmetry breaking with the Higgs field. Fermions and the Higgs boson itself also acquire mass by interacting with the Higgs field.

The Higgs boson was discovered at the Large Hadron Collider (LHC) in 2012 [1] [2]. The LHC constitutes one of the most important experiments devoted to the study of the SM. The generation of SM particles is produced by colliding protons at very high energies. These are detected by large scale particle detectors placed around the interaction points.

2.2 The Large Hadron Collider

The LHC is of special relevance in High Energy Physics (HEP) as it produces the highest energetic collisions in the world at the moment of this writing. It is situated at the European Laboratory for Particle Physics (CERN) between Lake Geneva and the Jura mountains. The LHC is a circular tube of 27 km that was constructed in a tunnel 90 m underground that was previously used for the Large Electron-Positron Collider (LEP) [7].

The LHC is employed to collide hadrons (protons or heavy ions) at very high energies. Protons are obtained by ionizing Hydrogen atoms. This proton plasma is separated in bunches that are injected into the accelerator complex shown in figure 2.2. Before reaching the LHC, particles are boosted by the LINAC 2 linear accelerator to 50 MeV and injected into the BOOSTER which accelerates them to 1.6 GeV. They are then transferred to the Proton Synchroton (PS) and accelerated to an energy of 26 GeV. These are then injected into the Super Proton Synchroton (SPS) and accelerated to 450 GeV. The 450 GeV protons are injected into the LHC where the bunches are split into two beams moving in opposite directions. Further acceleration occurs at a single location in the LHC. Eight Radio Frequency (RF) cavities oscillating at 400 MHz are set to boost protons up to an energy of 7 TeV. The oscillation frequency of the RF cavities is carefully designed to maintain the bunches at the desired energy. By contrast, protons arriving earlier or later than the expected arrival time are accelerated or decelerated accordingly.

The paths of the protons are bent by superconducting magnets that maintain the beam within the LHC. These magnets establish the upper limit of the center of mass energy ($\sqrt{s}$) of the

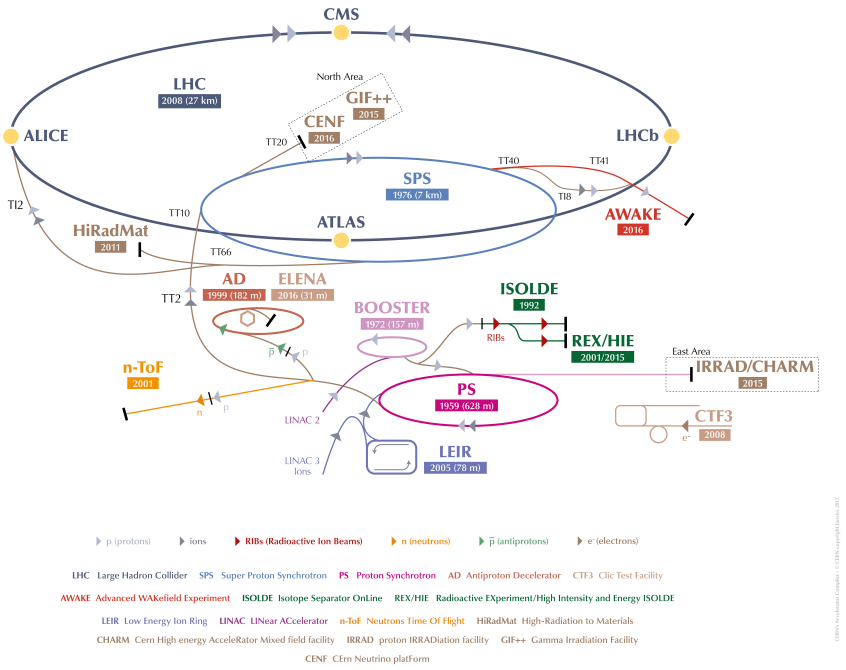

Figure 2.2: Sketch of the LHC (dark blue ring) and its main experiments. The smaller accelerators are used in a chain to help boost the particles to their final energies and provide beams to a whole set of smaller experiments, which also aim to uncover the mysteries of the Universe [8].

LHC to a maximum of 14 TeV, although the maximal operation energy at the moment is 13 TeV. The beams interact at four points, corresponding to the four experiments shown in figure 2.2: A Toroidal LHC Apparatus (ATLAS), Compact Muon Solenoid (CMS), A Large Ion Collider Experiment (ALICE) and Large Hadron Collider Beauty (LHCb).

The accelerator and detector systems at the LHC can handle a bunch crossing frequency of 40 MHz (every 25 ns). Each bunch contains around 1.15×10^{11} protons and has a dimension of 16.7 μm transversely and 7.55 cm longitudinally. The probability of proton-proton collision is given by the geometric cross-section (σ_c). The number of collisions per second is obtained by computing the factor $\mathscr{L}\cdot\sigma_c$, where $\mathscr{L}$, known as the instantaneous luminosity, is given by:

$$\mathscr{L} = \frac{N^2 n_b f_{rev}}{4\pi\varepsilon_n\beta^*}\gamma R \tag{2.1}$$

where ε_n is the normalized beam emittance, β^* a function related to the transverse size of the beam, $\gamma = E_{beam}/m_p$ is the relativistic factor and R is a geometrical reduction factor to correct for any collision angle. The integrated luminosity over a time period is a measurement of the amount of data recorded. Currently, the LHC records data at an instantaneous luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

The LHC still has not been operated at its design energy. A time-line of the LHC program is

shown in figure 2.3. After an initial incident (see [10]), the LHC was operated at $\sqrt{s} = 7$ TeV on

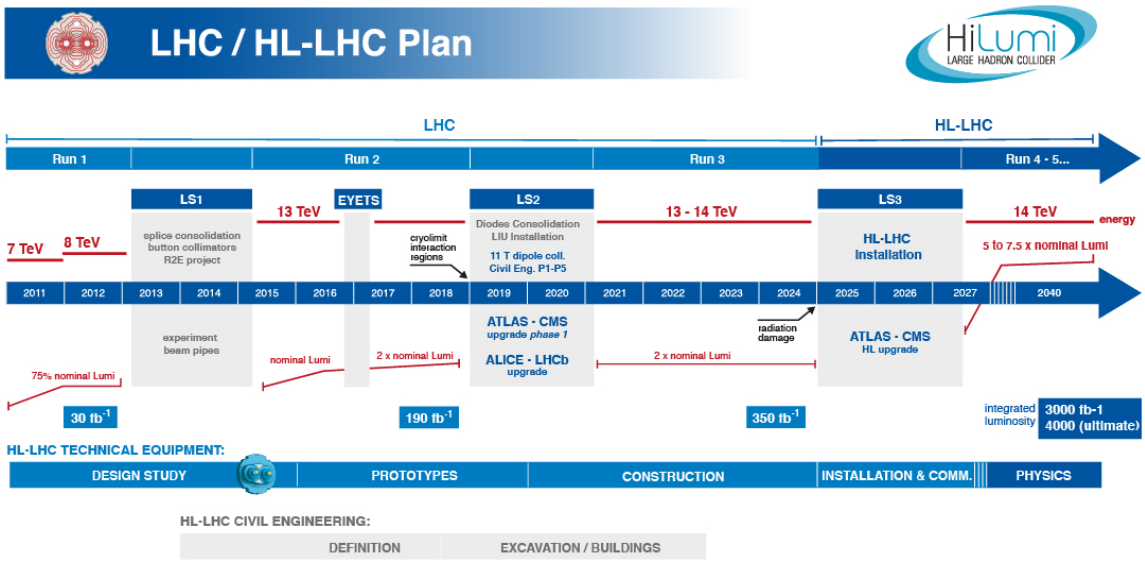

Figure 2.3: The LHC/HL-LHC schedule plan. Image reproduced from [9].

2011 and at $\sqrt{s} = 8$ TeV on 2012, at a reduced luminosity. After the Long Shutdown 1 (LS1), where reparation of the magnets took place, the LHC restarted its activity at the record-breaking center of mass energy of 13 TeV and at the nominal luminosity. Expectations beyond 2019 were that the data recorded at the nominal luminosity would not contribute significantly to increase the statistics. For this reason, the LHC will be upgraded to the High-Luminosity LHC (HL-LHC). Although some of the upgrade has been performed during the LS2, most of the work will be done on the LS3.

The integrated luminosity of the HL-LHC is expected to be increased by about an order of magnitude with respect to the one in the LHC. This will be achieved at a higher instantaneous luminosity, which will be increased by a factor of ~ 5. The increase in instantaneous luminosity will increase the pile-up of the LHC collisions from 25 to about 80 after the LS2 and to a final 200 in the HL-LHC era. Moreover, the increase in integrated luminosity translates into more radiation-induced damage on the four LHC experiments. The actual detector systems are not capable to resist this more challenging environment and, hence, have to be upgraded too. This thesis has been developed in the framework of ATLAS. In order to understand the improvements that need to be done on this experiment, a description of the ATLAS detector system is given in section 2.3.

2.3 The ATLAS detector system

The ATLAS detector system [11], depicted in figure 2.4, is one of the experiments in the LHC devoted to investigating the SM of particle physics. Its cylindrical shape has a diameter of

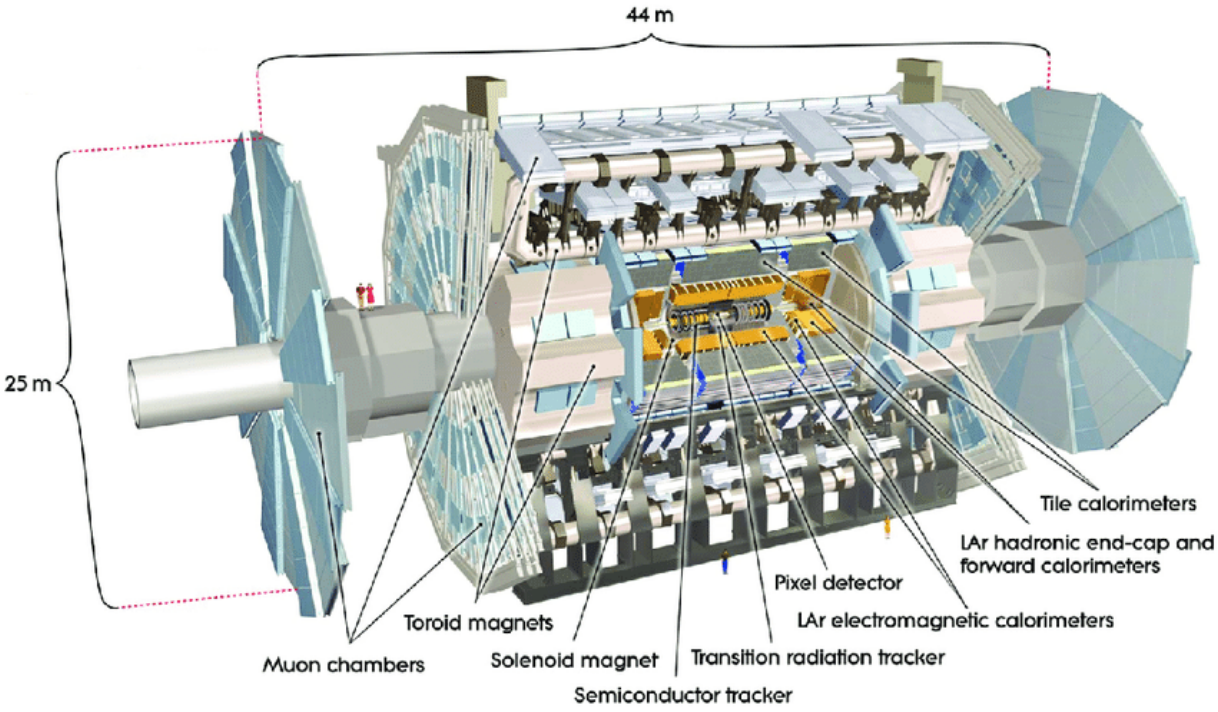

Figure 2.4: Overview of the ATLAS experiment. Image reproduced from [12].

25 m and a length of 44 m. It has multiple sub-detector systems distributed in layers. The LHC beam collides at the centre of the Pixel detector. This tracking system has a high granularity for the reconstruction of the particle track position and momentum that appear as a result of the collision. There are other layers of less precise semiconductor trackers which also help to reconstruct the particle movement. A list of the resolution and pseudorapidity[2] (η) coverage of the ATLAS trackers is given in table 2.1. The trackers are surrounded by calorimeter systems, in which the energy of most particles is measured. However, the energy of muons and neutrinos is not measured in the calorimeters and the first are detected in the muon Spectrometers, located outside the calorimeter. A more detailed discussion of the different sub-detector systems is given in the following sections.

2.3.1 The Inner Detector

The ATLAS Inner Detector (ID) is the detector module closest to the interaction point. Requirements at this location are to reconstruct the tracks of charged particles with the highest resolution. This precision has to be maintained even after large integrated particle fluence that results in a degradation of the performance of the sensors. A 2 T solenoid magnetic field is

[2]The pseudorapidity is a spatial coordinate commonly employed in particle physics to describe the angle between a particle momentum and the direction of the beam axis. It is defined as : $\eta \equiv -\ln[\tan(\theta/2)]$ with θ being the angle between the positive direction of the beam and the particle momentum.

System	Position	Area (m^2)	Resolution (μm)	η coverage
Pixels	Insertable B-layer	-	$R\phi$ = 10, z = 60	$\pm$3
	Removable B-layer	0.2	$R\phi$ = 12, z = 66	$\pm$2.5
	2 barrel layers	1.4	$R\phi$ = 12, z = 66	$\pm$1.7
	5 end-caps per side	0.7	$R\phi$ = 12, R = 77	1.7 - 2.5
Silicon strips	4 barrel layers	34.4	$R\phi$ = 16, z = 580	$\pm$1.4
	9 end-caps per side	26.7	$R\phi$ = 16, R = 580	1.4 - 2.5
TRT	Axial barrel straws	-	170 (per straw)	$\pm$0.7
	Radial end-cap straws	-	170 (per straw)	0.7 - 2.5

Table 2.1: Summary table of the resolution and η coverage for the different detector systems of the ATLAS Inner Detector. In the ATLAS co-ordinate system, the z-axis denotes the direction of the beam-line, R represents the radial distance to the beam, ϕ is the azimuthal angle about the z-axis.

employed to bend the trajectory of particles, allowing their momentum to be measured.

The ATLAS ID is organised in three layers of different detector technologies, as shown in figure 2.5. The Inner-most subdetector is made of Pixel Detectors, followed by the SemiConductor Tracker (SCT) [13] and the Transition Radiation Tracker (TRT) [14]. The Pixel Detector

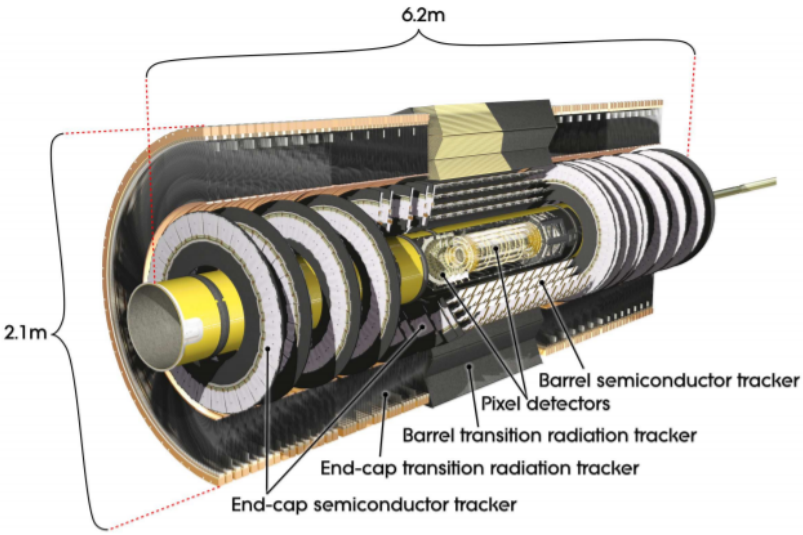

Figure 2.5: Cut-away view of the ATLAS Inner Detector. Image reproduced from [11].

consists of three barrel layers situated at 5.1, 8.9 and 12.3 cm from the beam-line with a length of 80 cm. Three end-cap disks are also placed at both ends of the pixel barrels. The detector technology mainly employed in this module is *hybrid* planar. In the hybrid technology, sensors are connected to a readout chip via tiny metal bumps, in a process called *bump-bonding*. In ATLAS, standard n-in-n planar sensors with an area of $50 \times 400\,\mu m^2$ and a thickness of $250\,\mu$m are bump-bonded into FE-I3 readout chips [15]. The spatial resolution achieved by pixel sensors

is given by the pitch/$\sqrt{12}$. Hence, spatial resolutions of around 115 μm are possible with these detectors. Sensors are operated at a reverse bias voltage of 150 V, which has to be increased up to 600 V after receiving radiation damage. Aside from the original pixel barrels, an extra inner layer, known as Insertable B-Layer (IBL) [16], was installed during the *Long Shutdown* 1 (LS1) on 2014. Being placed at 3.2 cm of the beam-pipe, its requirements in terms of radiation hardness and spatial resolution are even higher. Planar sensors provide 60 % of the ATLAS IBL, while the rest is made of novel n^+-in-p 3D sensors. The pixel size is reduced to $250 \times 50\ \mu m^2$ and their thickness varies from 200 μm in the planar sensors to 230 μm in the 3D. Similarly to the original Pixel Detector barrels, detectors in the IBL are bump-bonded with the FE-I4 readout chip [17]. The IBL detector is expected to resist the luminosity accumulated until the HL-LHC upgrade, in which the whole ID will have to be upgraded. The pixel modules installed in the end-cap disks are identical to those in the barrels. The six end-cap disks are concentric and perpendicular to the beam-line and at radii between 11 and 15 cm.

The SCT is similar to the Pixel Detector in terms of structure and technology. Four concentric barrels of 149 cm length at a radii of 30, 37.3, 44.7 and 52 cm and and nine end-cap discs form the SCT. The detector technology employed in this module is p-in-n silicon micro-strips. In the barrel layers, strips have 80 μm pitch and a length of 12 cm. 2D reconstruction is achieved by gluing two identical detectors rotated by 40 mrad. A precision of 17 μm in the transverse direction and 580 μm in the longitudinal are achieved with this configuration. The nine end-caps are located at a radii between 27.5 and 56 cm of the beam-line. The end-cap modules are analogous to the barrel modules but organised in concentric disks. The tracking precision achieved in the end-cap modules is the same than that of the barrels. Sensors in the SCT are biased at 150 V before irradiation and up to 350 V after being damaged by radiation.

The TRT is made of gas detectors. A total of 52544 axial straw-tubes with a diameter of 4 mm conform the 1.44 m length barrel that surrounds the SCT detector. The tubes are filled with a composition of Xenon, Carbon dioxide and Oxygen, in a 70/27/3 ratio. A gold-plated tungsten wire placed in the center of the tube serves as the charge collection anode of the sensor. The potential difference applied between the anode and the tube is of 1500 V. Particles crossing the tubes ionize the gas atoms creating an electron shower, which is multiplied at the boundaries of the anode creating a measurable signal. The precision achieved by these trackers is of 130 μm. There are also two end-cap disks made of 122880 straws parallel to the beam-axis.

2.3.2 Calorimeters

Calorimeters are employed in HEP experiments to measure the energy of particles. Most of the particles are fully stopped in this detector, depositing their energy and generating a proportional signal. Exceptions are muons and neutrinos. Calorimeters in the ATLAS detector consist of interleaving layers of a high density absorbing material and an active medium. Particles undergo what is known as a *particle shower* in the absorber material. A shower consists in the interaction

of the initial particle with the atoms of the absorber leading to other particles which, in turn, will interact again. The final result is a "shower" of electrons that ionise the active material generating a measurable signal. This shower is periodically sampled by the active material of the calorimeters. There are two calorimeter systems in ATLAS (see figure 2.6): the Liquid Argon calorimeter (LAr) [18] and the Tile Hadronic Calorimeter (TileCal).

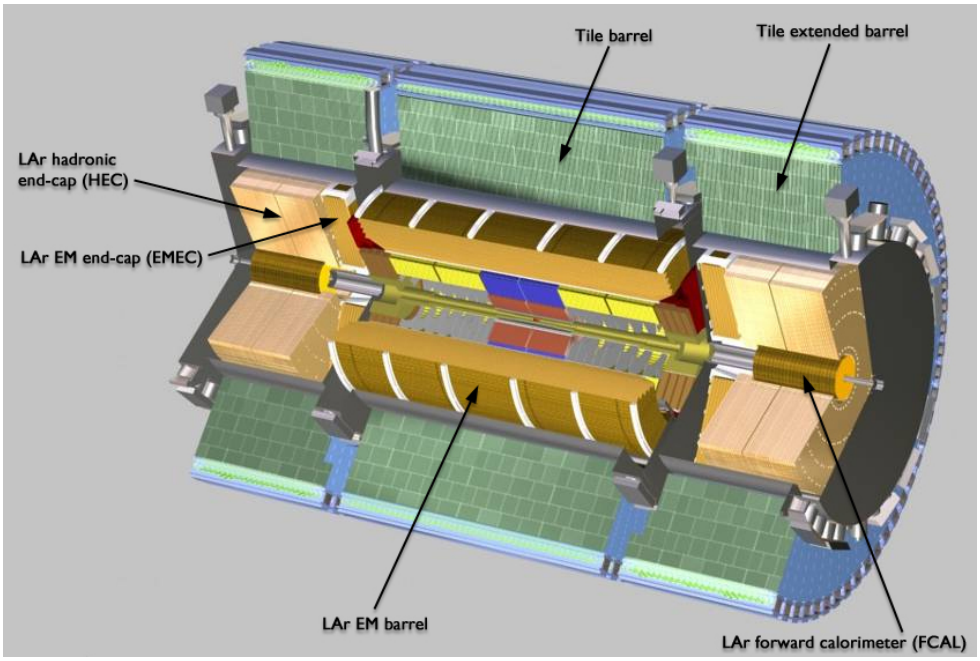

Figure 2.6: Cut-away view of the ATLAS calorimeter system. Image reproduced from [8].

The LAr is the closest calorimeter to the interaction point and is used to measure the energy of electrons and photons. In total, it has a thickness of ~ 22 radiation lengths. Its barrel is 6.4 m long and 53 cm thick and has 110000 cells to locate the shower. The materials used in the LAr calorimeter are lead as absorber and liquid Argon as the active material. The layers of liquid Argon and lead are arranged in an accordion-like shape. The gas is contained in a honeycomb structure that also supports the readout boards.

The LAr end-cap comprises the LAr ElectroMagnetic End-Cap (EMEC), the LAr Hadronic End-Cap (HEC) and the LAr forward calorimeter (FCAL). The EMEC structure is equivalent to the one of the LAr barrel. It consists of two concentric wheels with the same accordion-like distributed layers made of lead and liquid Argon. Instead, the HEC uses Copper as the absorber medium and has flat layers. It is structured in two wheels of 2 m radius and approximately 1 m thick. Finally, the FCAL is composed by three wheels of radius 0.455 m and a thickness of 0.450 m. Copper is again used as the absorber medium while detection is still done with liquid Argon.

The TileCal consists of a central barrel of 5.56 m length and two barrel extensions of 2.91 m both with a thickness of ~ 7 interaction lengths. Each module is segmented into 64 wedge-like structures containing steel plates, that act as the absorbing material, and scintillator tiles as the

active medium. Charged particles are detected in the scintillators by generating photons that, in turn, are detected by photomultipliers. The hadronic end-cap region is already covered by the HEC detector mentioned above.

2.3.3 The Muon spectrometer

The muon spectrometer, depicted in figure 2.7, is the outermost layer of the ATLAS detector system. It is designed to identify muons and precisely measure their momentum. In order to do so, four sub-detector systems are employed: the Monitored Drift-Tube (MDT), the Cathode Strip Chambers (CSC), the Resistive Plate Chambers (RPC) and the Thin Gas Chamber (TGC). These detectors are distributed in three cylindrical barrel layers and two end-cap wheels covering

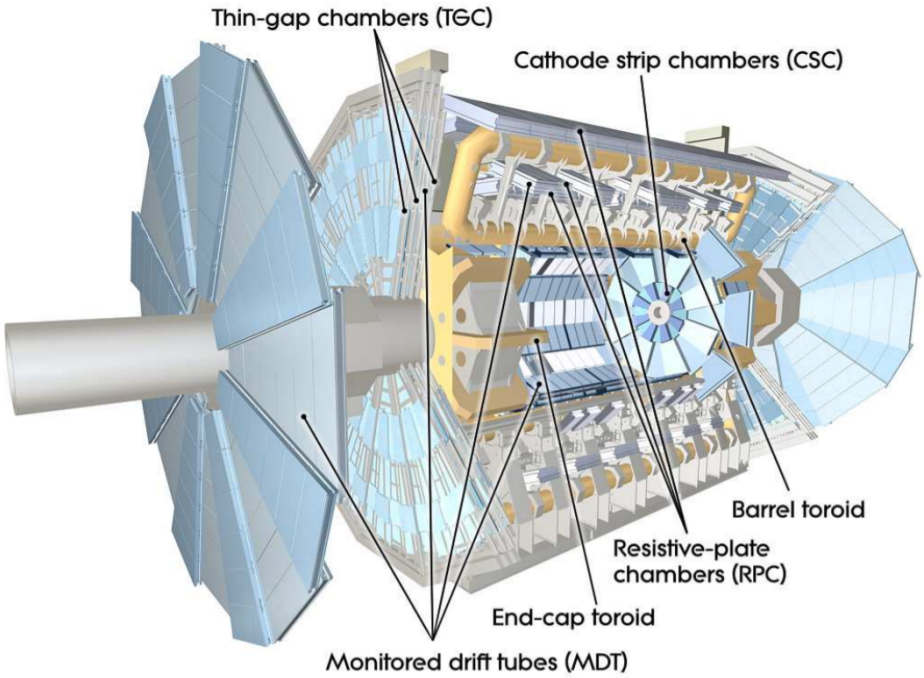

Figure 2.7: Cut-away view of the ATLAS Muon Spectrometer system. Image reproduced from [11].

different angular regions. The RPC and TGC serve as trigger systems for the MDT and CSC tracking chambers. The muon momentum is inferred from the bending of the particle under the effects of the toroidal magnet field.

2.3.4 ATLAS ITk upgrade for the HL-LHC

The LHC will undergo a phase II upgrade during the LS3 in which its instantaneous luminosity will be increased by a factor of ten. In the HL-LHC era the ATLAS detectors are expected to receive unprecedented pile-up and particle fluences. Around 200 proton-proton collisions per bunch-crossing are expected to occur with the objective of increasing the amount of data

collected from $\sim 400\,\text{fb}^{-1}$ to $\sim 4000\,\text{fb}^{-1}$ by the end of the HL-LHC program. This increase in particle fluence translates into an increase of radiation damage, especially in the regions closer to the interaction point. Figure 2.8 presents a simulation of the radiation damage expected in the ID region during the HL-LHC operational life-time, measured in 1 MeV neutron-equivalent damage. The ATLAS ITk is expected to receive $1.4 \times 10^{16}\,1\,\text{MeV}\,\text{n}_{eq}/\text{cm}^2$ over ten years of

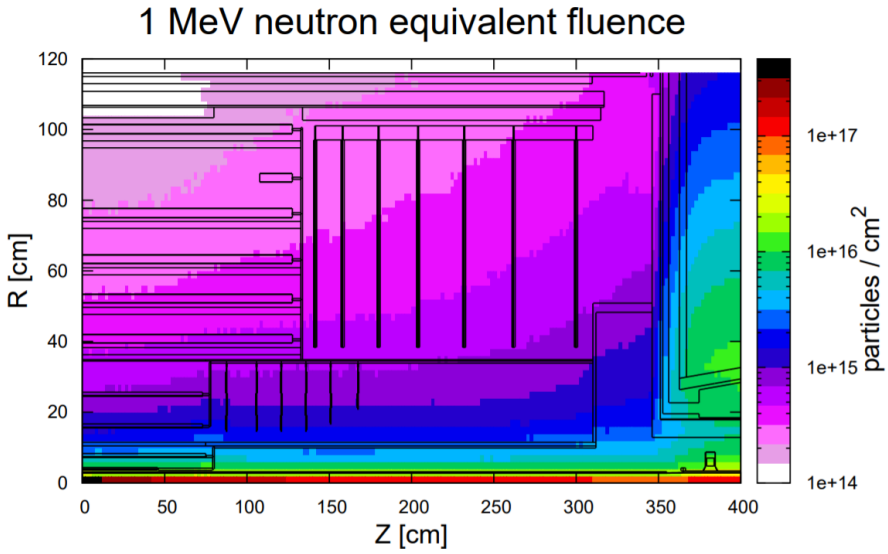

Figure 2.8: RZ map of the expected 1 MeV neutron-equivalent fluence received by the ATLAS ITk detector over ten years of operation at $\sqrt{s} = 14\,\text{TeV}$. Image reproduced from [19].

operation.

The current detector technology of the ATLAS ID is not prepared to resist this harsher environment and, hence, has to be substituted. Moreover, the larger pile-up is expected to saturate the TRT detector in terms of occupancy, demanding a more precise detector system. To meet these demands, the ATLAS ITk will be replaced by a fully Silicon-based detector system.

Several pixel technologies are being considered to replace the current ATLAS ITk: 3D, high resistivity planar and high voltage Complementary Metal-Oxide-Semiconductor (CMOS). Contrary to the planar case, in which the electrodes are implanted in the surface of the device, in the 3D technology the electrodes are distributed in columns that penetrate the sensor bulk perpendicularly to its surface. Typically, the distance between columns is shorter than the detector thickness leading into a shorter travel distance of the charge carriers generated by traversing particles (see section 3.8). This results into faster signals and increased radiation hardness. In the CMOS technology, the readout electronics required to amplify and shape the signal due to a particle hit are implanted in the sensing material. This is an advantage as compared to the 3D and planar technologies, in which the readout electronics have to be designed in a separate board and connected to the sensing material through a process called *bump-bonding*. CMOS detectors also have the advantage of being a relatively cost-effective technology as they do not

require the bump-bonding process and can be produced in industry. Hence, this technology is particularly interesting to cover the larger area of the outermost pixel layer of the ATLAS ITk. The performance requirements at this pixel layer are summarised in table 2.2.

Requirement	Value
Detection efficiency	$> 97\,\%$
Time resolution	25 ns
Particle Rate	$1\,\mathrm{MHz/mm^2}$
Non-ionising radiation fluence	10^{15} 1 MeV $\mathrm{n}_{eq}/\mathrm{cm}^2$
Ionising radiation dose	80 MRad
Power Consumption	$< 500\,\mathrm{mW/cm^2}$
Material budget	$< 2\,\%$ of x/X_0

Table 2.2: Main performance requirements for the pixel detectors to be installed in the outermost layer of the ATLAS ITk. The radiation damage requirements have been multiplied by a safety factor of 1.5. Table reproduced from [20].

Within the context of this thesis, CMOS detectors designed in novel processes are studied. The basic principles of semiconductor detectors and CMOS sensors in particular are reviewed in chapter 3.

2.4 Summary

The Standard Model of particle physics is the most prominent theory explaining the content of the universe and its interactions. The theory states that the matter content of the universe is formed by a set of particles called fermions, while their interactions are mediated by a set of particles known as bosons. The Higgs boson is the last particle to be predicted by Peter Higgs within the SM and was discovered at the LHC in 2012. The LHC is a particle collider consisting in a ring of 27 km employed to accelerate and collide protons or heavy ions at very high energies to produce and study the different particles of the SM. In order to do so, four detector complexes are placed around the four interaction points of the accelerator. This thesis has been developed in the context of one them: ATLAS. The ATLAS detector consists of different sub-detector systems placed in an onion-like shape, in multiple layers. The inner-most detector system, the ID, is made of eight layers of highly precise solid state detectors and one layer of a gas detector. The ID is devoted to reconstruct the particle-track positions and measure their momenta. This tracker system is surrounded by two calorimeters: one to measure the energy of electrons and photons and the other to measure the energy of the rest of charged particles. The last detector system consists of a muon spectrometer, employed to identify and measure the momentum of muons.

In 2025-2027, the LHC will undergo an upgrade to increase the number of particle collisions with the objective of collecting more data and obtaining more precise measurements. This

increase in particle rate has implications for the ATLAS detector, especially for its most inner layers in which the particle fluence is higher. Detectors will receive an unprecedented radiation damage that the current detector technologies are not capable to resist. In fact, the whole ATLAS ITk will be replaced by a fully Silicon-based detector system. Several sensor technologies are being considered to substitute this detector. One of the proposed solutions is the CMOS technology, which has the advantage to embed the readout electronics and the sensing material in the same substrate, removing the complex bump-bonding process. This, together with the fact that this technology is produced in industry in standard processes, makes this technology relatively cost-effective and very attractive for the outer-most layers of the ATLAS ITk. However, standard CMOS detectors are not radiation hard enough to resist the 80 MRad of ionising radiation damage and the 10^{15} 1 MeV n_{eq}/cm^2 of non-ionising radiation damage expected after ten years of operation at the outer-most layer of the ATLAS ITk. In this thesis, CMOS detectors fabricated in novel processes are tested with the objective to validate this technology for the HL-LHC environment.

Chapter 3

Introduction to Silicon Detectors

3.1 Introduction to Semiconductor Detectors

Particle tracking and identification is critical to investigate physical phenomena in High Energy Physics (HEP) experiments. Tracking technologies have evolved significantly in the last 50 years, from multi-wire proportional chambers invented in 1968 [21] to the current semiconductor-based sensor technologies. The major achievements in using silicon-based sensors are an increased spatial and energy resolution. Wire chambers have a typical spatial resolution of the order of mm, while Silicon sensors installed at the LHC can resolve tracks with the remarkable precision of 10 μm [22]. These improvements allowed to resolve tracks at unprecedented high particle densities and rates such as the ones in the LHC. Moreover, it allowed to resolve the interaction vertex (*primary vertex*) and a possible decay vertex (*secondary vertex*) of long living particles allowing to test particle decays described in the SM of particle physics.

The principle of operation of semiconductor detectors is analogous to that of ionization chambers. In ionization chambers, a gaseous medium is placed in between two metal plates with an applied voltage as shown in figure 3.1. When an ionizing particle crosses the medium, it generates electron-ion pairs that move in the presence of the electric field inducing a current on an external circuit. The attached electrodes can be segmented in different geometries to enhance the position resolution of the sensor, *pixels* [23] and *strips* [22] being the most commonly employed.

In the case of semiconductor detectors, the ionized material is a solid state semiconductor. The advantages offered by solid materials as compared to gases are their high density and low ionising energy (3.6 eV in Silicon vs >30 eV in gas). This translates to a higher number of generated charge carriers per traversed distance, allowing to fabricate very thin detectors. The charge generated by a *minimum ionizing particle* (MIP) in Silicon is 73 e-h/μm [24] which corresponds to a 4 fC signal in a typical 300 μm thick detector. This small signal is usually amplified and shaped by the readout electronics.

Among all the existing semiconductor materials, Silicon is the most widely used in sensor

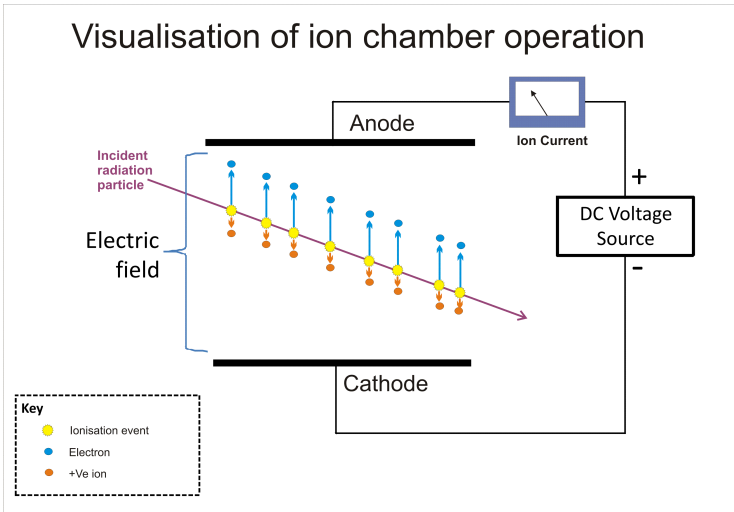

Figure 3.1: Simplified schematic of a ionization chamber.

fabrication for industrial and research applications. This is due to its abundance, as it constitutes 27.7 % of the Earth's crust mass [25], and its electrical properties (see sections 3.2 to 3.6). Furthermore, the extensive use of this material has led to a greater understanding of its properties and has lowered the costs of the manipulation and extraction processes.

The high success of Silicon-based sensing technologies has led to the generation of a wide range of detector prototypes and its development is still undergoing. The increase in luminosity expected in the HL-LHC upgrade motivated the creation of novel technologies more precise and resistant to radiation damage. Special R&D effort has been put towards the maturation of radiation hard technologies to be installed in the ATLAS ITk. In this thesis, the radiation hardness of novel CMOS detectors is studied. The basics of semiconductor detectors, the CMOS technology and the mechanisms of radiation damage are reviewed in this chapter. In section 3.2 the basic structure of Silicon is discussed.

3.2 Crystal structure

Crystals are characterised by a three-dimensional periodic pattern in their atomic distribution. This pattern is known as the *crystal lattice* and is related to some properties of the material. Electronic devices, such as radiation detectors, are commonly made of Silicon in a crystalline form. An isolated Silicon atom contains 10 electrons tightly bound to the nucleus, which forms the *core* of the atom, and 4 electrons that are more weakly bound and participate in chemical interactions, also known as the *valence electrons*. If several Silicon atoms are brought together, these valence electrons are responsible for an electrical bond with its four nearest neighbours forming the crystal structure. The lattice cell of a Silicon crystal is shown in figure 3.2. It has a cubic shape with an edge distance of 5.43095 Å [26]. Each cell contains 8 Si atoms in the corners (shared by 8 cells), 6 in the faces (shared by 2 cells) and 4 inside the cube (not shared).

Therefore, the atomic density of Silicon can be calculated:

$$\begin{aligned} n_{\mathrm{a}} &= \left(\frac{8}{8}+\frac{6}{2}+4\right)\cdot\frac{1}{(5.43095\times10^{-8})^3} \quad \left(\text{ atoms }/\mathrm{cm}^3\right) \\ &\approx 5.05\times10^{22} \quad \left(\text{ atoms }/\mathrm{cm}^3\right) \end{aligned} \tag{3.1}$$

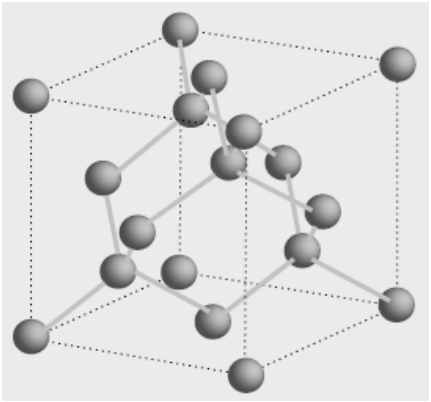

Figure 3.2: Structure of a Silicon lattice cell.

Although the real atomic structure of a Silicon crystal is the one shown in figure 3.2 it is usually more convenient to use a simplified model in order to visualise some structural properties of the crystal. This simplified model is known as the *bonding model*. In the bonding model, circles represent Si-atoms and lines correspond to the bonds between them, as shown in figure 3.3. It is important to note that the perfect structure shown in this figure is only achieved in a perfectly pure Silicon material at $T = 0\,K$. In a more realistic situation, valence electrons are thermally excited, breaking the atomic bonds. Moreover, crystals might have some imperfections in their periodic shape. Those imperfections are called *lattice defects* and might or might not be intentional. The bonding model will be used in the next sections to describe the effects of intentional defects.

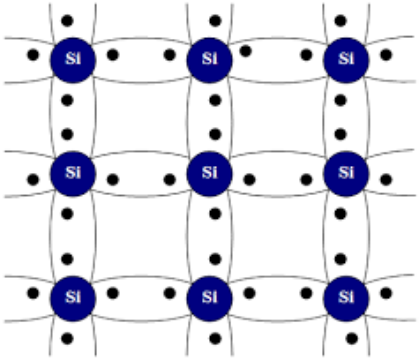

Figure 3.3: The bonding model for a Silicon crystal lattice.

3.3 Energy bands

Semiconductor detectors work by generating an electrical signal when a charged particle crosses them. This current is generated by valence electrons that have been given enough energy to break the covalent bond between atoms. In an ideal particle detector only external radiation should be the cause for exciting electrons into conduction energy states. However, in reality, electrons are also subject to thermal excitations. Hence, it is important to study the energy level distribution of electrons inside a material, since this is related to its capability to detect particles and to minimize thermal noise. This section is devoted to review the energy level distribution of crystalline solids and to justify the reason why Silicon is a good material for particle detection in terms of its electrical properties.

The electrical behaviour of a crystal is determined by the energy level distribution of electrons within it. As mentioned in section 3.2, an insulated Silicon atom contains 14 electrons: 10 that compose the core of the atom and are unperturbed in atomic interactions and 4 that occupy the energy states immediately after the core and participate in chemical interactions. In a crystalline solid, where atoms interact with their neighbours, it is expected that the electron wavefunction will be modified and, hence, its energy distribution. This modification does not affect the core electrons, since those are tightly bound to the atom nuclei. On the other hand, the wavefunction of valence electrons overlaps with the one of the other electrons in the material and becomes periodic, as predicted by the *Bloch theorem* [27]. By solving the energy level distribution of this new electron state, R.L. Kronig and W.G. Penney [28] found that the energy of valence electrons in a crystal is distributed in almost continuous regions (or *energy bands*) separated by a finite interval (*energy band gap*). The most energetic band is called *conduction band* while the less energetic is the *valence band*. A simplified diagram illustrating the energy band distribution is shown in figure 3.4. In contraposition with the isolated atom case, these new energy levels do not correspond to any atom in particular, but to the material as a whole. This means that electrons do not have a fixed position inside the material. Although this also applies to electrons belonging to the valence band, as a mean there are four electrons per atom forming the covalent bond between Si-atoms, as illustrated in figure 3.3. The valence electrons that are promoted to the conduction band are called *conduction electrons*. Due to the high density of free states in the conduction band, they can move more freely through the crystal lattice and, in the presence of an electric field, they generate a current.

When an electron is excited to the conduction band, a vacancy is left in the valence band. This vacancy is known as *hole* and can move within the material behaving like a pseudo-particle. Furthermore, a mass and an electrical charge can be associated to these entities. The charge that is associated to holes is the opposite of the electron charge. Holes and electrons are responsibles of charge transport in crystal materials and, hence, are also known as *charge carriers*.

It is important to mention that the band gap structure is present in any crystal. The difference between materials relies in the magnitude of its energy gap. Hence, electrons require different

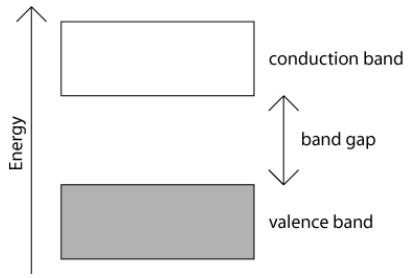

Figure 3.4: Simplified energy band diagram.

energies to traverse from the valence to the conduction band depending on the material, which affects its electrical conductivity. This fact allows us to classify materials in the following way according to the magnitude of their band gap:

- **Insulators:** Wide band gap (e.g. 5.45 eV for diamond [29]). Electrons at room temperature have an average energy of about 0.025 eV. Hence, at room temperature very few electrons are excited to the conduction band, which leads to a poor electrical conductivity.
- **Conductors:** Materials with a very little or even overlaping band gap. A large number of electrons occupy the conduction band even at low temperatures, leading to a good conductivity.
- **Semiconductors:** Middle term between insulators and conductors. A moderate number of electrons are excited to the conduction band.

Figure 3.5 illustrates the three different situations mentioned above.

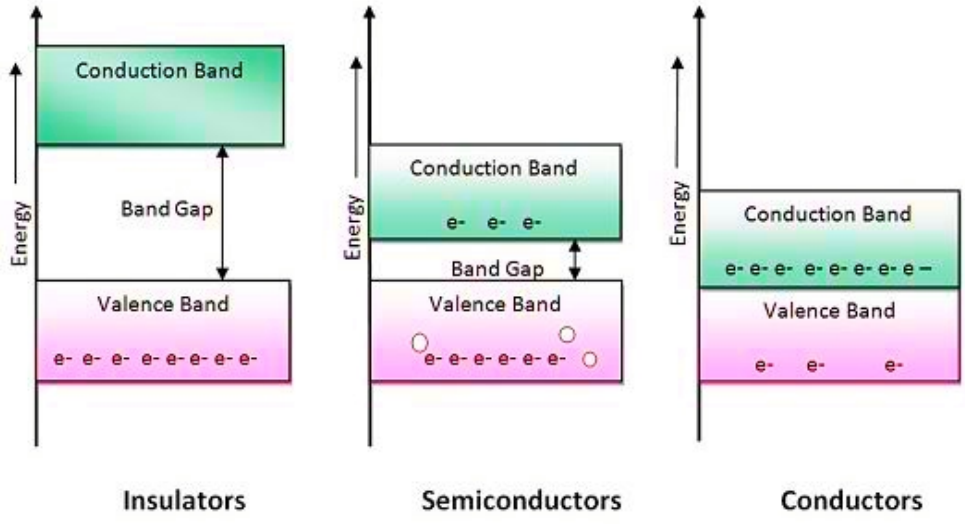

Figure 3.5: Energy band diagram for insulator, semiconductor and conductor materials.

Silicon has an energy gap of 1.12 eV [30] and is classified as a semiconductor. The mean energy required to ionize a single electron-hole pair differs from the band gap energy by a

factor of ~ 3 [31]. Figure 3.6 shows the pair-creation energy of different materials against their theoretical band gap. This difference is because part of the energy of an incoming particle is thermalized by phonon emission.

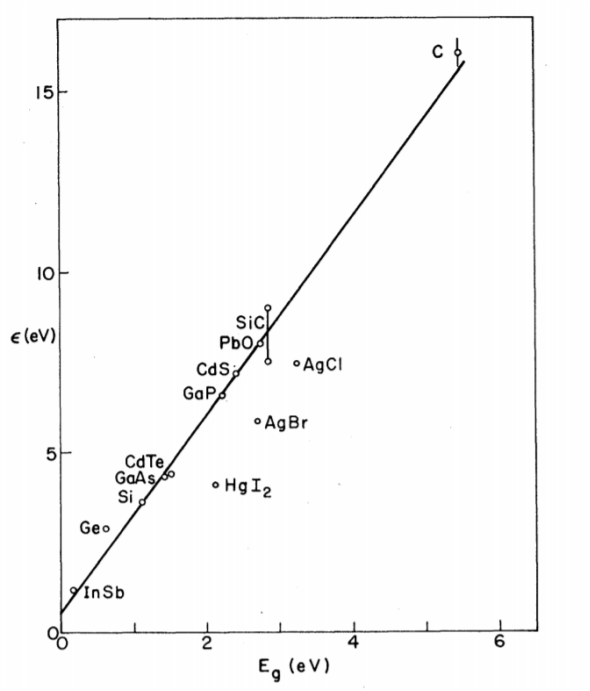

Figure 3.6: Empirically measured e-h pair creation energy (ε) versus the material energy band gap (E_g) for different semiconductor materials. ε for Silicon at room temperature is 3.6 eV. Image reproduced from [31].

The pair-creation energy of Silicon is 3.6 eV. Due to its relatively low pair creation energy, a large number of electrons can be excited to the conduction band due to incoming radiation, while not being conductive as in the case of conducting materials. This property makes Silicon, and semiconductors in general, a suitable material for particle detection. Another feature that makes Silicon a great material for particle detection, and more generally for electronic circuits, is the possibility to introduce impurities with the purpose of changing its conductivity. This effect is explained in section 3.4.

3.4 Semiconductors

Semiconductors are characterized by an intermediate energy band gap between that of insulators and conductors. Due to the relatively low energy required to excite an electron from the valence to the conduction band, semiconductor materials provide a large current per energy deposited by radiation if compared to insulator materials. However, electrons are also excited due to thermal interactions leading to a non zero number of charge carriers in the device. It is important to control this thermal charge carrier concentration, which is inversely proportional to the resistivity

of the material [32]. Large carrier concentrations lead to low resistivities which, in turn, lead to a large current under the influence of an external voltage. Such high currents and, hence, high current fluctuations have to be avoided in order to detect the $80\,e/\mu m$ that a typical particle deposits.

The amount of free charge carriers in a semiconductor material depends on the concentration of impurities that it contains. Adding impurities in a semiconductor is a technique (called *doping*) extensively used in semiconductor devices to control the number and type of charge carriers that a material contains. Perfectly pure semiconductors (or *intrinsic*) and doped semiconductors (or *extrinsic*) are discussed in the following sections.

3.4.1 Intrinsic Semiconductors

Intrinsic semiconductors are semiconductor materials without any or a negligible amount of impurities in their crystal lattice. Hence, the whole material can be represented by the bonding model of figure 3.3. The energy of electrons follows a Fermi-Dirac probability distribution, $F(E)$, given by the following equation [33]:

$$F(E) = \frac{1}{exp\left(\frac{E-E_f}{k_B T}\right)+1} \tag{3.2}$$

with E_f being the Fermi energy level, k_B the Boltzmann constant and T the absolute temperature.

From equation 3.2, the temperature dependence of the electron energy can be deduced. At $T = 0\,K$, the probability for an electron to have an energy $E > E_f$ is 0 and 1 if $E < E_f$. In intrinsic semiconductors E_f is at the middle of the energy band gap. Hence, at temperatures close to 0 K all electrons belong to the valence band, and there are no free charge carriers in the material. As the temperature increases, the probability that an electron occupies a conduction band energy state becomes higher. As explained in section 3.3, when an electron is excited to the conduction band a hole is created in the valence band. Thus, the concentration of conduction electrons and valence holes is the same in intrinsic semiconductors. The density of either electrons or holes (n_i) increases with temperature according to the following relationship [33]:

$$n_i = (N_c N_V)^{1/2} e^{-\frac{E_g}{2k_B T}} \quad \left(\text{ electrons or holes } /\text{cm}^3\right) \tag{3.3}$$

where,

$$N_{C,V} = 2\left(\frac{2\pi m^*_{n,p} k_B T}{h^2}\right)^{3/2} \tag{3.4}$$

In this equation, $m^*_{n,p}$ are the effective masses of electrons and holes respectively. The values of these effective masses are related to the charge carrier mobility inside the semiconductor and are temperature and material dependent. For Silicon at room temperature $m^*_n = 1.18\,m_e$ and $m^*_p = 0.81\,m_e$ [34] with m_e being the electron rest mass. Hence, the intrinsic charge carrier den-

sity at $T = 300\,K$ is $n_i \simeq 1 \times 10^{10}$ (e or h)/cm^3. Considering that there are 4 valence electrons per Si atom and that the Si density is 5.05×10^{22} atoms /cm^3 (see derivation in equation 3.1), this corresponds to 1 broken bond every 10^{13} atoms. Although the carrier density is small as compared to the number of Si atoms, it leads to a resistivity of $\rho = 400\,\mathrm{k}\Omega \cdot \mathrm{cm}$ [24]. Furthermore, due to crystal imperfections and small impurity concentrations, maximum resistivities of $20\,\mathrm{k}\Omega\,\mathrm{cm}$ are available in commercial Silicon wafers [24]. For a typical detector size of 300 μm thick and 1 cm^2 area this gives a resistance of 600 Ω. When applying a typical voltage of ~ 20 V this leads to currents of the order of 1 mA. This current scale has to be compared to the typical current of around 200 nA generated by a traverssing particle in a 300 μm thick detector. Thus, the low signal to noise ratio ($\sim 2\times10^{-4}$) in intrinsic Silicon semiconductor makes it unsuitable as a material for particle detection.

The free charge carrier concentration can be reduced by using a junction of p and n-type silicon. These materials are explained in section 3.4.2.

3.4.2 Extrinsic Semiconductors

Extrinsic semiconductors are intrinsic semiconductor materials whose properties are modified by adding impurities. Impurities are intentionally added to increase the concentration of either positive or negative charge carriers. This process is called *doping* and it is extensively used in the fabrication of semiconductor devices. An extrinsic semiconductor with an excess of holes is called *p-type* while if it has an excess of electrons it is known as *n-type*. In order to understand the effect of impurities on the charge carrier concentration it is useful to recall the bonding model explained in section 3.2. Impurities replace the position of Silicon atoms in the bonding model that are bonded by 8 valence electrons to the neighbouring atoms. The manipulation of charge carrier concentration relies on the number of valence electrons that the impurity elements have. Both cases will be discussed seperately.

- **N-type semiconductors**: Elements of the Column V (like phosphorus) of the Periodic Table are added to produce n-type semiconductors. These elements contain an extra valence electron compared to the Si atoms. Four valence electrons are used to form the usual bond between atoms and the extra valence electron is only loosely bound to the atom and is easily excited due to thermal interactions to the conduction band[1], generating a negative charge carrier in the material and a positively ionized atom. Since the Column V elements "donate" an electron to the Silicon crystal, these elements are also known as *donors*. N-type semiconductors are represented in figure 3.7 (left).

- **P-type semiconductors** In p-type semiconductors the impurities added to increase the number of holes are elements from the Column III in the Periodic Table of Elements (most

[1] excitation energy is 45 meV for Phosphorus

commonly Boron). These elements have one valence electron less than Si and, hence, an extra electron is needed to complete the covalent bond with the neighbouring Si atoms. This electron is taken from the conduction band and it creates a positively charged hole in the process that can move freely through the crystal lattice leaving a negatively charged atom. In this case, Column III elements "accept" an electron from the Silicon crystal and, hence, are known as *acceptors*. This process is summarized in figure 3.7 (right).

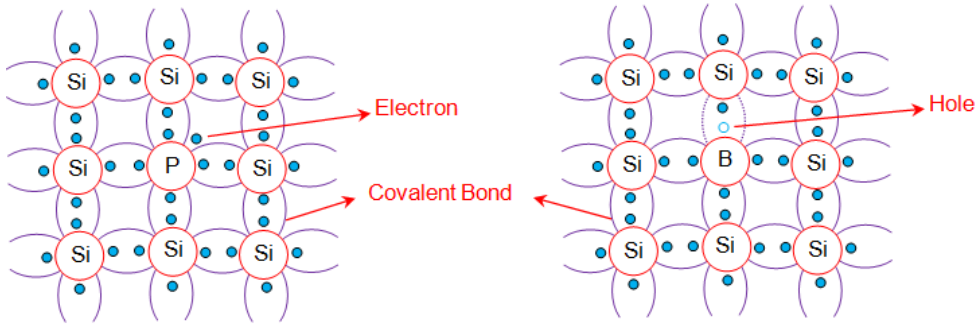

Figure 3.7: Bonding model for n-type (left) and p-type (right) semiconductors.

The number of impurities in n-type and p-type semiconductors is usually much higher than the intrinsic charge carrier concentration. Hence, effectively there is only one type of charge carrier present in extrinsic semiconductors and its concentration is equal to the number of either donor or acceptor atoms. Since the carrier concentration is increased in these materials, the discussion about large inherent currents and noise that was made for intrinsic semiconductors also applies in this type of semiconductor.

3.5 Charge transport in Semiconductor materials

The transport of charge carriers in Silicon plays a critical role in the formation of a p-n junction and the creation of signal in particle detectors. There exist three mechanisms of charge transport: thermal, diffusion and drift.

Charge carriers subject to a temperature T move in a random direction with a kinetic energy of $3/2k_BT$. Due to the random direction of thermally moving charge carriers no net signal is obtained.

The diffusion movement appears if there are regions with a different concentration of charge carriers. In that situation, charges migrate from the region with higher concentration to the lower one. This can be understood by considering two regions, A and B, with a different carrier concentration. As stated above, electrons in thermal equilibrium have a random movement in all directions with the same average velocity. If the majority of charge carriers belong in A, statistically more particles will cross from A to B than vice versa. This generates a net current

J_{dif} that can be expressed as [35]:

$$J_{n,dif} = eD_n \nabla n \tag{3.5}$$

$$J_{p,dif} = -eD_p \nabla p \tag{3.6}$$

where $D_{n/p}$ is the electron/hole diffusion coefficient.

The drift movement originates when the charge carriers move in an electric field. Electrons and holes are accelerated in the direction of the field while colliding with the semiconductor lattice. The current generated because of the drift movement is expressed as:

$$\vec{J}_{n,dr} = ne\mu_n \vec{E} \tag{3.7}$$

$$\vec{J}_{p,dr} = pe\mu_p \vec{E} \tag{3.8}$$

where n and p are the number of electrons and holes respectively and μ_n/μ_p is the electron/ hole mobility. The mobility is not an independent parameter but depends on the electric field. For large values of the electric field ($\sim 10^5$ V/cm) the charge carrier mobility decreases, leading to a saturation of the drift current. Compared to the diffusion current, the drift current is much faster. This has advantages in terms of radiation hardness and detector timing, as discussed in section 3.13.

The total drift current is given by the sum of equations 3.7 and 3.8 and is proportional to the electric field applied. The proportionality factor is known as *conductivity* σ and its reciprocal is the resistivity ρ of the material:

$$\sigma = \frac{1}{\rho} = ne\mu_n + pe\mu_p \tag{3.9}$$

The resistivity of a semiconductor depends on the number of free electrons and holes populating the material. Thus, it can be modified through the doping of the material (see section 3.4.2). In extrinsic semiconductors, the higher is the number of donors/acceptors the lower is its resistivity. As discussed in section 3.4.1 this is an issue as it leads to high currents as compared to the current generated by a ionising particle. The solution is to use a material depleted of free charge carriers which is achieved by making a junction of p-type and n-type materials. This junction is known as a *p-n junction* and is discussed in section 3.6.

3.6 The p-n junction

As it has been discussed in sections 3.4.1 and 3.4.2, the main problem in using intrinsic and extrinsic Silicon for the construction of a particle detector is the large number of mobile electrons and holes present in these materials. Depletion of free charge carriers is achieved by making a p-n junction and it constitutes the minimum building block for most of semiconductor devices. A

p-n junction itself is a two terminal device that can be used to detect radiation. The mathematical model of a p-n junction is reviewed in this section.

N-type and p-type silicon are characteristic for having an excess of electrons and holes respectively. When those two materials are brought together to form a p-n junction, the different concentration of charge carriers generates a diffusion current of electrons going into the p-type region and holes going into the n-type region. Around the junction, electrons and holes recombine creating a region depleted of free charge carriers, also known as a *depleted region*. Due to this charge movement, the non-mobile donors and acceptors generate an electric field that acts to oppose the movement of the electrons and holes due to diffusion, until equilibrium is reached where there is no net current flow. Hence, a p-n junction in equillibrium has a region around the junction depleted of free charge carriers which is positively charged in the n-type side due to the ionised donor atoms and negatively charged in the p-type due to the ionised acceptor atoms. The space charge neutrality condition gives the following relationship:

$$N_A x_p = N_D x_n \tag{3.10}$$

where N_A and N_D are the acceptor and donor concentrations respectively and x_p/x_n the coordinates of the space charge boundaries on the p-side/n-side respectively. Assuming that the only contribution to the net charge in the depleted region is due to the ionized donors and acceptors (i.e. there are no free charge carriers in that region), the charge density distribution can be written as:

$$\rho(x) = \begin{cases} -eN_A & -x_p < x < 0 \\ eN_D & 0 < x < x_n \end{cases} \tag{3.11}$$

where $x = 0$ is the position of the junction. The electric field due to this space charge can be calculated using the first Maxwell equation:

$$\frac{dE}{dx} = \frac{1}{\varepsilon}\rho(x) \tag{3.12}$$

By integrating equation 3.12 the expression of the electric field inside the depleted region is obtained:

$$E(x) = \begin{cases} \frac{-eN_A}{\varepsilon}(x+x_p) & -x_p < x < 0 \\ \frac{+eN_D}{\varepsilon}(x-x_n) & 0 < x < x_n \end{cases} \tag{3.13}$$

This electric field is naturally generated inside the depleted region of a p-n junction. With its minimum at the position of the junction, it is linearly decreasing towards the junction at the p-type and n-type sides. The potential associated to this electric field is known as the *built-in potential* since it is generated inside the junction without applying any external voltage. It can

be calculated by performing a second integration:

$$V_{bi} = \frac{e}{2\varepsilon}\left(N_A x_p^2 + N_D x_n^2\right) \tag{3.14}$$

Equations 3.10, 3.11, 3.13 and 3.14 are represented in figure 3.8.

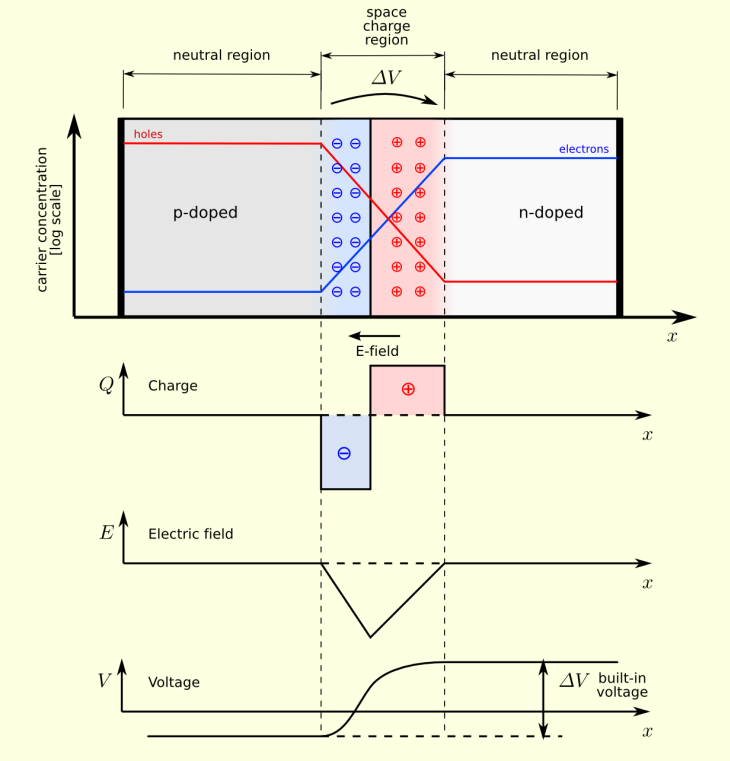

Figure 3.8: From top to bottom: Space charge distribution, space charge density, electric field and built-in voltage of a p-n junction. Image reproduced from [36].

As it has been mentioned, all the previous discussion was valid inside the depleted region i.e. in the interval ($-x_p$, x_n). This interval is also known as *depletion width* since it corresponds to the width of the region depleted of free charge carriers. The depletion width is given by:

$$W \equiv x_n + x_p = \sqrt{\frac{2\varepsilon}{e}\left(\frac{N_A + N_D}{N_A N_D}\right) V_{bi}} \tag{3.15}$$

The depletion width corresponds to the sensitive part of a particle detector, as it is a region with a low number of free charge carriers. Moreover, charge carriers within this region move due to the presence of the electric field leading to faster signals.

The depleted region can be further increased by applying an external voltage V_{ext}:

$$W \equiv x_n + x_p = \sqrt{\frac{2\varepsilon}{e}\left(\frac{N_A + N_D}{N_A N_D}\right)(V_{bi} + V_{ext})} \simeq \sqrt{\frac{2\varepsilon}{e}\left(\frac{1}{N_{A/D}}\right)(V_{bi} + V_{ext})} \tag{3.16}$$

Thus, the depletion width can be increased proportionally to the square root of the external voltage if a positive voltage (*reverse bias voltage*) is applied. The approximation in equation 3.16 can be used if $N_{D/A} \gg N_{A/D}$. This is the usual case in particle detectors and is denoted by p^+-n or n^+-p if the p/n is much more heavily doped. In this situation, it follows from equation 3.10 that the depleted region is much more extended in the less doped side. Equation 3.16 can be rewritten in terms of the resistivity as:

$$W \propto \sqrt{\rho(V_{ext} + V_{bi})} \tag{3.17}$$

Thus, high resistivity Silicon can be depleted at lower voltages and, hence, is a more desirable material for particle detectors. High resistivity substrates are becoming available in CMOS processes enabling the creation of fully depleted CMOS sensors, as described in section 3.11.

The voltage required to fully deplete the whole thickness of the detector is known as the *depletion voltage* (V_{dep}). Above this point, if a higher voltage is applied the electric field inside the junction will increase according to:

$$E(x) = \frac{2V_{dep}\,x}{d^2} - \frac{(V_{ext} - V_{dep})}{d} \tag{3.18}$$

where d is the sensor thickness.

A p-n junction stores or releases charge if the external voltage is modified. This behaviour is the same as that of a capacitor. In many cases, the capacitance of the depleted region is that of a parallel plate capacitor, since the charges are added or removed at its borders. Thus, the capacitance of a p-n junction is given by:

$$C_j = \varepsilon \frac{A}{W} = A\sqrt{\frac{\varepsilon e}{2(V_{bi} + V_{ext})} \frac{N_D N_A}{N_D + N_A}} \tag{3.19}$$

where A is the area of the junction.

3.7 Leakage Current

The *leakage current* or *reverse current* is a volumetric current present in the depleted region of a radiation sensor. It originates from thermal generation and the diffusion of charge carriers from the non-depleted region of the detector. Hence, the leakage current can be written in terms of its diffusion (I_d) and pair generation components (I_g) [37]:

$$I_R = I_d + I_g = q\sqrt{\frac{D_p}{\tau_p}\frac{n_i^2}{N_D}} + \frac{qn_iW}{\tau_g} \simeq \frac{qn_iW}{\tau_g} \tag{3.20}$$

where $\tau_{p/g}$ are the generation and recombination lifetimes respectively. The later approximation applies for silicon sensors, where n_i is small and the thermal generation dominates over the diffusion term. Due to its dependence on the depletion width, the reverse current saturates for reverse voltage values higher than the depletion voltage. Moreover, as its main contributor is thermal generation, leakage current has also a strong dependence on temperature [30]:

$$I_R \propto n_i \propto T^2 \exp\left(-\frac{E_g}{2k_B T}\right) \tag{3.21}$$

where k_B is the Boltzmann constant, T the temperature and E_g the Silicon band gap (1.12 eV).

The leakage current has a strong impact on the sensor noise and power consumption and hence, has to be kept as low as possible. This is achieved in the LHC by maintaining the sensors at low temperature. Moreover, non-ionising damage on the bulk of the device significantly increases this current. This is explained by the generation of intermediate levels in the Silicon band gap due to the radiation-induced traps, making thermal carrier generation more likely.

3.8 Charge generation

Particles generate an electrical signal in semiconductor detectors when they cross their sensitive bulk. This signal is induced by the generation of electron-hole pairs in the detector material. The trace and the generation mechanism of these charge carriers is very particle and energy dependent.

Moderately relativistic charged heavy particles crossing a material undergo electromagnetic collisions with valence electrons, ionising them if the transferred energy is high enough. The mean energy loss (stopping power) of these particles per unit distance is given by the Bethe-Bloch equation [24]:

$$\left\langle -\frac{dE}{dx} \right\rangle = 4\pi N_A r_e^2 m_e c^2 z^2 \frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\left(\frac{2m_e c^2 \beta^2 \gamma^2 T_{max}}{I^2}\right) - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right] \tag{3.22}$$

where N_A is the Avogadro's number, r_e is the classical electron radius, m_e is the electron mass, c is the speed of light, z is the charge of the traversing particle, Z is the atomic number of the crossed material, A is its atomic mass, $\beta = v/c$ with v being the velocity of the particle, $\gamma = 1/\sqrt{1-\beta^2}$, T_{max} is the maximum kinetic energy of the particle, I is the mean excitation energy of the material and $\delta(\beta\gamma)$ a density effect correction. An example plot of the stopping power for muons in Copper is shown in figure 3.9. The Bethe-Bloch equation is valid in the range $0.1 \lesssim \beta\gamma \lesssim 1000$ and for materials with intermediate Z. At $\beta\gamma \simeq 3$ the distribution finds a broad minimum. The position of this minimum only varies from $\beta\gamma = 3.5$ to 3 for materials with Z from 7 to 100 [24]. In high energy physics, most particles have a mean energy loss around this minimum. These are known as *Minimum Ionising Particles* (MIPs). A MIP crossing a bulk

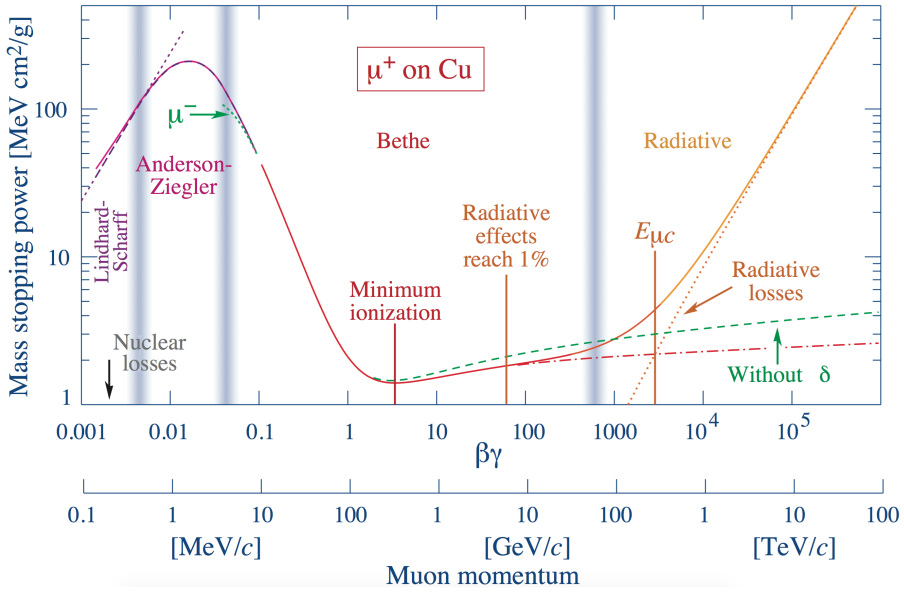

Figure 3.9: Energy loss rate for muons on copper. Figure reproduced from [24].

of Silicon generates 73 e/h pairs per μm as Most Probable Value (MPV).

Photon interaction with matter is also relevant in this work as low energy and high energy X-ray photons are used to study the performance of different sensors. Photons interact with matter through three different mechanisms: the photoelectric effect, Compton scattering and electron-positron pair production. The photoelectric effect occurs if the incident energy of the photon is higher than the ionisation energy of the atom. In this situation, the incident photon is absorbed by a valence electron, which is excited to the conduction band generating an electron-hole pair. In Compton scattering, the photon is deflected by an electron, transferring a fraction of its energy to the latter. This effect can ionise valence electrons generating electron-hole pairs. The pair production effect occurs if the photon energy is twice the center of mass energy of an electron. It consists of the conversion of the photon into an electron and a positron which, in turn, can ionise the material.

The probability of each effect to take place depends on the energy of the photon and the Z of the material, as shown in figure 3.10. Particle sensors are widely tested using fluorescence X-rays in the 2 to 10 keV energy range. In Silicon and for photon energies below 100 keV the photoelectric effect is the dominant interaction of photons with matter and, hence, the principal charge generation mechanism in these experiments. Due to this interaction, a photon beam of intensity I_0 gets exponentially attenuated through a material bulk according to:

$$I(x) = I_0 \exp(-x/\mu) \tag{3.23}$$

where x is the material thickness and μ is the penetration depth. For X-ray photons with an energy of 6 keV, 8 keV and 15 keV in Silicon μ is about 30 μm, 70 μm and 450 μm respectively [39]. Hence, for a typical 300 μm thick detector almost the totality of the beam is absorbed in

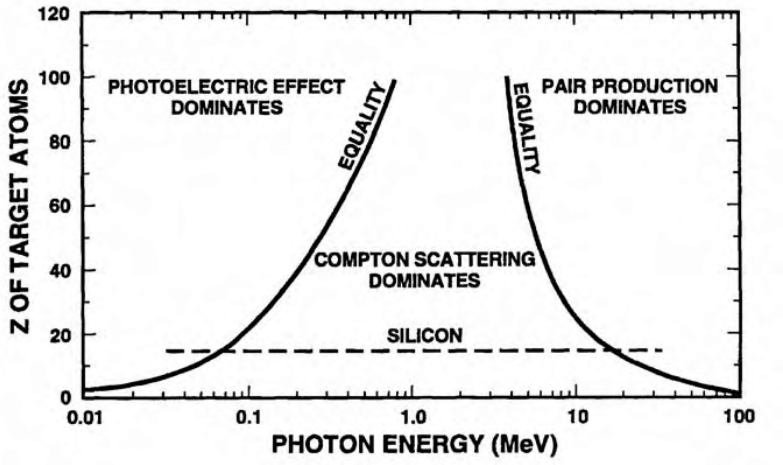

Figure 3.10: Regions of predominance of the three main processes of photon interaction with matter. The curves indicate when the two adjacent processes are of equal probability. Figure reproduced from [38].

6 keV and 8 keV X-rays while about 50% gets absorbed in 15 keV X-rays.

3.9 Signal creation in Particle Detectors

In section 3.8 the charge generation mechanisms on semiconductor devices have been described. External radiation induces free charge carriers in the sensor bulk that move in the presence of an electric field. These moving charge carriers are responsible for an electric pulse at the output of the device. The physics behind this phenomenon are reviewed in this section.

In order to understand the signal creation on semiconductor devices consider a simple p-n junction with two electrodes located at the top and bottom of the device. This is the simplest form of a semiconductor detector and is known as a *diode*. If an electron with charge $-q$ is induced in the junction, a charge $+q$ is induced at the electrode surface with a certain distribution. Moving this electron with respect to the electrode will result on a change in this distribution. These moving charges at the electrode surface generate a current, which can be detected with suitable electronics. The current vanishes when the electron stops moving or when it reaches the electrode. The relationship between a charge with velocity $\vec{v(t)}$ and the current $I_k(t)$ generated in the k-th electrode is established by the Shockley-Ramo theorem [40, 41]:

$$I_k(t) = q\vec{v}(t) \cdot \vec{\varepsilon_w} \tag{3.24}$$

where $\vec{\varepsilon_w}$ is the *weighting field*. $\vec{\varepsilon_w}$ defines the coupling between a moving charge and the corresponding electrode. Hence, its analytical expression depends on the structure of the detector and the shape of the terminal. Since the generated particles are driven by the drift field, its velocity

can be related to the electric field inside the junction:

$$\vec{v} = \mu_{e/h}\vec{E}_d \tag{3.25}$$

where $\mu_{e/h}$ is the electron/hole carrier mobility. Some considerations can be done regarding the Shockley-Ramo theorem. Note that any charged particle moving inside a sensor will induce a current on the attached electrodes. In a silicon sensor this current will have two contributions: moving electrons and moving holes. The time duration of the current corresponds to the travel time of the e-h pairs from its generation position to the readout electrodes. Electrons are collected earlier than holes due to their higher mobility and, hence, their contribution to the induced current will be finished before that of the holes. Moreover, e-h pairs generated closer to the electrodes will be collected sooner. As discussed in section 3.8, particles generate different charge distributions inside the Silicon. Hence, the current induced by these multiple e-h pairs is a convolution of the current induced by each individual charge carrier. The specific shape of a detector signal after a particle hit depends then on the type of particle and on the detector structure. Real-world detectors usually have complicated structures with segmented electrodes to allow for position resolution. Moreover, charge trapping from lattice defects can also affect the trajectory of charge carriers, modifying the induced current. For this reason, the description of induced signals is obtained from computer simulations for each particular device.

Measuring the charge generated by a traversing particle in a detector is critical to determine its energy. The total charge collected in a given event can be simply obtained by integrating the Shockley-Ramo current over the e/h collection time ($t_{e/h}$):

$$Q = \int_0^{t_e} I_{k,e}(t)\,dt + \int_0^{t_h} I_{k,h}(t)\,dt \tag{3.26}$$

in which the contribution from electrons and holes to the total charge of the event has been made explicit. Note that the sensor takes a time $\max(t_e,\, t_h)$ to collect the whole charge of the interaction. Detectors at the LHC must have a time response faster than the bunch-crossing time (25 ns). Timing of the sensor can be optimised by thinning the sensor to reduce the travel distance of charge carriers or by applying a high external voltage to increase the carriers velocity.

The signal described in this section can be read out by connecting electronics to the sensor terminals. In the majority of sensor technologies, electronics are designed in a different board and connected through a process called *bump-bonding* [42] in pixels or *wire-bonding* in strips. This is a complicated and expensive process that is avoided in the CMOS technology discussed in this thesis. In CMOS sensors, this is achieved by embedding the transistors that form the readout electronics in the bulk of the sensor. The physics behind transistors and a more comprehensive overview of the CMOS technology are given in sections 3.10 and 3.11.

3.10 Metal Oxide Field-Effect transistors

Field-Effect Transistors (or FET) [43] are electronic devices with three terminals in which the current flow between two of the terminals is controlled by the voltage applied to the third terminal. Metal Oxide Semiconductor Field-Effect Transistors (or FET) are a type of FET in which the terminal that controls the current flow is insulated by an oxide layer. Nowadays FETs are extensively used in electronic circuits and are the basic building block of the CMOS sensor technology. FETs (or MOS for short) perform as amplifiers or switches in analog and digital circuits. Specifically, they are used in the design of the CMOS sensors presented in chapter 4. Hence, the essential electrical characteristics of MOS transistors are examined in this section.

3.10.1 Structure of MOSFET transistors

A MOS transistor is a semiconductor device that contains two p-n junctions in its structure. FETs are distinguished between n-type (NMOS) and p-type (PMOS). Since the discussion about the structure and electrical properties of NMOS and PMOS transistors is analogous, only NMOS transistors are reviewed in this section and their differences are discussed when required.

The cross-section of an NMOS transistor is shown in figure 3.11. The bulk of the device is composed of a p-type semiconductor, known as the *substrate*, in which two heavily doped n^+-type regions are implanted. Usually, the substrate is connected to ground to give a reference voltage to the transistor. The n^+ regions have an ohmic contact attached to be able to function as electrodes of the device. The role of these electrodes is interchangeable and only depends on the voltage applied to them. The terminal with the higher voltage is known as *source* while the other one is called *drain*. The space between source and gate is known as the transistor *channel* and its distance between the implants, L in the figure, is the *channel length*. The orthogonal distance of the channel is the *channel width* and will be represented as W. A third terminal is placed above the transistor channel and is insulated from the p-type substrate by a Silicon-oxide layer. This terminal is known as the *gate* and acts as the control electrode of the transistor. The gate electrode is usually made of n^+-type polysilicon[2] contacted by a metal layer.

In the case of PMOS transistors, the structure of an NMOS transistor remains the same with all the doping types inverted. Hence, p^+-type implants forming the source and drain electrodes are embedded in a moderately doped n-type substrate. The gate is also made of p^+ polysilicon.

3.10.2 Electrical properties of MOSFETs

A FET device has different operating regions depending on the interconnections of its own terminals and the voltage applied to them. In this section, the voltage of the source terminal is

[2]The term polysilicon refers to a silicon material made of several silicon crystals. This is due to the fact that, being grown above Silicon Oxide, Silicon can not be built as a single crystal but as "grains" with diferent crystal orientations.

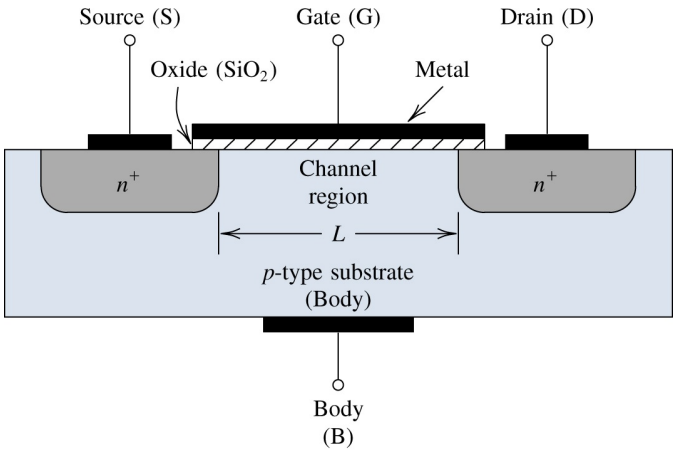

Figure 3.11: Cross section of a typical NMOS transistor.

taken as the reference potential. The gate-source voltage is denoted as V_{GS} and the drain-source voltage as V_{DS}.

Three different operating regions can be identified when modifying the V_{GS} voltage. If $V_{GS} < 0$, holes are attracted underneath the SiO_2 blocking the device conductivity. This region of operation is called *accumulation* and the device operates as a simple capacitor. No current can flow between source and drain in this configuration.

When V_{GS} is increased above zero, the holes populating the channel region are repelled creating a layer of negatively charged acceptor atoms. This is known as *depletion*. However, the acceptors are fixed in the silicon lattice and can not participate in the current flow if a potential is applied between source and drain.

The channel of an NMOS transistor is only enabled for current flow when V_{GS} is high enough to repel the holes that populate the channel region and to attract a significant number of electrons to the SiO_2-substrate interface. This situation is known as *strong inversion* since, effectively, an n-type region is created underneath the gate connecting the source and drain terminals. This n-type region is not permanent but disappears when the voltage applied to the gate is reduced again. The minimum voltage required to enable the transistor channel is defined as *threshold voltage* (V_{th}).

However, the current flow between source and drain does not stop abruptly when going below V_{th}. Instead, the current has an exponential dependence with V_{GS}. This current is generated by the electrons that cross the potential barrier at the source and diffuse to the drain due to the low concentration of electrons at the channel. This extra region of operation is known as *weak inversion*.

When the transistor channel is switched ON (i.e. $V_{GS} > V_{th}$) and V_{DS} is increased above zero the drain/substrate junction gets reverse biased, creating a depletion region around this area that is widened as V_{DS} increases. Due to the lateral electric field built into this depleted region (see section 3.6) a current starts flowing from source to drain. Two MOS configurations can be

defined according to the applied V_{DS} potential.

For small values of V_{DS} the electric field along the channel can be considered constant. This configuration is known as *linear region* since the current flow between source and drain (I_{DS}) varies linearly with the V_{DS} potential:

$$I_{DS} = \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{th})V_{DS} \equiv \frac{V_{DS}}{R_{ON}} \tag{3.27}$$

where $C_{ox} = \varepsilon_{ox}/t_{ox}$, with t_{ox} being the gate oxide thickness and ε_{ox} the SiO_2 dielectric constant, is the gate capacitance per unit area. According to Ohm's law, the linear factor can be associated to a resistance, with a value that depends on V_{GS}. Hence, a transistor operating in the linear region acts as a voltage-controlled linear resistor. If V_{DS} is increased further, the assumption of a constant electric field is no longer valid, but the voltage across the channel varies from 0 to V_{DS}. In this situation the current flow is given by:

$$I_{DS} = \mu_n C_{ox} \frac{W}{L}\left[(V_{GS} - V_{th})V_{DS} - \frac{V_{DS}^2}{2}\right] \tag{3.28}$$

and the transistor behaves as a voltage-controlled non-linear resistor.

When the drain implant is biased a depleted region around this electrode is created, counteracting the attraction of charge carriers by the gate terminal. Under this condition, the minimum gate voltage to support an inversion layer at a point x in the channel with a source voltage $V(x)$ is given by $V_{GS} - V_{th} - V(x) > 0$. When V_{DS} is increased above $V_{DS} > V_{GS} - V_{th}$ the inversion layer is no longer created, starting from the vicinities of the drain implant. This situation is known as channel *pinch-off*. The regions where the inversion layer is not formed do not contribute to the current flow. However, charge carriers from the inverted zones are still capable to traverse the pinched-off region and reach the drain. The voltage drop across the channel does not increase as the drain voltage is increased; instead the pinched-off region widens. This operating region is known as *saturation* as the drain current reaches a limit and ceases to increase. The drain current in saturation is given by:

$$I_{DS} = \frac{1}{2}\mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{th})^2(1 + \lambda V_{DS}) \tag{3.29}$$

where λ is known as the *channel length modulation parameter* and accounts for a residual dependence of the drain current with V_{DS} due to the effective decrease of the channel length in saturation. Note that, ignoring this residual effect, a transistor working in the saturation region works as a voltage-controlled current source. For reference, the I-V characteristics of the drain implant of an NMOS transistor from the linear to the saturation region are shown in figure 3.12.

An important parameter of the transistor is its *gate transconductance* g_m. This parameter quantifies how variations in the device current I_{DS} are related to variations of the gate voltage

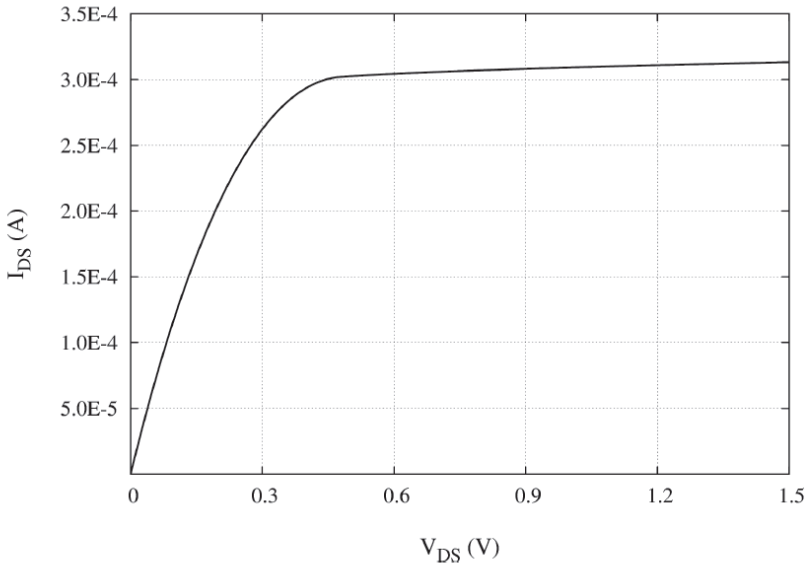

Figure 3.12: I-V characteristics of an NMOS transistor. Figure reproduced from [44].

V_{GS}. In saturation, and neglecting the channel length modulation (λV_{DS}), it can be calculated with the following expression:

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}} = \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{th}) = \sqrt{2\mu_n C_{ox} \frac{W}{L} I_{DS}} \tag{3.30}$$

Note that g_m has a dependence on the device aspect ratio W/L. Large values of the transconductance are desirable in transistors in which the noise contribution is required to be as low as possible. This will become apparent in section 3.12.

Another important parameter is the *output conductance* g_{ds}. Similarly to the transconductance, the conductance quantifies how changes in the drain to source voltage V_{DS} affect the drain to source current I_{DS}. It is defined as:

$$g_{ds} = \frac{\partial I_{DS}}{\partial V_{DS}} = \frac{1}{2}\lambda \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{th})^2 \tag{3.31}$$

Its reciprocal is known as the *output resistance* r_0 and is related to the gain achievable in MOS amplifiers.

MOS transistors form the basis of many electronic applications. In the next section it is pointed out the importance of such structures in the CMOS technology.

3.11 The CMOS technology

The Complementary Metal-Oxide Semiconductor technology (CMOS) finds its success in the capability to embed NMOS and PMOS transistors on the same substrate to create logic functions. The substrate is typically made of p-type Silicon in which the NMOS transistor is placed. In order to include PMOS transistors, an n-type region is made on the substrate, known as *n-well*. A simplified cross-section of a CMOS device including one NMOS and one PMOS transistor is shown in figure 3.13. The PMOS transistor is electrically isolated from the p-substrate by

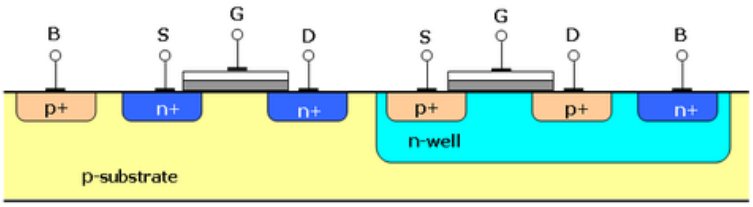

Figure 3.13: Simplified cross section of a typical CMOS device.

the n-well. This is not the case in the NMOS transistors, since the substrate and the transistor channel are made of the same type Silicon. Hence, the p-substrate has to be kept at the lowest voltage in order to avoid forward-biasing the p-n junction structures. Electrical isolation can also be achieved for the NMOS transistors by creating a *triple well* structure. A deeper n-well, known as *deep n-well*, is grown in the substrate to host the NMOS transistor p-type bulk. This is a requirement for low-noise applications although it has the disadvantage of occupying more space.

CMOS devices can contain parasitic transistors in their structure. Examples of this are the p^+ implants of the PMOS transistor, the n-well and the p-substrate forming a PMOS transistor oriented vertically or the n^+ drain terminal of the NMOS transistor, the p-substrate and the n-well, forming a horizontally oriented NMOS transistor. These two transistors act as amplifiers and are connected through the parasitic resistances on the p-substrate and the p-well forming a positive feed-back loop [44]. Hence, if a current flows through the p-substrate, the two parasitic transistors will amplify it indefinitely causing a large current that can potentially damage the device. This effect is known as *latch-up* and must be suppressed in the design. Typical safety measures are keeping the p-substrate and n-well resistances low and reducing the gain of the parasitic amplifiers by making the n-well wider. Device isolation can also be achieved by including a shallow trench filled with SiO_2 between transistors. This structure is known as *Shallow Trench Isolation* (STI). Although some parasitic paths can be suppressed with the use of STI oxides, transistors can still interfere through the common substrate. Moreover, STI layers have a detrimental impact in the radiation hardness of a circuit (see section 3.13).

Another way to prevent latch-up is by including a lightly doped *epitaxial layer* on a highly doped substrate. This arrangement is shown in figure 3.14. Typical scales for the epitaxial layer

are a few μm, while the p-substrate has a thickness of a few hundred of μm. By including this

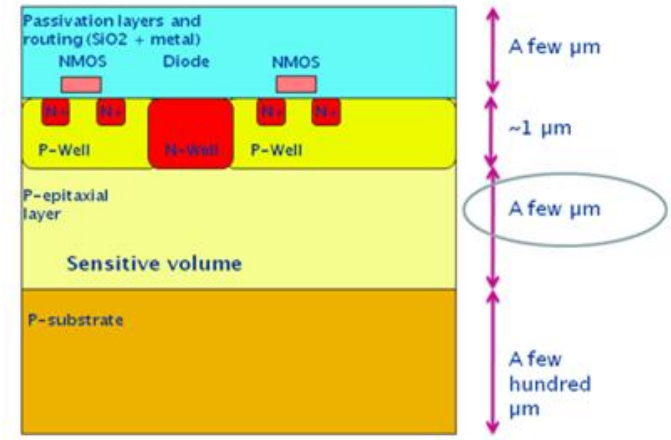

Figure 3.14: Simplified cross section of a CMOS device with an epitaxial layer.

extra layer, the resistance of the p-substrate is reduced as well as the gain of the vertical PMOS transistor [45]. These two improvements make the device less sensitive to latch-up and relax the requirements for the size of the n-wells, facilitating more compact circuits.

3.11.1 CMOS as particle detectors

The CMOS technology can be employed to fabricate particle detectors. This is of particular interest because of its potential to include the readout electronics on the sensing substrate during the fabrication process, avoiding the complex and expensive bump-bonding process. Typically, the front-end includes a pre-amplifier circuit capable to produce a signal of a few hundreds of mV for a MIP generating a charge of 200 e in the detector. Moreover, the fact that the CMOS technology is widely used in industry and can be mass produced also helps to reduce the cost of this technology.

In section 3.6 the basic structure employed to detect radiation has been described. In CMOS detectors, a p-n junction is also constructed to collect the free charge generated by incoming radiation. Hence, the collecting electrode is formed by an n-well with an n^+ contact to allow for external biasing. However, note that the n-well structures that accommodate the PMOS transistors can also act as collecting diodes causing a signal loss. In order to prevent this from happening, a deep p-well can be placed surrounding the n-well. CMOS sensors usually employ epitaxial layers since their high resistivity is easier to deplete. As a comparison, the resistivity of the epitaxial layer is usually about 20 Ω·cm while for the bulk material it is about 20-50 mΩ·cm.

3.11.2 Small electrode designs

The chips investigated in this thesis feature a small electrode design. Small electrode designs offer advantages in terms of power consumption and increased *Signal-to-Noise Ratio* (*SNR*).

The capacitance C of a p-n junction, described by equation 3.19, is proportional to its area. In particle detectors, the collection electrode is usually fabricated by growing a highly doped silicon material into a more lightly doped silicon bulk forming a p-n junction. Thus, the smaller the area of the electrode, the smaller is its capacitance. This has an impact on the generated signal, as the voltage level induced by a charge Q measured in the electrode is given by:

$$V = \frac{Q}{C} \tag{3.32}$$

Hence, small collection electrodes lead to larger signals. The electrode is usually connected to an amplification circuit. The main noise contribution of this circuit typically corresponds to the thermal noise (see section 3.12.1) of the input transistor [46], which is inversely proportional to the square root of its gate transconductance g_m. Thus, the *SNR* of this configuration is:

$$SNR \sim \frac{Q}{C}\sqrt{g_m} \tag{3.33}$$

Note that for lower values of C, associated to smaller electrode areas, the *SNR* is increased.

The transconductance is proportional to the bias voltage of the transistor (see equation 3.30). Given the assumption that the power consumption P of the circuit is dominated by this voltage, equation 3.33 can be written as:

$$P \sim \left(\frac{SNR}{Q}C\right)^2 \tag{3.34}$$

Hence, small capacitance values are crucial to achieve a low power consumption for a given *SNR*.

3.12 Noise in Semiconductor devices

Noise in detector systems and electronics is defined as the fluctuation of the sensor signal from its mean value. Noise is generated within the detector and the readout electronics through different mechanisms. All the devices in a detector system contribute in one way or another to the overall noise. These include the sensor leakage current, the transistors forming the front-end circuitry and the detector bias resistors. The correct understanding and characterisation of noise sources when designing a chip is critical to ensure its correct performance for a given application. The contribution to the total noise of each component is expressed in terms of *noise spectral density* (v^2), consisting in the power dissipated into a 1 Ω resistor (units of V^2/Hz or A^2/Hz). Depending on whether the noise spectral density has a frequency dependence or not, noise is considered *coloured* ($v^2(f)$) or *white*. Given a noise spectral density, the RMS of the associated signal

fluctuation can be calculated with the following expression [47]:

$$V_{RMS} = \sqrt{\int_{f_l}^{f_h} v^2(f)\, df} \tag{3.35}$$

where f_h and f_l are the highest and lowest noise frequencies taken into consideration.

The characterisation of noise in experimental measurements is done through the standard deviation, V_{std}, of the data sample [44]:

$$V_{std} = \sqrt{\frac{1}{N-1}\sum_{k=1}^{k=N}(V_k - V_{av})^2} \tag{3.36}$$

where V_k is the physical value of the k^{th} sample, V_{av} the mean value of the sample and N the total number of samples.

3.12.1 Thermal noise

Thermal effects generate random movements of charge carriers in conductors. Although these movements have a zero average value, they provoke instantaneous variations of signals within a circuit. This variation is referred to as *thermal noise* and is considered a white noise source. The thermal noise spectral density generated in a resistor with resistance R is modelled by the following expression:

$$v_{tn}^2 = 4k_B T R \tag{3.37}$$

Note that the depleted region of a p-n junction has an associated resistance. Thus, the thermal noise inside a transistor channel and a depleted particle sensor can be modelled by this equation. In the case of MOS transistors, the thermal noise can be written in terms of the device transconductance g_m as:

$$v_{tn}^2 = 4k_B T \alpha_w \gamma \frac{1}{g_m} \tag{3.38}$$

where γ takes into account the region of operation of the transistor being 1/2 in weak inversion and 2/3 in strong inversion [44]. α_w is a correction coefficient called *excess noise factor* and models an increase of the expected noise in short channel devices.

Thermal noise, together with shot noise, is the main source of noise in transistors.

3.12.2 Shot noise

Shot noise appears if there is a current flow across a potential barrier such as in p-n junctions. It originates from the discrete movement of the electrons and holes that generate the current. Electrons crossing the depleted region of a p-n junction are independently transmitted and collected randomly at the end of their path. This produces fluctuations in the measured current. The shot

noise spectral density is given by:

$$i_n^2 = 2qI \tag{3.39}$$

where I is the current under consideration and q the electric charge of its components. Contrary to the thermal noise, shot noise can not be suppressed by reducing the temperature.

3.12.3 Flicker noise and Random Telegraph Noise

The *Flicker*, or *1/f* noise, is one of the most prominent noise sources in MOS transistors. The origin of flicker noise is controversial in the literature [48]. The Hodge model affirms that flicker noise comes from random changes in the mobility of charge carriers [49], linking it to a bulk effect. In this representation, the 1/f spectral density (v_{nf}^2) is modelled by the following equation:

$$v_{nf}^2 = \frac{K_{fa}(V_{GS} - V_{th})}{C_{ox}WL}\frac{1}{f} \tag{3.40}$$

where K_{fa} is a constant that depends on the fabrication process, W and L are the dimensions of the transistor and f the frequency. Another possible explanation for the 1/f noise is the McWorther model [50]. It states that charge carriers can be randomly trapped at the Si-SiO_2 interface, varying the number of charge carriers in the current flow and, hence, generating noise. In this scenario, the flicker noise is expressed as:

$$v_{nf}^2 = \frac{K_{fb}}{C_{ox}^2 WL}\frac{1}{f} \tag{3.41}$$

Equations 3.40 and 3.41 differ in the bias voltage dependence factor. Experimental results have shown that such dependence is only seen in PMOS transistors. For this reason, flicker noise is better explained using the Hodge model for PMOS transistors and the McWorther model for NMOS transistors. The two equations have the same dependence on the transistor size but different fabrication constants.

Note that in both approaches, the noise spectral density increases as the transistor becomes smaller. In deep-submicron devices, the current flowing through the transistor channel is of the order of a few hundred electrons [44]. In this situation, a single electron being trapped on a trapping centre produces a noticeable variation of the measured current. This trapping and releasing of single electrons translates into randomly distributed pulses of constant amplitude. This phenomenon is known as *Random Telegraph Noise* (RTN) or *Random Telegraph Signal* (RTS) noise. From equations 3.40 and 3.41 it can be seen that RTS noise has a dependence on the transistor size, being larger for small transistors. Thus, the RTS noise poses a problem for transistors in the deep sub-micron scale [51] such as the ones employed in the TowerJazz 180 nm technology presented in section 4.1. The appearance of defects that lead to RTS noise is a statistical process that depends on the device fabrication as noted in equations 3.40 and 3.41

by the K_{fa} and K_{fb} coefficients. Hence, only some pixels in a chip will exhibit RTS noise.

RTS noise is also affected by Total Ionising Dose (TID) irradiation. Some studies have found an increase of RTS noise on FET devices after being exposed to ionising radiation [52]. This fact has been linked to the radiation-induced traps on the Si-SiO_2 interface, acting in a similar manner as the fabrication-induced defects. Due to these newly generated traps, the number of pixels in a device showing RTS noise behaviour is also expected to increase [53]. As discussed in section 3.13, annealing helps to remove these traps and, hence, to reduce RTS noise. The presence of RTS noise and its evolution after irradiation in Monolithic CMOS devices was studied within the context of this thesis. The results are presented in chapter 7.

3.13 Total Ionising Dose Effects

In sections 3.8 and 3.9 the mechanism for charge generation and signal creation in particle detectors has been discussed. Charged particles, such as high energy electrons, are responsible for charge generation within the detector bulk due to ionisation. This charge is collected by the readout electrode and does not cause any damage in this region of the device. However, electron-hole pairs are also generated in transistors that are part of the readout circuit. MOS transistors contain oxides and insulating layers, typically SiO_2, to electrically separate their electrodes from the transistor channel (see section 3.10). Holes created within these materials get confined causing a degradation of the electrical properties of the device. This effect is known as *ionising irradiation damage* and increases if the exposure to ionising radiation is higher. The term *Total Ionising Dose* (TID) denotes the total amount of absorbed radiation by a device. The SI unit employed to measure an absorbed TID dose is the Gray (Gy) and corresponds to the absorption of 1 J of energy per kg of matter. However, in the context of this thesis the unit *Radiation absorbed dose* (Rad) will be employed. One Rad corresponds to 1/100 Gy and will generate 8.1×10^{12} e-h/cm^3 in a SiO_2 insulating layer. CMOS detectors at the outer layer of the ATLAS ITk are expected to receive a TID of 80 MRad over ten years of operation [54]. Studying and understanding the effects of TID damage in particle detectors is critical to evaluate their viability in particle physics experiments. In this section the main mechanisms that lead to ionising damage are studied, as well as their effects on PMOS and NMOS transistors.

The physical processes leading to ionising damage are depicted in figure 3.15. Ionising radiation generates electron-hole pairs inside the SiO_2 that move in the presence of an electric field. Electrons drift towards the gate electrode while holes are pushed towards the SiO_2-Si interface. Before leaving the oxide, a fraction of electrons is recombined with holes. The remaining charge forms the so-called *charge yield* and corresponds to the fraction of charge that contributes to radiation damage. Holes move in the oxide through localised states, potentially getting trapped and generating a positively charged *oxide trap*. Either in the process of being trapped or while hopping through localised states, holes can release hydrogen ions that, in turn, drift towards

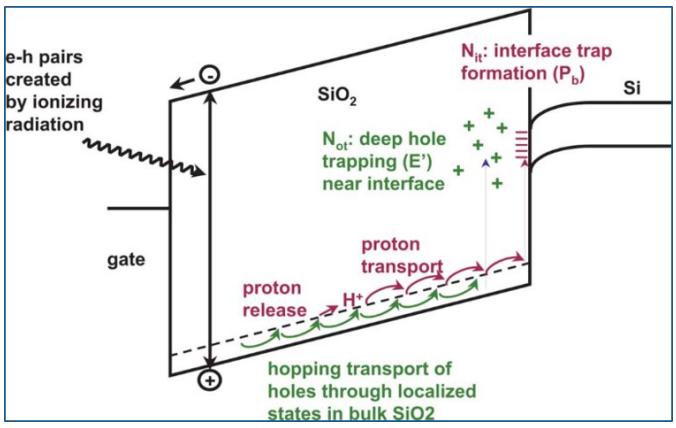

Figure 3.15: Illustration of the main processes that lead to TID damage in the energy band diagram of a MOS capacitor. The gate terminal is biased, generating the hole current and bending the energy band diagram in the MOS channel. Picture extracted from [55].

the interface. Hydrogen ions can get trapped there creating *interface traps*. Interface traps are positively charged on PMOS transistors and negatively charged on NMOS transistors.

The positive charge associated with oxide-traps gathers above the transistor channel causing a negative shift of the threshold voltage of both NMOS and PMOS transistors. Thus, the transistor will turn into its ON state at a lower gate voltage increasing the leakage current of the device at a given operating point. Interface traps contribute to a negative threshold voltage shift on PMOS transistors and compensate the effect of oxide traps on NMOS transistors. However, interface traps are created much slower than oxide traps, on time scales of thousands of seconds [55]. This delay on the creation of interface traps causes a characteristic shape in the threshold voltage vs TID dose plot of NMOS transistors commonly known as *TID bump*. This effect is discussed later in this section. In addition to the traps created on the gate oxide, radiation damage is also induced on the STI oxides. Charge created in this lateral oxide can open a conductive channel between source and drain enabling a flow of leakage current between these two terminals (see figure 3.16). Although the maximum value of the leakage current

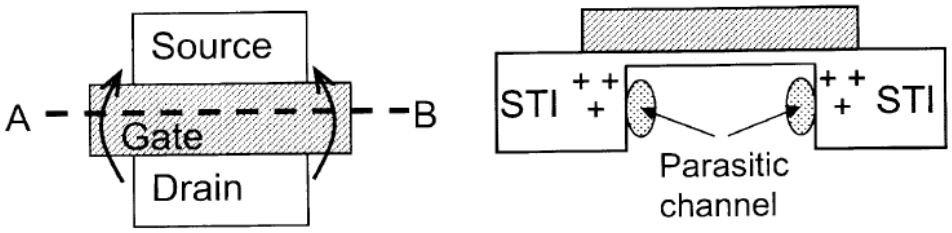

Figure 3.16: Top view of an open-layout NMOS transistor (left). On the right, the same transistor is viewed along the A-B line (from the source or drain). The radiation-induced positive charge trapping in the STI is represented by the multiple crosses. Picture reproduced from [56].

does not depend on the transistor width, the effective voltage threshold shift induced by the STI traps shows a strong dependence on it, being bigger for narrower transistors. This is known as

Radiation-Induced Narrow Channel Effect (RINCE) [56].

Aside from the creation of leakage paths between the source and drain terminals, TID damage also has an effect in the thermal and flicker noise of a MOS device. The increase of thermal noise is explained by the creation of interface states under the gate oxide, in the conductive channel. The carrier mobility is reduced in this region due to these radiation-induced traps, also reducing the transconductance of the device. It can be seen from equation 3.38 that a drop in g_m contributes to an increase of the thermal noise. The increase in flicker noise after irradiation is explained differently depending on the model under consideration. In the Hodge model, the flicker noise is directly related to the threshold voltage as stated by equation 3.40. Thus, a negative threshold voltage shift contributes to an increase of the flicker noise. In the McWorther model, the 1/f noise is caused by the trapping and releasing of single electrons in fabrication-induced traps located at the Si-SiO_2 interface. Radiation-induced oxide traps can also act as trapping centers, increasing the flicker noise of the device.

The negative effects of radiation can be partially "removed" through a process called *annealing*. Oxide trap annealing starts immediately after their creation, following a logarithmic dependence with time [55]. The neutralization of an oxide trap occurs due to either electron tunneling from the silicon into oxide traps or excitation of electrons from the oxide valence band to an oxide trap. Interface trap annealing only occurs at very high temperatures and, hence, induces a more lasting effect on transistors.

The fabrication process of MOS transistors can inherently contribute to its radiation tolerance. The thickness of the gate oxide in modern CMOS technologies is very thin, of the order of a few nm. In oxide thicknesses below 5 nm the annealing of oxide traps through tunneling is an efficient mechanism to remove this effect of radiation [57]. Hence, radiation damage in the gate oxide is negligible in modern CMOS technologies. However, the STI and field oxides are thicker and, thus, do not present this advantage. In order to mitigate its effects *Hardness-by-design* layout techniques are used. This consists on replacing the standard layout of a MOS transistor (represented in figure 3.11) by a more advanced layout. The layout most commonly employed is known as *Enclosed Layout Transistor* (ELT), shown in figure 3.17. In an ELT transistor, either the drain or the source terminal is enclosed by the gate and the remaining terminal surrounds the gate. Thus, the gate is not in contact with any STI oxide, avoiding the creation of leakage paths between the source and drain terminals as explained when discussing the RINCE effect. Any parasitic paths between neighbouring transistors is prevented by including a p^+ guard ring. Although ELT transistors can resist TID doses above 100 MRad [58], some drawbacks make it difficult to use them throughout the sensors. The most limiting factor is their increased area as compared to the standard layout transistors. In applications such as particle physics, the in-pixel circuits must be kept as compact as possible in order to achieve small pixel sizes and, thus, high spatial resolutions.

Radiation damage effects have been studied on transistors fabricated on the TowerJazz 180 nm

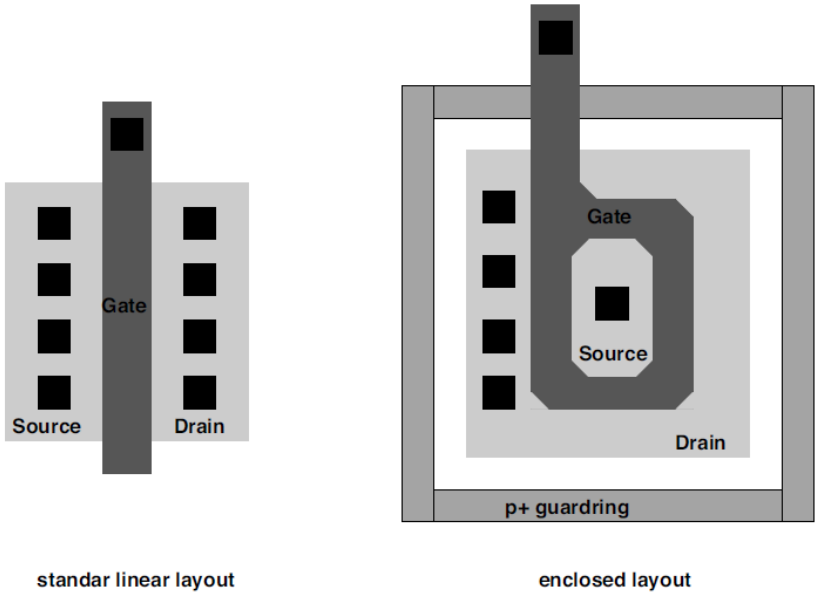

Figure 3.17: Layout of a typical ELT transistor (right). For comparison, the layout of a standard MOS transistor is shown on the left. Figure reproduced from [44].

technology [59]. Results for NMOS transistors with a minimum gate length (180 nm) and different widths are shown in figure 3.18. As expected due to the RINCE effect, minimum width

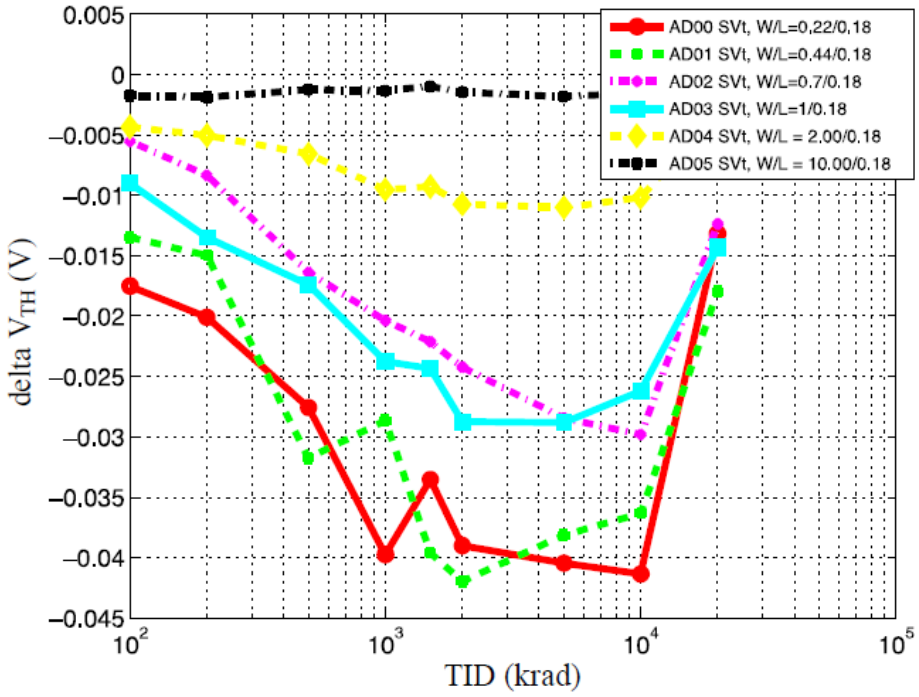

Figure 3.18: Threshold voltage shift as a function of TID for NMOS transistors with different gate width for a minimal gate length of 180 nm. The points at 2×10^4 krad correspond to 24 h of annealing after a dose of 1×10^4 krad. Figure reproduced from [59].

transistors are the most affected by TID in terms of threshold voltage shift. Degradation of structures wider than 1 μm is negligibly small. Note that the threshold shift saturates after a few MRad. This can be related to the slowly building interface traps starting to counteract the effect of oxide traps, leading to the characteristic TID bump mentioned above.

Transistors have different responses to TID damage depending on their structure, oxide thickness and actual fabrication process. The increase of noise in irradiated devices can impair the functionality of low-noise circuits. Hence, studying the radiation hardness of a particular technology is a requisite to determine its viability in a certain environment. Results on the TowerJazz 180 nm technology can be found in the literature and will be considered when analysing the results of chapter 7. Chapter 7 is devoted to study the effects of radiation damage on full demonstrator chips fabricated in this technology.

3.14 Summary

Solid state detectors are widely used in HEP experiments to measure the track position of different particles. Most detectors are made of Silicon, due to the abundance of this material and the good understanding of its electrical properties. Silicon atoms contain 14 electrons: 10 forming the core of the atom and 4 creating the electrical bond between neighbouring atoms. The latter are known as valence electrons, with an energy distribution divided into bands. Electrons in the valence band participate in the chemical bond between atoms, while electrons in the conduction band generate a current in the presence of an electric field. These are the so-called charge carriers and are responsible for the measurable electrical signal of the detector. However, electrons are excited to the conduction band at room temperature generating large currents as compared to that of a crossing particle. The solution is to use a junction of an n-type (Silicon with elements from the column V of the periodic table) and a p-type (Silicon with elements from the column III) material. This is known as a p-n junction and conforms the most fundamental block of most semiconductor devices. Around the juncture between the p-type and n-type materials, there is a region depleted of free charge carriers with a naturally generated electric field. Particles crossing this region generate charge carriers that move in the presence of the electric field and can be detected by attaching suitable electronics to the sensor terminals. These electronic circuits are typically required to amplify and shape the signal of incoming particles, and are made of transistors. Transistors are three-terminal devices in which the voltage of one of the terminals controls the current flow between the other two. Transistors are known as NMOS if the terminals are made of n-type silicon and the conducting channel is made of p-type silicon or as PMOS if it is the other way around.

The readout electronic circuits are usually designed in a different board and connected to the detecting material. However, the CMOS technology is capable of embedding NMOS and PMOS transistors in the sensing substrate, creating integrated devices capable to detect particles

and process its generated signal. CMOS detectors have an n-well implant acting as the charge collecting electrode of the device and a lowly doped p-type epitaxial layer, which creates a region in the bulk depleted of free charge carriers. The NMOS and PMOS transistors are usually isolated from the sensor bulk by different layers of n-type or p-type Silicon.

Transistors contain insulating layers made of SiO_2 in their structure which makes them sensitive to Ionising Dose effects. Incoming particles not only generate charge carriers in the detector bulk, but also in the transistors of the readout circuit. These charge carriers react around the SiO_2-Si interface generating two types of ionising radiation damage: oxide traps and interface traps. In PMOS transistors the effect of oxide traps and interface traps adds together while in NMOS transistors they counteract each other. This creates a characteristic shape in the performance of NMOS transistors after receiving different doses of radiation damage known as TID bump. Radiation damage results in an increase of the transistor noise and in a shift of the voltage level required to conduct current inside the transistor, the effects being more severe for smaller transistors. Different techniques can be employed to mitigate this type of damage such as using bigger transistors or more radiation hard transistor layouts. However, circuits have to be kept as compact as possible in order to achieve small detector sizes and, thus, high spatial resolutions.

In the outer-most layer of the ATLAS ITk, detectors are expected to receive ionising radiation damage of 80 MRad. The current CMOS detector technologies are not prepared to resist this harsh environment. Radiation damage produced in CMOS detectors fabricated in the TJ 180 nm technology is studied in the following chapters.

Chapter 4

CMOS sensors in TowerJazz 180 nm technology

4.1 Introduction to the TowerJazz 180 nm technology

CMOS sensors fabricated on the TowerJazz 180 nm technology have been employed for pixel tracking in the ALICE Inner Tracking System (ITS) [60]. Sensors in the ALICE experiment are expected to be subject to a moderate radiation level of 10^{13} 1 MeV n_{eq} / cm^2. The TowerJazz foundry offers flexibility in the doping profile of silicon materials that compose the CMOS sensors, which is required for tailoring its use in high radiation environments. A high resistivity epitaxial layer was used to increase the depletion region around the collection electrode. The epitaxial layer of the designs presented in this section feature a resistivity of over 1 kΩ and a thickness between 25 and 30 μm. Moreover, transistors of this technology have a thin oxide thickness of 3 nm, benefiting from the increased radiation tolerance of thin oxide technologies (see section 3.13). The design uses a small pixel collection electrode to minimise the capacitance, consequently reducing the noise and power consumption of the device. The electrode is several μm from the in-pixel transistors to reduce the lateral capacitance between the n-well and the wells containing the front-end transistors. A deep p-well is implemented to shield the PMOS transistors of the CMOS circuitry and avoid competing with the collection electrode for charge. A cross-section of the pixel design employed on the ALICE experiment is depicted in figure 4.1. As pointed out in the figure, the depletion region of this design extends from the collection electrode towards the p-epitaxial layer but does not cover the whole area of the pixel. A reverse bias can be applied to the sensor substrate to further increase the area of the depleted zone, although its lateral component is more difficult to extend [61]. Note that the voltage applied to the p^+ substrate is also seen by the p-well of NMOS transistors. Hence, the voltage applied to this implant is limited by the breakdown voltage of the source/drain junctions with the transistor bulk, which occurs at around 8 V. Charge collected outside the depleted region is still collected by diffusion but is slow and more sensitive to radiation damage (see section 3.5). Although this

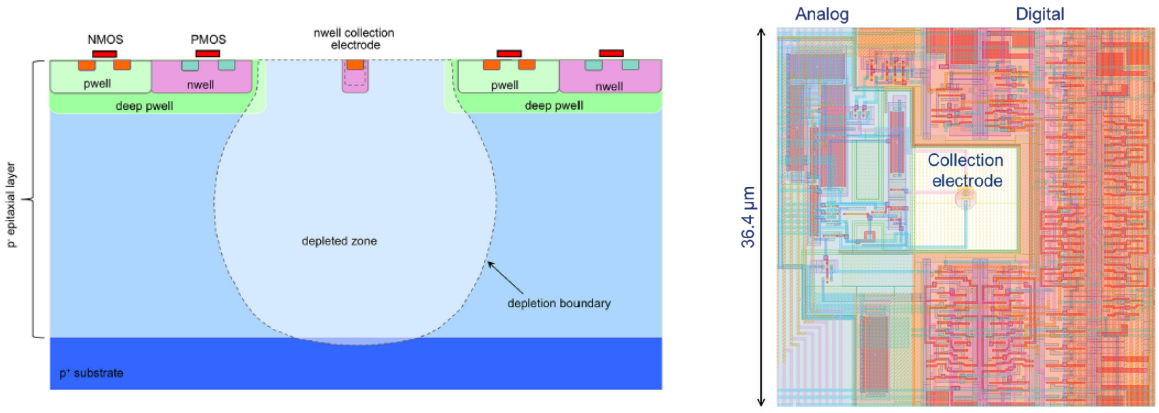

Figure 4.1: Cross-section of a pixel fabricated on the standard TowerJazz 180 nm technology (left). Figure reproduced from [61]. Top view of a pixel in MALTA showing the distribution of the Analog and Digital electronic circuits (right).

design satisfies the requirements of the ALICE experiment, it is not radiation hard enough to satisfy the more challenging environment of the ATLAS ITk.

A process modification was proposed aiming to achieve a more uniform depletion region across the pixel [61]. The proposed solution consists of placing a planar low doped n-type junction between the electrode and the p-type epitaxial layer. A new pixel design implementing this modified process is depicted in figure 4.2. In this case, depletion starts from the n-type layer

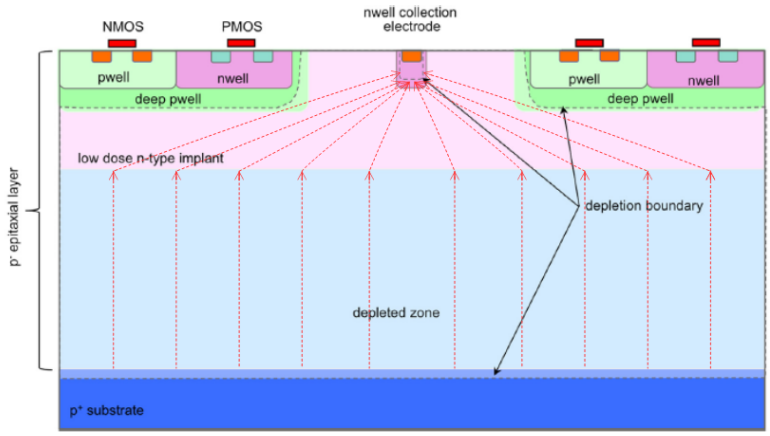

Figure 4.2: Cross-section of a pixel fabricated on the modified TowerJazz 180 nm technology. The red lines represent an approximation of the direction of the electric field. Figure reproduced from [61].

and uniformly extends to cover the full pixel area. The p-type epitaxial layer is depleted even at 0 V and further reverse bias can be applied to deplete the region around the collection electrode. The dose of the planar n-type implant is kept sufficiently low to be fully depleted at a few volts while maintaining a low sensor capacitance of a few fF. This design has the extra advantage of

insulating the deep p-well layer of the in-pixel electronics from the substrate, allowing to apply an independent bias on both materials.

4.2 Investigator1

The TowerJazz Investigator1 is a monolithic CMOS sensor designed in the standard TowerJazz 180 nm technology [63]. Its design was motivated by the intention to study the parameters that affect the pixel analog performance. For this purpose, the Investigator1 is divided in 134 mini-matrices featuring different pixel designs. Each mini-matrix is composed of 10 x 10 identical pixels: 8 x 8 active pixels surrounded by dummy pixels, implemented to minimise any possible edge effect (see figure 4.3). The dummy pixels are identical to the active pixels but are not connected to the readout chain. These are included to protect the active pixels from disturbances in the electric field, parasitic leakage currents and physical damage at the Silicon edges. A guard ring structure has been added around the matrix to prevent any charge flow from neighbouring matrices. The chip has a total area of $5.722 \times 5\,\mathrm{mm}^2$. The following pixel characteristics vary

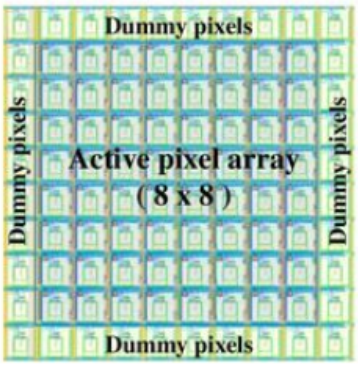

Figure 4.3: Pixel arrangement of the Investigator1 chip.

between mini-matrices:

- Pixel pitch: The pixels in the Investigator1 chip have a square shape with a pitch ranging from $20\times20\,\mu\mathrm{m}^2$ to $50\times50\,\mu\mathrm{m}^2$.
- Collection electrode size (Collsize): The collection electrode consists of an octagonal-shaped n^+-type implant which dimensions, specified by the parameter *collsize* in figure 4.4, range from 1 to $20\,\mu\mathrm{m}$. The octagonal shape is implemented to minimise the pixel capacitance, as it is the closest shape to a circle that can be implemented.
- Spacing: The *spacing* parameter refers to the space between the collection electrode and the deep p-well implants containing the transistors as noted in figure 4.4. It varies from 1 to $18.5\,\mu\mathrm{m}$.

- Opening: The pixel opening is a function of the spacing and collsize parameters, as can be seen in figure 4.4.
- P-well: The deep p-well implant is needed to shield the n-wells of NMOS transistors and the epitaxial layer. Its horizontal extension varies between minimum deep (only covering the n-well implants), maximum deep (covering all transistors) and medium deep (middle term between the two). Refer to figure 5.10 for a schematic of the three options.
- Reset: The electrode reset mechanism is implemented either as a simple diode or as a PMOS switch (also referred as *active*). More information about the two reset types is given in the text below.
- Transistor: Two different source followers (standard and special) located either inside or outside the collection electrode are implemented in different mini-matrices. The source follower transistor is marked as M2 in figure 4.6.

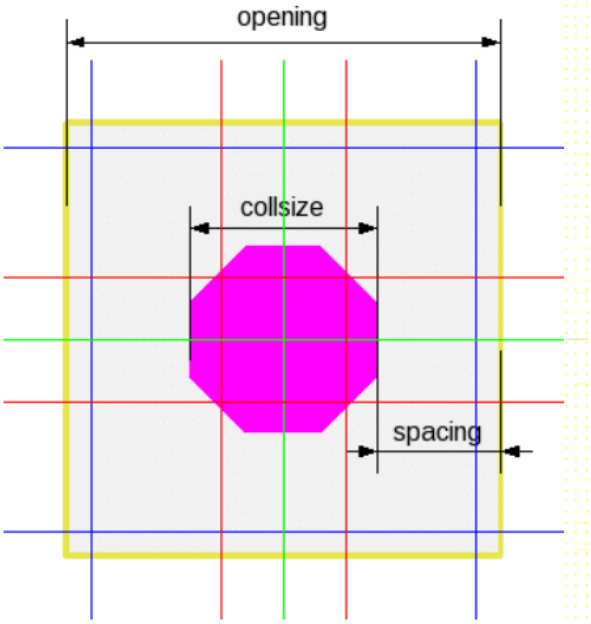

Figure 4.4: Schematic of the pixel cell and its defining measures.

Table 4.1 lists the combinations of the mentioned parameters for every mini-matrix. In terms of structure, the Investigator1 chip is divided into two identical matrices. Each matrix is organised in 15 columns of mini-matrices with increasing pixel pitch (from 20x20 μm^2 to 50x50 μm^2). This structure is shown in figure 4.5. The 8×8 active pixels of each mini-matrix are connected to wire-bond pads and can be externally read out, whilst the dummy pixels are left floating.

Pixel Pitch: 20 x 20 μm^2						**Pixel Pitch: 28 x 28 μm^2**					
MM N^o	**Collsize**	**Spacing (μm)**	**P-Well**	**Reset**	**Transistor**	**MM N^o**	**Collsize**	**Spacing (μm)**	**P-Well**	**Reset**	**Transistor**
0	3 x 3 oct	1	max deep	active	Standard Outside	**68**	min. oct	1	max deep	active	Standard Outside
1	3 x 3 oct	2	max deep	active	Standard Outside	**69**	min. oct	2	max deep	active	Standard Outside
2	3 x 3 oct	3	max deep	active	Standard Outside	**70**	min. oct	3	max deep	active	Standard Outside
3	3 x 3 oct	4	max deep	active	Standard Outside	**71**	min. oct	4	max deep	active	Standard Outside
4	3 x 3 oct	5	max deep	active	Standard Outside	**72**	min. oct	5	max deep	active	Standard Outside
5	min. oct	3	max deep	active	Standard Outside	**73**	2 x 2 oct	1	max deep	active	Standard Outside
6	2 x 2 oct	3	max deep	active	Standard Outside	**74**	2 x 2 oct	2	max deep	active	Standard Outside
7	4 x 4 oct	3	max deep	active	Standard Outside	**75**	2 x 2 oct	3	max deep	active	Standard Outside
8	5 x 5 oct	3	max deep	active	Standard Outside	**76**	2 x 2 oct	4	max deep	active	Standard Outside
9	3 x 3 oct	1	max deep	diode	Standard Outside	**77**	2 x 2 oct	5	max deep	active	Standard Outside
10	3 x 3 oct	2	max deep	diode	Standard Outside	**78**	3 x 3 oct	1	max deep	active	Standard Outside
11	3 x 3 oct	3	max deep	diode	Standard Outside	**79**	3 x 3 oct	2	max deep	active	Standard Outside
12	3 x 3 oct	4	max deep	diode	Standard Outside	**80**	3 x 3 oct	3	max deep	active	Standard Outside
13	3 x 3 oct	5	max deep	diode	Standard Outside	**81**	3 x 3 oct	4	max deep	active	Standard Outside
14	3 x 3 oct	1	max deep	diode	Special Inside	**82**	3 x 3 oct	5	max deep	active	Standard Outside
15	3 x 3 oct	2	max deep	diode	Special Inside	**83**	4 x 4 oct	1	max deep	active	Standard Outside
16	3 x 3 oct	3	max deep	diode	Special Inside	**84**	4 x 4 oct	2	max deep	active	Standard Outside
17	3 x 3 oct	4	max deep	diode	Special Inside	**85**	4 x 4 oct	3	max deep	active	Standard Outside
18	3 x 3 oct	5	max deep	diode	Special Inside	**86**	4 x 4 oct	4	max deep	active	Standard Outside
19	3 x 3 oct	1	max deep	diode	Special Outside	**87**	4 x 4 oct	5	max deep	active	Standard Outside
20	3 x 3 oct	2	max deep	diode	Special Outside	**88**	5 x 5 oct	1	max deep	active	Standard Outside
21	3 x 3 oct	3	max deep	diode	Special Outside	**89**	5 x 5 oct	2	max deep	active	Standard Outside
22	3 x 3 oct	4	max deep	diode	Special Outside	**90**	5 x 5 oct	3	max deep	active	Standard Outside
23	3 x 3 oct	5	max deep	diode	Special Outside	**91**	5 x 5 oct	4	max deep	active	Standard Outside
24	3 x 3 oct	1	max deep	active	Standard Inside	**92**	5 x 5 oct	5	max deep	active	Standard Outside
25	3 x 3 oct	2	max deep	active	Standard Inside	**93**	3 x 3 oct	1	min deep	active	Standard Outside
26	3 x 3 oct	3	max deep	active	Standard Inside	**94**	3 x 3 oct	2	min deep	active	Standard Outside
27	3 x 3 oct	4	max deep	active	Standard Inside	**95**	3 x 3 oct	3	min deep	active	Standard Outside
28	3 x 3 oct	5	max deep	active	Standard Inside	**96**	3 x 3 oct	4	min deep	active	Standard Outside
29	3 x 3 oct	1	min deep	active	Standard Outside	**97**	3 x 3 oct	5	min deep	active	Standard Outside
30	3 x 3 oct	2	min deep	active	Standard Outside	**98**	3 x 3 oct	3	medium deep	active	Standard Outside
31	3 x 3 oct	3	min deep	active	Standard Outside	**99**	3 x 3 oct	5	medium deep	active	Standard Outside
32	3 x 3 oct	4	min deep	active	Standard Outside	**100**	pALPIDEfs				
33	3 x 3 oct	5	min deep	active	Standard Outside	**101**	pALPIDEfs				
34	3 x 3 oct	3	medium deep	active	Standard Outside	**102**	pALPIDEfs				
35	3 x 3 oct	5	medium deep	active	Standard Outside	**103**	pALPIDEfs				
Pixel Pitch: 22 x 22 μm^2						**Pixel Pitch: 30 x 30 μm^2**					
MM N^o	**Collsize**	**Spacing (μm)**	**P-Well**	**Reset**	**Transistor**	**MM N^o**	**Collsize**	**Spacing (μm)**	**P-Well**	**Reset**	**Transistor**
36	3 x 3 oct	1	max deep	active	Standard Outside	**104**	3 x 3 oct	1	max deep	active	Standard Outside
37	3 x 3 oct	2	max deep	active	Standard Outside	**105**	3 x 3 oct	2	max deep	active	Standard Outside
38	3 x 3 oct	3	max deep	active	Standard Outside	**106**	3 x 3 oct	3	max deep	active	Standard Outside
39	3 x 3 oct	4	max deep	active	Standard Outside	**107**	3 x 3 oct	4	max deep	active	Standard Outside
40	3 x 3 oct	5	max deep	active	Standard Outside	**108**	3 x 3 oct	5	max deep	active	Standard Outside
41	min. oct	3	max deep	active	Standard Outside	**109**	min. oct	3	max deep	active	Standard Outside
42	2 x 2 oct	3	max deep	active	Standard Outside	**110**	2 x 2 oct	3	max deep	active	Standard Outside
43	4 x 4 oct	3	max deep	active	Standard Outside	**111**	4 x 4 oct	3	max deep	active	Standard Outside
44	5 x 5 oct	3	max deep	active	Standard Outside	**Pixel Pitch: 40 x 40 μm^2**					
45	3 x 3 oct	1	min deep	active	Standard Outside	**MM N^o**	**Collsize**	**Opening (μm)**	**P-Well**	**Reset**	**Transistor**
46	3 x 3 oct	2	min deep	active	Standard Outside	**112**	3 x 3 oct	30	max deep	active	Standard Outside
47	3 x 3 oct	3	min deep	active	Standard Outside	**113**	4 x 4 oct	30	max deep	active	Standard Outside
48	3 x 3 oct	4	min deep	active	Standard Outside	**114**	5 x 5 oct	30	max deep	active	Standard Outside
49	3 x 3 oct	5	min deep	active	Standard Outside	**115**	10 x 10 oct	30	max deep	active	Standard Outside
50	3 x 3 oct	1	medium deep	active	Standard Outside	**116**	15 x 15 oct	30	max deep	active	Standard Outside
51	3 x 3 oct	3	medium deep	active	Standard Outside	**117**	20 x 20 oct	30	max deep	active	Standard Outside
52	3 x 3 oct	4	medium deep	active	Standard Outside	**118**	3 x 3 oct	30	max deep	active	Standard Inside
53	3 x 3 oct	5	medium deep	active	Standard Outside	**119**	4 x 4 oct	30	max deep	active	Standard Inside
54	pALPIDE max WPD oct					**120**	5 x 5 oct	30	max deep	active	Standard Inside
55	pALPIDE max WPD sqr					**121**	10 x 10 oct	30	max deep	active	Standard Inside
56	pALPIDE min WPD oct					**122**	15 x 15 oct	30	max deep	active	Standard Inside
57	pALPIDE min WPD sqr					**123**	20 x 20 oct	30	max deep	active	Standard Inside
Pixel Pitch: 25 x 25 μm^2						**Pixel Pitch: 50 x 50 μm^2**					
MM N^o	**Collsize**	**Opening (μm)**	**P-Well**	**Reset**	**Transistor**	**MM N^o**	**Collsize**	**Opening (μm)**	**P-Well**	**Reset**	**Transistor**
58	3 x 3 oct	1	max deep	active	Standard Outside	**124**	3 x 3 oct	30	max deep	active	Standard Outside
59	3 x 3 oct	2	max deep	active	Standard Outside	**125**	5 x 5 oct	30	max deep	active	Standard Outside
60	3 x 3 oct	3	max deep	active	Standard Outside	**126**	10 x 10 oct	30	max deep	active	Standard Outside
61	3 x 3 oct	4	max deep	active	Standard Outside	**127**	15 x 15 oct	30	max deep	active	Standard Outside
62	3 x 3 oct	5	max deep	active	Standard Outside	**128**	20 x 20 oct	30	max deep	active	Standard Outside
63	min. oct	3	max deep	active	Standard Outside	**129**	3 x 3 oct	40	max deep	active	Standard Inside
64	2 x 2 oct	3	max deep	active	Standard Outside	**130**	5 x 5 oct	40	max deep	active	Standard Inside
65	4 x 4 oct	3	max deep	active	Standard Outside	**131**	10 x 10 oct	40	max deep	active	Standard Inside
66	5 x 5 oct	3	max deep	active	Standard Outside	**132**	15 x 15 oct	40	max deep	active	Standard Inside
67	3 x 3 oct	3	min deep	active	Standard Outside	**133**	20 x 20 oct	40	max deep	active	Standard Inside

Table 4.1: List of all Mini-Matrices (MM) in the TowerJazz Investigator1 chip and their properties. Refer to figure 4.4 for the definition of the opening, spacing and collsize parameters.

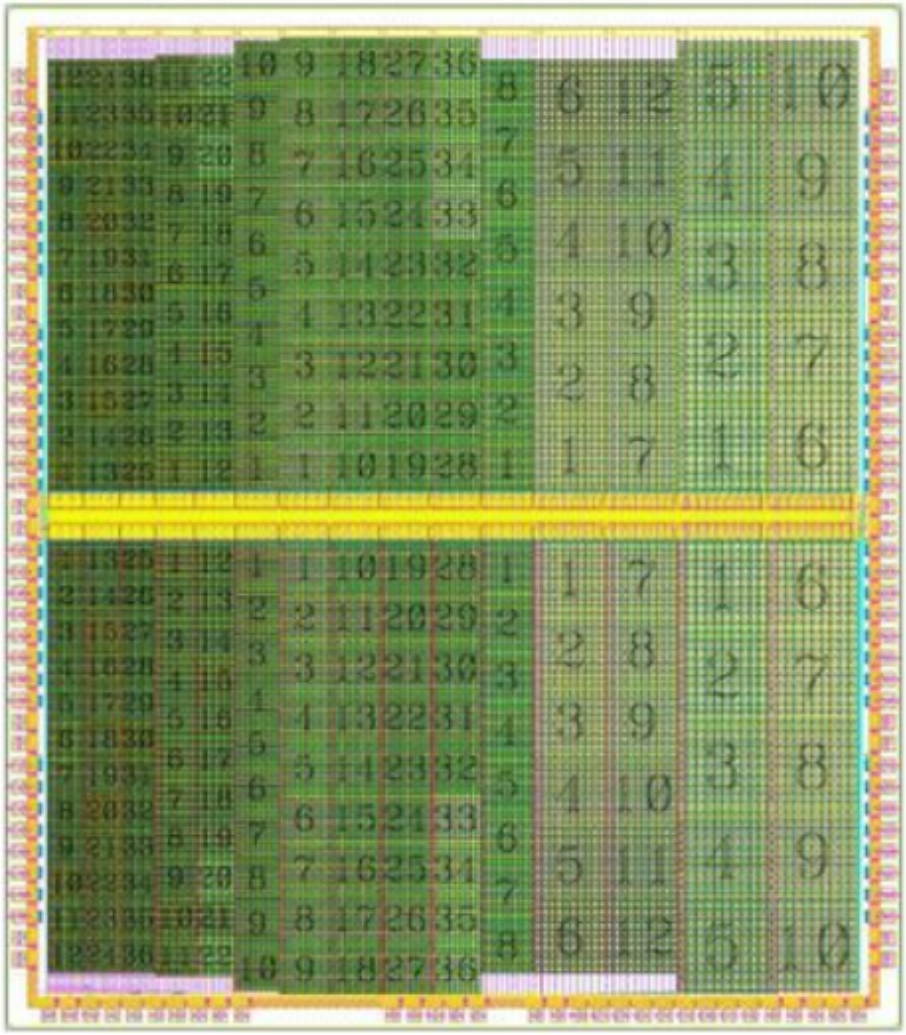

Figure 4.5: Investigator1 chip layout. The two matrices are separated by the yellow band. The mini-matrices with different pixel pitch are distributed in columns from 20x20 μm^2 in the left hand side to 50x50 μm^2 in the right hand side. The groups of mini-matrices with the same pixel pitch are numbered from 1 to N.

Principle of operation

A schematic of the pixel circuit, the column circuit and the output buffer circuit are presented in figure 4.6. There are variations of this circuit including different reset mechanisms and source followers (M2 transistor). The collection electrode, represented as a diode (D0) in the *Pixel circuit* block, is connected to a source follower (M2) which isolates it from the rest of the circuit. The electrode is initially biased with a voltage V_i. According to equation 3.32, the electrode stores a charge $Q_i = C_e V_i$. An incoming particle generates a charge Q_p in the epitaxial layer that is subtracted from the input charge in the electrode capacitor. Thus, the voltage level of the electrode varies to $V_f = (Q_i - Q_p)/C_e$. At the output node of the source follower, the voltage level is proportional to the voltage at D0, being the proportionality factor the gain of the transistor (g_S

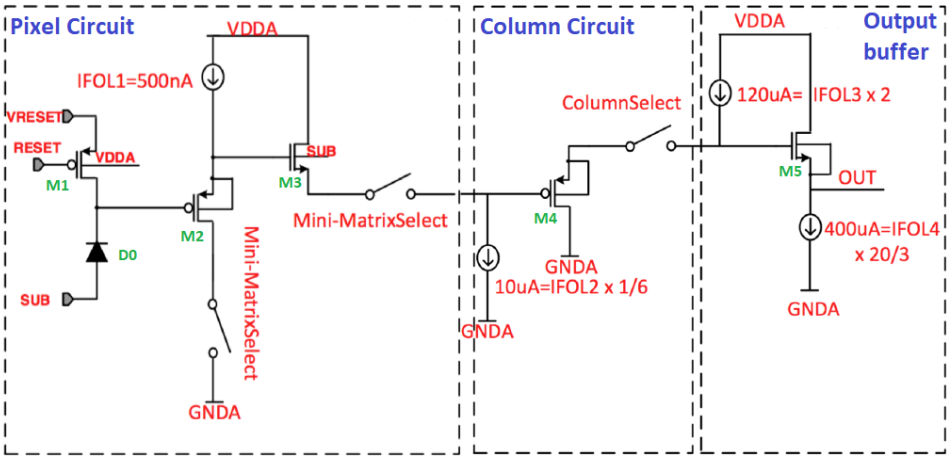

Figure 4.6: Schematic of the Investigator1 pixel readout circuitry with the active reset variation. Figure reproduced from [61] and modified.

$\simeq 1$). After a particle hit, the voltage at the output of M2 drops by:

$$\Delta V = g_S V_f - g_S V_i = -\frac{g_S}{C_e} Q_p \tag{4.1}$$

Hence, an incoming particle produces a fast voltage drop at the output of the front-end circuitry that is proportional to the charge deposited by itself. Note that, as discussed in section 3.11.2, small values of C_e associated to small electrode areas lead to large ΔV.

Aside from a particle hit, the sensing node is also discharged due to the leakage current, which produces a small and constant effect. Hence, a reset mechanism is needed to restore the default voltage value. The pixel reset strategy is varied between mini-matrices. Two types of reset mechanisms are implemented to restore the collection electrode. Those are specified as *diode* and *active* in table 4.1. A schematic of the diode reset solution is shown in figure 4.7 (left). In steady-state conditions, the D1 diode is biased by the small leakage current of the collecting electrode (D0). The electrode is thus biased by a voltage defined by the difference between VRESET_D and the voltage drop on D1 (around 500 mV). When a particle is detected the voltage drop in D0 causes D1 to be forward biased, increasing the current flow inside this device and charging the input node back to its original value. This process takes several hundreds of microseconds, and is slow if comparing it to the duration of the signal due to a particle hit.

The active reset strategy is shown in figure 4.7 (right). It is implemented as a PMOS transistor (M1) acting as a switch. A voltage pulse is periodically applied in the IRESET node to switch on the M1 transistor, connecting momentarily the collection diode (D0) with the VRESET_P voltage source. This restores the voltage of the diode until the next reset phase.

The output of the pixel circuit is connected to the column circuit through the *Mini-MatrixSelect* switch and this, in turn, is connected to the output buffer circuit through the *ColumnSelect* switch. These switches can be configured externally to select a specific mini-matrix to be read

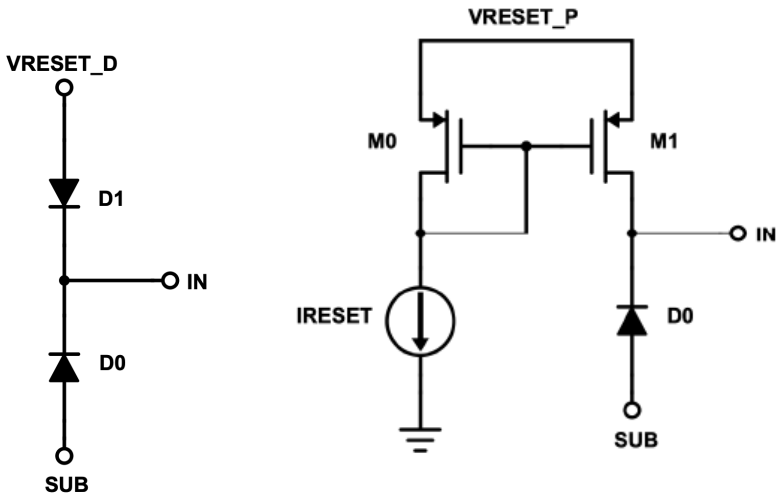

Figure 4.7: Schematic of the diode (left) and active (right) reset implementations of the TJ Investigator1.

out. The entire circuit is biased by four currents (IFOL1 - IFOL4) that are provided externally.

In this thesis, the Investigator1 is used as a tool to study how the different pixel parameters affect the collection of charge. These results are presented in chapter 5.

4.3 Demonstrators

4.3.1 TowerJazz MALTA

MALTA (*Monolithic from ALICE to ATLAS*) is a full size ATLAS ITk demonstrator for the HL-LHC. It is a monolithic active pixel detector designed in the modified TJ 180 nm technology. It contains a 512 x 512 pixel matrix with a 36.4 μm squared pixel size, and an active area of 18.3 mm^2. Each pixel is equipped with an analog test pulse, a masking capability, a signal shaper and a discriminator. The MALTA matrix is divided into eight sectors featuring different electrode sizes (2 or 3 μm), spacing distances (3.5 or 4 μm) and deep p-well horizontal sizes (maximum or medium deep p-well). A schematic of the MALTA pixel matrix layout is shown in figure 4.8. The choice of these parameters was done following results obtained with the Investigator1 chip (see chapter 5). Moreover, two reset mechanisms are implemented in the different sectors of the chip: a diode and a PMOS reset. The leading edge output of the discriminator is injected into the double column pixel logic, from which a coarse analog measurement through time-walk of the leading edge with respect to a reference signal can be performed on the periphery. The shaping time of the pixel is up to 20 ns, compatible with the requirements for bunch-crossing identification of the HL-LHC. In order to lower its power consumption, the readout of the chip is designed to be fully asynchronous for which no clock distribution is required.

This is explained in the corresponding section.

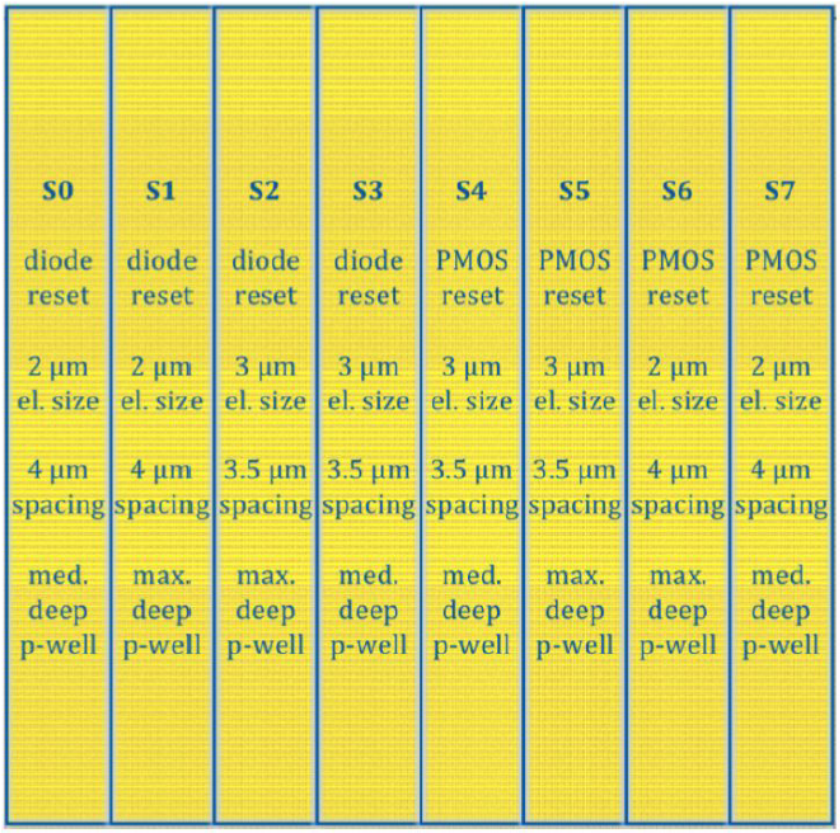

Figure 4.8: Schematic of the MALTA pixel matrix layout. The electrode size and spacing parameters are defined in figure 4.4. The schematic of figure 4.6 is represented with the active reset solution only showing the M1 transistor and the D0 electrode.

The analog front end

All pixel collecting electrodes of MALTA are connected to the input of an analog front-end circuit that amplifies and shapes the signal induced by a traversing particle. The full schematic and a simplified version of the front-end architecture are shown in figure 4.9. This architecture is based on the front-end designed for the ALPIDE CMOS sensor [64] for the ALICE ITS upgrade and optimised for a faster response and better TID tolerance. The front-end of MALTA consists of a charge-sensitive shaper-amplifier (transistors M0 to M6) and a discriminator (transistors M7 to M9).

Charge generated by a traversing particle induces a voltage drop on the collecting electrode (see equation 3.32). Voltage needs to be restored to its original value in order to allow a continuous operation of the chip. This is done through the two reset mechanisms implemented in MALTA: the simple diode and the PMOS current source. These two solutions were previously implemented in the Investigator1 chip (see section 4.2).

The generated signal goes through a pre-amplification stage to increase the SNR on the next

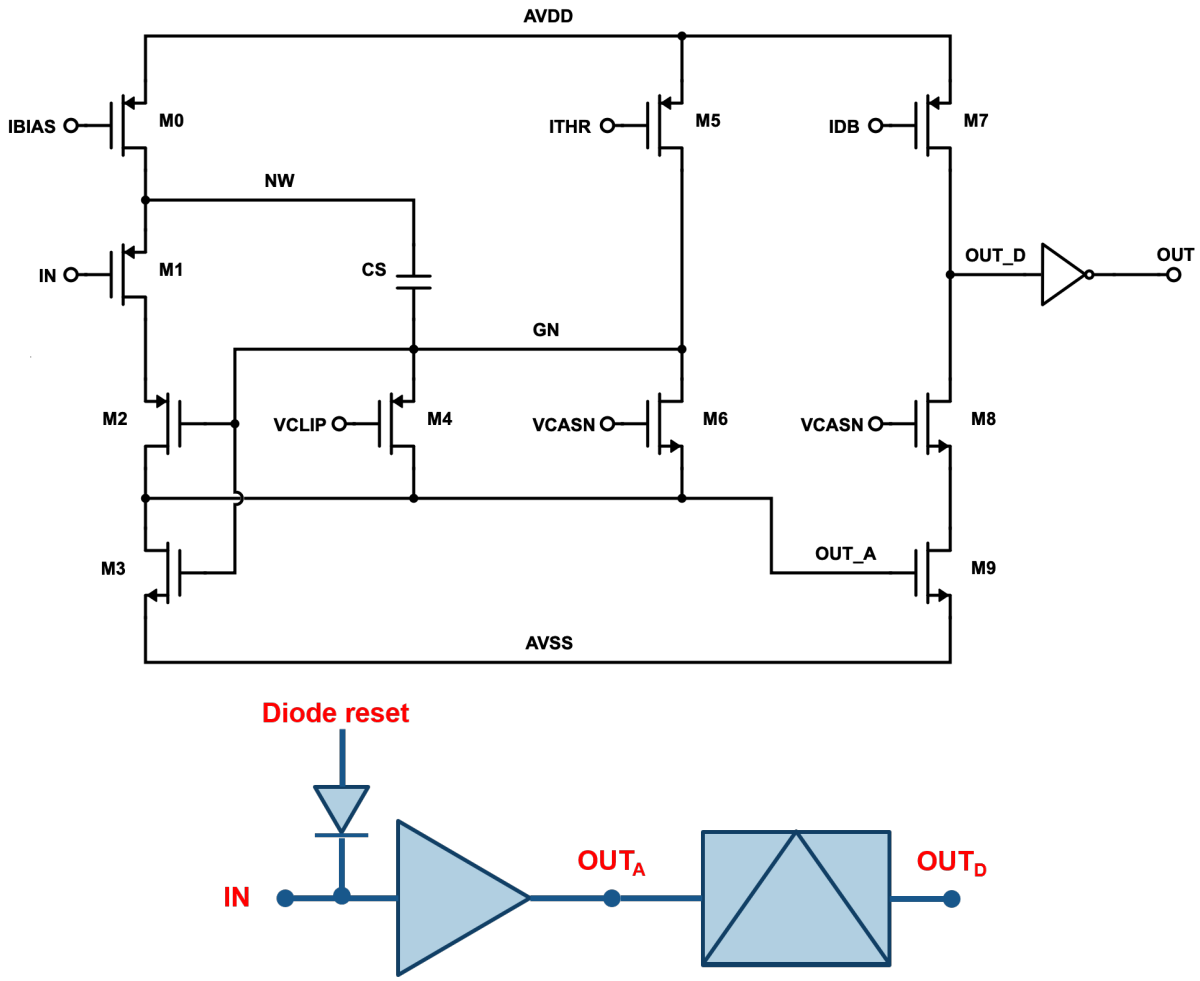

Figure 4.9: Schematic of the MALTA front-end.

operations. Shaping is also important to fulfill the timing requirements of the chip. Transistors M0 to M6 form the pre-amplification and shaping circuit. M0 provides a current (IBIAS) to the M1 transistor. The current of the input node is driven through the M1 transistor, which acts as a source follower and loads the capacitance C_S. The charge is then transferred to the parasitic capacitance of the OUT_A node, C_A. The voltage drop on OUT_A can be written as a function of the voltage drop on the IN node as [20]:

$$\Delta V_{OUT_A} = \frac{Q_S}{C_A} = \frac{C_S}{C_A}\Delta V_{IN} = \frac{C_S}{C_A}\frac{Q_{IN}}{C_{IN}} \tag{4.2}$$

Note that, due to the small value of the C_{IN} capacitance (~ 5 fF), the signal for a given Q_{IN} charge is already relatively large without amplification. The voltage difference is further enhanced if $C_S/C_A \gg 1$. C_S is provided by a large PMOS transistor with all its terminals connected together and has a value of around 150 fF. The only contributors to C_A are parasitic capacitance of the transistors connected to this node and has a value of around 5 fF, which is small compared to C_S. Hence, this circuit forms the pre-amplification phase of the front-end circuit. The M2 transistor

is placed to prevent capacitive coupling between the OUT_A and input nodes. Transistors M3, M4, M5 and M6 influence the shaping of the signal at the OUT_A node. M5 and M6 define the baseline voltage on OUT_A. When there is a rise in the OUT_A voltage, following a particle hit, M6 is turned off and the ITHR current charges the GN node. Current goes through the line and increases the M3 gate voltage. M3 starts conducting current, discharging the OUT_A node. Hence, shaping is done through this mechanism: pulse duration is controlled by the ITHR current and amplitude by the IBIAS current and the C_S capacitance. Further shaping for high input charges is provided by the clipping transistor M4, which shortens the duration of the pulse to maintain it within 25 ns. The voltage at which clipping is applied is controlled by the VCLIP bias.

Transistors M7, M8 and M9 form the discriminator circuit. The DC level of these transistors is set by the baseline voltage on OUT_A. M7 is biased by a current IDB which is set higher than the baseline current of the discriminator. In this situation, OUT_D is close to the AVDD voltage (1.8 V) and M7 is turned off. When voltage increases on OUT_A, the M9 transistor is turned on and starts to discharge the OUT_D node. As the OUT_A voltage decreases, the current driven by M9 gets smaller than IDB and the OUT_D node is charged again. Hence, threshold is controlled by the baseline level on OUT_A, which can be modified by changing VCASN, and the IDB current. In the final step, the signal is inverted to produce the digital pulse that is then fed to the digital electronics. A simulation of the signals on the IN, OUT_A and OUT_D nodes is shown in figure 4.10.

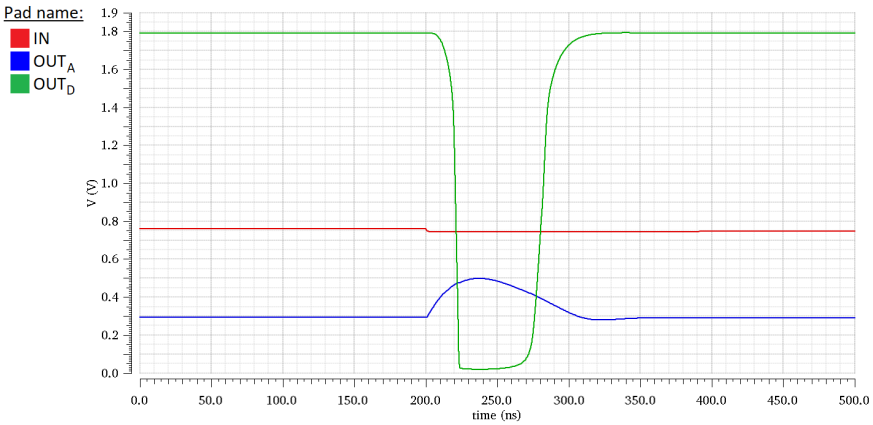

Figure 4.10: Transient response of the MALTA analogue front-end circuit: signals at the input node (IN), the output of the amplifier (OUT_A) and the output of the discriminator (OUT_D). Figure reproduced from [20].

Sizing of the front-end transistors plays a critical role on the timing, noise characteristics and radiation hardness of the circuit. All the transistors connecting to the OUT_A have to be kept small in order to maintain a small capacitance and good amplification on this node. However, timing has to fulfil the 25 ns bunch-crossing time of the LHC which puts an extra constraint on the sizing of some transistors. Transistor M1 has a big impact on the rise time of the signal and,

hence, a high transconductance (i.e. a large area) on this device is desired. M3 has an impact on the C_A capacitance and has to be kept small. However, if this transistor is too small the M1 source couples to the M3 gate through the capacitance C_S. A voltage drop on the M1 transistor can then drop the M3 gate voltage turning off this transistor and allowing more current to flow through the OUT_A node, effectively increasing the pre-amplifier gain. Another motivation to enlarge the transistor area when possible is the fact that the threshold voltage variation of transistors is reduced with $\sqrt{WL}$. Hence, larger transistors would reduce pixel to pixel variations on the front-end. Critical transistors are M5, influencing the gain of the amplifier and the DC level on OUT_A, and M6 only affecting the baseline. M5 and M6 have an area of 20 μm^2 and 1 μm^2 respectively, which is quite large compared to the area of the other transistors. This leaves transistors M3 and M9 as the main contributors to the charge threshold variation of the discriminator circuit. In any case, threshold dispersion was simulated [20] to be quite small and no in-pixel threshold tuning was included. Contributions to the overall noise of the front-end have also been simulated. The thermal noise of M1 is found to be the highest contributor to it (38 %), followed by the thermal noise (28 %) and 1/f noise (19 %) of transistor M3. Finally, for radiation hardness considerations, transistor M6 was made with an enclosed layout.

Some chip functionalities were included at the analogue front-end circuit level. These include pixel masking and test pulse injection. Pixel masking is implemented through three NMOS transistors acting as a switch that electronically disconnects the M9 transistor on the discriminator stage from the output of the pre-amplifier. The three switches correspond to the row, column and diagonal coordinates of the pixel. Digital signals control the ON and OFF states of the switches and can be addressed through the Slow Control module. Hence, this introduces the capability to deliberately disconnect the output of noisy or malfunctioning pixels and prevent large leakage currents that will affect the digital electronics.

The pulse injection capability is designed to induce user-controlled test pulses at the input node of the front-end allowing more control of the rate and value of the charge injected. A schematic of this capability is shown in figure 4.11. Pixels to be pulsed are selected through two PMOS switches, determining the column and row coordinates of the pixel (ROW_SEL and COL_SEL). V_{HIGH} and V_{LOW} are DC voltages set by the user through the Slow Control module. A pulse command can be sent through the same module to the enabled pixels, which would generate a voltage step of an amplitude V_{HIGH} - V_{LOW}. This voltage step generates a signal on the input node through the 230 aF capacitance (C_C). The amount of charge injected is given by [20]:

$$Q_{IN} = C_{IN}\Delta V_{IN} = C_{IN}\frac{C_C}{C_{IN}+C_C}\Delta V_{OUT} \approx C_C\left(V_{HIGH}-V_{LOW}\right). \tag{4.3}$$

Where it is assumed that $C_{IN} \gg C_C$. Hence, the charge injected to the input capacitance can be calculated for a given $V_{HIGH}-V_{LOW}$ value:

$$Q_{IN} = 2.30\times10^{-16}\frac{C}{V}\cdot 6.25\times10^{18}\frac{e^-}{C} = 1.43\frac{e^-}{mV}\left(V_{HIGH}-V_{LOW}\right). \tag{4.4}$$

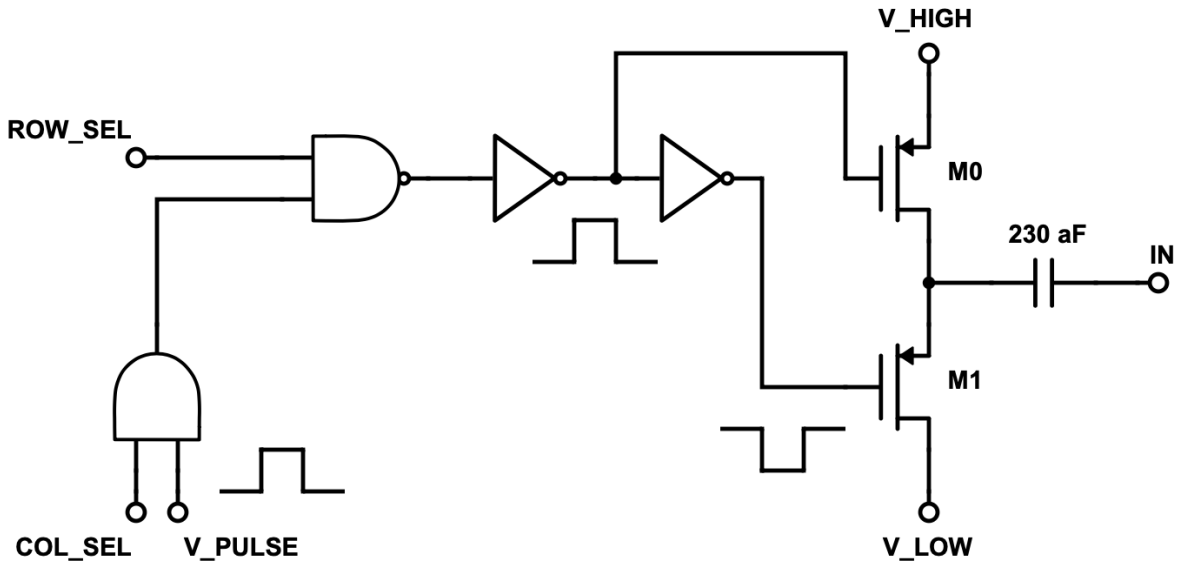

Figure 4.11: Schematic of the pulse injection capability circuitry. Figure obtained from [20].

Moreover, several pixels in MALTA have access to the OUT_A node, which outputs the analogue response of the pixel before the signal discrimination and pulse shaping stages.

Asynchronous readout architecture

The MALTA chip features a novel readout in which the hit information is transmitted asynchronously to the periphery circuit. This is done to avoid the propagation of a synchronisation clock within the pixel matrix, consequently reducing the power consumption of the digital circuit. The readout organised in columns of 2 x 512 pixels, referred as double columns. Every double column is composed of 32 "red" and 32 "blue" pixel groups of 2 x 8 pixels each, which alternate their readout bus chain as shown in figure 4.12. Every pixel within a group generates a hit signal of programmable width (0.5 ns to 2 ns) at the output of the pixel discriminator. This hit signal is input in one line of a 16 bit wide pixel bus, which transmits the pattern of pixels hit inside a group. In order to identify the pixel group, a static pixel group number is generated on the 5-bit group address bus with each hit. Red and blue groups are processed separately and in parallel, to minimise any hit information loss due to simultaneous signals transmitted to the double column bus. The 16 bit pixel bus and the group address are asynchronously transmitted on the column bus down to the end-of-column. The maximum latency of signal propagation along the double column is around 5 ns. Signals from all groups are merged onto a common bus at the end of the double column, where an arbitration logic is implemented to time-sort the pulses in case of simultaneous signals. The arbitration logic works delaying simultaneous signals, to avoid collisions and hit losses on the bus, and keeping track of the amount of delays for later correction. Column address and time-stamping information is added to each hit signal inside the periphery logic, which allows bunch crossing identification for later trigger processing. The

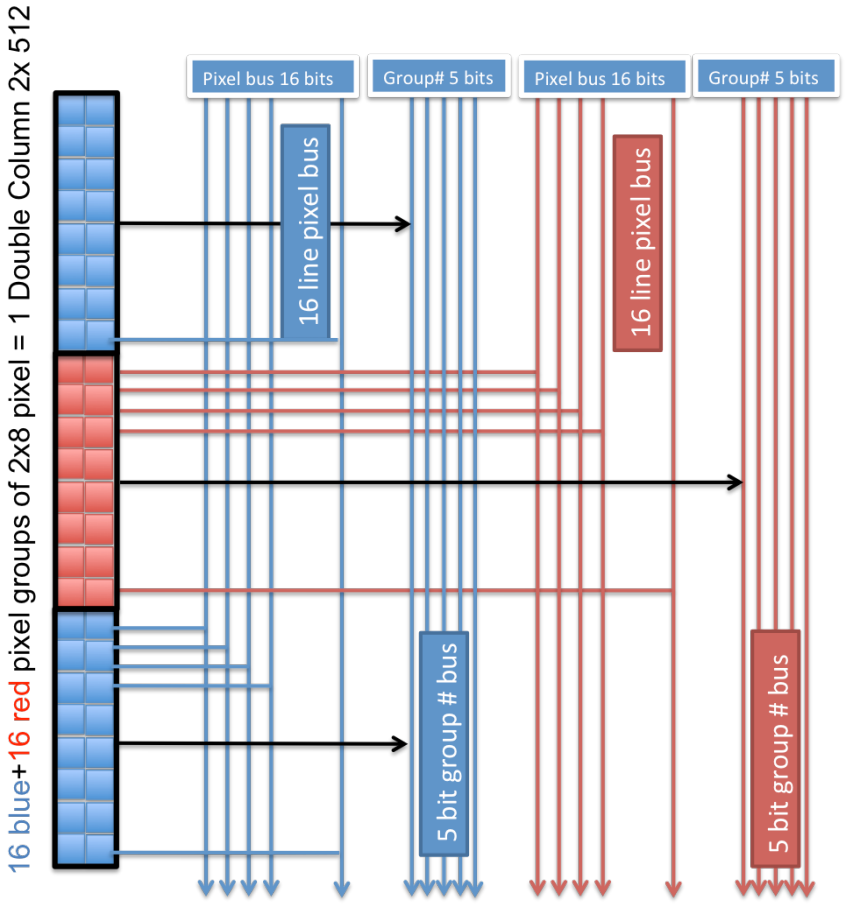

Figure 4.12: Schematic view of the MALTA double column readout.

process of arbitration and merging of signals is iterated over 10 levels to merge the signals of the 256 double-columns. The output of this process is a 40 bit wide chip bus (1 reference bit, 16 bits for the pixel address, 5 bits for the group address, 1 group identifier bit, 3 bits for the delay counter, 8 for the double-column address, 2 for the BCID timestamp and 4 for the chip address) that is asynchronously transmitted off the chip via LVDS drivers, designed to operate at up to 5 Gbps. These correspond to the hit information bits of the MALTA chip and are wire-bonded off the chip on 40 different differential output pads.

Slow Control block

The MALTA Slow Control (SC) block allows the operation of the MALTA chip by sending it 16-bit commands. The block consists of a finite state machine and a set of 16-bit registers. The register pool is used to momentarily store the SC commands and the finite state machine selects a specific MALTA register to implement the command. Each configurable feature of MALTA has an associated register that can be accessed from the SC.

The 16-bit configuration word is divided into action, value and non-information bits. The action bits define the specific action to be taken (e.g. enable masking of pixel row). The action

bits are, in turn, subdivided into option (4-bits), sub-option (variable length) and sub-sub-option (variable length) bits. The option bits define the general action to be taken ("masking"), the sub-option bits define a subgroup of this action (masking "a row") and the sub-sub-option bits define an extra option ("enable" masking a row). The value bits contain the value to be stored into the SC register defined by the action bits. In the example above, this would be the specific row to be masked. If a command does not require all 16-bits to be specified the remaining bits, referred as non-information bits, are ignored. A list of the most commonly employed MALTA commands and its corresponding SC address is given in table 4.2.

Command	Option	Sub-option	Sub-sub-option	Value
Reset	0000	00	-	-
Reset pulse	0000	01	-	clock cycles
Clear registers	0000	10	-	register address
Mask column (disable)	0001	00	0	0 to 511
Mask column (enable)	0001	00	1	0 to 511
Mask row (disable)	0001	01	0	0 to 511
Mask row (enable)	0001	01	1	0 to 511
Mask diagonal (disable)	0001	10	0	0 to 511
Mask diagonal (enable)	0001	10	1	0 to 511
Mask double column (disable)	0001	11	0	0 to 255
Mask double column (enable)	0001	11	1	0 to 255
Pulse column (disable)	0010	00	0	0 to 511
Pulse column (enable)	0010	00	1	0 to 511
Pulse row (disable)	0010	01	0	0 to 511
Pulse row (enable)	0010	01	1	0 to 511
Pulse fixed	0010	10	-	-
Pulse delay 2 ns	0011	0111	-	-
Pulse delay 1 ns	0011	1011	-	-
Pulse delay 750 ps	0011	1101	-	-
Pulse delay 500 ps	0011	1110	-	-
Reset VCASN	0100	00000	-	0 to 127
Reset VCLIP	0100	00001	-	0 to 127
Reset VPULSE_HIGH	0100	00010	-	0 to 127
Reset VPULSE_LOW	0100	00011	-	0 to 127
Reset VRESET_P	0100	00100	-	0 to 127
Reset VRESET_D	0100	00101	-	0 to 127
Reset ICASN	0100	00110	-	0 to 127
Reset IRESET	0100	00111	-	0 to 127

Table 4.2 continued from previous page

Reset ITHR	0100	01000	-	0 to 127
Reset IBIAS	0100	01001	-	0 to 127
Reset IDB	0100	01010	-	0 to 127
Enable IDB	0100	10101	0100	0 to 1
Enable ITHR	0100	10101	0101	0 to 1
Enable IBIAS	0100	10101	0110	0 to 1
Enable IRESET	0100	10101	0111	0 to 1
Enable ICASN	0100	10101	1000	0 to 1
Enable VRESET_D	0100	10101	1001	0 to 1
Enable VRESET_P	0100	10101	1010	0 to 1
Enable VPULSE_LOW	0100	10101	1011	0 to 1
Enable VPULSE_HIGH	0100	10101	1100	0 to 1
Enable VCLIP	0100	10101	1101	0 to 1
Enable VCASN	0100	10101	1110	0 to 1
Enable Merger	0110	0000	-	0 to 1
Read register	1000	0 to 31	-	0

Table 4.2: List of the MALTA SC commands available in the chip and its corresponding 16-bit address.

The communication protocol of the MALTA SC requires three signals: a serial input, a serial output and a 10 MHz clock. In order to write a command into MALTA, a 16-bit word has to be sent to the SERIAL_INPUT pad. This must be preceded by an additional low bit, indicating the start of the transmission. After decoding the input word and taking the specific action, the SC issues a reply word in the SERIAL_OUTPUT pad. The reply contains either a confirmation of the command or the required information. If a non-valid action is attempted, the input word is returned. Again, the reply word is acknowledged by an additional low bit preceding the data.

Communication with the MALTA SC block is established by the MALTA readout described in section 6.4. The readout serves as the interface between the computer and the MALTA chip. The user can send the 16-bit SC commands from the computer into the readout which are then transmitted to MALTA.

4.3.2 TJ mini-MALTA

The MALTA chip was redesigned to improve the radiation tolerance of the pixel structure and of the front-end circuit. Two modifications of the pixel layout were included: an extra-deep p-well implant and a gap in the n-type implant. A cross-section of pixels with these two implementations is shown in figure 4.13. TCAD simulations show that both process modifications enhance

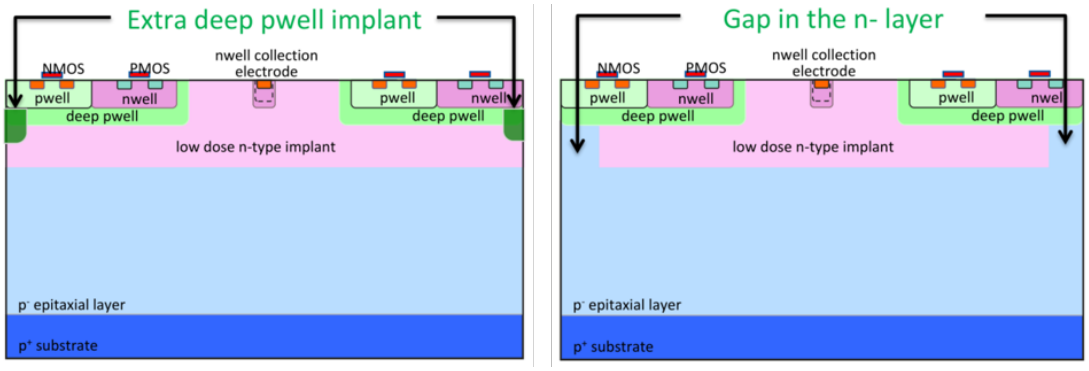

Figure 4.13: Cross section of a pixel with an additional extra-deep p-well near the pixel edges (left) and with a gap on the n-type implant (right).

the lateral component of the electric field improving the charge collection efficiency at the pixel edges after irradiation [20]. Simulations also show that the charge collection time is reduced by a factor of two as compared to the original modified process and that the total charge collected is three times higher at the edges of the pixel. However, the potential barrier between the p-wells and the p-substrate is reduced by effectively reducing the width of the n-type implant. This leads into punchthrough happening at lower voltages. From simulations, punchthrough happens at a substrate voltage of -8 V for the gap on the n-type implant solution and -10 V for the deep p-well solution at a p-well voltage of -6 V. Modifications of the pixel layout were motivated by charge collection efficiency results [65].

The front-end architecture was also modified to make it more resistant to ionising radiation. The most important change was an increase of the transistor M3 area (see figure 4.9), which dimensions (W/L) were increased from 1/0.18 μm to 1.22/0.38 μm. By increasing the size of this transistor, it is intended to reduce the RTS noise observed in some pixels of MALTA, which are dramatically increased in number after TID irradiation (see section 7.3). As seen in section 4.3.1, transistor M3 is the highest source of 1/f noise of the front-end circuit. Aside from reducing the RTS noise, doubling the area of transistor M3 is expected to reduce its output conductance leading to an increase of the pre-amplifier gain, simulated to be around 30 % higher. Penalty on the transconductance of M3 and the capacitance of the OUT_A node are supposed to be minimal. The second modification of the MALTA front-end is an increase of the area of the M4 transistor, in which the dimensions were modified from 0.44/0.18 μm to 0.88/0.18 μm. This was done to better control the clipping threshold.

The improvements mentioned above were implemented on a small-scale demonstrator named mini - MALTA. The mini-MALTA chip layout with the different components labelled is shown in figure 4.14 (left). A schematic of the pixel matrix is shown in picture 4.14 (right). It consists of 16x64 pixels and only contains the "medium" deep p-well pixel layout variation of MALTA. The matrix is subdivided in 6 sectors containing different modifications of the standard pixel layout and of the front-end architecture. Horizontally, the first half of the columns contain pixels

with the standard MALTA front-end circuit while the second half implements the new front-end with enlarged transistors. Vertically, the chip is divided in four sectors: two sectors featuring the two different pixel layout modifications and with a diode reset, one with the standard MALTA pixel layout and diode reset and one with the standard MALTA layout and PMOS reset. Hence,

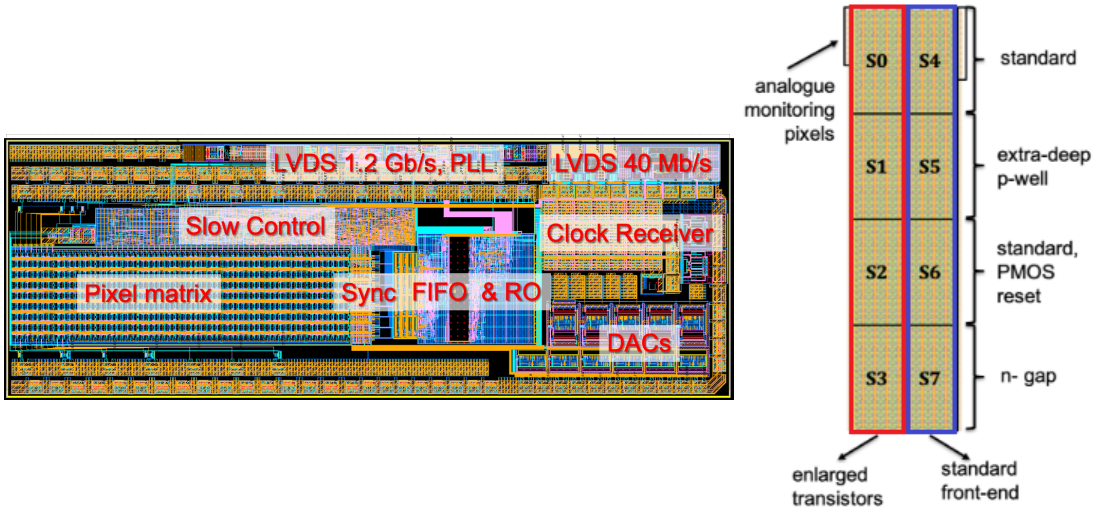

Figure 4.14: Mini-MALTA chip layout (left) and schematic of the mini-MALTA pixel matrix layout (right). Figure reproduced from [66] and modified.

sectors 4 and 6 contain pixels with the standard MALTA layout and front-end allowing direct comparison of these pixels with the improved ones on the same chip.

Readout architecture

The readout scheme of the mini-MALTA chip on the pixel matrix is the same as the one installed in MALTA. Each double column is divided in groups of 2×8 pixels alternating the "blue" and "red" readout lines. Again, two 22-bit words per double column are transmitted down the bus to the chip periphery. At the end of the column, an additional synchronisation block was included to synchronise the asynchronous hit information coming from the pixel matrix. This extra block simplifies the design of the digital periphery and of the off chip FPGA-based readout (see section 6.5) without increasing the power consumption of the matrix. Each colour group of each double column has its dedicated synchronisation block. A single block consists of a series of *Random-Access Memories* (RAMs) that take advantage of the alignment in time of the reference signal to enable the writing of the 21-bit pixel address. Time information is also added by storing the values of two different counters: a 40 MHz 3-bit BCID clock and a 640 MHz 4-bit clock for finer resolution. Storage of up to four consecutive data words is allowed by four rows of the RAM memory. Hence, a synchronisation block is composed of four 28-bit depth RAM cells and the corresponding addressing logic to route correctly each double column group to its corresponding

synchronisation block. Moreover, supplementary buffers and delay gates are included to allow misalignment between the reference signal and the address line of up to 400 ps.

Further encoding is applied after the synchronisation memories by the *priority encoder* logic. Synchronisation memories send a signal to the encoder logic to acknowledge that a hit has been stored and the priority encoder sends a signal back to read the memory. The 28-bit word containing information of the hit pixel address is then transferred to a 64-word depth FIFO, adding 4 bits for double-column identification and an extra BCID counter for a total word length of 48 bits. The hit information stored on the FIFOs is read out at a rate of 40 MHz. A further 8b/10b encoding is applied (8 bit word mapped to a 10 bit word) converting the 48-bit data word into the final 60-bit data word. This word is then serialised and transmitted off the chip via a data transmission unit (DTU) previously used on the ALPIDE chip. The output data is transferred in a single LVDS line at a rate of 1.2 Gb/s with double data rate (DDR). An independent readout system bypassing the 8b/10b encoder was also included as a backup feature. In this mode, the 48-bit word stored on the FIFO is serialised and fed at a rate of 40 MHz into the LAPA LVDS driver that was used in the MALTA chip. Although this mode is slower than the ALPIDE LVDS driver, it should support low particle rates (< 0.83 MHz).

Slow Control block

The Slow Control (SC) logic has been fully redesigned to substitute the partially non-working Slow Control of MALTA. This block establishes the communication protocol for interaction between the FPGA-based readout and the chip. It includes masking and pulsing capabilities, readout block selection, configuration of the DAC currents and voltages, configuration of the Phase-Locked Loop (PLL) and configuration of the receiver and transmitter of the LVDS and CMOS signals. The SC also includes a feedback functionality by receiving the number of *Single Event Upsets* (SEU) during the operation time of the chip.

The SC block consists of 30 registers which are written and read using a shift register. Each register is dedicated to the tuning of one of the chip parameters and has a default value that is loaded into the chip after power-up. Refer to table 4.3 for a given register functionality.

Register	Size	R/W	Description	Recommended
VCASN	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x07F
VCASP	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x000

VRESET_D	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x073
VRESET_P	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x04A
VPULSE_H	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x0FF
VPULSE_L	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x000
VCLIP	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x0FF
ICASN	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x014
IBIAS	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x064
ITHR	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x014
IDB	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x064
IRESET	10	R/W	9: Enable override 8: Enable monitoring [0:7]: Value one hot encoded	0x0FF
MASK_DIAG	16	R/W	[0:15]: Diagonal number	0xFFFF
MASK_COL	16	R/W	[15:0] Column number	0xFFFF
MASK_ROW	64	R/W	[63:0] Row number	0xFFFFFFFFFFFFFFFF
PULSE_COL	16	R/W	[15:0] Column number	0x0002
PULSE_ROW	64	R/W	[63:0] Row number	0x0000000000000003
PIX_MON_IDAC	8	R/W	[7:0] DAC settings of the monitoring pixels	0x5F

SEU_COUNT	16	R	Number of single event upsets	N/A

Table 4.3: List of the most commonly employed Slow Control registers used to configure the mini-MALTA chip.

The first 12 registers are used to set the value of the different DAC voltages and currents with the exception of VCASP, which is left floating in this mini-MALTA implementation. All of them include one bit for overriding and one bit for monitoring. When the overriding bit is set to 1 the currents and voltages set by the DAC operation can only be set through an external power supply. Moreover, when the monitoring bit is set to 1, the value of the voltage or current can be measured on the corresponding probing pads. The rest of the bits are used to set the value of the voltage/current. The next 5 registers are used for pixel masking and pulsing. For single pixel masking or pulsing only the column and row number is required for unique identification. However, in order to unequivocally mask or pulse multiple pixels, the diagonal coordinate has to be specified as well.

The mini-MALTA SC block has several pads to allow communication with the external readout (see section 6.5). These are described in table 4.4. All SC signals are synchronous to

PAD name	I/O	Description
CLK	Input	40 MHz clock signal
rstB	Input	If '0', reset registers to default value
ENABLE	Input	If '1', enable shift register to write data
START	Input	If '1', the register values are sent to SOUT pad
DATA	Input	Register values to write in a bit-stream of 480 bits
SOUT	Output	Register values to read as a serial bit-stream
SOUT_ACK	Output	Acknowledgement signal to confirm that the SOUT data was valid

Table 4.4: List of the mini-MALTA Slow Control pads. To enable SC communication, these pads are connected to the mini-MALTA FPGA-based readout.

the 40 MHz clock provided to the CLK pad. The rstB is an active low signal employed to reset the registers to their default value. Hence, it must be kept to '1' when operating the chip. During the writing and reading operations, the value of all registers must be input/output as a block.

In order to write a configuration, the ENABLE signal must be set to '1' until the operation is completed. In the same clock cycle, the bit-stream sent to the DATA pad is written to the registers (with the exception of SEU_COUNT) in the order specified in table 4.3. The mini-MALTA configuration can be read by setting the START signal to '1'. When this signal is received, the register values are sent to the SOUT pad. The SOUT_ACK signal switches to '1' with the first valid bit and stays constant until the reading operating is completed.

The mini-MALTA chip is controlled with the SC block described in this section. The user can access the SC configuration from the DAQ PC with the interface provided by the mini-MALTA readout described in section 6.5.

Monitoring pixels

Sixteen monitoring pixels have been included in the mini-MALTA pixel matrix that have access to either the input node of the front-end (i.e. the collection electrode) or the OUT_A node, after the pre-amplification phase. Half of the monitoring pixels are situated at the right side of the matrix and contain the standard MALTA front-end while the other half is situated at the left side and contain the version of the front-end with enlarged transistors. Moreover, different pixel layouts are included to study the differences between the standard modified process and its versions with the extra-deep p-well or the gap in the n-layer. Since these pixels lack the digital readout circuit, this empty area was filled with either a highly doped p^+ active area or with n-well fillings on the p^+ material simulating the n-wells of the digital circuitry. Pairs of pixels are multiplexed and their signals are sent to eight monitoring pads. The selection mechanism used to access a particular pixel within the multiplex is provided by the Slow Control block. On its default mode, only odd pixels are monitored. The monitored signal is buffered before reaching the monitoring pads. Two types of source follower buffers are used: a simpler two stage buffer (PMOS + NMOS source follower) and a more complex four-stage buffer (PMOS + NMOS + PMOS + NMOS source follower). Both buffers differ by their minimum reachable rise time (~ 10 ns for the two-stage and ~ 1 ns for the four-stage). Hence, the four-stage buffer is used to capture the fast rise of the signals at the input node while the two-stage buffer is used for the less demanding signal at the output of the pre-amplifier. Table 4.5 summarises the characteristics of each monitoring pixel and the pad label where it can be probed.

Mini-MALTA carrier board

A dedicated PCB was designed to host the mini-MALTA chip. A 3D simulation of the mini-MALTA carrier board is shown in figure 4.15. It is equipped with the necessary connections to provide power to the different chip domains, establish communication with the Slow Control configuring block, readout the data output of the chip and enable access to the testing capabilities of the sensor.

Power supplies can be connected to five Molex Mini-Fit Jr headers to power up the different power domains of the mini-MALTA chip. Pads are labelled DVDD (for the digital logic), AVDD (for the analog front-end), PVDD (clocks), PWELL (to power up the p-well implant) and SUB (to power up the pixel substrate).

Communication with the FPGA-based readout is allowed by an FMC connector situated at the back-side of the board. The different pins of the connector are used to send the data words

Number	Node	Process	Filling	Buffer	Pad
0	Output	Modified	Digital n-well	Two-stage	MON_OUT_LEFT[0]
1	Output	Modified	Digital n-well	Two-stage	
2	Input	Modified	Active area	Four-stage	MON_IN_LEFT[0]
3	Input	Extra deep p-well	Active area	Four-stage	
4	Output	Modified	Active area	Two-stage	MON_OUT_LEFT[1]
5	Output	Modified	Active area	Two-stage	
6	Input	Modified	Active area	Four-stage	MON_IN_LEFT[1]
7	Input	N-gap	Active area	Four-stage	
8	Output	Modified	Digital n-well	Two-stage	MON_OUT_RIGHT[0]
9	Output	Modified	Digital n-well	Two-stage	
10	Input	Modified	Digital n-well	Four-stage	MON_IN_RIGHT[0]
11	Input	Modified	Digital n-well	Four-stage	
12	Output	Modified	Active area	Two-stage	MON_OUT_RIGHT[1]
13	Output	Modified	Active area	Two-stage	
14	Input	Modified	Active area	Four-stage	MON_IN_RIGHT[1]
15	Input	Modified	Active area	Four-stage	

Table 4.5: List of the monitoring pixels available in the mini-MALTA chip and their characteristics. Pixels are connected to either the collection electrode (Input) or a node after the pre-amplification phase (Output). The different pixel layouts included are the standard TJ modified process (Modified), the variation with an extra deep p-well (Extra deep p-well) and the variation with a gap in the n-layer (N-gap). The lack of digital circuitry is filled with either a p^+ active area (Active area) or with n-well fillings (Digital n-well). Finally, two types of source followers are used: one using 2 transistors (Two-stage) and one using 4 transistors (Four-stage). More information is given in the text.

to the FPGA, send and receive the Slow Control commands and to send test pulses. A more detailed description of the mini-MALTA readout is given in section 6.5.

Several probing pads are included for testing and debugging purposes. The analogue output of the monitoring pixels is connected to 8 different pin connectors situated at the sides of the board. Two pads to monitor the VCASN voltage level through the two versions of the buffer are also included. A 23×2 pin connector allows for the monitoring and overriding of the DAC currents and voltages. There are also probing pads for the 40 and 640 MHz clocks provided by the FPGA, the data word output, the Slow Control output word and the Slow Control acknowledge signal.

The voltage levels on the FPGA are matched to the standard 1.8 V CMOS signals used on the chip side using a MAX3013 voltage level translator.

The mini-MALTA chip is wirebonded to the mini-MALTA test-board allowing to access all the chip functionalities. The chip is configured and operated through the FPGA-based readout presented in section 6.5. Some of the functionalities of the board presented in this section were used to perform the measurements presented in section 7.4.

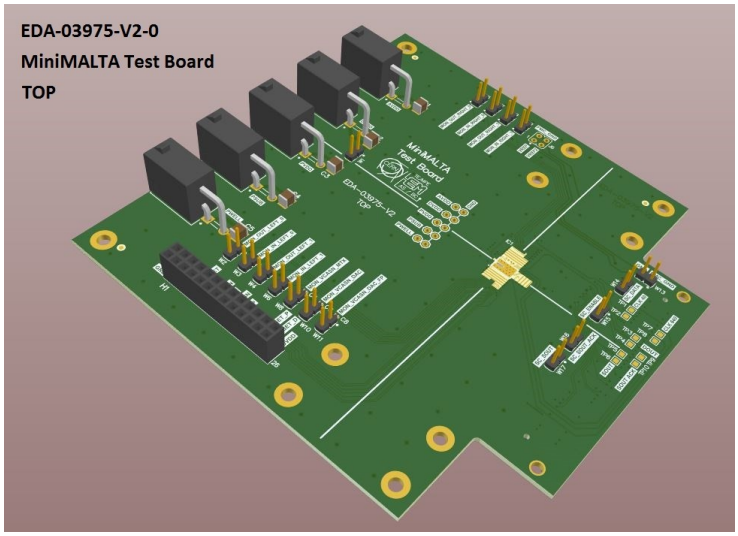

Figure 4.15: 3D schematic of the mini-MALTA carrier board.

4.4 Summary

CMOS sensors fabricated by the TowerJazz foundry have been used at the ALICE ITS for tracking purposes. In the standard TowerJazz 180 nm process, sensors have a small n-type implant acting as the collection electrode to minimise the pixel capacitance, a high resistivity epitaxial layer to facilitate the depletion of the sensor bulk and deep p-wells shielding the PMOS transistors of the front-end circuit. During the ALICE R&D process, an investigator chip has been designed to determine the parameters that affect the pixel analog performance. The chip, named TowerJazz Investigator1, has 134 mini-matrices of 10×10 pixels featuring different pixel sizes, collection electrode sizes, distances between the collection electrode and the transistors of the readout circuit, deep p-well sizes, reset circuits and source follower circuits. The front-end circuit of the chip only includes a pixel reset, a source follower and a mini-matrix selection circuit, giving access to the raw analog signal of the sensor for the selected mini-matrix. Due to the small collection electrode, the depletion region of the TJ standard process does not cover the whole sensitive area of the detector. Charge outside this depletion region is still collected by diffusion but is more susceptible to radiation damage. In the ALICE ITS detector the radiation hardness requirements are moderate and are satisfied by this design. However, this is not the case for the ATLAS ITk detector.

A process modification has been implemented in the TowerJazz 180 nm process with the purpose of achieving a more uniform depletion region, increasing this way the radiation hardness of the chip. In the TowerJazz 180 nm modified process, a low dose n-type implant is placed between the electrode and the p-type epitaxial layer. With this new design, the depletion region extends uniformly from the n-type implant without significantly increasing the capacitance of the device. A full demonstrator chip, named MALTA, has been designed in this modified technology. MALTA contains 512×512 pixels with an area of 36.4 μm^2. The chip has eight

different regions with variations of the electrode size, the distance between the electrode and the front-end transistors, the size of the deep p-well implants that shield the PMOS transistors and the reset circuit. The analog front-end circuit of MALTA can be summarised by a charge-sensitive shaper-amplifier and a discriminator and has been optimised to process signals within 20 ns, being compatible with the bunch-crossing identification requirements of the HL-LHC. Each pixel in MALTA is equipped with a masking circuit and an analog test pulse. Moreover, several pixels have access to a node after the amplification phase of the front-end circuit for monitoring purposes. The digital readout architecture of MALTA features a novel design in which the signal information is transmitted asynchronously to the chip periphery. With this arrangement, a synchronisation clock does not need to be provided to the pixel matrix reducing the power consumption of the chip. Signals of each pair of columns are propagated to the chip periphery in different readout lines that are merged in a common bus with the proper identification address and time stamps. The user can access the chip functionalities through the SC block. Each configurable feature of MALTA has an associated register that can be modified with a finite state machine. The registers are modified by sending a 16-bit command to the input pad of the SC preceded by an additional low bit to indicate the start of the transmission. The SC issues a reply word to its output pad either providing the required information or confirming the command.

Although the TJ modified process has increased radiation tolerance as compared to the TJ standard process, the charge collection efficiency at the edges of the pixel drops significantly after irradiation. Moreover, results have shown that the RTS noise associated to a transistor of the front-end circuit significantly increases after TID irradiation. Hence, the MALTA chip has been redesigned into a smaller demonstrator named mini-MALTA. The mini-MALTA chip has 16×64 pixels and eight different sectors. It includes sectors with an extra-deep p-well implant at the edges of the pixel and sectors with a gap in the low dose n-type implant. Both solutions are expected to improve the charge collection efficiency at the pixel edges. Moreover, half of the chip implements an improved version of the front-end circuit with two enlarged transistors while the other half maintains the MALTA front-end allowing a direct comparison. The digital readout architecture of the chip is similar to that of the MALTA chip, but an extra synchronisation block has been added at the end of column and the signals are synchronously propagated off the chip. The SC block has also been redesigned to substitute the partially non-working SC of MALTA. In mini-MALTA all the registers are written and read as a block. In order to modify the registers, a signal has to be sent to a specific pad. Then, a bitsream containing values for each register has to be provided in the input pad of the SC. Once the bitsream is processed the chip issues a response word that is acknowledged in a different pad.

Chapter 5

Fluorescence studies on the TowerJazz Investigator1 Chip

The TowerJazz Investigator1 chip, described in section 4.2, has several mini-matrices with different pixel geometries allowing to study how specific characteristics affect the analog performance of the chip. This has been used to determine the relationship between spacing, electrode size and P-well size with the analog output signal (section 5.4). The signals are induced using a technique called *X-ray fluorescence*, with the setup described in section 5.1. Signals are recorded by the oscilloscope and saved in a text file to be later analysed with a Python-based script, as shown in section 5.2. The linearity of the pixel response to different input charge is also studied and presented in section 5.3.

5.1 X-ray fluorescence Setup

The setup employed to characterise the charge collection performance on several pixels of the TowerJazz Investigator1 chip is presented in figure 5.1. For clarity, a block diagram of this setup is shown in figure 5.2. Measurements were carried out inside a Pb box to protect the environment from the X-rays employed.

X-rays are generated by a 50 kV Amptek mini X-ray source with a silver target. Its specifications can be found in [67]. An external fan is used to cool down the X-ray gun. The divergent X-ray beam generated is directed to thin targets of different materials. These X-ray photons can be absorbed by electrons of the K-shell of the target material, causing them to leave the atom. Electrons from higher energetic shells can decay to fill this vacancy, emitting the energy difference in the form of an X-ray photon. This process, known as *X-ray fluorescence*, is used to generate spectral lines of monochromatic X-ray photons. Different materials are employed to obtain different energy X-rays. Refer to table 5.1 for a list of the available materials and their most important X-ray energy. Targets are placed forming a 45° angle between the primary X-ray beam line and the Investigator1 chip to maximise the number of fluorescence photons hitting the

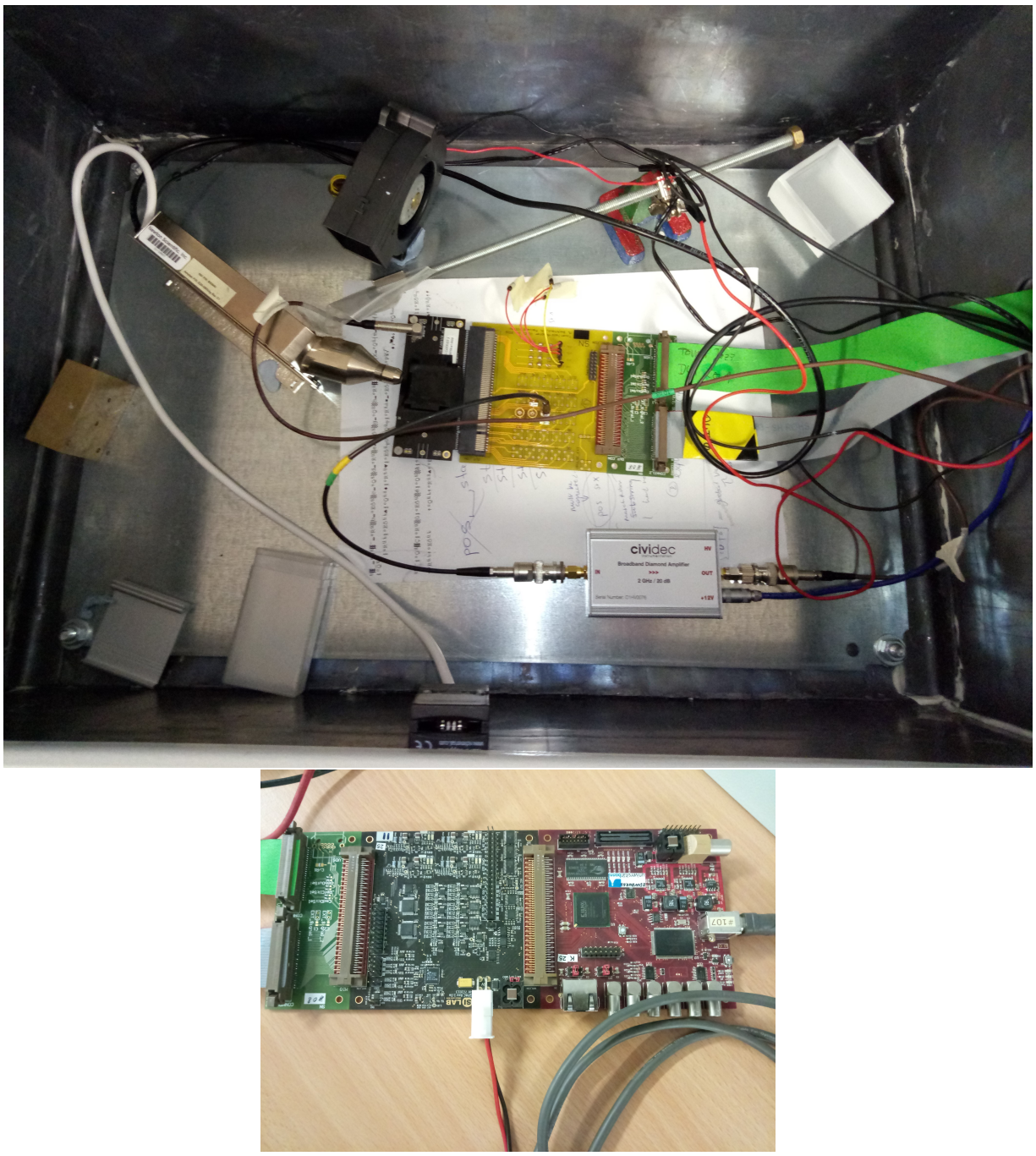

Figure 5.1: Picture of the X-ray fluorescence setup used to characterise the TowerJazz Investigator1 chip. Top figure shows the components inside the Pb box consisting of the mini X-ray source and the fluorescence target, the cividec amplifier, the Investigator1 chip and the TJ adapter card. Bottom figure shows the GPAC and MIO boards, which are placed outside the Pb box and connected to the TJ adapter card.

DUT.

Fluorescence X-ray photons generate charge carriers on pixels of the Investigator1 chip. Signals from the 64 active pixels of the selected mini-matrix are transmitted to the output buffer. The chip is then read out and controlled through three PCB connected in series: a Multi-IO

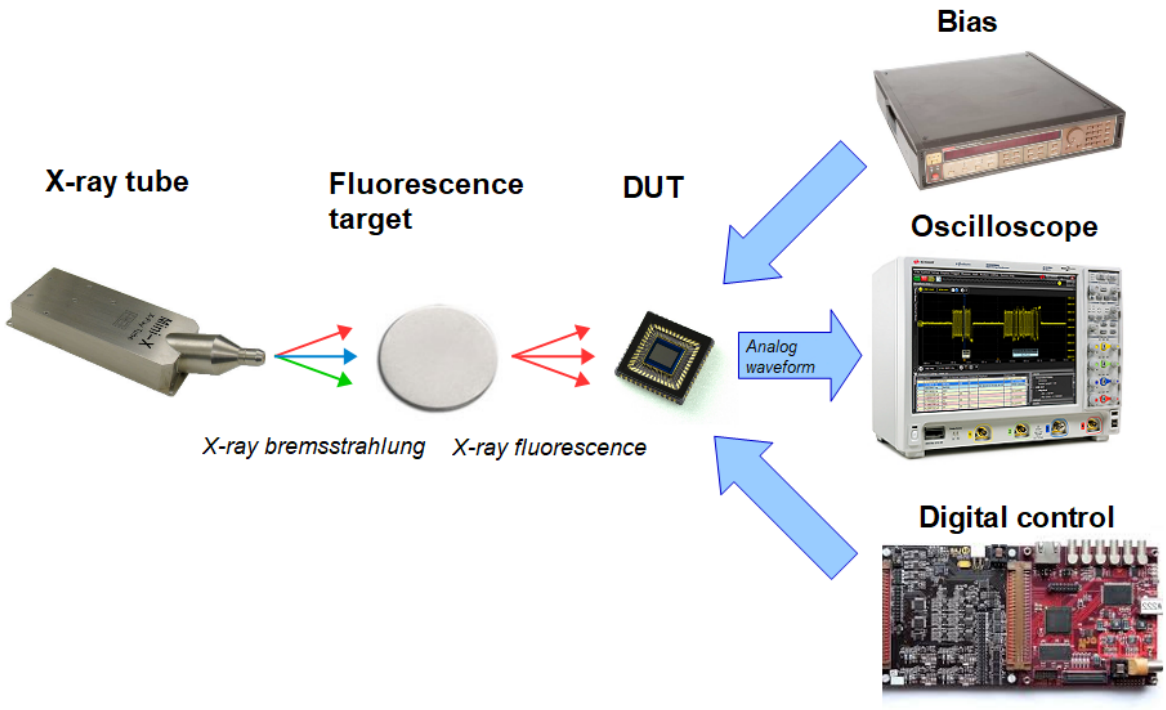

Figure 5.2: Block diagram of the setup employed to characterise the TJ Investigator1 chip. The chip is biased with Keithley 237 PSU and read out using a MSO9404A oscilloscope from Agilent technologies.

board (MIO), a General Purpose Analog Card (GPAC) and the TowerJazz Adapter card. Those are connected in series to the Investigator1 carrier board. The 64 output buffers from the selected mini-matrix are routed to 64 pins of a PCIe connector that connects the Investigator1 carrier board to the TJ Adapter board. 25 out of the 64 output signals are accessible through LEMO connectors on the TJ adapter card. Refer to figure 5.3 for the actual correspondence between pixel position and LEMO connector. LEMO cables are connected to a single-channel amplifier from the cividec Instrumentation GmbH C1HV series [68]. This is required to increase the S/N ratio, as no amplification is done in the Investigator1 front-end circuit. The amplifier is connected through a 50 Ω resistance to the readout oscilloscope with AC coupling. In the

Target	Energy (keV)
Ti	4.5
Fe	6.4
Cu	8.0
Nb	16.6
Sn	25.3
Ba	32.2

Table 5.1: List of X-ray targets and their respective X-ray fluorescence decay energies.

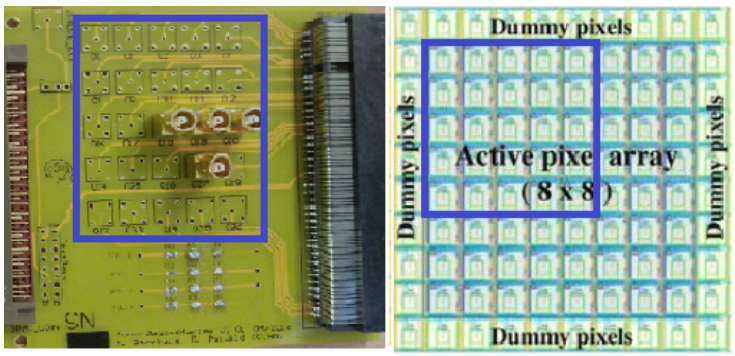

Figure 5.3: Picture of the TowerJazz Adaptor card and the correspondence between LEMO connectors and the pixel position within a mini-matrix. Only one pixel was connected at a time.

oscilloscope, a custom Labview-based program is employed to store and pre-analyse the data.

The chip is biased and controlled with the GPAC and MIO boards (see figure 5.1). The GPAC card provides all the bias voltages and signals needed to operate the chip (IFOL1 - IFOL4) with the exception of the pixel cell supply. This is supplied by a keithley 237 PSU set to -5 V. The chip is externally controlled with the MIO board. The board is equipped with an FPGA that can be operated using the pyBAR framework [69] designed in Bonn. By using this framework, specific mini-matrices can be connected to the readout chain by sending a 9-bit command. The reset signal used to reset the active pixels is also configured via pyBAR. This reset signal is routed to a channel of an MSO9404A oscilloscope from Agilent technologies for triggering purposes.

5.2 Signal description and analysis

Within the context of this thesis, only pixels with active reset were investigated. Thus, the signal described in this section only applies to pixels with active reset. Incoming particles generate a current on the device as described in section 3.8. This current charges the electrode capacitor, which keeps the induced voltage until the reset signal discharges it. Hence, the signal (or waveform) will have a sharp rise corresponding to a particle hit and then proceed almost flat until the reset signal clears it. An example of the waveform (green points) and the reset signal (purple) are shown in figure 5.4. It is noted that the reset signal generates a wiggling perturbation on the pixel response. In order to prevent false triggers, a double triggering strategy is followed. The first trigger is applied on the decaying edge of the reset signal, which occurs after the reset induced effect in the main signal. Then, a threshold is set in the waveform at three times the noise level. This trigger strategy is implemented at the oscilloscope level.

Each triggered waveform is stored in a text file and analysed using Python scripts. An example of an analysed waveform is shown in figure 5.5. The parameters that can be directly extracted from this waveform are listed below:

- Baseline: The baseline corresponds to the mean value of the points before the particle hit.

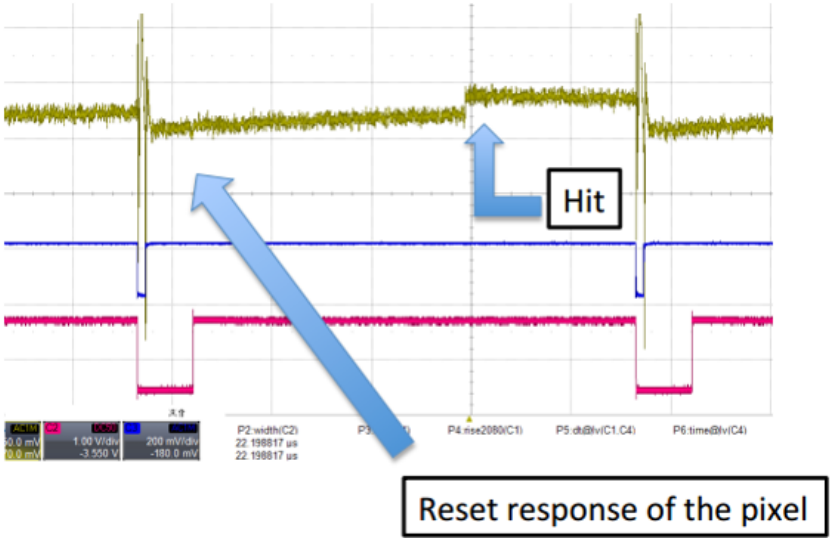

Figure 5.4: Example of a typical signal after a particle hit (green line) and the reset signal (purple line) recorded by an oscilloscope.

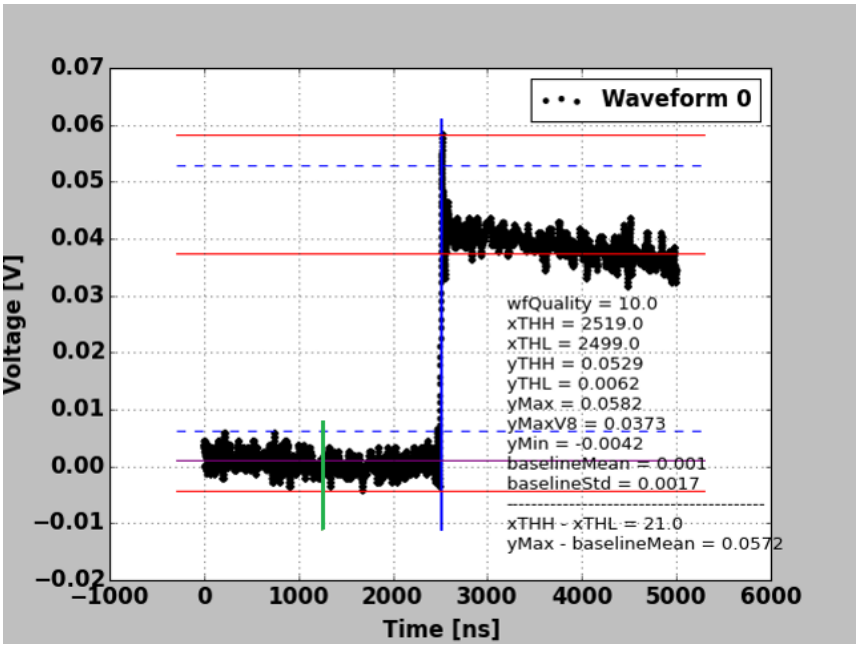

Figure 5.5: Waveform recorded and analysed using Python scripts.

The mean is calculated up to the green line, and the baseline level is marked by the purple line in figure 5.5.

- Signal amplitude: The charge collected of the event is extracted (in Volts) by measuring the waveform level after the particle hit. In practice this parameter was calculated by averaging 100 points in the waveform's plateau (middle red line in figure 5.5) and subtracting the baseline level from this value.

- Rise time: Is the time that the waveform takes to change from 10% to 90% of its maximum

amplitude. The dashed blue lines in figure 5.5 mark the two voltage levels and the blue solid lines mark their projection to the x-axis (time).

- Noise: The noise is calculated from the RMS of the baseline sample (up to the green vertical line in figure 5.5).

In the measurements presented in sections 5.3 and 5.4, around 1000 signals are recorded. The parameters listed above are plotted in a histogram. The mean value of the obtained distribution is obtained by fitting a Gaussian function to the data. An example of these plots is shown in figure 5.6. Within the context of this thesis, the signal amplitude as a function of different parameters

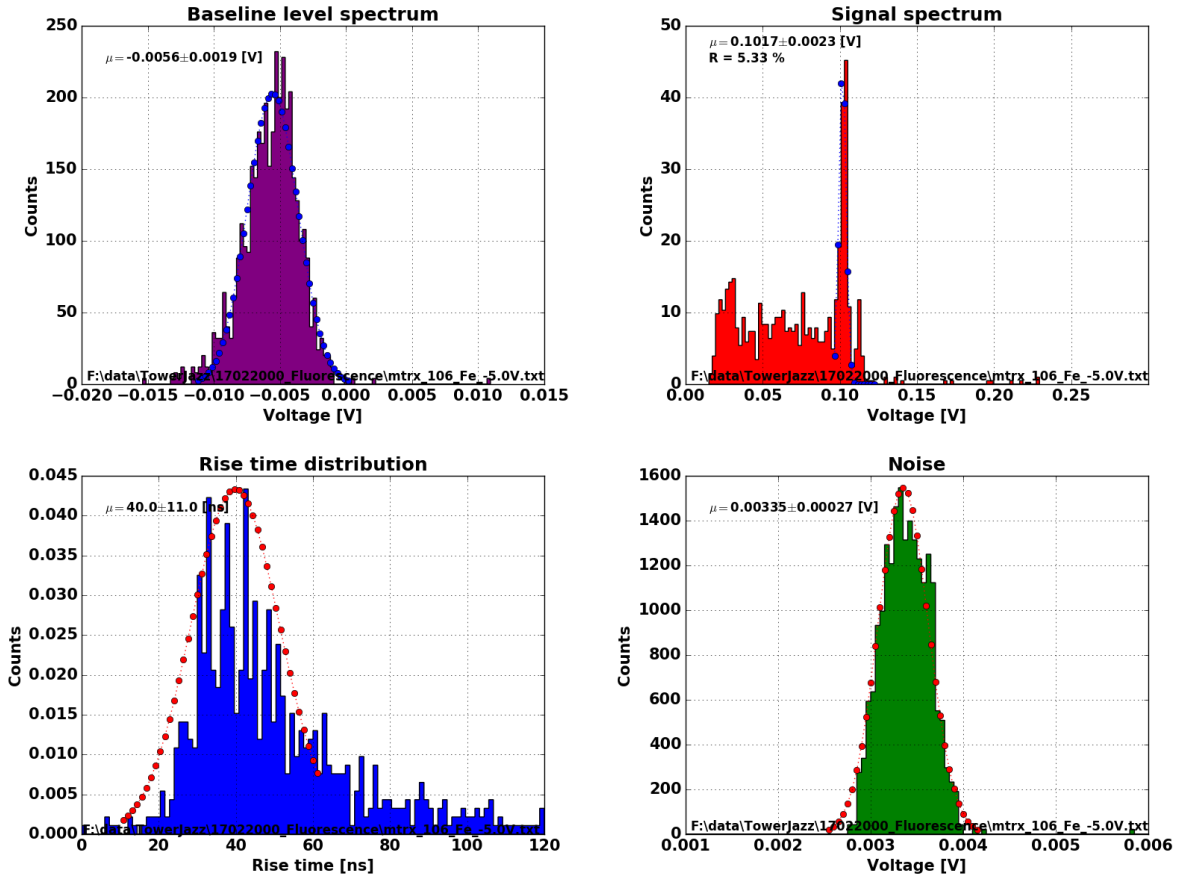

Figure 5.6: Histogram of the baseline (top left), charge collected (top right), rise time (bottom left) and noise (bottom right) distributions from ~ 1000 signals induced with Fluorescence X-rays using a Fe target on mini-matrix 106 of the Investigator1 chip.

is investigated. This distribution (top right plot in figure 5.6) is approximately homogeneous at low values, followed by a characteristic peak. At higher values fewer signals are recorded. The peak corresponds to the primary decay of the fluorescence X-ray photons (see section 5.1). The signals recorded below the peak are due to charge sharing with the neighbouring pixels. Finally, the values above the peak are due to an X-ray photon hitting the measured pixel plus charge sharing from a hit in the adjacent pixels, which is less likely to occur. The rest of the parameters

are not investigated. The rise time of the standard process is not optimal, as part of the charge is collected by diffusion. In the modified process, charge is collected by drift and the charge collection time fulfils the requirements of the ATLAS ITk [61]. Noise is expected to be affected by the pixel capacitance. However, no such dependence is observed in the measurements taken. A possible explanation is that noise is dominated by that of the readout chain.

The measurements presented in this chapter are intended to complement those presented in [62]. In that work, the sensor noise (without the readout chain) is estimated to be of around 0.24 mV. Since the measured noise is much higher (of around 3.5 mV) the hypothesis that the noise is dominated by that of the readout chain is corroborated. The results presented in the following sections are obtained by only measuring one pixel within the matrix. However, all results are similar to those presented in [62], increasing the their reliability.

5.3 Linearity of the pixel response

A relation between the pixel response in volts and the absolute charge induced in the device can be determined by measuring the signal obtained with different X-ray targets. The X-ray targets produce monochromatic X-ray lines of known energy (E_t). Considering that 3.6 eV are needed to generate a single electron-hole pair in Silicon (see section 3.3), the number of electrons generated by a given target is:

$$e = \frac{E_t(eV)}{3.6\,eV} \quad \text{electrons} \tag{5.1}$$

The response of six mini-matrices with different characteristics to four X-ray targets (Fe, Cu, Nb and Sn) is shown in figure 5.7. The bias voltage is set to -5 V for all measurements. For visualisation, an exponential function is fit to the data to account for saturation of the pixel response at high deposited energies. Pixels from mini-matrices 70, 79 and 106 present saturation of the pixel response at around 0.24 V for deposited charges higher than 5000 e. Pixels 114, 126 and 129, however, present a much lower gain and a linear behaviour for the whole investigated range. Signal saturation is caused by the PMOS transistors of the source follower circuit [70].

The expected deposited charge by a MIP in a 25 μm thick detector is approximately 1600 e. Mini-matrices 70 and 79 start showing deviations from the linear trend at around 2000 e while mini-matrix 106 shows deviations at around 3500 e. Hence, MIPs would produce a signal at the edge of the linear region for mini-matrices 70 and 79.

The matrices with high signal gain have small electrode sizes and spacing. MM070 has a collsize of $1 \times 1\ \mu m^2$ and a spacing of 3 μm, MM079 has a collsize of $3 \times 3\ \mu m^2$ and a spacing of 1 μm and MM106 has a collsize of $3 \times 3\ \mu m^2$ and a spacing of 3 μm. This is the opposite for matrices with small gain. MM114, MM126 and MM129 have a collsize of $5 \times 5\ \mu m^2$, $10 \times 10\ \mu m^2$ and $3 \times 3\ \mu m^2$ respectively and an opening of 30 μm. The geometry of the pixel is known to affect its capacitance which, in turn, affects the pixel gain. The effect of the different parameters to the pixel response is studied in section 5.4.

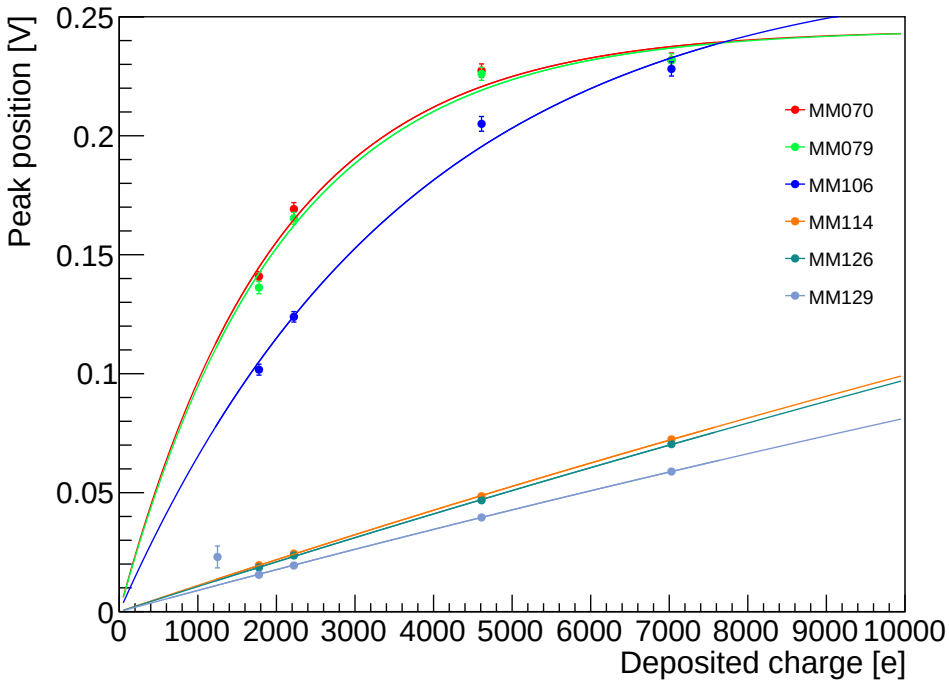

Figure 5.7: Charge collected (in mV) by pixels in the mini-matrices 70, 79, 106, 114, 126 and 129 as a function of the deposited charge (in electrons).

5.4 Charge collection studies

The study of how the different pixel parameters affect its analog signal output is presented in this section. The parameters studied are the electrode size, the spacing, the P-well coverage and the transistor type. This is done by selecting groups of mini-matrices with only one differing characteristic between them. Charge is generated from X-ray fluorescence using the Fe target (6.4 keV). The sensor is biased to -5 V in all measurements. As shown in section 5.3, this target produces a signal close to the linear region for all the measured pixels.

Electrode size dependence

The dependence between the electrode area and the output signal is described by equations 3.19 and 3.32. For a constant charge injected, the signal measured by the electrode is proportional to the depletion width (W) and the electrode area (A_{el}) as:

$$V \propto \frac{W}{A_{el}} \tag{5.2}$$

Thus, the measured signal is expected to increase for smaller electrode areas.

In order to study this relationship, mini-matrices 70, 75, 80, 85 and 90 are selected. Their common properties are a pixel area of 28x28 μm^2, a spacing of 3 μm, a maximum deep P-Well,

an active reset type and a standard outside transistor. Their collsize parameter varies from 1 to 5 μm. For convenience, the peak value of the X-ray fluorescence charge distribution is plotted against $1/A_{el}$. This is shown in figure 5.8.

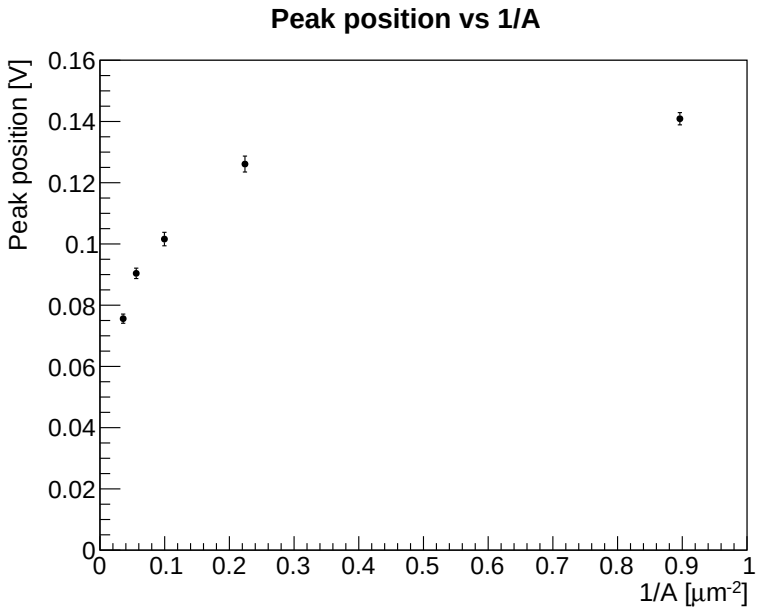

Figure 5.8: Charge collected as a function of $1/A_{el}$. Each electrode area corresponds to a measurement in a different pixel. Their common parameters are a pixel size of 28x28 μm^2, a spacing of 3 μm, a maximum deep P-Well, an active reset and a standard outside transistor. The sensor is biased at -5 V.

Note that, although the value of the signal measured increases with $1/A_{el}$ as expected, it does not follow a linear relationship with this parameter. Note that the source follower circuit saturates at around 0.22 V (see figure 5.7). Since the measured signals are below that value, this indicates that the depletion width decreases for smaller electrode sizes.

Spacing dependence

The spacing parameter has an effect on the capacitance associated between the electrode and the p-well implant that shields the in-pixel electronics. Thus, studying the impact of this parameter on the pixel gain is also relevant.

Mini-matrices 78 to 82 are selected to carry out this study. Their pixel characteristics are 28x28 μm^2 pixel area, 3x3 μm electrode size, maximum deep P-Well, active reset and standard outside transistor. The spacing varies from 1 to 5 μm. The pixel response as a function of the spacing is shown in figure 5.9. The pixel gain is found to follow a decreasing linear relationship

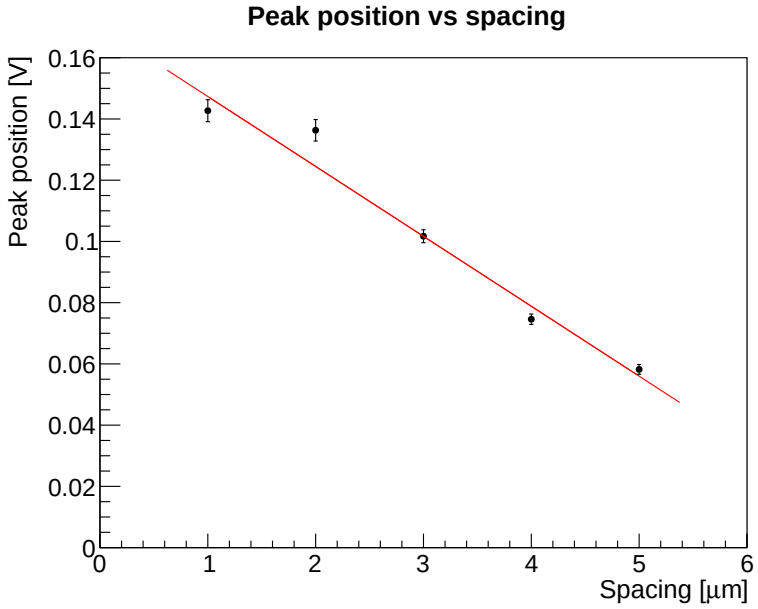

Figure 5.9: Charge collected (in mV) as a function of the spacing distance. Each point in this plot corresponds to a measurement in different pixels (from 78 to 82). All them have a pixel size of 28x28 μm^2, a 3x3 μm electrode size, a maximum deep P-Well, an active reset and a standard outside transistor. The sensor is biased at -5 V.

with the spacing parameter:

$$V = (0.1701 \pm 0.0031) - (0.02283 \pm 0.00079) \cdot spacing \tag{5.3}$$

This indicates that the spacing parameter also has an impact in the pixel capacitance, increasing for higher spacing values.

Dependence on P-Well size

The horizontal dimension of the deep P-well has an effect on the pixel capacitance, and hence gain, and on the charge sharing. Three P-well sizes are available: minimum, medium and maximum. The respective implants are shown in figure 5.10.

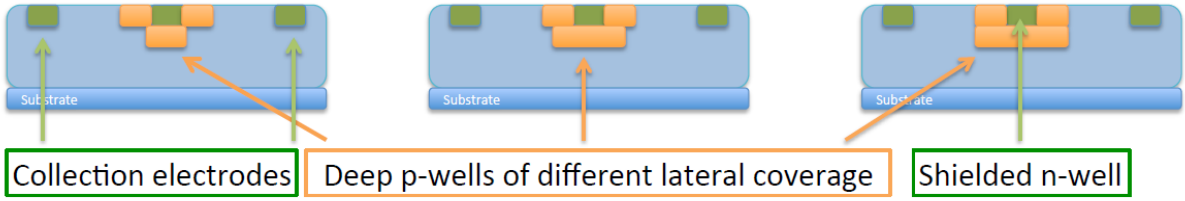

Figure 5.10: Different implementations of the P-well implant used to shield the n-wells from the epitaxial layer. The minimum size (left) is approximately the size of the electrode, the maximum size (right) is the size of the opening parameter and the medium size (center) is between the two extremes.

The P-well size is studied using pixels 80 (maximum), 95 (minimum) and 98 (medium). All of them have a pixel size of 28x28 μm^2, electrode size of 3x3 μm, a spacing of 3 μm, an active reset and a standard outside transistor. The Fe fluorescence spectrum of the three pixels is shown in figure 5.11.

The difference in peak position between the minimum and medium P-well is negligible and contained within the uncertainties. However, the signal of the maximum size is increased by $\simeq 15\,\%$ again due to differences on the pixel capacitance. Moreover, stronger charge sharing can be inferred from the increased contribution of the low amplitude hits on the spectrum. Note that each result is obtained for only one pixel, not taking into account possible pixel to pixel variations. However, similar results have been found in [62] corroborating the results presented in this chapter.

5.5 Summary

The different pixel geometries present in the TowerJazz Investigator1 chip provide a framework to study the effect of each pixel parameter on the overall performance. The signal was generated using monochromatic X-rays from X-ray fluorescence. The linearity of the induced signal with the total charge injected was tested in pixels from six different mini-matrices. Three of the measured pixels have an electrode size lower than $3\times3\,\mu$m^2 and a spacing of less than 3 μm. The other three have large spacing or electrode size parameters. Pixels of the first group have a high gain, and show a saturation of the signal amplitude for deposited charges higher than 5000 e. The second group of pixels show a relatively low gain and a linear correlation between the two parameters for the whole investigated range of energies. Signal saturation implies a reduction

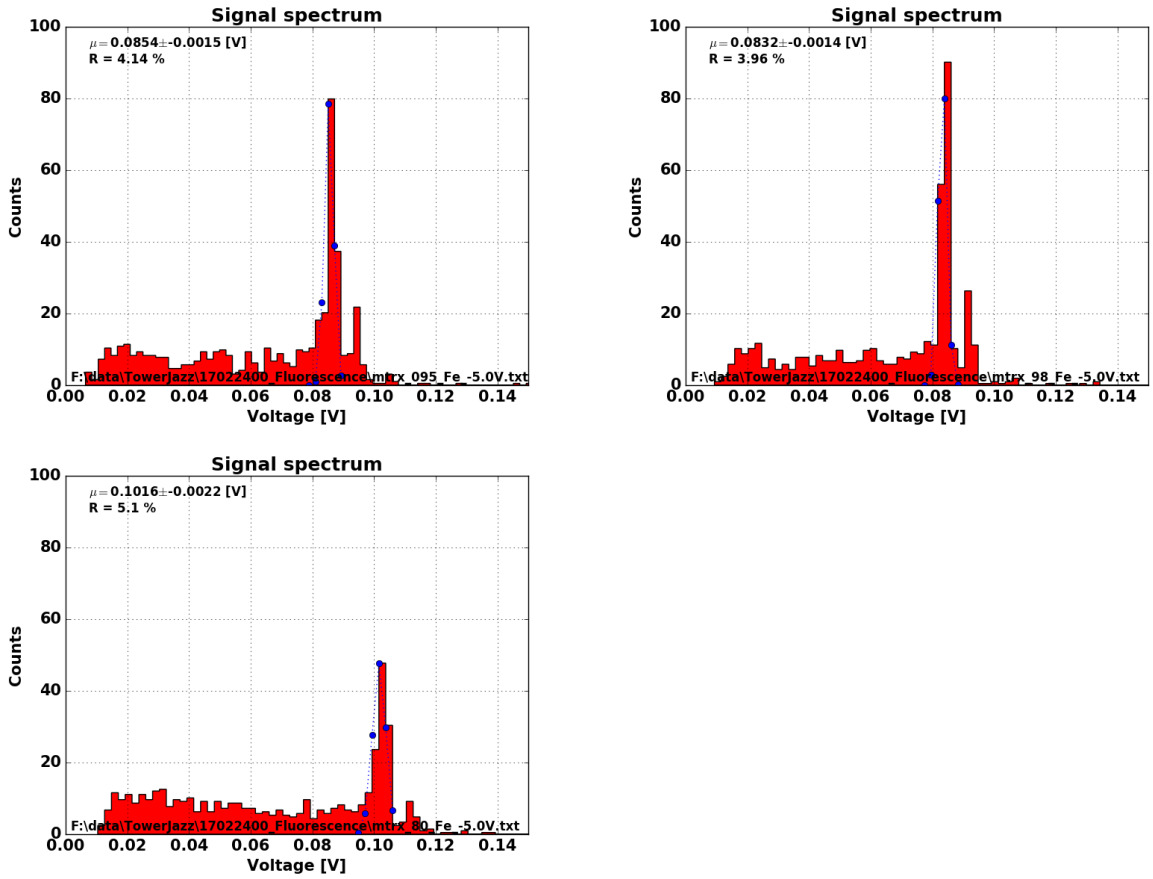

Figure 5.11: Spectrum out of 1000 waveforms of pixels 95 (top left), 98 (top right) and 80 (bottom). Pixel 95 has a minimum deep P-well, pixel 98 has a medium deep P-well and pixel 80 has a maximum deep P-well. All other properties are the same in all three.

of the energy resolution for high energetic particles. Hence, a linear correlation between sensor signal and charge deposition is preferred in the expected operating region of the detector.

The effects of the electrode area, the distance between the electrode and the surrounding electronics and the horizontal dimension of the P-well on the pixel gain were further investigated. All these parameters are expected to affect the overall capacitance of the pixel which, in turn, determines its gain. Signals were generated from fluorescence X-rays, using the less energetic target (Fe) in order to avoid saturation effects. Pixels with an electrode size ranging from $1\times1\mu m^2$ to $5\times5\mu m^2$ were measured. As expected, smaller electrode areas led to larger signals although the two parameters were not linearly correlated. This might be due to differences in depletion width. A decreasing linear correlation was found for the spacing parameter, which was measured from 1 μm to 5 μm. Thus, it is proved that smaller electrode sizes and spacing parameters lead to larger signals. Finally, the P-well horizontal distance was also found to have an effect on the pixel gain. Pixels with maximum P-well distance show a greater signal than pixels with medium and minimum P-well.

These measurements were used as input to the design of the MALTA chip described in

section 4.3.1. In MALTA, a compromise between a linear operating region and a high gain is sought. MALTA includes sectors with electrode sizes of $2\times2\,\mu m^2$ and $3\times3\,\mu m^2$, spacing parameters of 3.5 μm and 4 μm and medium and maximum P-well extensions. These characteristics are similar to those of MM 106, which shows good linearity up to a deposited charge of 3500 e.

Chapter 6

Readout development for TowerJazz Demonstrators

6.1 Introduction to Detector Readout

A particle detector is capable of converting a microscopic interaction into a sizeable signal that can be detected by the front-end electronics. However, these signals are not automatically analysed but stored in computers for later processing. Computers also hold the software interface used for user communication with the chip functionalities. However, commercial computers are not equipped with the hardware interface required to operate a particle sensor. The communication protocols vary from sensor to sensor, depending on the actual design and the implemented capabilities. Thus, tailored interfaces between the detector and the DAQ structure must be designed for each chip. This is known as the detector readout.

The readout circuits of the MALTA (section 4.3.1) and mini-MALTA (section 4.3.2) chips were developed within the context of this work. The MALTA chip has 37 differential output pads in which the asynchronous hit information is transmitted. The chip is operated and configured using the SC block. This block requires an external clock to be provided and receives/sends the SC commands through two dedicated pads. The writing and reading protocols of the SC configuration are activated with an additional active-low bit preceding the bit-stream. The MALTA readout is required to synchronise and process the 37-bit parallel data output of the chip and send it to the DAQ PC. It also needs to transmit and read back the SC configurations selected by the user. SC commands are input from the DAQ PC via Python-based scripts. These commands are arranged in the readout and sent to MALTA with the correct writing protocol. In order to read the SC configuration, the readout has to detect the active-low bit acknowledging the start of the transmission and upload the upcoming bit-stream back to the computer. The mini-MALTA chip has different protocols for data reading and SC transmission. Hence, different requirements are needed in its readout. The mini-MALTA data is synchronised to either a 640 MHz or a 40 MHz clock that is provided externally. In the slow readout mode, the valid data is ac-

knowledged by sending an active-low signal in a different pad for the duration of the data word. The SC block also needs a 40 MHz clock to be provided. The writing of a SC configuration is enabled when sending an active-high signal into the ENABLE pad. The reading of the configuration is triggered by sending another active-high signal in the START pad. The configuration is then transmitted off the chip in a serial bit-stream and acknowledged in a different pad. The mini-MALTA readout has to supply the 40 MHz and 640 MHz clocks needed in the chip. It also needs to detect the data acknowledgement signal, store the corresponding hit information and make it accessible to the user in the DAQ PC. The SC transmission protocols also have to be matched in the readout. Again, SC commands are sent via Python-based scripts. The readout has then to send the ENABLE signal and the configuration bit-stream simultaneously to the chip. When the reading protocol is activated by the user, the readout has to send the START signal to mini-MALTA , detect the acknowledgement bit and send back to the user the corresponding information.

Traditionally, readout systems were implemented in Application Specific Integrated Circuits (ASICs) which are dedicated boards targeted to a specific application. Despite some advantages of the ASIC designs, such as radiation hardness and the production of more compact circuits, the tendency is moving towards the use of Field Programmable Gate Arrays (FPGAs). Examples can be found on the COMPASS [72] and the ALICE [73] experiments at the LHC . The popularity of FPGA-based designs relies on their low cost and development time. However, FPGAs based on Static Random Access Memories (SRAMs) for their configuration, such as the ones employed in this thesis, are sensitive to radiation damage [74]. SRAM suffers from Single Event Effects (SEE) when exposed to radiation damage, specially Single Event Upsets (SEU) and Single Event Latchups (SEL).

For the purpose of creating a readout for a test device, FPGAs become the best option because of their reduced cost and development effort. SRAM-based FPGAs were chosen to develop the readout for the MALTA and mini-MALTA demonstrator chips. The MALTA readout is expected to be more complex due to the need of synchronizing the data after reaching the readout. Hence, it was developed in a Xilinx Virtex-7 FPGA, which is optimised for high performance applications. The mini-MALTA readout is implemented on a Xilinx Kintex-7 FPGA, which has less resources than the Virtex-7. Both FPGAs are discussed in section 6.3.1.

6.2 Introduction to FPGAs

Field Programmable Gate Arrays are configurable semiconductor devices in which digital circuits can be implemented. FPGAs are composed of arrays of Configurable Logic Blocks (CLB) interconnected by a stream of programmable routing fabric that can be configured by the user after its manufacture [76]. These are surrounded by series of Input/Output (IO) ports to connect the FPGA to an external board. This arrangement is shown in figure 6.1.

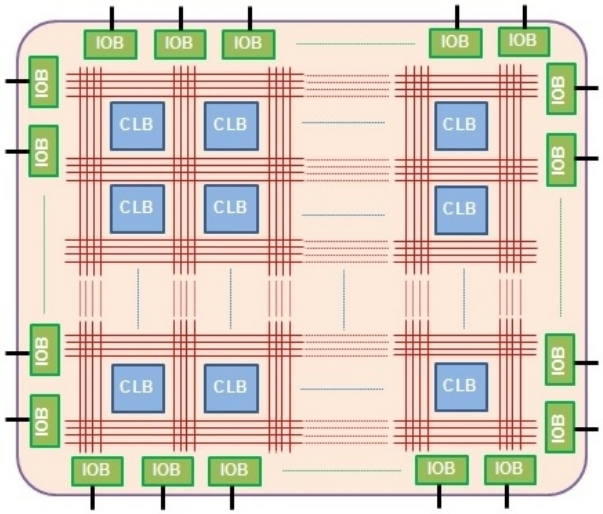

Figure 6.1: Schematic of a general FPGA device.

The programmability feature can be implemented in different ways, the flash, SRAM and anti-fuse approaches being more widely employed in modern FPGA technology [77]. SRAM-based FPGAs offer the advantages of being re-programmable and using standard CMOS technology processes on their fabrication. The use of CMOS processes allows the exploitation of the latest advances in this technology underlining the possibility to create more compact modules with low power consumption and high operational frequency. For these reasons, this option is more commonly found in industry and has been chosen to develop the MALTA and mini-MALTA readouts. SRAM FPGAs use static memory cells, such as the one shown in figure 6.2, to configure the routing and components of the FPGA. Routing is usually implemented in the form of multiplexers with lines that can be selected by the SRAM cell. The other memory cells are used to configure the state of CLBs that implement logic functions. SRAM cells can be

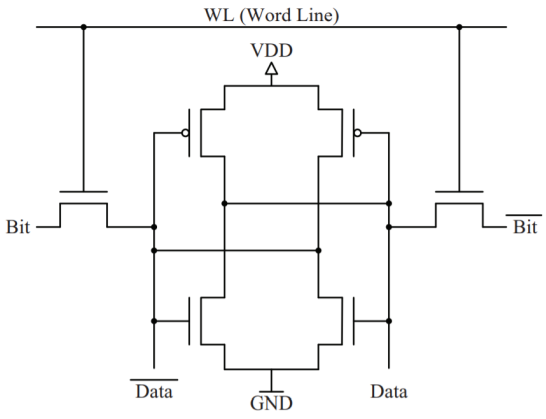

Figure 6.2: Schematic of a static memory cell [77].

programmed multiple times, offering the possibility to change or implement different circuits on the same FPGA [78]. However, due to the volatile nature of static memories, configuration

is lost upon the power down of the device. Due to the need for an external configuring element this can be a drawback for some applications, especially when the design is confidential. As this is not the case of the TJ-demonstrator readouts this fact did not constitute an issue for this design. Different circuits can be designed on an FPGA board using its programmability feature to configure the multiplexers and CLB modules. CLBs are blocks of logic elements, such as flip-flops, *Look-Up Tables* (LUTs) or multiplexers. These blocks can be internally configured to perform different operations (boolean, data-storage or arithmetic). The Xilinx 7-series FPGAs contain different types of CLB blocks tailored to specific functions. A more comprehensive discussion of the 7-series CLB elements and the specific resources employed in the design of the TJ demonstrator readouts is given in section 6.3.

6.3 The Xilinx 7-series FPGAs

The CLB elements of the Xilinx 7-series FPGAs [79] are arranged in two slices. These slices are connected to the general routing system through a switch matrix, as shown in figure 6.3. Each

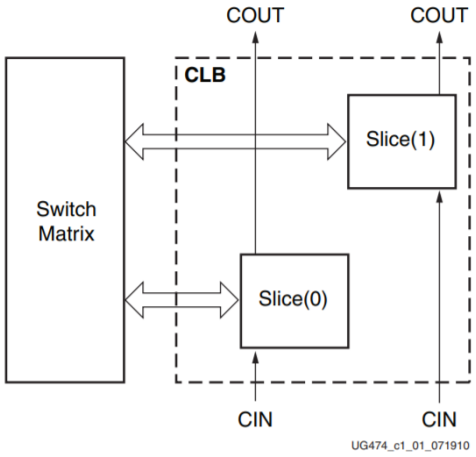

Figure 6.3: Distribution of slices and connections on a Xilinx 7-series CLB element [80].

slice is constituted by four 6-input LUT, eight flip-flops, wide multiplexers and arithmetic carry logic [80]. Each 6-input LUT can also be configured as two 5-input LUTs with the same inputs. In this case, the flip-flops can be used to store the outputs of the LUTs. Moreover, one-third of the slices are designed to use their LUTs as 64-bit memory RAMs or 32-bit shift registers.

Aside from the already mentioned CLB blocks, there are some resources in the 7-series Xilinx FPGAs that make easier the task of sampling an input signal and serializing or deserializing an input/output word. Those components are crucial in the MALTA and mini-MALTA readout firmwares and hence described in the next sections.

Input/Output buffers

A digital buffer is an electronic component that is used to isolate the input from the output, providing the same as the input voltage. Buffers are present next to the input/output ports of the FPGA and in the inner matrices for clock distribution. There are three different types of Input Buffers (IB) in the readout: a simple single input to single output buffer (IBUF), a two input to single output (IBUFDS) and a two input to two output (IBUFDS_DIFF_OUT) buffer. The schematic for those three cases is shown in figure 6.4. Everything is mirrored for the output buffers. The clock buffers are exactly the same as the input/output buffers but with a different naming (IBUFG). The I(O)BUF is automatically inserted in any signal directly connected to a top-level input (output) port if any other buffer is not specified. The I(O)BUFDS is an input (output) buffer that supports Low-Voltage Differential Signaling (LVDS). Its two ports (I and IB in figure 6.4) are opposite phases of the same logical signal. The I(O)BUFDS_DIFF_OUT also accepts differential signaling but differs from the IBUFDS in that it allows internal access to both phases of the differential signal.

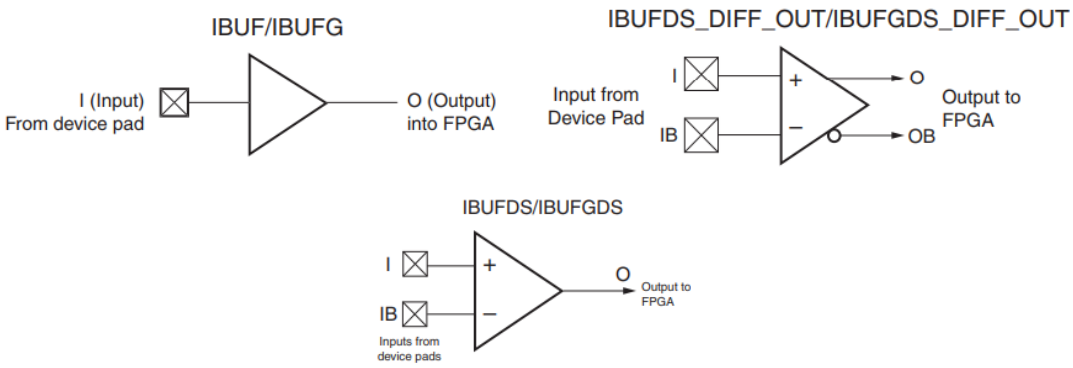

Figure 6.4: Schematic view of the IBUF (top left), IBUFDS_DIFF (top right) and IBUFDS (bottom) components of the 7-series Xilinx FPGAs. Figure reproduced from [81].

Clock Management Tiles (CMT)

Clock management and distribution in the 7-series FPGAs is driven by the Clock Management Tile (CMT) element. Each CMT is formed by a Mixed-Mode Clock Manager (MMCM) and a Phase-Locked Loop (PLL). PLLs are not included in the designs presented in this thesis and, hence, only MMCMs are discussed in this section. The MMCM functions include frequency synthesis, phase-shifting and clock deskew among others.

A block diagram of the MMCM module is shown in figure 6.5. A reference clock needs to be provided to the CLKIN[1:2] port from either the internal FPGA routing or an external source via the dedicated input pads. A programmable counter divider (D) divides the input clock frequency. Then, the phase and the frequency of the input and feedback clocks (CLKFB) are compared in the Phase-Frequency Detector (PFD). The PFD controls the Charge Pump (CP) and Loop Filter

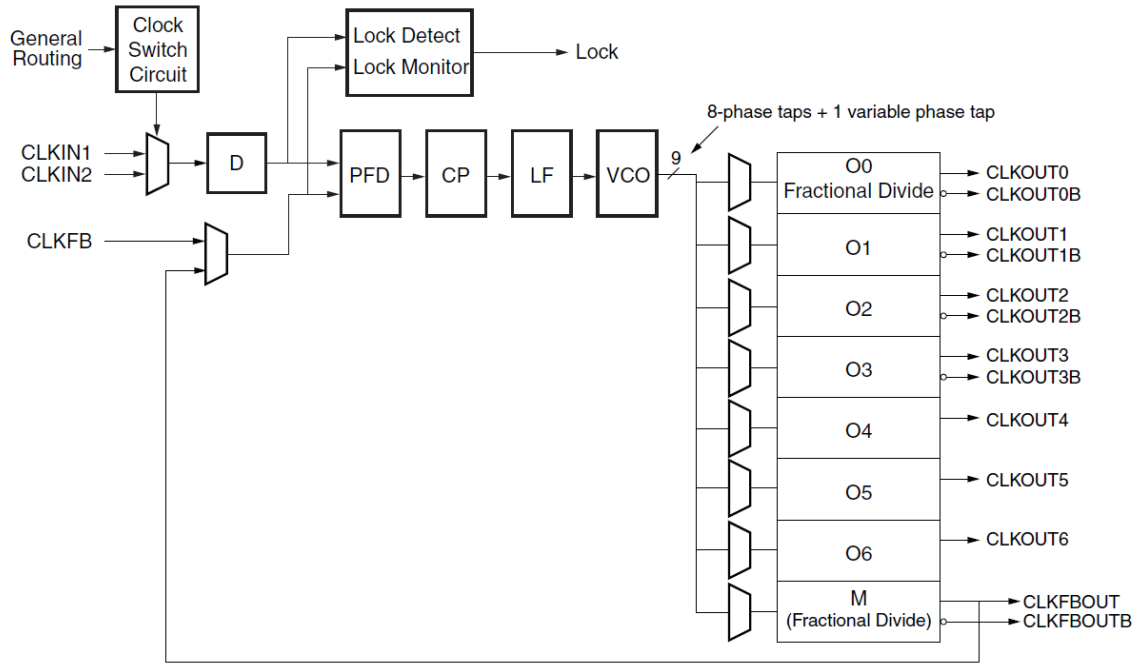

Figure 6.5: Block diagram of the MMCM module. Figure extracted from [82].

(LF) which, in turn, determine the frequency of the Voltage Controlled Oscillator (VCO). The VCO provides eight different clock phases and a variable phase that can be finely tuned. The eight VCO constant phases are 0°, 45°, 90°, 135°, 180°, 225°, 270°and 315°. Any of the nine VCO phases can be selected in each of the MMCM outputs (CLKOUT). Moreover, inverted versions (180°phase shift) of CLKOUT[0:3] are accessible in the CLKOUT[0:3]B ports.

The frequency of the VCO counter is determined by the frequency of the input clock (f_{clkin}) and the programmable D and M parameters [82]:

$$f_{VCO} = f_{clkin}\frac{M}{D} \tag{6.1}$$

The clock frequency at each MMCM output can be individually programmed by tuning the corresponding O divider. Thus, the output frequency is given by:

$$f_{out} = \frac{f_{VCO}}{O} \tag{6.2}$$

Note that the VCO frequency is common to all output counters, constraining the achievable output clock frequencies (as O only takes integer values). This constrain is lighter in the CLKOUT0 port, since it supports a fractional divide with a resolution of $f_{VCO}/8$ Hz.

IDELAYE2

Every IO block contains a programmable absolute delay element called IDELAYE2. The IDELAYE2 can be either connected to the input port of the ISERDESE1 component or driven di-

rectly into FPGA logic. The IDELAYE2 has 31 calibrated delay elements (*taps*). It allows to delay incoming signals on an individual basis. The IDELAYE2 works in conjunction with the IDELAYCTRL element, which provides a reference clock input that allows internal circuitry to derive a voltage bias, independent of process, voltage, and temperature variations. In order to define precise delay tap values for the associated IDELAYE2 component, the tap delay resolution is varied by selecting an IDELAYCTRL reference clock. The clock frequencies supported are 200 MHz, 300 MHz and 400 MHz which correspond to individual tap delay values of 78 ps, 52 ps and 39 ps respectively.

The IDELAYE2 module has different modes of operation: FIXED (static delay value), VARIABLE (for a dynamic increment/decrement of the delay value), VAR_LOAD (to dynamically load tap values) and VAR_LOAD_PIPE (for pipelined dynamically loadable tap values). All IDELAYE2 elements implemented in the MALTA readout are configured in FIXED mode. In this configuration, only the IDATAIN and DATAOUT ports are connected (see figure 6.6) and the rest is set to ground. The IDATAIN port is connected to the input signal to be delayed. This signal has to be driven from its associated IO port, contrary to the DATAIN connection in which the input is directly driven by the FPGA logic and cannot be driven from/to an IO. DATAOUT is the delayed data from either the IDATAIN or DATAIN input paths. It can be connected to an ISERDESE2 (see next section), input register or FPGA logic.

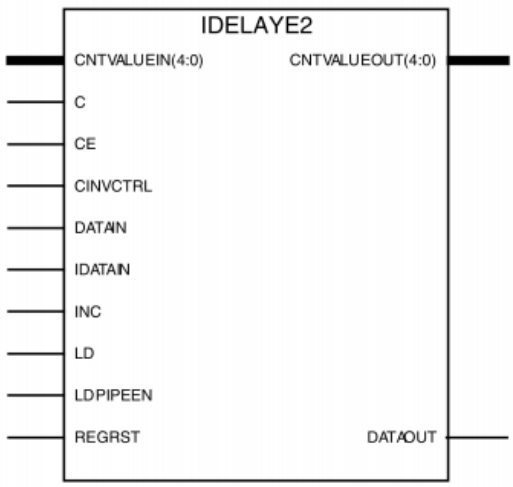

Figure 6.6: Schematic view of the IDELAYE2 component of the 7-series Xilinx FPGAs.

Input serial-to-parallel logic resources (ISERDESE2)

The ISERDESE2 is a dedicated deserializer/serial-to-parallel converter with an oversampling mode. This deserializer enables high-speed data without requiring the FPGA fabric to match the input data frequency. This feature is particularly useful to sample the asynchronous data of MALTA (see section 6.4.2). It supports both Single Data Rate (SDR) and Double Data Rate

(DDR). In SDR mode the output is a 2 to 8-bit wide parallel word. In DDR mode the output is a 4, 6 or 8 parallel word.

The possible configurations of the ISERDESE2 component are MEMORY, MEMORY_DDR3, MEMORY_QDR, OVERSAMPLE and NETWORKING. Only the OVERSAMPLE mode is used in the MALTA readout firmware. Figure 6.7 shows the internal circuitry of the ISERDESE2 when in OVERSAMPLE mode. This configuration is used to capture two phases of DDR data.

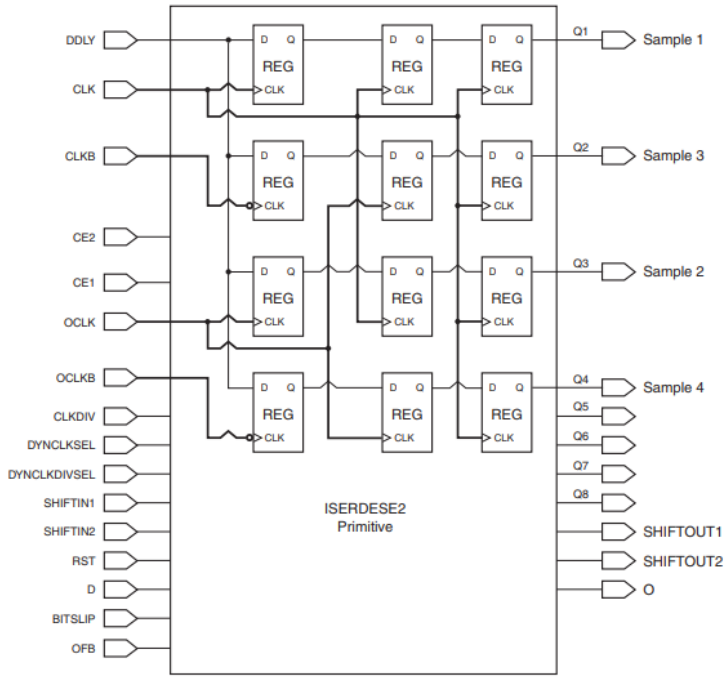

Figure 6.7: Schematic view of the ISERDESE2 component of the 7-series Xilinx FPGAs.

The data is captured on both the rising and falling edge of CLK and OCLK but is clocked out of the ISERDESE2 on the CLK domain. Four clocks need to be provided (CLK, OCLK, CLKB and OCLKB), which have to come from a single MMCM. All of them need to have the same frequency and a phase of 0°, 90°, 180° and 270° respectively. The serial input data port (DDLY) is the data input port for the ISERDESE2. This port only works in conjunction with the IDELAYE2 resource. The deserialized data is output in the Q1 to Q8 output ports. All the other ports are grounded.

The Virtex-7 FPGA contains a total of 74650 slices. Designing an efficient circuit including routing and timing considerations can be overwhelming. For this reason, the community has developed Hardware Description Languages (HDL).

6.3.1 Kintex-7 and Virtex-7 FPGAs

The Kintex-7 and Virtex-7 FPGAs belong to the Xilinx 7-series FPGAs. All FPGAs in this series have a common architecture, facilitating the migration of projects between devices, but

differ in size and performance. For comparison, the number of different resources available in the Kintex-7 and Virtex-7 FPGA models employed in the design of the MALTA and mini-MALTA readouts are specified in table 6.1. The Virtex-7 is basically a larger version of the

	FPGA product	**Kintex-7 XC7K325T**	**Virtex-7 XC7VX485T**
Logic Resources	Slices	50950	75900
	Logic Cells	326080	485760
	CLB Flip-Flops	407600	607200
Memory Resources	Total Block RAM (Kb)	16020	37080
Clock Resources	CMTs (1 MMCM + 1 PLL)	10	14
I/O Resources	Maximum Single-Ended I/O	500	700
	Maximum Differential I/O pairs	240	336
Integrated IP Resources	DSP Slices	840	2800
	PCIe Gen2	1	4
	GTX Transceivers (12.5 Gb/s Max Rate)	16	56

Table 6.1: Number of the different resources included in the Virtex-7 and Kintex-7 FPGAs used in this work. The MALTA readout uses 1.27 % of the Virtex-7 resources while the mini-MALTA readout utilises 2.97 % of the Kintex-7 resources.

Kintex-7 FPGA.

Both FPGAs are available in an evaluation board that provide most of the functionalities and connectivity commonly required by signal processing systems. The Virtex-7 FPGA is embedded in the VC707 evaluation board [83] while the Kintex-7 is mounted in the KC705 board [84]. These boards provide the hardware environment needed to operate the MALTA and mini-MALTA readouts and establish communication with a computer for DAQ and configuration.

VC707 and KC705 Evaluation boards

The VC707 and KC705 evaluation boards are equipped with standard connectors and user interfaces typically required in processing systems. The IO pins of the Virtex-7 (VC707) and the Kintex-7 (KC705) FPGAs are wire-bonded to these interfaces allowing easier access from the user. The capabilities utilized in the MALTA and mini-MALTA readouts are common in both boards. A block diagram of these interfaces is shown in figure 6.8. A technical description of these components as well as its use in the readouts is given in the following list:

- FMC HPC/LPC connectors: The VITA-57.1 [85] standard FPGA Mezzanine Connector (FMC) consists of a 40×10 pin connector. In the VC707 board two High Pin Count (HPC) FMCs are available. In the KC705, one FMC connector is substituted by a Low Pin Count (LPC) FMC. Both versions have a form factor of 40×10. However, in the HPC all 400 pins are present while the LPC is only populated with 160 pins. Each pins of the FMC is connected to different buffers of the FPGA. They are used to transmit all the signals

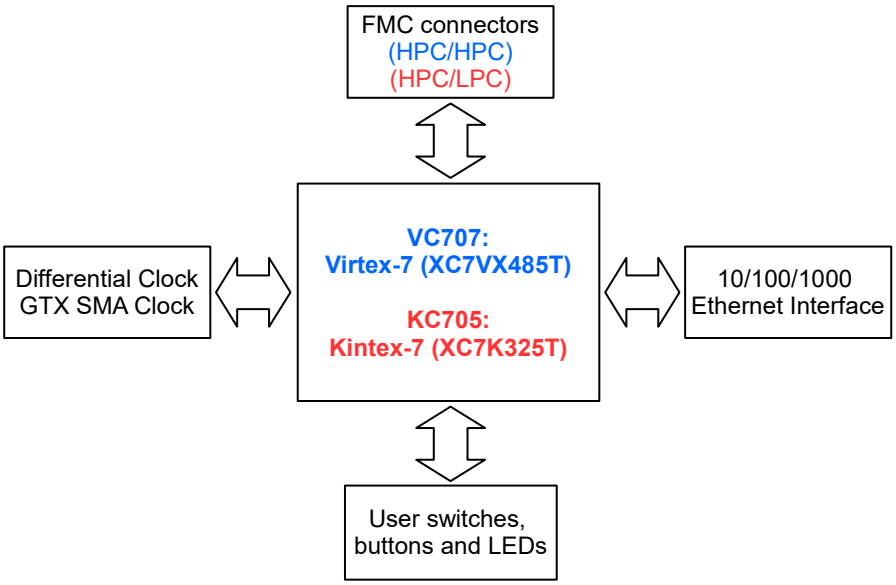

Figure 6.8: Block diagram of the resources of the VC707 and KC705 evaluation boards employed in the MALTA and mini-MALTA readout designs. The text in blue is for components exclusive to the VC707 board while the text in red is for components of the KC705 board.

required or sent by the MALTA and mini-MALTA chips. In the MALTA readout, two pins are used for the input and output slow control commands, one to send test pulses to the MALTA analog front end, and 37 pairs of LVDS signals to receive the data output from MALTA. In the mini-MALTA readout, two pairs of pins in the HPC FMC are used to transmit a 40 MHz and a 640 MHz differential clock, two pairs to receive the slow and fast data differential signals and one pair to receive the slow data acknowledgement signal. The signals required for SC operation described in table 4.4 and the test pulses are also transmitted through single FMC pins.

- Ethernet interface: The boards can support communications at 10, 100 or 1000 Mb/s via a Serial Gigabit Media-Independent Interface (SGMII) [86]. This interface is used in both readouts to establish communication with an external computer for user communication with the FPGAs. This allows the user to send input commands onto the chip (e.g. slow control commands, test pulses, reset signals and configuration parameters for the asynchronous oversampling module) and to read out the slow control configuration and data words.

- System clock: Both boards have a LVDS 200 MHz oscillator with 50 ppm[1] tolerance, soldered in the back side of the board and wired to an FPGA clock input. This clock is used as the input clock of the MMCM capability to produce all the clock resources needed by MALTA and mini-MALTA readout circuits. More details on the particular

[1] *parts per milion*

clock frequencies and usage are given in next sections.

- User LEDs: There are 8 user LEDs in the top of the VC707 and KC705 boards. They are used for monitoring the status of some FIFOs in the readout circuit.
- User buttons: Two pushbuttons in the VC707 board are used, one to reset MALTA manually and the other to reset MALTA and the readout FIFOs.

6.4 The MALTA Readout firmware

The readout for the MALTA chip is implemented in a Xilinx Virtex-7 FPGA. The MALTA signals are driven through an FMC connection to the VC707 board, in which the MALTA readout is implemented. A communication with the DAQ PC via ethernet connection allows monitoring of the readout status, data acquisition and sending control commands to the chip and the readout firmware. A schematic of this setup is shown in figure 6.9. A more detailed schematic of the MALTA readout firmware is shown in figure 6.10. It consists in two independent blocks: a slow control block, for external pulsing and configuration of the chip and an asynchronous oversampling (AO) block, to read out and synchronise the data output of the chip. The communication protocol with an external computer is established by the IP bus block. The IP bus is used to input the slow-control commands and to configure some parameters of the AO module.

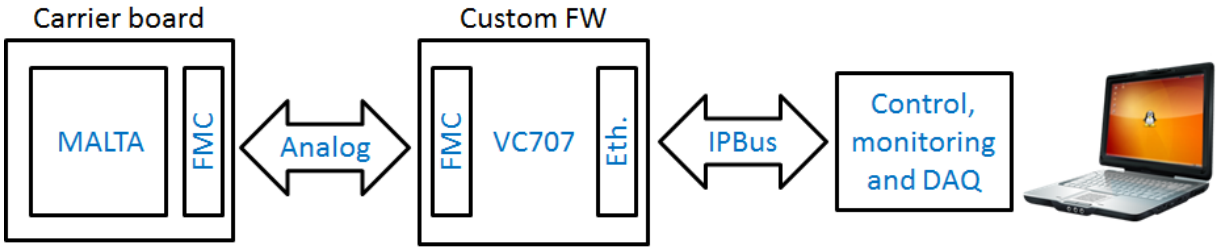

Figure 6.9: Sketch of the full MALTA readout setup. The MALTA chip is connected to the FPGA based readout via a FMC connection and the readout connects to a computer via ethernet. The computer is used to send signals to the firmware and to readout the data coming from MALTA.

The AO module takes the 37-bit LVDS data from the MALTA chip and samples it using the IDELAY and ISERDES components described in section 6.3. The sampled data goes into a data recovery unit (DRU) module that selects and formats it to be pipelined into a small FIFO that is, in turn, connected to a bigger FIFO in which timing information is also included.

The Slow Control module simply consists of an ISERDES serializer and an ISERDES deserializer. The serializer is used to write the slow control commands into MALTA and the deserializer to read them back in the FPGA. The bit patterns are sent from the DAQ PC via the IP bus module. The AO and Slow Control modules are explained in more detail in the following sections.

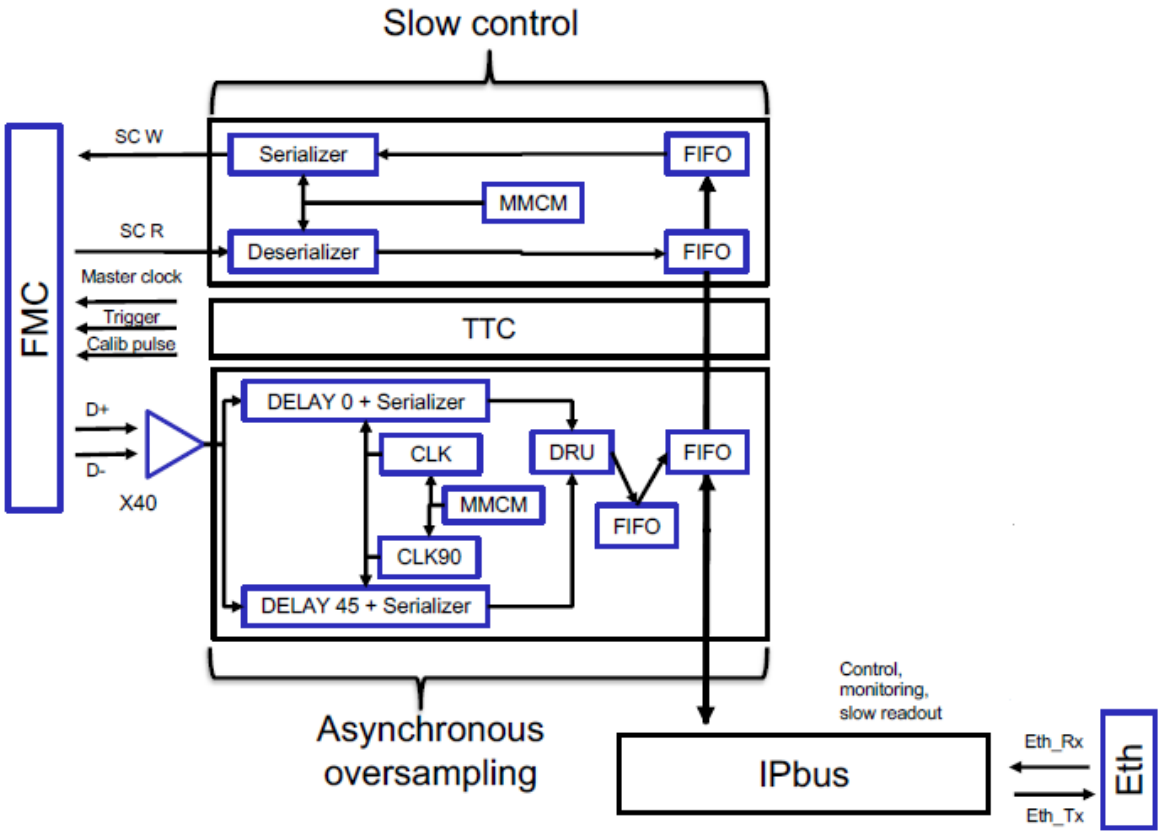

Figure 6.10: Schematic of the MALTA readout firmware implemented in the Xilinx FPGA. The three main blocks of the firmware (Asynchronous Oversampling, Slow Control and IP bus) including their main inner circuitry are shown in this schematic.

6.4.1 Slow Control

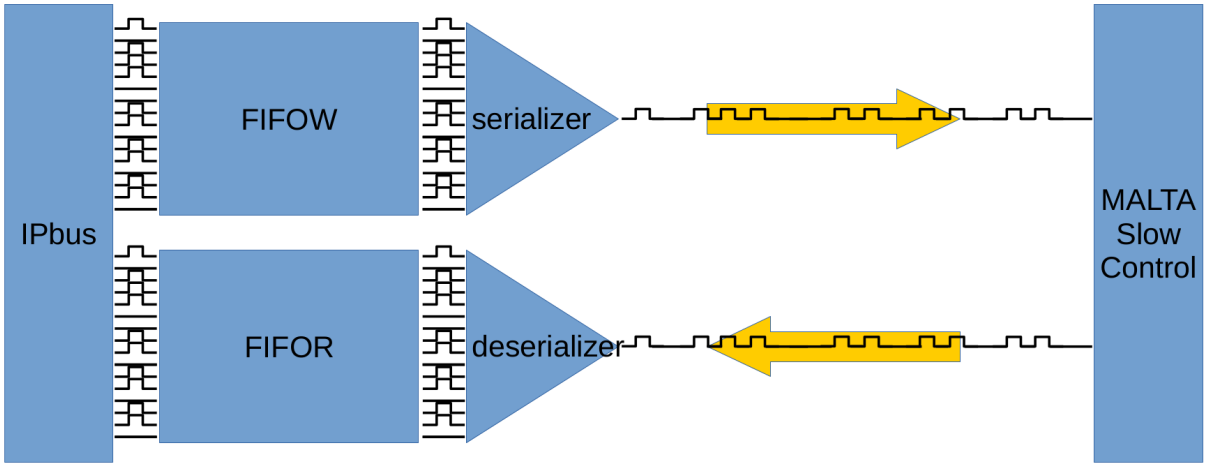

Figure 6.11: Sketch of the MALTA Slow Control communication strategy implemented in the FPGA-based readout.

Communication with the MALTA Slow Control is implemented in the FPGA-based MALTA

Readout system. A sketch of the MALTA Slow Control communication strategy is shown in figure 6.11.

The proper operation of the MALTA Slow Control communication requires the user to send and receive commands via the IPbus protocol. Once a command is sent, the software must wait for a few clock cycles of the Slow Control 10 MHz clock[2] and then issue a read request in order to check whether the Slow Control has returned the original word back in case the command was invalid.

The Slow Control commands are transmitted from the computer via the IPbus protocol in 32 bit words, out of which the 16 bit Slow Control word is extracted and stored in the writing FIFO (FIFOW). As soon as the FIFO is not empty, the serializer starts reading the FIFO and transmits each word present in the FIFO sequentially to the MALTA Slow Control via the SERIAL_INPUT line. As soon as the MALTA Slow Control issues a serial command (whose start is signaled by the SERIAL_OUTPUT line level going low for one clock cycle), the deserializer packs the command into a 16 bit word that is pushed to the reading FIFO (FIFOR). The FIFO is read out when the read request is sent from the IPbus.

6.4.2 Asynchronous oversampling

The technique used to recover the data from the MALTA chip is called *Asynchronous Oversampling* (AO). The term *oversampling* means to sample a digital signal with a frequency higher than twice its duration. If the sampled signal is not synchronous to the sampling clock, the method is called asynchronous oversampling. In the MALTA chip the data is asynchronous to any clock (see section 4.3.1). One of the critical parameters on the MALTA chip is the time between pulses, since it is related to the energy deposited in the event. Hence, good resolution in the signal edge is required. To achieve this, it is required to oversample the digital pulse. Aiming for a precision of 250 ps, it translates to a sampling clock of 4 GHz. Such a high frequency clock is very difficult to generate and not supported by the FPGA. Hence, a strategy to implement the AO technique in the FPGA is required.

The AO technique is implemented by duplicating and delaying the data and clock signals, as shown in figure 6.12. The duplicated clock and data are delayed by 90 and 45 degrees respectively. Then, the rising and decaying edges of both clocks are used to sample the original and the delayed data. Using this technique and a clock frequency equal to the data frequency it is possible to sample each data bit four times. This circuit can be constructed using an MMCM, an IDELAYE2 and an ISERDESE2, as shown in figure 6.13. A single MMCM provides all the clocking resources needed for the module: two 500 MHz clocks with a phase difference of 90 degrees between each other and a 300 MHz clock for the IDELAYCTRL component. The 500 MHz clocks are duplicated and inverted to achieve the 180 and 270 degrees clocks required by the ISERDES components. Each pair of LVDS signals from the MALTA datalines is driven

[2] such condition is most likely to be naturally fulfilled by the latency of the PC itself

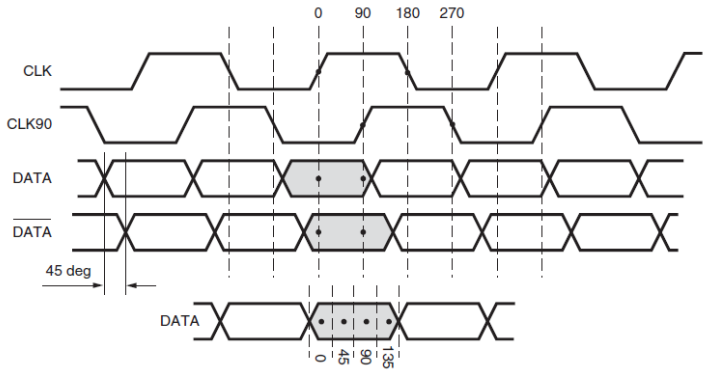

Figure 6.12: Effective sample edges when duplicating and dephasing the sample clock and the data.

into a different FPGA input. These are taken separately into an IDELAYE2 component. The negative copy of the signal is delayed 5 taps (260 ps) obtaining a version of the data delayed by $\sim$45 degrees. The positive copy goes through an IDELAYE2 module with a tap value of 0 to compensate for the intrinsic delay that this component applies to any signal passing through it. The 0 and 45 degree phase data passes to two ISERDES modules in which they are sampled with the 0, 90, 180 and 270 phased clocks. Eight sampled bits are created by combining the 4 clock phases and the 2 data phase samples. This 8 bit output is driven outside this module into the Data Recovery Unit (DRU) module. This same structure is applied for each of the 37 data lines from MALTA.

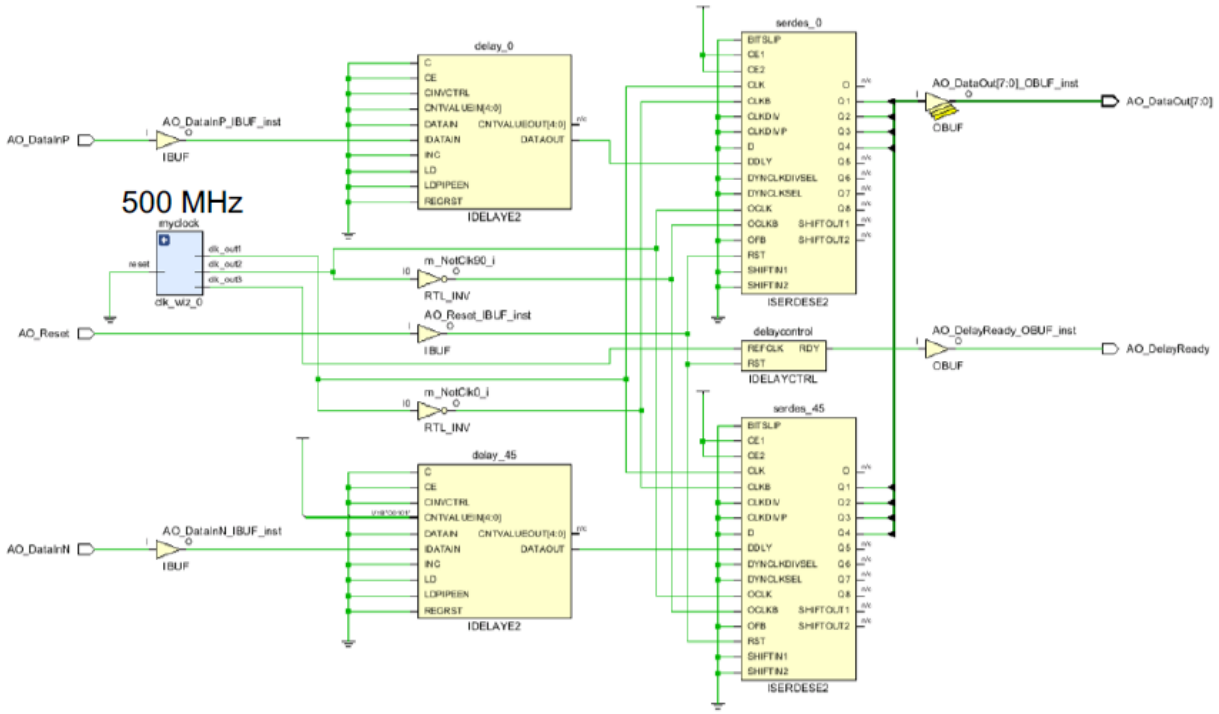

Figure 6.13: Schematic of the Asynchronous Oversampling module.

The AO module is also used to compensate for possible delays in between FMC channels

that come from either the architecture of the chip or the connections in the MALTA PCB . This can be corrected using the IDELAYE component and applying a different tap value for each of the data channels. This delay can only be measured by applying a simultaneous pulse to each channel and measuring the difference between the data channels. The different tap values are calculated offline through dedicated python-based scripts and input using the IP bus. In figure 6.14 there is an example of this methodology using as an input another FPGA and sending 300 signals to each of the FMC pins used for MALTA and passing it through the FPGA based MALTA readout. As seen in the picture on the right, good alignment can be achieved.

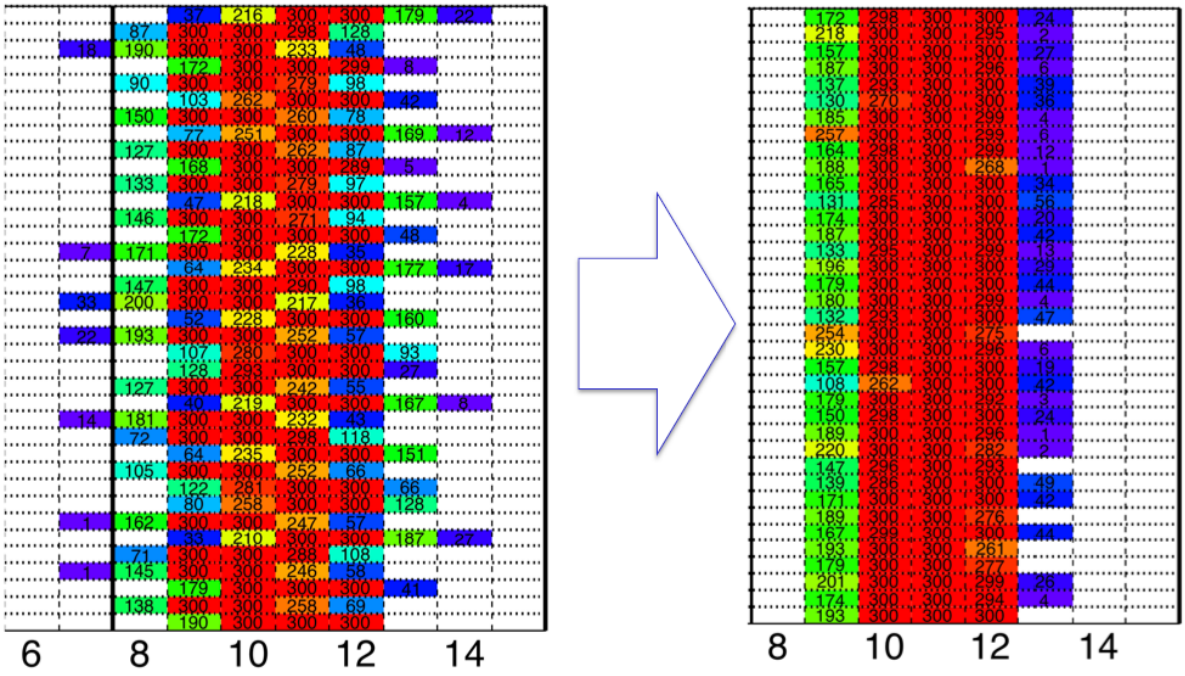

Figure 6.14: 8 bit output of the 37 channels of MALTA after going through the AO module before (left) and after (right) aligning them. The data sent to the readout was a single bit to all the 37 channels using a FPGA-based MALTA emulator. This was repeated 300 times to get better statistics.

A schematic of the DRU module is shown in figure 6.15. It consists of two blocks: one to determine and synthetise the position of the rising edge of the reference signal (Edge Detector module) and the other to, whenever there is a hit in the reference signal, determine what is the status of the other 36 channels (Position Detector module). The design of the two modules is similar: they both assume that the signals are 1 ns long and, hence, that only 4 of the 8 bits will be 1 if the full signal is contained within this sampling window. Considering this assumption, the only possible cases to be considered are in table 6.16. In the yellow and green cases, the edge detector module sends a trigger signal to the position detector module. Those are the cases where the rising edge of the signal is within this sampling window. In the red cases no trigger is sent since that should be done in the previous sampling. The position detector module takes this trigger signal and checks if the channel is up or down. The same cases that appear in figure 6.16 have to be considered in order to not sample signals corresponding to a consecutive hit

that overlaps with the previous sampling window. Both modules are implemented in a series of LUTs to consider each of these cases individually and then send the information about the position of the rising edge in a 3 bit word and a 36 bit word for the status of the other MALTA data lines. These bits give information about the pixel that was fired and the energy deposited. Also notice that the modules described above assume a very accurate alignment for the signals coming from different FMC pins, which is assured by the AO module.

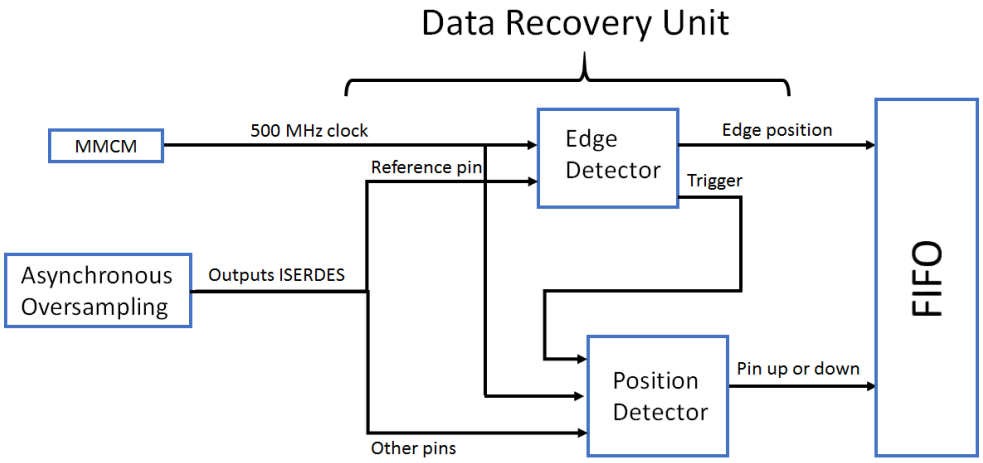

Figure 6.15: Schematic of the Data Recovery Unit (DRU) firmware.

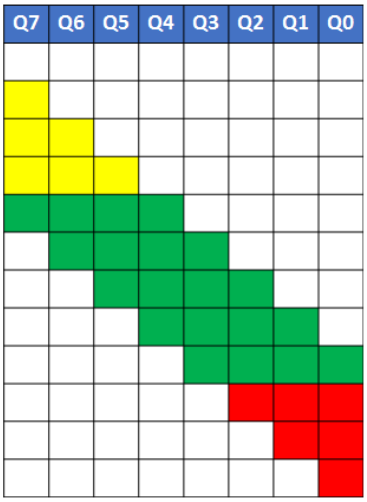

Figure 6.16: Possible situations for a single bit after coming out of the ISERDES component. The headers of the table (Q0-Q7) represent the 8 outputs of the ISERDES component. The color legend is the following: green bits are accepted in this sampling window without any conflict, yellow bits are accepted but another signal could overlap in the same window and red bits are not accepted because they are accepted in the previous window.

The full AO module including the DRU has been tested using an emulator for the MALTA chip. The emulator was capable to send different signal patterns through the FMC connector

as well as different signal widths. This was used to measure how many times a signal is misunderstood. When sending a single bit at a time, a false positive happens less than 2×10^{-6} % (0 in 500k events) and a false positive happens less than 1×10^{-6} % of the times (0 in 1 million events). Two consecutive bits were also sent in the same sampling window to see if there is any error in the edge detection or in the status of the different bits. The same error rate and false positive rate is found in this case.

6.5 The mini-MALTA Readout firmware

The Mini-MALTA chip readout is implemented in a Xilinx Kintex®-7 FPGA. The Kintex-7 is slower than the Virtex-7 family. This reduction in the performance of the FPGA does not constitute an issue, as the oversampling of data is no longer needed in the Mini-MALTA chip. A schematic of the Mini-MALTA readout firmware is shown in figure 6.17. The project is divided

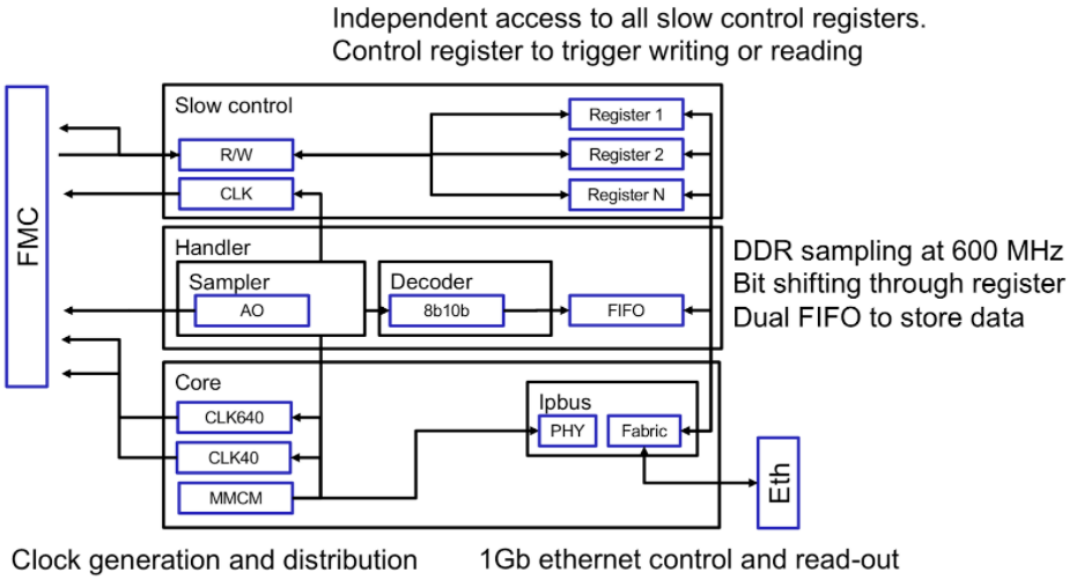

Figure 6.17: Block diagram of the mini-MALTA readout firmware implemented in a Xilinx Kintex-7 FPGA.

in three blocks: the core, the handler and the Slow Control. The core block contains the generation of the required clocks and the ipbus module for ethernet communication with the DAQ PC. The Slow Control block enables the SC communication with the chip by reading/writting the 30 SC registers described in table 4.3. The handler block is used to sample the hit information coming from the chip. The next sections are devoted to describe the SC and Handler blocks in more detail.

6.5.1 Slow Control

The Slow Control block of the mini-MALTA firmware enables the user to read and write the configuration of the chip. Similarly to the MALTA SC firmware, the mini-MALTA SC works by serializing and deserializing the configuration bit-stream, as represented in figure 6.11.

The serializer module is employed to upload a specific configuration to the chip. This module has three input and two output pads, as shown in figure 6.18. The 30 SC registers can be accessed

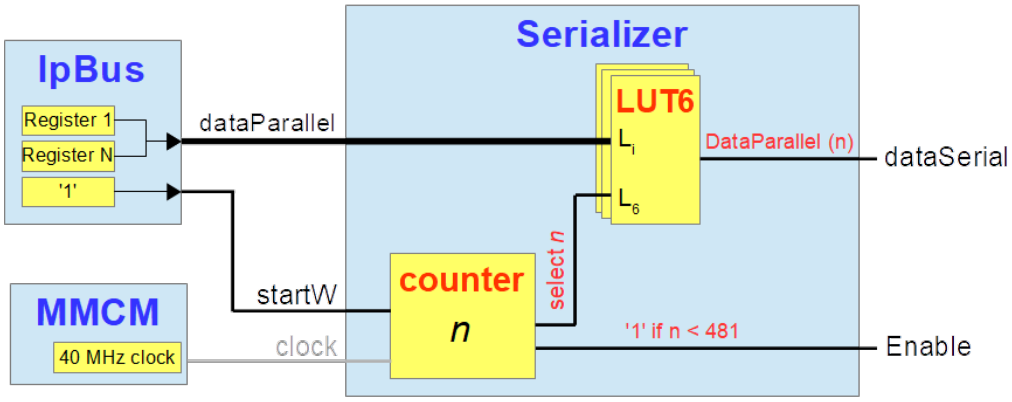

Figure 6.18: Simplified block diagram of the serializer module included in the SC block of the mini-MALTA readout.

independently from the software by selecting the corresponding IPbus address. However, the configuration is written in a block and, hence, a value must be specified to each register. The configuration word is transmitted to the FPGA via ethernet and then input in parallel into the *dataParallel* port of the serializer module. In order to initiate the serialization of the SC word, a digital '1' must be sent to the *startW* port for one clock cycle. Once this signal is received, an internal counter starts to increase with every clock cycle. This counter drives the serialization of data by selecting the corresponding bit of the parallel word, which is sent out of the module through the *dataSerial* port. This port is connected to the DATA pad of the mini-MALTA SC module (see table 4.4). The *Enable* output port is connected to the ENABLE pad of the SC and is set to '1' for the whole duration of the writing operation.

The deserializer module is used to split the serial SC word into its different registers and send it in parallel to the IPbus module. The value of each register can be then accessed via ethernet from the DAQ PC. A schematic of the deserializer module is shown in figure 6.19. It has four input and three output ports. The *clock* port is connected to the general 40 MHz SC clock. The *Isoutack* and *IdataSerial* ports are connected to the SOUT_ACK and SOUT pads of the mini-MALTA SC (table 4.4). Finally, the *IalignPattern* port is connected to the value of the first two registers sent by the user to the serializer module. The START signal required to activate the reading of the mini-MALTA configuration is sent from the IPBus by

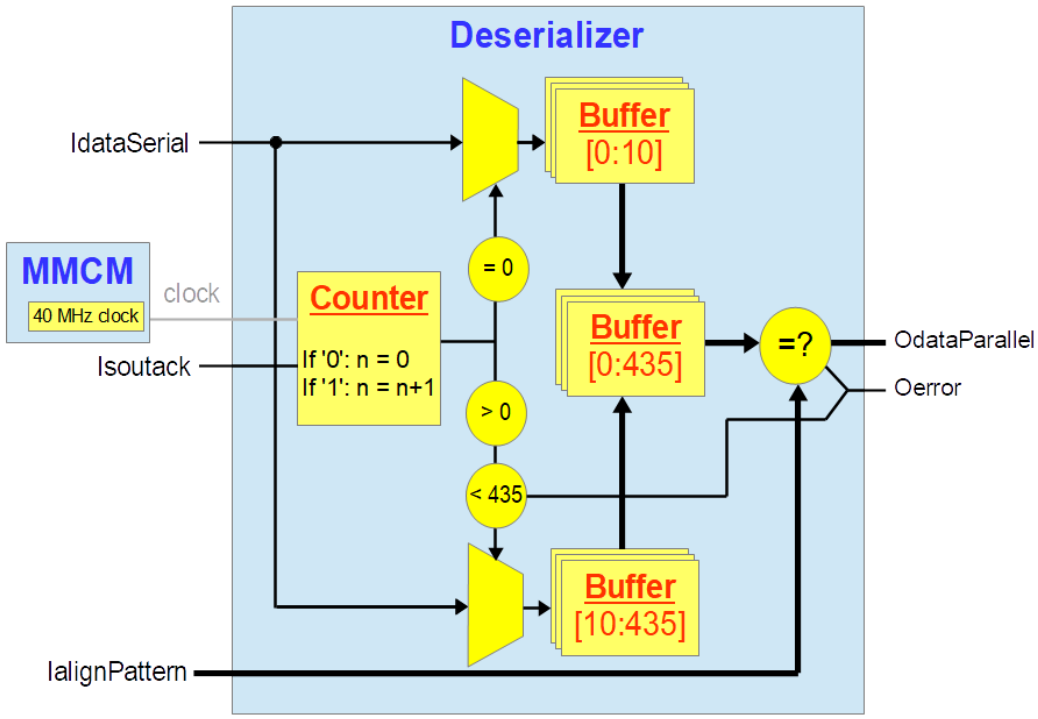

Figure 6.19: Simplified block diagram of the deserializer module included in the SC block of the mini-MALTA readout.

an external module. This signal takes several clock cycles to reach the mini-MALTA chip and start buffering the register values to the SOUT pad. While awaiting for the acknowledgement signal (*Isoutack*), the *IdataSerial* data is buffered into a 10-bit deep pipeline. Once *Isoutack* is received, the values of *IdataSerial* are buffered for 425 clock cycles. The deserializer module also accounts for possible desynchronizations between the *Isoutack* and *IdataSerial* signals. After 425 clock cycles, or when the *Isoutack* signal turns low, the buffered data is compared to the value of the first two registers previously written by the serializer module. Once this bit pattern is found, the valid 416 bit-stream is sent in parallel to the *OdataParallel* ports. If no coincidence can be found, an error is reported. The deserializer module can also detect errors in the acknowledgement line. If the *Isoutack* signal turns low earlier than expected (before 416 clock cycles), an error code is sent to the *OdataParallel* port and no data is read.

6.5.2 Data Handler

The Data Handler module is devoted to processing and storing the hit information coming from the mini-MALTA chip. The mini-MALTA readout version described in this work is not compatible with the mini-MALTA fast read-out mode. This version was made to simplify the readout system and avoid any potential issues in the 8b/10b encoder.

The mini-MALTA data word of the slow read-out mode consists of 48 bits arranged in the following way: 47 to 44 for the synchronisation memory address, 43 to 28 for the pixel address, 22 to 19 for the group address, 22 to 19 for fine time information, 18 for the reference, 17 to 15 for the Gray encoded BCID and 14 to 0 for an additional BCID time stamp. This data is synchronized to an external 40 MHz clock in the end of column logic and transmitted off the chip in series via a dedicated pad. The duration of the data word is marked by an acknowledgement signal, which turns to '1' while the word is being transmitted.

The Data Handler module is subdivided in three blocks that identify, process and store the mini-MALTA data. These are named *Slow Data Processor*, *Time Stamp Handler* and *Slow Data Storage*. A basic schematic of the Data Handler module is presented in figure 6.20. The

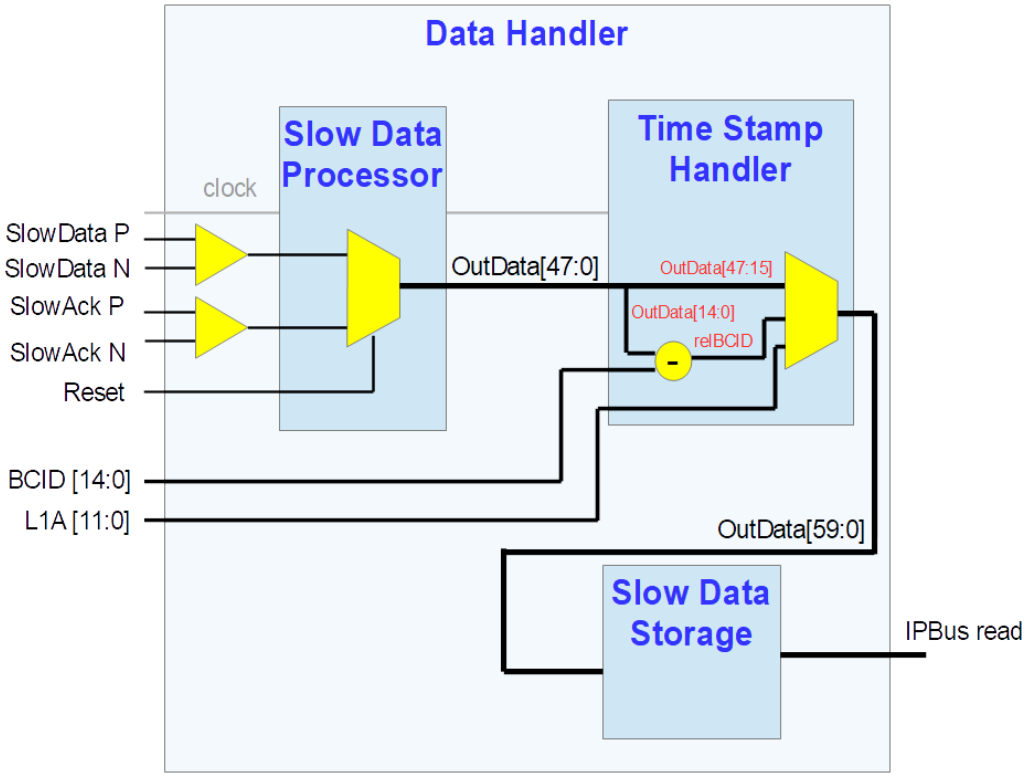

Figure 6.20: Simplified block diagram of the Data handler module of the mini-MALTA readout.

48-bit mini-MALTA word and its corresponding acknowledgement signal are connected to two IBUFDS buffers, which merge these LVDS signals (SlowData P/N and SlowAck P/N) into a single signal phase. The buffer outputs are connected to input ports of the Slow Data Processor module. Within this module, the Slow Data signal is constantly buffered for 57 clock cycles. When a transition from high to low is detected in the acknowledgement signal, the last 48-bits are sent in parallel into the *OutData* port of this module. The reading of data can be stopped and reset by sending a signal in the *Reset* port.

The mini-MALTA word is further processed in the Time Stamp Handler module. This mod-

ule is used to add a 12-bits coarse time stamp (*L1A*) in the mini-MALTA data word and modify the mini-MALTA BCID. The L1A time stamp and an additional 14-bits BCID are generated in an external module. The L1A counter is increased when receiving an external signal. This signal can be generated by either the DAQ PC in laboratory tests or an external counter (e.g. an FEI4 telescope in testbeams). The BCID counter increases with a 40 MHz clock and is reset when the L1A counter increases. This BCID is employed to recalculate the mini-MALTA BCID. The mini-MALTA BCID (*mmBCID*) is substituted for the BCID generated in the readout (*readBCID*) for the first data word received after a BCID reset. The consecutive mini-MALTA words receive a relative BCID (relBCID) calculated with the following expression:

$$relBCID = readBCID + \Delta mmBCID \tag{6.3}$$

where $\Delta mmBCID$ is the increment of the mmBCID since the first data word received. The final mini-MALTA word is then 60-bits long containing 12-bits for the L1A time stamp, 33-bits for the hit address and 15-bits for the modified BCID time stamp. This mini-MALTA word is written in an IPbus register in the *Slow Data Storage* module, which can be later accessed via ethernet from the DAQ PC.

6.6 Summary

The signals detected by particle detectors are processed by the readout circuits and stored in computers. Moreover, communication with the SC block controlling the functionalities of the chip is also accessed by the user from a computer. The interface between the chip and the computer, known as the detector readout, is specifically designed for each detector. Detector readouts were traditionally designed in ASIC boards, although the tendency is moving towards the use of FPGAs. FPGAs are reconfigurable boards in which digital circuits can be implemented. The readouts of the MALTA and mini-MALTA chips have been designed in a Xilinx Virtex-7 and Kintex-7 FPGAs respectively. Both FPGAs are wire-bonded in evaluation boards that provide different interfaces to be connected to the carrier board of the chip and to the computer. The MALTA and mini-MALTA signals are transmitted via an FMC connector and communication with the computer is enabled with an ethernet interface.

The MALTA readout is divided in two different blocks: the AO block to read out and synchronise the data output and the SC block to establish communication with the SC block of MALTA. The AO module is used to oversample the digital signal coming from the MALTA chip. The 40 parallel information bits are input in 40 different LVDS inputs of the FPGA, in which they are duplicated, dephased and sampled with two different clocks. With this technique, the 1 ns long signals of MALTA can be sampled 4 times. The AO module also includes an alignment feature to compensate for possible delays in the different data lines. The sampled signals

are driven into another module to extract the position (in time) of the signal and to determine the full address of the hit. The SC block of the MALTA readout consists of a simple serialiser, to serialise the 16-bit register value coming from the computer and to be input in MALTA, and a deserialiser, to deserialise the 16-bit response word coming from MALTA and send it to the computer.

The mini-MALTA readout consists of a data handler block to process the signals and a SC block to allow the communication with the SC block of mini-MALTA. The data handler module is used to identify, process and store the mini-MALTA data. The 48-bit mini-MALTA word and its corresponding acknowledgement signal are buffered into the FPGA in which they are deserialised. Then, the mini-MALTA time stamp is modified to match the counter of the mini-MALTA readout and 12-bits of coarse time stamp is added to the data. The SC block of the mini-MALTA readout is designed to match the SC block of mini-MALTA. The mini-MALTA registers can be written by the user in a custom software environment. These are transmitted in parallel to the FPGA, in which they are sorted and serialised in a single bus. The writing of data is enabled by sending a separate signal into the chip. In order to read the SC data, a signal has to be sent to the mini-MALTA SC to start the transmission of the 416 bit-stream containing the value of all mini-MALTA registers. After the start signal is sent, the SC block of the mini-MALTA readout awaits for the acknowledgement signal indicating that the received data is valid and starts deserialising and splitting the data corresponding to each register. Finally, the value of each register is sent to the computer to be accessed by the user.

Chapter 7

Investigation of the effect of ionising dose in TowerJazz Demonstrators

7.1 Description of TID setup

The X-ray irradiation facility at the University of Glasgow is based on a 100 kVp X-ray tube from Thales, with a tungsten (W) target, and a 3.2 kW HV generator from Gulmay. They are installed inside a 1.5 x 1.5 x 1.0 m^3 Lead (Pb) chamber. The system is controlled by a Gulmay MP1 integrated controller and the potential applied can be varied to change the energy of the X-rays and, effectively, the dose rate. The dose rate delivered by the system is calibrated. Dose rate

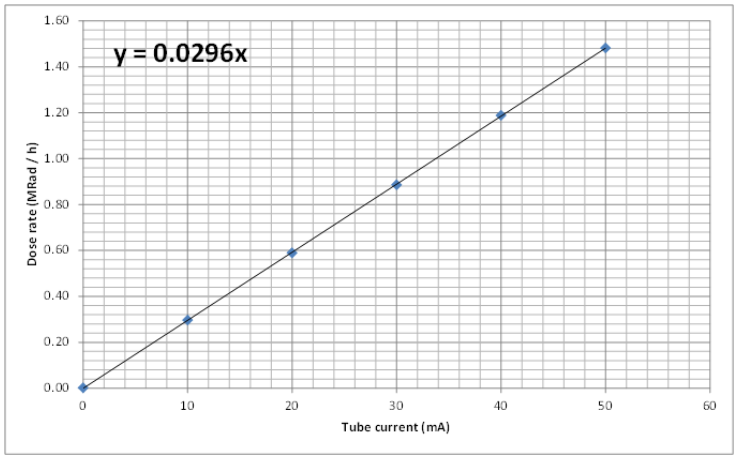

Figure 7.1: Dose rate delivered by the X-ray tube as a function of the tube current applied. This dependence is measured at a distance of 20 cm from the X-ray generator.

received is compared to a calibrated PIN diode provided by RAL [87]. At a distance of 20 cm from the X-ray source, the dose rate as a function of the tube current is shown in figure 7.1. The X-rays used to irradiate the device generate signals that are readout by the chip, increasing the analog and digital current consumption. In MALTA, the digital current reaches its compliance value at very low TID values (see section 7.3.2). Thus, a low dose rate of 0.25 MRad/h was

selected in order to avoid generating too much electrical activity on the chip, as it is powered during irradiation. The beam spot at a distance of 20 cm is shown in figure 7.2. It is measured

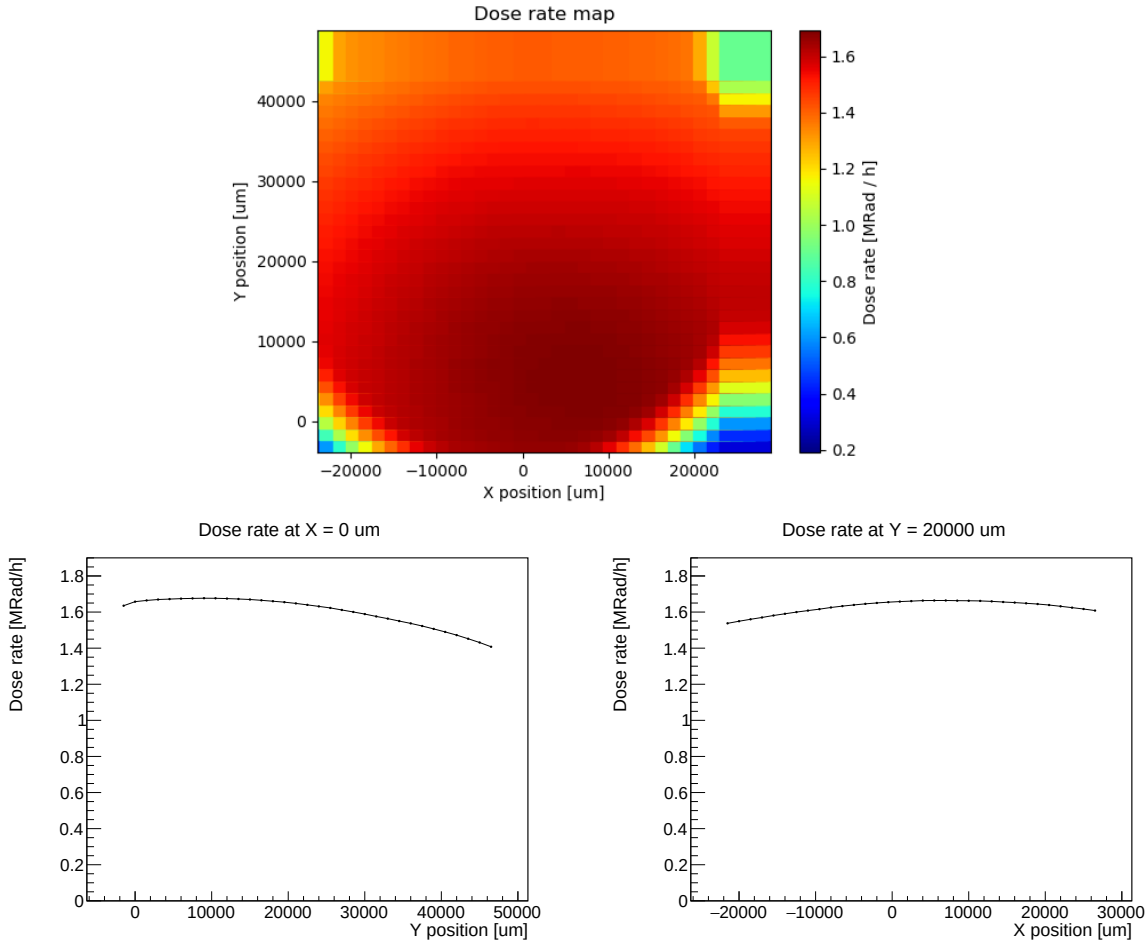

Figure 7.2: Dose rate map as a function of area delivered at 20 cm from the X-ray generator (top). Projections at $x = 0\,\mu$m (bottom left) and $y = 20000\,\mu$m (bottom right) are also presented.

by placing the RAL diode in a X-Y stage and measuring the current delivered by the diode at each position. Then a conversion factor is applied to convert current to dose rate. In the map shown in figure 7.2, 32×32 positions are measured with a step size of 0.15 cm. The projections along the X and Y axis in figure 7.2 show that the beam spot is uniform at the 3 % level in a circular shape of 2 cm of diameter. Hence, uniform irradiation damage can be applied to the full MALTA chip.

7.2 DUT characterisation techniques

Different tests were developed in collaboration with the design team of the MALTA and mini-MALTA sensors to characterise the different functionalities of each chip. Detailed descriptions of every technique are given in this chapter to provide the reader with a better understanding of

the studies presented in this thesis.

7.2.1 Threshold calibration scan

The purpose of this test is to measure the value of the threshold level applied at the discriminator stage of the front-end circuit. The threshold level is controlled by the IDB current, which can be provided either through the DAC circuitry or from an external PSU (see section 4.3.1). A desired value of the IDB current is selected and kept constant during the test. The threshold of a single pixel is calculated by injecting different charge values at the input of its front-end and measuring the value for which the pulse goes below the threshold. In order to have control over the amount of charge injected and the injection rate, signals are induced through the internal pulsing capability. This also allows to select a specific pixel to be measured. A number of test pulses (typically 100) above the selected threshold are sent by tuning the V_{HIGH} and V_{LOW} voltage levels as discussed in section 4.3.1. Then, the V_{LOW} voltage is gradually increased, to reduce the charge of the pulse (see equation 4.4), until no hits are observed at the data output of the chip. The number of pulses observed at the output of the front-end are computed for each V_{LOW} voltage, leading to the characteristic S-shaped curve of figure 7.3. Three regions can be

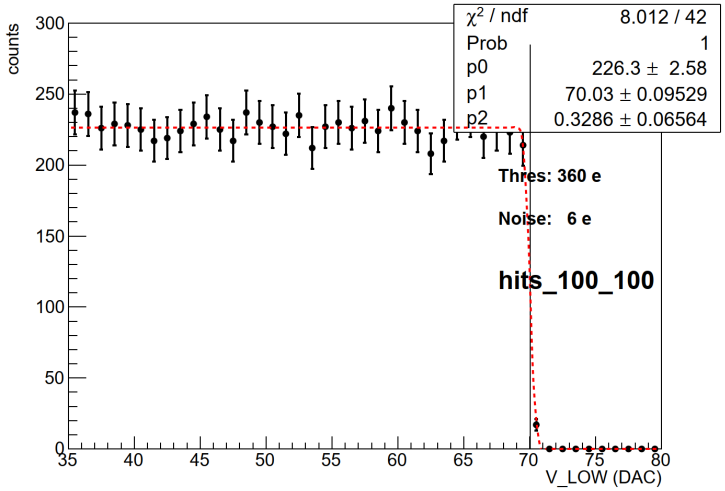

Figure 7.3: An example of a threshold calibration scan for a single pixel. This measurement was taken on the MALTA chip at an IDB voltage of 40 mV.

distinguished in this plot: A plateau where the number of counted hits is the same as the number of pulses sent, a section where no hits are observed and a transition region. The regions with full and no hits correspond to the test pulses being above and below the threshold respectively. On the transition curve, only a fraction of hits are above the threshold. This is explained by considering the random noise induced by both the sensor and the front-end electronics. Random fluctuations of the waveform amplitude can cause some pulses near threshold to fluctuate above threshold and be counted and other pulses to fluctuate below the threshold and not be counted.

Hence, the middle value of the transition curve is a measurement of the threshold in V_{LOW} DACs units. For better accuracy, the middle point is extracted by fitting an error-function to the data and extracting its mean value. Pixel noise can be extracted from the width of the transition curve, corresponding to the RMS of the error function fit.

The pixel threshold and noise calculated using this technique are obtained in terms of V_{LOW} DACs. This unit is very arbitrary since it depends on the particular front-end architecture and DAC circuitry design. Hence, a conversion factor to a more absolute unit is necessary. As mentioned in section 4.3.1, 1 DAC unit corresponds to 7 mV. Knowing that the amount of charge injected by the pulsing capability is 1.43 e^- per mV of difference between V_{HIGH} and V_{LOW}, the following conversion factor is obtained:

$$\mathrm{Thr}(e^-) = \frac{1.43\,e^-}{1\,\mathrm{mV}} \cdot \frac{(V_{\mathrm{HIGH}} - V_{\mathrm{LOW}}) \cdot 7\,\mathrm{mV}}{1\,V_{\mathrm{LOW}}\ \mathrm{DAC}} \cdot \mathrm{Thr}(V_{\mathrm{LOW}}\ \mathrm{DAC}) \tag{7.1}$$

The threshold calibration scan has been automated for both chips using python-based scripts. This allowed the threshold and noise to be measured over different sectors of the chip. This method has been used to study the radiation hardness of the threshold-related circuitry (sections 7.3.5 and 7.4.3) and the increase of noise with irradiation (sections 7.3.6 and 7.4.4). However, some limitations are found. The pulsing capability used to induce charge on the device is known to be affected by ionising radiation consequently modifying the conversion factor of equation 7.1. Moreover, the value of 1.43 e^- per mV does not account for pixel to pixel variations of the metal-to-metal capacitor used to inject pulses. Although the dependence of the pulsing signal with TID can be measured on the monitoring pixels, calibration for specific capacitance values of every pixel is not possible to obtain. The methodology used to record pulsing waveforms on monitoring pixels is explained in section 7.2.3.

7.2.2 Noise rate and number of noisy pixels as a function of threshold

Although the noise level at a certain threshold is obtained by performing the scan described in section 7.2.1, information about the noise rate and its dependence with the applied threshold is still missing. The aim of this test is to characterise the overall noise of the chip and determine the minimum operable threshold that can be applied.

This test is conducted by waiting a fixed amount of time while enabling the data acquisition sequence of the chip. All possible sources of charge generation are removed from around the chip. Hence, all hits recorded in this time window are due to electronic noise exceeding the applied threshold. The threshold is then scanned over a range of values containing the average noise level of all the pixels on the chip. Figure 7.4 shows hit-map plots obtained using the technique described for two different thresholds: one above the average noise level and one below. At very high thresholds, electronic noise does not exceed the threshold and therefore no hits are observed. Once the threshold is lowered, random noise exceeds the threshold more often

and is seen on the readout.

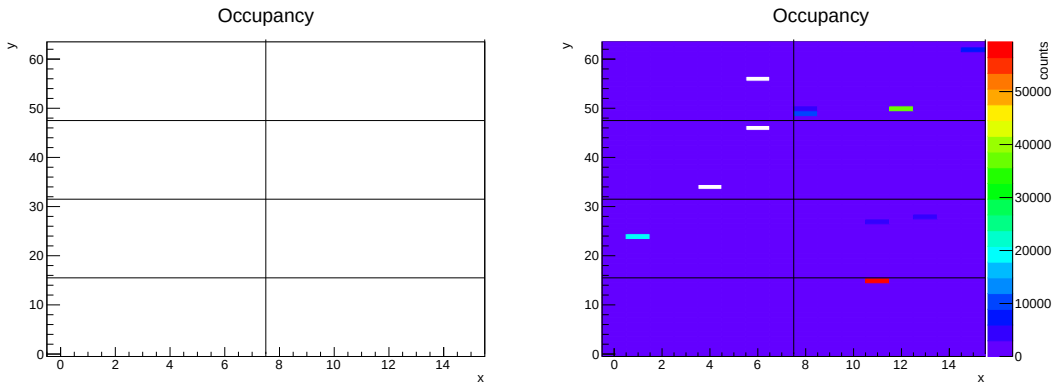

Figure 7.4: Hit-maps obtained on the mini-MALTA chip by waiting a fixed amount of time and recording noise hits. Left plot is at a high threshold while right plot is at a threshold close to the noise level.

The performance of the chip can be significantly affected if a large fraction of its pixels are noisy. Hence, the quantities computed are the number of noisy pixels and the noise rate as a function of threshold. The noise rate is calculated dividing the number of hits by the recording time. Examples of these two plots are shown in figure 7.5. The test was run by executing a defined IDB voltage range, the acquisition time and the number of repetitions to increase statistics and executing the corresponding PyROOT script.

7.2.3 Recording signals on monitoring pixels

The monitoring pixels can be accessed via dedicated probing pads present on the MALTA and mini-MALTA carrier boards. The pads allow signals to be applied to the input of the pre-amplifier and to measure the output of the pre-amplifier. Two types of signal were measured on the monitoring pads: test pulses using the internal circuitry and monochromatic X-rays from fluorescence. A schematic of the setup used in both cases is shown in figure 7.6.

Internal pulse injection

Internal pulse injection is used in different scans. Hence, it is important to measure their stability and evolution with irradiation. Characterisation of the pulsing signal is done by sending multiple pulses at a fixed V_{HIGH} and V_{LOW} voltage. After inducing a pulse, analogue signals can be measured by connecting a differential probe on the corresponding pads and reading their voltage output with an oscilloscope. Waveform information was stored using the functionalities of the scope. Analysis was carried out offline using PyROOT based scripts to extract the baseline and the amplitude of the pulse. The baseline is obtained from the average over a number of data points taken over a time range before the pulse starts. The amplitude is defined as the difference between the peak value of the waveform and the baseline level and is a measurement of the

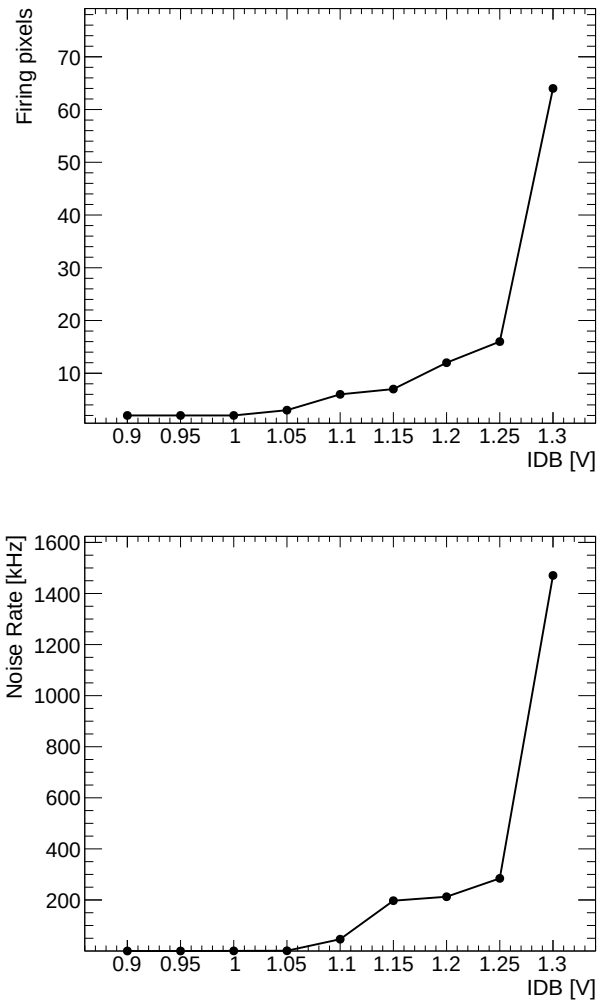

Figure 7.5: Number of pixels with > 0 noise hits as a function of threshold (top). Global noise rate as a function of threshold (bottom). The threshold level is inversely proportional to the applied IDB voltage.

energy deposited. For clarification, these parameters are sketched for an example waveform in figure 7.7.

It is important to monitor these parameters since a decrease in the baseline would lead to an effective increase of the threshold level and a drop in the amplitude would mean that the ability of the sensor to detect low energy pulses would be degraded. Measurements of the baseline are independent of the energy source used to induce the signal and variations with TID are purely due to the radiation damage induced on the transistors that control this voltage level. However, variations in the amplitude can be related to the pre-amplifier transistors or to TID effects on the pulse injection circuitry. To study the effect damage from ionising radiation requires a charge injection mechanism not affected by the TID damage, isolating this way the effects in the pre-amplifier. This is discussed below in section 7.2.3.

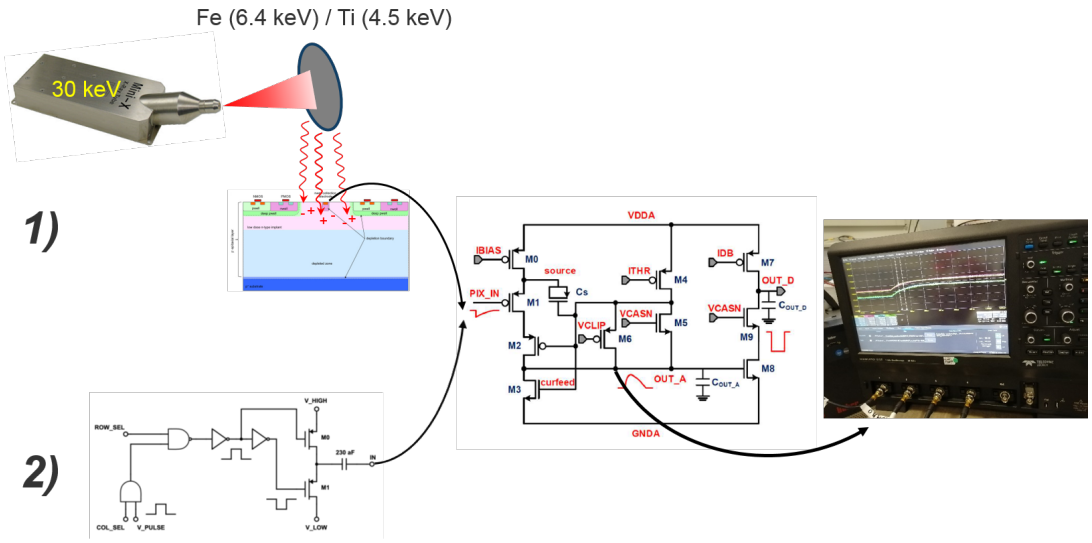

Figure 7.6: Sketch of the technique employed to readout analogue signals inducing charge with 1) X-ray fluorescence and 2) internal pulsing capability.

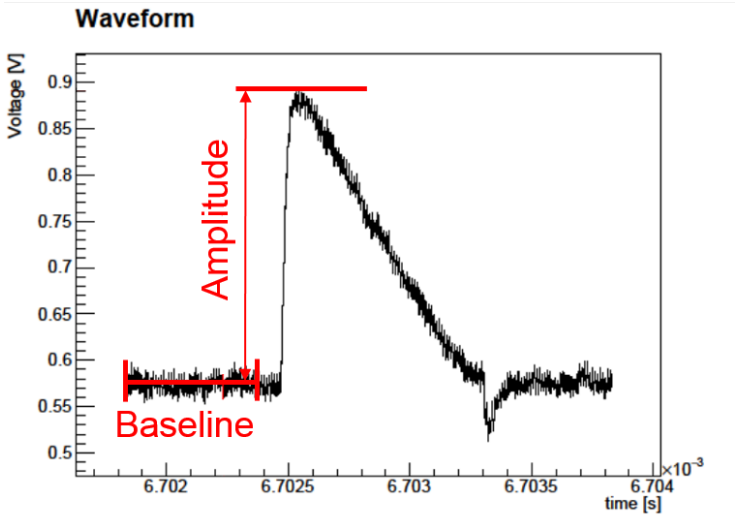

Figure 7.7: Waveform obtained by injecting a pulse through the pulsing capability on a monitoring pixel with access to the OUT_A node. Data was stored by the oscilloscope and transferred to a PC for offline analysis.

X-ray fluorescence signals

Fluorescence X-rays were induced by illuminating an Iron (Fe)/Titanium (Ti) circular plate with 30 keV X-rays generated with a mini-X-ray gun. The energy of the fluorescence X-rays is 6.4/4.5 keV and does not produce significant radiation damage in the sensor. Alignment between the X-ray gun, the fluorescence target and the DUT is critical to ensure a good hit rate while ensuring that the X-rays coming from the X-ray generator do not hit the detector. The waveforms are again recorded by an oscilloscope using a differential probe. However, due to the wide range of energies delivered by the X-rays, a large amount of signals ($\sim$ 10000) have to be recorded

in order to obtain a good quality spectrum. Each waveform data file takes up 663 kB of disk space, which would lead to 6.63 GB of data per energy spectrum. This huge amount of data is very difficult to manage and would take a large amount of processing time to store on the oscilloscope. Hence, the oscilloscope was set to only record the peak value of the waveforms in a histogram form. A measurement of the signal baseline is also done before every scan to calculate the amplitude of the waveform. An example of the histogram obtained from X-ray fluorescence with a Fe target in the MALTA chip is presented in figure 7.8. The histogram represents the energy distribution of the target material. This distribution has a peak at the most probable decay energy (6.4 keV for Fe and 4.5 keV for Ti). A Gaussian function is fit around the energy peak to extract its value in mV. Note that the number of entries is higher below the peak than above. Amplitude values below the peak are due to charge sharing when a particle hits the neighbouring pixels. Amplitude values above the peak correspond to charge sharing due to multiple hits on neighbouring pixels or hits on the measured pixel plus charge sharing, which is less likely to happen.

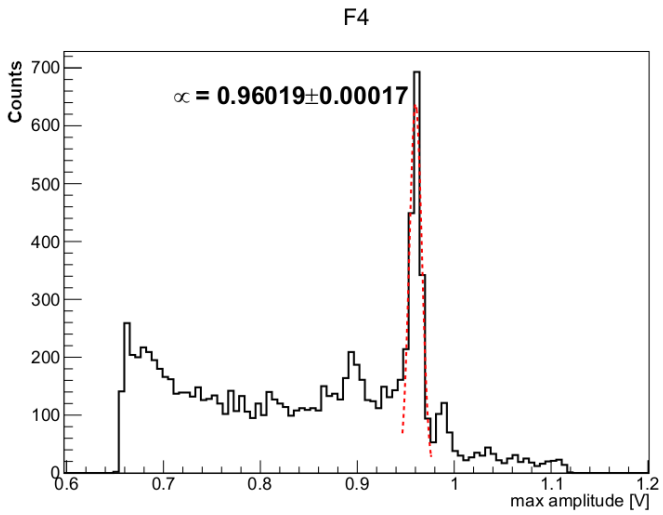

Figure 7.8: Example histogram obtained from the peak value of 10000 waveforms induced from X-ray fluorescence with a Fe target on the mini-MALTA chip. A Gaussian function was fit around the peak of the distribution.

X-ray fluorescence represents an independent source of energy that can be used to study the radiation damage induced on the pre-amplifier circuitry. Moreover, due to the known energy deposited on the device, it can be used to calibrate the response of the pixel with the charge generated on the device.

7.3 MALTA TID measurements

The MALTA chip (see section 4.3.1) has to withstand a TID dose of 80 MRad in order to fulfill the radiation hardness requirements for the upcoming upgrade of the ATLAS detector. In order

to understand the overall effects of TID damage on the MALTA chip, irradiation campaigns have been carried out at CERN (Switzerland) and at the University of Glasgow (UK). At CERN, sensors were irradiated at different TID doses. Chips were kept unpowered and no measurements were taken during the irradiation process. Post-irradiation studies showed poor performance of the chip at 5 MRad. This motivated the need for a deeper understanding of the damage effects on different parts of the front-end at finer TID dose steps.

The irradiation campaign carried out in Glasgow aimed to measure critical parameters of the chip with smaller dose steps. MALTA was irradiated using the X-ray facility of the University of Glasgow described in section 7.1. In order to simulate the operational conditions of the chip at the LHC, MALTA was irradiated powered and at low temperature. Low temperature also minimized annealing effects when testing the sensor. During the course of irradiation the different current domains of the chip were monitored (i.e. Analog, digital, DAC + LVDD and substrate + P-well) in order to observe the power consumption of the chip as a function of TID. A detailed time-line of the irradiation campaign is shown in figure 7.9. Every 0.125 MRad,

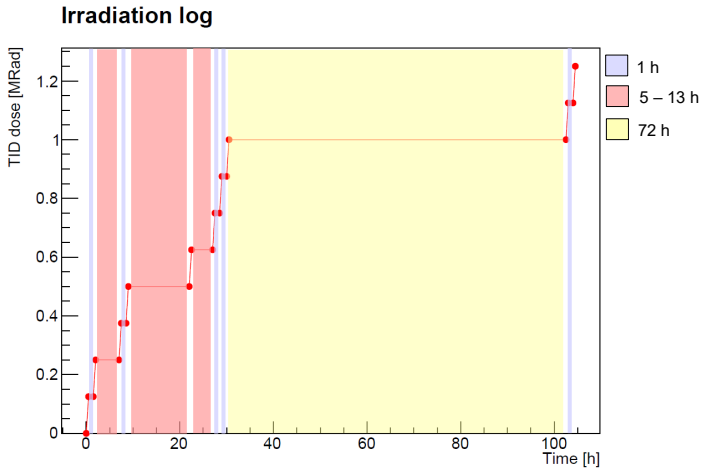

Figure 7.9: Plot summarising the MALTA irradiation log.

irradiation was stopped for $\sim$ 1 h to perform basic functionality tests. These included noise rate at different thresholds and analog response to the pulsing signal. Longer pauses were done at 0.25 MRad ($\sim$ 5 h), at 0.5 MRad ($\sim$ 13 h), at 0.625 MRad ($\sim$ 5 h) and at 1 MRad ($\sim$ 72 h). During these periods longer measurements were done, sometimes coinciding with the overnight break. The extra scans are threshold value per pixel at a fixed IDB voltage and pixel gain from fluorescence measurements. The results of all these measurements are presented in this section.

7.3.1 Description of the setup

The setup used to irradiate the MALTA chip is described in this section. This setup is designed to fulfill all the technical requirements needed to perform the measurements described in the

introduction.

In order to ensure constant TID dose rate delivered, MALTA has to be at a fixed position during irradiation and aligned to the center of the X-ray beam. Furthermore, the chip has to be kept at low temperature to minimize annealing effects when the sample is irradiated. A low humidity environment is also needed in order to prevent condensation, which would damage the sensor. Hence, the MALTA chip is firmly screwed onto optical stages mounted inside a box and aligned with the X-ray beam. Alignment of the DUT with the X-ray beam is done using a laser pointer, as shown in figure 7.10 (left). The box is equipped with a cooling system based on a circulation of silicone oil at low temperatures. The silicone oil cools down a metal plate which, in turn, is in contact with the MALTA PCB. The silicone oil can be cooled down to a minimum temperature of -30 ^{o}C, which translates to a chip temperature of 0 ^{o}C when the chip is powered. The low humidity environment is achieved by pumping dry air inside the box. The box is closed and sealed to prevent any leak of dry air and, hence, to maintain the environment with low humidity. Humidity values below 1% are achieved with this method. Temperature and humidity inside the polyester box are monitored by an AM2303 sensor. The temperature closer to the chip position is measured by a DS18B20 digital thermometer. Both sensors can measure temperature with a precision of $\pm$ 0.5 ^{o}C and ranges from $\sim$ - 40 to 125 ^{o}C.

The box used to contain the MALTA chip has different apertures for the power cables, the FMC connector to the Virtex 7 FPGA and the tubes for silicone oil circulation of the cooling system to come in from outside. A probe is also attached to the analog monitoring pixels pad and routed out of the box to a digital scope.

Radiation damage to other electrical components needed to operate MALTA and/or to perform the tests has to be prevented. Hence, the power supplies are placed outside the X-ray box and the cables are taken inside through a dedicated aperture. The DAQ PC is also placed outside the box and connected via ethernet to the FPGA board. The FMC cable that connects the MALTA PCB to the FPGA-based MALTA readout has a length of $\sim$ 20 cm and, hence, the Virtex7 FPGA has to be placed inside the X-ray box. Precautions are taken by placing the FPGA outside of the X-ray beam-line and shielding it with lead plates in order to avoid any reflected X-rays.

Six power supply channels are needed to operate MALTA: four Keithley 2410 and one dual TTi (see figure 7.10, right). Four channels are used to power up the different power domains of MALTA. Due to an issue with the slow control capability some DAC voltages had to be applied externally. Hence, one PSU channel was used to override the IDB voltage (to modify the in-pixel threshold) and one to override the VPULSE_LOW voltage (to control the low voltage level of the internal pulse). Table 7.1 summarizes the voltages applied by each power supply.

The DAQ PC executes different functions. Python-based scripts designed at CERN are used to load the chip configuration and send commands into MALTA. Scripts to perform the tests described in the following sections were also developed. The PSU are remotely controlled by

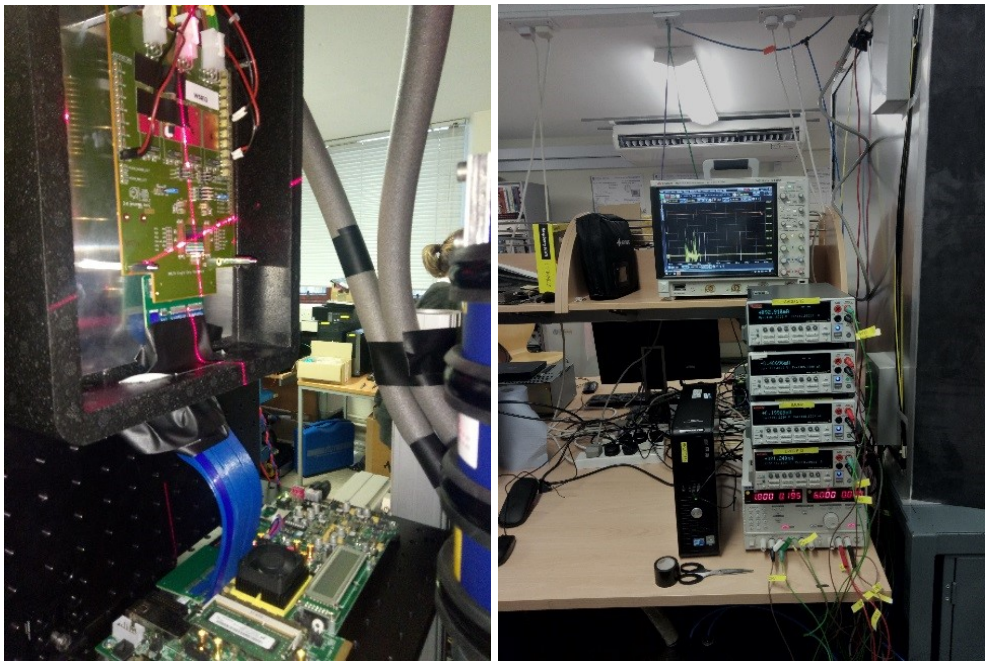

Figure 7.10: Picture of the polyester box that contains MALTA, placed inside the X-ray box and ready for irradiation (left). Picture of the components of the MALTA setup that are placed outside the box (right). These components are 4 Keithley 2410 SMU, 1 QL355TP dual TTI and the DAQ PC.

the PC. This allows automation of turning on/off procedures, scanning pixel threshold (IDB) and pulse amplitude (VPULSE_LOW) voltages and monitoring the current. Pixel hit information of MALTA is stored and analysed on the DAQ PC.

The MALTA chip is powered and configured during the irradiation process. Configuration is done such that noise and pixel activity is minimized. Due to an issue with the Slow Control of MALTA, communication with the chip via this capability is only possible with a small fraction of the pixel matrix. Hence, configuration is set by applying the voltages externally using the override mode. The IDB and VPULSE_lOW voltages are set using an external power supply, as they are modified during the scans. The rest of voltages are set using the variable resistance (trimmer) of the MALTA PCB.

7.3.2 Current evolution as a function of TID

The current consumptions of the MALTA analog (I_{AVDD}), digital (I_{DVDD}) and periphery circuits (I_{LVDD}) have been monitored during the X-ray irradiation process. Prior to X-ray irradiation, and after configuration of the chip, the current values were $I_{AVDD} = 0.0897A$, $I_{DVDD} = 0.0211A$ and $I_{LVDD} = 0.195A$. Pixel activity increases the digital and analog current consumption due to the constant charging and discharging of the collection electrode, with a signal that is transmitted along the two circuits. The activity generated by the X-rays increases these two currents to $I_{AVDD} = 0.1306A$ and $I_{DVDD} = 0.0332A$, while I_{LVDD} does not significantly change. These

PSU model	Voltage domain	Applied voltage (V)
Keithley 2410	IDB	0.6 - 1.2
Keithley 2410	VPULSE_LOW	1.05 - 1.45
Keithley 2410	Analog	1.8
Keithley 2410	Digital	1.8
TTi Ch. 1	PWELL + SUB	1.8
TTi Ch. 2	DAC + LVDS	1.8

Table 7.1: List of each power supply channel and the power delivered to each domain. Two Keithley 2410 were used to modify the IDB and VPULSE_LOW voltages during measurements and the rest to set the voltages of the different power domains of the chip.

results are comparable to the estimated power consumption ($P_{XVDD} = I_{XVDD} \times V_{XVDD}$) expected at the ATLAS ITk [88]. Note that the digital current consumption is much lower than the analog. This is a result of not transmitting a clock over the pixel matrix in this novel asynchronous readout solution.

The power consumption of the chip varies with the TID dose received, as shown in figure 7.11. No change is observed in I_{LVDD}. The analog and digital currents increase as the chip is

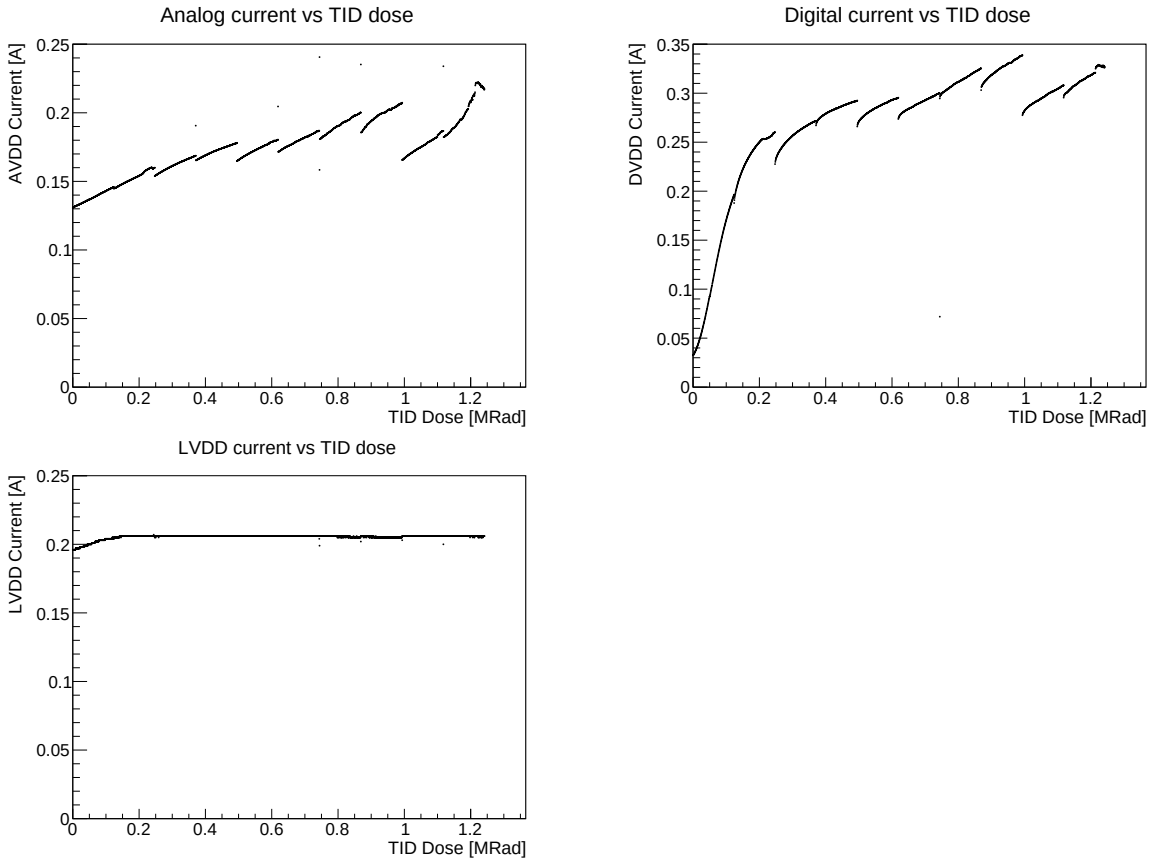

Figure 7.11: Evolution of the analog (top left), digital (top right) and periphery (bottom) currents as a function of TID. The current consumption shown in these plots corresponds to the periods where X-ray irradiation was being carried on. The drops in current consumption observed every 0.125 MRad are due to annealing effects that happened during the measurement periods.

irradiated. This is due to an increment in the number of noisy pixels, generating more pixel activity and, hence, current. The current drop every 0.125 MRad is a consequence of the annealing effects that occur during the measurement period. The increase is more prominent in the digital current at the first stages of the irradiation (up to 0.25 MRad).

The digital current reaches its compliance value (0.35 A) at 1 MRad. This current compliance is set to prevent permanent damage to the chip due to current overflow. The irradiation process is aimed to be done under the same conditions in order to obtain results that could be compared. This includes having similar annealing periods between measurements and setting a fixed configuration of the chip. Hence, leaving the chip to anneal for several days or change its configuration (e.g. to set an even higher threshold) in order to reduce the current would have led to inconsistent results. Thus, irradiation was stopped after a TID dose of 1.25 MRad.

7.3.3 Fluorescence measurements

The effects of radiation damage in the pre-amplifier circuit of MALTA were studied by measuring the analog pixel response in one of the monitoring pixels. The signal was induced with fluorescence X-rays from a Fe target, following the procedure described in section 7.2.3. Fluorescence scans were taken at 0, 0.25, 1 and 1.25 MRad. The peak of the obtained amplitude distribution (see figure 7.8) in every irradiation step is shown in figure 7.12.

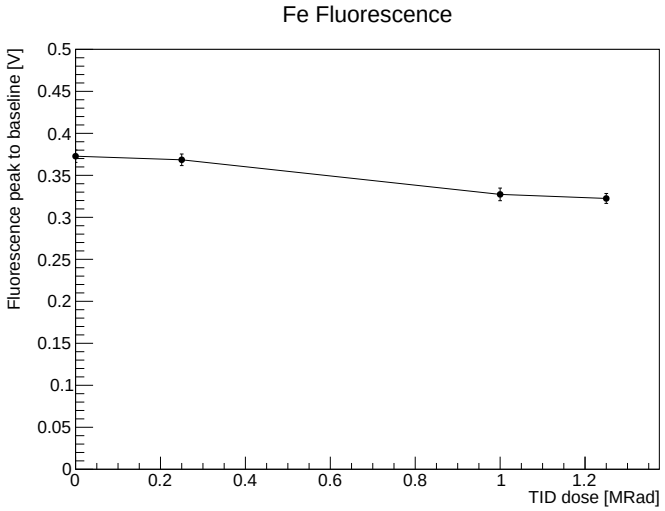

Figure 7.12: Amplitude of the signal obtained from X-ray fluorescence using a Fe target as a function of TID. The signal was obtained in one of the MALTA monitoring pads which are connected to the output of the front-end pre-amplifier.

The signal amplitude drops to $\sim 12\,\%$ of its initial value after 1 MRad of TID damage and up to $\sim 14\,\%$ after 1.25 MRad. The measurement at 1 MRad is taken after 65 h of annealing and should not be compared to measurements done with a very different annealing time. Since the source used to generate the signal is not affected by TID, the decrease in amplitude is linked to

a decrease of the front-end pre-amplifier gain. The first stage amplification circuit is similar to a common-source amplifier. Thus, its gain can be expressed as $A_{s1} = g_m(M1)r_{ds}(M3)$. Changes in the M1 transconductance or M3 output resistance produce a variation in the front-end gain. As explained in section 3.13 the transconductance of a transistor is reduced with increasing TID. Hence, a reduction of the output signal amplitude with TID is expected. A decrease in the pre-amplifier gain supposes a reduction of the SNR, leading to an increase in power consumption.

The characterisation of the threshold applied by the front-end circuit is done using the internal pulsing capability. Since this circuitry is also affected by TID, any deviations of the input charge for a fixed configuration have to be measured. Results are presented in section 7.3.4. The amplitude of the signal induced using internal pulsing is also measured at the output of the pre-amplifier. Thus, a drop by about 12 % at 1 MRad (+ 65 h of annealing) and by about 14 % at 1.25 MRad is anticipated if the pulsing circuitry is not affected by TID.

7.3.4 Study of the Internal Pulsing signal

The analog response to the test pulse signal was recorded every 0.125 MRad up to 1.25 MRad. The internal pulsing capability is used to calibrate the applied in-pixel threshold, since the charge injected by a single pulse can be calculated. Hence, it is important to study its evolution as a function of irradiation dose. The pixel analog response to internal pulsing is measured on one of the monitoring pixels. A differential probe is connected to the dedicated MALTA PCB pad and read out in an oscilloscope. Full waveform information is stored at each irradiation step for further offline analysis. Waveforms at some irradiation doses are shown in figure 7.13. From these waveforms, baseline level and peak to peak position is extracted. Plots are shown in figure 7.14. Although the baseline level remains constant, the amplitude decreases linearly by 30% at a TID of 1.25 MRad as compared to its pre-irradiation value. Deviations from the linear trend of this plot might be due to different annealing times between measurements. The highest annealing time happened at 1 MRad, of about 65 h. The amplitude at this dose is measured to decrease by ~ 15 %. This differs from the expected decrease of ~ 12 % at 1 MRad and of ~ 14 % at 1.25 MRad due to the decrease in the pre-amplifier gain observed in section 7.3.3. Hence, for a fixed $V_{HIGH} - V_{LOW}$ value the charge injected in the pixel decreases by ~ 3 % (the 15 % observed minus 12 % due to the drop in the front-end gain) at 1 MRad and by 16 % at 1.25 MRad. Note that the result obtained at 1 MRad is highly affected by annealing and has to be handled with care.

Threshold calibration is done assuming a constant charge injected for a given $V_{HIGH} - V_{LOW}$ (see equation 7.1). Thus, corrections have to be made when calculating the threshold at different TID doses. However, these results are obtained for only one pixel and, hence, can only be used qualitatively.

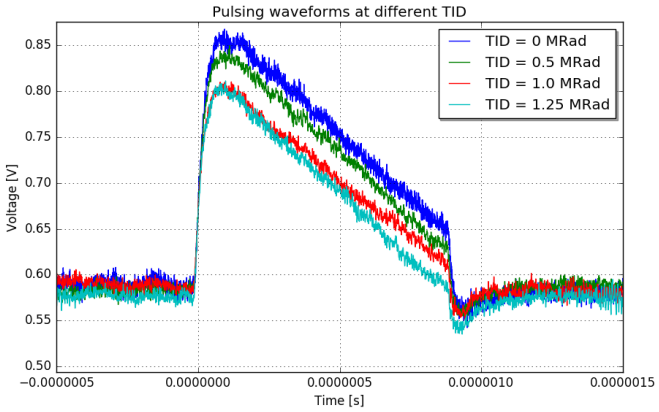

Figure 7.13: Analog response of monitoring pixel 0 to internal pulsing signal at different irradiation doses.

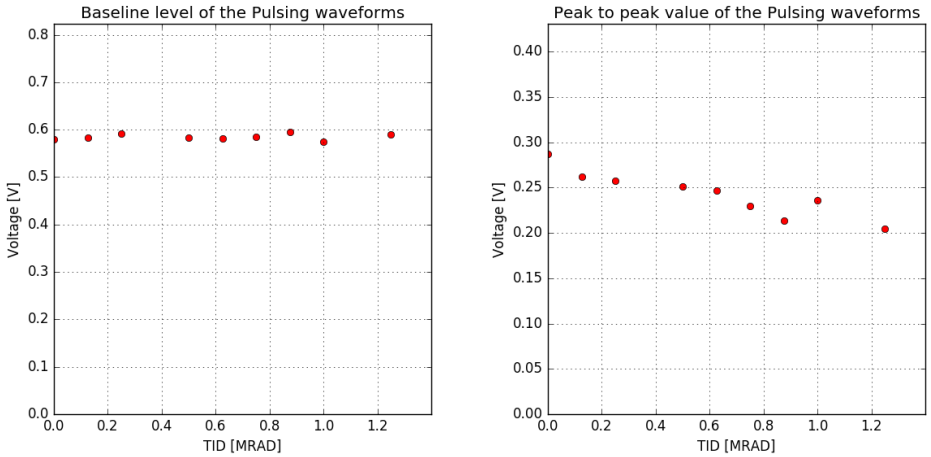

Figure 7.14: Baseline level (left) and peak to peak position (right) of pixel analog response to internal pulsing signal against TID dose.

7.3.5 Threshold calibration through internal pulsing

The threshold applied by the front-end discriminator for a given chip configuration can be measured using the procedure described in section 7.2.1. This scan requires to induce a signal in several pixels using the internal pulsing circuitry. Due to an issue with the SC communication of MALTA, pulses can only be sent to one pixel at a time. Hence, the scan duration escalates with the number of pixels measured. The X-ray irradiation capability has limited availability. In order to take maximum profit of the given irradiation time it was aimed to minimise any delay between irradiation steps while achieving good statistics. For this reason, only around 200 pixels are measured at each irradiation step. To further minimise the time between measurements, threshold calibration is only done at IDB = 1 V.

A histogram showing the distribution of measured threshold when applying 1 V to the IDB pad for an unirradiated chip is shown in figure 7.15. The different MALTA sectors are repre-

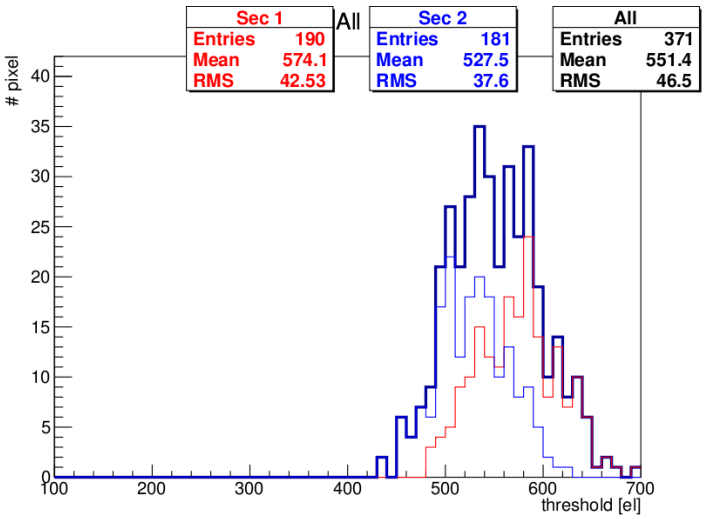

Figure 7.15: Histogram of the threshold applied in pixels of the MALTA sector 1 (red) and 2 (blue) when setting the IDB voltage at 1 V. The sum of both sectors is highlighted in black. This measurement is taken at a TID of 0 MRad.

sented in different colours: red binning for sector 1 and blue for sector 2. The sum of both is shown in black. A total of 371 pixels were measured: 190 pixels from the sector 1 and 181 from the sector 2 of MALTA. Measured thresholds were 574.1 and 527.5 electrons respectively. This is a relatively high threshold if comparing to the expected operation at 300 e^- at the HL-LHC. The threshold can be lowered by applying a higher IDB voltage. However, at a lower threshold some pixels become very noisy (see section 7.3.6), which can not be masked due to the issue with the propagation of SC commands over the pixel matrix.

Threshold calibration was done at 0, 0.5 and 1 MRad. The threshold mean as a function of TID is shown in figure 7.16. The applied threshold drops by about 10 % at 0.5 MRad and

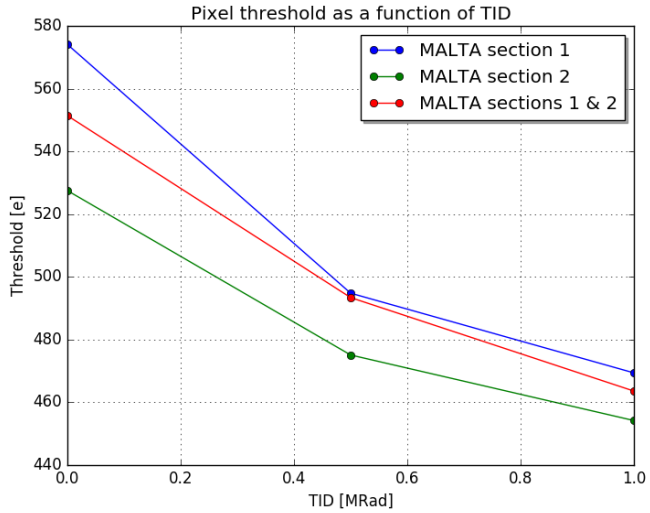

Figure 7.16: Mean value of the pixel threshold at different TID doses. This measurement is taken at an IDB voltage of 1 V.

by about 14 % at 1 MRad. In order to obtain the applied threshold in electrons, the threshold measured with the arbitrary V_{LOW} DAC value units is multiplied by the simulated 1.43 e^-/mV factor (see equation 7.1). This factor corresponds to the charge injected by the pulsing circuitry per mV of difference between V_{HIGH} and V_{LOW}. However, the results presented in section 7.3.4 show that, for a fixed V_{HIGH} - V_{LOW} value, the charge injected by the internal pulsing circuitry decreases with TID. This would shift the threshold measured at 0.5 and 1 MRad to even lower values. The actual correction can not be done since the gain using X-ray fluorescence is not measured at 0.5 MRad and the measurements obtained at 1 MRad have high annealing times.

This negative shift in the applied threshold contributes to an increase of noise at a given MALTA configuration. The noise is measured every 0.125 MRad at different IDB voltages. Results are shown in section 7.3.6.

7.3.6 Noise measurements as a function of threshold

The noise of transistors in the front-end circuit can increase as a function of TID (see section 3.13) if not properly mitigated. The noise of individual pixels can be measured by doing the scan described in section 7.2.1. Its mean value over few pixels ($\sim$ 300) at different TID doses is shown in figure 7.17 (top). Noise increases from about 9 e^- at 0 MRad to about 12 e^- at

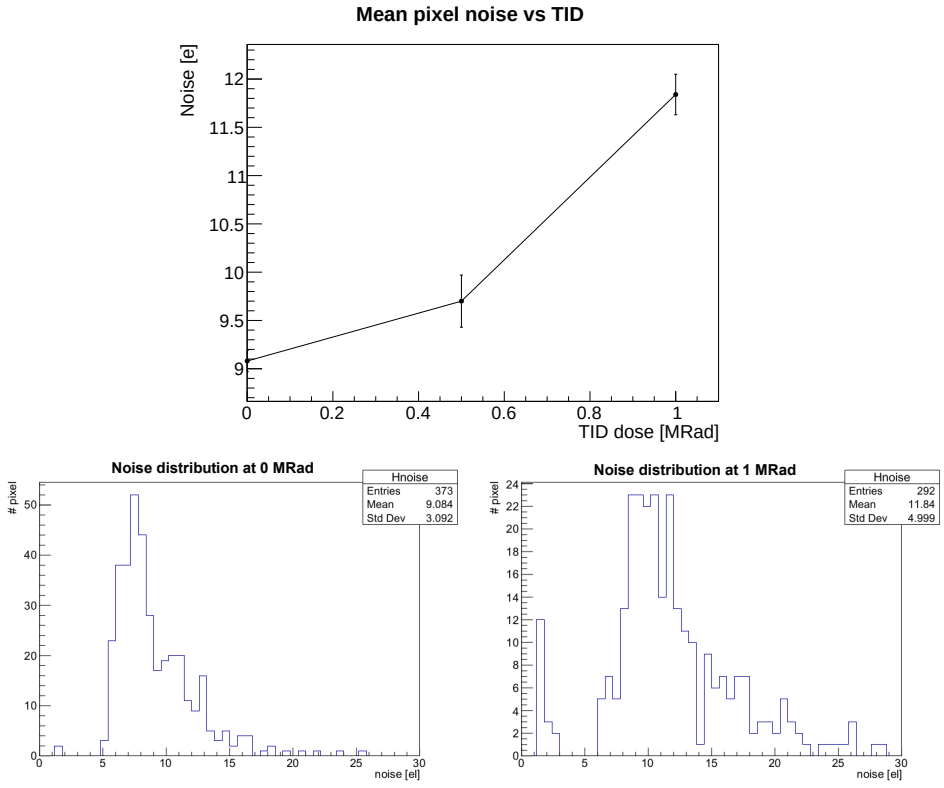

Figure 7.17: Mean value of the pixel noise of MALTA at different TID values (top). The specific pixel noise distribution at 0 (bottom left) and 1 MRad (bottom right) is also shown.

1 MRad. Although its mean value roughly matches the simulations presented in [20], the noise

distributions shown in figure 7.17 (bottom) do not have a Gaussian profile as would be expected. At 0 MRad, there are some pixels with high noise values that form a long tail in the noise distribution which is further increased at 1 MRad. This noise tail can be explained by the RTS noise generated by the M3 transistor. Voltage pulses in the gate of this transistor are amplified generating a significant noise in the OUT_A node. In turn, this generates a high noise hit rate at the output of the discriminator if the threshold is not set to a high enough value.

The noise rate has been measured at different thresholds every 0.125 MRad. The threshold level is adjusted by varying the IDB voltage. Threshold changes exponentially with IDB, becoming lower at higher voltages. Since the threshold calibration was done at IDB = 1 V, a range containing this value was scanned. Noise rate is calculated at each IDB voltage by waiting for a fixed period of time and counting the number of hits seen by MALTA. Then the number of hits is divided by the time period to obtain rate and this is plotted against IDB as seen in figure 7.18. The left plot corresponds to the noise hit rate at 0 MRad while in the right plot noise hit rate at

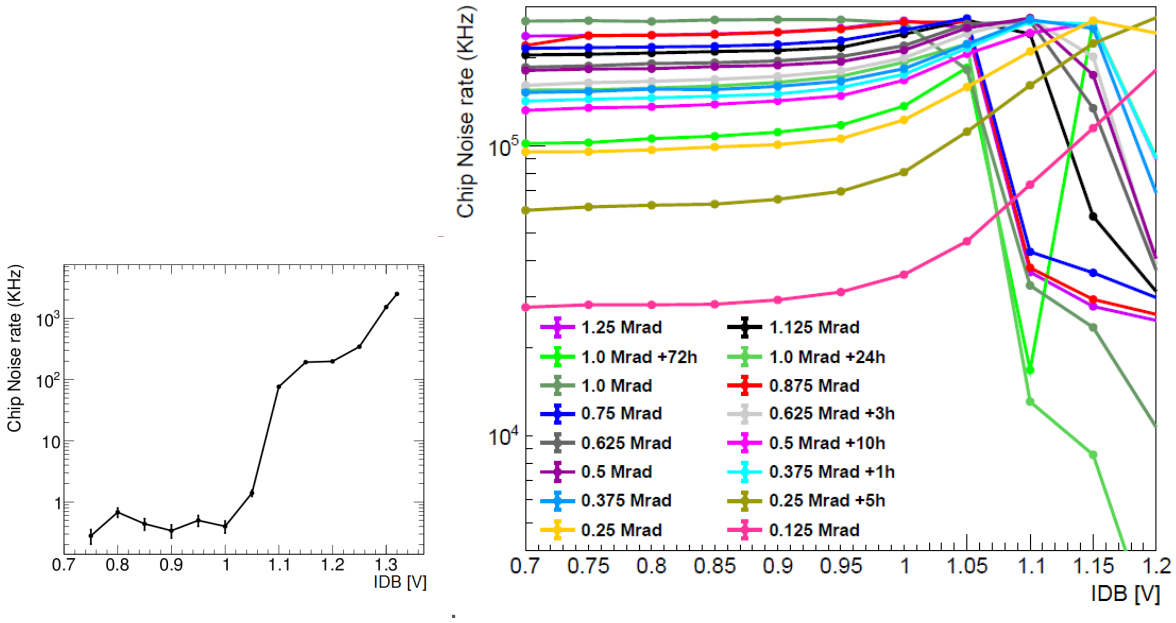

Figure 7.18: Noise rate as a function of IDB voltage. Left plot shows the noise rate at 0 MRad. Right plot shows noise rate at different irradiation doses and after chip annealing.

each irradiation step is shown. After long scans the noise measurement was repeated to see any effect from annealing. This is marked in the legend accordingly.

At 0 MRad the noise rate is effectively 0 kHz up to IDB = 1 V. At higher IDB values the noise rate increases, since the threshold reduces and gets closer to the pixel noise level. At the first irradiation step (0.125 MRad), the noise level increases to 30 MHz. Noise increases at higher TID doses until reaching a plateau at 250 MHz at a dose of 1 MRad. This plateau is given by a current overflow of the Digital domain and consequently a drop of the digital voltage. Operating at a lower DVDD voltage than the nominal is not considered, as the operating point

of transistors in the front-end would not be clear and irradiation is aimed to be done without changing the chip configuration. After annealing for 72 h at 1 MRad, the noise decreases close to its value at 0.25 MRad. This is also seen in the analog and digital current consumption (see figure 7.11). Again, reducing the current consumption by annealing the chip for long periods of time is not used as a strategy to continue irradiation to higher doses, as this would lead to inconclusive results.

The evolution of the number of noisy pixels and its noise rate with TID at IDB = 0.9 V is given in figure 7.19 top and bottom respectively. Only measurements done a short period of time after irradiating the chip are included in these plots. As noted before, the noise rate

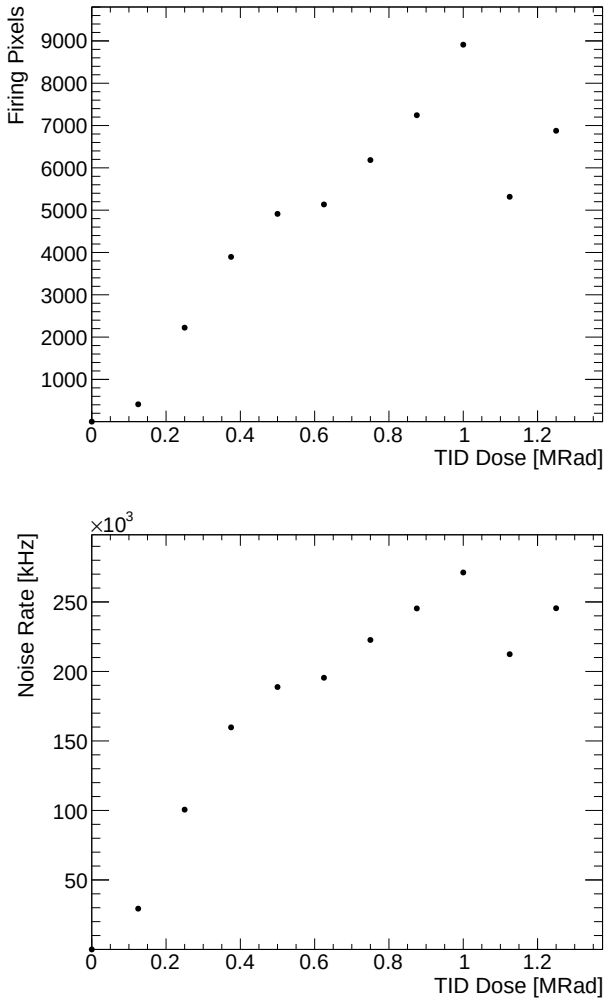

Figure 7.19: Number of noisy pixels (top) and the overall noise rate of the chip (bottom) as a function of TID at IDB = 0.9 V. The measurements presented in this plot are done with minimum annealing time.

increases to 30 MHz at 0.125 MRad and reaches 250 MHz at 1 MRad. The noise rate decreases after 1 MRad due to the 72 h period elapsed before restarting the irradiation, where annealing effects took place. The huge noise rate at 1 MRad is produced by approximately 9000 noisy

pixels. Considering that there are 262144 pixels in MALTA, this corresponds to 3 % of the total. Hence, a small fraction of noisy pixels produces an overflow of the digital current preventing the operation of the chip at a TID higher than 1 MRad without long annealing periods. The problematic pixels can not be masked due to an issue with the propagation of slow control commands through the pixel matrix.

7.3.7 Summary

The MALTA chip was submitted to TID irradiation using the X-ray irradiation facility of the University of Glasgow, aiming for a total dose of 80 MRad. Irradiation was done with the chip powered and configured and at a temperature of 0 °C. The X-rays were stopped every 0.125 MRad to carry out a characterisation of the chip. Currents of the analog and digital domains as well as of the periphery circuits were monitored during irradiation. Only the analog and digital currents increased with TID, while the periphery current stayed almost constant during the whole irradiation. The current of the digital domain hit its compliance value of 0.35 A at a dose of 1 MRad. Letting the chip anneal for 72 h helped to significantly reduce the current and irradiation was continued up to a dose of 1.25 MRad. The increase in power consumption was linked to huge increase in noise hit rate coming from a few pixels in the matrix. The noise hit rate and the number of noisy pixels were measured every 0.125 MRad. At 1 MRad, 3 % of the total number of pixels was responsible for a noise rate of 250 MHz at IDB = 0.9 V. A possible explanation for this huge noise in a few pixels of the matrix is that it is generated by the RTS noise of transistor M3.

The in-pixel threshold when applying an IDB voltage of 1 V was measured in pixels of two MALTA sectors at 0, 0.5 and 1 MRad. A decrease of around 14 % was observed after a dose of 1 MRad. This decrease in the threshold level contributes to increase the noise hit rate at the same MALTA configuration.

The technique employed to measure the in-pixel threshold uses the internal pulsing circuitry, which could be affected by TID. The amplitude of the signal generated from internal pulsing was measured every 0.125 MRad. The signal amplitude was measured to drop by ~ 17 % after a dose of 1 MRad. Hence, the measured threshold must be corrected accordingly.

A drop in the signal amplitude could be caused by either TID effects in the internal pulsing circuitry or in the pre-amplifier circuitry. In order to disentangle the two effects, fluorescence X-rays using a Fe target were used to generate a signal. Measurements were done at 0, 0.25, 1 and 1.25 MRad. An amplitude drop by 14 % after 1 MRad was observed in this case, meaning that the FE amplifier is affected by TID.

7.4 mini-MALTA TID measurements

The mini-MALTA chip has been presented in section 4.3.2. An improved version of the MALTA front-end is implemented in the left sectors of the mini-MALTA matrix. Sectors in the right are kept with the standard MALTA front-end to allow direct comparison. The RTS noise is expected to be significantly reduced both before and after TID irradiation in the sectors with the new front-end. As a side effect, enlarged transistors are expected to induce variations in the pre-amplifier gain, leading to a lower threshold when setting the same DAC voltages. For these reasons, threshold and analog noise are more frequently monitored than in the MALTA irradiation campaign. The analog response to internal pulsing and to X-ray fluorescence has also been measured to allow charge calibration and an absolute measure of the pixel gain.

In order to minimize annealing effects the irradiation campaign has been done as continuously as possible. Figure 7.20 shows the irradiation and measurement time-line from 0 to 10 MRad. From 0 to 2 MRad measurements are taken every 0.125 MRad and overnight breaks are taken every 0.5 MRad. From 2 to 10 MRad the measurement rate is increased to 1 MRad. In this situation, only one measurement is possible before the overnight break. After 10 MRad irradiation is only stopped to carry out the electrical functionality tests and for a longer time at 28 MRad to perform maintenance of the X-ray tube.

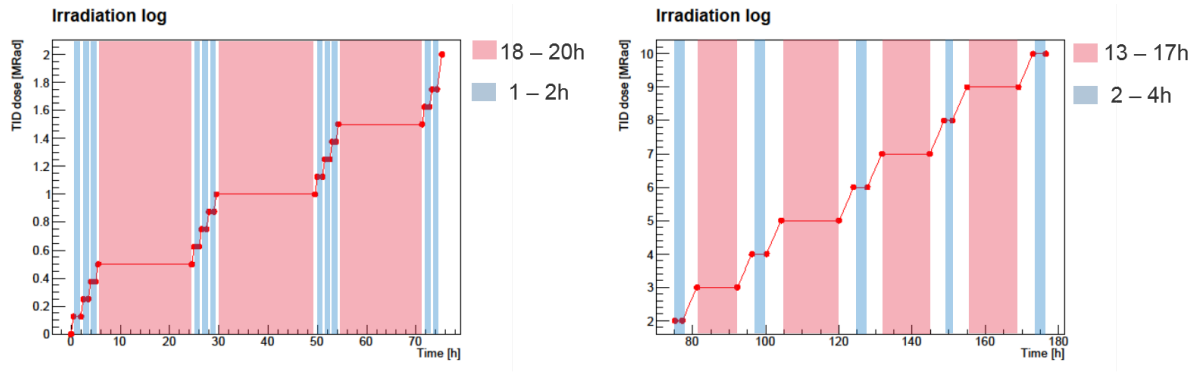

Figure 7.20: Plot summarising the mini-MALTA irradiation log from 0 to 10 MRads.

Since TID damage is expected to affect the front-end functionality rather than the charge collection efficiency, all results are presented comparing the sectors with standard MALTA front-end and the new version with enlarged transistors.

The setup employed to irradiate the mini-MALTA chip is analogous to that used in the MALTA irradiation campaign (see section 7.3.1). The only difference is that in the mini-MALTA setup all the configuration voltages are set and modified through the internal DAC circuitry, as this is fixed in this new iteration of the chip. Hence, only two TTi PSUs are used to provide power to the four voltage domains needed to operate the chip (AVDD + PVDD, DVDD, DAC and substrate). Moreover, the Virtex-7 FPGA is substituted with the Kintex-7 FPGA containing the mini-MALTA readout firmware described in section 6.5.

7.4.1 Study of the internal pulsing signal

The internal pulsing circuitry is capable of sending test pulses to the input of the front-end by applying a voltage difference over a capacitor. This voltage difference is controlled by the V_{LOW} and V_{HIGH} DAC parameters. The pulsing capability is expected to be affected by radiation damage and is a key tool used to perform the tests described in the following sections. Hence, calibration of the pulsing signal during the irradiation process is necessary. In every irradiation step, a differential probe is connected to two analog pixel monitoring pads, one from a sector with MALTA front-end and one from a sector with enlarged transistors front-end, to measure the analog response to test pulses. The V_{LOW} and V_{HIGH} voltage levels are set to 50 and 149 DAC units respectively. 128 waveforms are collected and stored using an oscilloscope to then extract the baseline level and waveform amplitude. More information about this test can be found in section 7.2.3.

The waveform amplitude as a function of TID is shown in figure 7.21. It is dominated

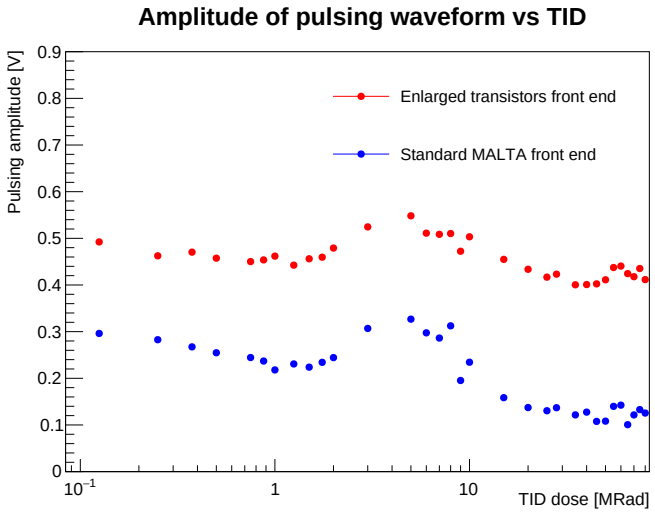

Figure 7.21: Signal amplitude, measured from the baseline to the waveform peak position, as a function of TID. Waveforms are induced by test pulses from the internal circuitry. The red dots correspond to one of the monitoring pixels with improved front-end while the blue ones correspond to a pixel with the non-modified MALTA front-end.

by the pre-amplifier gain, which is determined by the transistor M3, and by the input signal amplitude, given by the test pulse input charge. Since the transistor M3 has been enlarged in the mini-MALTA front-end to better tolerate TID damage, a difference of the gain on left and right mini-MALTA sectors is expected. Simulations show that the gain is increased by 30 % on pixels with enlarged transistors. At 0 Mrad, the gain of the pixel with enlarged transistors is found to be 1.66 times higher than that of the pixel with standard transistors. This result does not match the predicted increase from simulations [20] and is being discussed with the foundry. The behaviour of the waveform amplitude after TID irradiation is a convolution of the radiation damage effects in the internal pulsing circuitry and in the M3 transistor. Results show

that the waveform amplitude on pixels with improved front-end only drop to 93 % of its value after 80 MRad, while pixels with standard front-end drop to 63 % of its original value. Since the charge injection mechanism is the same in the two pixels, the difference in performance between the two pixels is correlated to the different size of the transistor M3. Hence, it can be concluded that the mini-MALTA front-end has a higher gain and is less affected by radiation damage in terms of pixel gain than the MALTA front-end, confirming the expectations. The increase on pulse amplitude at around 5 MRad is explained by a decrease of the baseline level.

The baseline level as a function of TID is shown in figure 7.22. The baseline level is defined

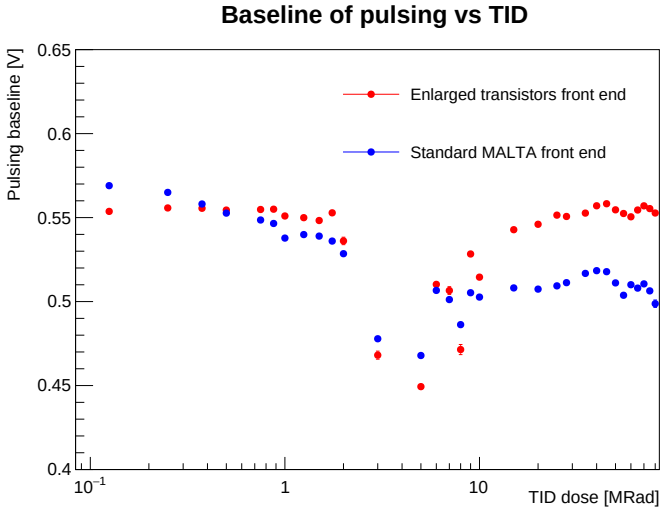

Figure 7.22: Comparison of the baseline level of the mini-MALTA sectors with standard (blue) and improved (red) front-end as a function of TID.

by the V_{GS} voltage of the VCASN transistor (M7). Although transistor M7 is identical on both front-ends, its V_{DS} voltage is set by the gate voltage of transistor M3. Since the output characteristics of a FET operating in the saturation region are not exactly flat, a variation on the V_{DS} voltage will lead to a slightly different V_{GS} on transistor M7 and, hence, to a different baseline. Because the enlarged M3 transistor is less affected by radiation damage, the baseline level will be less degraded on pixels with the mini-MALTA front-end. At 0 MRad, the difference between the baseline level of the two pixels is 0.04 V. Pixel to pixel variations could produce this small difference and, thus, it can not be linked to differences between the two front-ends. However, on the pixel with standard front-end, the baseline started to decrease as soon as the device was subject to irradiation. At 5 MRad the baseline drops sharply by $\sim$0.1 V to then recover to 90 % of its original value. The baseline of the pixel with enlarged transistors is approximately constant throughout the irradiation, except for a drop at 5 MRad, as is observed for the other pixel. This drop in the baseline level can be explained by interface traps cancelling the effects of oxide traps, as explained in section 3.13. An extra evidence is that this "bump" in the measured baseline matches the TID bump position observed in the NMOS transistors of the TJ 180 nm technology (see figure 3.18).

These results show that increasing the M3 transistor size leads to a higher gain and mitigates the damage due to ionising radiation. To disentangle the effects of the signal injection mechanism (which is also damaged by the X-rays) on the waveform amplitude, an independent charge injection mechanism is required. Results from X-ray fluorescence are presented in the next section. Baseline level degradation is also slightly improved by increasing the M3 transistor size. This is due to the fact that the output characteristics of transistor M3, which sets the M7 V_{DS} voltage, are less affected by TID for larger transistor size.

7.4.2 Measurement of gain from fluorescence

As it has been mentioned before, the gain of the pre-amplification phase is expected to increase as a consequence of enlarging the M3 transistor. The charge that is injected at the input of the front-end in the measurements of figure 7.21 is expected to vary with irradiation, since the pulse injection mechanism is also subject to irradiation. To have an absolute measurement of the pixel gain, pulses are injected using X-ray fluorescence. Concretely, a monochromatic X-ray beam from Titanium (Ti) fluorescence is used to induce a signal in two analog monitoring pixels: one with standard front-end and the other with enlarged transistors. The output pad of the analog monitoring pixels is connected to an oscilloscope via a differential probe. Around 10000 waveforms are recorded in each pixel and its amplitude distribution is measured. The peak of the distribution corresponds to the signal induced by the K-α decay of Ti, which has an energy of 4.8 eV. A schematic of this setup and a more detailed description of this test is presented in section 7.2.3. The amplitude distributions of figure 7.23 show that, at 0 MRad, the pixels with improved front-end have a 1.6 times gain increase with respect to the pixels with standard front-end. This corroborates the result obtained in the previous section, where charge was injected through the internal pulsing circuitry.

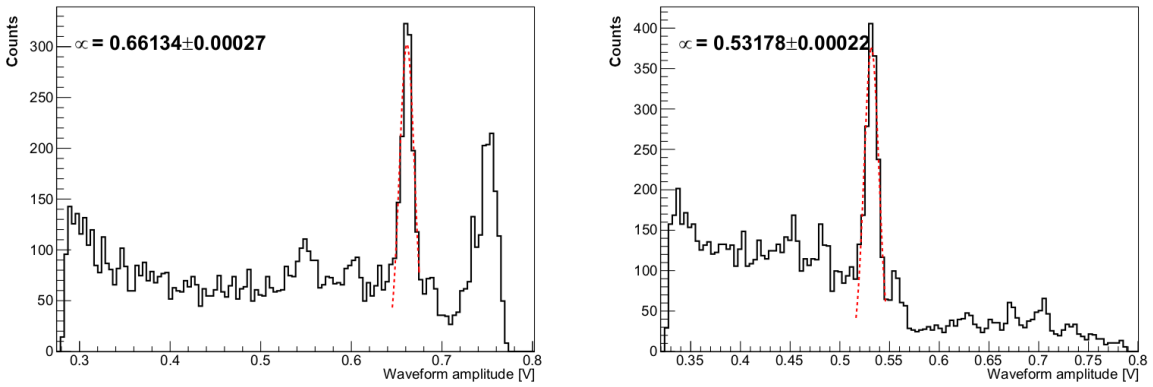

Figure 7.23: Distribution obtained when doing X-ray fluorescence with a Ti target and measuring the analog signal on a monitoring pixel with enlarged transistors (left) and standard (right) front-end. The second peak in the left-hand plot is due to the signals reaching the saturation of the source follower circuit.

The pixel gain is measured at different TID doses. Results are presented in figure 7.24. The evolution of the pixel gain with TID irradiation from X-ray fluorescence measurements is

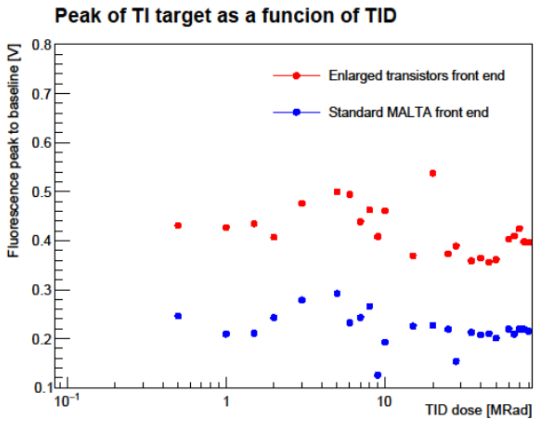

Figure 7.24: Amplitude of the signal obtained from X-ray fluorescence using a Ti target as a function of TID in two monitoring pixels of the mini-MALTA chip. The red dots correspond to a pixel with enlarged front-end while the blue dots correspond to a pixel with standard front-end.

very similar to the one in figure 7.21. The significant increase in amplitude at around 5 MRad followed by a further decrease on both mini-MALTA sectors is again linked to a drop of the baseline level. At 80 MRad, the signal amplitude on the pixel with standard front-end decreases to 73 % of its value at 0 MRad, while on the pixel with improved front-end it only drops to 86 % of its initial value. In the pixel with new front-end, the drop in amplitude when inducing waveforms through X-ray fluorescence is 6 % lower than when inducing them through internal pulsing. This difference might be explained by the fact that the pulses induced through internal pulsing were close to the saturation region of the pre-amplifier and, hence, the amplitude is not decreasing linearly. On the pixel with standard front-end a difference of 10 % is observed between the drop on amplitude when inducing the waveforms through internal pulsing and X-ray fluorescence. Since the X-ray fluorescence measurements were done systematically from one to two hours after the measurements with internal pulsing, this improvement might be due to annealing effects. Because enlarged transistors are more resistant to ionizing radiation, annealing effects are not that important in the pixel with the new front-end.

Results from internal pulsing and X-ray fluorescence have shown that the pixel gain is increased by a factor ~ 1.65 in pixels with the new front-end and that signal degradation is also improved. This increase in gain is expected to reduce the minimum achievable threshold. Results of the in-pixel threshold value as a function of TID are shown on the next section.

7.4.3 Threshold calibration

The threshold has been measured at different stages of the irradiation process. The threshold is measured by sending pulses of different amplitudes through the internal pulsing capability

at a fixed IDB current (150 dacs) as described in section 7.2.1. The threshold distribution of pixels with standard front-end and enlarged transistors front-end at 0 MRad is shown in figure 7.25. As it has been discussed in the introduction of this chapter, enlarging the M3 transistor leads to a higher gain of the pre-amplifier which, in turn, leads to a lower threshold value. Since the threshold dispersion decreases with the threshold mean, this value is also reduced in the sectors with the new front-end. One can observe that, by increasing the dimensions (W/L) of the M3 transistor from 1/0.18 μm to 1.22/0.38 μm, the mean value and the RMS of the threshold distribution is reduced by a factor of ~ 1.8. Again, this is in line with the increase in gain presented in the previous sections.

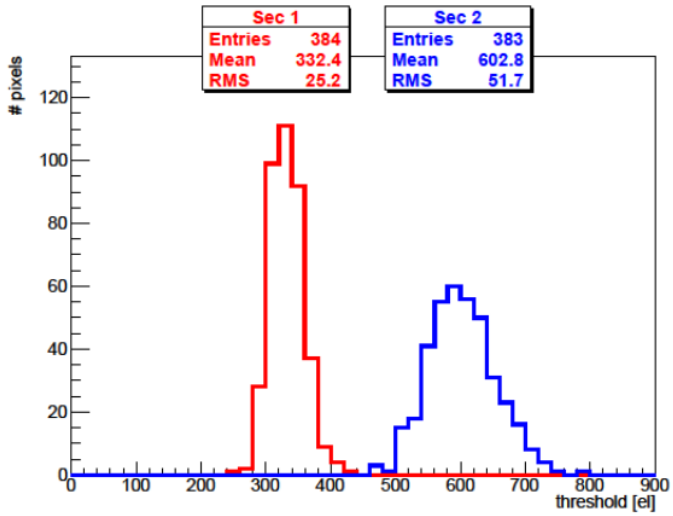

Figure 7.25: Threshold distribution of the mini-MALTA sectors with standard (blue) and enlarged transistors front-end (red) at 0 MRad.

The applied threshold and its pixel-to-pixel dispersion is expected to be affected by ionising radiation. The standard and the new front-ends have identical transistors setting the threshold level. Thereby, similar evolution is expected as the TID dose increases. The evolution of the threshold mean and RMS is monitored at different stages of the irradiation campaign. The mean of the threshold distribution as a function of TID dose is shown in figure 7.26. Effectively, a similar trend is observed in the two front-end versions. The mean value linearly decreases up to 1 MRad. At 60 MRad the threshold mean stabilises in both front-ends and drops by 25% with respect to its value at 0 MRad.

The threshold RMS is also reduced as a consequence of the threshold mean decrease. Results are shown in figure 7.27 (left). The RMS follows a similar trend as the threshold mean: a linear increase from 0 to ~ 2 MRad and it reaches a constant value at ~ 60 MRad. However, since this might be an effect of the absolute value of the threshold applied, it is more convenient to weight the RMS to its mean value. This result is shown in figure 7.27 (right). In this situation, the trend of the two curves is much more similar, only diverging slightly as the TID dose increases. At 80 MRad the RMS/Mean value of the improved front-end is 20% lower than for the sectors with

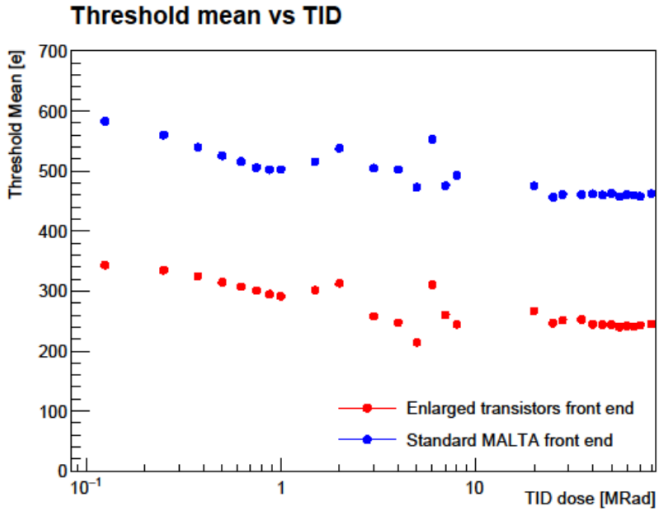

Figure 7.26: Mean value of the threshold distribution for the mini-MALTA sectors with standard (blue) and enlarged transistors front-end (red) from 0 to 80 MRad.

standard front-end.

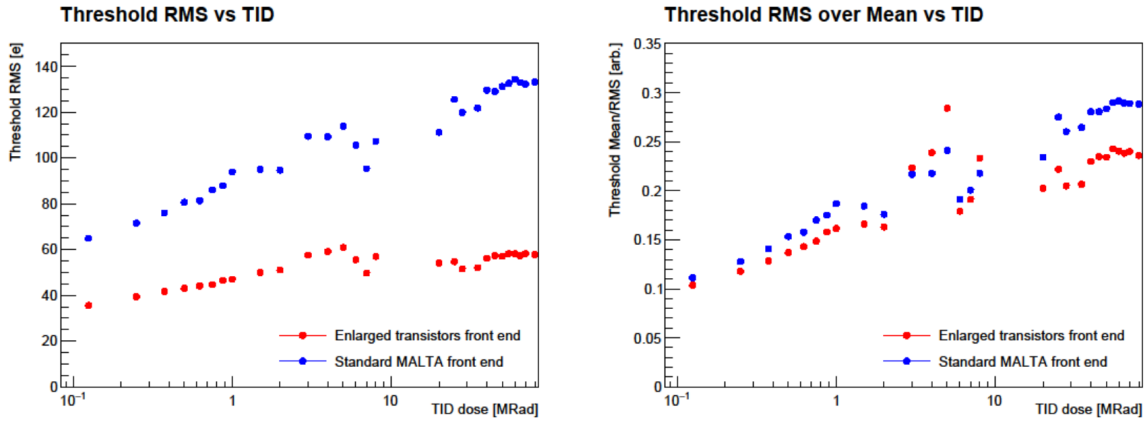

Figure 7.27: RMS of the threshold distribution (left) and RMS normalised to the threshold mean (right) for the mini-MALTA sectors with standard (blue) and enlarged transistors front-end (red) from 0 to 80 MRad.

7.4.4 Measurement of Noise through internal pulsing

The noise of a single pixel is obtained from the standard deviation of the transition curve from 0 to 100% pulses detected when sending pulses of different amplitudes through the internal pulsing circuitry (see section 7.2.1). The distribution of the pixel noise for all the pixels of sectors with standard transistors (blue) and with enlarged transistors (red) is shown in figure 7.28. Although the most probable value of the two distributions is very similar, the tail on the standard sector distribution significantly increases its RMS. The fact that this noise tail is not present on the sectors with the improved front-end confirms that the RTS noise is reduced by

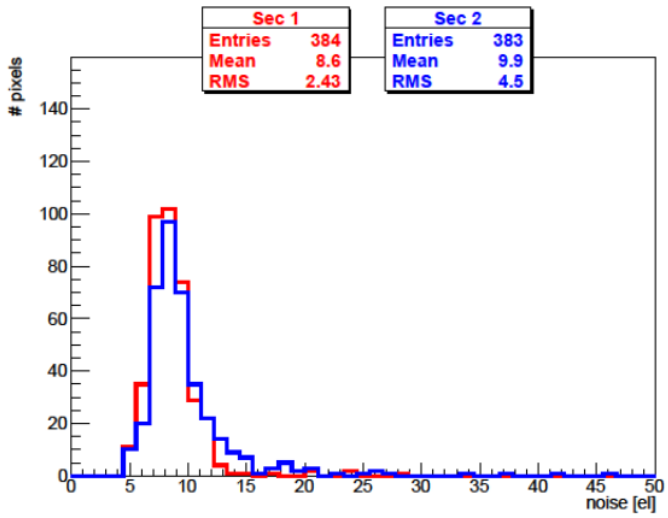

Figure 7.28: Noise distribution for sectors with standard (blue) and enlarged transistors (red) front-ends for a mini-MALTA chip at 0 MRad.

enlarging the M3 transistor, almost disappearing.

Recalling the results obtained in the MALTA chip, the RTS noise is expected to increase when irradiating the chip to a few hundreds of kRad. The evolution of the peak value and the number of pixels in the distribution tail (pixels with RTS noise) for both mini-MALTA sectors are shown in figure 7.29 (left) and (right) respectively. The MPV of the noise distribution is

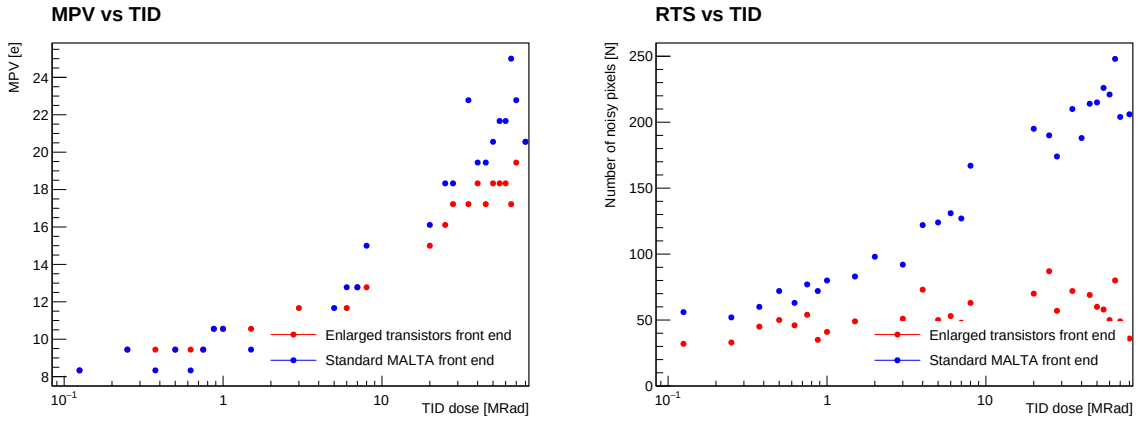

Figure 7.29: Most probable value (left) of the noise distribution and number of pixels with RTS noise (right) for the mini-MALTA sectors with standard (blue) and enlarged transistors front-end (red) from 0 to 80 MRad.

basically the same at 0 MRad for the two mini-MALTA sectors and evolves very similarly up to 80 MRad. The pixel noise stays roughly constant at 8 e^- up to 1 MRad and then starts to increase linearly up to a value of $\sim 20\,e^-$. The difference between sectors is more evident when looking at the number of pixels in the tail of the distribution. Already at 0 MRad the number of pixels with a noise value well above the MPV of the distribution is almost half in the sectors where the M3 transistor has been enlarged. After 1 MRad, the number of pixels with RTS noise starts to

increase linearly in the sectors with standard front-end while is maintained roughly constant on the sectors with enlarged transistors. This result proves that enlarging the M3 transistor not only reduces the RTS noise of unirradiated chips but also prevents a further increase after receiving TID damage.

7.5 Summary

The mini-MALTA chip has been irradiated to a TID of 80 MRad using the X-ray irradiation facility of the University of Glasgow. Similarly to the MALTA irradiation campaign, mini-MALTA was irradiated at 0° C with the chip powered and configured. The characterisation of the chip was done in regular time periods in order to have similar annealing between measurements. The performance of the MALTA front-end was compared to that of the new front-end with enlarged transistors. The investigated chip functionalities were the pixel gain, the effects in the internal pulse circuit, the in-pixel threshold applied and the pixel noise.

The gain of the front-end circuit was measured in two pixels with the different front-end implementations using fluorescence X-ray photons with a Ti target. Results showed that the gain is ~ 1.6 times higher in the pixel with the new front-end. After 80 MRad of TID, the gain of the mini-MALTA front-end decreases by 14 % while in the pixel with the MALTA front-end it drops by 27 %. Hence, by enlarging the M3 transistor the pixel gain is increased and less degradation is observed after irradiation.

The charge injection mechanism has also been tested in the two pixels mentioned above. Since this circuit is identical in both pixels, the differences observed are associated to changes in the pixel gain. The results corroborate the difference in gain by a factor of 1.6 between the two front-ends at 0 MRad. After 80 MRad the signal decreases by 7 % in the mini-MALTA front-end and by 37 % in the MALTA front-end. A flat difference between these measurements and the measurements using X-ray fluorescence was expected, caused by the damage in the internal pulse circuitry. Although this difference is not the same in the two front-ends, deviations might be due to different annealing effects for the two pixels.

The increased gain of the mini-MALTA front-end leads to an effective decrease of the in-pixel threshold, as for the same input charge the signals are higher. The in-pixel threshold was measured in the mini-MALTA sectors with the modified and the standard MALTA front-end. For a given configuration, the sectors with the new front-end show a 1.8 times lower threshold than the sectors with the MALTA front-end. This is in line with the measured increase in gain. The threshold dispersion over all sectors is also reduced as a consequence of the decrease in the threshold mean. The threshold mean in both front-end versions follows a similar trend after TID irradiation. However, its RMS is much less affected in the sector with improved front-end.

The RTS noise contribution of transistor M3 is supposed to be reduced in the improved front-end as a consequence of enlarging it. The RTS noise is only observed in some pixels of

the matrix. The noise was measured in the mini-MALTA sectors with improved and standard front-ends. Already at 0 MRad, the RTS tail of the noise distribution is reduced by a factor of ~ 2. When irradiating the chip, the MPV value of the noise increases equally in the two sectors. However, the tail of the distribution, associated to RTS noise, is roughly constant in the sectors with the new front-end while it linearly increases in the sectors with the MALTA front-end.

Chapter 8

Conclusion

The TowerJazz Investigator chip with process modification was tested within the context of this thesis. The goal of these measurements was to determine the optimal pixel parameters for the design of demonstrators. The chip consists of 134 blocks of pixels with different characteristics designed to optimize the sensor properties for better pixel gain. Several matrices were characterised allowing a comparison between pixels with different electrode sizes, different distances between the electrode and the n-well implant and different deep p-well horizontal distances. Charge was induced using X-ray fluorescence photons and the analog response was recorded using an oscilloscope. Measurement results have shown that these geometrical parameters have an impact on the detector gain. The detector gain is known to have a dependence in the capacitance of the collection electrode, being larger for smaller capacitance values. The sensor capacitance decreases with smaller electrodes and larger distances to the neighbouring electronics. As expected, pixels with smaller electrode size and maximum opening in the P-well implant present the maximum gain. The size of the deep P-well also has an impact in the pixel gain and the charge sharing. Measurements have shown higher charge sharing in pixels with maximum P-well horizontal distance. However, its gain is 15 % larger due to its effect on the pixel capacitance.

Results from the Investigator1 chip were taken into account to design a full scale ATLAS ITk demonstrator named MALTA. The MALTA sensor was produced in the modified TowerJazz 180 nm process and contains a matrix of 512×512 pixels, each pixel having an area of $36.4\,\mu\text{m}^2$. Each pixel is equipped with a low-power and low-noise analog front end that amplifies and discriminates signals from particle hits. The readout architecture is a novel asynchronous approach in which no clock is propagated over the pixel matrix, significantly reducing the power consumption of the digital circuit. This novel readout scheme requires an off-chip readout to synchronise the digital hits and store the data into an external memory system. The solution was implemented in a Xilinx Virtex7 FPGA board by using asynchronous oversampling. Measurements have proven that the novel asynchronous readout implemented in the MALTA chip is viable and minimises the power consumption of the chip.

First measurements on the MALTA sensors also demonstrated the functionality of the analog and digital circuits. However, high levels of RTS noise linked to a transistor of the analog front-end circuit were observed. The pixel-to-pixel threshold variation was found to be higher than expected. These two facts prevented operation of the chip at low threshold values. X-ray irradiation studies made at the University of Glasgow have shown an increase of pixels with RTS noise and an increase in the pixel-to-pixel threshold spread at doses as low as 1 MRad. These results were complemented with beam tests in Bonn and CERN in which unirradiated and irradiated sensors were tested. Chips irradiated up to $10^{15} n_{eq}/cm^2$ were inefficient near the pixel edges. This was related to a lack of electric field towards the collecting electrode at this region and the trapping of the slowly moving charge carriers.

A second generation of chips was designed to correct the RTS issue, the high threshold variation and the efficiency loss at the pixel boundaries. The new chip was a smaller version of MALTA, named mini-MALTA, with increased size of critical transistors to correct for the RTS and threshold variation issues and two different solutions to enhance the lateral electric field. The novel asynchronous propagation of signals was maintained but an additional digital synchronizer was implemented at the periphary of the chip. To allow a direct comparison between the MALTA chip structure and the new implementations, mini-MALTA is divided into 8 substructures including all possible combinations between the standard MALTA chip and the different improvements.

An exhaustive TID irradiation campaign was carried out at the University of Glasgow comparing the evolution of the standard MALTA and the new mini-MALTA sectors of the chip. The total dose achieved was 80 MRad, which is the expected dose received by the sensors at the outermost layers of the ATLAS ITk during the lifetime operation at the LHC. Different parameters of the pixel performance were studied: the pixel gain, the threshold mean and its pixel-to-pixel spread and the RTS noise. The gain of the mini-MALTA sectors is 1.65 times higher than in the MALTA side. This is an effect of increasing the size of the M3 transistor, which was implemented to address the RTS noise observed in the MALTA chip. The threshold mean has been reduced due to the increase in gain. More importantly, the increase in the pixel-to-pixel threshold with increasing ionising dose has been significantly reduced, being roughly constant during the whole irradiation, while in the MALTA sectors it doubled its value. The new mini-MALTA sectors also show an 80% reduction in the number of pixels with RTS noise. While the mean value of the in-pixel noise is roughly the same and increases with the TID dose in both sectors, the outliers in the distribution linked to pixels with RTS noise only increase in the MALTA sector.

This work has proved the viability of Monolithic CMOS sensors in high radiation environments. Sensors designed in the modified TowerJazz 180 nm technology are radiation hard enough to resist 80 MRad of TID damage and $10^{15} n_{eq}/cm^2$ of NIEL damage. A next generation of monolithic devices is being developed to include in-pixel threshold tuning with the aim of

further reducing the operational threshold of the chip.

Bibliography

[1] The CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B 716 30, (2012). arXiv:1207.7235.

[2] The Atlas Collaboration, *Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC*, Phys.Lett. B716 pages 1-29, (2012). arXiv:1207.7214.

[3] F. Englert and R. Brout. *Broken Symmetry and the Mass of Gauge Vector Mesons*, Physical Review Letters, 13:321–323, 1964.

[4] P.W. Higgs. *Broken symmetries, massless particles and gauge fields*, Physics Letters, 12(2):132–133, 1964.

[5] P. W. Higgs. *Broken Symmetries and the Masses of Gauge Bosons*, Physical Review Letters, 13:508–509, 1964.

[6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. *Global Conservation Laws and Massless Particles*. Phys. Rev. Lett., 13:585–587, 1964.

[7] L. Evans and P. Bryant, *LHC Machine*, JINST, 3 (2008), S08001.

[8] CERN, "CERN public webpages", http://cdsweb.cern.ch/.

[9] G. Apollinari, I. Béjar Alonso et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN Yellow Reports: Monographs. CERN, Geneva, (2017). doi:10.23731/CYRM-2017-004.

[10] M Bajko et al. *Report of the Task Force on the Incident of 19th September 2008 at the LHC*, Technical Report LHC-PROJECT-Report-1168. CERN-LHC-PROJECT-Report-1168, CERN, Geneva, (March 2009). https://cds.cern.ch/record/1168025/.

[11] The ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation, 3(08):S08003, (2008).

[12] ATLAS experiment at CERN, Oct. 2017. http://atlas.cern.

[13] Jackson JN. *The ATLAS semiconductor tracker (SCT)*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. April 2005;541(1):89-95.

[14] Vogel A. *ATLAS Transition Radiation Tracker (TRT): Straw tube gaseous detectors at high rates*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. December 2013;732:277-80.

[15] I. Peric et al. *The FE-I3 readout chip for the ATLAS Pixel Detector*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 565(1):178-187, (September 2006).

[16] M. Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*, ATLAS TDR 19, CERN/LHCC 2010-013, CERN, Geneva, (2010). https://cds.cern.ch/record/1291633.

[17] 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):S155-S159, April 2011.

[18] H Wilkens and the ATLAS LArg Collaboration. *The ATLAS liquid argon calorimeter: An overview*, Journal of Physics: Conference Series, 160(1):012043, 2009.

[19] The ATLAS Collaboration, *Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment*, Technical Report CERN-LHCC-2012-022. LHCC-I-023, CERN, Geneva, (Dec 2012).

[20] I. Berdalovic, *Design of radiation-hard CMOS sensors for particle physics applications*, 2019 (Zagreb), doctoral thesis.

[21] G. Charpak. *Development of multiwire proportional chambers*, CERN Courier, 6:174–176, 1969.

[22] J.B.A. England, B.D. Hyams, L. Hubbeling, Joseph C. Vermeulen, and P. Weilhammer. *A silicon strip detector with 12 μm resolution*, Nucl. Inst. and Meth., 196:149–151, 1986.

[23] L. Rossi, P. Fischer, T. Rohe, and N. Wermes. *Pixel Detectors: From Fundamentals to Applications*, Springer, Berlin, Heidelberg, 2006.

[24] Tanabashi M. et al. *Review of Particle Physics*, Physical Review D. 2018;98(3).

[25] Monroe J, Wicander R. *The Changing Earth: Exploring Geology and Evolution*, Cengage Learning; 2008. 753 p.

[26] Shimura, F. (2014). *Semiconductor Silicon Crystal Technology*, Saint Louis: Elsevier Science.

[27] Shigeji Fujita, K. (2010). *Quantum Theory of Conducting Matter*, Springer New York.

[28] R. De L. Kronig, Penney W. *Quantum Mechanics of Electrons in Crystal Lattices*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 1931;130(814):499-513.

[29] O'Donnell K, Chen X. *Temperature dependence of semiconductor band gaps*, Applied Physics Letters. 1991;58(25):2924-2926.

[30] H. Spieler, *Semiconductor detector systems*, 978-0198527848: Oxford science publication, 2014.

[31] R. C. Alig, S. Bloom, *Electron-Hole-Pair Creation Energies in Semiconductors*, Phys. Rev. Lett. 35, 1522 – Published 1 December 1975

[32] Simon M. Sze, Kwok K. Ng *Physics of Semiconductor Devices*, 3rd ed. Chichester: John Wiley and Sons; 2006.

[33] M. Rudan, *Physics of semiconductor devices*, 2nd ed. Springer, 2014.

[34] Pierret R., *Semiconductor Device Fundamentals*, 2nd ed. Addison-Wesley, 1996.

[35] Neamen, D. A., *Semiconductor physics and devices: basic principles*, 4th ed. McGraw-Hill, 2011.

[36] p–n junction, available at: https://en.wikipedia.org/wiki/P-n_junction

[37] Sze S. M. *Semiconductor Devices: Physics and Technology*, 2nd Edition, Wiley 2002.

[38] J. R. Schwank. *Basic mechanisms of radiation effects in the natural space environment*, Notes of the short course of the 1994 IEEE Nuclear and Space Radiation Effects Conference, Tucson (Arizona), 1994, Section II, pp. 1-109.

[39] L.M.P. Fernandes, *Characterization of large area avalanche photodiodes in X-ray and VUV-light detection*, Journal of Instrumentation, August 2007. doi: 10.1088/1748-0221/2/08/P08005.

[40] W. Shockley, *Currents to conductors induced by a moving point charge*, Journal of Applied Physics, 9:635-636, 1938.

[41] S. Ramo, *Currents induced by electron motion*, Proceedings of The Institute of Radio Engineers, 27:584-585, 1939.

[42] Lozano M, Cabruja E, Collado A, Santander J, Ullán M. *Bump bonding of pixel systems*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 2001 Nov 1;473(1):95–101.

[43] Agajanian, A. H., *MOSFET technologies*, Springer US, (1980).

[44] Rivetti A. *CMOS: Front-End Electronics for Radiation Sensors*, Taylor & Francis, (2015).

[45] Estreich DB, Ochoa A, Dutton RW. *An analysis of latch-up prevention in CMOS IC's using an epitaxial-buried layer process*, In: 1978 International Electron Devices Meeting. 1978. p. 230–4.

[46] 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.

[47] Baker, R. J., *CMOS: circuit design, layout, and simulation*, 3th ed., ser. IEEE Press series on microelectronic systems. Newark, NJ: John Wiley and Sons, 2019.

[48] Jindal RP., *Compact Noise Models for MOSFETs*, IEEE Transactions on Electron Devices, 2006.

[49] Hooge FN., *1/f noise sources*, IEEE Transactions on Electron Devices, vol. 41, pp. 1926-1935, (1994). doi:10.1109/16.333808.

[50] S. Christensson, I. Lundstrom, and C. Svensson. *Low frequency noise in MOS transistors - I Theory*, Solid-state Electronics, 11:797–812, 1968.

[51] Z. Shi, J. P. Mieville, and M. Dutoit, *Random telegraph signals in deep submicron n-MOSFETs*, IEEE Trans. Electron Devices, vol. 41, no. 7, pp. 1161–1168, (Jul. 1994).

[52] Muto H, Tsai M-H, Ma T-P. *Random telegraph signals in small MOSFETs after X-ray irradiation*, IEEE Transactions on Nuclear Science. 1991;38(6):1116-23.

[53] Chao CY-P, Wu TM-H, Yeh S-F, Chou K-Y, Tu H, Lee C-L, et al. *Random Telegraph Noises in CMOS Image Sensors Caused by Variable Gate-Induced Sense Node Leakage Due to X-Ray Irradiation*, IEEE Journal of the Electron Devices Society. 2019;7:227-38.

[54] 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.

[55] Schwank, James R., et al. *Radiation Effects in MOS Oxides*, IEEE Trans. Nucl. Sci. 55.4 (2008): 1833-1853.

[56] Faccio F., Cervelli G. *Radiation-induced edge effects in deep submicron CMOS transistors*, IEEE Transactions on Nuclear Science, Vol. 52, No. 6, December 2005, pp. 2413-2420.

[57] Snoeys, W. et al., *Integrated circuits for particle physics experiments*, IEEE Journal of Solid-State Circuits, Vol. 35, No. 12, December 2000, pp. 2018-2030.

[58] W. Snoeys et al. *Layout techniques to enhance the radiation tolerance of standard CMOS technologies demonstrated on a pixel detector readout chip*, Nucl. Instr. Meth., A439:349–360, 2000.

[59] Hillemanns H. et al. *Radiation hardness and detector performance of new 180 nm CMOS MAPS prototype test structures developed for the upgrade of the ALICE Inner Tracking System*, IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC), October 2013, pp. 1-5.

[60] Aglieri, G. et al., *Monolithic active pixel sensor development for the upgrade of the ALICE inner tracking system*, JINST, Vol. 8, No. 12, December 2013, pp. C12041.

[61] Soleys, W. et al., *A process modification for CMOS monolithic active pixel sensors for enhanced depletion, timing performance and radiation tolerance*, Nuclear Inst. and Methods in Physics Research, A 871 (2017) 90–96.

[62] Riegel C., *Performance tests of depleted CMOS sensors for application at HL-LHC*, 2018 (Bergische Universität Wuppertal), doctoral thesis.

[63] H. Pernegger et al *First tests of a novel radiation hard CMOS sensor process for Depleted Monolithic Active Pixel Sensors*, 2017 JINST 12 P06008.

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

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

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

[67] *Miniature X-ray tube specification manual*, www.amptek.com/-/media/ametekamptek/documents/resources/mini-x-specifications.pdf?la=en.

[68] CIVIDEC Instrumentation GmbH, www.cividec.at.

[69] https://github.com/SiLab-Bonn/pyBAR

[70] C. Gao et al., *A Novel Source-Drain Follower for Monolithic Active Pixel Sensors*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 831 (2016), 147-155. doi: 10.1016/j.nima.2016.03.074.

[71] The ATLAS collaboration, *ATLAS level-1 trigger: Technical Design Report*, CERN-LHCC-98-014; ATLAS-TDR-12.

[72] Bai Y, Bodlak M, Frolov V, Jary V, Huber S, Konorov I, et al. *Overview and future developments of the FPGA-based DAQ of COMPASS*, JINST. February 9, 2016;11(02):C02025-C02025.

[73] K. Red, *Single event upsets in SRAM FPGA based readout electronics for the time projection chamber in the ALICE experiment*, Ph.D. dissertation, The University of Bergen, Bergen, Norway, 2009.

[74] Placinta V-M, Cojocariu LN. *Radiation Hardness Studies and Evaluation of SRAM-Based FPGAs for High Energy Physics Experiments*, Proceedings of Topical Workshop on Electronics for Particle Physics — PoS(TWEPP-17).

[75] Wirthlin M. *High-Reliability FPGA-Based Systems: Space, High-Energy Physics, and Beyond*, Proceedings of the IEEE. 2015;103(3):379-89.

[76] R. Dubey, *FPGA Devices*, Springer London, 2009.

[77] I. Kuon, R. Tessier, J. Rose, *FPGA Architecture: Survey and Challenges*, Foundations and Trends in Electronic Design Automation, vol. 2, no. 2, pp. 135-253, 2008.

[78] Xilinx Inc. *Partial Reconfiguration User Guide*, XAPP538 (v1.0) April 4, 2012.

[79] https://www.xilinx.com/support/documentation/data_sheets/ds180_7Series_Overview.pdf

[80] Xilinx Inc. *7 Series FPGAs Configurable Logic Block*, UG474 (v1.8) September 27, 2016.

[81] Xilinx Inc. *7 Series FPGAs SelectIO Resources*, UG471 (v1.10) May 8, 2018.

[82] *7 Series FPGAs Clocking Resources*, UG472 (v1.4) July 30, 2018.

[83] *VC707 Evaluation Board for the Virtex-7 FPGA*, User Guide, UG885 (v1.8), (February 20 2019).

[84] *KC705 Evaluation Board for the Kintex-7 FPGA*, User Guide, UG810 (v1.9), (February 4 2019).

[85] https://www.vita.com/

[86] Yi-Chin Chu, CISCO SYSTEMS, *Serial-GMII Specification*, ENG-46158 (V1.8), (April 27 2005).

[87] RAL Space CMOS Active Pixel Sensor technology [Internet]. Ralspace.stfc.ac.uk. 2019 [cited 22 May 2019]. Available from: https://www.ralspace.stfc.ac.uk/Pages/CMOS-active-pixel-sensor-technology.aspx

[88] Cardella, R. et al., *MALTA: an asynchronous readout CMOS monolithic pixel detector for the ATLAS high-luminosity upgrade*, JINST, Vol. 14, No. 06, jun 2019, pp. C06 019, available at: https://doi.org/10.1088%2F1748-0221%2F14%2F06%2Fc06019.