Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

TECHNISCHE UNIVERSITÄT WIEN

Dissertation

Monolithische aktive Pixelsensoren für Detektoren mit hohen Teilchenraten

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines

Doktors der Technischen Wissenschaften Technische Physik

eingereicht an der

Technischen Universität Wien, Fakultät für Physik

unter der Leitung von

Privatdoz. Dipl.-Ing. Dr.techn Christoph Schwanda

Univ.Lektor Dipl.-Ing. Dr.techn. Bergauer Thomas

von

Dipl.-Ing. Patrick Sieberer BSc

Matrikelnummer: 01328786

Unterschrift Betreuer

Unterschrift Student

Wien, am 21.3.2023

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

TECHNISCHE UNIVERSITÄT WIEN

Doctoral Thesis

Monolithic Active Pixel Sensors for High Rate Tracking Detectors

submitted in satisfaction of the requirements for the degree of

Doctor of Science in Engineering Sciences Technical Physics

submitted to

TU Wien, Faculty of Physics

under supervision of

Privatdoz. Dipl.-Ing. Dr.techn Christoph Schwanda

Univ.Lektor Dipl.-Ing. Dr.techn. Bergauer Thomas

by

Dipl.-Ing. Patrick Sieberer BSc

Matrikelnummer: 01328786

Signature Supervisor

Signature Student

Wien, on 21.3.2023

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Kurzfassung

Verarmte monolithische aktive Pixelsensoren (in Englisch: Depleted Monolithic Active Pixel Sensors, DMAPS) sind eine vielversprechende Technologie für zukünftige Spurdetektoren. Das Konzept hinter einem monolithischen Ansatz ist die Integration der Messelektrode eines Silizium Detektors und der Ausleseelektronik im selben Stück Silizium. Die Technologie hat zahlreiche Vorteile gegenüber konventionellen hybriden Pixelsensoren. Eine verbesserte Spurrekonstruktion aufgrund geringeren Materialbudgets und einfachere, sowie billigere Produktion, da keine fehleranfällige Flip-Chip Montage nötig ist, sind wesentliche Vorteile von DMAPS.
Die CERN RD50-HVCMOS Gruppe erforscht diese neue Technologie und setzt dabei einen Fokus auf Strahlenhärte, einer der größten Herausforderungen für zukünftige Hochenergiephysik-Detektoren aufgrund des Luminositätsanstieges. Der Fokus dieser Dissertation ist das Digitaldesign und die Möglichkeit zur Integration in größere Experimente, was den nächsten wesentlichen Entwicklungsschritt darstellt.

Im Zentrum dieser Dissertation steht eine erhebliche Erweiterung des Vorgängerchips um eine digitalen Auslese. Der Chip ist in einem kommerziellen CMOS Prozess mit einer Strukturgröße von 150 nm entwickelt und von LFoundry auf einen hoch-resisitiven Substrat gefertigt, um die Verarmungszone mit höherer Vorspannung zu vergrößern und die Strahlenhärte zu erhöhen. Eine essenzielle Komponente des neues Chips, RD50-MPW3, ist die digitale Peripherie, welche eine 640 MHz Datenausgabe bereitstellt, die eine Trefferrate von zirka 1 MHz/cm^2 auslesen kann. Der Fokus liegt auf Integration in ein Auslesesystem zur schnellen Datenauslese, Kommunikation und Synchronisation mit anderen Detektoren. Dies beinhaltet die Implementierung eines weit verbreiteten I2C Busses sowie definierte Datenpakete und Kodierung des Datenstroms für eine zuverlässige, verlustfreie Datennahme. Die neu entwickelte Peripherie arbeitet mit der Design-Frequenz und bislang wurden keine Probleme gefunden.

Die Arbeit beinhaltet eine detaillierte Auswertung des Vorgängers, RD50-MPW2, sowie des neu entwickelten RD50-MPW3. Die Charakterisierung von ersterem fokussiert auf die analoge Leistungsfähigkeit des Sensors, wofür eine Bestrahlungskampagne durchgeführt wird, um die Strahlenhärte zu testen. Das Augenmerk bei der Auswertung von RD50-MPW3 liegt auf der digitalen Auslese, wobei der Chip in einen größeren Aufbau integriert und mit anderen Subdetektoren im Rahmen eines Strahltests synchronisiert wird. Die Analyse dieser Daten beendet diese Dissertation. Den Abschluss der Dissertation bildet die Analyse der Daten.

Die Ergebnisse aller Studien zeigen, dass die DMAPS Technologie eine vielversprechende Wahl für viele Experimente in der Hochenergiephysik ist. Strahlenhärte, Materialbudget und Auslösung sind bereits nahe an den Anforderungen für verschiedene, geplante Detektoren nach 2035. Die Technologie ist für eine Integration in derzeitige Systeme bereit. Um die vollen Kapazitäten der DMAPS Technologie auszuschöpfen, sind Forschungsdesiderate im Bereich der Integration in große Systeme, wie beispielsweise die Akzeptanz eines externen Triggers, zu beheben.

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Abstract

Depleted monolithic active pixel sensors (DMAPS) are a promising technique for future solid-state tracking detectors. The concept behind a monolithic approach is to integrate the sensing electrode of the silicon detector and the readout electronics into the same piece of silicon. The technique has numerous advantages over conventional hybrid pixel sensors. Improved tracking performance due to less material budget and potentially simpler and cheaper production as no error-prone bump-bonding is needed are essential benefits of DMAP sensors.
The CERN RD50-HVCMOS group aims to investigate this new technology, focusing on radiation hardness, one of the most critical challenges of future high-energy physics detectors due to the luminosity increase. This thesis focuses on digital design and integration features as the next essential step toward integration in a large-scale experiment.

The main focus of this thesis is the design of a significantly extended chip, including a digital readout. As with all the predecessors, the chip is designed using a commercial CMOS process with a 150 nm feature size from LFoundry, fabricated on a high resistivity substrate to enlarge the depletion region using a higher bias voltage and thus radiation hardness. A major new compound of the new chip, RD50-MPW3, is the digital periphery, which provides a fast 640 MHz output stream, capable of reading out a hit rate of around 1 MHz/cm^2. The focus is on integrating the chip into a DAQ system for efficient readout, communication, and synchronization with other detectors. This includes implementation of a communication interface via the widely known I2C bus, and framing and encoding the output stream to ensure reliable data recording. The newly developed digital periphery does operate at design speed without any issues found so far.

The work includes a detailed evaluation campaign of the predecessor, RD50-MPW2, and the newly designed RD50-MPW3. The characterization of RD50-MPW2 focuses analog performance of the sensor front end, and an irradiation campaign is conducted to study radiation hardness. Performance studies of the new digital readout are the main focus during the evaluation of RD50-MPW3. The chip is integrated into a large setup and synchronized with other subdetectors during a test beam campaign. The analysis of this campaign concludes this thesis.

The results of all studies prove that the DMAPS technology is a promising choice for many future experiments in high-energy physics. Radiation hardness, material budget, and resolution are already close to the demands for various detectors planned to be built after 2035. Although the technology is ready to be integrated into current detector systems, more R&D with a focus on integration into large systems, for instance, accepting an external trigger decision, is needed to exploit the full capabilities of DMAPS.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Contents

1 Introduction 1
1.1 Future Challenges of Advanced Detector R&D . . . 1
1.2 European Initiative for Future Particle Detectors . . . 1

2 Silicon Sensors and Radiation Damage 5
2.1 Signal Generation in Silicon Sensors . . . 5
2.1.1 Semiconductor Materials and Doping . . . 5
2.1.2 PN-Junction and Depletion Zone . . . 7
2.1.3 Charge Generation and Collection . . . 8
2.2 Particles in Matter . . . 10
2.2.1 Energy Loss at Low Energies . . . 11
2.3 Radiation Damage . . . 12
2.3.1 Non-Ionizing Energy Loss . . . 12
2.3.2 Defect Classification . . . 13
2.3.3 Annealing . . . 13
2.3.4 Negative Effects of Radiation Damage . . . 14

3 Track and Vertex Detectors for Future Experiments 15
3.1 Track and Vertex Reconstruction . . . 15
3.1.1 Tracking . . . 15
3.1.2 Vertexing . . . 15
3.2 Future Challenges for Tracking Detectors . . . 16
3.2.1 Luminosity and Granularity Increase . . . 16
3.2.2 Radiation Environment . . . 17
3.2.3 Multiple Scattering . . . 17
3.2.4 Trigger Information . . . 18
3.2.5 Beam Telescopes . . . 18
3.2.6 Translation to Tracking Parameters . . . 19

4 Depleted Monolithic Active Pixel Sensors within RD50 21
4.1 Depleted Monolithic Active Pixel Sensors for Future Tracking Sensors . . . 21
4.1.1 MAPS . . . 21
4.1.2 DMAPS . . . 23
4.2 RD50 and the RD50-MPW series . . . 25
4.2.1 Initial Design: RD50-MPW1 . . . 25

5 Digital Design Implementation 27
5.1 Requirements of Digital Design for CMOS Fabrication . . . 27
5.1.1 Design Entry: RTL Code and Constraints . . . 27
5.1.2 Process Entry: Physical Information, Timing Libraries and Design Rules 28
5.2 Fundamentals of Digital Design Implementation . . . 29
5.2.1 RTL-to-GDSII Flow . . . 29
5.2.2 Software Tools . . . 33
5.2.3 Chip-level Implementation, Verification and Production . . . 33
5.2.4 Requirements Engineering and Industrial Verification . . . 35
5.3 Conclusion . . . 38

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

6 DAQ and Testbeam for RD50-MPW2 39
6.1 Introduction to RD50-MPW2 . . . 39
6.1.1 Analog-only Pixel Matrix . . . 39
6.1.2 Other Features of RD50-MPW2 . . . 41
6.2 Readout System of RD50-MPW2 . . . 41
6.2.1 Hardware . . . 42
6.2.2 Firmware . . . 43
6.2.3 Software . . . 43
6.3 Test Campaign of RD50-MPW2 . . . 44
6.3.1 Radiation Hardness Tests . . . 44
6.3.2 Timing . . . 46
6.3.3 RD50-MPW2 Testbeam at MedAustron . . . 47
6.4 Conclusion . . . 51

7 Design of the RD50-MPW3 Chip 53
7.1 Overview and Analog Design of RD50-MPW3 . . . 53
7.1.1 Goals and New Functionalities . . . 53
7.1.2 Pixel Design . . . 54
7.1.3 Double Column Architecture . . . 55
7.2 Digital Readout Design of RD50-MPW3 . . . 56
7.2.1 Overview of Digital Blocks . . . 56
7.2.2 End Of Column . . . 59
7.2.3 Control Unit . . . 62
7.2.4 Transmission Unit . . . 63
7.2.5 Critical Timing Path . . . 65
7.3 Conclusion . . . 66

8 Testbeam for RD50-MPW3 67
8.1 Adapting the Readout System for RD50-MPW3 . . . 67
8.1.1 MPW3 firmware . . . 68
8.1.2 Time-Stamping . . . 69
8.2 Testbeam setup at CERN . . . 71
8.2.1 Accelerator and Beam Settings . . . 71
8.2.2 Reference Detectors . . . 71
8.2.3 Detector Setup . . . 72
8.3 Data Analysis . . . 73
8.3.1 Event Definition . . . 73
8.3.2 Clustering . . . 73
8.3.3 Alignment . . . 74
8.3.4 DUT Analysis . . . 76
8.4 Conclusion . . . 79

9 Conclusion 81
9.1 Digital Design . . . 81
9.2 DMAPS for Tracking in Future Collider Experiments . . . 83

Acknowledgements 85

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

1 Introduction

1.1 Future Challenges of Advanced Detector R&D

Modern high-energy physics experiments put more and more challenging demands on detectors and instrumentation to discover the nature of the world. Significant improvements in technology are needed to process the signals and handle the amount of data produced by future detectors. The Office of Science from the US Department of Energy summarizes the needs in four 'Grand Challenges' in their report on high energy physics detector research [1]. The following list converts them into requirements for future high-energy physics detectors:

- The sensitivity of detectors has to improve to the maximum possible by technology to detect the tiniest signals. This is not only needed for detecting dark matter signals, which are possible in the eV range or even below, but also for better differentiation of signal and noise in collider experiments.
- Future detectors must be scalable in order to find rare phenomena. This does include an increase in the active pixel area and the granularity of the readout. The electronics also need to be scalable to handle the increased data rate.
- Novel materials for signal generation need to be studied to open doors beyond the current technological limits.
- Extreme environments will be a key technology for future detectors. This includes especially a harsh radiation environment for collider-type experiments with increasing luminosity and operation at very low temperatures, especially for dark matter detectors.

These challenges can only be addressed with an extensive R&D campaign, which needs large, international and diverse collaborations.

1.2 European Initiative for Future Particle Detectors

The Conseil européen pour la recherche nucléaire (CERN) council updates the European strategy of particle physics every five to ten years. The last update happened in 2020 when the CERN council asked the European Committee for Future Accelerators (ECFA) to propose a future European detector development strategy. The outcome was published as ´The 2021 ECFA Detector Research and

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Development Roadmap´ in October 2021 and can be found in [2]. This document is the baseline for future work on detectors in particle physics. Based on its proposals, many initiatives and research projects are currently adapted to fit the proposed scheme.

The ECFA detector roadmap is closely related to the future timeline of accelerators and colliders. Although the major focus is on the development of detectors for the next five to ten years of accelerator upgrades and initiatives, the long-term prospect of large-scale colliders is also considered. In the short-term upgrades, a focus is put on fixed target experiments at the CERN Super Proton Synchrotron (SPS) and collider experiments like the Belle II detector in Japan or ALICE and LHCb detectors at CERN. In parallel, the strategy involves the development of non-collider experiments, mainly focusing on smaller-scale dark matter and neutrinos experiments. The long-term outlook does include research and development (R&D) work for large-scale non-collider and collider experiments, such as the LHC upgrades, the International Linear Collider (ILC), or the Future Circular Collider (FCC).

The roadmap document divides detector research into nine task forces listed below:

TF#1: Gaseous Detectors

TF#2: Liquid Detectors

TF#3: Solid State Detectors

TF#4: Photon Detectors & PID

TF#5: Quantum & Emerging Technologies

TF#6: Calorimetry

TF#7: Electronics & On-Detector Processing

TF#8: Integration

TF#9: Training

Those task forces have yet to be built, and research groups are about to decide which task forces they want to contribute. The idea of the ECFA roadmap scheme is to provide overall support, have a common guideline everybody can use as the baseline for their work, and find and use synergies between different working groups within the same task force and across them. Thus, this thesis aims to find suitable task forces for the project group for future collaboration and development targets, as the project will end after 2023.

Attractive task forces for this project are TF3 and TF7, so a closer look into them is given in this sections, including the main goals and detector R&D themes

TU Bibliothek WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

(DRDTs) for each task force. An overview of technologies used in each task force is given, while details about them can be read in the roadmap document [2].

The task force on solid state detectors (TF3) focuses on silicon detectors as they achieve excellent position resolution nowadays due to lithographic processing, and they play a key role not only in tracking but also in segmented calorimeters and photonic detectors. Four DRDTs have been identified within the task force. DRDT 3.1 focuses on monolithic active pixel sensors (MAPS), which include sensing and processing within the same piece of silicon. Commercial CMOS imaging (CIS) technology should be explored for tracking, while passive sensors are a good candidate for calorimetry. The exploration of 4D sensors, including a precise timing measurement, is the aim of DRDT 3.2. This includes research with low gain avalanche diodes (LGADs), 3D sensors and timing capabilities of MAPS. Operation of the sensor at extreme fluences is the topic of DRDT 3.3, which includes wide bandgap (WBG) materials and 2D structures as sensors. 3D interconnections using through-silicon vias (TSVs) are already widely used in industry but have yet to be used in particle detectors. The exploration of this technology is the goal of DRDT 3.4, as it promises a lighter and more compact sensor, which reduces power consumption.

Electronics is and has always been one of the most significant challenges of large-scale detectors. The readout of modern detectors requires state-of-the-art electronics combining speed, low power and radiation requirements. Thus, TF7 deals with R&D on future electronics for detector readout systems that utilize novel developments for particle physics. The DRDTs of this task force cover the development of high data rate application-specific integrated circuits (ASICs) and technologies for intelligent detectors. These should utilize versatile front-end electronics and data reduction techniques like machine learning (ML) and artificial intelligence (AI). Technological support for 4D or even 5D techniques, including fast analog to digital and time to digital converters and support for electronics operating under extreme conditions (radiation hardness) are part of this task force as well in order to deal with new technologies mentioned in TF3 besides exploration of emerging electronic technologies from industry.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

2 Silicon Sensors and Radiation Damage

Silicon is the bulk material for the sensors used in this project. Thus, this chapter deals with the relevant physical aspects of semiconductors and silicon in detail.

2.1 Signal Generation in Silicon Sensors

Silicon as a detector material is introduced at the beginning of this chapter, followed by an explanation of measuring signals in solid-state silicon detectors.

2.1.1 Semiconductor Materials and Doping

The border between conductors, semiconductors and insulators is fluent. [3] gives a typical value of $10^{-3} - 10^{8}\,\Omega\text{cm}$ for the resistivity of semiconductors as criteria for this border. This category includes silicon and germanium as elemental semiconductors of the fourth group and a few other composite materials for elements of the third group combined with the fifth. Also, silicon carbide and diamond (as an insulator) are nowadays investigated as sensor materials. While all of them have certain advantages and disadvantages, silicon is still the most used. The following sections deal with silicon as an example, although most of the physics presented can be easily adapted for other materials.

Silicon has a diamond lattice structure and four conduction electrons as an element in the fourth group. Figure 2.1 shows a two-dimensional representation of the lattice, where each conduction electron is shared between two silicon atoms. This binding is relatively weak, and thermal excitation at temperatures $> 0K$ is enough to free one electron, which can travel around in the lattice. Also, holes can travel by being occupied by a neighboring electron. The amount of charge carriers in such pure silicon is relatively low, and [3] gives a typical value of $\sigma \approx 2.8 \cdot 10^{-4} (\Omega m)^{-1}$ as conductivity for this so-called intrinsic silicon.

The electronic characteristics can be improved for the use as particle detector by *doping* the silicon. During the doping process, a silicon atom is replaced by an atom from the third or fifth group.

Figure 2.2a shows the replacement of a silicon atom with boron, an element of the third group. In this configuration, an electron is missing for the covalent bond; thus, a hole is left in the lattice. This can hole can easily travel around in the lattice and serves as a positive charge carrier. It must be noted that the whole crystal is still neutral because there are as many positive as negative charge carriers in the crystal.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

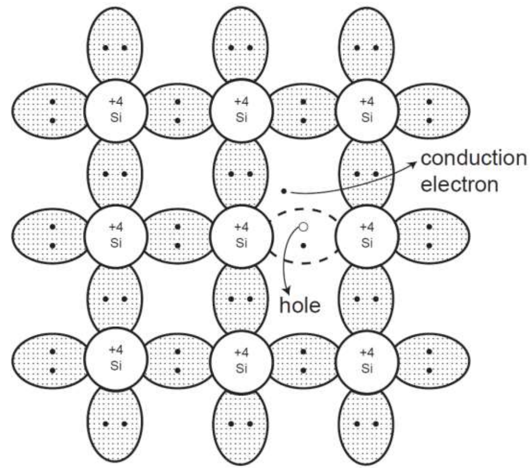
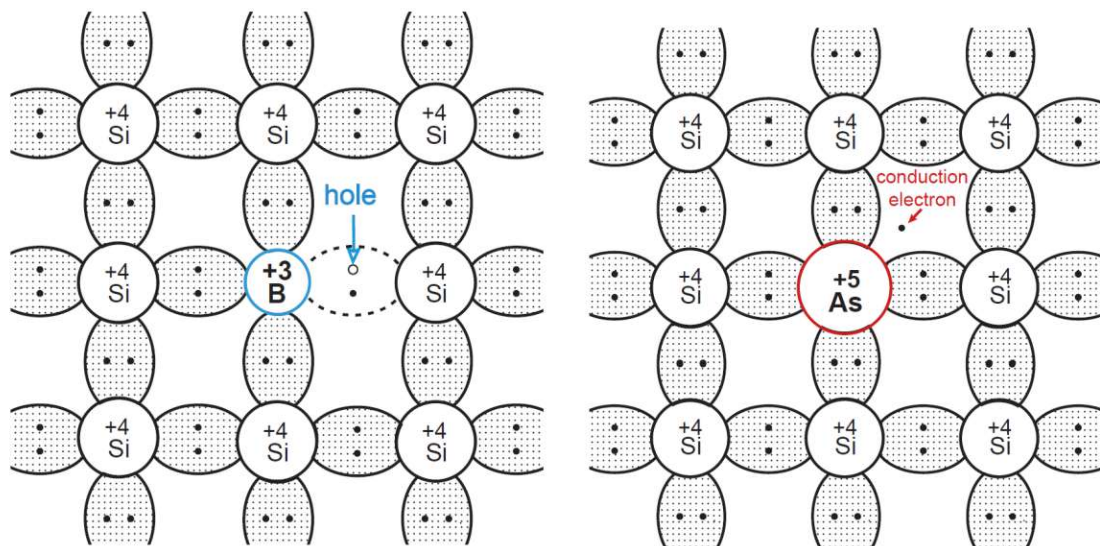


Figure 2.1: Each atom of silicon has four electrons in the outer shell. Each electron is shared with another electron from a neighboring atom with a covalent bond. This bond can be broken by thermal excitation, which frees a single electron that can be used as a charge carrier. Picture taken from [3].

The second possibility is depicted in figure 2.2b, where a silicon atom is replaced with arsenic, an element of the fifth group. In this case, there is one electron too much in the lattice, which can travel around quite freely as it does not have a bonding partner.
The conductivity of such so-called extrinsic silicon depends on the doping concentration, and doping can change it by orders of magnitude.



a): p-doped silicon with a boron implant

b): n-doped silicon with an arsenic implant

Figure 2.2: The two different types of doping are depicted. An example of p-type silicon is shown in figure2.2a, while n-type silicon is shown in figure 2.2b. Picture taken from [3].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2.1.2 PN-Junction and Depletion Zone

P-type and n-type silicon are already quite good conductors. A combination of both types of silicon is needed to form a sensor that can collect an induced charge. The transition between an n-type and a p-type silicon is called a pn-junction. Figure 2.3 illustrates what is happening at the transition zone for two differently doped silicon layers. Electrons from the n-layer will fill the holes of the p-layer, as they diffuse into the other layer due to the inhomogeneous distribution of charge carriers. No charge carriers are available in this border region anymore as electrons fill all holes. Thus this zone is called *depletion zone*. However, the space-charge is not zero anymore, as there are too many electrons in the p-type silicon, which generates a negative *space-charge region* (another naming for depletion zone) in the p-layer and a positive space-charge region in the n-layer. In this example, the doping concentration for the p-type is less than the doping concentration of the n-layer. Thus a larger area of the less doped p-type layer is depleted. Nevertheless, the amount of positive and negative charge carriers are still equal, and the overall crystal is still neutral. The constant space-charge density translates to a linear electric field (with a kink) by using the Maxwell equations and demanding the same charge in both regions. The electric potential is quadratic, as it is given by the integral of the electric field.

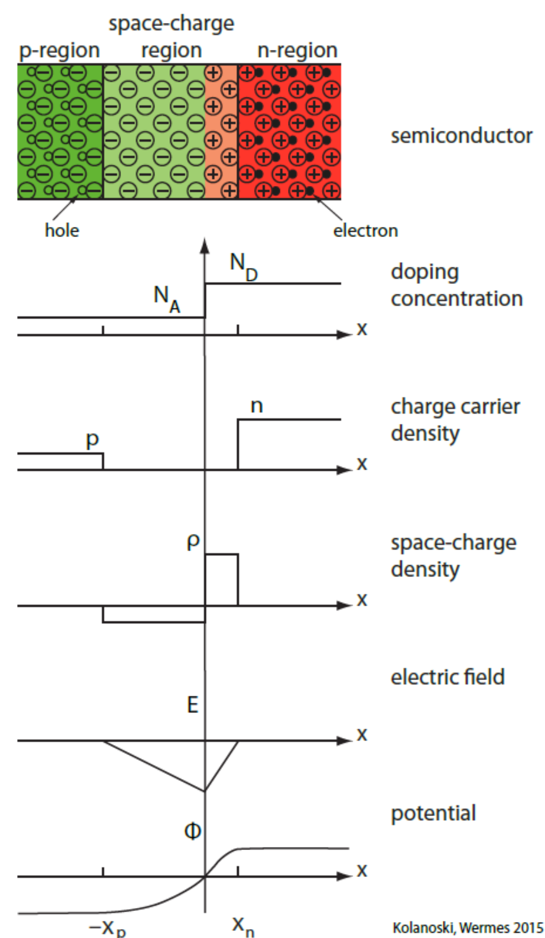


Figure 2.3: A symbolical view and important physical parameters of a pn-junction are depicted. An electric field is only present inside the space-charge region. Thus no charge can be collected far away from the junction. Acceptors are atoms from the third group with a negatively charged lattice position and a positive hole. Those are drawn with a ⊖-symbol and the amount is noted as N_A
Donors are atoms from the fifth group with a positively charged lattice position and a negative electron surrounding it. Those are drawn with a ⊕-symbol and the amount is noted as N_D
Picture taken from [3].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2.1.3 Charge Generation and Collection

The basic principle of charge generation in a pn-junction is rather simple. A traversing ionizing particle with enough energy creates an electron-hole pair in the space-charge region. These two charges automatically drift away from each other due to the electric field inside the space-charge region. If the p and n regions are connected using a metalization, this current can be measured as an electrical signal. However, a simple pn-junction is not a useful sensor yet. A modern silicon sensor has many modifications for mechanical or electrical reasons.

The whole ingot is typically doped with one flavor to make fabrication easier. For instance, sensors have a p-doped bulk (p-bulk) material with only small n-doped implants on one side. A typical thickness, limited by mechanical cutting techniques of ingots, is 300 µm. Such a sensor is depicted in figure 2.4, while the reverse type also exists, and all concept works accordingly by changing polarity. The n-implants are structured either in one-dimensional strips or two-dimensional pixels. The n-implants are metalized and serve as one electrode, while an entire backside metalization gives the other electrode. Each implant is separately connected to the readout electronics via alternating coupling (AC) or direct coupling (DC). The pn-junction is between every n-implant and the p-bulk instead of just one big junction over the whole sensor to get more granular position information of the traversing particle.

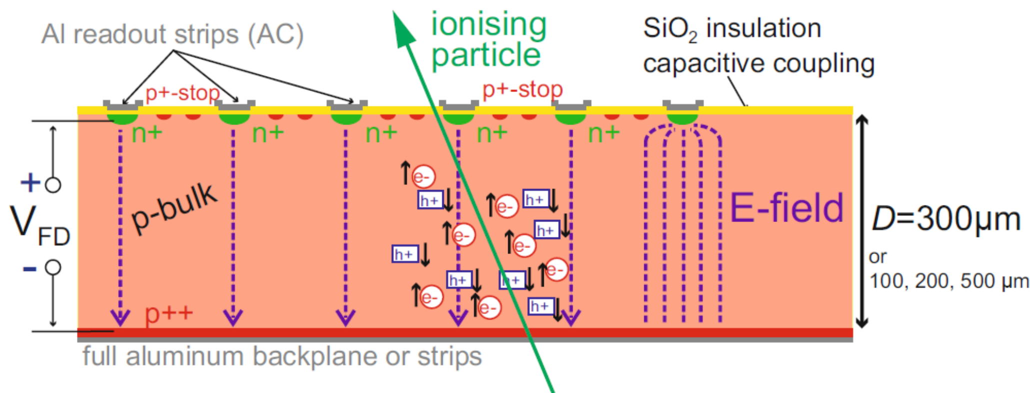


Figure 2.4: An illustration of charge generation for a p-doped bulk material is depicted. N-doped regions are printed in green, while red is chosen for p-doped regions. Darker color corresponds to higher doping. Picture taken from [4].

Another problem of the n-implant in p-bulk material is the surface between the bulk and the isolation layer of silicon oxide (drawn in yellow in figure 2.4). Due to broken bonds, positive charges at these surfaces attract the electrons that accumulate between the strips and short the n-implants. A higher doped p-implant is needed to dissipate the electrons along the sensor and isolate the n-implants. Various techniques of this p-implant exist, where p-stop is one of them. This modification is not needed with p-in-n sensors (p-implants in an n-bulk), as holes move to this surface, which are repelled by the positive charges in the surface.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Due to the electrical field, electrons travel toward p-type layers. They can fill up a hole before reaching the collection electrode, which is called recombination. As this is a loss in signal, one has to counteract this behavior to produce an efficient sensor. To stop the recombination, one has to reduce the number of free charge carriers outside the depleted region using one of the two following techniques:

1. *Cooling.* Cooling of the sensor leads to decreased mobility, resulting in fewer free charge carriers due to less thermal excitation.

2. *Reverse Biasing.* The space-charge region can be extended by applying a reverse bias to the sensor, connecting a positive potential to the negative implant and a negative potential to the positive implant. In an ideal sensor, the depletion zone fills the whole sensor, which is called full-depletion (FD) and is also shown in figure 2.4.

Typically, both techniques are exploited to some extent, while the latter option has the advantage of a strong electric field, which equals a shorter path of the charge carriers and a faster sensor. Moreover, charge collection is dominated by the drift induced by the electric field rather than diffusion, which is the random lateral movement.
However, reverse biasing leads to another problem. If the applied voltage is too high, the current increases rapidly, as seen in the IV curve in figure 2.5. The breakdown voltage has multiple dependencies, mainly on the sensor's geometry. The current below V_{BD} follows a $\sqrt{V}$ dependence, which is explained in [4] and is shown in the small plot in figure 2.5. V_{BD} must be higher than the full depletion voltage V_{FD} to fully deplete the sensor.

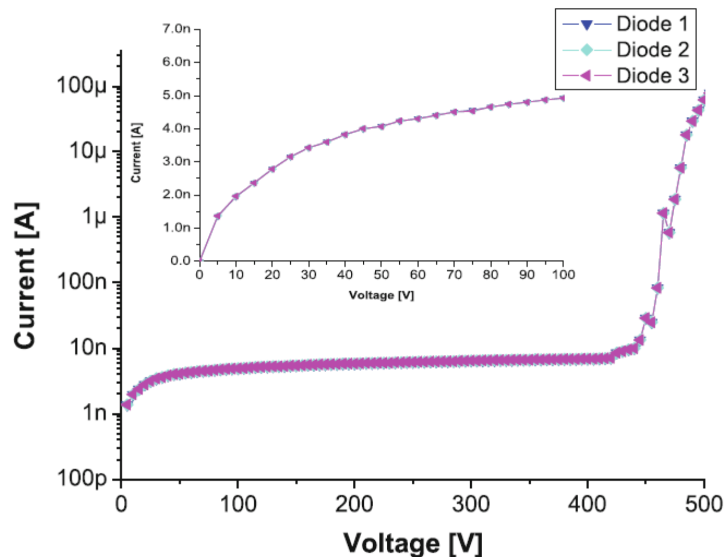


Figure 2.5: The IV-curve in the reverse bias region of a pn-junction is shown. Picture taken from [4].

Due to induction, the charge is already measured before the generated electron-

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

hole pairs reach the electrode. This theorem is called Shockley-Ramo theorem; details about it can be found in [3] or [4]. As a matter of concept, the signal starts when the particle produces the first electron-hole pair and ends when the charge is collected by the electrode.

2.2 Particles in Matter

The generated charge in the space-charge region of solid-state detectors depends on the detector material and the traversing particle. Silicon detectors can only detect ionizing particles that deposit enough energy to create at least one electron-hole pair. Ionization happens mainly by the passage of charged, heavy particles and is described by the energy loss (or stopping power in the case of hadrons) of the incident particle.
The Bethe-Bloch equation is an semi-empirical function describing the energy loss of heavy particles in matter. A detailed derivation of the non-empirical terms can be found in [3] and [5]. Equation 2.1 shows a slightly adapted version from [5]. The stopping power depends on the energy of the incident particle and is drawn as a function of it in figure 2.6. A detailed derivation is waived in this thesis, as it can be found in the references. Instead, an interpretation of the terms relevant to tracking devices is given.

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = 4\pi N_A r_e^2 m_e c^2 z^2 \frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\left(\frac{2m_e c^2\beta^2\gamma^2 E_{kin,max}}{E_{ex}^2}\right) - \beta^2 - \frac{\delta(\gamma)}{2}\right] \tag{2.1}$$

Variables used in equation 2.1 and 2.2 are listed and explained here:

$$\frac{\mathrm{d}E}{\mathrm{d}x} = \text{Energy loss per length unit}$$
$$N_A = \text{Avogadro's constant}$$
$$r_e = \text{Electron radius}$$
$$m_e = \text{Electron mass}$$
$$z = \text{charge number of incoming particle}$$
$$Z = \text{Atomic number}$$
$$A = \text{Atomic mass}$$
$$\beta = \frac{v}{c}$$
$$\gamma = \text{Relativistic } \gamma\text{-factor: } \frac{1}{\sqrt{1-\beta^2}}$$
$$E_{kin,max} = \text{The maximum kinetic energy a free electron can have after a collision (= maximum of transmitted energy)}$$
$$E_{ex} = \text{mean excitation energy}$$
$$\delta = \text{density effect correction factor}$$

Figure 2.6 shows a wide energy range of the incident particle, while the border regions are not described in the basic Bethe-Bloch formula and are sometimes added with additional terms. Particles in high-energy physics typically have an

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

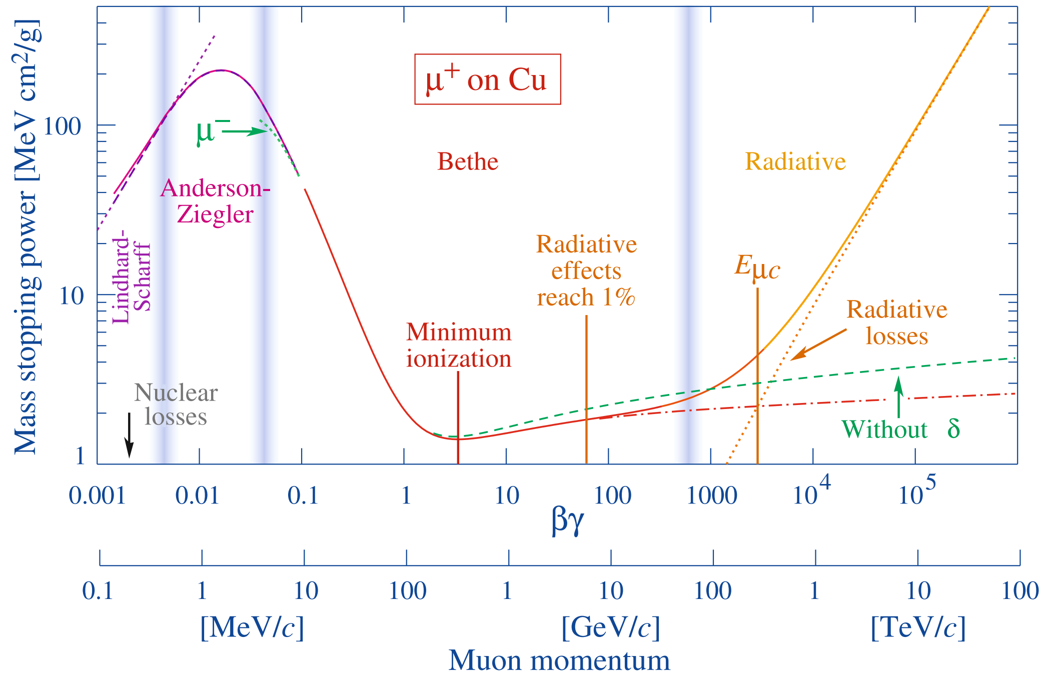


Figure 2.6: As an example of the Bethe-Bloch equation, the stopping power of muons in copper is drawn. Effects are marked, while the Bethe-Bloch equation only holds in the middle of the shown energy scale. The x-axis is given in $\beta\gamma$, which is proportional to $\frac{p}{m}$. Picture taken from [5].

energy of 100 MeV to a few GeV; thus the standard Bethe-Bloch formula covers the relevant region, and borders are neglected hereafter. The β^{-2} term is dominating for lower energies (or velocity v), resulting in a quadratic loss. For high energies, the $\ln(\beta^2\gamma^2)$ increase counteracts, leading to a minimum in the energy-loss curve. This point is called minimum ionization, and particles with this energy are called minimum ionizing particles (MIPs). There are two reasons why this point is essential for tracking devices: Firstly, detectors must be sensitive enough to detect MIPs to record every possible traversing particle. Secondly, the energy loss is nearly constant and quite close to the value of MIPs, keeping in mind the logarithmic scale in figure 2.6. The nearly constant region covers the range of typically achieved energies in high-energy physics. The assumption of a constant energy loss for all energies is often exploited in experiments, especially for calibration of sensors.

2.2.1 Energy Loss at Low Energies

For lower energies still in the Bethe-Bloch range, one can assume $\beta << 1$ and neglect the δ-term needed for high energies. With this assumption and replacing βc by v in equation 2.1, the energy loss function can be simplified and is shown in equation 2.2. This limit for small energies is sensible at medical facilities, which reach much lower energies than particle accelerators for high-energy physics.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = 4\pi N_A r_e^2 m_e c^4 z^2 \frac{Z}{A}\frac{1}{v^2}\left[\frac{1}{2}\ln\left(\frac{2m_e v^2 E_{kin,max}}{E_{ex}^2}\right)\right] \tag{2.2}$$

2.3 Radiation Damage

Radiation hardness is one of the critical characteristics of silicon sensors, as damage in the lattice introduces various adverse effects mentioned at the end of this section. Increasing the sensor's radiation hardness is one of this project's primary goals; thus, an introduction to radiation damage is given here.

2.3.1 Non-Ionizing Energy Loss

Radiation can cause damage to the surface or the bulk of the silicon sensor. The space-charge region is usually buried and not at the surface, thus, the latter is much more important. Charged, hadronic interaction is mainly a coulomb interaction with electrons in the atomic shell leading to ionization of the sensor. Such interactions are typically reversible as the generated electron-hole pair is simply replaced with electrons and holes from the electrodes. This ionizing radiation does not cause permanent damage to the silicon bulk. Instead, so-called non-ionizing energy loss (NIEL) is the main mechanism for causing damage. [6] describes the primary mechanism as displacement of a *primary knock on atom (PKA)* which leads to an interstitial atom and a vacancy in the lattice. It is a reaction of high-energetic, uncharged particles with the nucleus.

The NIEL hypothesis can be found below and is explained in detail in the reference.

> The basic assumption of the NIEL hypothesis is that displacement-damage induced change in the material scales linearly with the amount of energy imparted in displacing collisions, irrespective of the spatial distribution of the introduced displacement defects in one PKA cascade, and irrespective of the various annealing sequences taking place after the initial damage. (Michael Moll, 1999 in [6])

The hypothesis has been tested over a wide range of energies nowadays, and many radiation damage calculations are based on it. The mentioned annealing effect is covered in section 2.3.3 and uses the NIEL hypothesis.

As a consequence of the NIEL theorem, the change in current between irradiated and unirradiated silicon sensors is constant, given a static volume. Section 5.1.2 in [6] gives a formula for this behaviour:

$$\frac{\Delta I}{V} = \alpha \times \Phi_{eq} \tag{2.3}$$

The *current related damage rate* α is constant and the NIEL hypothesis has been tested in-depth by measuring this parameter for many different types of silicon.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

2.3.2 Defect Classification

The Frenkel pair has already been mentioned as a displacement of a lattice silicon atom (PKA). The Frenkel pair is one possibility of so-called point defects. Interstitials (one additional silicon atom in the pure lattice) and vacancies are other options for pure silicon.
Furthermore, damage can also be caused by impurities leading to additional interstitials or substituting a regular lattice atom.
Lastly, a combination of all defects mentioned above is possible and illustrated in figure 2.7.

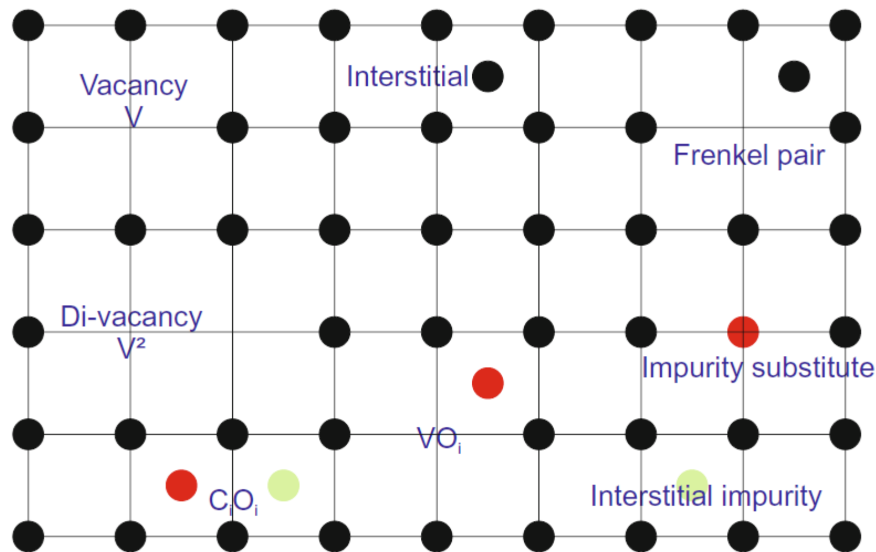


Figure 2.7: An illustration of the most common possible lattice damages is shown. The Frenkel Pair is the most elemental damage, shown on the top right. Impurities of carbon atoms are printed in red, while oxygen atoms are printed in green. Picture taken from [4].

An accumulation of point defects is called a cluster. Clusters can be arbitrarily complex and combine different point defects. Thus, they are hard to study.

2.3.3 Annealing

The diffusion process following lattice damage is called annealing. Diffusion is caused by thermal movement and can lead to the recombination of Frenkel pairs or other defects. Moreover, defects can move to the surface and annihilate there. The defects can introduce additional charge carriers in the bulk of the silicon, shifting the effective doping concentration. This effect can lead to so-called *type-inversion* where the effective doping concentration changes polarity. Before type-inversion, the effective doping concentration decreases, which is called beneficial annealing. After type-inversion, the effective doping concentration increases, which is called reverse annealing.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

2.3.4 Negative Effects of Radiation Damage

Leakage Current

Defects create additional energy levels less than the ionization energy of pure silicon. The mobility of charge carriers increases as additional energy levels can be used. The thermal energy needed to occupy these levels is lower than for ionization, leading to an increased leakage current of the sensor. This effect can be mitigated by cooling the sensor. This effect can be reduced by annealing, as fewer levels are available if defects annihilate due to thermal excitation. However, the annealing process does saturate over time.

Full Depletion Voltage

The full depletion voltage V_{FD} is proportional to the effective doping concentration, as fewer charge carriers are available, which have to recombine to reach full depletion of the sensor. During reverse annealing this becomes the most critical effect, as there is no way of counter-acting it, and sensors have to be designed to operate at a higher voltage towards to end of their lifetime, so one can still fully deplete the sensor after irradiation.

Charge Collection Efficiency

Defects serve as trapping centers for charge carriers, reducing charge collection efficiency as some of the generated charges are trapped by defects and not collected by the electrodes. This is a minor effect as long as the sensor is fully depleted, as the present electric field in the space-charge region attracts charges before they are trapped. The effect can be mitigated by applying a higher voltage, even above V_{FD}.

TU Bibliothek WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

3 Track and Vertex Detectors for Future Experiments

The chapter starts with an introduction to tracking and vertexing. Afterward, future challenges of tracking detectors are expressed, and the key characteristic which needs to be achieved are listed.

3.1 Track and Vertex Reconstruction

In order to optimize performance for the reconstruction of tracks and vertices, most modern tracking systems consist of two different devices for those tasks. Requirements for both differ from each other and are explained in this section.

3.1.1 Tracking

A tracking system consists of many sensing layers, which give a set of 3-dimensional points in space. Depending on the readout system, a fourth coordinate, the event time, is also available at a reasonable precision. Track reconstruction or, in short, tracking is the process of determining points originating from the same traversing particles and bundling these together. A track-fitting algorithm is used for each bundle to reconstruct the final track.

A crucial part of tracking is the alignment procedure. As tracking sensors deliver information in their local coordinate system, it has to be converted into global coordinates. This is done by aligning the sensors along a predefined, global coordinate system. This global system is usually chosen to be suitable for the whole detector and thus dependent on the geometry of all parts of it. For instance, if the detector has a cylindrical shape, cylindrical coordinates are chosen to simplify conversion.

Bundling points is called *track finding*, whereas fitting a track for one bundle is called *track fitting*. Those are two different mathematical problems, and various techniques exist for solving them. A comprehensive explanation of them can be found for example in [7].

3.1.2 Vertexing

Vertex reconstruction, or, in short, vertexing, is the process of combining reconstructed tracks into a track-bundle, where all tracks of the bundle originate from

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

the same collision point. The collision point is called a vertex. Vertices are classified into two categories. Primary vertices are, by definition, the points of collision of two beam particles, whereas secondary vertices are all other interaction points. Secondary vertices can either be decay points of unstable particles or interaction points of a particle with the detector. As a matter of fact, secondary vertices depend on primary vertices.

Contrary to track reconstruction, vertex reconstruction benefits from a-priory information. The collision point is either defined by the bunch size of the particle bunches in a collider experiment or by the target's position in a fixed-target experiment. However, this is just course information, typically in the order of centimeters to millimeters, as beam spots are not arbitrarily small. Thus, a precise vertex reconstruction using mathematical models is needed.

Techniques for vertex reconstruction are again divided into two mathematical problems. The process of bundling tracks that originate from the same point is called *vertex finding*, while the process of determining the origin is called *vertex fitting*. The latter is quite similar to track fitting, which is why the algorithms are similar. A precise explanation of used algorithms can also be found in [7].

3.2 Future Challenges for Tracking Detectors

A generic overview of challenges for future detectors has been given in the introduction. This chapter focuses on future challenges of tracking and vertexing detectors. An overview of the main challenges, which are equal for both types of detectors, is given first. Afterward, the challenges are converted into quantitative measures for both detector types. This section focuses, but not limits itself, on LHC experiments as a clear plan for the future upgrades of LHC is available.

3.2.1 Luminosity and Granularity Increase

The biggest challenge for future tracking devices originates from the High Luminosity LHC (HL-LHC) upgrade at CERN, which aims to increase the luminosity by a factor of ~10 after *Run 3* as seen in Figure 3.1. According to [8], this leads to a pileup of 200 collisions per event compared to around 20 at most nowadays.

The increased luminosity demands a more granular tracking system to distinguish tracks and vertices from each other. This can be achieved by using a smaller pixel size or upgrading the tracking system to a fourth dimension and including a precise timing measurement. A combination of both is desirable to reach the best possible results. Keeping in mind that the luminosity frontier will not stop with HL-LHC, the requirements for experiments in the farther future will be even tighter. For example, the SuperKEKB[1] at the High Energy Accelerator Research Organization (KEK) in Japan pushes the luminosity frontier way beyond LHC already today.

[1] https://www-superkekb.kek.jp/

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Figure 3.1: Integrated luminosity is measured in inverse femtobarn (fb^{-1}). A femtobarn is an area unit and corresponds to $10^{-13}m^2$, thus a luminosity of $3000\,fb^{-1}$ corresponds to 3000 particle per $10^{-13}\,m^2$. Picture taken from the High Luminosity LHC Project. (https://hilumilhc.web.cern.ch/)

3.2.2 Radiation Environment

Radiation by charged and uncharged particles damages silicon sensors, as shown in section 2.3. The very high luminosity, especially when colliding hadrons, leads to a high radiation environment, which is studied in detail before construction so that sensors can be adapted accordingly. For example, figure 3.2 shows the radiation study for the CMS tracker during the phase-2 upgrade. A detailed study can only be done when the detector geometry is known, so only rough predictions are available for planned detectors in the far future.

3.2.3 Multiple Scattering

Particles interact not only during collisions but also with the detector material. While this is necessary to measure the particle's energy in calorimeters, it is unfavorable for tracking detectors. Collisions with the material of the detector result in multiple scattering of the particles, which is more challenging to model than a straight line (or a curved line, accounting for the typically present magnetic field). A higher material budget leads to more multiple scattering; thus, the analysis must fit more parameters, which increases the possible fit errors.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek, WIEN Your knowledge hub

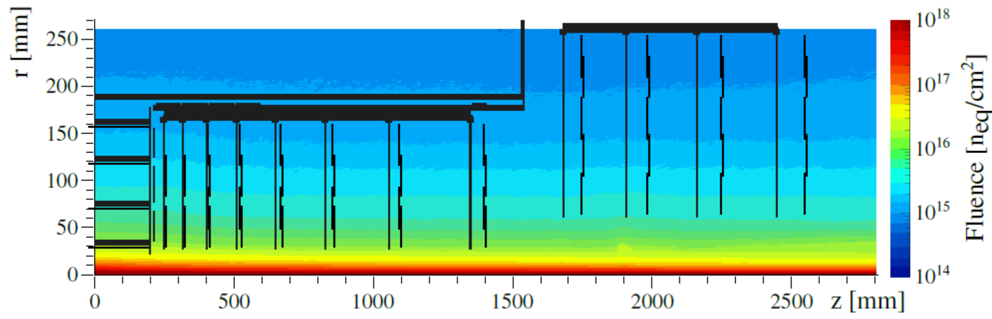


Figure 3.2: The particle fluence after an integrated luminosity of $3000\,fb^{-1}$. Fluence is typically given in 1 MeV-neutron equivalent per square-centimeter (n_{eq}/cm^2); thus, the impact of all traversing particles is converted to the impact of 1 MeV neutrons. As most particles graze each other, instead of colliding head to head, the highest fluence occurs close to the beam pipe at low scattering angles. Picture taken from [9].

3.2.4 Trigger Information

As large detector systems cannot store every piece of information they record, an event selection process, called a *trigger*, must be applied to reduce the amount of stored data. This decision must be fast, as sub-detectors must buffer information until the trigger decision is made. More precise and fast input to the trigger system from the tracker to improve the discriminating power of the event selection and increase the rate of interesting output is highly desired for future experiments. Trigger information from the tracker specifically allows the exploitation of single, isolated tracks and investigation of the pileup.
Front-end electronics need to be able to contribute to the trigger decision and output fast enough signals.

3.2.5 Beam Telescopes

Improvements in tracking detectors are not only needed for future large-scale experiments. The requirements of reference detectors, which are used to characterize other sensors, are also increasing to test the improved features of new sensors efficiently. So-called *beam telescopes* are detectors that characterize tracking sensors by providing a reference track, which the device under test (DUT) should also detect. Thus, higher spatial and timing resolution and less material budget are needed for detectors used in telescopes as well.
The readout rate should be as high as possible to reduce the time consumption of the tests. Furthermore, a rather new feature is demanded to increase the output rate of useful tracks even further: Future telescopes need to be able to define a region of interest (ROI), so they select only tracks that pass the DUT and thus reject all tracks which are just passing by the DUT and cannot be used for analysis.
A standardized data acquisition (DAQ) system and a standardized way of syn-

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

chronizing detectors using a trigger logic unit (instead of an entire trigger system for large-scale experiments) are also required. Consequently, future sensors' front-end readout electronics need to interact with such a readout chain. A standardized system makes setting up the experiment and comparing results easier.

3.2.6 Translation to Tracking Parameters

Translating the above-mentioned beam characteristics into performance parameters of track and vertex detectors depends on the physics goals one wants to achieve. A detailed table of future experiments and their target performance parameters can be found in [2], while a table of relevant parameters for selected experiments is given in table 3.1.

		BelleII (2026)	CMS Phase 3 (~2035)	ILC (~2040)	FCC-ee (~2045)	FCC-hh (>2045)
Vertex detector	σ_{hit} (µm)	$\lesssim$5	$\lesssim$15	$\lesssim$3	$\lesssim$3	$\simeq$7
	hit rate (GHz/cm^2)	$\lesssim$0.1		$\simeq$0.05	$\simeq$0.05	$\simeq$30
	σ_t (ns)	100	$\lesssim$0.05	500	25	$\lesssim$0.02
	NIEL fluence ($\times 10^{16} n_{eq}/cm^2$)		$\simeq$2			$\simeq$100
Track Detector	σ_{hit} (µm)			$\simeq$6	$\simeq$6	$\simeq$10
	hit rate (GHz/cm^2)					
	σ_t (ns)			$\lesssim$0.1	$\lesssim$0.1	$\lesssim$0.02
	NIEL fluence ($\times 10^{16} n_{eq}/cm^2$)					$\lesssim$1

Table 3.1: Selected tracking parameters of selected future experiments are shown. Table adapted from [2].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4 Depleted Monolithic Active Pixel Sensors within RD50

4.1 Depleted Monolithic Active Pixel Sensors for Future Tracking Sensors

Depleted Monolithic Active Pixel Sensors (DMAPS) are a promising option for future tracking sensors. Thus, they are mentioned as a key technology in the ECFA roadmap [2]. Various other technologies with different advantages exist as well; however, this thesis covers only DMAPS.
Conventional hybrid sensors implement readout electronics and the sensing diode on two different chips. While this approach leaves many possibilities to optimize the performance of both parts individually, it has some disadvantages which are overcome by monolithic sensors.
A monolithic approach includes readout electronics and the sensing diode within the same piece of silicon, which has various advantages. A comprehensive list is given here:

- No complicated and costly bump-bonding is needed, which is used in a hybrid approach to connect the sensor to a separate readout chip.
- Large-scale classical hybrid sensors are difficult to manufacture, and only a single supplier is capable of manufacturing modern sensors.
- Using just one layer of silicon instead of two reduces the material budget and thus multiple scattering.
- Commercial CMOS technology is used to manufacture them, which is very cheap, also for large-scale sensors, and offered by many foundries.

4.1.1 MAPS

Monolithic Active Pixel Sensors (MAPS) are based on camera chips and fabricated in a CMOS process to use standard CMOS electronics for readout. Figure 4.1 depicts a typical sensor profile. An n-type implant in a p-doped bulk serves as a collection electrode. Standard CMOS wafers use a cheap low ohmic substrate. Thus, the sensor can only be depleted in a small area under the collection electrode. The breakdown voltage V_{BD} is the limiting factor. The relative area of the depletion zone, which is the effective pixel area, is called *fill factor*. As many photons hit the sensor homogeneously spread in photon detectors, the fill factor can be rather small. However, this is a limiting factor for single particle detection,

as lots of the charge is not or only partly collected, which leads to inefficiency. Moreover, a charge collection by drift in the electric field of the depletion zone is beneficial, as it is much faster than the random diffusion of charge carriers in the sensor.
In order to mitigate this behavior, a thin epitaxial layer is grown on top of the bulk. This layer can have higher doping and resistance, which increases the depletion zone and makes the sensor more efficient. However, this layer can be grown very thin only, and this process is expensive, so the efficiency increase is limited.

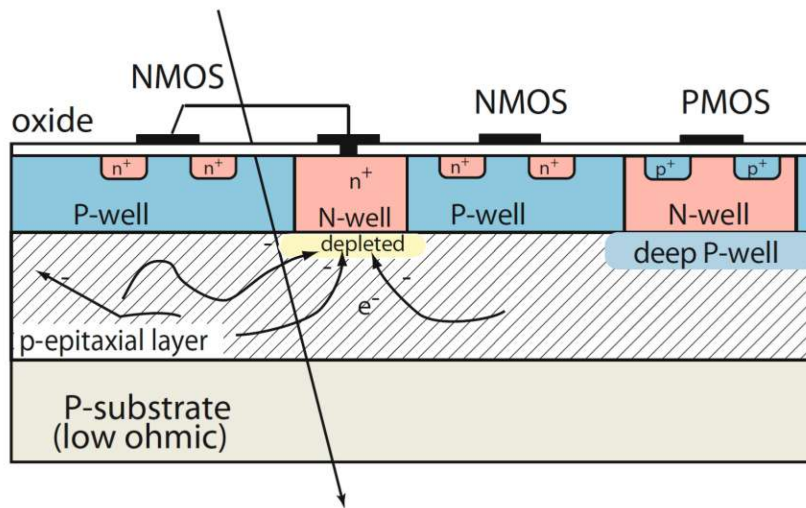


Figure 4.1: An n-type implant, called N-well, is used as a collection electrode in MAPS with p-doped bulk material. Other n-implants of the CMOS electronics must be additionally shielded as they concur with the collection electrode. Picture taken from [3].

An improvement to standard CMOS technology is the High Voltage CMOS (HVCMOS) technology, where the p-wells and n-wells of the CMOS electronics are embedded in a deep n-well, which is also used as a collection electrode. Such a sensor is depicted in figure 4.2. P-wells and n-wells can be used without a limitation here. Although a higher voltage can be applied before breakdown is reached due to the beneficial layout, the depletion zone is still relatively small, and charge collection is relatively slow.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

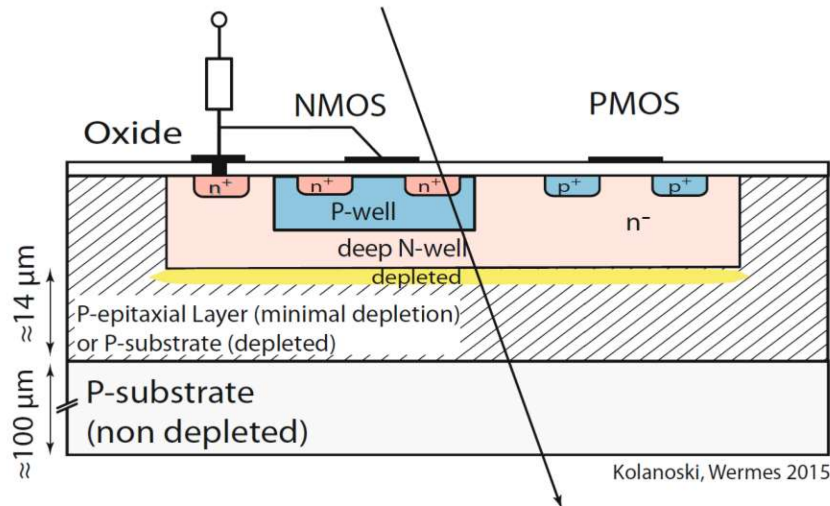


Figure 4.2: The collection electrode of HVMAPS is sufficiently larger than in standard MAPS detectors. However, the depletion zone is still rather thin due to the low resistivity of the substrate. Picture taken from [3].

4.1.2 DMAPS

Further performance improvement of MAPS sensors can be achieved by using a high-resistivity substrate. The depletion zone can grow much deeper for high-resistivity substrates as a higher bias voltage can be applied before the sensor breaks down. Thus, a bigger volume can be depleted, which leads to more efficient sensors. Due to the large depletion zone, these sensors are called Depleted MAPS (DMAPS). Such sensors are faster than standard MAPS, as the electrical field is larger, and charge collection is dominated by drift. If the sensor is thin enough, it is possible to fully deplete DMAPS before reaching the breakdown voltage.

Two different geometries, similar to MAPS, can be used. A small fill factor geometry is depicted in figure 4.3 while a large fill factor geometry is depicted in figure 4.4. The naming is analog to MAPS. The advantage of the large fill factor design is an increased radiation tolerance, as the strong, relatively linear and high electric field, as well as the large electrode decrease the mean drift distance, leading to less trapping of charges at defects.
The disadvantage of a large fill factor design is a high intrinsic capacitance between the deep N-well and the deep P-well, which leads to more noise, as discussed in detail, for example, in [10].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

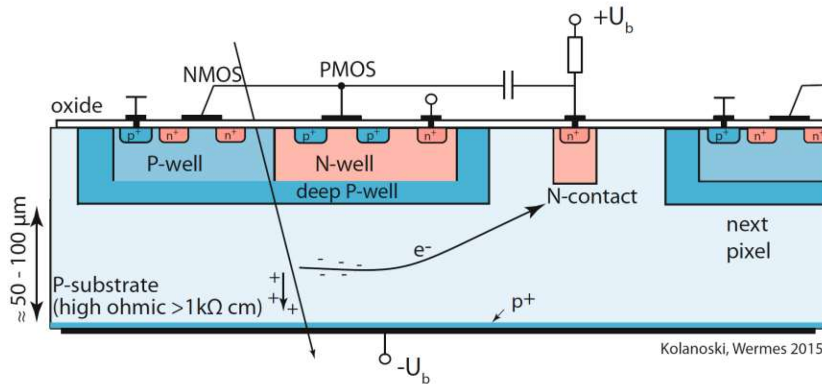


Figure 4.3: The drift path of a small fill factor design is rather high on average, leading to a dominant charge collection by thermal diffusion. However, process modifications, such as more complex implants, can decrease the drift path. Picture taken from [3].

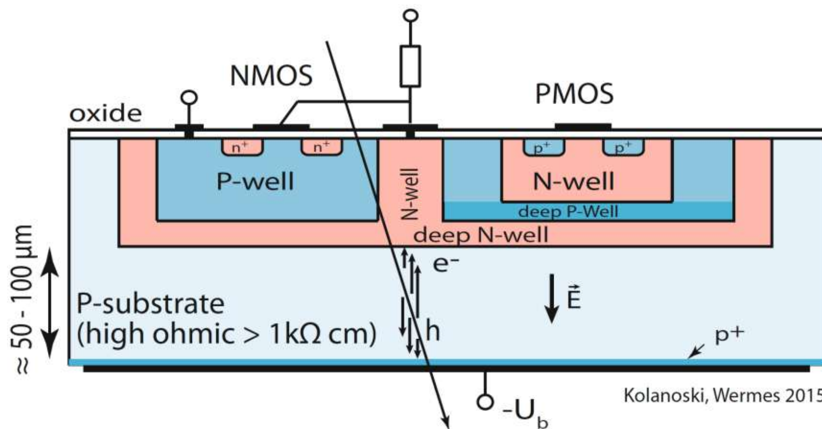


Figure 4.4: The drift path of charge carriers is short on average, leading to a dominant charge collection by drift in the electric field. Picture taken from [3].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.2 RD50 and the RD50-MPW series

The focus of the CERN-RD50 group is the development and characterization of radiation-hard semiconductors for very high luminosity collider experiments. The R&D program includes defect characterization for an in-depth understanding of lattice defects and their impact on the sensor, detector characterization to understand the impact of different process modifications, the study of new sensor structures such as 3D detectors, low-gain avalanche diodes (LGADs) and HVCMOS sensors, and full detector systems for test beam and irradiation campaigns.

The RD50-HVCMOS group is part of the R&D on new structures and designs and characterizes DMAPS based on an HVCMOS process. The R&D includes ASIC design, performance evaluation, and technology computer-aided design (TCAD) studies for simulating charge transport inside the detector and DAQ development. A 150 nm process from LFoundry[1], called LF15A, is used for all chips designed by the group. A large fill factor design is chosen, as the main interest of RD50 is radiation hardness. All designs are submitted using multi-project wafer (MPW) shuttle runs. Many designs from different contractors are produced on the same wafer in such a production to reduce costs, especially for small prototyping chips.

Three chips have been designed so far within the RD50-MPW series called RD50-MPW1, RD50-MPW2, and RD50-MPW3. The first iteration is not part of this thesis and is only introduced in the next section, as it is the baseline for all other iterations. The scope of this thesis covers the characterization of RD50-MPW2, the design of RD50-MPW3, and the characterization of RD50-MPW3.

4.2.1 Initial Design: RD50-MPW1

RD50-MPW1 is the very first design of the group and aims to understand the technology and get hands-on experience with the new sensor concept. The chip includes two analog active matrices with very limited digital readout and two front-end flavors. In addition, passive test structures are implemented, wholly decoupled from the matrices. Details about the design can be found in [11] and are not presented in this chapter. However, the design of the pixels in the successor chips is very similar and can is explained in the respective sections. The analog pixel front-end works quite well, and the digital pixel front-end can output two 8-bit timestamps for the leading and trailing edges. Moreover, a priority chain for the readout is developed, and a simple serializer for the output is also implemented. Both concepts are kept for future iterations; however, the next iteration does not have a digital readout, and the serializer is not implemented in this iteration.

Figure 4.5 is an essential output plot of the measurement campaign. The current-to-voltage (IV) curve of a single pixel, measured using one of the passive test structures, shows a leakage current (I_{leak}) of a few µA and a breakdown voltage of -56 V. However, a higher breakdown voltage is desired for greater radiation

[1] LFoundry S.r.l, www.lfoundry.com

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

hardness, and the leakage current per pixel is too high. In-depth TCAD studies have been conducted to understand where the issues arise. The results show that the spacing between the bias implant and the collection well is the main reason for the low V_{BD}, and a suboptimal guard ring scheme is the issue for the high I_{leak}.

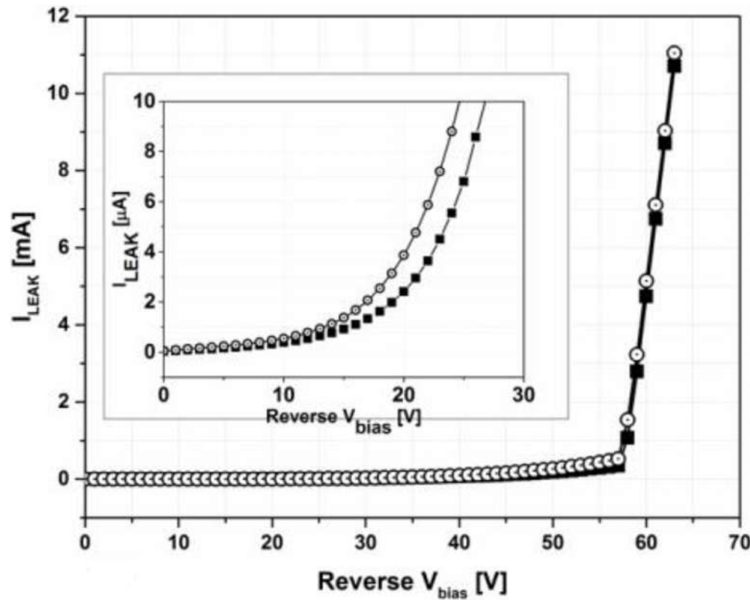


Figure 4.5: The IV curve of RD50-MPW1 shows a typical behavior as described in figure 2.5. While the breakdown voltage V_{BD} is within exceptions from the design, the leak current I_{leak} is much higher than expected. Picture taken from [11].

To sum up other measurements of RD50-MPW1, the analog pixel front-end was working quite well, while the digital readout had some serious issues. Thus, the decision was made to tackle the two issues separately. Thus, a small prototype chip, RD50-MPW2, with an analog-only matrix and passive test structures, was developed to optimize the low breakdown voltage and high leakage current. Measurements results are discussed in chapter 6.
The next iteration, RD50-MPW3, deals with the digital readout and incorporates lessons learned from the two previous iterations. The digital design is wholly renewed and discussed in chapter 7 while measurements can be found in chapter 8.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

5 Digital Design Implementation

The design of digital application-specific circuits (ASICs) has never been done at HEPHY. Thus a central part of the work for this thesis is to introduce the basic concepts of digital design and set up a digital design flow. This chapter is not a complete guide to learning digital design; instead, it gives an overview of the design flow and verification, including a short glimpse of industrial digital design.

5.1 Requirements of Digital Design for CMOS Fabrication

This small section briefly discusses what is needed in order to start fabricating a CMOS chip with a modern silicon foundry.

5.1.1 Design Entry: RTL Code and Constraints

A modern design process usually starts with some specification or at least an idea of what functionality the chip has to provide. Based on this document, the design of the hardware is written. This is done in a so-called hardware description language (HDL), which is more about describing what the chip should do than programming functions as in software. The most common HDLs are SystemVerilog and VHDL. VHDL is much more commonly used in Europe and not widespread in America, while SystemVerilog is equally used in both continents. Digital ASICs are typically written in SystemVerilog, while VHDL is more often used for programming FPGAs.

Implementation of the logic typically starts with a functional design, which describes all functionalities, logical blocks (macros), interfaces, and input-output pads (IOs). A rather short code is used, which can contain non-synthesizable objects such as time delays, in order to make this first implementation easily understandable. Such a code cannot be fabricated, as non-synthesizable objects do not have a representation in hardware. Thus, the next step is to go to the register transfer level (RTL) code, which contains only synthesizable code, so code with a representation in hardware. A time delay must be transformed into a counter, for instance. This step is where hardware design differs a lot from software design. Registers are updated in parallel, (software) loops do not exist, and many more typical software constructs, such as classes and other high-level objects, cannot be used (efficiently) here. The pendant to the compiler in software is the synthesizer in hardware, and it does depend to a certain extent on the synthesizer which objects can be used and not used.

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

In parallel to the RTL implementation, one has to think about the interfaces to the outer world. The outputs of the ASICs have to know what load they have to drive, and the inputs need to know the input capacitance together with the maximum and minimum allowed delays. Moreover, internal clock domains and other constraints (such as clock-gated blocks) must be defined. This is done by writing a constraints file in the Synopsis Design Constraints (SDC) format. The synthesizer needs this information to choose cells that can meet these constraints, for example, are strong enough to drive a specific load at an output.

5.1.2 Process Entry: Physical Information, Timing Libraries and Design Rules

Modern silicon foundries offer a couple of different fabrication processes, which differ significantly in feature size, wafer production, implantation, and metal lines for signal routing. Wafers can be produced with the Czochralski or the float zone technique, where the latter typically has fewer defects and can only be produced in smaller sizes. The implant layers are the most crucial choice for designing a CMOS sensor. A deep n-type implant is needed in the RD50-MPW series to create a depletion zone for signal generation, as shown in figure 6.2. Not all processes support such a layer, as it is not needed for most commercial applications. Furthermore, enough metal layers must be available on top of the silicon wafer to route all the signals. Different metals with different properties, such as resistivity, are used in different processes. High-quality metals with rather high resistance are typically used for the bottom metals and signal routing. The top metals are thicker and use low resistivity metal for routing power.

All of this information is delivered within a so-called Process Design Kit (PDK), which contains all of the global physical information mentioned above and the physical and timing information of so-called standard cells. Those standard cells are the provided logic gates, storage elements, clocking cells and other cells dependent on the PDK. In digital design, these gates are composed of certain transistors, where some variants are available, but transistors cannot be chosen arbitrarily as in analog design. The only information needed for the design is the physical layout of these cells and the timing properties, such as propagation delay of the cell, maximum driver capacity, and setup- and hold- time. This timing information is stored in timing libraries, typically in the liberty format (.lib). The big advantage of digital design is that those timing parameters can easily be looked up in these libraries for each standard cell. In contrast, in analog design, one has to run a transistor-level analog simulation and look at the transient, which is much more time-consuming.

Each process has limitations, as all manufactured structures can have only a specific minimum size due to instruments used to fabricate implants or metal lines. Thus, there are rules in every process for the minimal size of wells, the minimum spacing between metal lines, a minimum metal density per layer to avoid thermal

TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

stress and many more depending on the fabrication process. The foundry provides those so-called Design Rules Constraints (DRCs), which modern design tools can check for during implementation.

5.2 Fundamentals of Digital Design Implementation

This section aims to give a somewhat detailed overview of the steps needed to implement a digital design, together with used software tools. Moreover, verification is explained, and a comparison to industrial implementation is given.

5.2.1 RTL-to-GDSII Flow

The chain needed to implement a RTL-level design into a silicon wafer is referred to as RTL-to-GDSII flow. A GDSII (Graphic Database System 2) file is the final layout file, including the position and dimension of all transistors, storage elements, and metal connections of a design given to the foundry for fabrication. GDSII is slowly but steadily replaced by the newer and much smaller OASIS (Open Artwork System Interchange Standard) format; however, the flow is still named after the old format nowadays.

Specification and Design Entry

Figure 5.1 shows an overview of all design steps and how to verify the design after certain steps. Design specification and RTL coding are mentioned in the previous section. During and after RTL coding, a functional simulation and verification need to be done. A so-called testbench (which does not need to be synthesizable) needs to be written to test all design functionalities. It is important to use only I/O signals and no internal signal for these tests, so one can re-use the same tests to verify later design steps. The scope of these tests can differ a lot between industry and research. Functional verification using high-level software such as UVM (Universal Verification Methodology) integrating many tests is the standard in the industry. At the same time, academic research often risks submitting a chip with less verification in order to save development time. The choice of how much functional verification is needed is based on the complexity of the design. It is always a trade-off between time (which includes money) and the chip's reliability.

Synthesis

Synthesis converts an RTL-level design into standard cells consisting of basic logic gates and storage elements. A timing constraint file is needed in order to calculate the timing properties of all I/Os while timing parameters from logic gates are read from the liberty file of the PDK.
Synthesis is done completely by the tool and consists of synthesis to generic gates,

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

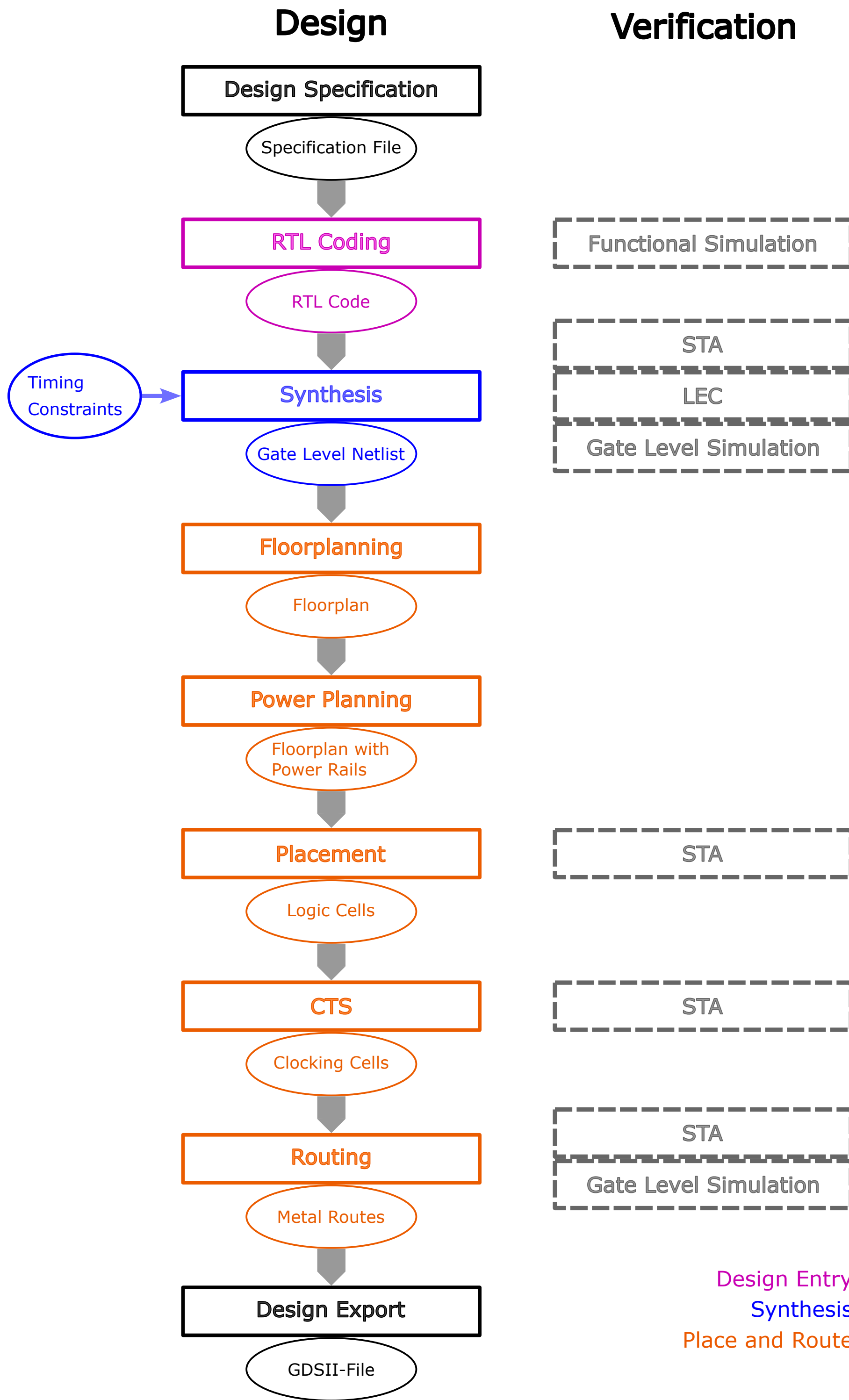


Figure 5.1: Overview of the RTL-to-GDSII flow. Implementation steps are given in rectangular shapes, while the output of the respective step is shown in circles. Where applicable, verification of the design step is also mentioned. The shown flow is a trade-off between design time and verification time.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek, WIEN Your knowledge hub

mapping to gates from the PDK, and optimization. In the first step, the RTL code is converted to a gate-level netlist, representing the logic only using low-level logic gates instead of higher-level abstraction, such as state machines typically used in the RTL design. Thus, a gate-level netlist is often not human-readable anymore as it is just a very long list of gates.
As a second step, mapping to real gates from the process is performed. This is the first step, where information from the foundry is needed (liberty files) together with the constraints files. Generic gates are replaced by real gates, considering the driving capabilities of connected cells. A coarse wire load model (WLM) estimates the timing parameters (resistance and capacitance) of the wire connecting two cells. Those WLMs estimate the wire length based on the number of logic cells of the design and the overall design space, considering maximum utilization. Parameters of the real wire replace these models after routing, also considering parasitic capacitance to neighboring wires and other properties. However, those values are not available at this step of the design.
The third step is optional but highly recommended as it might greatly reduce the design size. Logic optimization includes many algorithms for logic pruning, gate sharing, and other timing optimization techniques, which all focus on reducing area, time, or power consumption while keeping the same constraints.
Various verification steps are highly recommended after synthesis. A logic equivalent check (LEC) should be done as the optimization step prunes, removes, and adds some gates. Various tools exist for this, and all of them require a *golden design*, typically the RTL code to which the current design is compared. Those tools run a series of tests and check if both designs produce equal logic output when given the same input. Static timing analysis (STA) generates timing paths from all timing start points (inputs cells and the clock input of flip-flops) to all timing endpoints (output cells and data inputs of flip-flops) and sums up all timing parameters of the path. The time of all logic cells is looked up in the liberty file and summed up with the wire delay calculated from the WLMs. Together with the signals' rise- and fall time, the sum is compared to the maximum allowed time of the path given by the clock period. These checks are typically done for the setup- and hold- case. Static timing analysis is done automatically by the design tools and repeated after most steps. Lastly, one should rerun the functional analysis with the gate-level netlist, including time delays. The latter can be extracted after synthesis and are typically given in the standard delay format (.sdf files).
After those steps, all gates and storage elements are fixed and can be placed.

Placement

Floorplanning, power planning and logic cell placement are often referred to as placement steps.
Floorplanning includes a definition of the size of the chip, which is usually given by mechanical constraints of the whole detector design. Moreover, placement constraints for I/O cells, which typically have to be placed in a certain area given by external constraints for bonding, and additional constraints for routing and placement are part of the floorplanning step. Typically, a placement constraint is set for certain time-critical logic blocks to place them close to their I/Os, which

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

makes the job for the automatic placement tool easier. A logic row represents connected pwell and nwell implants, which have the height of standard cells. All standard cells have the same height given by the PDK and the same thickness for pwell and nwell implants, but the width is different according to the number of transistors needed for the gate. The PDK defines a minimum width, and all cells must be multiples of this width to fill the whole area with wells.
Power planning consists of defining the main power routing lines, which are usually thick metal lines on the topmost layers of the design. If the design has multiple power domains, they must also be defined during this step. Detailed power routing, which connects the main power lines from the top metal to the silicon implants, is usually done automatically by the tool.
At the end of this phase, logic cells are placed onto the pre-defined rows. During this step, a course routing is done, which inserts metal lines to the standard cells, but they are not yet connected to the implants. This is needed to estimate better the wire lengths and their timing properties for more accurate timing analysis. Notably, the placement does leave some space for clocking cells, which can be controlled by setting a particular placement utilization. Typically, 70-80% are utilized at this stage.
After placement, a simple verification by checking the static timing analysis with updated WLMs is recommended. The routing step can still vary parameters to improve timing paths, but the capability is limited. If there are significant violations of static timing parameters, one has to go back to the floorplanning phase or even the RTL design.

Routing

Clock tree synthesis (CTS) and detailed routing are usually referred to as routing steps. CTS starts with the generation of the clock tree, considering a possible maximum latency constraint. As a single clock buffer can drive only a certain amount of gates, a clock tree must be implemented, so the clock arrives simultaneously at every flip-flop. To be precise, one wants the clock tree to differ by a few picoseconds between different flip-flops, as it would produce an enormous spike in the power consumption to flip all flip-flops simultaneously. This difference in time is then considered in further static timing analysis checks. Buffers and inverters are placed to handle the delay. Moreover, clock-gating cells are also placed during this step, if defined. The clock is the most critical net, so it is routed first.
After CTS, signal and power lines are routed during detailed routing. In this procedure, the design is split into small areas where the details router routes signal lines to connect all cells. If needed, the router can shift standard cells slightly or replace them with faster (but usually more power-consuming) cells to meet timing. Before design export, a few more steps have to be done. The most important one is placing decoupling capacitors to reduce power spikes and filler cells containing pwell and nwell implants without gates. Those cells are placed in holes, still left between the standard cells, to connect all pwells and nwells of a standard row. Placement of antenna cells and metal fill to reduce mechanical stress are further steps.
Lastly, a detailed extraction of timing parameters, the RC-extraction, is done to

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

compute accurate timing parameters for all wires, including coupling capacitance to neighboring wires, which replace WLMs for final static timing analysis. It is recommended to repeat gate-level simulation with timing parameters after this extraction. RC extraction is typically the most time-consuming step in the full implementation. However, the precision of modern tools is impressive and needed for high-performance chips. This step is typically referred to as Sign-Off analysis and is the last step before handing the digital implementation to the chip-level verification and production explained in section 5.2.3.

5.2.2 Software Tools

The software toolchain used in this thesis is the Cadence[1] digital implementation flow. Thus only Cadence tools are mentioned here. All hardware design software vendors often change their tools' functionality or merge into the same tools, so this list needs to be updated quickly. It is recommended to check the RTL-to-GDSII flow of the vendor, which typically states the tools used for each step.

Figure 5.2 gives an overview of the Cadence tools for each step. Many text editors and IDEs support SystemVerilog and VHDL code nowadays, so Cadence only provides a powerful elaboration and simulation tool called *Xcelium* (and no text editor). The synthesis tool is called *Genus* and works with Xcelium to simulate a synthesized netlist. To my knowledge, Cadence plans to integrate the Genus tool into Xcelium entirely in 2023. The logic equivalent checker is called *Conformal* and still runs as a separate tool but can be called during the Xcelium compilation flow for easy integration into the flow. The powerful physical implementation tool is called *Innovus* and includes all other functionalities from Floorplanning to detailed routing. *Quantus*, a separate tool fully integrated into Innovus, is used for Sign-Off level RC parameter extraction.

5.2.3 Chip-level Implementation, Verification and Production

The RTL-to-GDSII flow, as described in 5.2.1 is about block-level design. At this level, a single, fully functional design block is implemented. However, a modern ASIC typically has more than one block. All of these blocks work independently from each other and only have a few interconnections. These blocks can be digital or analog. Moreover, I/O pads have to be implemented and connected to the respective blocks. All of this is the purpose of chip-level implementation. It can be done using analog or digital tools and is referred to as analog-on-top or digital-on-top design. Analog-on-top designs require manual connection of all blocks and I/O pads, while digital-on-top designs can do this automatically. However, analog designs have to be ported into the digital domain, which is much more difficult than vice versa, as one has to calculate the digital timing libraries.

In both cases, foundries often ask for additional steps after routing before the design can be produced. This typically includes a special metal structure at

[1] Cadence Design Systems, https://www.cadence.com/

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

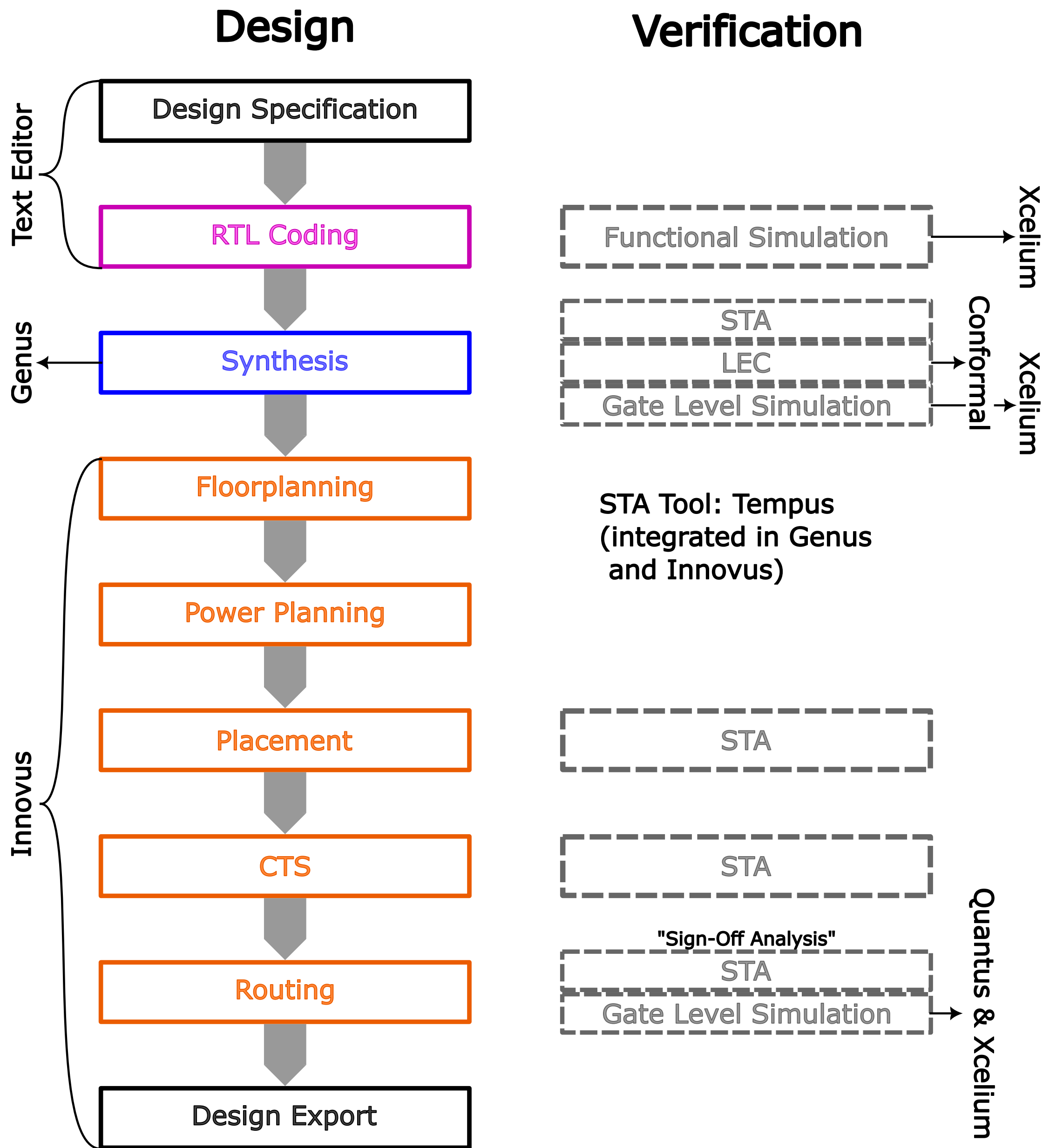


Figure 5.2: Cadence software tools are given in this figure for various design steps. All of these tools are highly integrated, and Cadence provides a common GUI, called *Stylus-UI* for all of them.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

the chip edges and metal fill to reach a certain metal density in every layer for mechanical stability, isolation trenches between wells, passivation layers, and more process-specific needs. Some might be done by the foundry, while others have to be done by the user.
At the very end, a final verification step is done, which includes two more essential checks, the Layout Versus Schematic (LVS) and the Design Rule Check (DRC).

Layout versus Schematic

The LVS check is the analog pendant to the purely digital LEC. This check aims to compare the routed design with the original schematic. This is especially needed for analog designs as routes are drawn manually, and one must ensure that every wire is connected properly. LVS is run for every analog block during block-level design and has to be repeated on the chip level as well.
Simulation at the chip level is often not feasible for ASICs with much functionality, as analog simulations require solving huge linear equation systems. Thus, simulations and LVS are run for a simple block where such a simulation is feasible to guarantee functionality, and the LVS on the chip-level guarantees connections between blocks.
LVS can be completely replaced by LEC and full-chip simulation for a purely digital design. However, as mentioned, any possible analog cell must be ported to the digital domain. This means running analog simulations for different input signals and writing down the output timing characteristics in digital liberty files for each analog block. Cadence provides a tool called *Liberate* for this, which needs a separate license.

Design Rule Check

A DRC aims to check if the foundry can produce the ASIC. The foundry does provide a list of requirements within the PDK, including minimal and maximal width of implants and metals, spacing between metals and vias, metal density requirements, maximum length of metal lines per layer, and more rules to guarantee producibility. These rules guarantee that the design can be produced and heavily depend on the process. Luckily, modern digital and analog design tools can read and follow these requirements during the placement and routing steps or at least issue a warning if a particular part violates design rules. The foundry will run DRC on the submitted design. However, it is advised for the user to run chip-level DRC before tape-out as well. It is possible to submit a design violating these rules. Negotiations and iterations with the foundry are often needed to waive some design rules, taking the risk that something might fail.

5.2.4 Requirements Engineering and Industrial Verification

The industry has different demands on ASIC design than research. Typically, the industry has much more human resources and money available, which are often critical sources in academic research. As research often operates completely in the prototype phase, it is possible to neglect some industry verification standards to

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

save resources. On the contrary, the industry usually targets a bigger and more costly production phase where a mistake in the design is not affordable; thus, verification plays a much more crucial role than in research.

This section tries to compare both implementation flows. The industry does implement quite a lot of standards, and some of the principles might be useful in research as well. Two of the most complex industrial design and verification standards for electronics are the space and aviation standards, the ECSS-Q-ST-60-02C [12] and DO-254 [13]; most of the concepts presented in this chapter are taken from them. Both of them implement a strict design and verification flow, which is not presented in detail here because of its complexity. Instead, a fundamental concept of ASIC design requirements and a framework for verification are introduced. Both are used or recommended in these standards.

Requirements Engineering

The International Requirements Engineering Board (IREB) defines requirement engineering as *"The systematic and disciplined approach to the specification and management of requirements with the goal of understanding the stakeholders' desires and needs and minimizing the risk of delivering a system that does not meet these desires and needs."* in [14], page 11. Industry standards such as the ECSS or DO-254 define how requirement engineering is performed for a specific process. A requirement has to be well defined, which means it should have the following characteristics, which are taken and adapted from [15]:

- Completeness: A requirement should sufficiently describe the objective without additional information.
- Correctness: A requirement should accurately describe the objective.
- Unambiguity: A requirement should be written so that it cannot be interpreted in multiple ways.
- Singularity: A requirement should only state a single certain objective.
- Feasibility: A requirement should be realizable within the system.
- Verifiable: A requirement should be provable to the customer.
- Quantifiable: A technical requirement should be quantifiable to enhance verifiability.

Using the above definition in an electronics design context, one can group requirements at different levels. A requirement can be defined by the end user, who might not even know the implementation details. Moreover, it can be defined for the whole system level (example given: overall power consumption or mechanical restrictions), a subsystem like a single ASIC (example given: input clock frequency, the maximum allowed power consumption for this module), or a single

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

component (example given: States of a certain finite state machine, bus specification).
There are various types of requirements in electronics design which all need a different form of verification. An example of types is listed below:

- Functional: A requirement stating what a specific component should deliver as (logic) output.
- Performance: A requirement stating maximum throughput values
- Interface: A requirement defining an interface between two components or the outside world
- Environmental: A requirement defining external conditions such as temperature, humidity, or radiation level
- Quality: A requirement defining safety or reliability objectives.
- Usability: A requirement defining the purpose of a certain component is used.
- Physical: A requirement stating mass, dimension, or power consumption restrictions

All requirements are written into a requirements specification document, from which a test plan should be derived. A test plan treats each requirement separately and defines which tests need to be performed in order to validate it. The type of test hugely depends on the type of requirement. Requirements and tests should be traceable, so it must be clear from which level the requirement is defined and from which other requirements it might originate. Many low-level requirements form one higher-level requirement, which is automatically fulfilled if all lower-level ones are fulfilled. Various requirement management tools exist in order to make the relationships easier manageable.

ASIC Verification using UVM

A verification framework for in-depth ASIC implementation has to fulfill various requirements for efficient and reliable verification. Firstly, advanced logging support is beneficial, as reading a terminal output is inefficient. Secondly, checking algorithms, such as scoreboards, are useful for a quick decision if a certain condition or requirement is met. Thirdly, a solid software development kit (SDK), including randomization, file I/O, and high-level data constructs, is needed to test requirements quickly. Lastly, it should support a variety of simulators and platforms. Considering all of them, an object-oriented language with high-level abstractions makes the most sense.

The Universal Verification Methodology (UVM) uses the object-oriented SystemVerilog language and provides many packages fulfilling all the abovementioned needs. It is the most complete (and thus complex) verification framework today,

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

and many other frameworks are based on the ideas of UVM. Moreover, as the name suggests, it defines a methodology where reusability is the key goal, so UVM is more than just a collection of abstract layers.

No UVM introduction is given here, as it would already fill a whole university lecture, and enough literature is available[1]. However, the most basic concept of UVM is shortly explained here, so one can get an expression on verification is done. UVM operates on the so-called Transaction Level Model (TLM), where a transaction is modeled as a class object for easy reusability and scalability. Thus, we must convert our RTL model, which is based on modules, into the class-based TLM world. The critical component of UVM doing this is called *agent*. An agent takes care of the low-level bit-banging of a single interface of the ASIC, thus handling all the timing constraints of the protocol. A driver, monitor, or both are implemented within an agent based on the interface direction (input, output, inout). A monitor gets all signals from the outputs of the ASIC and converts them into a class object, respecting interface timing, while a driver does the exact opposite. After implementing an agent for every interface, one does not need to care about signal timing and can fully operate in the object-based TLM world. A testbench generates a sequence of test objects given to the agents while it looks at the monitor outputs and compares the output with the expected result. A reference model must be implemented (and validated!) to check the results correctness.

5.3 Conclusion

Digital design and verification of complex, large-scale ASICs have become more and more challenging tasks in the modern world. Sophisticated verification frameworks are already a standard in the industry but have yet to be in research. However, the complexity of ASICs used in research steadily increases, industry-level requirements will have to be written, and complex verification algorithms will have to be used in research in the near future. Getting familiar with a typical design flow and knowing industry standards is essential for digital design nowadays. Thus, this chapter is utterly needed in a thesis about digital design.

[1] e.g., https://www.chipverify.com/uvm/uvm-tutorial

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

6 DAQ and Testbeam for RD50-MPW2

As the first prototype of the RD50-MPW series suffered from high leakage current and low breakdown voltage, as mentioned in 4.2.1, a second prototype is designed to tackle these issues and thus focuses on analog design. This thesis aims to develop a complete digital readout chain, including hardware, firmware, and software. It includes configuration and readout of the chip and is proper for laboratory evaluation and test beam integration.

6.1 Introduction to RD50-MPW2

The RD50-MPW2 is the second prototype chip with many different test structures focusing only on analog features. To total size of the chip is 2.1 mm × 3.2 mm and a thickness of 300 µm. Figure 6.1 shows a picture of it. The main interest of this thesis is in the analog-only pixel matrix. Various substrate resistivities of 10 Ω·cm, 200-500 Ω·cm, 1.9 kΩ·cm and 3 kΩ·cm are available for an intense radiation study which is also presented in this thesis.

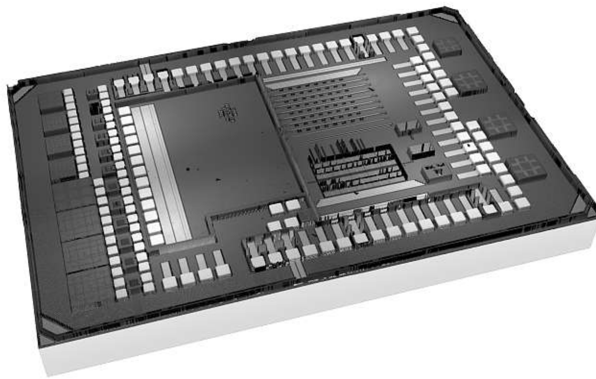


Figure 6.1: 3D Picture of RD50-MPW2 taken with a confocal microscope.

6.1.1 Analog-only Pixel Matrix

The analog pixel matrix consists of eight-by-eight pixels with a pitch of 60 µm by 60 µm. A cross-section of a pixel cell is shown in figure 6.2. A deep n-type implant called DNWELL serves as collection electrode. Electronics are implemented inside this DNWELL and thus shielded from the applied high voltage, which is applied using p-type implants for biasing from the top side of the chip. No backside processing and substrate contact are needed for this biasing scheme. The

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

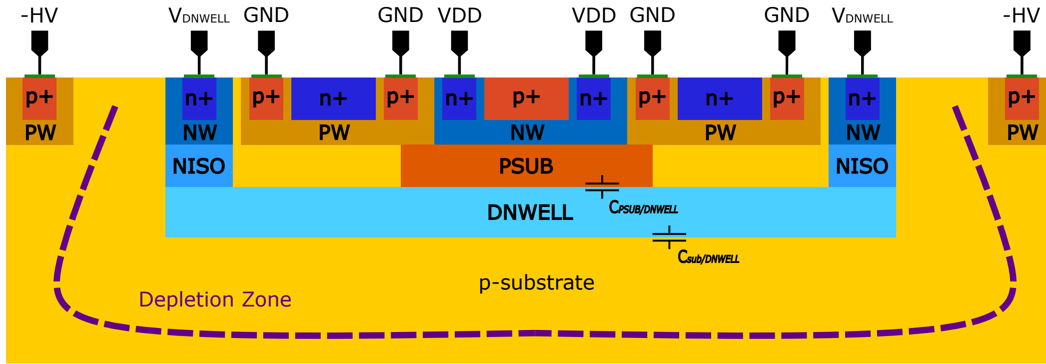
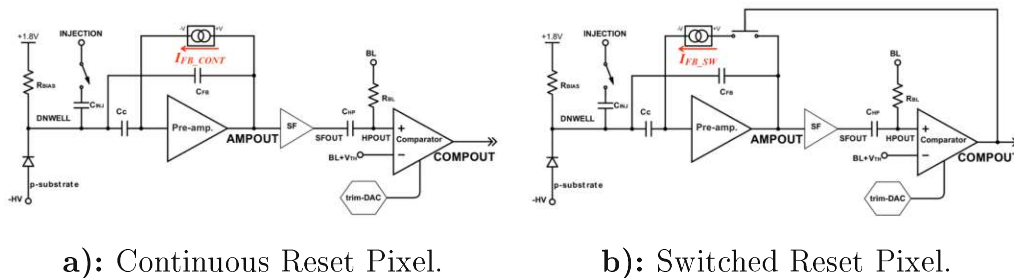


Figure 6.2: A cross-section of a pixel showing different implants and the depleted volume where the charge is collected is shown. The sensor capacitance adds up from two principal components ($C_{PSUB/DNWELL}$ and $C_{sub/DNWELL}$). Simulations have shown that both have about the same capacitance.

pixel is designed using a large fill factor design as introduced in section 4.1.2 to put a focus on radiation hardness. The charge is collected using the $C_{sub/DNWELL}$ while $C_{PSUB/DNWELL}$ is an unwanted capacitance from the sensor perspective but needed to separate the DNWELL from the n-wells of the electronics.

Analog electronics inside the pixel consist of a charge-sensitive pre-amplifier (CSA), a source follower buffer (SF), and a comparator, as depicted in figure 6.3. An injection circuit, an analog output (AMPOUT), a digital output (COMPOUT), and a trimming DAC (trim-DAC) to fine-tune the comparator threshold are implemented for each pixel. Two different pixel flavors are implemented: the



a): Continuous Reset Pixel. **b):** Switched Reset Pixel.

Figure 6.3: In-pixel electronics for the two-pixel flavors in RD50-MPW2 are shown. The current source of the switched reset pixels (6.3b) is much stronger ($I_{FB_SW} > I_{FB_CONT}$) in order to reset the amplifier faster. Picture taken from [16].

so-called continuous reset pixel and the switched reset pixels. Figure 6.3a shows the continuous reset pixel, which resets the amplifier relatively slowly and thus has a readout signal long enough to be used as coarse charge measurement. On the contrary, the switched resets pixel shown in figure 6.3b is reset relatively fast and is thus much more suitable for spatial tracking.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

6.1.2 Other Features of RD50-MPW2

RD50-MPW2 comprises other features not covered by this thesis but are mentioned for completeness. Those structures are depicted in figure 6.4 and are listed below:

1) + 6) passive test structures to study different pixel layouts

2) a bandgap voltage reference

3) a single event upset tolerant memory to study radiation hardness

4) the pixel matrix

5) an analog buffer

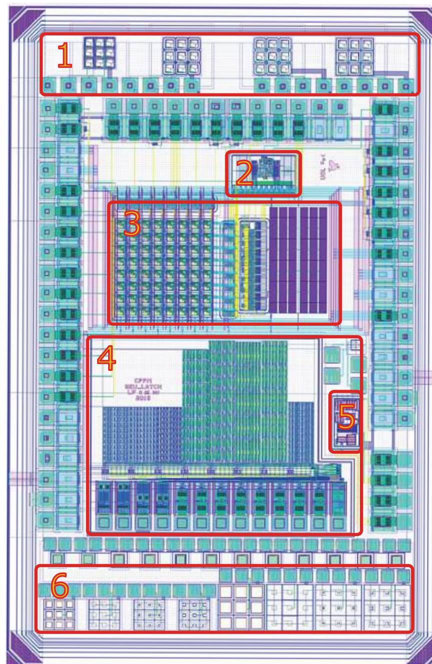


Figure 6.4: An overview of all elements implemented in RD50-MPW2 is depicted.

Notably, the test structures implement different corner shapes for the pixels and can be measured easily with a probe station and needles. Measurements shown in section 6.3.1 use these passive structures. The analog buffer is needed to drive the chip's analog output, and the bandgap reference is a test for a future in-chip voltage reference.

6.2 Readout System of RD50-MPW2

The readout system is based on the Caribou framework [17]. Klemens Flöckner wrote a master's thesis focused ond data acquisition as part of this project. The

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Caribou implementation can be found in chapter 6 of his work [18] and is not repeated here. The entire readout system covering hardware, firmware, and software is primarily developed by the RD50 group at HEPHY.

6.2.1 Hardware

PCBs as part of Caribou

Caribou is based on the Xilinx ZC706 System on Chip (SoC) board which is commercially available from the vendor. The control and readout board (CaR) is developed by the Caribou group. It implements a lot of different integrated circuits (ICs), which are typically needed in order to control, power and readout a pixel sensor. The so-called chipboard is part of the Caribou hardware and is a chip-specific board that carries the sensor. It has been developed by IFIC Valencia[1] and consists of all connections between the sensor and the readout board, including signals lines and power. Moreover, basic filters, signal converters and direct outputs for measurements with a scope are implemented. Figure 6.5 shows a picture of all parts.

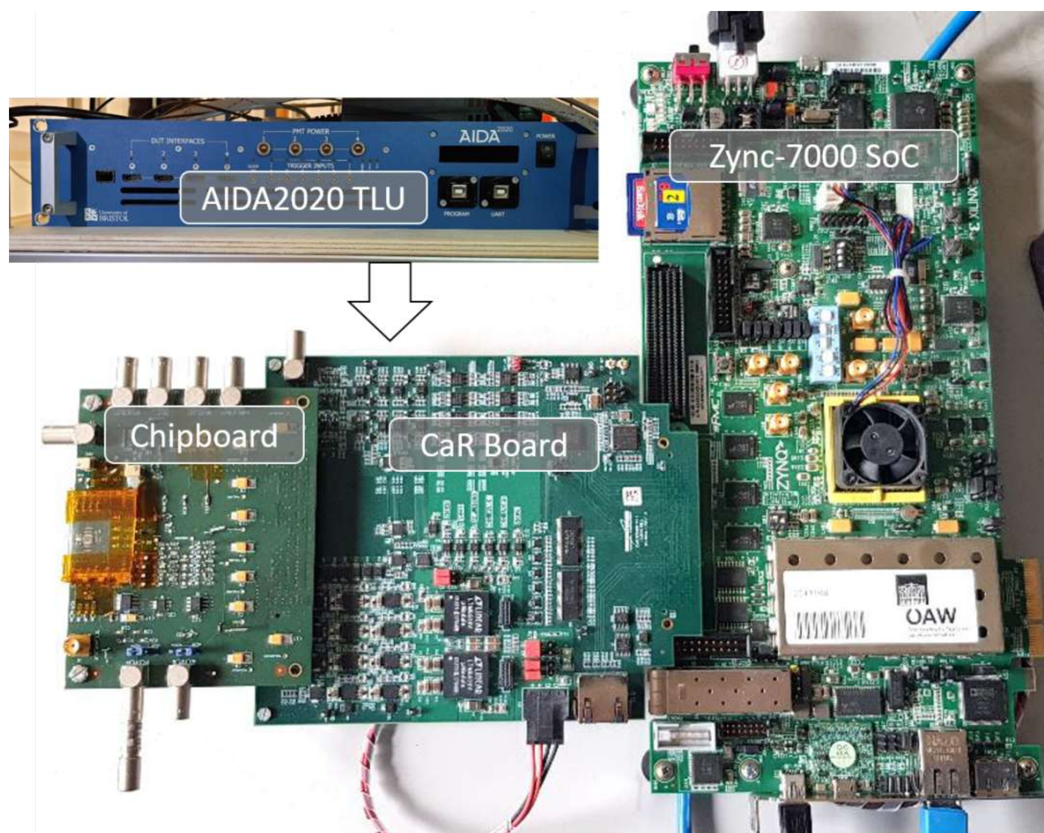


Figure 6.5: Hardware for reading out RD50-MPW2 is shown. Data is stored on the SD card on the Zync evaluation board.

[1]Instituto de Física Corpuscular, Valencia, Spain

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Trigger Logic Unit AIDA2020

In a test beam environment, the device under test, the RD50-MPW2, is usually operated together with other detectors for reference. Thus, in order to synchronize and trigger the readout of different detectors, a trigger-logic unit (TLU) is needed. The RD50-CMOS group decided to use the widely used AIDA2020 TLU. Details of it can be found in [19].

6.2.2 Firmware

RD50-MPW2 does not contain any digital signal processing; thus, it is implemented entirely in the FPGA of the SoC. The digital output signal of the sensor is sampled by a 200 MHz clock for a charge measurement. In parallel, an asynchronous hit counter is implemented. The firmware consists of two different versions, a standalone version meant for laboratory evaluation only and a test beam version, which includes a mechanism to synchronize the readout of the sensor with the TLU. Details of both versions have been published in paper [20]. It is worth mentioning that the FPGA has to take over quite some tasks to read out RD50-MPW2 . This framework is the first full implementation of a sensor from the RD50-MPW series with a digital readout capable of synchronizing with other detectors.

6.2.3 Software

DAQ Software: Peary

Peary is the software part of Caribou, which implements an interactive shell to control and read out the sensor. It uses a hardware abstraction layer (HAL) which includes all low-level drivers to devices from the CaR board and some components of the SoC which are needed to configure the CaR board. Peary gives an easy-to-use and quick starting point for developing DAQ software.

On the one hand, the sensor is integrated into peary; thus, it needs to be defined as a peary device, and primary drivers to read and write to the sensor are implemented. On the other hand, software routines for readout have been defined. The most important ones are a calibration routine to determine the ideal trim-DAC value using the injection capability of RD50-MPW2, an open shutter measurement to active a pixel and readout all signals it gets, which is handy for laboratory evaluation and a routine to synchronize with the TLU and issue a readout of a single pixel, which is needed for test beam measurements. Details of these tests can be found in [18].

DAQ Control Framework: EUDAQ2

The software framework used to control the DAQ chain, especially during test beam measurements, is called EUDAQ2. It implements a standardized way to communicate with many devices connected to it, configure them and start a measurement.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Detailed information can be found in [21], where a quick summary of essential tasks is given here. The fundamental structure of EUDAQ2 is a finite state machine, which has four primary states called *Initialized*, *Configured*, *Running*, and *Stopped*, which are rather self-descriptive in what they do. These basic commands are sent via the GUI module, the so-called *Run Control*. All other modules connect to this Run Control. Essential modules needed for every measurement are so-called *Producers* to control a specific device and *Event Collectors*, which store collected data on disk.

Analysis Framework: Corryvreckan

The modular and widely used tracking analysis framework Corryvreckan [22] is used to analyze test beam data. It targets analysis of small to medium-scale test beam setups based on pixel sensors and simultaneously supports 3D and 4D reconstruction. The most important features are listed below (in order). These are also the ones used in the analysis of all test beam data throughout this thesis.

1. The *EventLoader* loads data from all devices into the framework.
2. The *EventDefinition* module is needed to define an event in Corryvreckan. Per definition, an event is given by a start and an end time.
3. The *Clustering* module finds pixel clusters.
4. The *Correlation* module correlates hits from different detector planes. This information is used for a course alignment of all active detector planes.
5. The *Prealignment* module shifts all correlations to 0 to get a course-aligned position of each detector plane.
6. The *Alignment* module performs track-based alignment to fine-align all planes.
7. The *DUT-Association* module tries to associate hits in the device under test (DUT) with tracks from the reference detectors.
8. Various analysis modules are used to analyze spatial resolution, efficiency, and other characteristics of the DUT and reference detectors.

6.3 Test Campaign of RD50-MPW2

The analog performance of RD50-MPW2 was tested intensively. The focus is on studies for radiation hardness, as this is the main interest of RD50. However, tracking studies to get spatial resolution and timing studies are also tried.

6.3.1 Radiation Hardness Tests

An irradiation campaign is conducted where RD50-MPW2 sensors have been irradiated up to $2 * 10^{15}$ N_{eq}/cm^2 using the nuclear reactor at JSI-RIC in Ljubljana[2]. However, HEPHY only tested sensors with fluences up to $1 * 10^{14}$ N_{eq}/cm^2.

[2]Institute Jožef Stefan - Reaktorski Infrastrukturni Center

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

At first, simple IV curves using the test structures are recorded to measure leakage current and breakdown voltage V_{BD}. The measurements for the unirradiated sensors can be found in [23] and prove that RD50-MPW2 has a V_{BD} two orders higher than RD50-MPW1 due to the increased spacing between the bias and ground implants and due to an improved guard-ring. IV-curves of irradiated samples are shown in figure 6.6 and agree with measurements for unirradiated samples. The IV-curves show an increase of I_{LEAK} by around an order of magnitude when going up with fluence by order of magnitude. It is expected that V_{BD} does not depend on fluence, while the leakage current does.

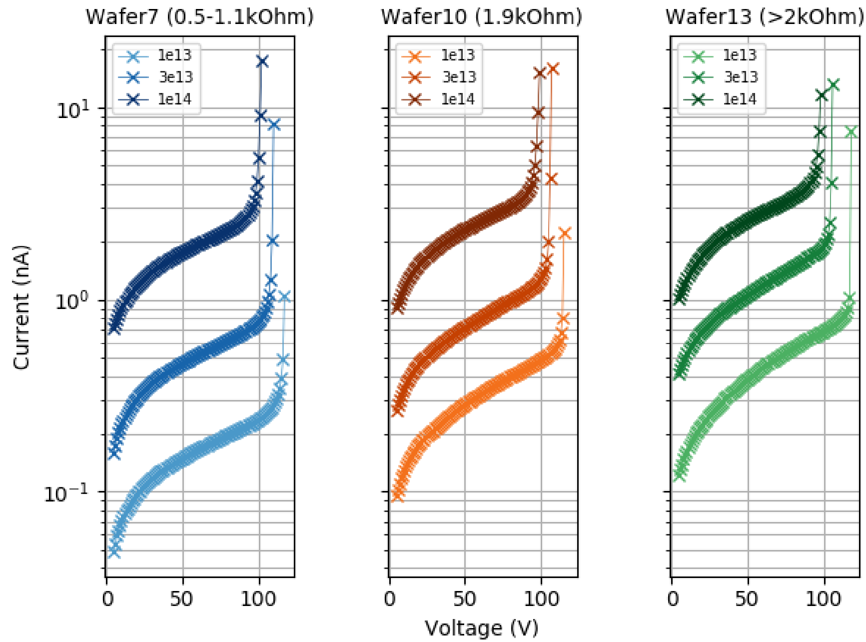
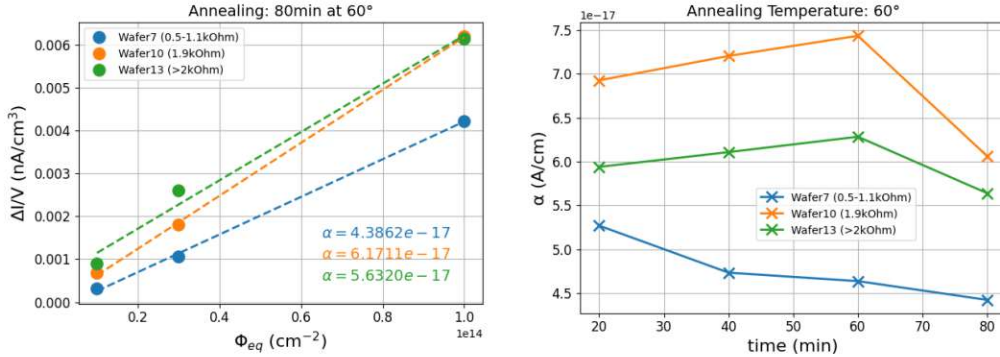


Figure 6.6: IV curves of irradiated RD50-MPW2 samples from wafers of three different resistivities for fluences of $1 * 10^{13}$ to $1 * 10^{14}$ N_{eq}/cm^2 are depicted.

As shown in section 2.3.1, the current after irradiation depends on fluence by a factor α, following equation 2.3. According to [6], α should be in the order of 5×10^{-17} for different types of silicon after an annealing time of 80 min at $60°C$. This factor has been measured for all available substrate resistivities and fluences by measuring the current difference before and after irradiation and using formula 2.3. Figure 6.7a shows this measurement for three different resistivity substrates, while figure 6.7b shows the time dependency. A value of around $\mathbf{5 \cdot 10^{-17} A/cm}$ has been measured for α. The current before irradiation is about four orders of magnitude lower, thus not considered in this measurement. One must note that these curves are usually calculated when the sensor is fully depleted, where the volume does not depend on the applied high voltage. This is not the case for the RD50-MPW sensors, as they break through before full depletion is reached. Thus, the current is normalized to the volume for these plots, where a volume given by the size of the pixel times the depletion depth (which is measured us-

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

ing TCT studies as explained in section 6.3.2). Though this volume calculation could be more precise, a good agreement can be reached with the results in [6], considering that different samples have been used instead of irradiating the same sample multiple times.



a): Measurement to determine α by fitting a straight line after an annealing time of 80min.

b): Time dependency of the α value for different annealing times.

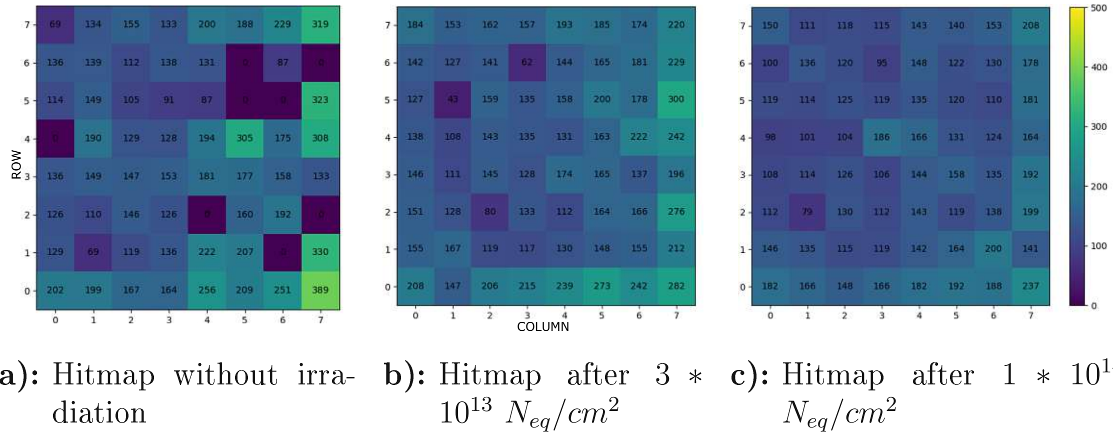
Figure 6.7: The measurement of the α factor for different annealing times is plotted. The difference between resistivities comes most likely from different depleted volumes, which are all assumed to be the same as there is no precise measurement.

The most complex measurement of the RD50-MPW2 test series was performed on the active pixel matrix, utilizing the digital readout in the FPGA. In the so-called *Open-Shutter* measurement, a single pixel is measured by opening a readout window of 20 s and counting every hit for the pixel during this time. In order to induce charge, a radioactive ^{90}Sr source is put on top of the sensor. This measurement is done successively for all pixels. As seen in figure 6.8, the number of hits per pixel decreases with increasing fluence. This is expected due to more significant radiation damage, which leads to a less efficient sensor. Furthermore, these measurements prove that the developed digital readout in the firmware is working. A more detailed explanation of this measurement can be found in [18].

6.3.2 Timing

Timing studies utilizing a laser-induced charge have been conducted with RD50-MPW2 using the comparator output (COMPOUT) with the so-called (top-) transient current technique (TCT). The time from the charge impact until the readout electronics detects the signal is called Time of Arrival (ToA). Together with the Time over Threshold (ToT), which gives the pulse length of the signal, the so-called *time walk* can be measured. The rising edge of the signal is less steep for smaller signals. Thus the discriminator is triggered later than for large signals. This behavior is called *time walk*, and one needs to correct for this time offset depending on the input charge to measure a precise signal. The result of these measurements can be found in [24] and show a constant time walk of 8 ns for

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



a): Hitmap without irradiation

b): Hitmap after $3 * 10^{13}\ N_{eq}/cm^2$

c): Hitmap after $1 * 10^{14}\ N_{eq}/cm^2$

Figure 6.8: Hit-maps of different samples irradiated to different fluences. Different sensors, which are all fabricated on the highest resistivity substrate are shown.

signals above $10ke^-$ (corresponding to a ToT of ≈150 ns). This can be done with the analog output only, as it requires a fast time-to-digital converter (TDC) to get a digital output, a complex structure that does not fit in a single pixel of such a small size.

One can measure the depletion depth and its dependency on the input voltage by shooting the laser into the side of the sensor instead of shooting from the top. With this so-called edge-TCT method, the depletion depth is determined to be around 100µm at an input voltage of 100V. These results are also published in [24] and used in the previous section to calculate the depleted volume.

6.3.3 RD50-MPW2 Testbeam at MedAustron

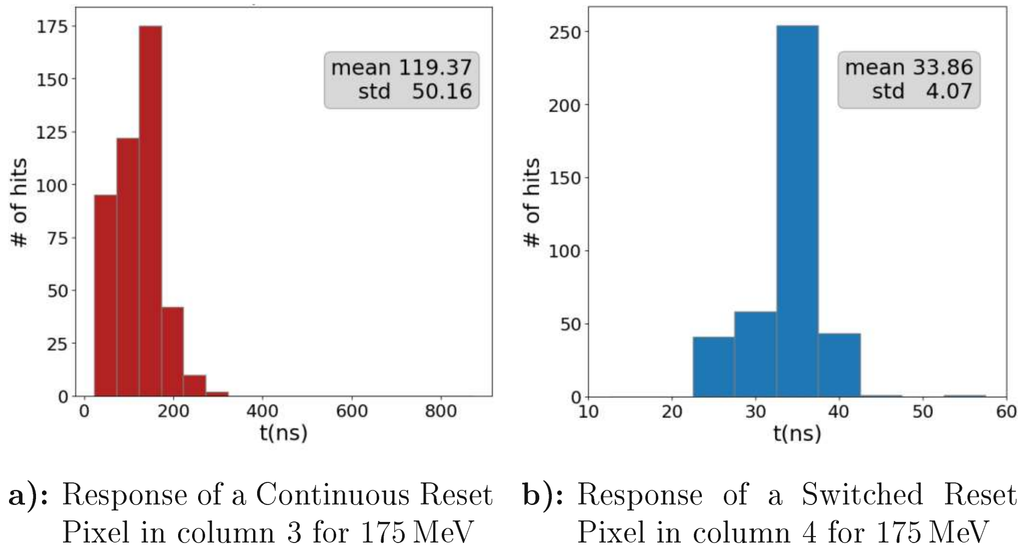
Multiple test beams for RD50-MPW2 are conducted using the accelerator from MedAustron, a clinical accelerator for cancer treatment. MedAustron provides a dedicated room for clinical and non-clinical research. Usually, the accelerator is operated in a so-called medical configuration with particle rates up to $\approx 10^{12}$. Since these medical rates are too high for HEP purposes, dedicated so-called *low-flux* settings with reduced particle flux were developed for physics research. Seven discrete energy steps from 60 to 250 MeV are available, plus a proton beam at 800 MeV. Details of the beam parameter can be found in [25].

A telescope made from four layers of double-sided silicon strip detectors (DSSDs) is also available in the non-clinical research room, which provides reference tracks for tracking studies. It has been developed by HEPHY as part of a system for medical imaging and is therefore called tracker. Details can be found in [26]. It does use the AIDA2020-TLU and is integrated into the EUDAQ2 framework, as presented in sections 6.2.1 and 6.2.3.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Charge Dependency

As mentioned in 2.2 the deposited charge in silicon sensors follows the Bethe-Bloch equation. RD50-MPW2 does have a course charge measurement by measuring ToT. It has a granularity of 5 ns, and the measurement can be done using the digital readout. Such a measurement can be found in figure 6.9 for both pixel flavors. The measurements are done with a 175 MeV proton beam.



a): Response of a Continuous Reset Pixel in column 3 for 175 MeV

b): Response of a Switched Reset Pixel in column 4 for 175 MeV

Figure 6.9: The feedback loop to reset the pre-amplifier has a stronger current source for switched reset pixels as mentioned in 6.3. Thus the pulse length is almost constant at ≈35 ns. The slower continuous reset pixels do have a higher dead time, but are much more suitable for measuring charge as the readout circuit needs less precision for measuring ToT.

Chapter 17.2 in [3] states that ToT is (linear) proportional to the deposited energy if the total charge is collected. The latter is always the case in RD50-MPW2 because only one pixel is connected to the readout at a time. With this assumption and the approximation for low energies mentioned in 2.2, the ToT should follow equation 2.2. Thus, function 6.1 (with linear proportional parameters only) is chosen as a fit function for the ToT distribution.

$$ToT(ns) \approx a + \frac{b}{v^2}\ln\left(c \cdot v^2\right)$$

With v = velocity of incoming particle and parameters a,b and c. (6.1)

This behavior has been measured for both pixel flavors by stepping through all available energies of the proton beam, and results are given in figure 6.10. Although the granularity of ToT is not high enough for precise measurement, one can conclude that ToT and, as a consequence, also energy distribution follow the Bethe-Bloch equation.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

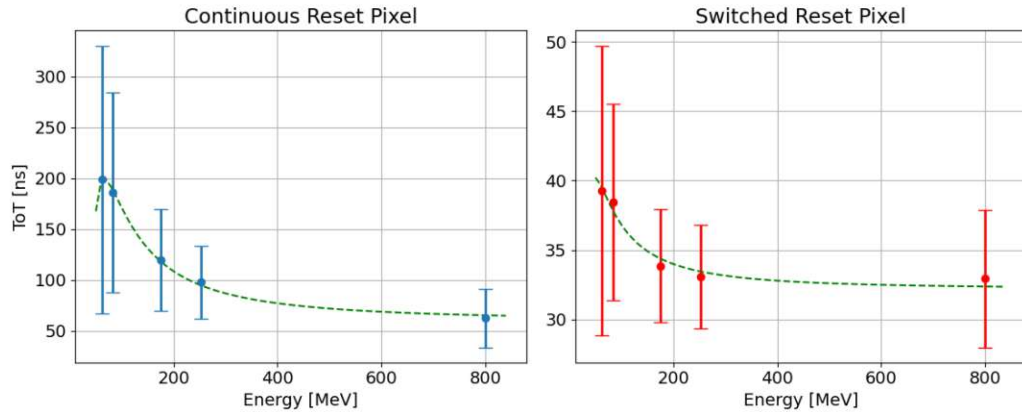


Figure 6.10: ToT distribution of both pixel flavors. 5000 measurements per energy step are plotted and a function following equation 6.1 is fitted.

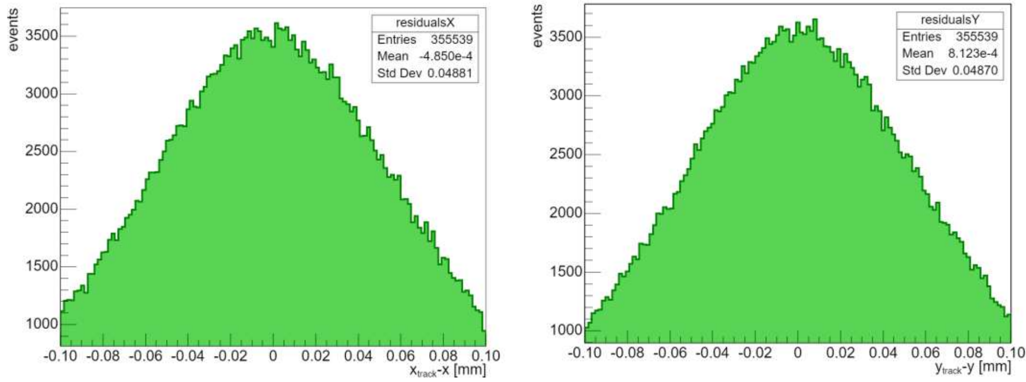
Tracking

An attempt to measure typical performance parameters for tracking has also been tried. The setup consists of the reference telescope with four planes and scintillators for triggering. A detailed drawing and explanation can be found in [18], section 7.2.1.

Measurement runs without the RD50-MPW2 have been taken to align the telescope. Analysis of tracker data at clinical energies does have a few implications. Firstly, the analysis framework corryvreckan is meant to be used for pixel sensors; thus, the hits on the double-sided strips need to be converted to (single-sided) pixels. In order to avoid ghost hits, events with multiple hits are discarded. Moreover, one has to account for multiple scattering in air due to the low energy of incoming particles. Taking all of this into account, the alignment of the telescope is possible according to the steps mentioned in section 6.2.3. The offset between a reconstructed track and a hit on a fully aligned detector plane is called residual and can be found in figure 6.11. It should be centered around zero with a width in the order of one-third of the pixel pitch (which is the distance between two strips in this case) according to [3], section E. However, this value does not consider scattering in air, which does make track reconstruction much more difficult.

The sensor is treated as the device under test (DUT) in this measurement and aligned separately after the telescope. Unfortunately, RD50-MPW2 is reaching its limits here, as only one pixel can be read out at a time. This has multiple implications. On the one hand, getting the geometry correct is not trivial, as one cannot know the orientation of the sensor just by looking at one pixel. Furthermore, corryvreckan does not support analyzing multiple runs together, so one cannot have more than a single pixel in the analysis. Furthermore, charge-sharing information is completely lost as only one pixel is read out. Taking these implications into account, correlation can still be computed, and a course alignment can be done by shifting them to zero. The result can be found in figure 6.12. However,

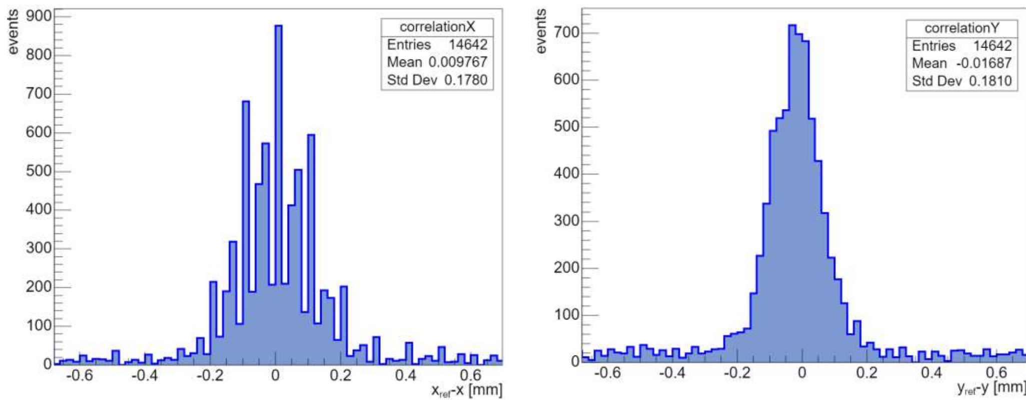
TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



a): Residuals of plane 1 in X-direction **b):** Residuals of plane 1 in Y-direction

Figure 6.11: After coarse and fine alignment of the telescope, the residuals are well below the strip pitch of 100 µm (in the x-direction) and 50 µm (in the y-direction). Plane 2 is chosen as the reference plane as it is closest to the DUT.

track-based alignment does not add any information due to the aforementioned implications. A summary of these measurements, including a detailed description of firmware and measurement setup can be found in [20].



a): Correlations of DUT with reference plane in X-direction **b):** Correlations of DUT with reference plane in Y-direction

Figure 6.12: Correlation after a manual shift to center them around 0. Statistics are low, as only one pixel can be activated.

A so-called *window-run* has been conducted with the rough alignment to see if the whole analysis procedure is working. This study analyzes the same measurement with a single pixel active but using different bundles of tracks from the telescope. All tracks within a pixel plus the surroundings of another half of the pitch are bundled. More hits are recorded for a bundle closer to the active pixel as more tracks go through the pixel due to the overlap in the area. Moreover,

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

some tracks from the neighboring pixel are recorded due to charge sharing. As the first contribution is constant, a qualitative difference in charge sharing can be measured. Results have been published in [27] and show a clear difference between a pixel in columns 0-3 and 4-7. This is expected as those two regions correspond to the two different pixel flavors mentioned in section 6.1.1.

6.4 Conclusion

The first main achievement of the performed laboratory and test beam measurements is the characterization of the analog performance of RD50-MPW2. In particular, the sensor does work as expected, even after irradiation. Problems of its predecessor (low V_{BD} and high leakage current) have been tackled and are overcome in RD50-MPW2. Therefore, the RD50-MPW series continues to use the analog front end of this sensor for future developments.
The second main achievement is the development of a full digital readout chain and verification of it during test beams. Big parts of DAQ hardware, firmware and software can be re-used, and the analysis framework, as well as the beam characteristics at MedAustron are well understood.
The limitation of a single-pixel readout will be overcome in the successor chip, RD50-MPW3, which provides a bigger matrix and digital readout.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

7 Design of the RD50-MPW3 Chip

This thesis focuses on the digital part of RD50-MPW3. Thus, analog design is only touched. Nevertheless, the essential concepts of frontend signal processing are described in the beginning as the digital readout chain is based on the frontend implementation.

7.1 Overview and Analog Design of RD50-MPW3

As all its predecessors, RD50-MPW3 is fabricated in the 150 nm LF15A high voltage CMOS (HV-CMOS) process by LFoundry.

7.1.1 Goals and New Functionalities

After the very successful RD50-MPW2 submission, which focussed on analog performance, the next iteration in the RD50-MPW series focuses on digital performance and the expansion of the analog pixel design to a full size matrix. The well-working analog part is kept, and new digital features are implemented. The list of design requirements for RD50-MPW3 includes:

A) A full matrix implementation big enough to be used for tracking studies

B) A double-column architecture in the pixel matrix to alleviate routing congestion

C) A separation of digital and analog grounds in the pixel to avoid cross-talk

D) Shielding lines between high-activity digital signals in the matrix to avoid cross-talk

E) A digital high speed readout at up to 640 MHz

F) A digital periphery running at 40 MHz, driven by a single 640 MHz clock

G) An idle pattern when no data is sent for digital phase finding

H) Framing of the data with an automated, variable frame length for different transmission rates adjusted to the current hit-rate

I) 8-bit/10-bit encoding for DC balancing

J) Differential pads for glitch-safe signal transmission

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

K) Analog-on-Top design flow, as the PDK does not support a fully digital design

L) A readout chain, allowing to integrate the chip into a larger, heterogenous HEP detector experiment, capable of synchronizing with other detectors

7.1.2 Pixel Design

The analog in-pixel front end is taken from RD50-MPW2. A digital readout is added, consisting of two 8-bit random access memory (RAM) cells for storing the leading and trailing edge (LE and TE) of the signal after the comparator. When the edge-detector detects the rising edge of the comparator output, LE is written by writing the current value of the global timestamp (TS) into the RAM cell. TE is written when the signal falls below the threshold and the comparator outputs a falling edge. The circuit of the data path is depicted in figure 7.1, which corresponds to the continuous reset pixel flavor of RD50-MPW2. Details of signal generation and the analog frontend are given in section 6.1.1 and are published for the RD50-MPW3 chip in [28].

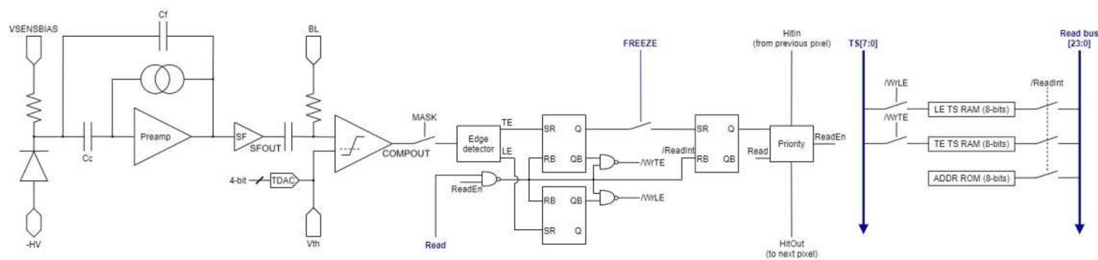


Figure 7.1: In-pixel electronics for the data path of an RD50-MPW3 pixel are shown. Contrary to its predecessor, only one pixel flavor is available due to a lack of design time. The injection circuit is not shown. Picture taken from [28].

Eight configuration bits are stored in every pixel and can thus be individually tuned.
The EN_INJ bit connects the preamplifier's input to a common input pad to inject a signal. This can be used for calibrating the pixel and is not shown in figure 7.1.
The EN_SFOUT and EN_HB bit enable an analog output line before and after the comparator. The SFOUT and the HB signals, are each connected together for all pixels of the matrix and routed to two output pads. Thus, in the normal operation of the chip, the signals should be enabled for only one pixel at a time. Both signals are used for studying the analog behavior of the sensor.
Four trimming DAC (TDAC) bits are used to fine-tune the threshold of the comparator separately for each pixel. These values have to be adjusted during calibration.
The MASK bit is used to enable or disable the readout of a pixel and is particu-

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

larly useful for disabling noisy pixels.

The configuration circuit of a pixel is adapted to a digital peripheral readout and consists of a data flip-flop (DFF) with a 2-to-1 multiplexer (MUX) in front of it as shown in figure 7.2. The implementation of this DFF did not fit into the pixel size of RD50-MPW2, so the pixel size of RD50-MPW3 is increased by 2 µm in both dimensions to 62 µm by 62 µm. The configuration circuit has two modes: In the first mode, values are shifted into the DFF by asserting the SHIFT_EN signal and toggling the clock 128 times, as 128 pixels are connected together to form a shift register. Thus, only the first pixel is connected to the digital periphery. During this mode, all LD switches, which connect the output of the DFF to the eight SRAM cells to store eight configuration bits, are open. In the second mode, the SHIFT_EN is de-asserted, and one of the LD switches is closed by activating the corresponding LD line. This line is connected to 128 pixels, so the same configuration bit is written for all 128 pixels during a single write circle. Although this style of configuration does require quite a lot of signal lines routed over the full matrix, it is favorable over other as less information (example given: The current RAM bit which is written) has to be stored in the pixel, which reduced the pixel size. A read operation is also supported as the serial output of the last pixel is routed to an output pad. In order to read values from a certain SRAM, one has to invert the order of the two modes.

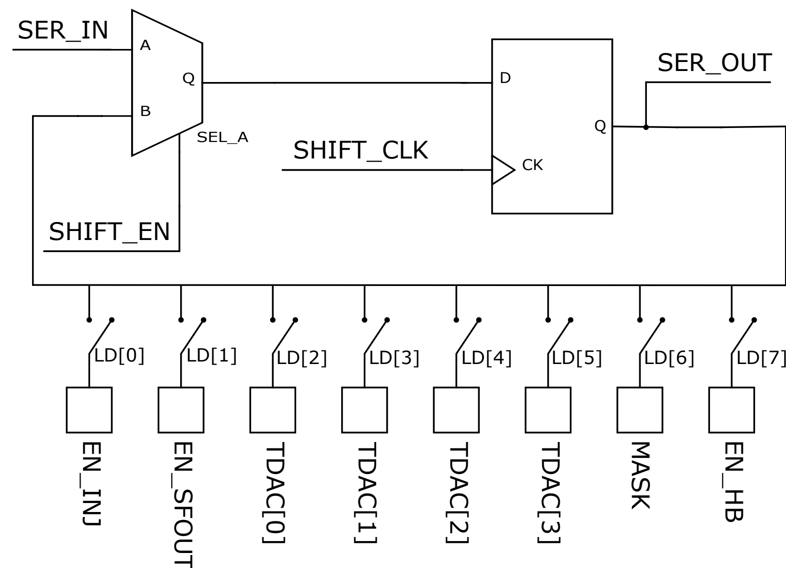


Figure 7.2: In-pixel configuration circuit for an RD50-MPW3 pixel is shown. The design focuses on minimizing area.

7.1.3 Double Column Architecture

Pixels in the matrix are arranged in a double-column architecture to alleviate routing congestion. A double column consists of 64 rows with two pixels per row,

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub
TU WIEN

thus 128 pixels altogether. Two pixels of a row are placed back to back with each other, so digital and analog lines can be shared as shown and explained in [28]. Figure 7.3 shows an overview of the readout lines of an entire double column. The data bus is shared between all pixels; thus, one pixel can be read out at a time. This is sufficient, as a readout can happen every 100 ns, which equals four clock cycles, as shown in section 7.2.2. The readout is triggered by a simple hit flag per pixel, which are all connected together. These hit flags are routed snake-like to connect both columns with just one signal line. Configuration follows the same snake-like routing but in the other direction.

Two readout modes are available, which differ in the priority chain of the pixels. The priority chain is implemented with a token that is either passed to the next pixel if the current one does not have data to be read out or, in case there is data to be read out, kept in the current pixel. Routing of the token line follows the snake-like manner as the hit-flag again. In the first mode, pixels at the end of the chain (thus, with a higher address) are prioritized. This is quite easy to implement but introduces a bias towards high address pixels in case of high hit rates. In order to overcome this problem, a second mode, the so-called FREEZE mode, is been implemented. In this mode, the FREEZE signal is asserted when the first hit reaches the peripheral readout, which disconnects the input from the priority logic, as shown in figure 7.1. This stops the processing of new events in every pixel but still keeps the last hit in the RAM cells for readout if more than just one pixel was asserted within the last readout circle. Every pixel in the double column that has a hit is read out before the FREEZE signal is released, which allows the processing of new hits. In this mode, there is no prioritizing towards pixels with a high address.

In order to alleviate crosstalk between the high-activity digital readout lines, shielding lines connected to ground are implemented between all digital lines.

7.2 Digital Readout Design of RD50-MPW3

Digital signal processing is done in the so-called digital periphery, which is located next to the pixel matrix. It is designed to meet the requirements listed in section 7.1.1 and connects to the pixel matrix following the double-column readout scheme.

7.2.1 Overview of Digital Blocks

An overview of all blocks of the periphery can be found in figure 7.4, where the most essential signals connecting the blocks are drawn as well.

Each double column of the matrix has a corresponding end of column (EOC), which handles the readout and configuration of a double column. This complex block is described in section 7.2.2.
The EOCs are controlled by a control unit that handles the readout of the EOC

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

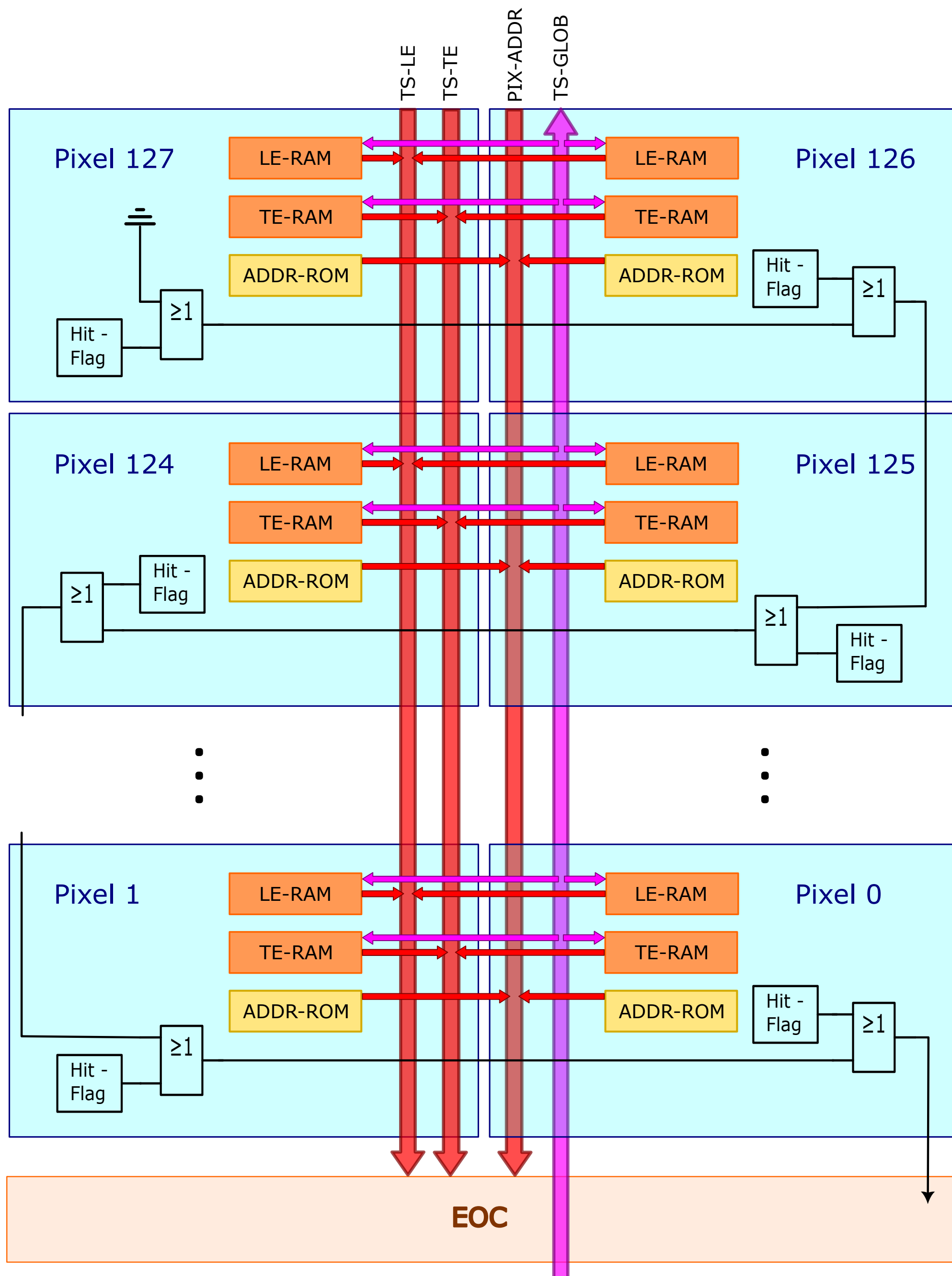


Figure 7.3: The snake-like routing scheme for the hit flag is depicted. Configuration follows the same scheme, while data readout is routed through straight lines in the middle.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

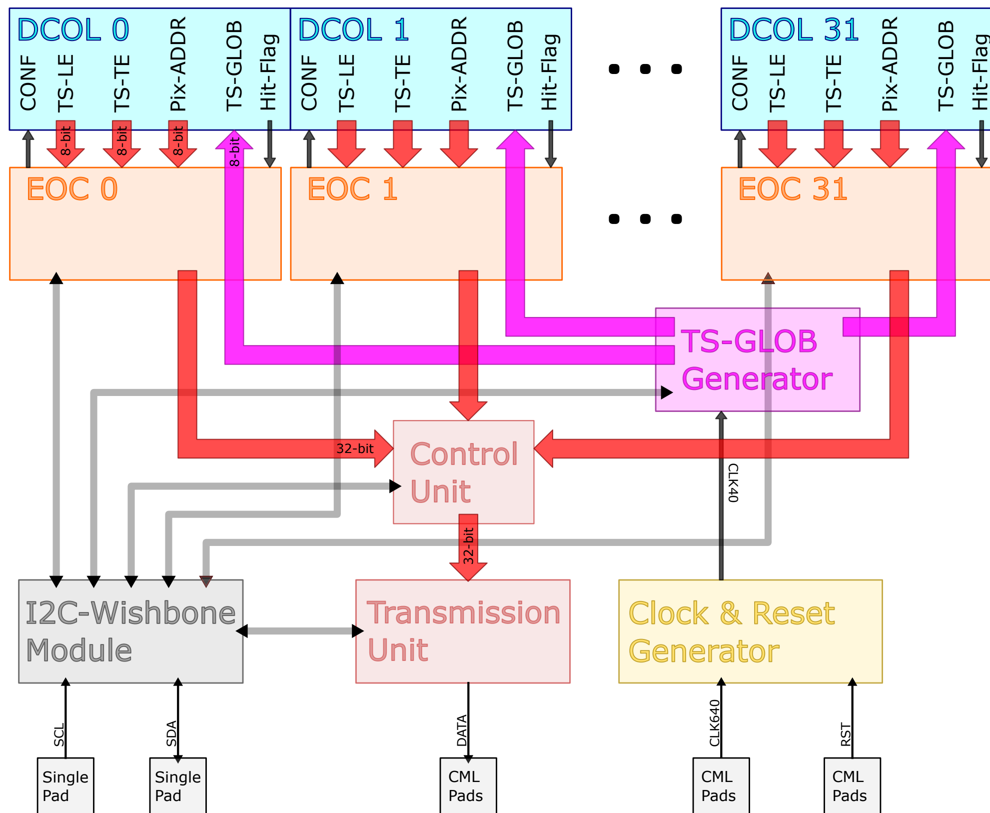


Figure 7.4: An overview of all digital blocks, including the most important signals, is depicted. Thin lines correspond to single-bit signals, while broad lines correspond to bus signals.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

and pushes data into the transmission FIFO. A more detailed description can be found in section 7.2.3.
The transmission unit is responsible for framing, encoding, and serialization and is described in section 7.2.4.
The clock- and reset generator has two tasks. As there is only one 640 MHz clock incoming, it divides the clock into the internally used 40 MHz clock, which drives most of the digital blocks. An external 40 MHz clock can also be fed to the ASIC, which is meant as a backup, in case the clock divider does not work properly. The second task is the acceptance and distribution of asynchronous reset signals. Two different reset signals can be applied, where the *global reset* does reset all registers and logic blocks. The *auxiliary reset* does only reset the logic blocks, including all state machines, while keeping the current configuration, so one does not need to re-configure the whole chip. Since a reset can be asserted asynchronously, resets are synchronized to both internal clocks to avoid metastability.
A global timestamp (TS-GLOB) generator is driven by the 40 MHz clock and generates a global timestamp with a granularity of 25 ns. The timestamps are fanned-out to all 32 double columns. An essential feature of the global timestamp generator is a fast reset signal, which is needed for synchronization. As the other resets, this signal is synchronized to avoid metastabilities.
All digital blocks are connected to a wishbone bus, which is used internally to configure all blocks. Data-width of this wishbone bus is one byte, while the address width is two bytes. This large address space leaves some space for future chip extensions. An I2C-Wishbone module converts the wishbone control signals to externally used I2C signals and vice versa. The I2C bus is chosen for external communication to save output pads as it only needs two.
All clock and reset pads and fast data pads are implemented as differential current mode logic (CML) pads to avoid crosstalk and common mode noise, and allow high bandwidth transmission.

7.2.2 End Of Column

The end of column (EOC) is the central component of the digital periphery, as its tasks are the readout and configuration of the pixel matrix. The implementation of a double column and the EOC are closely related. All signals are carefully chosen to mitigate the low area of a double column and implement a readout speed of around 10 MHz hit rate.

Overview of EOC

The logic of an EOC is divided into two parts, which are responsible for two different tasks. Figure 7.5 visualizes this separation. Configuration is handled by a 128-to-1 multiplexer, which switches through 128 values in 16 8-bit configuration registers that store information, which will be written into a certain SRAM cell of the double column. These pixel configuration registers are connected to the internal wishbone bus, which is used for communication inside the ASIC as mentioned in section 7.2.1. Read-back from values stored in the pixels to the configuration register is not supported. Instead, the data to be read is shifted to

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

an output pad together with a start-of-stream signal. Read-back and checking the configuration are done using an FPGA and software.
An additional register for the configuration of the EOC is implemented. It does store information about:

- The current value of the LD switch which determines what data is in the pixel register and thus pushed to the double column (three bits)
- One bit to enable or disable readout
- One bit to enable or disable configuration
- One bit to read or write configuration
- One bit to enable or disable FREEZE mode
- One bit to enable or disable the hit bus of the double column for analog monitoring

The central component on the readout side is a 32-words deep FIFO, which is used as a buffer for derandomizing data from a single double column. A single data word from the double column consists of leading edge, trailing edge and pixel address with eight bits each, which is why the FIFO is 24 bits wide. A simple token handler indicates when the FIFO of the EOC is read out; its functionality is described in section 7.2.3. Data to be read out is stored in an intermediate output register, which consists of 32 bits, where the last eight bits represent the EOC address and are static and incremented for each EOC.

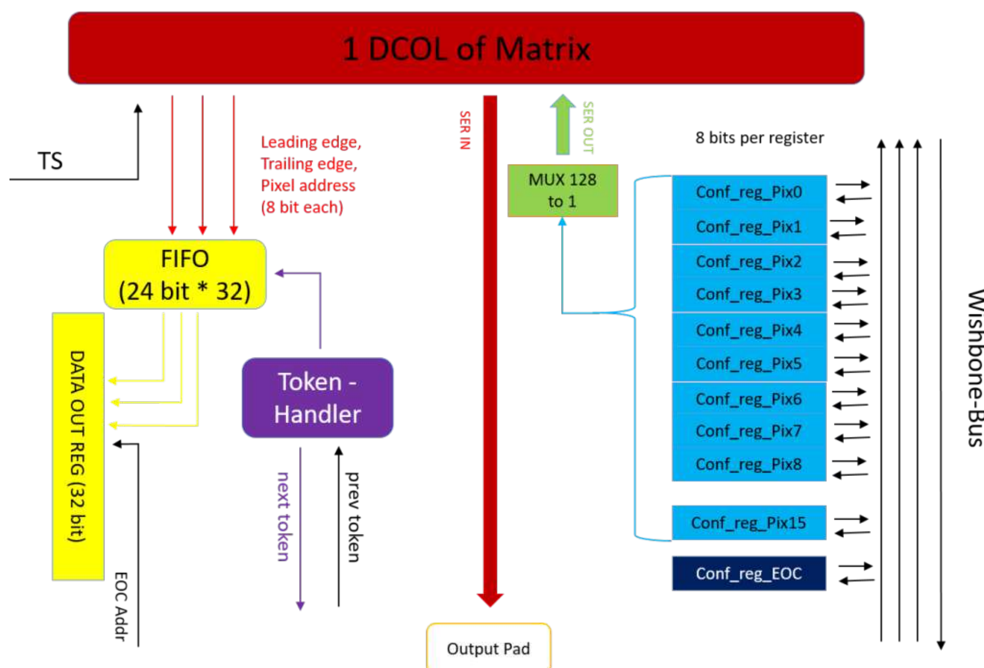


Figure 7.5: A schematic overview of all features of a single EOC is shown. The left side shows blocks needed to read out the matrix, while the right side dedicated to configuration. Picture taken from [28].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The timestamp is not part of the logic of an EOC, but it is routed through the EOC for easier buffering.

The logic of a single EOC is implemented in a finite state machine (FSM), which is depicted in figure 7.6. The default state is IDLE, as the FSM falls automatically into this state (via the transient UNINIT state) after a reset is issued. The EOC can either go to the configuration state called SHIFT or to the readout state called READ_EOC. Thus, it is impossible that a double column is configured while data is read out to avoid confusion about the configuration during data taking. The algorithm to read out a double column is described in the next section; it is active during the READ_EOC state. The SHIFT state is a transient state and exactly 130 clock cycles long. 128 cycles are needed to shift the 128 values through the shift register formed by the double column, and two cycles are needed to read or write the data into the SRAM. A *shift_done* flag is used to prohibit automatic restarting of the shift cycle. It needs to be reset via a slow control command in order to start a new configuration cycle.

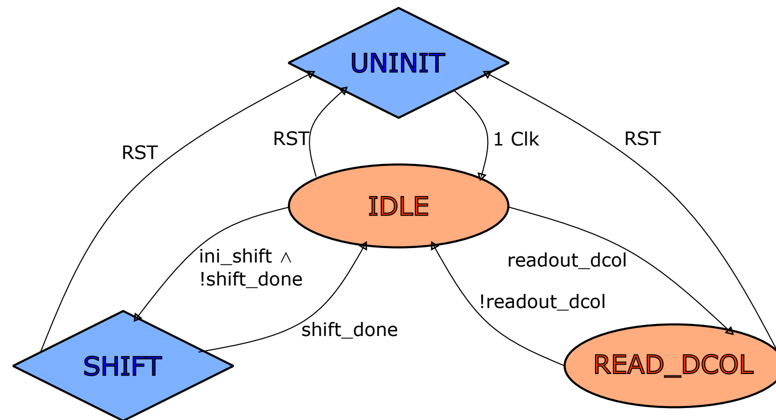


Figure 7.6: The FSM of the EOC has four states, as depicted. SHIFT and UNINIT states are transient states, as they fall into IDLE automatically.

Double Column Readout Signals

The readout of a double column follows the so-called FEI3 style[29], adapted from the ATLAS tracker. The needed signals are depicted in figure 7.7. A readout cycle is triggered when the *hit_out* flag from the double column is high. As this is a completely asynchronous signal, it is synchronized to the clock before it is evaluated. Readout begins one clock cycle after *hit_out* is received by pulling all readout lines to ground for one cycle to discharge all readout lines by issuing the *pull_down_enable* flag. This is needed, as the FEI3 implementation of a pixel does contain only transistors to pull up readout lines. The transistors to pull down the lines are not implemented in order to save area. After discharging, the *read* signal issues the readout of the double column, for which the pixel with the highest priority puts data on the bus, which is then stored in the FIFO of the EOC. Toggling of *pull_down_enable* and *read* is repeated until the *hit_out* flag

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

returns to zero, which indicates that no pixel has data for readout.

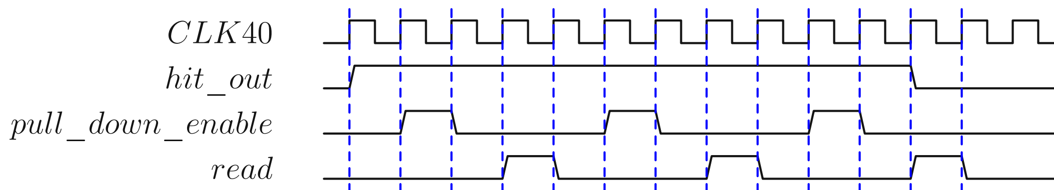


Figure 7.7: Most important readout signals for the FEI3-style readout. The waiting time of one clock cycle between *pull_down_enable* and *read* is fixed to one clock cycle.

The FREEZE signal can be set separately for each EOC by the slow control and is not shown in figure 7.7, as it does not actively participate in the readout cycle. It does only change the priority of pixels to be read out.

7.2.3 Control Unit

The Control Unit (CU) is responsible for reading out the EOCs and pushing data in the transmission FIFO for derandomizing data of double columns. This is done by a rolling shutter algorithm, which asks all EOCs to put one data word into their output register (if they have any in the internal FIFO) and then pushes the data subsequently into the transmission FIFO. This process is very similar to the FREEZE mode of the double columns.

Figure 7.8 shows the state diagram for the control unit. The IDLE and UNINIT states are equivalent to the ones of the EOC. The CU has two states for operation called SEND_DATA and DEBUG. In the first state, the above mentioned readout cycle is repeated infinitely, while in the debug mode, the cycle happens only once and stops afterwards.

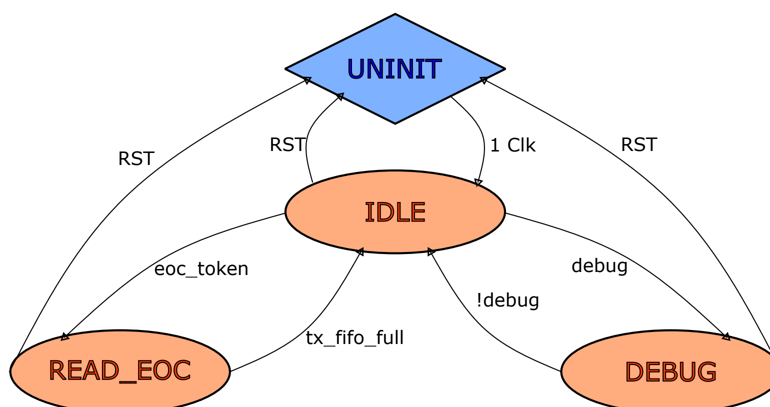


Figure 7.8: The FSM of the CU has four states as depicted.The UNINIT state is a transient state, as it fall into IDLE automatically.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

7.2.4 Transmission Unit

The transmission unit (TX Unit) is responsible for reading out the transmission FIFO (TX-FIFO) and sending data via the serial link. Before serialization, data words are framed and encoded. A schematic view of the logic is depicted in figure 7.9, where one can see three types of data words: Payload data is stored in the TX-FIFO, which has a depth of 128, while start-of-frame (SOF), end-of-frame (EOF) and debug words are static and configurable over the slow control. In addition, an IDLE data word is available. Each data word is 32 bits long and encoded using an aurora 8b/10b encoding; thus an encoded data word has 40 bits. The IDLE word contains four configurable comma characters of the encoding scheme to distinguish it easily from other words. SOF and EOF use data words of impossible addresses per default for differentiation, while the DEBUG word can be arbitrary and is chosen via the slow control. After encoding, data words are pushed into a serializer, implemented as a shift-register with optional, parallel write input to each register.

Serialization happens at 640 MHz, meeting the design specification. This is why an external clock of 640 MHz needs to be provided to the ASIC. Dividing this speed by the length of a data word (40), the logic before the serializer must run at precisely 16 Mhz. This causes three clock domain crossings (CDCs), quickly introduced hereafter.

- The CDC between the 40 MHz domain on the input side of the TX-FIFO and the 16 MHz domain at the output side is resolved by using an asynchronous FIFO implementation which can handle two different clocks for the push and pop side. This is an IP block taken from cadence.

- The CDC between the 16 MHz domain before the serializer and the 640 MHz domain after it is resolved by generating an asynchronous 16 MHz clock from a ring counter of 40 bits running at 640 MHz. The 16 MHz clock is simple one of the values of the ring counter; thus it is only high for a brief period. The clock is depicted in figure 7.11.

- The different data registers and the selection register for switching the MUX use the slow control and thus run at 40 MHz. This causes another CDC between the output register of the encoder (running at 16 MHz) and the input registers. Another complication is that the encoding logic needs at least three cycles of the 640 MHz clock. The solution is a particular *enable* signal, with is active during this period and covers the two possible timing scenarios between the 40 MHz and the 640 MHz clock. This is described in great detail in [27]. The timing of this *enable* signal is the most crucial part of the whole digital periphery, and timing constraints are written carefully using multi-cycle paths, which guarantee that the logic has more than one clock cycle before the output needs to be ready.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

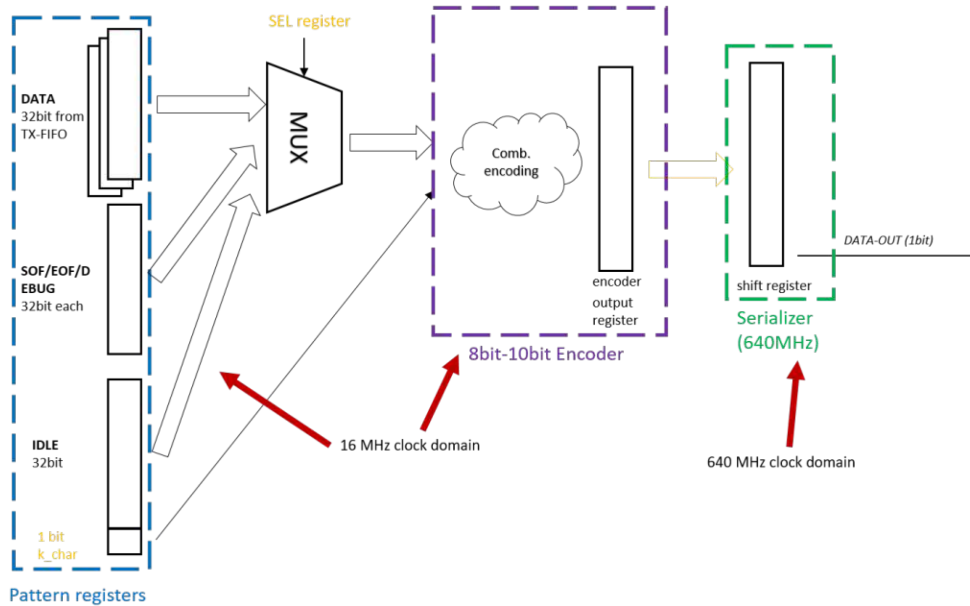


Figure 7.9: A schematic view of the transmission unit is shown. Picture adapted from [27].

The FSM of the transmission unit has four different states, which are equivalent to the ones of the control unit. The UNINIT state is equivalent to the respective state of the CU and the EOCs. The IDLE state is different, as the transmission unit does still send the IDLE pattern instead of idling. The SEND_DATA state can only be reached if the TX-FIFO has data to guarantee a constant data output without glitches or jumps. The debug mode works equivalent to the one of the CU. The debug pattern is still framed and encoded, though, so the DEBUG state sends out three words (SOF, DEBUG, EOF) exactly once.

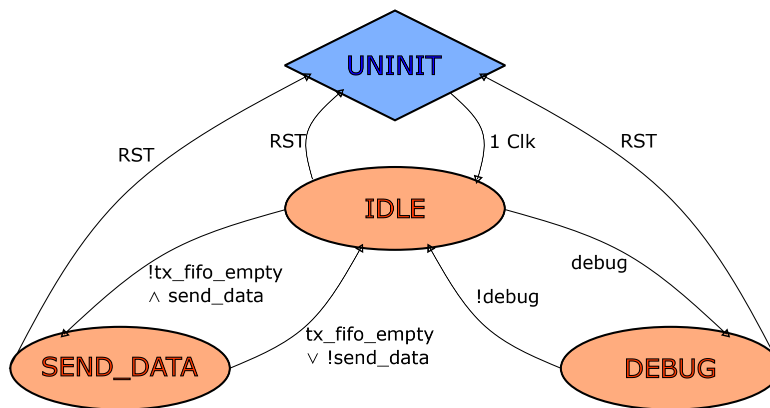


Figure 7.10: The FSM of the transmission unit has four states, as depicted. The UNINIT state is a transient state, as it falls into IDLE automatically.

An example waveform can be found in figure 7.11, which shows the different states to read data words and flags needed to switch between them. The serial

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

output is implemented carefully, as no jumps or glitches must occur. Thus, it is always driven, and changes between the state happen seamlessly.

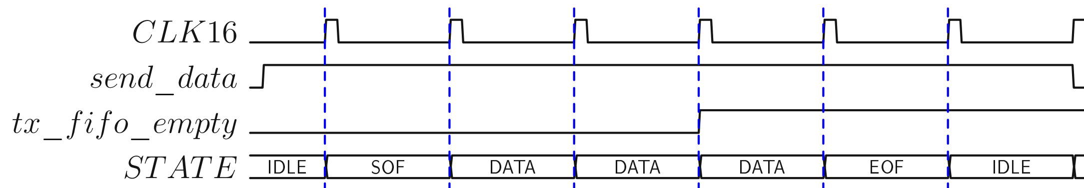


Figure 7.11: The waveform shows the important control signals of the FSM states for an example of three data words in the transmission FIFO.

The IDLE pattern is essential for the backend FPGA reading out the ASIC. It is not feasible to read such a high-speed signal by sending out a sampling clock next to the data line, as the delay between the two lines has to be very low, which is difficult to route on the PCB. Instead, one can implement either an algorithm for clock recovery out of the data stream or a phase-finding algorithm and shift the FPGA clock in phase until it is aligned. The latter option is chosen for RD0-MPW3, which does require a constant, known pattern. The IDLE pattern is used for a digital phase finding algorithm, and thus the ASIC must send this pattern once it is turned on and before data is sent out. Thus, it cannot be turned off by the slow control.

7.2.5 Critical Timing Path

The critical timing path of the RD50-MPW3 digital logic is located in the clock & reset generator. The clock divider responsible for dividing the 640 MHz clock by 16 is implemented as a binary counter with four bits. The most critical path is the feedback loop of the register holding the highest bit, as the logic needed to calculate each bit increases with the place value of the bit. Figure 7.12 shows a schematic of the critical path. A timing path can only start and end at the output of a sequential logic cell, for instance, a register, or an I/O-pin. The register in the critical path is clocked by the fast external clock and has the same register, Clock_Div[3], as the start and end point.

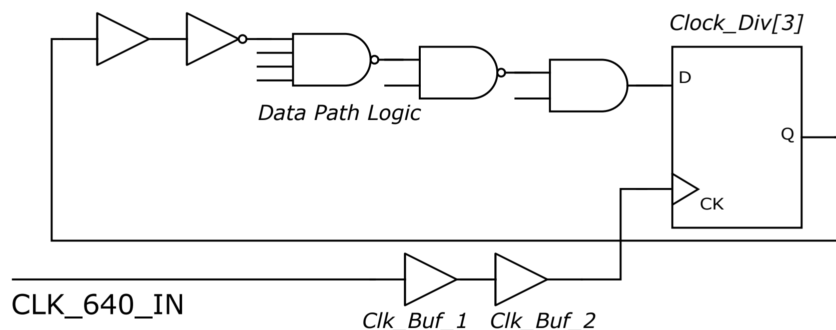


Figure 7.12: A schematic view of the critical path is shown. The time path is a feedback loop to the same register.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Due to a non-disclosure agreement, details of the components, including their timing properties, cannot be given. Nevertheless, an explanatory calculation of the critical path is given. Setup analysis checks for the maximum possible delay of a timing path to guarantee data has arrived at the input of a register before the capture clock arrives. The *required time* (t_{req}) is the point in time when data is captured and given by the clock period (t_{CLK}) plus the latency of the capture clock (t_{CC}) (= insertion delay). In our case, the latter could be ignored as the capture and launch clock have the same delay (apart from a possible minimal uncertainty in the input latency, e.g., jitter) as it is a loop timing path. The *arrival time* (t_{arr}) is the actual time of arrival of the data. It is given by the data delay of the data path logic (t_{data}) plus the clock-to-output delay of the flip flop ($t_{C->Q}$) plus the latency of the launch clock (t_{LC}). In addition, each register has a setup time(t_s), which is the minimum time the data signal should be held constant before the clock edge to guarantee reliable sampling. Thus, the setup time needs to be deducted from the required time. An uncertainty t_u is typically added to the arrival time to have some margin in each timing path. The arrival time must be shorter than the required time so data arrives at the register before it is captured. The difference is called *negative slack* and must be positive for all time paths to avoid timing violation. An example calculation for the critical path shown in figure 7.12 is given in equation 7.1. Values are taken from post-layout static timing analysis.

$$\begin{aligned} t_{req} &= t_{CLK} + t_{CC} - t_s = 1809ps \\ t_{arr} &= t_{data} + t_{C->Q} + t_{LC} + t_u = 1081ps \\ \text{Negative Slack} &= t_{req} - t_{arr} = 728ps \end{aligned} \tag{7.1}$$

7.3 Conclusion

The design concept of RD50-MPW3 and the actual implementation constitute a significant part of this thesis. Requirements have been stated at the beginning of the design, and a conclusion if the requirements mentioned in section 7.1.1 are all met is drawn at the end of this thesis.
Implementing the transmission unit, including timing constraints and resolving its CDCs, is the most crucial part of the design. However, the digital implementation works perfectly, and no issue has been found so far, as measurements show in the next chapter.
During the design, communication with the analog designer who implemented the matrix and the engineer who implemented the readout in the FPGA was crucial and very fruitful. A more decent requirements list and test plan are definitely recommended for more complex ASICs.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

8 Testbeam for RD50-MPW3

This chapter mentions changes in the readout system for RD50-MPW3 compared to its predecessor. Sensor-level tests for radiation hardness, charge dependency of the signal, and similar are not yet performed, as the thesis focuses on digital performance during test beams. Moreover, no new behavior is expected, as the analog front-end did not change, and a radiation campaign needs quite some time; thus, these tests are currently in preparation, and results are not yet available. Afterward, the test beam setup for a CERN test beam is explained, and analysis results are presented.

8.1 Adapting the Readout System for RD50-MPW3

The DAQ system of RD50-MPW3 is adapted from RD50-MPW2 and is again based on the Caribou system and uses the same auxiliary devices mentioned in section 6.2. The changes are briefly listed hereafter:

- A new chipboard is developed, which includes buffers, level-shifters and routing of signal lines to the CaR board
- The firmware is completely renewed and explained in the next section
- The software is upgraded and distinguishes between an *initialize* phase for powering the chip and a *configuration* phase for sending slow control commands to the chip
- A monitor has been added to the software, which shows a live hit map
- A GUI has been implemented to handle the more complex configuration options
- change from a trigger-synchronous readout system to a time-synchronous readout system

The software of RD50-MPW3 is explained in great detail in [30] and not repeated here.

The AIDA-TLU introduced in section 6.2.1 has two readout modes. The time-synchronous readout mode, AIDA mode, is used to synchronize RD50-MPW3. The idea is to synchronize the timestamp generator of RD50-MPW3 with the AIDA-TLU by using the TLU-clock for generating the RD50-MPW3 input clock

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

and a so-called t_0 signal at the beginning of data taking, which resets time counters of the TLU and RD50-MPW3 synchronously. On the contrary, the trigger-synchronous EUDET mode demands counting trigger numbers at the same rate from all detectors, as done in RD50-MPW2. Synchronized data taking is only possible in AIDA mode for RD50-MPW3 as the chip cannot count external trigger numbers, making EUDET mode synchronization impossible. The main advantage of the time-synchronous mode is the independence from other detectors in the system, as every detector can record at its maximum speed. Moreover, there is no need to receive a trigger number from the TLU, which costs time. The disadvantage is the increased complexity of the readout system, as this timestamp must be global to all detectors in the system, which requires a certain length of the timestamp as it needs to cover the whole data-taking period. This period is typically a few hours, which equals around 40 bits of the timestamp at 40 Mhz. Distributing a 40-bit timestamp in the chip is not feasible due to routing congestion, so the readout system has to consider this problem, taking arbitrary overflows into account. A detailed explanation of how this is done for RD50-MPW3 during offline analysis is given in section 8.3.4. The AIDA-TLU stores the internal timestamp together with the trigger number. Thus, one can convert between the two, and connected detectors can use different synchronization modes, which can be synchronized in the offline analysis.

8.1.1 MPW3 firmware

Due to the digitization of data in the chip, no analog readout of the sensor is required. Therefore the RD50-MPW3 firmware is completely renewed containing only a digital readout. Control happens via an I2C bus and uses the I2C controller of the Caribou framework, so there is no need for configuration circuits in the FPGA either.

RD50-MPW3 has a single data output stream synchronized to a clock running at the same speed but generated outside the chip. Thus, a phase scan is necessary to align the data stream's phase and the sampling clock. An automated digital phase detector is implemented using the known IDLE pattern of the chip. The phase detector calculates the needed phase shift, which is applied to the input data stream to align it with the internal clock.

Figure 8.1 shows a schematic of the readout block in the firmware. Data is initially delayed by the phase shift determined by the digital phase detector. Afterwards, a dual-edge triggered data flip-flop (DFF) is used to convert the single data stream into a double data rate (DDR) transfer to reduce the needed clock speed of the circuitry afterward. A de-serializer converts the stream into ten parallel bits, then decoded into eight bits by an 8-bit/10-bit decoder. Before storing the data, the Frame-Finder detects if the current word is a SOF, EOF, or a regular data word. In the first two cases, it replaces a word with an eight-bit overflow counter, which is mentioned in the next section for data synchronization. Data is stored in two different FIFOs. All of the data is written into a user

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

datagram protocol (UDP) FIFO, a fast, standardized protocol for transferring data to disk via a one gigabit small form-factor pluggable (SFP) port of the Caribou system. A configurable fraction of data is sent to the AXI bus of the SoC for processing with the CPU of the Caribou SoC board. This slow transfer is used for monitoring data and laboratory measurements.

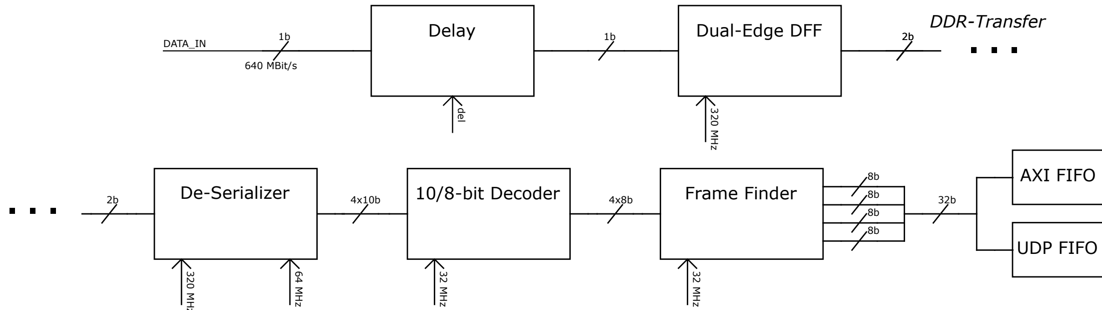


Figure 8.1: A schematic view of the firmware for RD50-MPW3 is shown. The different clock domains are a result of the double data rate (DDR) transfer used in the FPGA to half the clock speed. Decoding and frame-finding are done per byte but after each other and not in parallel.

8.1.2 Time-Stamping

Due to an issue in the readout of the matrix, the periphery has to be slowed down for RD50-MPW3. Thus, a clock period of 50 ns (instead of the design period of 25 ns) is used for the calculations here and for the analysis results later on.

As explained in section 7.2.4, the chip sends data in frames, where each frame starts with a SOF word, ends with an EOF word, and has an arbitrary amount of data words in-between. All words are 32 bits long.
As each data word contains only eight bits for the start time (and eight more for the end time), the FPGA is counting overflows of this timestamp. The chip has an output signal indicating when the timestamp has an overflow, making it rather easy to implement. To prolong the timestamp, eight bits in SOF and EOF are reserved for this overflow value, and all events get timestamps of 16 bits by extending the eight bits from the pixel by the overflow value. A simple extension only works if there is no overflow of the lower eight bits during a frame. As seen in the offline analysis, this is the case for all RD50-MPW3 measurements done so far. However, the analysis system is designed to implement a pre-processing step, allowing a sophisticated calculation of this 16-bit timestamp in case of overflows.

Figure 8.2 shows how data is stored in the UDP FIFO and how the software calculates the global event timestamp. The UDP FIFO can store 4096 words, which corresponds to a bit more than 1000 frames at most. As mentioned above, 16 bits are not enough for a global timestamp. Thus, the FPGA has an internal 64 bit timestamp running at the same speed. Those 64 bit are sent in two words

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

after the UDP packet is sent. The timestamp calculation now takes the 16 bits from the pixel and SOF/EOF (for start and end time) and replaces the last 16 bit of the 64 bit counter. The result is a global, 64 bit timestamp, which is sufficiently long as it covers almost 30 years.
A UDP package is sent when the FIFO is either half full, or by a software query.

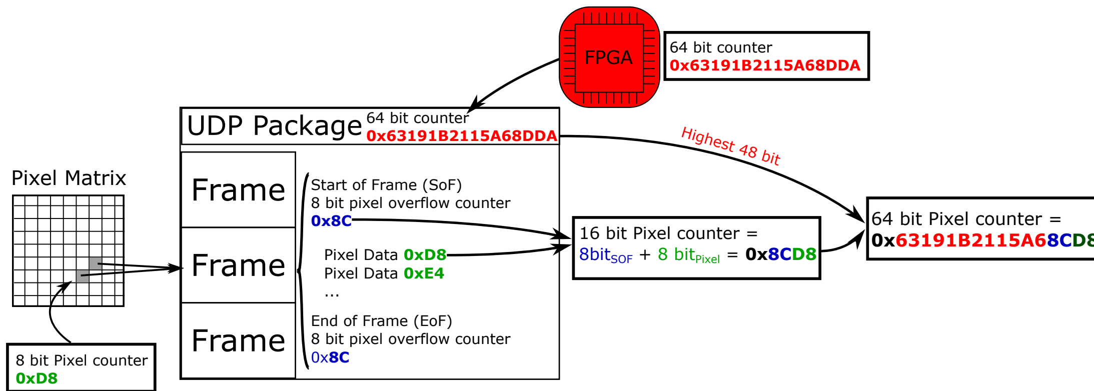


Figure 8.2: The Timestamp calculation for RD50-MPW3 is shown, while values from different sources have different colors. Figure privately provided by Uwe Krämer.

In both cases, 2048 words are sent, while unused words are filled with zeros. This leads to a lot of unnecessary data for the low data rates during the testbeam, but it leaves space for much higher data rates or a bigger matrix. In both cases, neither the periphery, nor the firmware require changes due to this data buffer.

One issue with this method arises if there is an overflow of the 16-bit counter within the UDP package. These cases cannot be distinguished, as the 64 bits are only assigned once per UDP packet and not per frame. Such an overflow occurs every $2^{16} \times 50\,ns = 3.2768\,ms$ and would require changing the 17th bit, which is impossible as the 64-bit counter overwrites it. It has to be taken into account by shifting the data in multiples of 3.2768 ms trying to find matches with a reference track provided by an external telescope. The telescope and method are explained in more detail in section 8.2.2 and 8.3. The event duration is $230\,us$, which is defined by the reference telescope and makes double counting unlikely, as those hits would only overlap when the event window overlaps during the time of the overflow, which has a probability of $\sim 7.0\%$ and needs to be taken into account by the analysis as well. This estimation is calculated by dividing the time it takes for the counter to overflow by the event duration.

For synchronization, the UDP counter, the overflow counter in the FPGA and the pixel timestamp are reset when the TLU issues a t_0 signal, which happens at the beginning of each run. In this time-synchronous mode, the timestamp of the TLU and the detector run synchronously at the same speed and no exchange of trigger numbers or other signals is needed anymore, which makes the readout fast and independent.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

8.2 Testbeam setup at CERN

RD50-MPW3 is tested intensively during a test beam at CERN. The H6 beam line of the super proton synchrotron (SPS) is used for this purpose, as it has a pre-installed beam telescope, the AIDA-Telescope.

8.2.1 Accelerator and Beam Settings

The particle beam of the H6 beamline is a secondary beam produced through the interaction of the primary SPS beam with a target. This allows for different particle types and energy, typically consisting of a mix of protons, charged and uncharged hadrons and leptons. The beam passes through multiple areas, and users agree on detailed beam settings. The energy changes during the data-taking period within a range of 100 GeV - 300 GeV. Therefore, the impact is always close to a MIP. The overall data rate is highly dependent on the current *super-cycle*, which decides the target of the SPS beam. The rate during a spill of particles is in the range of a few kHz. The spot size is bigger than all the detectors used to take data and can not be measured.

8.2.2 Reference Detectors

The reference detectors used during the CERN test beam are similar to the ones used for RD50-MPW2 at MedAustron. The same AIDA-TLU is used for synchronization and a pre-installed telescope is used.

AIDA-Telescope

The H6 beamline has the AIDA-telescope pre-installed, a EUDET-type beam telescope fully integrated into the EUDAQ2 framework. All EUDET-type telescopes are based on the Mimosa26 (M26) MAPS. Various identical telescopes exist at various institutes, and a reference is given in [31]. Part of the telescope, including the RD50-MPW3 mounted as DUT in the middle, can be seen in figure 8.3.

Timing Reference Plane and Auxiliary Devices

The AIDA-TLU is used for synchronization as it supports a time-synchronous mode (needed for RD50-MPW3) and a trigger-synchronous mode (needed for the telescope) at the same time. Reference has been mentioned in section 6.2.1 and can be found there.
Two devices are tried as a timing reference, but only one is working during the test beam. The so-called FEI4 plane delivers a timestamp with a 25 ns precision, which the telescope cannot record. Unfortunately, the device has some bugs and dramatically reduces the data rate. In addition, due to an observed low efficiency seen in the offline analysis, possibly due to incorrect operational settings, the decision is made to exclude it from the analysis. The second device, a Timepix3-plane, does work well, but there needs to be more time to properly integrate into the EUDAQ2 framework during the test beam, so data can not be synchronized.

TU Bibliothek, WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

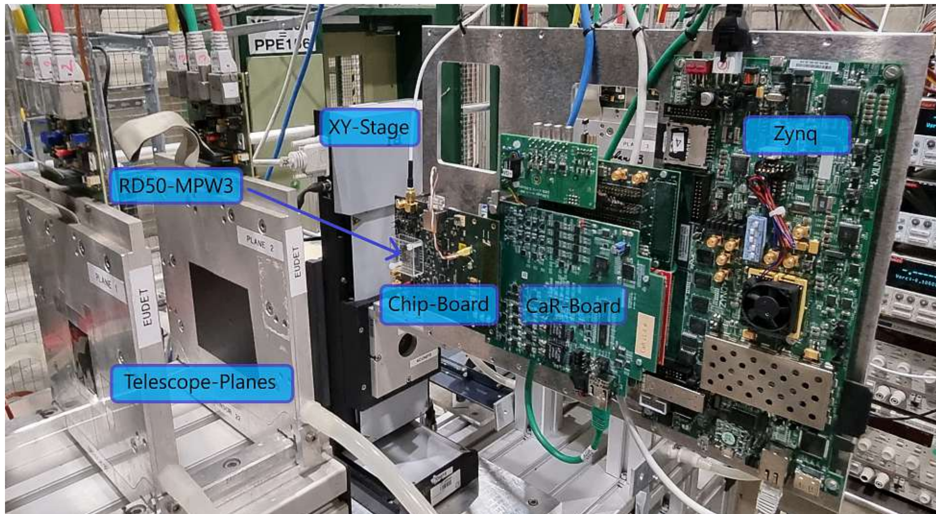


Figure 8.3: Picture of the front layers of the EUDET-type telescope and RD50-MPW3 including the Caribou readout system.

Tracking with a timing reference is less challenging as it gets easier to distinguish tracks during the same event. However, this is not a big problem for low data rates as those encountered during the test beam campaign, and differentiation of tracks is possible during the analysis even without precise timing information for the used rates.

8.2.3 Detector Setup

Three different types of detectors are used during the CERN test beam, drawn in figure 8.4. The AIDA-telescope consists of six Mimosa26 planes, of which the first plane was broken and could not be used. The RD50-MPW3 is put in the middle of the six planes as DUT to get the best possible track resolution at the sensor's position. The FEI4 mentioned above was screwed on the back of the fifth M26 plane.

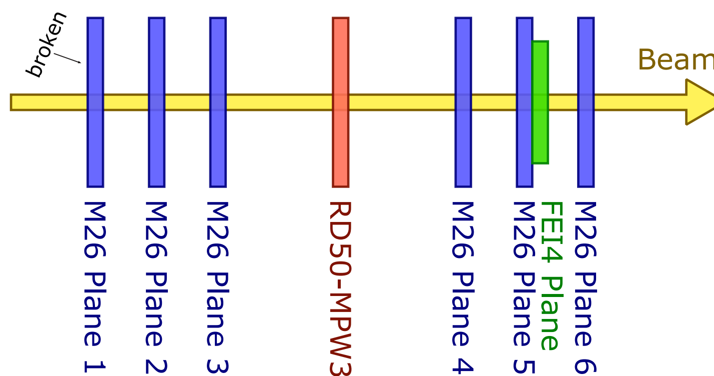


Figure 8.4: Eight layers of active detectors have been used for data taking. The M26 planes have additional drill holes on both sides, allowing them to screw small detectors directly onto the planes.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

8.3 Data Analysis

The test beam data analysis is based on the Corryvreckan framework. A quick overview is given in section 6.2.3.
Contrary to RD50-MPW2, the full pixel matrix can be read simultaneously, enabling more capabilities for data analysis, including more tracking parameters.

Due to an issue in the readout of the matrix, the periphery has to be slowed down for RD50-MPW3. Thus, a clock period of 50 ns (instead of the design period of 25 ns) is used here and for the calculations in section 8.1.2. Moreover, all presented results use only the matrix's upper half, as a significant amount of noise is found during the data taking, making it impossible to find a suitable threshold for the whole matrix.

Details about data flow and software implementation for the RD50-MPW3 analysis can be found in [30].

8.3.1 Event Definition

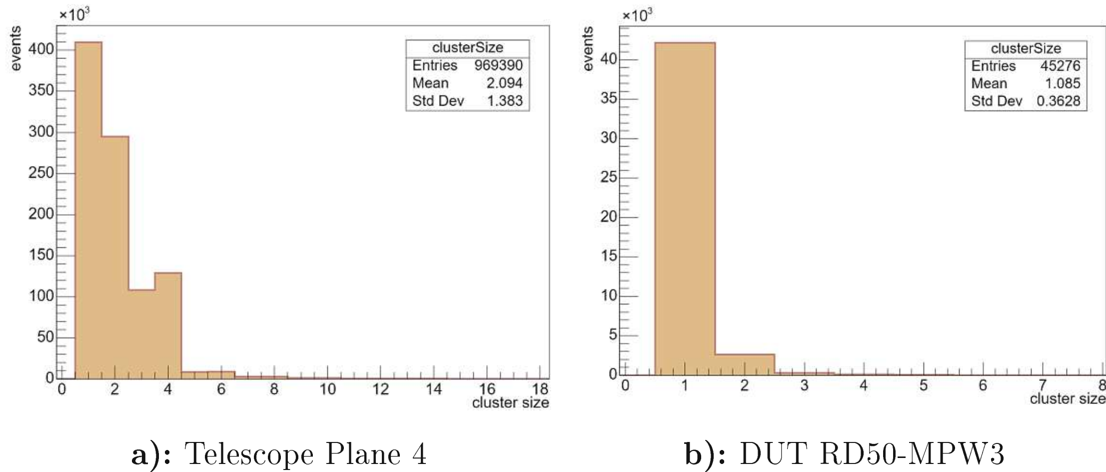
The analysis of RD50-MPW3 is event-based, while an event is defined by a start and end time. Data runs can only be analyzed one after the other, as the global time is reset at the beginning of each run. Each run consist of around 30 minutes of data taking.
Reference tracks are compared to the output of the DUT for data analysis. The telescope defines reference tracks; thus, the telescope also defines events. The Mimosa26 sensors have a readout time of 230 µs per event; therefore, this time window defines an event. An event can have multiple tracks. Any hits on the DUT with timestamps that fall within the timeframe defined by the Event start and end time are assigned to the corresponding event. Timestamps that fall outside any event windows defined by the telescope are discarded in the analysis.

8.3.2 Clustering

Clustering is the process of finding hits that originate from the same particle impact. Due to charge sharing, multiple neighboring pixels can be hit by one particle. Clustering is done individually for each detector plane. Thus, the clustering module looks for a neighboring pixel within an event and plane and matches them to one hit. Only pixels sharing an edge are counted as neighbors, while pixels sharing a corner are not. This hit gets the mean coordinates from all pixels in the cluster as the hit position. It will calculate the weighted mean for even better position information if charge information is available. The clustering also considers a time window. Only hits that lie within the same time window are clustered together. While only of minor importance in this work due to the low-rate environment, it is vital to cluster hits accurately in a high-rate environment.

Plot 8.5 shows the cluster size distribution for the RD50-MPW3 as DUT and for the telescope plane 4, directly after the DUT as an example for the cluster size of a Mimosa26 plane. The clusters in RD50-MPW3 consist primarily of single

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub



a): Telescope Plane 4 **b):** DUT RD50-MPW3

Figure 8.5: Two pixels are hit on average by a single particle in the telescope, while the DUT has almost only single-pixel clusters.

pixels only, while the Mimosa26 sensor has multiple-pixel clusters. The cluster size is mainly dependent on the threshold and the depletion zone, which is, in turn, dependent on the geometry and bias voltage of the sensor.

8.3.3 Alignment

Alignment of the detector planes is done in five steps, listed below:

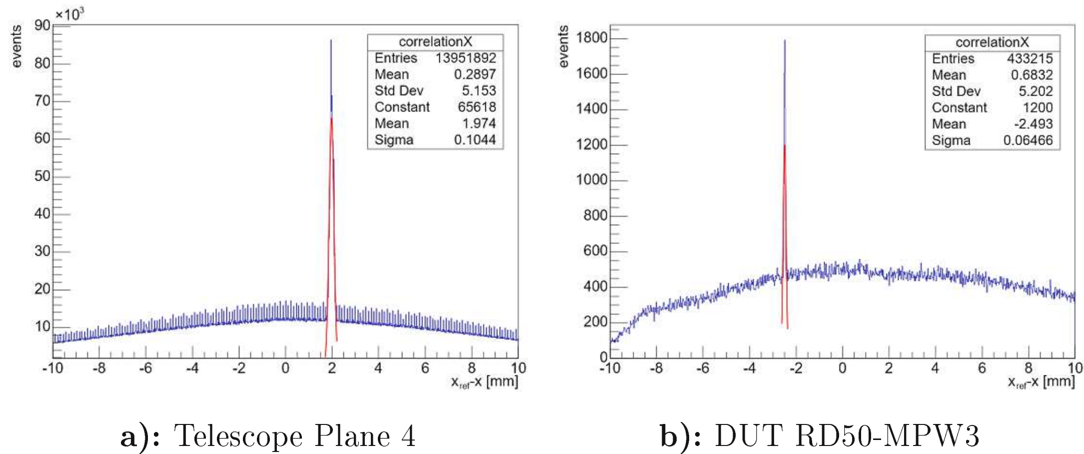
1. Manual laser alignment of the DUT relative to the telescope in x and y.
2. Measurement of z-positions with a ruler. As the z-position is a weak mode for the track, it is no longer corrected by the software.
3. Coarse alignment using correlations.
4. Precise track-based alignment of the telescope by minimizing the track fit χ^2.
5. Precise track-based alignment of the DUT by minimizing the DUT residual.

The origin is set at the center of the first plane, and plane two is chosen as the reference; thus, the x and y position of the second plane define positions (0,0) in the coordinate system, and all other planes are shifted with respect to it.

Correlations and Prealignment

The reference plane's position minus the plane's position defines correlation. Correlations are calculated for each plane separately. The clustering module defines the position of a hit, and all hits within an event are matched. The result should peak at the position of the plane. Prealignment shifts the plane by the coordinates of the correlation peak, so all correlations are centered around zero afterward.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub



a): Telescope Plane 4 **b):** DUT RD50-MPW3

Figure 8.6: The correlations show that the telescope plane needs to be shifted by 2 mm and the DUT by 2.5 mm. The first three values in the legend correspond to the full distribution, while the last three values are the fit parameters for the gauss-fit (in red) only. The underground is given by noisy pixels, which are correlated as well. The repeated pattern in figure 8.6a most likely originates from high-noise pixels.

Planes are typically aligned by a precision of around 100 µm using this technique. Plot 8.6 shows the correlations of telescope plane four and the DUT. To get the precise position of the peak, a gaussian is fit to the peak, and the mean of the fit is used as correction for the x and y position in this coarse alignment for this coarse alignment.

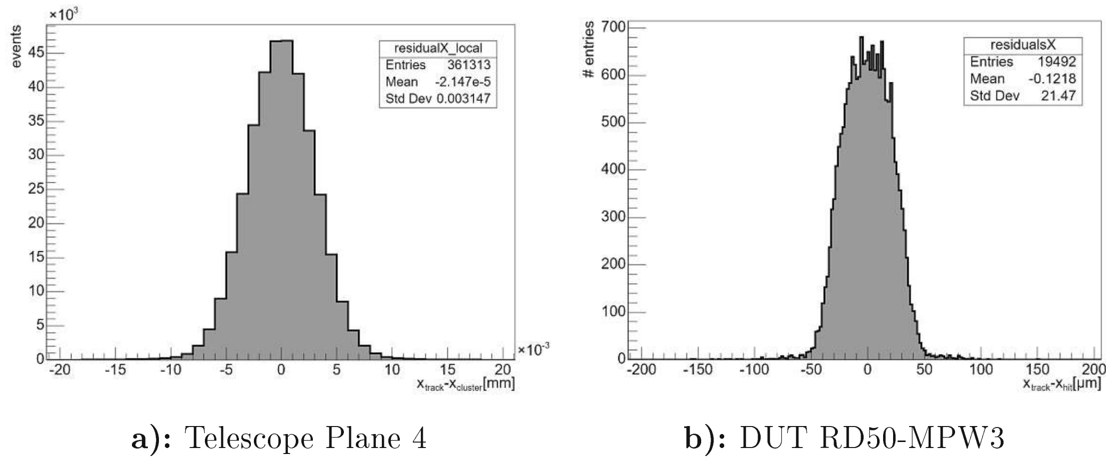
Precise Track-based Alignment

Track reconstruction is done by fitting a straight line through the positions measured in each plane of the reference telescope. In case of multiple hits within one event, the closest hit to the track is taken. In order to not match hits from a different track or noise, a spatial cut and a χ^2-cut are used in addition. The method only works if the planes are already coarsely aligned.

The fine alignment is based on these track reconstructions. It uses the coarsely aligned positions as starting point and performs track optimization based on the Minuit2 algorithm, a numeric minimization library. It reduces the sum of all distances between the reconstructed tracks and measured positions of each plane by iteratively moving around the planes. The iterative process is stopped once the precision is in the order of the least precise device in the alignment process. In these measurements, the Mimosa 26 planes and RD50-MPW3 DUT are aligned to a precision of a few microns and tens of microns, respectively, according to their expected binary resolution.

Plot 8.7 shows the residuals after the fine alignment. The reconstructed track's

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



a): Telescope Plane 4 **b):** DUT RD50-MPW3

Figure 8.7: The track-based alignment centers the offset of pixel hits around 0, while an ideal fit, hitting all points on the track, is given by an offset of 0.

distance minus the hit's position gives the residual. The track reconstruction is re-done after the alignment procedure with the updated positions. Those histograms peak around 0.
Hits considered for a track need to be in a search window of five times the plane's resolution, which equals around 100 µm for telescope planes and the DUT. One pixel of the cluster within this window is enough for a hit to be considered for the track, so the cluster center does not necessarily need to be within the window. This explains possible residuals outside this window as in figure 8.7b.

The χ^2 is computed to evaluate the precision of the track fit. It is divided by the number of degrees of freedom (ndof) for comparison to other fit methods. There are six degrees of freedom (three for the direction and three for the rotation). As said above, for the track generation via the reference, there is also a cut on the χ^2. It should be as low as possible and is given in plot 8.8.
After the alignment, the analysis uses the fine-aligned positions of all planes and only considers hits that fall within the aforementioned search window to filter out most noise hits.

8.3.4 DUT Analysis

The two most essential parameters for a tracking sensors are efficiency and spatial resolution, calculated in this section.

Efficiency

The efficiency is defined by the number of associated hits in the DUT divided by the number of tracks hitting the DUT. Tracks are extrapolated to the z-position of the DUT for this analysis. Due to the aforementioned search window of 100 µm (see section 8.3.3), the extrapolated position of the track and the cluster in the DUT are allowed to differ by 100 µm only. Otherwise, the hit in the DUT is not

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

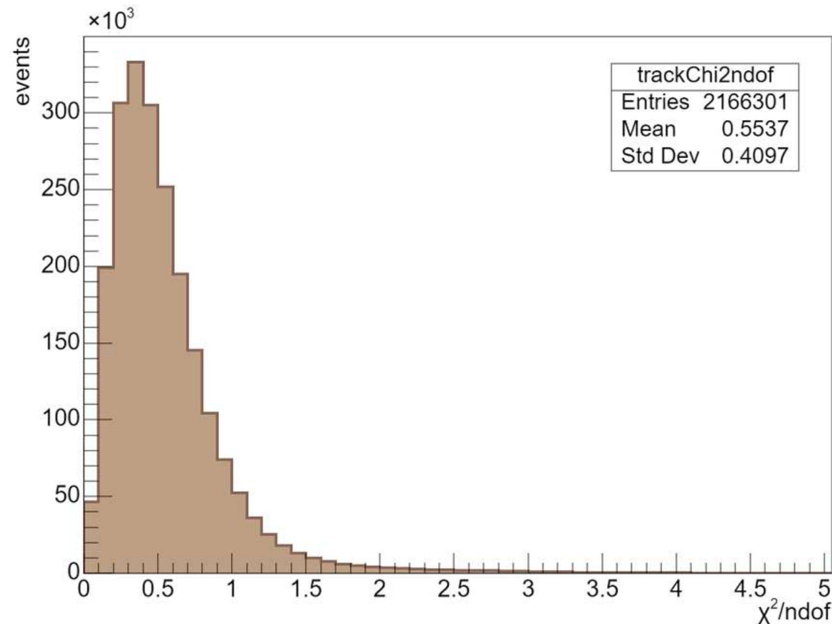


Figure 8.8: The sum of squared random variables defines the χ^2. The random variables for the fit are given by the distances between the track position and hit position in all three dimensions on each plane.

counted. Figure 8.9 shows a two-dimensional map of these deviations, showing a standard deviation of around 21 µm for most events.

The efficiency calculation must consider the overflow problem mentioned in section 8.1.2. An event from the DUT can only have one particular time instead of multiple, so analysis needs to run for all overflows separately. Table 8.1 shows the output of eight runs, which all use the same data set while only shifting the time of DUT by one overflow value. These runs consist of 300.000 events with ~2.100.000 tracks each. Out of these, the same 107.702 tracks are considered for the analysis for each time shift. Most other tracks are discarded, as they do not hit the DUT.

The efficiency of the sensor is quoted as the sum over all time shifts, which equals 55.24%. These values assume no double-counting between the time shifts, which has to be considered to get a more realistic value. The double-counting can only happen when an event is considered for two time-shifts. This case has a probability of 7%, as discussed in section 8.1.2 and should be deducted from the overall efficiency. Thus, the final quote for the **efficiency of RD50-MPW3** is given by $55.24\% * 0.93 = \mathbf{51.4\%}$.

Spatial Resolution

The spread of the residuals gives the spatial resolution of the DUT. Typically, the standard deviation is quoted as spatial resolution. Plot 8.10 shows the residuals for RD50-MPW3 in both dimensions.

The result in plot 8.10 contains the telescope's uncertainty. Thus, the telescope's spatial resolution is typically subtracted to quote a reasonable estimation

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

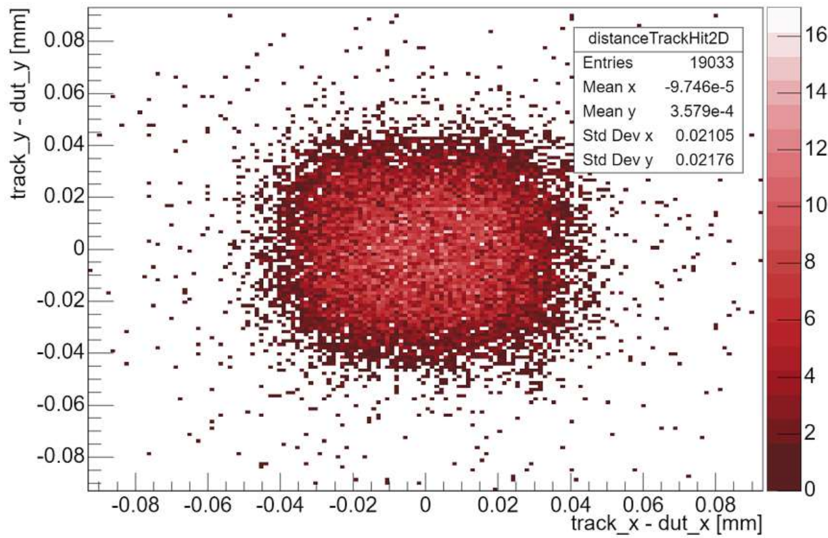
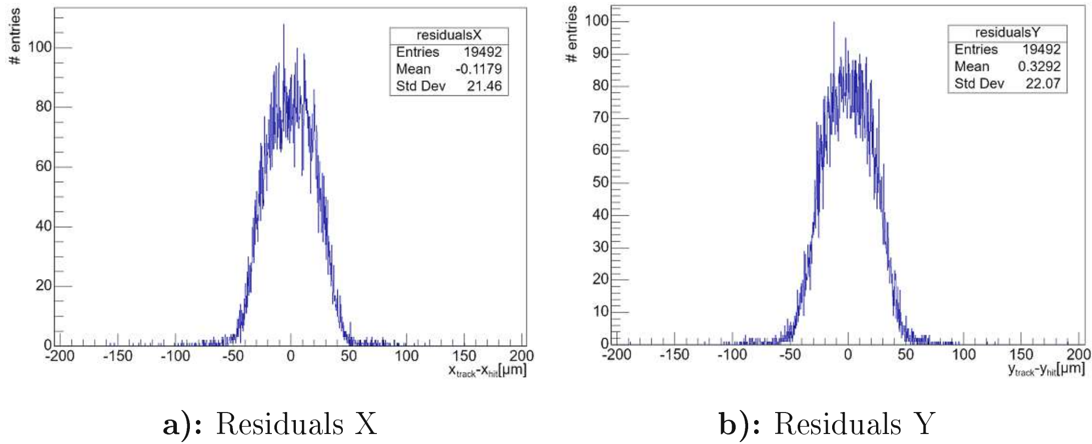


Figure 8.9: The difference between the reconstructed track and the measured hit position at the DUT is at most 40 µm in both dimensions.

Overflows shifted	Matched hits	Efficiency
-2	568	0.53%
-1	1439	1.34%
0	8540	7.93%
1	14869	13.81%
2	19033	17.67%
3	11106	10.31%
4	2879	2.67%
5	1054	0.98%

Table 8.1: The efficiency for various overflows is given. A shift by one overflow equals a time shift of 3.2768 ms. The efficiency has also been measured of a random time shift of 1.4 ms, where only matches with noise hits are expected. The efficiency for such a shift is calculated at 0.352%. Additional overflows than the listed ones are not considered, as the efficiency gets close to the value for a random shift.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub



a): Residuals X **b):** Residuals Y

Figure 8.10: The standard deviation is similar for both dimensions, which is expected due to the quadratic layout of the pixel. Residuals are shown for a single time shift corresponding to 3 overflows only.

of the resolution of the DUT. The telescope's spatial resolution is in the order of 1.83 µm, as stated in [31]. Considering error propagation, the resolution of the DUT (σ_{DUT}) is calculated with formula 8.1. With this correction, the final result for the **resolution of RD50-MPW3 is 21.38 µm in x-direction and 21.99 µm in y-direction**.

$$\sigma_{DUT} = \sqrt{\sigma_{total}^2 - \sigma_{telescope}^2} \tag{8.1}$$

8.4 Conclusion

Work done to develop the readout of RD50-MPW2 has been very valuable while implementing the readout chain for RD50-MPW3. No hardware needs to be adapted apart from the small carrier board of the ASIC. The firmware needs a complete re-design from an analog conversion and digital readout to a high-speed, digital-only system. Changing to a time-synchronous readout mode adds more complexity to the system but significantly increases the possible readout rate.
The test beam setup is nearly identical to RD50-MPW2 as both setups use a telescope integrated into the EUDAQ2 framework, requiring no framework change. This greatly demonstrates the usefulness and feasibility of a commonly used, standardized data acquisition framework and trigger logic unit.

The presented data were taken during the group's first large-scale test beam. Thus, quite a few problems were solved during this test beam, which had the commissioning of the setup as the most important goal. The number of steps taken to synchronize the data and the effort put into the analysis, figuring out how to do the time shifts, did cost much time. Nevertheless, some runs with good data have been taken and analyzed in this chapter. Due to some problems during data taking, all runs are rather short, so the statistics of the presented data have yet to be improved during further test beams.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

The reported efficiency of RD50-MPW3 is below expectation. However, due to a lack of preparation time, high noise in the sensor was found during the test beam only, as explained in [30]. Therefore, the threshold of the discriminator had to be set 200 mV higher than expected from measurements of RD50-MPW2. Furthermore, half of the pixel matrix is masked to find a common threshold for all active pixels. Consequently, the charge sharing is quite low, which can be seen by the high amount of single-pixel clusters. If the charge is shared between two pixels, the threshold is not reached most likely anymore, and no hit can be detected. The high noise is currently being investigated, and more test beams are in preparation to mitigate this behavior.

The spatial resolution corresponds very well with the expected result for a binary readout, which is given by $pitch/\sqrt{(}12)$, according to [3], chapter E.1. The pitch for RD50-MPW3 is 62 µm, thus the expected value of the spatial resolution is 17.9 µm; which in good agreement with the measurement.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

9 Conclusion

This work describes the first efforts in establishing ASIC and DMAPS design at HEPHY. Much effort was invested in getting experience in the necessary tools and building the toolchain. This is described in chapter 5, following an introduction to tracking in high energy physics and the DMAPS technology itself.

9.1 Digital Design

A quick introduction to digital design in industry, including requirements engineering, is given in section 5.2.4. Evaluating the design process of RD50-MPW3, a definition of detailed requirements before starting the design would have been useful. The design requirements of RD50-MPW3 listed in section 7.1.1 are too vague and lack important definitions. Most crucially, a precise separation of the analog and digital tasks is missing, including an interface definition. This did not only cause many questions and discussion rounds during the design process but also led to a not working interface between the periphery and the matrix. It was found out during testing that a readout signal of the double column is arriving too fast, violating timing constraints of the analog double column. The issue was resolved by slowing down the clock of the periphery. Fortunately, this is not a problem for RD50-MPW3, as the data rate is about an order of magnitude lower than the capability of the periphery. However, this mistake could have been avoided by a clear interface definition as a design requirement in the beginning.
A detailed verification may find such issues. RD50-MPW3 has some basic tests implemented, including drivers and monitors for the most important external interfaces. This is very useful for quickly writing tests and is similar to the definition of UVM agents. However, those tests did not cover analog signals nor the interface between the matrix and the periphery. The periphery itself does work at design speed, and not even a single issue was found during operation of the periphery alone. The same is true for the operation of the matrix. To conclude the verification process of RD50-MPW3, much more mixed-signal verification, covering especially the interface between the matrix and periphery, would have been beneficial. On the contrary, simulations of the analog matrix and basic tests for the digital periphery seem to be enough for testing the individual parts, so implementing a full UVM framework does not seem to be needed.

At the end of this section, a recapitulation of the design goals listed in section 7.1.1 is given. A short version of the list is given hereafter for quick reference. Details about the goals can be found in section 7.1.1.

A) full matrix for tracking

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

B) double-column architecture

C) separation of grounds

D) shielding lines

E) high-speed readout at up to 640 MHz

F) digital periphery running at 40 MHz, driven by a single 640 MHz clock

G) idle pattern

H) framing of the data

I) 8-bit/10-bit encoding

J) differential pads

K) analog-on-Top design flow

L) synchronizable readout chain

Shielding lines and a double column architecture have been implemented as shown in section 7.1.3, meeting design requirements B) and D).
The readout is working at 640 MHz and only requires a single input clock at the same speed, fulfilling requirements E) and F). The speed has been tested upon reception of the chip. Due to the communication issue mentioned above, all measurements use a 320 MHz clock.
The phase finding algorithm described in section 8.1.1 is working; otherwise, no digital data could be read out. The automated frame length and encoding are also needed for the readout chain, for instance, to provide a 16-bit timestamp, as mentioned in section 8.1.2. Thus, requirements G), H) and I) and implicitly covered by the functional digital readout chain.
Differentials pads have been used to fulfill requirements J). No glitches have been observed so far.
The design uses an analog-on-top approach, meeting requirement K). Hundreds of signal lines, especially at the interface between the matrix and periphery, require manual routing, which is already much work for this chip size. A digital-on-top approach might be indispensable for a bigger chip size.
Investigations into the high noise problem found an unwanted connection between the digital ground and the analog circuitry. This is a strong hint for the noise problem; however, more measurements and simulations are needed to prove this. In any case, requirement C) is not met.
Lastly, the essential design goals A) and L) to develop a matrix capable of tracking hits and implementing a synchronizable readout system have been reached, as demonstrated during the beam tests performed at CERN in section 8.3.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

9.2 DMAPS for Tracking in Future Collider Experiments

Table 3.1 does list the most important requirements for future tracking sensors at various experiments.
With a spatial resolution of 21 µm, RD50-MPW3 is suited for the CMS Phase 3 upgrade, especially considering that this value will decrease once the noise problem is fixed and charge sharing can be measured. A further reduction of spatial resolution can only be reached with an even higher granularity which is not possible without compromising other features, as the pixel size has to be reduced. Choosing a process with a smaller feature size is one option to achieve the goals for the other detectors.
The maximum hit rate has not been measured so far, but considering a readout speed of 640 MHz and the encoding, which increases the amount of data to be sent, a theoretical maximum of 16 Mhz hit rate over the full matrix can be handled by the periphery. Given an active area of ≈16 mm^2, this equals to 1 MHz/cm^2, which the periphery can handle. Thus, RD50-MPW3 is not suited for any of the mentioned detectors with its current readout. However, increasing the number of transmission lines can easily increase the maximum possible hit rate.
The time resolution of RD50-MPW3 is limited by the speed of the internal timestamp and has a precision of 25 ns. Thus, RD50-MPW3 is suited for BelleII, ILC and FCC-ee. Using an arbitrarily high timestamp frequency is impossible, as this increases power consumption.
RD50-MPW2 has been characterized after irradiation up to $2 * 10^{15}\ N_{eq}/cm^2$ by other institutes and is still operational at this fluence. Unfortunately, the efficiency loss could not be measured, as RD50-MPW2 has no digital readout. Those measurements are planned with RD50-MPW3 and irradiation up to even higher fluences will be done. It is expected that RD50-MPW3 will work up to a fluence of at least $1 * 10^{16}\ N_{eq}/cm^2$, although this value still has to be confirmed. With this radiation tolerance, RD50-MPW3 could be used in all of the mentioned experiments except FCC-hh.

Summing up the results, RD50-MPW3 has a high radiation tolerance, which is the most crucial target of the RD50 group. Furthermore, the DMAPS technique has much space for improving time resolution and hit rate. Thus, RD50-MPW3, implementing a large fill-factor design including a binary digital readout in every pixel, is a promising candidate for future tracking sensors under a harsh radiation environment.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Acknowledgements

At first, my gratitude goes to my head supervisor of this thesis, Privatdoz. Dipl.-Ing. Dr. Christoph Schwanda of the Austrian Academy of Sciences for showing me the correct (scientific) direction and allowing this project to be my own work. Furthermore, I would like to thank my second supervisor Univ.-Lektor Dipl.-Ing. Dr. Thomas Bergauer, for being my mentor and supporting my scientific career for almost 5 years by now.

I like to offer my deepest gratitude to Christian Irmler, who work with me during the whole time of the project to make it possible. I am looking back at countless hours of video calls during the design of RD50-MPW3, which enlightened almost every day during this pandemic-prone phase of the project.

I had the honor to help supervising the master thesis from Klemens Flöckner and Bernhard Pilsl. Both of you did an excellent job and I think I have been lucky to get the two of you as support for my project. Klemens and Bernhard, I wish you all the best for your future, hopefully scientific, career.

I am thankful to the colleagues from the RD50 collaboration and the RD50-CMOS group for supporting my work and fruitful discussion about results. A very special thanks goes to the groupleader, Dr. Eva Vilella, for supporting my work. Moreover, I would like the thank the CERN testbeam crew for the help and countless hours spend together taking data. Bernhard, Bojan, Chenfan, Christian, Douwe, Sam and Uwe: I enjoyed every minute with you at CERN.

My gratitude also goes to my dearest colleagues and friends at HEPHY, Maximilian Babeluk, Andreas Bauer, Kostas Damanakis, Klaus-Dieter Fischer, Stefanie Kaser, Veronika Kraus, Helmut Steininger, Felix Ulrich-Pur, Moritz Wiehe and Hao Yin for always lightening up the mood.

A special mention also goes to Christian Irmler, Uwe Krämer, Bernhard Pilsl and Elisabeth Sieberer for proofreading this thesis.

For providing motivational support to ensure completion of this thesis, I would like to thank my parents Gerhard and Elisabeth and my love Sarah♡.

This work has been partly performed in the framework of the CERN-RD50 collaboration. It has received funding from the Austrian Research Promotion Agency FFG, grant number 878691, and from the European Union's Horizon 2020 Research and Innovation program under grant agreement 101004761 (AIDAinnova).

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Bibliography

[1] US Department of Energy - Office of Science. *Basic Research Needs for High Energy Physics Detector Research & Development.* Report. 14th Dec. 2019. URL: https://science.osti.gov/-/media/hep/pdf/Reports/2020/DOE_Basic_Research_Needs_Study_on_High_Energy_Physics.pdf.

[2] ECFA Detector R&D Roadmap Process Group. *The 2021 ECFA detector research and development roadmap.* Tech. rep. Geneva, 2020. DOI: 10.17181/CERN.XDPL.W2EX. URL: https://cds.cern.ch/record/2784893.

[3] Hermann Kolanoski and Norbert Wermes. *Particle detectors : fundamentals and applications.* eng. Oxford New York, NY: Oxford University Press, 2020. ISBN: 0198858361.

[4] Frank Hartmann. *Evolution of Silicon Sensor Technology in Particle Physics.* 6th Nov. 2017. ISBN: 978-3-319-64436-3. DOI: https://doi.org/10.1007/978-3-319-64436-3.

[5] C. Patrignani et al. (Particle Data Group). *Chin. Phys. C, 40, 100001 (2016) and 2017 update.* 2017. URL: https://pdg.lbl.gov/2017/reviews/contents_sports.html (visited on 03/02/2023).

[6] Michael Moll. 'Radiation Damage in Silicon Particle Detectors: Microscopic Defects and Macroscopic Properties'. Hamburg University, 1999.

[7] Rudolf Frühwirth and Are Strandlie. *Pattern Recognition, Tracking and Vertex Reconstruction in Particle Detectors.* 17th Feb. 2021. ISBN: 978-3-030-65771-0. DOI: https://doi.org/10.1007/978-3-030-65771-0.

[8] Apollinari G. et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1.* CERN Yellow Reports: Monographs. Geneva: CERN, 2017. DOI: 10.23731/CYRM-2017-004. URL: https://cds.cern.ch/record/2284929.

[9] *The Phase-2 Upgrade of the CMS Tracker.* Tech. rep. Geneva: CERN, 2017. DOI: 10.17181/CERN.QZ28.FLHW. URL: https://cds.cern.ch/record/2272264.

[10] Leonardo Rossi et al. *Pixel Detectors.* 8th Apr. 2006. ISBN: 978-3-540-28333-1.

[11] Eva Vilella Figueras. 'Recent depleted CMOS developments within the CERN-RD50 framework'. In: *Proceedings of The 28th International Workshop on Vertex Detectors - PoS(Vertex2019).* Vol. 373. 2020. DOI: 10.22323/1.373.0019.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

[12] European Cooperation for Space Standardization. *ECSS-Q-ST-60-02C – ASIC and FPGA development*. 31st July 2008. URL: `https://ecss.nl/standard/ecss-q-st-60-02c-asic-and-fpga-development/` (visited on 18/01/2023).

[13] RTCA. *DO-254 ELECTRONIC*. 18th Jan. 2023. URL: `https://www.rtca.org/products/` (visited on 18/01/2023).

[14] International Requirements Engineering Board. *Handbook for the CPRE Foundation Level according to the IREB Standard. Education and Training for Certified Professional for Requirements Engineering (CPRE) Foundation Level*. Vol. 1.1.0. 1st Sept. 2022. (Visited on 18/01/2023).

[15] Jeremy Dick, Elizabeth Hull and Ken Jackson. *Requirements Engineering*. eng. 4th ed. 2017. Cham: Springer International Publishing Imprint: Springer, 2017. ISBN: 3319610732. URL: `10.1007/978-3-319-61073-3`.

[16] Chenfan Zhang et al. 'Development of RD50-MPW2: a high-speed monolithic HV-CMOS prototype chip within the CERN-RD50 collaboration'. In: vol. PoS TWEPP2019. Mar. 2020. DOI: `10.22323/1.370.0045`.

[17] Tomas Vanat. 'Caribou – A versatile data acquisition system'. In: *PoS* TWEPP2019 (2020). DOI: `10.22323/1.370.0100`.

[18] Klemens Flöckner. *Development of a DAQ System for Depleted Monolithic Active Pixel Sensors*. eng. Wien, 2022.

[19] P. Baesso, D. Cussans and J. Goldstein. 'The AIDA-2020 TLU: a flexible trigger logic unit for test beam facilities'. In: *Journal of Instrumentation* 14.09 (Sept. 2019). DOI: `10.1088/1748-0221/14/09/p09019`.

[20] Patrick Sieberer et al. 'Readout system and testbeam results of the RD50-MPW2 HV-CMOS pixel chip'. In: *Journal of Physics: Conference Series* 2374.1 (Nov. 2022). DOI: `10.1088/1742-6596/2374/1/012096`.

[21] Y. Liu et al. 'EUDAQ2—A flexible data acquisition software framework for common test beams'. In: *Journal of Instrumentation* 14.10 (Oct. 2019). DOI: `10.1088/1748-0221/14/10/p10033`.

[22] Dominik Dannheim et al. 'Corryvreckan: a modular 4D track reconstruction and analysis software for test beam data'. In: *Journal of Instrumentation* 16.03 (Mar. 2021). DOI: `10.1088/1748-0221/16/03/p03008`.

[23] Ricardo Marco Hernández. 'Latest Depleted CMOS Sensor Developments in the CERN RD50 Collaboration'. In: *Proceedings of the 29th International Workshop on Vertex Detectors (VERTEX2020)*. DOI: `10.7566/JPSCP.34.010008`.

[24] Bojan Hiti et al. 'Characterisation of analogue front end and time walk in CMOS active pixel sensor'. In: *Journal of Instrumentation* 16.12 (Dec. 2021). DOI: `10.1088/1748-0221/16/12/p12020`.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

[25] Felix Ulrich-Pur et al. 'Commissioning of low particle flux for proton beams at MedAustron'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1010 (2021). ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2021.165570`.

[26] F. Ulrich-Pur et al. 'Imaging with protons at MedAustron'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 978 (2020). ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2020.164407`.

[27] Patrick Sieberer et al. 'Design and characterization of depleted monolithic active pixel sensors within the RD50 collaboration'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1039 (2022). ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2022.167020`.

[28] P. Sieberer et al. 'RD50-MPW3: a fully monolithic digital CMOS sensor for future tracking detectors'. In: *Journal of Instrumentation* 18.02 (Feb. 2023), p. C02061. DOI: `10.1088/1748-0221/18/02/C02061`.

[29] Ivan Perić et al. 'The FEI3 readout chip for the ATLAS pixel detector'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 565.1 (2006). Proceedings of the International Workshop on Semiconductor Pixel Detectors for Particles and Imaging. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2006.05.032`.

[30] Bernhard Pilsl. *Data-Acquisition-Systems for Depleted Monolithic Active Pixel Sensors*. Wien, 5th Dec. 2022. URL: `https://doi.org/10.34726/hss.2022.105002`.

[31] Hendrik Jansen et al. 'Performance of the EUDET-type beam telescopes'. In: *EPJ Techniques and Instrumentation* 3.1 (Oct. 2016).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub