Karlsruhe Institute of Technology

Development and characterisation of monolithic silicon detectors

Zur Erlangung des akademischen Grades eines

DOKTORS DER INGENIEURWISSENSCHAFTEN
(Dr.-Ing.)

von der KIT-Fakultät für Elektrotechnik und Informationstechnik des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von

M.Sc. Ekaterina Trifonova-Ehrler

geboren in Wolschsk (Mari El, Russland)

Tag der mündlichen Prüfung:	28. November 2025
Hauptreferent:	Prof. Dr. rer. nat. Marc Weber
Korreferent:	Prof. Dr. rer. nat. Ivan Perić
Korreferent:	Prof. Dr. rer. nat. Steven Worm

Karlsruhe Institute of Technology

Entwicklung und Charakterisierung von monolithischen siliziumbasierten Detektoren

Zur Erlangung des akademischen Grades eines

DOKTORS DER INGENIEURWISSENSCHAFTEN
(Dr.-Ing.)

von der KIT-Fakultät für Elektrotechnik und Informationstechnik des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von

M.Sc. Ekaterina Trifonova-Ehrler

geboren in Wolschsk (Mari El, Russland)

Tag der mündlichen Prüfung:	28. November 2025
Hauptreferent:	Prof. Dr. rer. nat. Marc Weber
Korreferent:	Prof. Dr. rer. nat. Ivan Perić
Korreferent:	Prof. Dr. rer. nat. Steven Worm

Abstract

Semiconductor devices have become essential tools in both fundamental research and applied technologies. Their applications span from microelectronics and power devices to optoelectronics and sensors. Broadly defined, a sensor is any device that detects or measures physical signals from the external environment and converts them into readable or processable information. Many conventional semiconductor devices, including photodiodes, transistors, and solar cells, function as sensors by responding to electrical, optical, or radioactive stimuli. Among the various materials used for sensor development, silicon remains the dominant platform due to its favourable electronic properties, process maturity, scalability, and availability.

This thesis presents the development and characterisation of advanced silicon-based particle detectors for accelerator-driven applications in both high-energy physics (HEP) and medical physics. The work covers the full detector development chain, from conceptual design and physical modelling to laboratory testing and complete module assembly. While solid-state detectors are well-established in high-energy physics, where performance requirements continue to grow in terms of spatial resolution, timing precision, and radiation hardness, there is also increasing interest in their implementation in medical applications such as hadron therapy, where their fast response and precision can improve beam monitoring systems.

Two core Monolithic Active Pixel Sensors (MAPS) technologies are explored: High-Voltage CMOS (HV-CMOS) for use in large-scale particle physics experiments and medical accelerators, and Bipolar CMOS (BiCMOS) as a base for high-speed, radiation-tolerant detectors. A detailed Technology Computer-Aided Design (TCAD) simulation framework was developed to optimise device geometry, to evaluate breakdown behaviour, and to predict behaviour under irradiation. These results guided the design and further chip development. Extensive characterisation of the fabricated sensors was performed using radioactive sources, charge injection, X-ray tubes, and particle beams to validate simulation models and optimise detector performance. This work demonstrates how predictive modelling and systematic testing can enable robust, scalable sensor designs tailored to the specific requirements of both HEP and medical applications.

In addition to sensor development, this work includes hands-on experience in high-density interconnection techniques and full module assembly, as demonstrated through contributions to the Silicon Tracking System (STS) of the Compressed Baryonic Matter (CBM) experiment. This highlights the importance of integrating practical knowledge and production expertise in real-world detector systems.

Altogether, the thesis demonstrates the necessity of addressing all stages of development, from physical modelling and process optimisation to experimental validation and system integration, to advance the next generation of silicon-based detectors for particle physics and beyond.

Zusammenfassung

Sowohl aus der Grundlagenforschung als auch aus der angewandten Wissenschaft sind Halbleiterbauelemente nicht mehr wegzudenken. Ihre Anwendung reicht von der Mikro- und Leistungselektronik bis hin zur Optoelektronik und zu Sensoren. Allgemein gesprochen ist ein Sensor ein nicht weiter definiertes Gerät, das ein physisches Signal aus seiner Umgebung detektiert oder misst und in lesbare oder prozessierbare Information übersetzt. Viele konventionelle Halbleiterelemente, einschließlich Photodioden, Transistoren und Solarzellen, fungieren als Sensoren, indem sie auf elektrische, optische oder radioaktive Stimuli reagieren. Trotz zahlreicher Alternativen ist Silizium wegen seiner günstigen elektronischen Eigenschaften, des ausgereiften Produktionsprozesses, der Skalierbarkeit und der Verfügbarkeit weiterhin das dominierende Material im Sensordesign.

In dieser Doktorarbeit werden die Entwicklung und Charakterisierung von fortschrittlichen siliziumbasierten Teilchendetektoren für die Beschleunigerphysik sowohl in der Hochenergiephysik (HEP) als auch in der Medizinphysik vorgestellt. Sie umfasst dabei die gesamte Detektorentwicklungskette von konzeptuellem Design und physikalischer Modellierung bis hin zu Labortests und Modulbau. Während Festkörperdetektoren in der Hochenergiephysik, wo die Leistungsanforderungen hinsichtlich Ortsauflösung, Zeitauflösung und Strahlenhärte stetig wachsen, seit langem etabliert sind, gibt es auch ein wachsendes Interesse daran, sie in medizinische Anwendungen wie der Hadronentherapie einzusetzen, wo ihre schnelle Ansprechzeit und Präzision die Strahlmonitorsysteme verbessern können.

Zwei verschiedene Monolithische Aktive Pixel-Sensortechnologien (MAPS) werden untersucht: Hochvolt-CMOS (HV-CMOS) für den Einsatz in großen Teilchenphysikexperimenten und medizinischen Teilchenbeschleunigern einerseits und Bipolar-CMOS (BiCMOS) als Grundlage für strahlenharte Detektoren mit hoher Zeitauflösung andererseits. Ein detailliertes Technology Computer-Aided Design (TCAD) Simulationsframework wurde zur Sensorsimulation entwickelt, um deren Geometrie zu optimieren, das Durchbruchverhalten zu untersuchen und ihr Verhalten unter Bestrahlung vorherzusagen. Dessen Simulationsergebnisse leiteten den Designprozess und die weitere Chipentwicklung an. Ausgiebige Charakterisierung der produzierten Sensoren wurde unter Einsatz von radioaktiven Quellen, Ladungsinjektionen, Röntgenröhren und Teilchenstrahlen durchgeführt, um die Simulationsmodelle zu verifizieren und die Detektorleistung zu optimieren. Diese Arbeit demonstriert, wie Modellierung zur Verhaltensvorhersage und systematisches Testen, das robuste und skalierbare Design von Sensoren ermöglichen, die auf die spezifischen Anforderungen sowohl von HEP als auch von medizinischen Anwendungen zugeschnitten sind.

Zusätzlich zur Sensorentwicklung beinhaltet diese Arbeit auch praktische Erfahrungen aus der Verbindungstechnik für hochintegrierte Schaltungen und aus dem Detektormodulbau, wie

durch die Mitwirkung am Silizium Tracker System des Komprimierte Baryonische Materie Experiments (CBM) gezeigt wurde. Dies hebt die Wichtigkeit von Wissen über die praktische Integration und Produktion von tatsächlichen Detektorsystemen hervor.

Zusammengefasst zeigt diese Arbeit auf, wie entscheidend es ist, alle Schritte einer Entwicklung von der physikalischen Modellierung und Prozessoptimierung bis hin zur experimentellen Validierung und Systemintegration im Blick zu haben, um die nächste Generation von siliziumbasierten Detektoren für die Teilchenphysik voranzubringen.

To my family — in the past, in the present, and in the future.

Contents

Abstract . . . . . **v**

Zusammenfassung . . . . . **vii**

1 Introduction **1**

I Development of silicon sensors **9**

2 Fundamentals of silicon sensors **11**
2.1 Semiconductor physics 11
2.1.1 Semiconductor materials and crystal structures 12
2.1.2 Energy bands and carrier concentrations in equilibrium 15
2.1.3 Carrier transport 19
2.1.4 Generation and recombination 23
2.1.5 Basic equations 24
2.1.6 The $p - n$ junction 26
2.2 Silicon as a primary material for semiconductor particle sensors 30
2.2.1 Working principle of silicon sensors 31
2.2.2 Radiation damage in silicon sensors 33

3 Design and fabrication of silicon sensors **37**
3.1 Manufacturing of silicon sensors 37
3.1.1 Crystal growth 37
3.1.2 Front-end processing steps 39
3.2 Design foundations for silicon sensors 45
3.3 Silicon sensors segmentation 46
3.3.1 Strip sensor 46
3.3.2 Pixel sensor 47
3.4 Coupling and biasing of silicon sensors 48
3.4.1 Coupling schemes 48
3.4.2 Biasing techniques 48
3.5 Readout concepts in segmented silicon sensors 50
3.5.1 Hybrid sensors 50
3.5.2 Monolithic sensors 51
3.6 Isolation and guarding techniques 52

4 Simulations of semiconductor devices — Technology Computer-Aided Design (TCAD) 55
4.1 Modelling of device geometry . . . 59
4.1.1 Parameters . . . 60
4.1.2 Geometrical structure . . . 61
4.1.3 Doping . . . 61
4.1.4 Contacts . . . 64
4.1.5 Mesh . . . 66
4.2 Numerical simulations and physical models . . . 69
4.2.1 Device section . . . 70
4.2.2 File section . . . 71
4.2.3 Electrode section . . . 72
4.2.4 Physics section . . . 73
4.2.5 Plot section . . . 80
4.2.6 System section . . . 81
4.2.7 Math section . . . 84
4.2.8 Solve section . . . 87
4.3 Beam simulation and radiation damage model . . . 92
4.3.1 Minimum Ionising Particle (MIP) simplification in TCAD . . . 93
4.3.2 Traps and radiation damage modelling . . . 95
4.3.3 Transient analysis and solvers . . . 98
4.4 Summary for TCAD simulations and applications . . . 100

II Development and investigation of monolithic HV-CMOS sensor technology 103

5 The HV-CMOS technology . . . 105

6 HV-CMOS sensors for ion beam monitoring — HitPix . . . 109
6.1 Application description . . . 109
6.2 Detector description . . . 114
6.3 Chip description . . . 117
6.4 Chip design evolution . . . 119
6.5 Sensor design and TCAD modelling . . . 121
6.6 Operational conditions and choice of substrate . . . 128
6.6.1 Substrate selection and depletion region analysis . . . 128
6.6.2 Power consumption evaluation . . . 132
6.7 Electric fields and breakdown simulations . . . 134
6.7.1 Breakdown optimisation through design adjustments . . . 135
6.7.2 Electric field simulation . . . 136
6.7.3 Electrostatic potential simulations . . . 138
6.7.4 Conclusion and recommendation . . . 140
6.8 Charge collection and timing using beam simulations . . . 140
6.8.1 Charge carrier dynamics in TCAD simulations . . . 142
6.8.2 Transient simulation . . . 145
6.8.3 Charge collection efficiency calculation . . . 146

6.8.4 Final observations . . . 147
6.9 Summary for HitPix development . . . 149

7 HV-CMOS sensors for future gamma-ray telescopes — AstroPix . . . 151
7.1 Application description . . . 151
7.2 Chip description . . . 153
7.3 Evolution of chip designs . . . 156
7.4 Sensor design and TCAD modelling . . . 157
7.5 Choice of substrate and depletion region analysis . . . 159
7.6 Breakdown optimisation and enhancement of design stability . . . 164
7.6.1 Metallisation . . . 165
7.6.2 Metal–insulator–semiconductor (MIS) physics and its implications for sensor design . . . 171
7.6.3 Influence of p-stop on electric field distribution . . . 175
7.7 Outcome and recommendation . . . 176

8 Investigation of heavy-ion beam influence on HV-CMOS sensors . . . 179
8.1 Efficiency loss under heavy-ion beam irradiation . . . 179
8.2 Influence of single-event effects (SEE) on CMOS architecture . . . 182
8.2.1 The funnelling effect . . . 183
8.2.2 Simulation of the funnelling effect for the HV-CMOS designs . . . 184
8.2.3 Susceptibility to single-event latch-up (SEL) . . . 186
8.3 Investigation of SEE influence on HV-CMOS . . . 188
8.3.1 Analysis procedure of the ion beam tests results . . . 190
8.3.2 Performance of the AstroPix chip family . . . 193
8.3.3 Performance of the ATLASpix1 chip family . . . 196
8.3.4 Performance of MuPix8 chip . . . 202
8.3.5 Performance of the HitPix chip family . . . 205
8.4 Summary and suggestions . . . 211

9 Summary for HV-CMOS sensor technology development . . . 213

III Bi-CMOS technology for high-speed monolithic sensors for particle physics 217

10 Motivation for technology application . . . 219
10.1 The BiCMOS technology . . . 219
10.1.1 Heterojunctions . . . 220
10.1.2 Heterojunction Bipolar Transistors (HBT) . . . 221
10.1.3 Bandgap engineering . . . 222
10.1.4 Properties of SiGe alloys . . . 223
10.1.5 The SiGe HBT . . . 224
10.1.6 The SiGe HBT BiCMOS . . . 226
10.2 BeMiPix concept and project motivation . . . 227

11 Pixel sensor with miniature collection node — BeMiPix . . . 231
11.1 Sensor design and TCAD simulation . . . 231
11.1.1 Development of the structure . . . 232
11.1.2 Readout electronic design . . . 235
11.1.3 Optimising pixel pitch for breakdown behaviour and depletion . . . 237
11.1.4 Charge collection and charge sharing simulations . . . 238
11.2 Layout and production . . . 242
11.3 Characterisation and evaluation . . . 243
11.4 Sensor current and breakdown investigation . . . 245
11.4.1 Dark current measurements . . . 246
11.4.2 Sensor current evaluation with X-ray photons . . . 246
11.5 Temperature dependence measurements . . . 250
11.5.1 Thermal generation of charge carriers and leakage current . . . 251
11.5.2 Temperature-dependent IV characteristics . . . 252
11.6 Waveform analysis procedure . . . 254
11.7 Source measurements with Iron-55 and Strontium-90 . . . 255
11.7.1 Charge collection performance and depletion depth evaluation . . . 256
11.7.2 Hit rate evaluation . . . 258
11.7.3 Rise time characterisation . . . 259
11.7.4 Iron-55 calibration and charge response validation . . . 261
11.8 Laser scanning measurements for time and spatial resolutions . . . 264
11.8.1 Measurement procedure . . . 264
11.8.2 Sensitivity analysis and depletion characterisation . . . 266
11.8.3 Time resolution analysis . . . 269

12 Summary of BeMiPix development in SiGe BiCMOS technology . . . 273

IV Semiconductor detectors production for high-energy physics experiments 277

13 High-density interconnection techniques . . . 279
13.1 Introduction to high-density interconnection techniques for sensor assembly . 279
13.1.1 Wire bonding . . . 282
13.1.2 Flip chip bonding . . . 285
13.1.3 Tape Automated Bonding (TAB) . . . 287
13.1.4 Gold bonding at IPE KIT — infrastructure and machines . . . 289
13.2 Ultra-thin gold wire bonding for small pad-size SiGe devices — UniGe . . . 293
13.3 Gold bump bonding for The Pixelated Silicon Drift Detector — PixDD . . . 296

14 Compressed Baryonic Matter experiment at the future Facility for Antiproton and Ion Research . . . 301
14.1 Introduction to the Facility for Antiproton and Ion Research (FAIR) . . . 301
14.2 The Compressed Baryonic Matter (CBM) experiment . . . 303
14.3 The Silicon Tracking System (STS) . . . 304

14.4 Assembly of STS modules at KIT — infrastructure and workflow 306
14.4.1 Preparation and die and wire bonding of the LDOs to the FEBs 308
14.4.2 Silicon microstrip sensor 309
14.4.3 Silicon Tracking System-X-Y-Time-Energy (STS-XYTER) readout ASIC 311
14.4.4 Tape automated bonding of microcables to ASIC 313
14.4.5 Chip-cable to FEB and sensor assemblies 314
14.4.6 Quality assurance and storage 315
14.5 Progress and challenges 319

15 IPE KIT detector packaging for HEP summary 321

V Conclusion 325

Acknowledgements 333

Bibliography 335

1 Introduction

The field of semiconductor devices is vast and foundational to modern science and technology. Although they rely on the same fundamental building blocks, such as $p - n$ junctions, heterojunctions, and metal–insulator–semiconductor (MIS) interfaces, and are governed by the same laws of solid-state physics, semiconductor devices support an exceptionally diverse range of applications.

Semiconductor technology began with the invention of diodes and bipolar transistors, devices that were pivotal to the Third Industrial Revolution (1947–2000) [1]. These developments laid the groundwork for microchips, integrated circuits based on the metal–oxide–semiconductor field-effect transistor (MOSFET), microprocessors, and dynamic random-access memories (DRAMs). Since 2001, semiconductor memory, particularly floating-gate memory, has become the principal technological driver, enabling rapid progress in global connectivity, mobile computing, artificial intelligence, big data, cloud technologies, robotics, and solid-state storage [2]. Another important direction in semiconductor development stems from field-effect transistors adapted for power and high-frequency applications. These devices have enabled high-speed, high-impedance amplifiers, power electronics, and monolithic microwave integrated circuits (MMICs). A third and equally critical branch includes photonic devices and sensors, containing technologies such as light-emitting diodes (LEDs), lasers, solar cells, and radiation detectors. These systems uniquely exploit the ability of semiconductors to detect, generate, or convert signals between electrical and non-electrical forms [2].

The concept of a **sensor** extends beyond semiconductor devices and covers a broad range of technologies. In general terms, a sensor is defined as a device that detects or measures physical signals originating from the external environment. A sensor may be designed to respond to a variety of stimuli, including chemical, biological, thermal, mechanical, optical, magnetic, or radioactive signals. While these input types are diverse, the underlying goal is consistent: to extract information from a physical environment and translate it into a readable or processable signal. Many conventional semiconductor devices can, in fact, be interpreted as sensors. For example, a bipolar transistor senses variations in input current, while an LED may be viewed as a sensor for electrical-to-optical conversion. Solar cells convert photons into electrical current and can be considered sensors for light intensity. This conceptual framing demonstrates the versatility of semiconductor materials and the wide range of physical effects that can be exploited in device design.

Among the many available semiconductor materials, **silicon** has become the primary material in both industrial electronics and scientific applications. Its favourable properties, well-established processing techniques, and wide availability have made it central to technologies such as computing, telecommunications, photovoltaics, and, of course, sensors [3].

Within experimental physics, silicon sensors have become the standard for particle tracking, particularly in high-energy physics experiments. Their technological maturity, reliable radiation hardness, and scalability make them an ideal choice for this application. With ongoing technological progress, a diverse range of silicon-based detectors has been developed, where each is optimised for specific goals such as spatial resolution, time precision, integration level, or resistance to radiation damage. These advances have enabled ever more compact, precise, and robust detection systems across both scientific and industrial domains.

This work focuses on the **accelerator-driven development of silicon sensors**, with two major application domains explored: particle physics experiments and medical physics. The concept of particle acceleration dates back over a century. Over the past 120 years, the maximum energy of accelerated particles has increased by more than nine orders of magnitude—from a few keV to several TeV. Today, the culmination of this technological evolution is the Large Hadron Collider (LHC) at CERN, a 26.7 km long circular accelerator capable of reaching energies in the multi-TeV range. Accelerators are now a key enabling technology across many domains, including scientific, industrial, and medical applications [4].

Silicon sensors for particle physics

The evolving needs of the high-energy physics (HEP) community are guided by strategic frameworks such as the European Strategy for Particle Physics, which is regularly reviewed by the CERN Council. The latest strategy, adopted in June 2020, outlines long-term experimental goals and emphasises the need for a coordinated roadmap for detector R&D in Europe. It calls for balancing efforts across established and emerging technologies, and fostering synergies with adjacent fields [5].

Solid-state detectors (SSDs), particularly those based on silicon, represent a cornerstone of detector development for HEP. Technologies such as planar pixels, planar strips, and 3D pixels are implemented in nearly all modern collider experiments. Their compatibility with photolithographic techniques enables precise segmentation and excellent spatial resolution, making them indispensable for vertex reconstruction and charged particle tracking.

To meet the stringent demands of future facilities, including the future hadron colliders, such as Future Circular Collider FCC-hh and proposed e^+e^- Higgs factories, significant improvements in the performance of solid-state detectors are required. Key priorities include reduced material budgets, lower power consumption, high spatial and timing resolution, and extreme radiation tolerance. For instance, the innermost detectors at FCC-hh are expected to withstand fluences up to $1{\times}10^{18}$ n_{eq}/cm^2, with even higher doses in forward calorimeters, reaching up to 5000 MGy and 5×10^{18} n_{eq}/cm^2 over the lifetime of the experiment [5].

In response to these challenges, silicon detectors manufactured using commercial CMOS imaging processes, known as **MAPS**, short for **Monolithic Active Pixel Sensors**, are becoming increasingly prominent. These technologies, once developed primarily for consumer imaging applications, are now deployed in experiments such as STAR, Mu3e and ALICE, and are planned for CBM and LHCb. MAPS offer the advantages of low mass, fine segmentation, and high granularity, particularly well-suited for electron-based colliders.

Current R&D efforts are focused on achieving full monolithic integration of sensing and readout electronics within CMOS-based pixel arrays. The Technology Roadmap emphasises the need for ultra-thin, large-area sensors with high resolution and cost-effective fabrication. Passive CMOS sensors, where a commercial pixel array is bonded to an external readout ASIC, also offer promising paths for calorimetry, where extreme precision is not required but cost and scalability are critical. Additionally, commercial CMOS imaging sensor (CIS) technologies are under active investigation for tracking and vertexing applications, bridging the gap between industrial and HEP requirements [5].

Silicon detectors in medical physics

While medical physics is a broad field, this thesis focuses on accelerator-based applications, particularly hadron therapy, where the interest in implementing solid-state detectors has grown. Although not yet widely adopted in this domain, such detectors are actively being investigated as promising candidates for future medical beam monitoring systems. This focus is further motivated by the strong conceptual and technological overlap with high-energy physics.

Hadron therapy has become a well-established clinical technique, with more than one hundred treatment centres operating worldwide. A leading example of such clinical implementation is the Heidelberg Ion Beam Therapy Centre (HIT). Proton beams and accelerator-produced isotopes are routinely used for cancer therapy, offering highly localised and effective treatment with minimal side effects. The principal advantage of hadrons is the pronounced Bragg peak in their depth–dose profile, which allows precise energy deposition in tumour tissue while sparing surrounding healthy structures [6].

To fully exploit the advantages of hadron therapy, extremely precise control and monitoring of the therapeutic beam is essential, creating a clear opportunity for the adoption of solid-state sensors in medical accelerator systems. Before the beam reaches the patient, it passes through a monitoring system designed to measure the beam's position and dose rate in real time. This feedback is critical for the therapy control system, and the monitoring device must operate with minimal interaction with the beam to avoid scattering and energy loss. Traditionally, multi-wire chambers (MWCs) and gas-filled ionisation chambers have served as the primary beam position and dose monitors. However, these systems face limitations, especially when rapid response times are required. For example, drift time in MWCs introduces latency, which can cause dose uncertainties and make this technology unsuitable for flash irradiation. Furthermore, emerging techniques such as MRI-guided ion beam therapy introduce new challenges, such as magnetic field tolerance and resistance to mechanical vibrations. Solid-state detectors offer promising solutions to these limitations, with silicon-based devices providing improved spatial resolution, fast signal response, mechanical stability, and radiation hardness.

From concept to validation: modelling and testing

This thesis focuses on the development of silicon-based particle detectors. Behind every successful detector concept lies a sequence of validation stages, ranging from theoretical formulation and simulation to experimental characterisation and prototype testing. These stages not only require technical expertise in electronics and design but also a deep understanding of the underlying physical processes. A well-structured development cycle, involving both numerical modelling and empirical testing, is crucial to ensure the detector's functionality, reliability, and suitability for its intended application.

Physical modelling plays an indispensable role in modern sensor development. It allows for a detailed description of a system's behaviour using a mathematical model, which is useful for prototyping, debugging fabricated devices, and improving functionality. This work employs **Technology Computer-Aided Design (TCAD)**, a specialised branch of electronic design automation (EDA) used to model semiconductor processes and device operation. TCAD enables detailed simulations of fabrication steps (e.g., doping, thermal treatments) and device characteristics (e.g., electric fields, carrier densities, charge collection), offering insight into internal mechanisms that are not always easily accessible through direct measurement. For this thesis, the commercial TCAD platform developed by Synopsys was used. The toolchain comprises multiple modules with dedicated solvers and scripting interfaces. Simulation results were validated against experimental data and played a key role in guiding design decisions for monolithic detector prototypes, particularly in evaluating substrate configurations, electric field shaping, depletion depth, and radiation response.

In sensor development, predictive simulations help identify early on which design and production steps should be optimised. Properties such as substrate resistivity, layer thickness, and interface traps strongly affect electronic performance and must be considered in advance. TCAD modelling enables the evaluation of these dependencies before fabrication, thereby accelerating development and aligning design expectations with experimental results.

Device modelling alone cannot guarantee real-world detector performance. **Experimental validation** remains an essential step to confirm, refine, or challenge the assumptions made during modelling. Prototype testing enables direct observation of how a device behaves under operational conditions and is vital for identifying limitations, bugs, or unexpected behaviours. These measurements serve a dual role. First, they provide the ground truth required to validate simulation models. Second, they offer feedback for successive design iterations. In particular, discrepancies between simulated and measured behaviour can highlight missing physical effects or model inaccuracies, leading to improved fidelity in future simulations. In practice, sensor prototypes are tested under various conditions to replicate their intended use cases. In this thesis, I employed several testing methods, including electrical characterisation, measurements with radioactive sources such as ^{55}Fe and ^{90}Sr to evaluate signal response and noise, as well as particle beam tests at accelerator facilities to study timing resolution, radiation tolerance, and detection efficiency.

Through the integration of modelling and testing, the development process becomes a continuous **optimisation loop**. This approach is central to the sensor concepts explored in this thesis,

enabling informed design choices and paving the way towards robust, application-specific detector technologies.

Scope of this work and technologies investigated

During the course of my doctoral research, I was actively involved in the full development chain of silicon-based particle detectors: from conceptual modelling to functional module integration. This comprehensive process included:

- simulation and optimisation of sensor structures using physical and numerical models;
- laboratory characterisation of prototypes using radioactive sources, electrical tests, X-ray tubes, and particle beams;
- sensor integration through high-precision bonding techniques to establish electrical and mechanical interfaces with readout electronics;
- full detector module assembly, including packaging and system-level evaluation.

The core focus of this work lies in two leading **monolithic sensor technologies**: High-Voltage Complementary Metal-Oxide-Semiconductor (HV-CMOS) and Bipolar Complementary Metal-Oxide-Semiconductor (BiCMOS), both of which are widely regarded as promising platforms for next-generation tracking detectors in high-energy physics. Monolithic sensors integrate both the sensing diode and readout electronics on a single silicon die, reducing interconnect complexity and material budget while enabling high integration density and cost-effective large-area coverage.

HV-CMOS technology was first proposed for high-energy physics by Prof. Dr Ivan Perić in 2007 [7], and has since been implemented in prototypes for experiments such as ATLAS and Mu3e [8]. HV-CMOS MAPS provide fine spatial resolution and high granularity, and their timing performance has steadily improved. However, they continue to face limitations in radiation tolerance and noise performance when compared to hybrid sensors. Despite their higher cost and increased material budget, hybrid designs remain the standard for applications that demand the highest timing precision and radiation hardness [8].

The second major platform investigated in this thesis is **SiGe BiCMOS**, which combines the high-speed analogue performance of bipolar transistors with the logic integration capabilities of CMOS. My work focused on the design, implementation, and evaluation of a new sensor concept called BeMiPix (Bipolar experimental Miniature Pixel, developed using the SG13G2 130 nm SiGe BiCMOS process from IHP. SiGe BiCMOS offers key advantages for pixel sensors, such as excellent time resolution and promising radiation hardness.

The content of this thesis is organised into five parts.

Part I introduces the foundations of silicon sensor development.

Chapter 2 presents the fundamental principles of semiconductor physics, with a focus on silicon as the primary material used in particle detectors.

Chapter 3 explores the essential aspects of silicon sensor design and fabrication, including readout techniques, sensor segmentation, and the simulation of device behaviour.

Chapter 4 provides a detailed overview of the Synopsys Technology Computer-Aided Design (TCAD) tool suite, with particular emphasis on the role of numerical simulation in understanding and optimising device performance. Section 4.1 describes the principles behind device geometry modelling, while Section 4.2 focuses on numerical device simulation and the incorporation of physical models. Finally, Section 4.3 discusses beam simulations and the implementation of radiation damage models. Each TCAD tool presented in this chapter is discussed both theoretically and practically, with code examples and insights from my own work to guide future users and demonstrate the real-world applicability of simulation in sensor design.

Part II presents my work on the development and investigation of monolithic HV-CMOS sensor technology.

Chapter 5 provides a general introduction to the principles, advantages, and design features of HV-CMOS technology.

Chapter 6 is dedicated to the development of a monolithic active HV-CMOS hit-counting pixel sensor named HitPix, intended for medical applications in high-rate ion and proton beam therapy. Sections 6.1 to 6.5 describe the overall chip concept and application, while the main focus lies on the sensor development process using TCAD, and its functional characteristics as a semiconductor detector. Detailed simulations were performed to optimise the operational conditions of the sensor, presented in Sections 6.5 and 6.6, including substrate selection and analysis of the depletion region. Special attention was devoted to improving the breakdown limits, discussed in Section 6.7. A key advantage of TCAD simulations is their ability to model charge transport and collection under realistic beam conditions, as demonstrated in the evaluation of sensor detection performance in Section 6.8.

Chapter 7 introduces another application of HV-CMOS technology: the development of the monolithic CMOS active pixel sensor, named AstroPix, for use in future gamma-ray space telescopes. This chapter presents sensor optimisation strategies aimed at future production iterations, including breakdown improvement and enhanced design stability, covered in Section 7.6.

Chapter 8 investigates a performance limitation encountered in HV-CMOS sensors under exposure to high-intensity heavy-ion beams: the loss of detection efficiency. This effect was first observed during the development of HitPix sensors and led to the hypothesis that a form of Single Event Effect (SEE), such as Single Event Latch-up (SEL), could be responsible. The theoretical basis and supporting TCAD simulations are presented in Sections 8.1 and 8.2. To

investigate the origin and mechanism of this behaviour, a comparative study was carried out using five different HV-CMOS sensor designs. The results of this study are summarised in Section 8.3, with individual performance evaluations provided for the AstroPix, ATLASpix, MuPix, and HitPix chip families.

Chapter 9 summarises HV-CMOS sensor technology development and outlines design strategies for future developments.

Part III covers another important modern monolithic sensor technology that is receiving growing attention in particle physics for high-speed applications: BiCMOS technology. As part of this work, I developed and characterised a prototype chip named BeMiPix (Bipolar Experimental Miniature Pixel).

Chapter 10 introduces the motivation behind this development and the fundamental technological principles of BiCMOS processes.

Chapter 11 presents the design and simulation of the BeMiPix sensor. Sections 11.1 and 11.2 describe the concept, architecture, and layout of the chip, including TCAD simulations used to optimise its performance. The chip underwent extensive characterisation, including electrical breakdown and leakage current measurements (Section 11.4), temperature-dependent behaviour studies (Section 11.5), and response measurements to radioactive sources such as ^{55}Fe and ^{90}Sr (Section 11.7). In addition, laser scanning tests were performed to assess timing and spatial resolution (Section 11.8).

Chapter 12 provides an extended summary of the BiCMOS technology and key insights gained during the prototype development of BeMiPix.

Part IV focuses on the practical aspects of semiconductor detector production for high-energy physics experiments, including cleanroom work, sensor interconnection, and detector module assembly.

Chapter 13 presents my work on high-density interconnection techniques in two different applications. The first project, described in Section 13.2, involves ultra-thin gold wire bonding for SiGe chips developed at the Université de Genève. The second project, detailed in Section 13.3, covers gold bump bonding for the Pixelated Silicon Drift Detector (PixDD).

Chapter 14 documents my contribution to the Compressed Baryonic Matter (CBM) experiment, part of the future Facility for Antiproton and Ion Research (FAIR). Sections 14.1–14.3 introduce the experiment's goals and the Silicon Tracking System (STS), the subsystem where my contribution was focused. Section 14.4 details the sensor module assembly work carried out at KIT IPE, including all stages of construction and the quality assurance procedures that I implemented.

Chapter 15 summarises the detector packaging activities performed at KIT IPE, including the complete set of projects described in this part. These achievements were made possible by access to advanced cleanroom technologies and facilities available at IPE KIT.

Part V concludes the thesis by summarising the main results and contributions across all stages of development and characterisation of silicon-based detectors for particle physics applications.

Part I

Development of silicon sensors

2 Fundamentals of silicon sensors

This chapter introduces the basic principles of semiconductor physics, with a focus on the material that remains dominant in both industry and scientific research: silicon. Despite the development of alternative semiconductors, such as GaAs or SiC, silicon remains the primary material used in particle physics detectors due to its well-studied electronic properties, mechanical stability, mature processing technology, and cost efficiency. This chapter reviews the key physical concepts necessary for understanding how semiconductor devices work, especially silicon-based sensors.

The first part of the chapter summarises general semiconductor properties that are used throughout this thesis as the basis for interpreting and modelling sensor behaviour. It covers the crystal structure of semiconductors, energy bands, intrinsic and extrinsic carrier concentrations, charge transport, and the basic theory of the $p-n$ junction. As the fundamental building block of most semiconductor devices, it provides a starting point for understanding signal formation, amplification, and charge separation in sensors. The second part focuses specifically on silicon. It begins with a short overview of why silicon is the preferred material for most particle detectors, followed by a discussion of the operation principles of silicon sensors. The chapter concludes with a section on radiation damage in silicon detectors, including both bulk and surface effects.

2.1 Semiconductor physics

Solid-state physics studies how the macroscopic properties of solid materials arise from their microscopic structure and interactions. It combines concepts from crystallography, quantum mechanics, electromagnetism, and thermodynamics. As the largest field within condensed matter physics, it provides the theoretical foundation for materials science and technologies such as transistors, integrated circuits, and photonic devices.

The behaviour of semiconductor devices is directly linked to the physical properties of the materials from which they are built. Semiconductor physics is a broad research area, with topics ranging from simple on/off switches to advanced transistors, power electronics, and optoelectronic systems. This section provides an overview of the key physical concepts relevant to semiconductor behaviour, independent of specific device structures or applications. The focus is on fundamental topics such as crystal structure, energy bands, carrier statistics, charge transport, and the $p-n$ junction. These provide a common base for understanding semiconductor sensors and their operation in different environments. While only the essential ideas are summarised here, links to more advanced research and device-level effects are

pointed out where relevant. The primary reference used throughout this section is "Physics of Semiconductor Devices" by Sze, Li, and Ng [2].

2.1.1 Semiconductor materials and crystal structures

Semiconductors are a class of materials whose electrical properties lie between those of conductors and insulators. They can be formed from either elemental materials (such as silicon and germanium) or compound materials (such as gallium arsenide). In semiconductors, atoms are positioned in a regular, repeating structure, where each atom forms stable bonds with its neighbours. The atomic structure ensures that each atom is effectively surrounded by eight electrons. This configuration results from covalent bonding, where each bond involves two atoms sharing a pair of electrons. In typical semiconductor lattices, each atom forms four covalent bonds with four nearest neighbours, resulting in a total of eight shared electrons per atom. These bonds give the crystal its mechanical, optical and electronic properties.

A crystalline solid is defined by its periodic arrangement of atoms in space. This periodicity results in each atom experiencing an identical local environment, which can be described using the concept of a *Bravais lattice*. There are 14 distinct Bravais lattices, classified into seven crystal systems according to the geometrical relationships between unit cell edges and angles. The *unit cell* is the smallest building block of the crystal that can reproduce the entire structure by periodic translation. Its dimensions are described by the lattice parameters a, b, and c. In many cases, the most basic repeating unit is the *primitive cell*, which contains the minimum number of atoms required to generate the full lattice [9, 10].

Most technologically relevant semiconductors adopt the diamond or zincblende crystal structures. Both structures are based on the face-centred cubic (fcc) Bravais lattice and belong to the tetrahedral coordination system, in which each atom is surrounded by four nearest neighbours located at the vertices of a tetrahedron. These configurations are shown in Figure 2.1. In the diamond lattice (e.g., Si, Ge), all atoms are of the same species, whereas the zincblende structure (e.g., GaAs, GaP) features alternating atoms of two different elements, such as gallium and arsenic in GaAs, from groups III and V of the periodic table, respectively. This tetrahedral coordination leads to high mechanical strength and directional bonding properties, which are essential for electronic device applications.

The properties of crystalline materials are highly dependent on the orientation of the crystal planes and directions, especially for surface-sensitive technologies such as semiconductor fabrication. The orientation of a crystal plane is conventionally expressed using *Miller indices* (hkl), which are determined by taking the reciprocal of the plane's intercepts with the unit cell axes and reducing the values to the smallest integers. A given plane intersects the crystal lattice at regular intervals, which significantly affects material behaviour, for example, charge carrier mobility, and defect propagation [2].

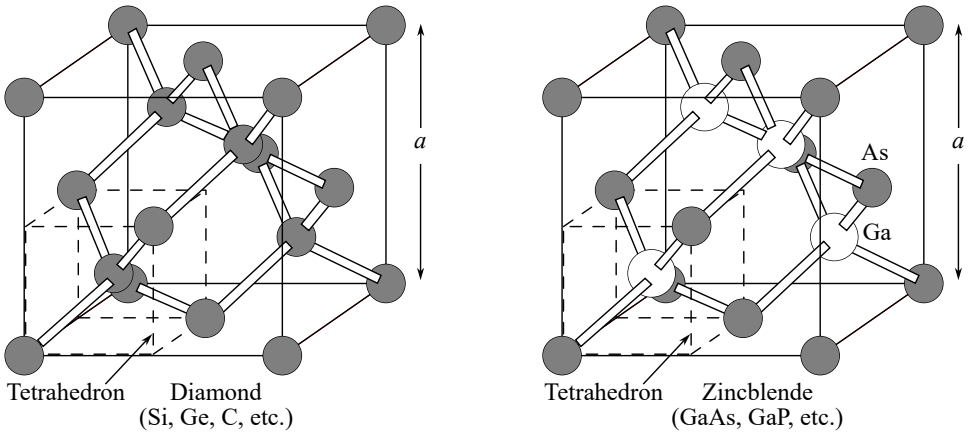


Figure 2.1: Primitive cells of the diamond (left) and zincblende (right) lattice structures. Both are based on a face-centred cubic lattice. In the diamond structure, all atoms are identical (e.g., Si, Ge), while in the zincblende structure, alternating atoms, such as Ga and As, occupy the two interpenetrating sublattices, characteristic of III–V semiconductors. The lattice constant a defines the unit cell edge length [2].

The direct lattice, also called the real lattice, describes the periodic arrangement of atoms in physical space. It can be described using three primitive basis vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$. The direct lattice sites can be defined by:

$$\vec{R} = m\vec{a} + n\vec{b} + p\vec{c} \tag{2.1}$$

where m, n, and p are integers.

However, to describe electron behaviour, particularly wave phenomena such as diffraction, band structure, and phonon interactions, one must shift from real space to mathematical space. To describe wave-related properties, the *reciprocal lattice* is used. This lattice exists in mathematical space, often called reciprocal space or k-space, where each point represents a wavevector. The reciprocal lattice is mathematically defined as the dual of the direct lattice. It reflects the periodicity of the crystal in terms of wave behaviour. The reciprocal basis vectors $\vec{a}^*$, $\vec{b}^*$, and $\vec{c}^*$ corresponding to the real-space basis vectors are defined as [2]:

$$\vec{a}^* \equiv 2\pi \frac{\vec{b} \times \vec{c}}{\vec{a} \cdot (\vec{b} \times \vec{c})} \quad \vec{b}^* \equiv 2\pi \frac{\vec{c} \times \vec{a}}{\vec{a} \cdot (\vec{b} \times \vec{c})} \quad \vec{c}^* \equiv 2\pi \frac{\vec{a} \times \vec{b}}{\vec{a} \cdot (\vec{b} \times \vec{c})} \tag{2.2}$$

These definitions ensure that $\vec{a} \cdot \vec{a}^* = 2\pi$, etc. The volume V_c^* of the primitive cell of the reciprocal lattice is related to the volume V_c of the direct lattice by [2]:

$$V_c^* = \frac{(2\pi)^3}{V_c} \tag{2.3}$$

A general reciprocal lattice vector can be expressed as [2]:

$$\vec{G} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^* \tag{2.4}$$

where h, k, and l are integers. This relationship implies that every reciprocal lattice vector is normal to a family of real-space lattice planes, fulfilling [2]:

$$\vec{G} \cdot \vec{R} = 2\pi \times \text{Integer} \tag{2.5}$$

The primitive cell of a reciprocal lattice can be represented by a *Wigner–Seitz cell*. The Wigner–Seitz cell in reciprocal space is referred to as the first *Brillouin zone*. It is constructed by drawing perpendicular bisectors from a central reciprocal lattice point to its nearest equivalent reciprocal lattice sites. The Brillouin zone contains all unique $\vec{k}$-vectors not reducible by reciprocal lattice translations and plays a central role in solid-state physics. Figure 2.2 shows the first Brillouin zone of a face-centred cubic lattice.

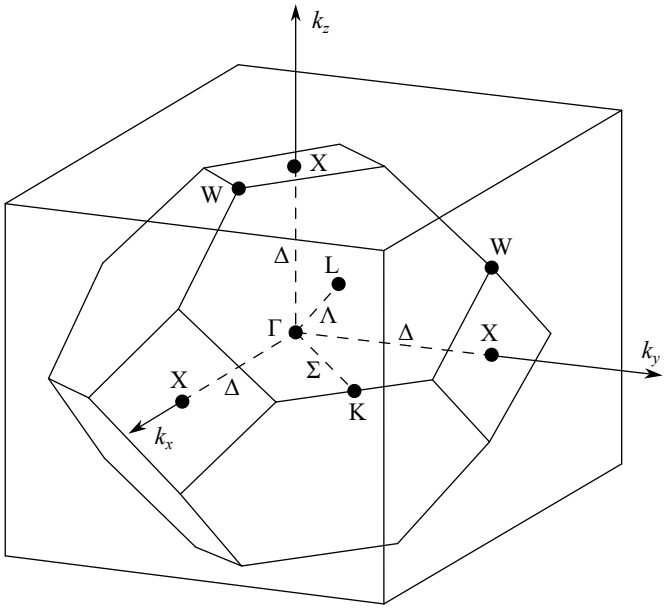


Figure 2.2: First Brillouin zone for a face-centred cubic (fcc) lattice. High-symmetry points such as Γ, X, L, K, and W represent critical points in the band structure of semiconductors. These zones and points define allowed and forbidden regions for electron wave propagation and are used extensively in band structure calculations [2].

The structure of the Brillouin zone directly determines the band structure of a material. The reciprocal lattice is useful to visualise the E–k relationship when the coordinates of the wave vectors $\vec{k}$ are mapped into the coordinates of the reciprocal lattice. As most semiconductors used in electronic and photonic applications have either diamond or zincblende structures, the corresponding fcc-based Brillouin zones are of particular importance for device physics. Understanding the connection between real and reciprocal lattice structures is essential for the analysis of carrier transport, scattering phenomena, and optical absorption in semiconductor devices.

2.1.2 Energy bands and carrier concentrations in equilibrium

The band structure of a semiconductor defines how electrons behave in a crystal and determines how they interact with external excitations. It also explains the existence of energy gaps, allowed states, and the distinction between conductors, semiconductors, and insulators. The band structure, that is the E–k relationship, is usually calculated by solving the one-electron Schrödinger equation under the approximation of a periodic potential $V(\vec{r})$ such as experienced by electrons in a crystalline semiconductor lattice [11]:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r})\right]\psi(\vec{r}) = E\psi(\vec{r}) \tag{2.6}$$

where:

- $\psi(\vec{r})$ is the wavefunction of the electron,
- E is the energy eigenvalue,
- m is the effective electron mass,
- $\hbar$ is the reduced Planck constant.

As it has been mentioned, the potential $V(\vec{r})$ has the periodicity of the crystal lattice [2]:

$$V(\vec{r}) = V(\vec{r} + \vec{R}) \tag{2.7}$$

for lattice vector $\vec{R}$.

Considering Bloch's theorem, the electronic states in a periodic crystal lattice can be expressed as wavefunctions of the form [11]:

$$\psi_{n,\vec{k}}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}} u_{n,\vec{k}}(\vec{r}) \tag{2.8}$$

where the function $u_{n,\vec{k}}(\vec{r})$ is lattice-periodic:

$$u_{n,\vec{k}}(\vec{r}) = u_{n,\vec{k}}(\vec{r} + \vec{R}) \tag{2.9}$$

The index n is the band number, and $\vec{k}$ is the wavevector in reciprocal space.

The Schrödinger equation for these Bloch states becomes:

$$\left[-\frac{\hbar^2}{2m}(\nabla + i\vec{k})^2 + V(\vec{r})\right] u_{n,\vec{k}}(\vec{r}) = E_n(\vec{k}) u_{n,\vec{k}}(\vec{r}) \tag{2.10}$$

where the energy $E_n(\vec{k})$ now depends explicitly on the wavevector $\vec{k}$.

Due to the periodicity in $\vec{k}$-space, the energy values are typically restricted to the first Brillouin zone, which is the Wigner–Seitz cell of the reciprocal lattice. This allows any wavevector to be mapped into a unique point within the Brillouin zone using the reduced zone scheme.

In semiconductors, the energy gap between the conduction band and the valence band is a fundamental parameter. This band gap E_g can be either direct or indirect, depending on whether the minimum of the conduction band and the maximum of the valence band occur at the same value of $\vec{k}$ in the Brillouin zone. In a direct band gap semiconductor, such as GaAs, transitions between conduction and valence bands can occur without a change in momentum, allowing, for example, efficient photon emission. In indirect gap semiconductors, such as Si or Ge, these transitions require a phonon to conserve momentum, making photon emission less efficient [2]. Figure 2.3 shows this difference. In GaAs (right), the conduction band minimum and valence band maximum both occur at Γ ($k = 0$), while in Si (left), they are at different $\vec{k}$ points.

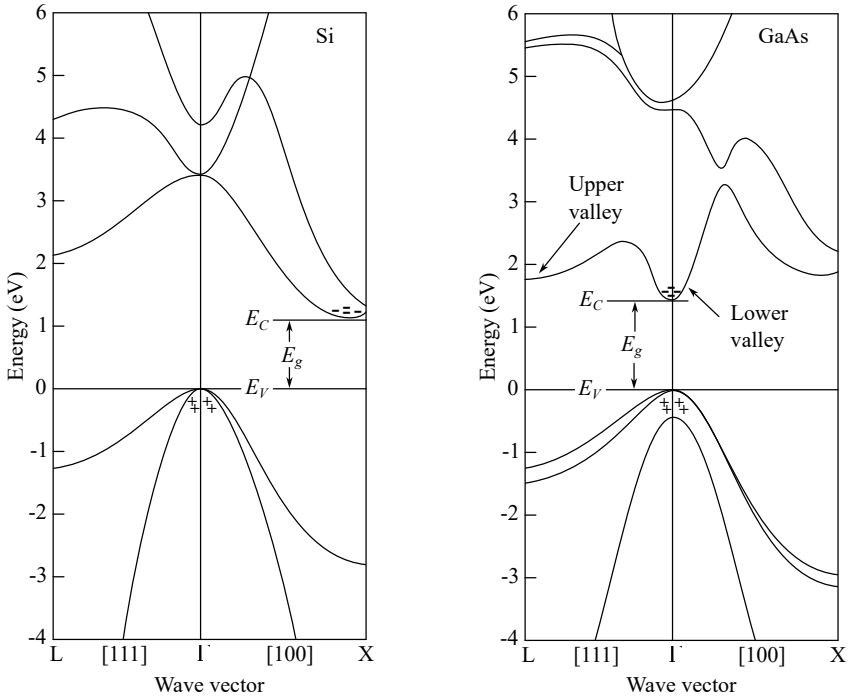


Figure 2.3: Energy-band structures of indirect (left) silicon and direct (right) GaAs. E_g is the energy band gap. Plus signs (+) indicate holes in the valence band; minus signs (–) indicate electrons in the conduction band [2].

In all semiconductors, a forbidden energy range exists where no allowed states are present. Below this gap lies the *valence band*, and above it, the *conduction band*. The top of the valence band is labelled E_V, and the bottom of the conduction band is E_C. The bandgap energy is then defined as:

$$E_g = E_C - E_V \tag{2.11}$$

For small values of $\vec{k}$ near the band edges, the E–k relationship can be approximated by quadratic equation [2]:

$$E_n(\vec{k}) = \frac{\hbar^2 k^2}{2m^*} \tag{2.12}$$

where m^* is the effective mass of the carrier. This effective mass reflects how the electron behaves under the influence of the crystal potential and is different from the free electron mass.

One of the most important properties of semiconductors is that their electrical conductivity can be modified by introducing controlled amounts of impurities. These impurities, known as dopants, introduce additional energy levels within the bandgap and significantly affect carrier concentrations. Unlike in metals or insulators, where the carrier concentration is fixed, semiconductors allow precise tuning through doping.

In a pure (intrinsic) semiconductor, each atom forms covalent bonds with its neighbours by sharing its valence electrons, resulting in a fully bonded crystal lattice. At absolute zero, no free charge carriers are present, but at higher temperatures, some bonds are thermally broken, leading to the generation of electron–hole pairs. When a donor impurity atom with one more valence electron than the host atoms is introduced into the lattice, it forms covalent bonds with surrounding atoms but contributes an additional, weakly bound electron. This electron can be thermally excited into the conduction band, leaving behind a positively charged donor ion. The resulting material is called n–type, where electrons are the majority carriers. Conversely, when an acceptor atom with one fewer valence electron replaces a host atom, it forms an incomplete bond. To complete the bonding structure, an electron is drawn from a neighbouring bond, which results in the creation of a mobile hole in the valence band. The acceptor atom becomes negatively charged, and the material is classified as p–type, with holes as majority carriers.

At thermal equilibrium, the probability that an energy level E is occupied by an electron is given by the *Fermi–Dirac* distribution [2]:

$$f(E) = \frac{1}{1 + \exp\left[\frac{E - E_F}{k_B T}\right]} \tag{2.13}$$

where E_F is the Fermi level, k_B is Boltzmann's constant, and T is the absolute temperature.

The position of the Fermi level E_F is crucial, as it indicates the probability distribution of electrons over available energy states. Semiconductors are classified as nondegenerate or degenerate based on the position of the Fermi level. Semiconductors are called nondegenerate when the Fermi level lies more than several kT below E_C. In this case, carrier concentrations can be approximated using Boltzmann statistics. If the Fermi level is closer than several kT or lies within a band, the semiconductor is degenerate, and full Fermi–Dirac statistics must be used.

In intrinsic semiconductors, the Fermi level lies near the middle of the bandgap. The intrinsic carrier concentration n_i describes the equilibrium number of electrons in the conduction band and holes in the valence band and is given by:

$$np = n_i^2 = N_C N_V \exp\left(\frac{-E_g}{k_B T}\right) \tag{2.14}$$

known as the mass-action law, where n and p represent electron and hole concentrations, respectively, n_i is the intrinsic carrier concentration, E_g is the energy bandgap, and N_C, N_V are the effective densities of states in the conduction and valence bands, respectively.

Adding doping introduces discrete energy levels within the bandgap. A donor level E_D lies slightly below E_C and is neutral when occupied by an electron; it becomes positively charged if empty. An acceptor level E_A lies above E_V, neutral when unoccupied, and negatively charged when filled by an electron.

Figure 2.4 shows simplified schematic diagrams of intrinsic, $n-$type, and $p-$type semiconductors. From left to right: the band diagram with occupied levels, the density of states $N(E)$, the Fermi–Dirac function $f(E)$, and the resulting carrier concentrations n and p.

For nondegenerate semiconductors under the assumption of full ionisation, the carrier concentrations can be written using exponential approximations involving the Fermi level. For an $n-$type semiconductor (electrons are the majority carriers) [2]:

$$n_{n0} \approx N_D = N_C \exp\left(-\frac{E_C - E_F}{k_B T}\right) \tag{2.15}$$

$$p_{n0} = \frac{n_i^2}{n_{n0}} \approx \frac{n_i^2}{N_D} \tag{2.16}$$

and for a $p-$type semiconductor (holes are the majority carriers) [2]:

$$p_{p0} \approx N_A = N_V \exp\left(-\frac{E_F - E_V}{k_B T}\right) \tag{2.17}$$

$$n_{p0} = \frac{n_i^2}{p_{p0}} \approx \frac{n_i^2}{N_A} \tag{2.18}$$

Here, the subscript "0" indicates equilibrium values. These expressions clearly show how carrier concentrations are exponentially dependent on the Fermi level position within the bandgap. This behaviour is fundamental for understanding charge transport, junction formation, and signal generation in semiconductor sensors.

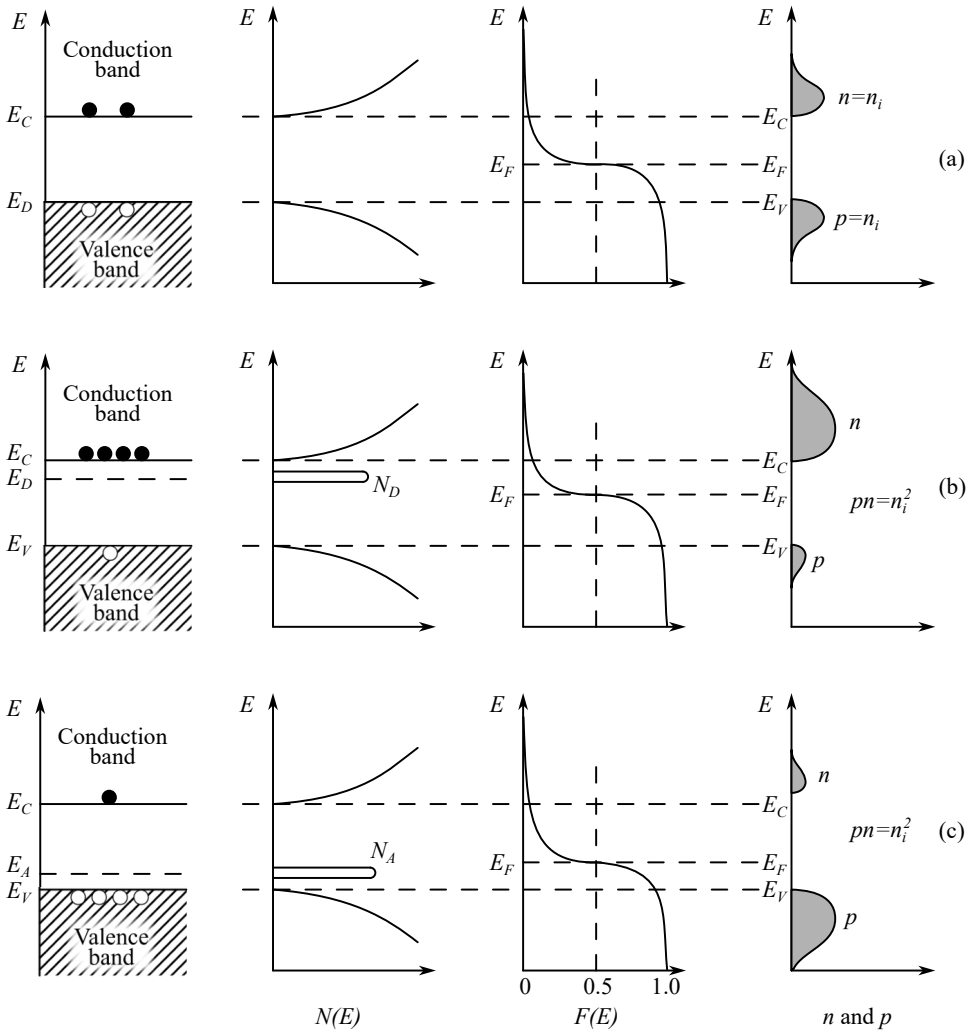


Figure 2.4: Schematic band diagrams, density of states $N(E)$, Fermi–Dirac distribution $f(E)$, and carrier concentrations for (a) intrinsic, (b) $n-$type, and (c) $p-$type semiconductors at thermal equilibrium. The shaded regions in E_C and E_V indicate the carrier populations. For all three cases, the mass-action law $pn = n_i^2$ holds [2].

2.1.3 Carrier transport

The transport of charge carriers is fundamental to the operation of all semiconductor devices. Their response to external electric fields, concentration gradients, and material inhomogeneities defines key performance characteristics such as current flow, switching behaviour, signal speed, and charge collection efficiency.

Carrier transport is typically described by mobility μ, which quantifies how quickly a charge carrier drifts under the influence of an electric field. It is also known as *conductivity mobility* or *drift mobility*. Mobility is strongly influenced by scattering mechanisms that impede the motion of carriers. The dominant scattering types are:

- Lattice (phonon) scattering: caused by thermal vibrations of the crystal lattice, which dominates at high temperatures.
- Ionised impurity scattering: significant at low temperatures or in highly doped materials, where carriers interact with charged dopant atoms.

The temperature dependence of mobility can be approximated by a power law:

$$\mu \propto T^{-\gamma} \tag{2.19}$$

where the exponent γ depends on the material and doping level, and typically ranges from 1.5 to 2.5.

For low impurity concentrations, the mobility is limited by phonon scattering with $\gamma = 1.5$, and it decreases with increasing temperature. However, due to the interplay of different scattering mechanisms, the actual temperature dependence deviates from a simple $T^{-1.5}$ behaviour. For example, in high-purity silicon near room temperature, measured mobilities vary approximately as $T^{-2.42}$ for electrons and $T^{-2.20}$ for holes [2]. At room temperature, the low-field mobility values for common semiconductors show band gap dependence. Direct bandgap materials such as GaAs exhibit significantly higher mobility than indirect bandgap semiconductors like silicon. In particular, the electron mobility of GaAs is roughly six times higher than that of Si. This makes GaAs and related III–V compounds attractive for high-speed and high-frequency applications, such as in high-electron-mobility transistors (HEMTs).

In semiconductors, carrier transport occurs primarily through two mechanisms: *drift* and *diffusion*. Drift transport refers to the motion of charge carriers under the influence of an electric field.

At low electric fields, the carrier drift velocity $\vec{v}$ is proportional to the field strength $\vec{E}$:

$$\vec{v} = \mu\vec{E} \tag{2.20}$$

where μ is the mobility of the carrier in $\mathrm{cm}^2/\mathrm{V{\cdot}s}$.

At higher electric fields, impact ionisation becomes significant. The drift velocity no longer increases linearly with the field; instead, it gradually saturates to a finite speed and becomes largely independent of impurity concentration. Additionally, nonlinearities in mobility are observed. For silicon, the saturation velocity for both electrons and holes is approximately 1×10^7 cm/s. In contrast, GaAs exhibits a more complex behaviour: electron velocity increases up to a certain field, reaches a peak, and then decreases. The saturation velocity for GaAs at high fields approaches 6×10^6 cm/s. This field-dependent velocity behaviour is crucial for high-speed and high-frequency device design [2]. Temperature also influences drift velocity: as temperature rises, phonon scattering increases, which lowers the saturation velocity in both Si and GaAs.

In semiconductors with both types of charge carriers, the total drift current density under an applied field is:

$$\vec{J} = \sigma\vec{E} = q(\mu_n n + \mu_p p)\vec{E} \tag{2.21}$$

where q is the elementary charge, n and p are the electron and hole concentrations, and μ_n and μ_p are their respective mobilities.

The individual drift current densities are given by:

$$\vec{J}_n^{\text{drift}} = qn\mu_n\vec{E} \tag{2.22}$$

$$\vec{J}_p^{\text{drift}} = qp\mu_p\vec{E} \tag{2.23}$$

One additional important property related to mobility is the *resistivity* ρ, which quantifies a material's opposition to electrical conduction and plays a central role in determining its electrical behaviour. It is defined as the inverse of conductivity σ:

$$\rho = \frac{1}{\sigma} \tag{2.24}$$

with total conductivity:

$$\sigma = q(n\mu_n + p\mu_p) \tag{2.25}$$

Thus, resistivity can be expressed as:

$$\rho = \frac{1}{q(n\mu_n + p\mu_p)} \tag{2.26}$$

These expressions highlight the role of doping in resistivity control. Increasing the dopant concentration increases the carrier density, thereby reducing resistivity. However, at very high doping levels, mobility decreases due to enhanced impurity scattering, partially compensating for the increase in conductivity. Resistivity is also temperature-dependent. At low temperatures, carrier freeze-out increases resistivity. At high temperatures, phonon scattering dominates, which reduces mobility and increases resistivity. Between these extremes lies a minimum-resistivity point, often used in temperature-sensitive device applications. Figure 2.5 shows measured resistivity versus impurity concentration for Si and GaAs. The dependence is not linear because mobility varies with doping level. These curves allow one to estimate the impurity concentration from resistivity measurements, though the actual carrier concentration may differ due to incomplete ionisation.

In contrast to drift, which is driven by electric fields, *diffusion* occurs due to spatial gradients in carrier concentration. During the discussion about drift, carrier populations were assumed to be spatially uniform. However, in many realistic situations, excess carriers are locally introduced, such as through optical illumination, injection from a $p-n$ junction, or particle interactions. These create nonuniform carrier distributions. Whenever such a concentration

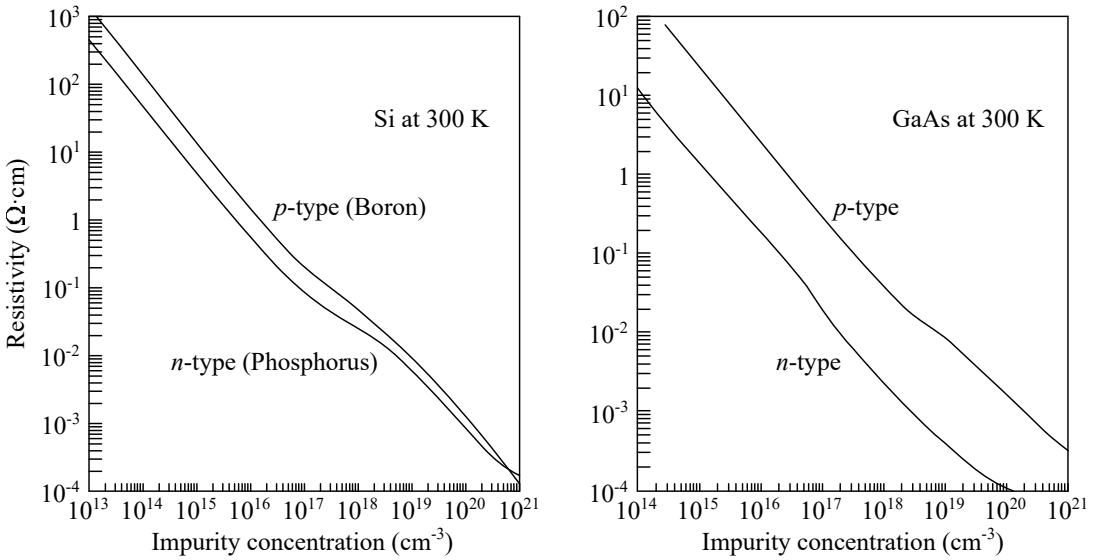


Figure 2.5: Resistivity as a function of impurity concentration for (left) Si and (right) GaAs [2]. The non-linear behaviour is due to mobility degradation at high doping levels.

gradient exists, carriers tend to migrate from regions of high concentration to regions of low concentration, attempting to restore equilibrium. This process is described by Fick's law and gives rise to a diffusion current.

The diffusion current densities for electrons and holes are given by:

$$J_n^{\text{diff}} = qD_n \nabla n \tag{2.27}$$

$$J_p^{\text{diff}} = -qD_p \nabla p \tag{2.28}$$

where D_n and D_p are the diffusion coefficients for electrons and holes, respectively. The signs reflect the direction of net carrier movement: electrons flow down the concentration gradient, while holes flow in the opposite direction [11].

The diffusion coefficient and mobility are related through:

$$D_n = \frac{k_B T}{q} \mu_n \tag{2.29}$$

$$D_p = \frac{k_B T}{q} \mu_p \tag{2.30}$$

where k_B is Boltzmann's constant, T is the absolute temperature, and μ_n, μ_p are the carrier mobilities. This relation links random thermal motion to the carrier's response to a potential gradient.

Aside from the standard conduction mechanisms, such as drift and diffusion, several additional processes are worth mentioning, as they become significant under specific conditions, including high electric fields, sharp potential barriers, or quantum confinement.

Thermionic emission is a conduction mechanism in semiconductors, typically involving majority carriers that overcome a potential barrier. It becomes dominant when the barrier width is small compared to the carrier mean free path. This is common in metal–semiconductor junctions, such as Schottky diodes. The key parameter controlling thermionic emission is the barrier height, not its shape. Once carriers have sufficient thermal energy ($\sim k_B T$) to cross the barrier, they contribute to a current. For this mechanism to dominate, carrier scattering within the barrier must be negligible, and the region beyond the barrier should not limit the current [2].

Tunnelling is a quantum-mechanical effect in which carriers pass through a potential barrier even if their energy is lower than the barrier height. Classically, such carriers would be completely confined, but quantum mechanics allows for a non-zero probability of barrier penetration due to the wave nature of particles. The electron wavefunction does not abruptly vanish at the barrier; instead, it decays exponentially and can extend through it. The tunnelling probability depends on both the height and width of the barrier. It becomes significant in structures with thin insulating layers or high electric fields, such as tunnel diodes and MOS interfaces [2].

2.1.4 Generation and recombination

Whenever the thermal equilibrium of a semiconductor is disturbed, meaning that the product of carrier concentrations deviates from its equilibrium value ($pn \neq n_i^2$), generation and recombination processes act to restore balance. These processes are recombination when $pn > n_i^2$ and thermal generation when $pn < n_i^2$. They determine the carrier lifetime and influence all dynamic behaviours in devices. The main recombination mechanisms, illustrated in Figure 2.6, are summarised below [11]:

- *Band-to-band recombination* (direct recombination), where an electron in the conduction band directly recombines with a hole in the valence band.
- *Recombination through trap states* (indirect recombination), in which the transition is mediated by defect or impurity states located within the bandgap. This process is described by Shockley–Read–Hall statistics.
- *Surface recombination*, which takes place at surfaces or interfaces with high defect densities, where disrupted atomic bonds act as recombination centres.

In direct bandgap semiconductors, such as most III–V compounds, band-to-band recombination is highly probable. The energy released during an electron transition from E_C to E_V is conserved either by photon emission (radiative recombination) or by transferring the energy to another carrier via an Auger process [12]. These processes are the inverse of optical absorption and impact ionisation, respectively. In indirect bandgap semiconductors, like silicon and germanium, the dominating transition types are through trap levels located within the bandgap. These trap-assisted transitions, typically described by Shockley–Read–Hall (SRH) statistics, mainly function in two ways: electron capture and hole capture via recombination centres [13, 14, 15]. The overall recombination rate depends on the density, energy level, and capture cross-section of the traps.

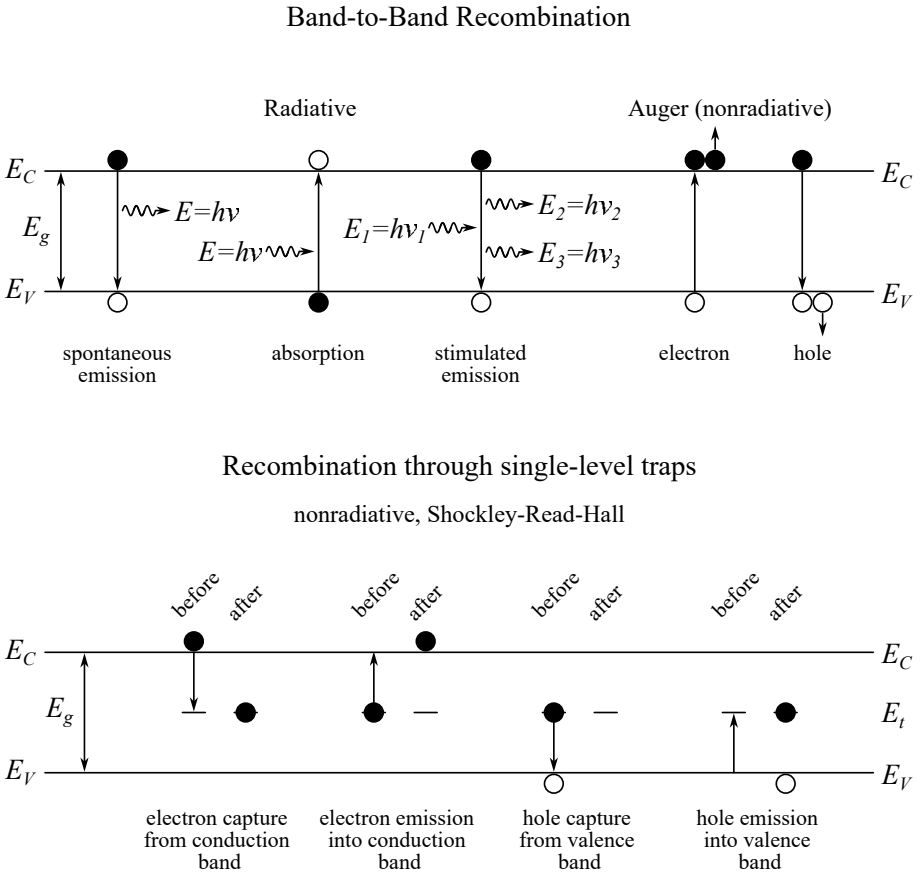


Figure 2.6: Main recombination mechanisms in semiconductors: (top) direct radiative recombination in direct bandgap materials, and (bottom) trap-assisted recombination (SRH process) in indirect bandgap semiconductors.

A critical parameter describing the impact of recombination is the *carrier lifetime*. The carrier lifetime is defined as the average time it takes for a minority carrier to recombine. The process through which this is done is typically known as minority carrier recombination. Carrier lifetimes depend strongly on doping concentration, temperature, and the presence of traps or defects. In detector applications, where signal strength is proportional to collected charge, the carrier lifetime directly affects charge collection efficiency, particularly for minority carriers in low-field regions.

2.1.5 Basic equations

The behaviour of semiconductor devices under external influences, such as applied voltage, temperature change, or optical excitation, is governed by a set of coupled equations that describe charge transport, carrier generation and recombination, and electrostatics. These fundamental relations are known as the drift-diffusion (DD) model and consist of three groups

of partial differential equations: the Poisson equation, which governs the potential, together with the current-density and continuity equations, which describe carrier transport and time-dependent variations in carrier concentrations [2]. When thermal gradients are significant, these equations can be coupled to the heat flow equation, forming a thermoelectric model used in advanced simulations.

The *current-density equations* describe the total current densities for electrons and holes, including contributions from both drift (movement in response to electric fields) and diffusion (movement due to concentration gradients). The general forms are:

$$\vec{J}_n = qn\mu_n\vec{E} + qD_n\nabla n \tag{2.31}$$

$$\vec{J}_p = qp\mu_p\vec{E} - qD_p\nabla p \tag{2.32}$$

where q is the elementary charge, n and p are the electron and hole concentrations, μ_n and μ_p are the respective mobilities, D_n and D_p are the diffusion coefficients, and $\vec{E}$ is the electric field. These expressions are valid under low-field conditions and are central to device models for transport phenomena.

Continuity equations describe transient behaviour, such as injection, recombination, or dynamic response. The continuity equations deal with time-dependent phenomena. They are derived from charge conservation and take the form:

$$\frac{\partial n}{\partial t} = G_n - R_n + \frac{1}{q}\nabla \cdot \vec{J}_n \tag{2.33}$$

$$\frac{\partial p}{\partial t} = G_p - R_p - \frac{1}{q}\nabla \cdot \vec{J}_p \tag{2.34}$$

where G_n, G_p and R_n, R_p are the generation and recombination rates for electrons and holes, respectively. The divergence terms represent net current flow into or out of a volume.

The electrostatic potential ψ in the semiconductor is described by the *Poisson equation*, which relates it to the spatial charge distribution:

$$\frac{d^2\psi}{dx^2} = -\frac{dE}{dx} = -\frac{\rho}{\varepsilon} = -\frac{q}{\varepsilon}(p - n + N_D^+ - N_A^-) \tag{2.35}$$

where ε is the permittivity of the material, and N_D^+ and N_A^- are the concentrations of ionised donors and acceptors, respectively.

Together, these equations, Poisson, current-density, and continuity, form the complete drift-diffusion model. This system is used in both analytical and numerical simulations to describe the static and dynamic behaviour of semiconductor devices under a variety of conditions.

2.1.6 The *p* – *n* junction

The basic theory describing the current-voltage (IV) behaviour of $p-n$ junctions was first established by Shockley in the 1950s [13, 16]. The general device equations introduced in Subsection 2.1.5 are applied here to derive both steady-state and dynamic characteristics of the junction. The $p-n$ junction is a two-terminal structure formed by bringing into contact a $p-$type and an $n-$type semiconductors. Depending on the geometry, doping gradient, and external biasing, the junction can fulfil a range of electronic functions. It is the fundamental building block of most semiconductor devices.

The behaviour of the $p-n$ junction at equilibrium and under bias conditions is visualised through energy band diagrams, which are shown in Figure 2.7. When these two parts are brought into contact, electrons from the $n-$type region diffuse into the $p-$type region, while holes diffuse in the opposite direction [17]. At equilibrium, the conduction and valence bands bend across the junction, aligning the Fermi level E_F throughout the structure. A depletion region forms where mobile electrons and holes have recombined, leaving behind fixed charges. This space charge creates an electric field and a potential barrier, known as the *built-in potential* ψ_{bi}. Using Boltzmann statistics for non-degenerate semiconductors, the built-in potential can also be written in terms of doping concentrations:

$$\psi_{\text{bi}} = \frac{kT}{q}\ln\left(\frac{n_{n0}}{n_i}\right) + \frac{kT}{q}\ln\left(\frac{p_{p0}}{n_i}\right) \approx \frac{kT}{q}\ln\left(\frac{N_D N_A}{n_i^2}\right) \tag{2.36}$$

where n_{n0} and p_{p0} are the equilibrium majority carrier concentrations on the $n-$ and $p-$sides, respectively, and n_i is the intrinsic carrier concentration.

Under *forward bias*, an external positive voltage is applied to the $p-$type side, and the external field opposes the built-in potential. This reduces the potential barrier, narrowing the depletion width and allowing a significant number of carriers to diffuse across the junction. Electrons from the $n-$side diffuse into the $p-$region and recombine with holes. Similarly, holes from the p-side defuse into the $n-$region. This results in an exponential increase in current. In contrast, under *reverse bias*, the external voltage strengthens the built-in potential. The depletion region widens, and the energy barrier increases. Carrier injection is suppressed, and only a small reverse saturation current flows due to minority carriers or thermally generated pairs. At high reverse voltages, breakdown mechanisms may occur.

As mentioned before, carrier diffusion leads to the mutual annihilation of free electrons and holes at the $p-n$ junction interface. This process reveals the fixed, ionised charge near the junction, resulting in the formation of a space-charge region, known as the *depletion region* and is given by [2]:

$$W = \sqrt{\frac{2\varepsilon_s}{q}\frac{N_A + N_D}{N_A N_D}(\psi_{\text{bi}} - V)} = x_n + x_p \tag{2.37}$$

where ε_s is the semiconductor permittivity, q is the elementary charge, x_n and x_p are the widths of the depletion region extending into the $n-$type and $p-$type sides of the junction, and N_A,

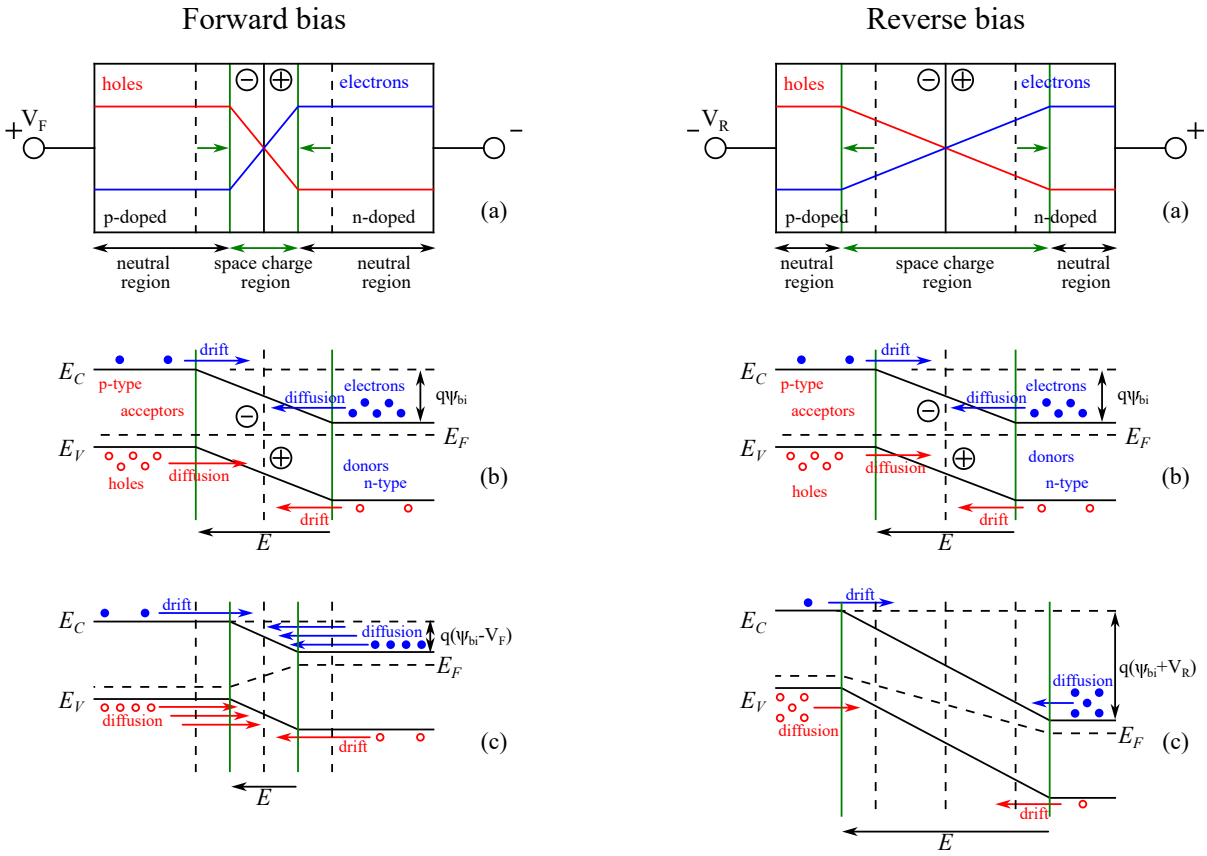


Figure 2.7: Energy band diagrams of a $p-n$ junction: (a) schematic of the structure, (b) energy bands at equilibrium, (c) under forward (left) and reverse (right) bias conditions.

N_D are the acceptor and donor doping concentrations, respectively. From this, the voltage-dependent depletion-layer capacitance per unit area C_D, often referred to as the *junction capacitance*, is defined as:

$$C_D = \frac{dQ_D}{dV} = \frac{\varepsilon_s}{W} = \sqrt{\frac{q\varepsilon_s}{2}\frac{N_A N_D}{N_A + N_D}}\left(\psi_{\text{bi}} - V - \frac{2kT}{q}\right)^{-1/2} \tag{2.38}$$

A correction factor $2kT/q$ comes from two majority-carriers distribution tails: electrons in $n-$side and holes in $p-$side.

For small-signal analysis, the inverse-square capacitance method is commonly employed. The relationship is given by:

$$\frac{1}{C_D^2} = \frac{2}{q\varepsilon_s N}\left(V_{\text{bi}} - V - \frac{2kT}{q}\right) \tag{2.39}$$

Differentiating this with respect to voltage gives:

$$\frac{d(1/C_D^2)}{dV} = -\frac{2}{q\varepsilon_s N} \tag{2.40}$$

Plotting $1/C_D^2$ against V thus results in a straight line, as illustrated in Figure 2.8a. The slope allows extraction of the doping concentration N, and the intercept on the voltage axis provides the built-in potential (offset by $2kT/q$).

The IV characteristics of the $p-n$ junction under forward and reverse bias are fundamental to understanding of semiconductor devices. In the ideal case, the $p-n$ junction behaves according to the Shockley equation, derived under four main assumptions [2]:

- The abrupt depletion-layer approximation, where the built-in and applied voltages are supported entirely by the depletion region with neutral quasi-neutral regions outside.
- The validity of the Boltzmann approximation, assuming non-degenerate carrier statistics.
- The low-injection condition, meaning that injected minority carrier densities remain small compared to majority carrier densities.
- No generation or recombination occurs within the depletion region, and the electron and hole currents remain constant across it.

Under these assumptions, the minority carrier injection governs the total current. Hole current is determined by injection from the $p-$side into the $n-$side, and vice versa for electrons. The total current density becomes [2]:

$$J = J_p + J_n = J_0 \left[\exp\left(\frac{qV}{kT}\right) - 1 \right] \tag{2.41}$$

with the reverse saturation current J_0 defined as:

$$J_0 = \frac{qD_p n_i^2}{L_p N_D} + \frac{qD_n n_i^2}{L_n N_A} \tag{2.42}$$

In practical diodes, deviations from ideal behaviour occur due to several mechanisms:

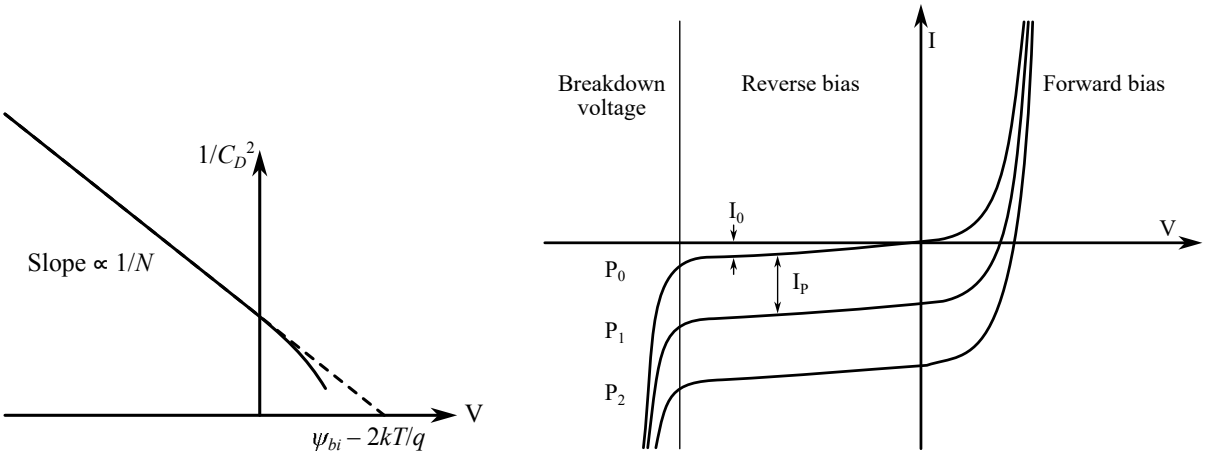
- generation–recombination (G–R) currents within the depletion region, facilitated by trap states;
- high-injection effects at larger forward voltages, violating the low-injection assumption;
- parasitic resistances;
- junction breakdown at large reverse bias.

To accommodate these deviations, the ideal diode equation is generalised by introducing an empirical parameter known as the *ideality factor* n, which reflects the dominant transport mechanism. The modified current expression becomes:

$$J = J_0 \left(\exp\left(\frac{qV}{nkT} \right) - 1 \right) \tag{2.43}$$

where n typically ranges from 1 (indicating pure diffusion) to 2 (indicating recombination-dominated current in the depletion region).

The IV characteristic is illustrated in Figure 2.8b. Under external excitation, such as illumination, the reverse current increases due to the generation of electron–hole pairs, resulting in a photocurrent I_P. The dark current I_0 corresponds to the diode's response in the absence of light. The labelled points P in Figure 2.8b indicate current levels at different light intensities, with P_0 marking the unilluminated state.



(a) Capacitance–voltage CV characteristic: A $1/C^2$ vs V plot enables extraction of V_{bi} and doping concentration N.

(b) Ideal current–voltage characteristic of a $p-n$ junction diode with and without illumination.

Figure 2.8: Electrical characteristics of the $p-n$ junction under bias.

When a reverse bias is applied across a $p-n$ junction and the electric field exceeds a critical threshold, the junction undergoes breakdown. This regime is characterised by a rapid and potentially destructive increase in reverse current. Three principal mechanisms are responsible for breakdown in semiconductors: thermal instability, quantum mechanical tunnelling, and avalanche multiplication. While the first two mechanisms are typically secondary in well-designed silicon devices, they are briefly summarised below before focusing on avalanche effects, which are dominant in most practical cases.

Thermal breakdown arises from the mutual influence of reverse current and temperature rise. As the reverse bias increases, so does the leakage current, resulting in the generation of heat. This temperature increase further raises the reverse saturation current -J_0, resulting in a feedback loop known as thermal runaway. If the generated heat cannot be efficiently dissipated, the junction temperature rises even more [2].

Tunnelling breakdown, often referred to as Zener breakdown, can be caused by quantum mechanical tunnelling in highly doped junctions, where the depletion region becomes narrow enough for carriers to tunnel through the barrier. The breakdown voltage due to tunnelling exhibits a negative temperature coefficient, as the bandgap narrows with increasing temperature, thereby lowering the energy barrier for tunnelling. This mechanism is particularly relevant in devices engineered for low breakdown voltages and becomes less significant as the breakdown voltage exceeds the tunnelling threshold [2].

In moderately to lightly doped junctions, the dominant breakdown mechanism is *avalanche multiplication*. Thermally generated carriers in the depletion region gain kinetic energy from the high electric field. If this energy becomes sufficient, these carriers can ionise atoms via impact, generating additional electron–hole pairs. This secondary generation leads to a chain reaction: each new carrier accelerates and produces further ionisation events. The process continues until the carrier population diverges, defining the breakdown condition [11].

Avalanche multiplication is governed by the impact ionisation coefficients α_n and α_p, which describe the rate of ionisation for electrons and holes, respectively. When the total multiplication factor approaches infinity, breakdown is said to occur. By combining the breakdown condition with the Poisson equation, critical quantities such as the breakdown voltage, the maximum electric field, and the depletion width at breakdown can be determined. An approximate universal expression for breakdown voltage, summarising the empirical behaviour across many materials, is given by [2]:

$$V_{\mathrm{BD}} \propto 60 \left(\frac{E_g}{1.1\ \mathrm{eV}} \right)^{3/2} \left(\frac{N}{10^{16}\ \mathrm{cm}^{-3}} \right)^{-3/4} \tag{2.44}$$

Here, E_g is the bandgap energy and N is the doping concentration. The breakdown field typically reaches values around 3×10^5 V/cm in silicon. Avalanche breakdown is crucial for several device applications. It limits reverse biasing in radiation detectors and bipolar transistors, but it is also intentionally exploited in devices such as avalanche photodiodes and modern Low Gain Avalanche Diodes (LGADs).

In planar structures, *edge effects* play a critical role in breakdown estimations. Localised high-field regions form at corners or curvature discontinuities, leading to premature breakdown. Techniques such as guard rings, field plates, or graded junctions are used to smooth electric field lines and elevate breakdown thresholds. These techniques are introduced in Section 3.6. The influence of curvature on breakdown voltage has been quantitatively modelled [18, 19] and is further discussed in the context of radiation sensor design in Subsection 7.6.2.

2.2 Silicon as a primary material for semiconductor particle sensors

Silicon is the most widely used material for semiconductor radiation detectors, especially in high-energy physics (HEP) applications. Its dominance comes from a well-balanced combi-

nation of electrical performance, mechanical robustness, and technological maturity. With a bandgap of 1.12 eV, silicon offers a low intrinsic carrier concentration at room temperature, which minimises leakage current and supports stable operation without the need for advanced cooling. It also exhibits a high breakdown electric field and excellent crystalline purity, which contribute to efficient charge collection and reproducible device behaviour. Compared to alternative semiconductors, silicon shows a balance between performance and practicality. Germanium, for example, offers higher charge carrier mobilities and superior energy resolution, but its narrow bandgap (0.66 eV) results in high thermal carrier generation and the necessity of cryogenic cooling to suppress leakage current. Gallium arsenide (GaAs) offers higher radiation hardness [20], but it suffers from higher material costs and limited availability of large, defect-free wafers.

A key advantage of silicon lies in the ability to control its electrical properties through doping. Introducing impurity atoms from group III or group V elements allows for precise tuning of the majority carrier type and concentration. The most common substrates for HEP sensors are boron-doped p–type wafers, with doping levels in the range of 10^{11} cm^{-3} to 10^{15} cm^{-3} for high-resistivity material. The doping profile directly determines the electrical field distribution and depletion behaviour when reverse bias is applied across a junction. These characteristics are crucial for defining the active sensing volume and ensuring full charge collection. Extensive characterisation of boron doping in silicon is available in the literature, including detailed discussions of diffusion mechanisms, activation, and compensation effects [21].

In addition to its intrinsic physical advantages, silicon benefits from decades of industrial development through the microelectronics and CMOS fabrication sectors. The resulting infrastructure enables the cost-effective production of high-quality wafers on a large scale with consistent material properties. Processing steps such as oxidation, ion implantation, and photolithography are well-established and highly optimised for silicon substrates. Chapter 3 "Design and fabrication of silicon sensors" is dedicated to this topic.

Silicon detectors are particularly well-suited to the detection and measurement of ionising radiation from charged particles and photons. Their use allows for precise spatial, temporal, and energy resolution when integrated into carefully engineered sensor architectures. This versatility has made silicon indispensable in the design of tracking detectors, calorimeters, timing layers, and pixel sensors across numerous HEP experiments [8].

2.2.1 Working principle of silicon sensors

The operation of silicon detectors in high-energy physics is fundamentally based on the interaction between charged particles and matter. As a charged particle passes through a medium, it loses energy primarily through ionisation, interacting with an atom's electrons via the Coulomb force. In semiconductors, these interactions result in the generation of electron–hole pairs along the particle's trajectory. If the particle's trajectory is inclined with respect to the sensor plane, this ionisation trail can extend across multiple pixels. This is an effect known as charge sharing. Such spatial spreading also occurs in orthogonal incidence, particularly near pixel boundaries [22].

The average energy loss of a particle per unit distance in a material is described by the Bethe–Bloch equation [23, 24]:

$$-\frac{dE}{dx} = \frac{4\pi n z^2}{m_e c^2 \beta^2} \cdot \left(\frac{e^2}{4\pi\varepsilon_0}\right)^2 \cdot \left[\ln\left(\frac{2m_e c^2 \beta^2}{I(1-\beta^2)}\right) - \beta^2\right] \tag{2.45}$$

where v is the particle's velocity, c is the speed of light, $\beta = v/c$ is the relativistic velocity factor, E is the particle's energy, x is the distance travelled, e is the elementary charge, z is the particle charge number, ε_0 is the vacuum permittivity, n is the electron density, m_e is the electron mass, and I is the mean excitation potential of the material [8]. This stopping power determines how much energy a particle loses in the detector material and, hence, how many electron–hole pairs are produced.

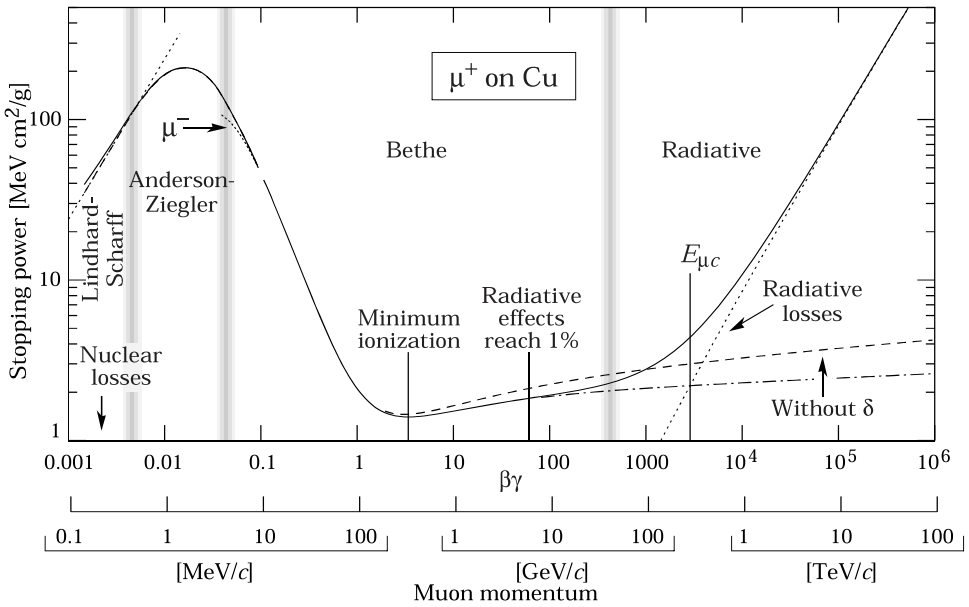


Figure 2.9: Stopping power of muons in copper as a function of momentum, illustrating the Bethe–Bloch behaviour. A similar shape is observed in silicon. The minimum corresponds to minimum ionising particles (MIPs), which are common in HEP detectors. At higher energies, radiative losses (e.g., Bremsstrahlung) dominate, while low-energy particles are typically absorbed early and are less relevant in tracking systems. Adapted from [25].

As shown in Figure 2.9, the stopping power has a minimum, which is defined as the minimum ionising particle (MIP). MIP deposits approximately constant energy per unit length. The shape of the stopping power curve is universal across materials, although the absolute values differ. Radiative losses (Bremsstrahlung) increase at high energies, but their contribution to the detector signal is limited, as most generated photons escape without further interaction.

The basic building block of a silicon detector is a reverse-biased $p-n$ junction. Figure 2.10 illustrates the basic detection principle using a monolithic sensor with top-side biasing. The applied bias voltage removes free carriers from the bulk, forming a depletion region that acts

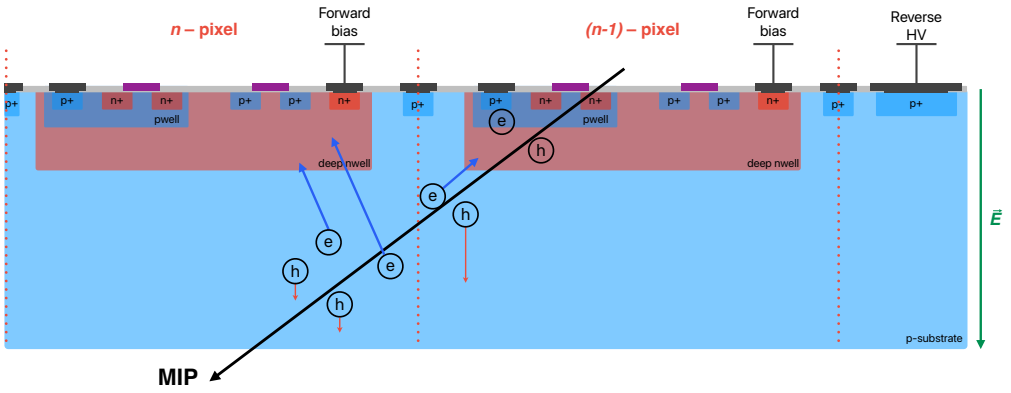


Figure 2.10: Schematic illustration of a charged particle traversing a monolithic silicon sensor. Electron–hole pairs created in the depletion region are separated by the internal electric field. Electrons are collected by the pixel electrode, while holes are collected at the backside substrate contact. Carriers generated outside the depletion zone are subject to diffusion and may recombine before collection.

as the sensitive volume. Electron–hole pairs generated within this region are separated by the electric field and collected efficiently. The motion of these charges induces a transient current, which is then amplified and digitised in the readout electronics [22]. In this example, for simplicity, the full p–type bulk is depleted. However, the charge generated in undepleted areas (e.g., heavily doped regions or outside the space charge zone) may recombine or become trapped before reaching the electrodes. In some cases, carriers originating near the interface may still contribute to the signal or induce secondary effects, such as parasitic currents or crosstalk.

The amount of collected charge is directly proportional to the deposited energy. In silicon, an average energy of 3.6 eV is required to generate one electron–hole pair, despite the bandgap being only 1.12 eV. This low ionisation energy results in large signal amplitudes for each particle hit. For comparison, gas detectors typically require several tens of eV to generate a single ion pair. A minimum ionising particle loses approximately 390 eV/µm in silicon, corresponding to around 80 electron–hole pairs per micrometre. This relation forms the basis for energy calibration, allowing particle identification and tracking based on the deposited charge and signal shape. The high density of silicon also reduces the range of secondary electrons (delta rays), allowing for highly localised energy deposition and, consequently, precise spatial resolution [26].

2.2.2 Radiation damage in silicon sensors

Silicon detectors exposed to prolonged radiation suffer performance degradation due to the formation of defects in both the material bulk and surface layers. These defects alter the sensor's electrical characteristics and limit its operational lifetime, particularly in high-fluence environments. Radiation damage is categorised into two principal types: bulk damage, predominantly

caused by non-ionising energy loss (NIEL) from hadrons and neutrons, and surface damage, mainly resulting from ionising radiation such as X-rays photons and charged particles.

Surface damage is caused by ionising radiation interacting with insulating layers, especially silicon dioxide SiO_2, which generates electron–hole pairs. While the mobility of electrons is sufficiently high so that they can move to a nearby electrode, holes become trapped, leading to a net positive charge at the Si/SiO_2 boundary [26]. The accumulated positive charge can induce an electron accumulation layer in the silicon beneath the interface, degrading isolation between electrodes and, in extreme cases, triggering breakdown events. Surface generation currents caused by interface traps are field-dependent and evolve over time due to the very high sheet resistance of the Si/SiO_2 surface. The result can be a breakdown after several days of operation or a humidity-dependent breakdown voltage [27]. Surface damage is strongly dependent on technology, dose rate, and annealing behaviour, which can occur even at room temperature. Although surface effects can often be mitigated through proper layout and passivation techniques, they remain a source of long-term instability if not properly accounted for during the design process.

Bulk radiation damage arises from the displacement of silicon atoms from their lattice positions by high-energy particles. A minimum energy transfer of approximately 25 eV is required to displace a single atom in silicon. For energy transfers above one keV, the silicon atom itself can initiate secondary displacements, leading to defect clusters; for energies above ~12 keV – multiple clusters may form. These defects are mobile at room temperature and may either anneal, migrate to surfaces, or interact with impurities to form stable, electrically active centres [28]. The main consequences of bulk damage include an increase in reverse (dark) current due to the formation of generation–recombination centres, trapping of signal carriers, which reduces the charge collection efficiency (CCE), and a modification of the internal electric field caused by changes in the effective doping concentration. For irradiation fluences above 10^{14} cm^{-2}, defect densities can exceed the original doping concentration, leading to significant changes in material properties [26].

To characterise the microscopic nature of defects, deep-level transient spectroscopy and related techniques have been used to determine defect levels, charge states, and capture cross-sections for electrons and holes [28]. Based on these findings, effective defect models are constructed to support numerical device simulation. Modern Technology Computer-Aided Design (TCAD) tools are capable of simulating the impact of radiation-induced defects on detector performance. As shown in Figure 2.11, different parameter sets can produce markedly different simulation outcomes, highlighting the need for careful calibration. While many aspects of post-irradiation behaviour can be captured by TCAD models, it is important to recognise that no current model is universally predictive, especially at high fluences or for complex detector geometries [28]. Further details on the direct implementation of radiation damage models in TCAD simulations can be found in Section 4.3.

Model	Type	Level [eV]	$\sigma_{e,h}$ [cm^2]	η [cm^{-1}]	Comment
EVL 2002	A	$E_C - 0.525$	1×10^{-15}	–	Tool: Microsoft Excel
	D	$E_V + 0.48$	1×10^{-15}	–	
	–	$E_C - 0.65^{(*)}$	1×10^{-13}	0.4	$^{(*)}$level for current generation, no space charge
Perugia 2006	A	$E_C - 0.42$	$2 \times 10^{-15}, 2 \times 10^{-14}$	1.613	Tool: Silvaco
(p-type sensors)	A	$E_C - 0.46$	$5 \times 10^{-15}, 5 \times 10^{-14}$	0.9	
	D	$E_V + 0.36$	$2.5 \times 10^{-14}, 2.5 \times 10^{-15}$	0.9	
	A	$E_C - 0.42$	$2 \times 10^{-15}, 1.2 \times 10^{-14}$	13	
(n-type sensors)	A	$E_C - 0.50$	$5 \times 10^{-15}, 3.5 \times 10^{-14}$	0.08	
	D	$E_V + 0.36$	$2 \times 10^{-18}, 2.5 \times 10^{-15}$	1.1	
Glasgow 2008	A	$E_C - 0.42$	$9.5 \times 10^{-15}, 9.5 \times 10^{-14}$		Tool: Synopsys
	A	$E_C - 0.46$	$5 \times 10^{-15}, 5 \times 10^{-14}$	0.9	model adapted from *Perugia 2006*
	D	$E_V + 0.36$	$3.23 \times 10^{-13}, 3.23 \times 10^{-14}$	0.9	simulation of p-type 3D sensors
KIT 2013					Tool: Synopsys
(protons)	A	$E_C - 0.525$	$1 \times 10^{-14}, 1 \times 10^{-14}$	–	$\eta_A = 1.189\ \text{cm}^{-1} \times \phi - 6.454 \times 10^{13}\ \text{cm}^{-3}$
	D	$E_V + 0.48$	$1 \times 10^{-14}, 1 \times 10^{-14}$	–	$\eta_D = 5.598\ \text{cm}^{-1} \times \phi - 3.949 \times 10^{14}\ \text{cm}^{-3}$
(neutrons)	A	$E_C - 0.525$	$1.2 \times 10^{-14}, 1.2 \times 10^{-14}$	1.55	
	D	$E_V + 0.48$	$1.2 \times 10^{-14}, 1.2 \times 10^{-14}$	1.395	
Delhi 2014	A	$E_C - 0.51$	$2 \times 10^{-14}, 2.6 \times 10^{-14}$	4	Tool: Silvaco
	D	$E_V + 0.48$	2×10^{-14}	3	
Perugia 2016	A	$E_C - 0.42$	$1 \times 10^{-15}, 1 \times 10^{-14}$	1.613	improving *Perugia 2006*
(p-type sensors)	A	$E_C - 0.46$	$7 \times 10^{-15}, 7 \times 10^{-14}$	0.9	$\phi_{eq} \leq 7 \times 10^{15}\ \text{cm}^{-2}$
	–	–	$3 \times 10^{-15}, 3 \times 10^{-14}$	–	$7 \times 10^{15}\ \text{cm}^{-2} \leq \phi_{eq} \leq 1.5 \times 10^{16}\ \text{cm}^{-2}$
	–	–	$1.5 \times 10^{-15}, 1.5 \times 10^{-14}$	–	$1.5 \times 10^{16}\ \text{cm}^{-2} \leq \phi_{eq} \leq 2.2 \times 10^{16}\ \text{cm}^{-2}$
	D	$E_V + 0.36$	$3.23 \times 10^{-13}, 3.23 \times 10^{-14}$	0.9	

Figure 2.11: Overview of representative TCAD defect models developed for simulating radiation damage in silicon detectors. These models differ in the number and type of effective defect levels (donor/acceptor) used to reproduce post-irradiation electric field profiles and charge transport. Adapted from [28].

3 Design and fabrication of silicon sensors

3.1 Manufacturing of silicon sensors

The manufacturing of silicon sensors involves a series of complex, high-precision processes aimed at forming the active and passive structures on a silicon wafer. This production workflow can be divided into three main phases. The first one is the *crystal growth*. Although not typically considered part of wafer processing itself, crystal growth is a fundamental step for the fabrication of silicon sensors. The next phase, referred to as the *front-end fabrication*, covers wafer fabrication. It is a sequence of complex steps that defines the sensor's active structures, implant profiles, and interconnections. The third, known as the *back-end fabrication*, involves the assembly and packaging of the diced chips, ensuring mechanical stability and electrical connectivity. A detailed discussion of back-end techniques, including high-density interconnection and bonding technologies, is provided in Chapter 13, within the context of semiconductor detector assembly for high-energy physics applications. The crystal growth and front-end processing steps are discussed further in this chapter.

3.1.1 Crystal growth

Crystal growth is a highly specialised discipline involving thermodynamics, kinetics, materials chemistry, and fluid dynamics. The goal is to produce high-purity, monocrystalline silicon crystals with precise crystallographic orientation, minimal defects, and tailored electrical properties suitable for electronic applications.

The number of available crystal growth techniques is large and diverse, reflecting the wide range of material systems and application requirements. A common way to classify these methods is by the phase of the nutrient medium from which the crystal is grown. Single crystals may be formed from a liquid phase, such as a melt or solution, from a vapour phase, through condensation, sublimation, or chemical reaction, or directly within a strained solid. Regardless of the nutrient phase, crystal formation requires the creation of a stable solid–liquid (or solid–vapour) interface. In practice, a seed crystal is typically used to initiate and stabilise the growth front, enabling better control over crystal orientation and structural quality [29].

Crystal growth from the melt is the most commonly used approach for silicon. The two principal methods for producing detector-grade material are the Czochralski (CZ) and Float-Zone (FZ) techniques [29]. Both methods belong to the same group, as crystals are formed from the liquid phase. Still, they differ significantly in terms of thermal management, impurity control, and achievable crystal quality. A detailed overview of these and other crystal growth

methods, along with their historical development and technical foundations, can be found in the "Handbook of Crystal Growth" [30] and in "Silicon Semiconductor Technology: Processing and Integration of Microelectronic Devices" by Ulrich Hilleringmann [31].

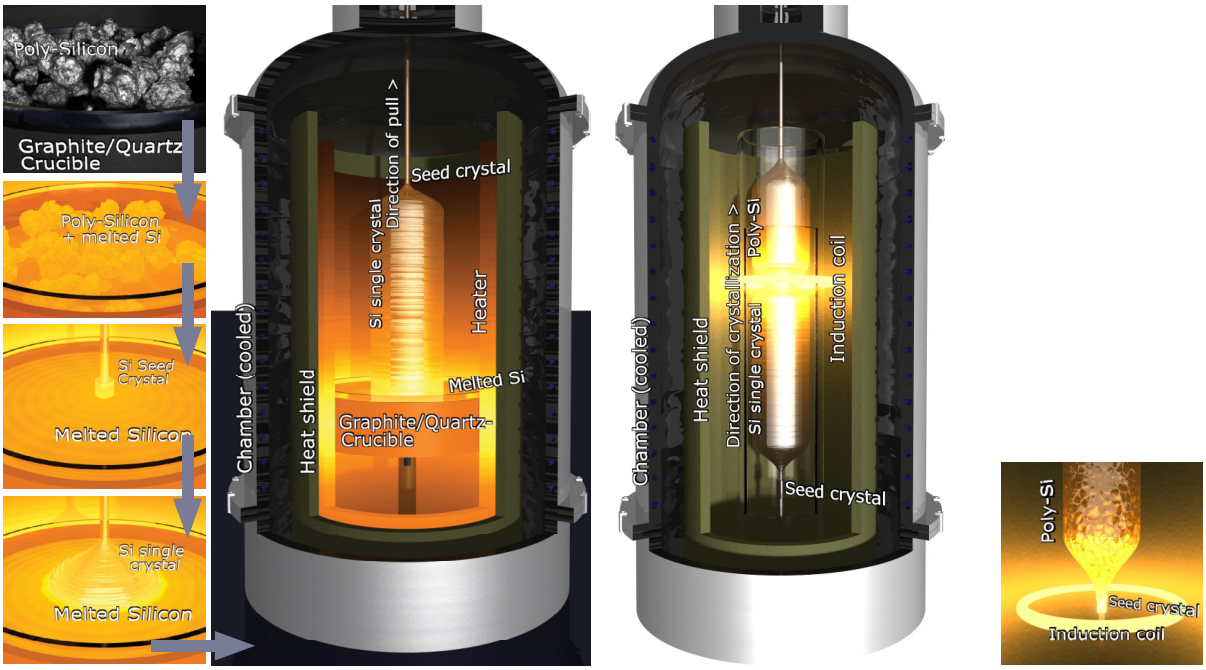
In the *Czochralski process*, polycrystalline silicon is melted in a rotating quartz crucible heated at approximately 1425 °C. A seed crystal is dipped into the molten silicon, initiating solidification at the interface. The crystal is slowly pulled upwards while rotating counter to the crucible, maintaining a stable temperature gradient near the solid–liquid boundary. The formed crystal is named *ingot*. Its cross-section gets smaller the faster the ingot moves up, and the usual growth rate is 3–20 cm/h [32, 31]. The scheme of this process is shown in Figure 3.1a.

Typically, transistor and circuit integration use $n-$type or $p-$type silicon wafers. The dopants, phosphorus or boron, define the desired electrical properties of the crystal. The corresponding doping material is already dissolved in the melt, so the doping atoms are incorporated into the lattice during crystal growth [31].

The Czochralski process enables the growth of large-diameter crystals, up to 18 inches, which is 46 cm, making it the standard method for the microelectronics industry. The diameter of the growing ingot is controlled by the pulling rate, and the uniformity is improved through opposing rotations of the seed and crucible. However, contamination from the crucible walls leads to incorporation of oxygen and carbon, which is typically in the range of 10^{18} cm^{-3} and 10^{17} cm^{-3}, respectively, limiting the material's minority carrier lifetime. Typical resistivities for CZ silicon range from 0.01 Ω·cm to 50 Ω·cm, though higher values are attainable under controlled growth conditions and method modifications [32, 31].

The *Float-Zone method* offers an alternative crystal growth approach with significantly lower impurity incorporation. This process is shown in Figure 3.1b. A radio-frequency (RF) induction coil is used to locally melt a narrow zone of a vertically mounted polycrystalline ingot. A seed crystal initiates crystallisation at one end, and the molten zone is slowly moved along the ingot length. As solidification proceeds, the melt inherits the orientation of the seed and the crystal forms progressively along the axis. Impurities are rejected into the melt and carried toward the tail of the ingot, which is later removed [32].

This method avoids contact between the melt and the quartz crucible, significantly reducing oxygen and carbon concentrations and allowing for the production of extremely high-resistivity (1–10 kΩ·cm) material with long carrier lifetimes. For this reason, FZ silicon is commonly used for detector-grade applications, particularly in high-energy physics, where high resistivity is essential. However, it is limited to smaller crystal diameters, typically up to 8 inches, which is about 20 cm, and the production cost per wafer is considerably higher than with the CZ method.



(a) Illustration of the Czochralski process. Polycrystalline silicon is melted in a quartz crucible. A seed crystal initiates growth as it is slowly pulled upward, forming a monocrystalline ingot with defined orientation [32].

(b) Illustration of the Float-Zone method. An induction coil locally melts a zone in a polysilicon rod. The molten region moves along the ingot while crystallisation follows the seed orientation [32].

Figure 3.1: Crystal growth methods for silicon ingot production: (a) the Czochralski and (b) the Float-Zone methods

3.1.2 Front-end processing steps

Front-end processing involves the parallel fabrication of many identical circuits, or *dies*, across a single silicon wafer. This is a multi-step sequence, but not strictly linear. Several operations are repeated at different stages to achieve the desired complexity of the structure. The full production flow, excluding the crystal growth, is summarised in Figure 3.2, which illustrates the sequence of key front-end processing steps involved in sensor fabrication.

Starting material: pre-doped wafer as a substrate

The fabrication of silicon sensors begins with a high-purity monocrystalline wafer, which is sliced from an ingot. These wafers are delivered by the supplier already doped to a specified resistivity range. The initial doping is introduced in a controlled manner to tailor the electrical conductivity of the material. It is done either during crystal growth, by dissolving dopants into the molten silicon, or through diffusion processes applied to the wafer bulk after growth. Phosphorus and arsenic are common $n-$type dopants, contributing free electrons, whereas boron is widely used as a $p-$type dopant, creating mobile holes in the valence band [33].

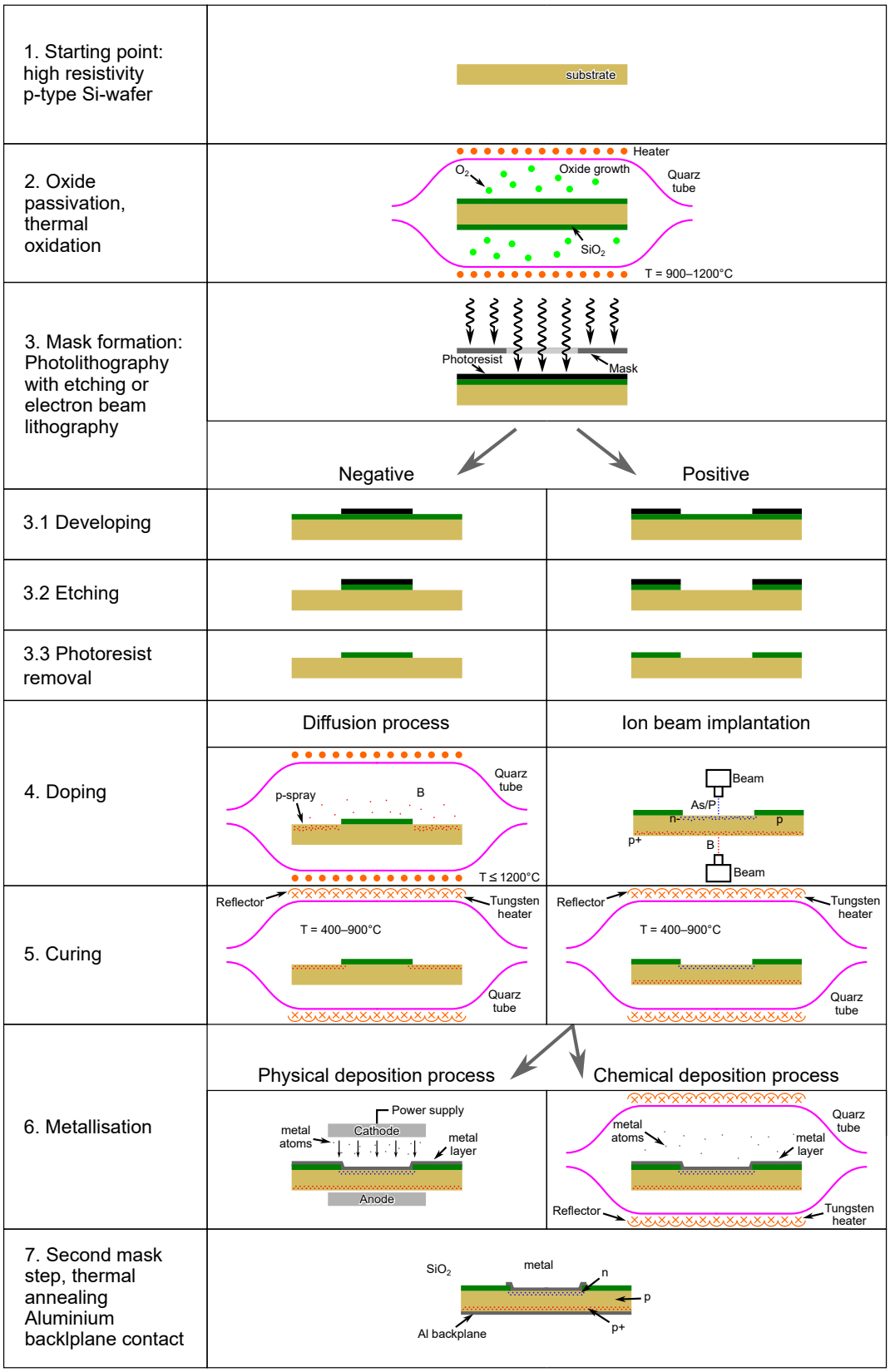


Figure 3.2: Overview of the typical front-end production flow for silicon sensor fabrication, including oxidation, photolithography, etching, doping, metal deposition, and passivation. Several steps are repeated across multiple layers to build up the final sensor structure.

Surface oxidation and passivation

A fundamental step in front-end processing is the formation of a thin silicon dioxide (SiO_2) layer on the wafer surface through thermal oxidation. This oxide serves multiple purposes: it acts as an insulating layer, a diffusion barrier, a mechanical and chemical passivation layer, and a hard mask for subsequent photolithography. The oxidation method, oxide thickness, and quality must be tuned for the specific electrical, optical and mechanical requirements. For sensor fabrication, the oxide also provides alignment marks for mask placement, protects the surface from contamination, and ensures that previously established doping profiles are maintained during high-temperature steps [31].

The oxide layer is typically grown in a quartz-tube furnace operating at temperatures between 900 °C and 1200 °C. Thermal oxidation occurs when oxygen, acting as the reaction gas, comes into contact with a heated silicon surface. A chemical reaction takes place between the oxygen and the silicon substrate, forming a layer of amorphous silicon dioxide SiO_2 on the wafer surface. Two primary thermal oxidation techniques are used in sensor processing: dry and wet oxidation.

Dry oxidation of silicon takes place in an atmosphere of pure oxygen and follows the chemical reaction:

$$Si + O_2 \rightarrow SiO_2 \tag{3.1}$$

This reaction typically proceeds at high temperatures in the range of 1000–1200 °C to achieve practical growth rates. This method is suitable for forming thin, dense oxide films, which have high breakdown voltage, making them ideal for high-field applications such as the gate oxide in MOS transistors [31]. In high-energy detector fabrication, the oxide layer is typically around 200 nm thick and grown via dry thermal oxidation at approximately 1030 °C [34].

In *wet oxidation*, oxygen gas is passed through a bubbler vessel containing water heated to 90–95 °C. The resulting water-saturated oxygen stream is introduced into the oxidation furnace. At the heated semiconductor surface, water vapour reacts with silicon to form silicon dioxide according to the chemical reaction:

$$Si + 2H_2O \rightarrow SiO_2 + 2H_2 \tag{3.2}$$

This whole process typically takes place at temperatures between 900 °C and 1100 °C. Even at relatively low temperatures, the growth rate of silicon dioxide in wet oxidation is significantly higher than in dry oxidation. This makes it ideal for forming thick oxide layers, such as field oxides or masking layers, in applications where oxide thickness is prioritised over electrical quality [31].

Lithography and pattern transfer

After the wafer surface has been passivated by silicon dioxide, the local definition of device features is made through lithography. The lithography technique includes a sequence of

process steps: dehydration of the wafer surface, application of an adhesion promoter (primer), resist coating, solvent outgassing through prebaking, exposure of the resist via a photomask or electron beam, development of the coating, post-exposure bake for resist hardening, and finally, optical inspection of the patterned structures.

Exposure is traditionally performed with ultraviolet light at wavelengths of 436 nm (G-line) or 365 nm (I-line), emitted by mercury vapour lamps. More advanced systems use excimer lasers, X-ray or electron-beam lithography. The last one enables direct patterning without a mask.

An additional importance has been given to a photomask. It is a thin, radiation-sensitive film, usually an organic photoresist layer, which is responsible for creating structure patterns. Lithography in general can be divided into two techniques: positive and negative resist. During development, the photoresist dissolves in the exposed regions for *positive resists*, leaving the non-exposed areas intact on the wafer surface. In contrast, for *negative resists*, the exposed regions remain after development, while the unexposed resist is removed. Negative resists generally offer greater chemical and thermal stability compared to positive coatings. Modern negative resists now support sub-micrometre patterning, although positive resists remain preferred for the finest critical layers because they avoid swelling during post-exposure processing.

In *electron beam lithography*, a focused, computer-controlled electron beam is scanned directly over the resist-coated wafer, similar to the method used for photomask fabrication. The resist used in this technique requires a minimum dose of electron irradiation to become soluble in the developer. During exposure, electrons deposit their energy locally in the resist layer, selectively altering its dissolvability. Regions that should remain unexposed are actively blanked during scanning, enabling highly precise patterning without the use of a physical mask [31].

Etching

Once the resist pattern is defined, it must be transferred into the underlying material stack. Etching techniques are used either to remove a material layer completely or to transfer the lithographically defined photoresist pattern into the underlying substrate. Together, lithography and etching form the core pattern-transfer loop that is repeated many times throughout front-end processing, progressively building up the transistors, sensing and interconnect structures. There are two main etching techniques: wet chemical etching and dry (plasma) etching.

Wet chemical etching selectively dissolve specific materials. It converts the solid material of the layer to be removed into a liquid compound using an acidic or alkaline solution. This etching process has an isotropic effect, leading to lateral under-etching beneath the mask edges. Nevertheless, the method offers very high selectivity between different materials and is still favoured for total oxide removal and certain metal etches. However, as an exception, the element silicon can only be etched with low selectivity to silicon dioxide.

Dry etching or *plasma etching* enables highly controlled, well-reproducible, homogeneous removal of almost all materials in silicon semiconductor technology, with good selectivity to

both mask and substrate. It supports both anisotropic and isotropic etch profiles and achieves precise pattern transfer with minimal reagent usage. The process uses gaseous media excited by a gas discharge in an alternating radio-frequency field in the vacuum range of approximately 1 to 100 Pa. Despite the high equipment cost, dry etching has largely replaced wet methods for defining transistor gates and performing metallisation etches due to its superior accuracy and reproducibility [31].

Doping: diffusion and ion implantation

Doping introduces donor or acceptor impurities into selected wafer regions to tailor the electrical properties and form device-grade $p - n$ junctions. Two main methods are employed: thermal diffusion and ion implantation.

Diffusion introduces dopants into silicon by exposing the wafer to vapour, liquid, or solid sources at temperatures between 800 °C and 1200 °C. The wafers are loaded into a high-purity quartz tube, uniformly heated by resistance elements and monitored by thermocouples, ensuring that the axial temperature variation is less than 0.5 °C over a 50 cm zone. In modern practice, diffusion is almost exclusively a flow-through process: a carrier gas (Ar, N_2, or O_2) enriched with donor or acceptor source passes over the wafers, establishing equilibrium between the furnace atmosphere and the silicon surface. The typical gases for diffusion with gas sources are phosphine (PH_3), diborane (B_2H_6), or arsine (AsH_3) [31].

Although diffusion is straightforward and suitable for forming broad, lightly doped regions, such as p-spray layers, it offers limited control over junction depth and lateral spread. Uniform doping across multiple wafers requires strictly laminar gas flow; however, turbulence near the wafer stack often induces local variations in dopant partial pressure, resulting in sheet-resistance non-uniformities. Dopant uptake by the quartz tube further degrades batch-to-batch reproducibility. Moreover, dopants can enter the silicon not only vertically through the oxide window but also laterally beneath the masking layer.

For applications that demand abrupt, well-defined $p - n$ junctions, such as the charge-collection diodes, the limited depth and lateral control of diffusion make it unsuitable; *ion implantation* is therefore the method of choice. In this process, dopant ions are generated in an ion source, accelerated through an electric field, and directed onto the wafer. The implanted dose is set with high precision by measuring the total delivered charge, while the accelerating voltage fixes the projected range of the ions in silicon. Consequently, both dopant concentration and junction depth can be tailored with sub-micrometre accuracy [31]. A more detailed discussion of dopant-profile modelling with TCAD is provided in Section 4.1.3.

Post-implantation annealing and dopant activation

Ion implantation leaves most dopant atoms in off-lattice positions, electrically inactive and accompanied by significant point-defect damage. Thermal annealing is therefore required for two purposes: to repair the lattice and to drive the dopants onto substitutional sites. The key parameter is the peak temperature rather than the duration of the temperature treatment.

Below 400 °C only 1–10% of dopants become active, whereas temperatures of 900–1000 °C yield nearly complete activation even at high doses [31].

Deposition processes and metallisation

Deposition processes aim at the production of homogeneous, particle-free layers with high electrical quality, minimal impurity incorporation and mechanical stress. Depending on the application and technique, the deposited films can be monocrystalline, polycrystalline, or amorphous. Two classes of techniques are available: chemical deposition and physical deposition [31]. A fundamental distinction exists between these two methods. Chemical methods involve the formation of solid films through surface reactions with gases. In contrast, physical techniques rely on direct transfer of material from a source to the substrate. Both approaches are extensively used in semiconductor fabrication, for example, for metallisation, dielectric formation, and protective passivation layers.

Chemical deposition techniques are used extensively in semiconductor technology to deposit crystalline or amorphous layers with specific chemical and electrical properties. When the deposited crystalline layer adopts the atomic structure of the substrate, the process is referred to as *epitaxy*. This is further classified into *homoepitaxy*, where the substrate and layer are of the same material (e.g. silicon-on-silicon), and *heteroepitaxy*, where different materials are involved, such as the deposition of SiGe alloys or silicon-on-insulator (SOI) structures grown on sapphire or spinel substrates [31].

Such epitaxial techniques are particularly important for the fabrication of heterojunction bipolar transistors and the development of BiCMOS technology. These topics are part of the research in this thesis and are discussed in detail in Part III "Bi-CMOS technology for high-speed monolithic sensors for particle physics". In the same work, the influence of different homoepitaxial silicon layers on detector performance was also investigated.

A widely employed chemical method is *Chemical Vapour Deposition (CVD)*, in which reactant gases decompose thermally or under plasma excitation to form solid layers on the wafer surface. The substrate itself acts only as a mechanical support and does not participate chemically. CVD allows the formation of essential materials like silicon dioxide, silicon nitride, silicon oxynitride, and polycrystalline silicon, commonly used for gate electrodes and interconnects.

Physical deposition includes Molecular Beam Epitaxy (MBE), evaporation and sputtering. *Molecular Beam Epitaxy (MBE)* enables the deposition of ultra-thin epitaxial layers under ultra-high vacuum conditions ($\sim 10^{-8}$ Pa). *Evaporation* is a widely used method for depositing metallic films onto silicon wafers. Depending on the heat source, it is categorised as thermal or electron beam evaporation. In both methods, metal is evaporated in a high vacuum and condenses onto the substrate. Due to the long mean free path of the particles, the material travels in straight lines and deposits perpendicularly onto the wafer surface, leading to poor step coverage. Aluminium is commonly used for evaporation. *Sputtering* addresses some limitations of evaporation, particularly in terms of step coverage and compositional accuracy. In this method, argon ions are accelerated towards a target material in a vacuum chamber. Upon impact, atoms from the target are ejected and deposited onto the substrate. Sputtering

occurs at higher pressures (1 Pa), which leads to shorter mean free paths and more scattered particle trajectories. This results in better coverage of non-planar surfaces.

The overall physical deposition efficiency is low. Sputtering generally produces more penetrable films, which may require post-deposition annealing to improve their properties. Nonetheless, the oxide quality achieved via sputtering is still inferior to that obtained with Low-Pressure CVD techniques [31].

Electrical connections to the sensor structures are formed by metallising both the front and back sides of the wafer. In the context of front-side metallisation, aluminium or aluminium alloys are commonly deposited using sputtering or CVD. These conductive layers form the read-out pads and routing lines. Photolithographic patterning followed by etching removes unwanted metal regions.

The backplane is commonly coated with a continuous metal layer to ensure uniform biasing across the entire sensor area. A full-area metallisation, often deposited by sputtering, is followed by a mild annealing step, which is typically at 400–450 °C, to improve adhesion and reduce the contact resistance between the metal and underlying silicon [35].

3.2 Design foundations for silicon sensors

The performance of a solid-state detector is often evaluated in terms of its signal-to-noise ratio (SNR). A high-quality device must maximise the charge signal collected from a traversing particle while simultaneously minimising internal noise. These two goals, however, impose contradictory requirements on the detector material.

To generate a large signal, the material should have a low mean ionisation energy, which implies a small band gap E_g. This ensures that many electron–hole pairs are produced by each ionising particle. Conversely, to achieve low noise, the material must have very few thermally excited carriers at room temperature, which favours a large band gap. An ideal compromise between these demands would require a band gap of approximately $E_g \approx 6$ eV, which is a value characteristic of diamond. However, even chemical vapour deposited (CVD) diamond remains prohibitively expensive for large-scale tracking applications [35, 36].

A basic calculation illustrates how silicon performs in this context. Consider a silicon detector with thickness $d = 300$ µm (0.03 cm) and area $A = 1$ cm^2. The average energy required to produce a single $e - h$ pair in silicon is $I_0 = 3.62$ eV, and a minimum ionising particle (MIP) deposits energy at a rate of $\frac{dE}{dx} = 3.87$ MeV $\cdot$ cm^{-1}. The number of $e - h$ pairs generated by MIP in such a detector, simply speaking, a signal is given by:

$$N_{\text{signal}} = \frac{\frac{dE}{dx} \cdot d}{I_0} = \frac{3.87 \times 10^6 \text{ eV cm}^{-1} \times 0.03 \text{ cm}}{3.62 \text{ eV}} \approx 3.2 \times 10^4 \; e - h \text{ pairs} \tag{3.3}$$

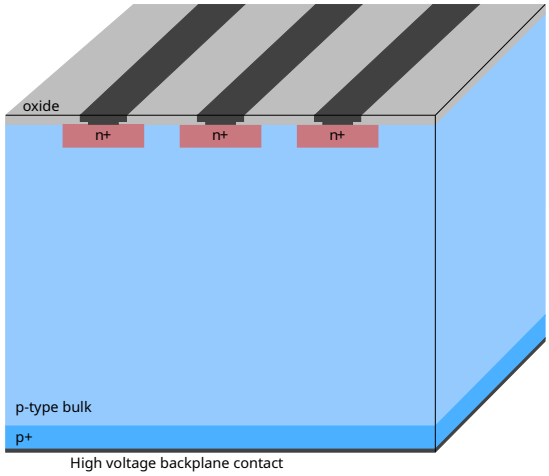
In contrast, the number of thermally generated carriers, meaning noise, in the same detector volume at $T = 300$ K is estimated using the intrinsic carrier concentration, n_i, which is equal to 1.45×10^{10} cm^{-3}:

$$N_{\text{noise}} = n_i \cdot d \cdot A = 1.45 \times 10^{10}\ \text{cm}^{-3} \times 0.03\ \text{cm} \times 1\ \text{cm}^2 \approx 4.35 \times 10^8\ e-h\ \text{pairs} \tag{3.4}$$

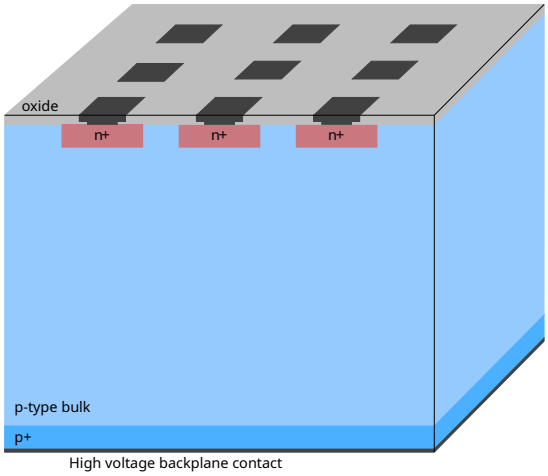
Thus, without depletion, thermally generated $e-h$ pairs outnumber the signal carriers by approximately four orders of magnitude. This highlights the essential role of full depletion, typically achieved by applying a reverse bias across a $p-n$ junction, to suppress intrinsic noise and enable particle detection [35]. This principle drives the operation of standard $p-n$ junction detectors. While such devices may differ in segmentation, geometry, or read-out architecture, they all rely on this fundamental concept.

3.3 Silicon sensors segmentation

To determine the precise position of a traversing charged particle, silicon sensors are spatially segmented into electrically independent readout structures. The two principal approaches are the strip detector, which provides one-dimensional (1D) spatial information, and the pixel detector, which offers a two-dimensional (2D) resolution.



(a) Illustration of a single-sided strip detector



(b) Illustration of a planar pixel detector

Figure 3.3: Examples of sensor segmentation in planar silicon detectors: (a) strip detector architecture and (b) pixel detector layout. Both structures are based on the $p-n$ junction principle and designed for charge localisation.

3.3.1 Strip sensor

Figure 3.3a shows the schematic of a single-sided strip sensor. These devices are fabricated on n–or p–type silicon substrates. A $p-n$ junction is formed between the bulk and the

segmented front-side implants, which are either p^+ or n^+, depending on the substrate type. Under reverse bias, the depletion region expands from this junction towards the backside. The bias contact is typically realised via a full-area n^+ (for $n-$type bulk) or p^+ (for $p-$type bulk) implant on the backplane, although top-side biasing schemes are preferable in some designs. The sensor surface is coated with a SiO_2 passivation layer, which electrically isolates the implants and protects the device against environmental damage. In some cases, additional passivation materials such as silicon nitride or polyimide are employed to improve mechanical robustness or radiation tolerance [37].

In a strip detector, the readout electrodes are implemented as long, parallel implants with uniform pitch. When a particle traverses the sensor and deposits energy in the active volume, the corresponding strip collects the resulting charge carriers, allowing reconstruction of the hit position along one coordinate. Strip sensors typically incorporate 500–800 independent readout channels [38].

An important issue in strip sensors is microdischarge, which can occur due to strong localised electric fields at the strip edges, particularly on the side carrying the full depletion voltage. Although this effect is usually non-destructive, it causes the affected preamplifier channels to saturate, forcing them to be disabled during data acquisition [37]. Despite this, strip detectors remain widely used in applications where the particle density is moderate and large sensor areas are required, such as outer tracking layers in collider detectors, thanks to their relatively low material budget, simpler architecture, and mature fabrication technology.

3.3.2 Pixel sensor

Figure 3.3b illustrates a typical pixel sensor layout, which consists of a two-dimensional matrix of readout cells. While based on the same $p-n$ junction principle, the segmentation into isolated pixels enables fine-grained charge localisation. Each pixel is connected to a dedicated front-end channel, typically through bump bonding, or can include integrated electronics. The small electrode size reduces parasitic capacitance, improving signal-to-noise performance. This configuration enables precise spatial resolution in both coordinates, making it ideal for environments with high track multiplicities or overlapping events [38]. The dimensions and geometry of pixel cells can be tailored to the needs of the experiment, offering a high degree of design flexibility.

However, this architecture introduces manufacturing complexity. In hybrid pixel detectors, each pixel must be capacitively coupled or bump bonded to its corresponding readout cell on the ASIC. The latter involves tens of thousands of micro-bumps. This process is difficult to rework in the case of bonding failures. When the high-voltage side of the sensor faces the electronics, careful passivation is required to prevent surface breakdown. The hybrid integration of sensor and readout ASIC contributes significantly to the material budget, with each layer adding approximately 1.5% X_0, primarily due to power delivery and cooling infrastructure [37].

Recent research efforts aim to reduce this material budget and improve system integration. Strategies include the use of thinner substrates, on-detector DC–DC conversion, serial powering schemes, and embedded power conductors. A more transformative approach involves

Monolithic Active Pixel Sensors (MAPS), where both sensing and signal processing elements are integrated in a single CMOS die. This eliminates the bump bonding step, which reduces the input capacitance and enables ultra-thin detector designs. A detailed comparison of hybrid and monolithic implementations is provided in Section 3.5.

Taken together, there is no such thing as a universal solution. Both segmentation strategies have their place. While pixel detectors offer unmatched granularity and radiation hardness, their increased complexity and material requirements pose integration challenges. Strip detectors, in contrast, provide a cost-effective and technically robust solution for large-area tracking where occupancy is lower and spatial resolution demands are moderate.

3.4 Coupling and biasing of silicon sensors

In silicon strip and pixel detectors, each segment (strip or pixel) must be individually read out while maintaining a well-defined biasing scheme for efficient depletion and signal collection. This section introduces the principles and implementation techniques of both coupling and biasing, with emphasis on modern methods employed in high-energy physics detectors.

3.4.1 Coupling schemes

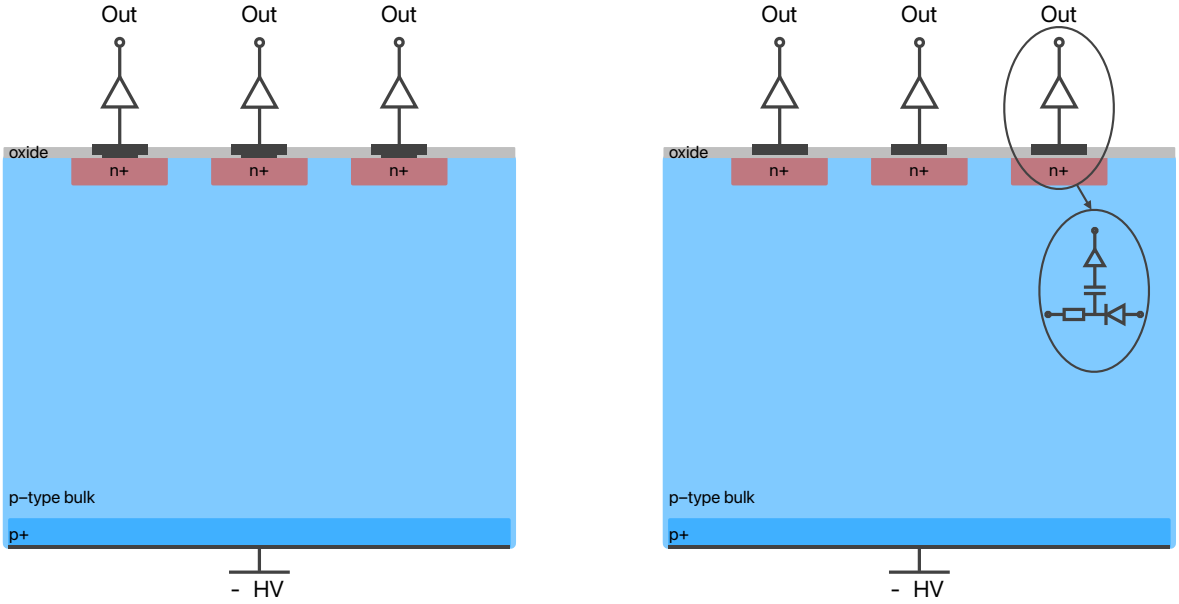
Each strip or pixel can be connected to its dedicated readout channel via one of two methods: DC coupling or AC coupling. Both coupling schemes are illustrated in Figure 3.4.

In the case of *DC coupling*, the readout metallisation is directly connected to the implant. This straightforward configuration allows the signal to be transmitted without additional capacitive interfaces. However, it also means that the leakage current from the sensor flows directly into the front-end amplifier. While this current is relatively small for pixels due to their limited area, it can become substantial in irradiated strip detectors, often exceeding the signal current. Since the leakage current varies with both radiation damage and bias voltage, designing an amplifier that tolerates such a wide range of input currents is challenging [38].

To mitigate this issue, many designs use *AC coupling*, which electrically isolates the DC leakage current from the AC signal, allowing only the signal to reach the readout. In this configuration, the coupling capacitor is formed by placing a dielectric, typically SiO_2, between the implant and the readout metal layer. It ensures that only the AC signal reaches the amplifier, preserving signal integrity and preventing saturation. An RC low-pass filter is usually used to decouple the bias voltage.

3.4.2 Biasing techniques

In segmented silicon detectors, proper biasing and signal decoupling are essential for reliable charge collection and readout. Biasing ensures that all strip or pixel implants are held at a



(a) DC-coupled structure. The implant is directly connected to the readout electrode. While simple, this approach requires the amplifier to sink the leakage current.

(b) AC-coupled structure. A capacitor is formed between the implant and the readout metal, with the DC leakage current diverted via a bias resistor.

Figure 3.4: Schematic comparison of DC and AC coupling methods for strip sensors. Both designs are based on a reverse-biased $p - n$ junction in a p-type bulk.

defined potential relative to the substrate. During operation, the bias line is typically grounded, while the bias voltage is applied to the backside of the detector.

In AC-coupled devices, a dedicated biasing scheme is required to connect each implant to a common bias line while maintaining electrical isolation between individual channels. Several methods exist for biasing AC-coupled sensors, each offering a trade-off between simplicity, radiation tolerance, and fabrication complexity [35].

- *Polysilicon biasing method* provides a high-resistive path, which is established between each strip implant and the bias line using a strip of polycrystalline silicon (polysilicon). It offers excellent stability and radiation tolerance, but requires additional processing steps.
- *Punch-through structures* establish conduction between neighbouring high-doped regions once the inter-strip potential exceeds the punch-through voltage. Punch-through structures avoid additional fabrication but may become unreliable under radiation exposure.
- The *Field-Oxide Field-Effect Transistor (FOXFET)* method dynamically controls the bias resistance using a gate electrode over field oxide. This biasing method provides integration flexibility, though its performance depends heavily on oxide quality and processing control. It has simplified processing but offers lower radiation robustness.

In DC-coupled detectors, the bias potential is supplied directly through the readout electronics, which act as the ground reference and eliminate the need for additional bias resistors or rings. This approach definitely simplifies the sensor design. Nonetheless, for strip detectors operating in high-radiation or high-channel-count environments, AC coupling combined with polysilicon biasing remains the most robust and widely adopted solution.

3.5 Readout concepts in segmented silicon sensors

Readout architectures play a critical role in defining the performance, complexity, and scalability of silicon detectors. In modern high-energy physics experiments, pixel and strip sensors are not only required to detect individual particle hits with high spatial and temporal precision but also to transmit this information efficiently to the data acquisition system. Two primary architectures have been developed to address these requirements: hybrid readout, where sensor and electronics are manufactured separately and connected via bump ot wire bonding, and monolithic readout, where both are integrated into a single CMOS substrate. Each approach carries its own advantages and constraints in terms of radiation tolerance, resolution, power consumption, and production cost. The following subsections provide an overview of both hybrid and monolithic sensor technologies as employed in silicon tracking systems.

3.5.1 Hybrid sensors

The hybrid sensor system consists of two physically separate chips: a high-resistivity silicon sensor for detection, and a dedicated readout chip responsible for signal amplification and processing. The two components are electrically and mechanically connected via bump bonding, wire bonding or microcables. This design enables independent optimisation of the sensor and electronics, providing radiation tolerance and fast, localised signal processing.

A representative example of a hybrid system is the Silicon Tracking System (STS) in the CBM experiment, which employs double-sided strip sensors with readout via front-end electronics mounted on external boards. The assembly of this detector is highly complex and serves as a benchmark for hybrid strip systems. Detailed information on this architecture is provided in Chapter 14.

Hybrid pixel detectors introduce additional integration challenges. Each pixel must be individually connected to its readout counterpart via bump bonding. This process involves multiple high-precision steps, including: deposition of an under-bump metallisation layer, photolithographic patterning of the contact pads, deposition of the bump material, etching to remove excess metal, and thermal reflow to shape the bumps into robust solder spheres. An example of such an assembly can be found in Section 13.2, where the Pixelated Silicon Drift Detector (PixDD) is discussed. A more detailed overview of interconnection technologies, including bump bonding, wire bonding, and flip chip techniques, is provided in Chapter 13.

The performance of hybrid pixel detectors is defined by the low leakage current per pixel, usually ~1 pA, and a small input capacitance around 1 fF, which together yield a high signal-to-noise ratio of up to 150 [35]. The spatial resolution is directly related to the pixel pitch, which is commonly 100×150 µm^2, with achieved resolutions in the order of 15 µm. However, hybrid architectures come with several drawbacks. The large number of readout channels increases power consumption and system-level complexity. The bump bonding process is costly, requiring tight alignment tolerances and cleanroom infrastructure. Moreover, scaling these systems to large-area coverage becomes increasingly expensive and technically demanding.

In conclusion, while designing hybrid sensors, several practical considerations must be addressed. The connection between the sensor and electronics must have low capacitance to minimise serial noise, especially important for devices like drift detectors. Shielding of the front-end input from external capacitive pickup is essential, as is thermal management of the electronics and, in some cases, the sensor itself [39].

3.5.2 Monolithic sensors

Monolithic sensors combine both the sensing volume and readout electronics within a single silicon substrate. Unlike hybrid architectures, which require bump bonding between a separate sensor and ASIC, monolithic devices integrate all components on a common wafer. This architecture significantly reduces material budget, simplifies assembly, and enables cost-effective large-area production.

In conventional Monolithic Active Pixel Sensors (MAPS), the active volume consists of a low-doped (high-resistivity) layer grown atop a high-doped (low-resistivity) bulk, allowing charge generated by ionising particles to be collected by drift and diffusion. Signal processing circuitry, which is usually produced in CMOS technology and consists of NMOS and PMOS transistors, is implemented within isolated wells on the same substrate. Signal routing and logic elements are defined using photolithographic layers for wells, polysilicon, diffusion, contacts, and metal interconnects.

This co-integration eliminates the need for bump bonding and reduces the interconnect capacitance. In systems where external connections would otherwise introduce several picofarads of parasitic capacitance, monolithic integration allows detector-to-electronics capacitance to remain in the 10–100 fF range, greatly improving the signal-to-noise performance of the front-end amplifiers [39].

However, the integration of sensor and electronics is not straightforward. Standard CMOS processes are not optimised for the high-resistivity substrates required for particle detection, which is an essential property for low leakage current and efficient charge collection. Moreover, electronics fabrication often involves high-temperature steps and heavily doped substrates, both of which can degrade detector performance. To address these challenges, some MAPS developments have focused on building CMOS-compatible electronics directly within the constraints of detector-compatible processing. This required the development of new transistor structures and circuit designs optimised for low noise and minimal power consumption.

All these challenges drove the development of specialised CMOS variants such as High-Voltage CMOS (HV-CMOS) and High-Resistivity CMOS (HR-CMOS), which enhance depletion depth and radiation tolerance by introducing deep implants or specially engineered wells [35]. In this thesis, monolithic sensors are a major focus of study. Several different prototypes were developed and tested, including HitPix (Chapter 6) and AstroPix (Chapter 7), both based on HV-CMOS technology. In addition, BeMiPix (Chapter 11), developed in a BiCMOS process, highlights the potential of advanced monolithic concepts in future detector systems.

3.6 Isolation and guarding techniques

In segmented silicon detectors, especially those based on the *n*-in-*p* design, effective electrical isolation between neighbouring collection nodes' implants is crucial for maintaining signal integrity and minimising cross-talk. This need arises due to the presence of an electron accumulation layer that can form beneath the SiO_2 surface in $p-$type substrates. If left unaddressed, this conductive channel can electrically short neighbouring n^+ readout electrodes. To prevent this, additional isolation implants are introduced during fabrication.

Three principal methods are commonly employed: p-stop, p-spray, and moderated p-spray isolations [38]. These approaches differ in processing complexity, robustness against breakdown, and radiation hardness.

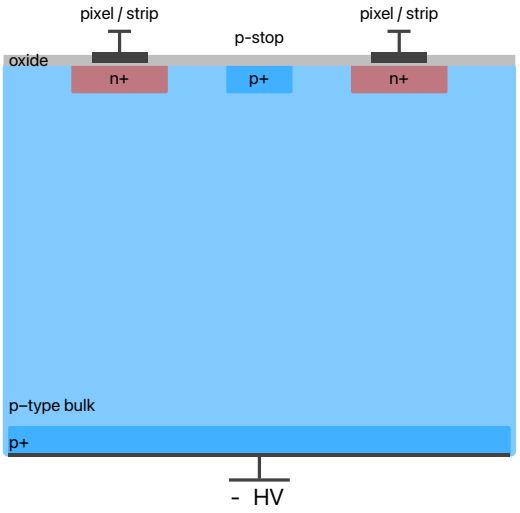
In the *p-stop* technique, highly doped p^+ regions are implanted between neighbouring n^+ readout electrodes. This requires an additional photolithographic mask but ensures effective isolation by interrupting the surface electron layer. However, the geometry and doping of the p-stop structures must be carefully optimised, as too narrow or too doped regions can significantly reduce the sensor's breakdown voltage.

The *p-spray* method introduces a shallow, lightly doped $p-$type layer across the entire sensor surface. This type of implant suppresses the formation of the conductive surface channel without requiring an extra photolithographic mask. While it simplifies processing, its effectiveness is sensitive to the implantation dose: if the dose is too low, isolation may be insufficient; if too high, it can degrade breakdown performance.

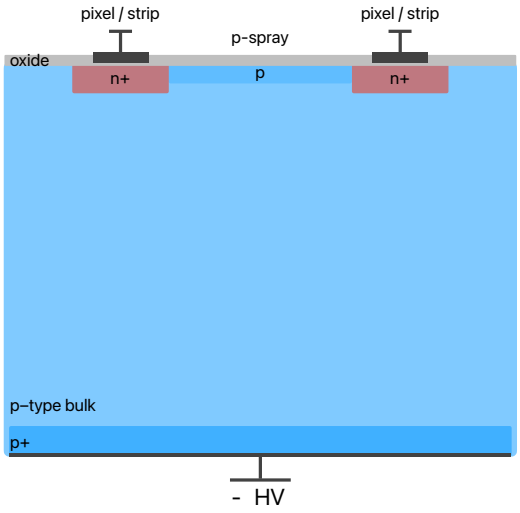
A more advanced approach is the *moderated p-spray* technique, which combines the advantages of both methods. Here, a localised high-dose boron implant is placed between neighbouring electrodes to guarantee inter-pixel or inter-strip isolation. At the same time, a surrounding low-dose p-spray provides a gradual doping profile optimised for high-voltage performance. This graded profile reduces electric field peaks and improves breakdown characteristics, especially in sensors operating at high reverse bias.

These three isolation schemes are illustrated in Figure 3.5. The diagrams show cross-sectional layouts of strip or pixel implants with corresponding p-stop or p-spray isolation structures on a p-type bulk.

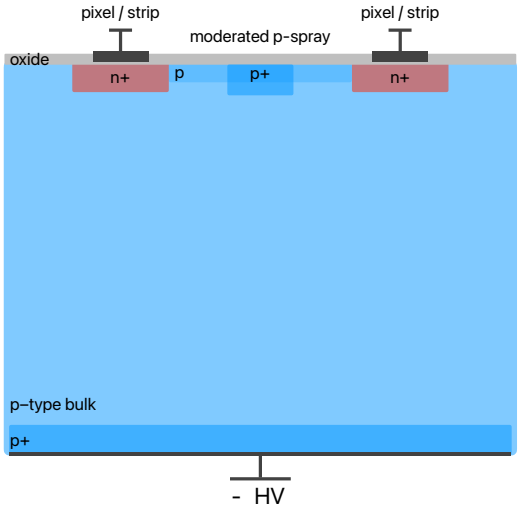
In addition to implant-based isolation, most detectors also incorporate *guard rings (GRs)*, which surround the active sensor area. These are implants of the same doping type as the readout



(a) p-stop isolation



(b) p-spray isolation



(c) moderated p-spray isolation

Figure 3.5: Common surface isolation techniques in *n*-in-*p* silicon detectors. These structures aim to prevent the formation of an electron inversion layer beneath the SiO_2 that could short adjacent readout electrodes.

electrodes and are typically equipped with their own metal contacts. Although referred to as "rings", the term describes their closed-loop layout rather than a circular shape. Guard rings may follow any closed geometry around the sensor periphery. Their primary functions are to shape the electric field at the edges of the sensor, reduce leakage current originating from outside the active volume, and suppress high-field concentrations that could otherwise lead to avalanche breakdown. The number, spacing, and potential of guard rings are carefully tailored to the sensor's operating voltage and geometry. In high-voltage detectors, multiple floating or grounded guard rings are often used to create a graded potential drop from the active area to the edge, ensuring stable and reliable operation over time.

4 Simulations of semiconductor devices — Technology Computer-Aided Design (TCAD)

Physical modelling is commonly used to study prototypes' behaviour, debug produced devices, and improve functionality. It describes the physical behaviour of a system by an abstract mathematical model. The two main approaches in physical modelling are analytical modelling and numerical simulation [40]. Analytical modelling is the representation of the physical property of the system by a lumped-parameter model. This approach is efficient. It allows for the discovery of the essential concepts of device physics and the quick resolution of individual aspects of design problems. However, it has been proven insufficient in solving problems with complex physics or geometry. Numerical simulations, on the other hand, describe the device behaviour through the numerical solution of the differential equations, which represent the device physics in a given geometrical domain. This allows us to address more complex design problems that do not have a closed-form solution. The result tends to be more reliable but challenging to interpret.

Performing numerical simulations requires complex software architectures, expensive licenses of commercial tools, and high computational resources. Analytical modelling and numerical simulations are complementary techniques, often used together but with different specific aims. In semiconductor chip development, these approaches are used during multiple design flow steps and depend on each other's results. Usually, analytical modelling covers circuit simulations and immediate screening of experimental results. At the same time, numerical simulation is often used to understand advanced device physics, scaling analyses and manufacturing processes. This chapter will cover the numerical simulations of semiconductor devices, using the tool called TCAD, which stands for "Technology Computer-Aided Design".

TCAD is a branch of Electronic Design Automation (EDA) that models semiconductor fabrication and semiconductor device operation. TCAD gives us a deep and detailed understanding of the device and its physics. With this tool, it is possible to:

- predict and scale the technology needed for developing a new device;
- optimise the device features when analytical calculations are complicated or impose unacceptable assumptions;
- measure and describe physical quantities of the semiconductors, which are not feasible to experiments, for example, local carriers distribution and local electric field [41].

TCAD can be applied to analyse and design both semiconductor processes and devices. Analysis of the device's experimental data or production processes is important in the first stage of device model development. Careful comparison with experimental data is needed to develop a suitable model. Once the physical model has been created, the behaviour of a system can be simulated. That helps to understand the dependence of system performance on physical parameters. After a deeper investigation of physics models and production processes, the built model can be used to devise more suitable device architectures to achieve the desired functionality. The Technology Computer-Aided Design (TCAD) tool by SYNOPSYS[1] was selected for this thesis. This design software consists of different tools, each with its own programming language and syntax. Logos and main applications of these tools are shown in Figure 4.1.

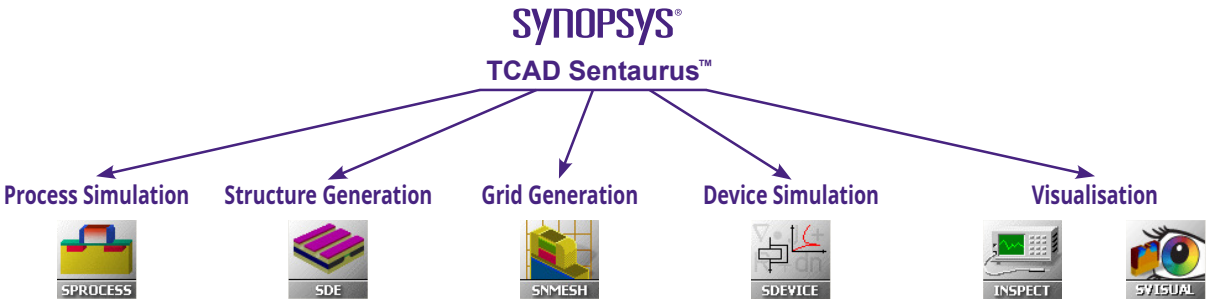


Figure 4.1: Illustration of the most commonly used TCAD Sentaurus tools, including Process Simulation, Structure Generation, Grid Generation, Device Simulation, and Visualisation [40].

In the typical design tool flow, the creation of a device structure, including its doping, is made by process simulation *Sentaurus Process* or process emulation *Sentaurus Structure Editor* tools. Using a process emulation tool, the device structure is realised through idealised procedures that mimic the production process flow. It is used for first-order device analysis, and process simulations can be done once a new device architecture has been optimised and established [40]. The creation of the device structure is followed by the grid generation for numerical computation. This is called meshing and can be performed using *Sentaurus Structure Editor* or *Sentaurus Mesh*. After that, simulation of the device's physics, electrical, thermal, optical and mechanical characteristics is done in *Sentaurus Device*. Finally, *Sentaurus Visual* is used to visualise the two-dimensional and three-dimensional output from the simulation, and *Inspect* is used to plot the electrical characteristics [42]. Further, each of these tools will be discussed in detail.

Sentaurus Process is an advanced, highly flexible, multidimensional process simulator suitable for silicon and non-silicon semiconductor devices. It simulates all standard manufacturing steps, such as lithography, diffusion, ion implantation, oxidation, etching, deposition, and silicidation, and computes mechanical stress. *Sentaurus Process* also allows users to create individual models and to solve customised diffusion equations, including dopant, defect, impurity, and oxidant diffusion equations. This requires detailed modelling of the physical principles of manufacturing and modelling of the specific types of equipment. The real-life calibration of these models needs expensive experiments, whereas the TCAD simulation can

[1] https://www.synopsys.com/manufacturing/tcad.html

dramatically decrease the expenses. During device fabrication, several etching and deposition steps are necessary. Such steps can be modelled in the *Sentaurus Process* using simple geometric operations or mathematical formulations without simulated physical processing. The creation of complex three-dimensional shapes can be assisted by an interface to *Sentaurus Structure Editor* [43].

Sentaurus Structure Editor can be used as a two-dimensional (2D) or three-dimensional (3D) structure editor and as a three-dimensional (3D) process emulator to create TCAD devices. The structures are generated and edited interactively using the graphical user interface (GUI) or in batch mode using scripts. Doping profiles and meshing strategies can be defined in the same ways. The two-dimensional or three-dimensional structure editor covers the following aspects of device creation: generation of the model geometry, Boolean operations between created geometrical bodies, definition of the contact regions, adding of the constant, analytic, and externally generated doping profiles to the model, definition of the local refinements, and inclusion of external sub-meshes. It also provides an interface for the Synopsys meshing engine *Sentaurus Mesh*. In addition, device structures can be defined using emulated process steps such as import or definition of the mask layouts, substrate generation, patterning, isotropic and anisotropic etching and deposition, with or without shadowing and directional effects, polishing and fill operations, and implantation. All three modes, the 2D and 3D structure editor and 3D process emulator, share the same software infrastructure and internal data representation and can be combined freely. The 2D model can be extruded or swept along a curve to generate the 3D model. Afterwards, process steps can be performed on the generated 3D model. Similarly, a 2D slice can be generated from the 3D model and saved in a 2D boundary file [44]. More details on this tool can be found in Section 4.1.

Sentaurus Mesh is a set of tools that produces finite-element meshes for semiconductor device simulations, process simulations, and electromagnetic simulations. It reads the input geometry from a boundary file, and the grid can be adapted to analytic profiles generated by the *Sentaurus Structure Editor* or profiles generated by the *Sentaurus Process*. Calculated and recorded meshes are used by the *Sentaurus Device* to solve physical equations. *Sentaurus Mesh* has several mesh generation engines: an axis-aligned mesh generator, an offsetting mesh generator, and a tensor-product mesh generator. Choosing which mesh generator to use for a particular application depends mainly on the device's geometry. For devices where the most important surfaces are axis-aligned, the axis-aligned mesh generator gives the highest quality elements with minimal node count for such devices. In one dimension, the meshes contain segments; in two dimensions, the meshes contain triangles; in three dimensions, the meshes comprise tetrahedra. This meshing engine generates input data, which is used for the *Sentaurus Device* and the *Sentaurus Process*. For devices where the main surfaces are non-axis-aligned or curved, such as a MOS-type structure with a non-planar channel, it is recommended to use the offsetting mesh generator. The tensor-product mesh generator is used for electromagnetic simulations using *Sentaurus Device Electromagnetic Wave Solver* [45].

Sentaurus Device numerically simulates the electrical behaviour of a single semiconductor device in isolation or several physical devices combined in a circuit. It applies to different types of semiconductor devices, ranging from submicron MOSFETs to bipolar power structures. It can also be used for the III–V compound homostructure and heterostructure devices.

Simulations are performed using advanced physical models and various numerical methods covering the electrical, thermal, optical and mechanical behaviours of these devices. An actual semiconductor device is represented in the simulator as a 'virtual' device whose physical properties are divided into a non-uniform mesh of nodes. Any physical quantity calculated by *Sentaurus Device* at any point between nodes can be obtained by interpolation. *Sentaurus Device* is also a mixed-mode device and circuit simulator. In this use case, multiple devices, such as diodes or transistors, are combined in one simulation, where different mesh and physical models can be applied to individual devices. Additionally, *Sentaurus Device* provides advanced radiational models, which allow us to simulate semiconductor particle detectors [42]. More detail on this tool can be found in Sections 4.2 and 4.3.

All the described TCAD Sentaurus simulation tools can be integrated into one environment using *Sentaurus Workbench*. The *Sentaurus Workbench* tool is the primary graphical front end, providing a convenient framework for designing, organising, and automatically running complete TCAD simulation projects. The organisation of simulations into projects and folders provides a clear overview of the overall simulation environment and gives flexibility in the configuration of the simulation flow. The main advantage of it over the terminal simulation session is the possibility of parameterising input files to automate the project execution. An intuitive graphical user interface simplifies the editing and handling of complex simulation projects [46]. The *Sentaurus Visual* tool is part of *Sentaurus Workbench* visualisation. *Sentaurus Visual* uses both a graphical user interface and a scripting language for automated tasks [47].

Figure 4.2 shows the typical TCAD simulation flow. The diagram gives the step-by-step approach for silicon sensor design, including the type of TCAD tool responsible for the action. Additionally, possible and more valuable results are listed on the right side of the plot. Sections 4.1 to 4.3 provide more detailed view on the design process. The development of the HitPix chip (Chapter 6), developed for Heidelberg Ion-beam Therapy Center (HIT), is used as a code reference. Section 4.1 describes the creation of the geometrical structures, their doping and meshing. Section 4.2 explains how to use different physical models and which calculation methods fit them. Additional attention is paid to the *Radiation Models* used in TCAD simulations due to the focus on particle detector applications in this thesis. Section 4.3 illustrates the choice of the models and Minimum Ionising Particle (MIP) approximation approach.

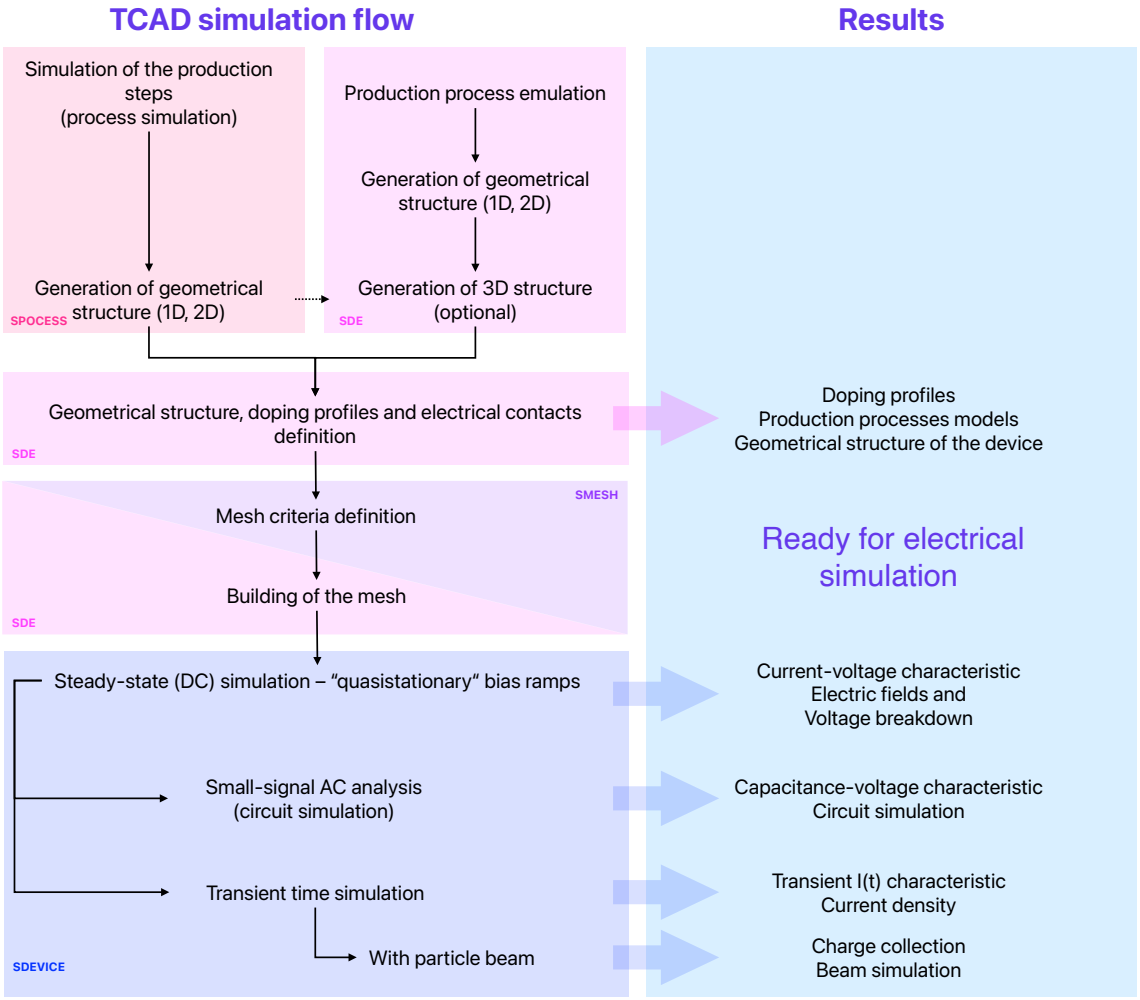


Figure 4.2: Diagram illustrating the typical TCAD simulation flow for silicon sensor design. The flow includes key steps such as process simulation, structure generation, meshing, and various simulation types (e.g., steady-state, small-signal AC, and transient simulations). Each step is associated with the specific TCAD tool responsible for the action, and potential results (e.g., doping profiles, current–voltage characteristics, and charge collection) are highlighted on the right.

4.1 Modelling of device geometry

The design of any semiconductor device begins with defining its structure. *Sentaurus Structure Editor* was chosen as the programming environment for this purpose. In principle, the process of structure definition follows these steps:

1. Set of the zero point and coordinate system.
2. Generate the model geometry.

3. Set of the doping regions.
4. Add constant, analytic, or externally generated doping profiles to the chosen regions.
5. Define contact regions.
6. Include external submeshes.
7. Create the mesh through the interface to the *Sentaurus Mesh* engines.
8. Define local refinements.
9. Run and export of the structure, for further use in *Sentaurus Device*.

Based on these steps, the code can be divided into five main sections: parameters, geometrical structure, doping, contacts, and mesh. Step 9, "Run the simulation and export the structure for further use in *Sentaurus Device*," is supported by the *Sentaurus Workbench* GUI. Although Synopsys provides comprehensive manuals covering a wide range of design tasks, some code specifications are addressed in the following sections.

4.1.1 Parameters

Parameterisation plays a crucial role in structure creation, enabling a flexible and adaptable design through generalised scripting. Both fixed and variable parameters can be used, allowing users to modify the structure dynamically. To use variable parameters, the complete project must be built within *Sentaurus Workbench*.

A practical example for parameter definition is:

```
(define pitch 200)
(define pitch @pitch@)
```

The first line defines a pixel pitch of 200 µm, where µm is the default metric unit in TCAD. The second line enables the assignment of different values later within *Sentaurus Workbench*, providing flexibility for parameter variation.

Beyond geometrical parameters, such as the pixel pitch in this example, other values can also be advantageous to parameterise. For instance, when describing a doping profile, its values can be set in a way that gives the opportunity to vary and promptly adjust the doping. The following code demonstrates this concept, where the doping penetration depth is defined as a variable parameter:

```
(define this-profile-peak-position 0)
(define this-profile-peak-value 4.65e+17)
(define this-profile-depth-value 1.18e+17)
(define this-profile-depth @this-profile-depth@)
```

These parameters can then be efficiently utilised for doping profile emulation, enabling precise control over the semiconductor structure.

4.1.2 Geometrical structure

In *Sentaurus Structure Editor*, geometric objects can be freely drawn. However, for most applications, it is necessary to explicitly define the coordinates of the objects to ensure the precise alignment of different regions. An example of defining the sensor bulk using this approach is shown below:

```
(sdegeo:create-rectangle (position 0 0 0) (position pitch thickness 0)
"Silicon" "bulk")
```

This code creates a rectangular shape with a starting point at (0 0 0) and an ending position at (pitch thickness 0). The length of the rectangle is determined by the parameter `pitch`, and its width is set by the parameter `thickness`. The material assigned to this shape is silicon, and "bulk" is the designated name for this specific simulation region.

Most semiconductor devices consist of multiple regions. One approach to constructing such a multi-regional structure is by stacking basic two-dimensional shapes. In this case, newly added regions may intersect with pre-existing ones. If an overlap occurs, a predefined scheme is used to resolve the intersection. Boolean operations can be applied to these regions to merge or modify overlapping shapes. For instance, when adding a new structure with different material properties to an area occupied by an existing structure, the old structure must be adapted accordingly. In this case, the overlapping section of the old structure is removed, and the boundary of the new structure replaces the original boundary.

This corresponds to the following command:

```
(sdegeo:set-default-boolean "ABA")
```

This method is particularly useful for creating complex structures that incorporate multiple geometric forms, facilitating a more flexible and precise device design.

4.1.3 Doping

In *Sentaurus Structure Editor*, doping can be defined using various methods, including constant profiles, analytical functions, Gaussian functions, error functions, or one-dimensional external profiles. This functionality is accessible via both the GUI and scripting. The general process of material doping involves defining the doping profile, specifying its key characteristics, such as concentration or dose, and applying it to the designated physical regions.

In this work, two primary types of doping profiles are used: constant and Gaussian. Constant doping profiles are suitable for bulk doping or the formation of uniform epitaxial layers, whereas Gaussian profiles are primarily used for ion implantation emulation.

The code below is an example of a constant doping profile with a concentration without a gradient.

```
(sdedr:define-constant-profile "def-dop-const-p" "BoronActiveConcentration"
doping-value-p-bulk)
(sdedr:define-constant-profile-region "place-dop-const-p" "def-dop-const-p"
"bulk")
```

This code consists of two lines, where the first one, with the command “define-constant-profile”, refers to the doping profile with non-gradient concentration, and the second one, with the command "define-constant-profile-region", refers to putting the pre-defined doping style and concentration into a fixed region without considering the situation of dopant diffusion.

Syntax of the command “define-constant-profile”:

- “def-dop-const-p” refers to the name of this specific doping profile, which is assigned manually.
- “BoronActiveConcentration” refers to the dopant itself, and here it is boron.
- "doping-value-p-bulk" (without quotation marks in the code) refers to this doping concentration. This value can be defined directly in the code line or during parameter definition.

Syntax of the command “define-constant-profile-region”:

- “place-dop-const-p” refers to the name of this action, which is user-chosen.
- “def-dop-const-p” refers to the name of this specific doping profile, the same one as used in line one, as a connecting element between two commands.
- “bulk” is the name of the geometrical region where the doping is placed and generated in advance.

To summarise, the command “define-constant-profile-region” creates the action called "place-dop-const-p" to apply the profile "def-dop-const-p", which is determined by the command “define-constant-profile-region”, into the geometric shape “bulk”, which is generated by the command "sdegeo:create-rectangle".

Unlike constant doping, analytical doping profiles require a reference window instead of being applied to the entire geometric body. The process is typically divided into two steps. The first step defines a reference window (in coding referred to as Ref/Eval window) to establish the baseline for lateral extent and peak depth positioning. The second step specifies the shape of the doping profile. Each analytical doping profile type requires different parameters and parameter combinations.

A Gaussian doping profile is characterised by the following minimum parameters:

- peak concentration C_{peak} in cm^{-3};
- peak position y_{peak} in µm;
- length $GLength_y$ in µm or standard deviation $stdDev_y$ in µm.

For a Gaussian function, the length represents the distance from the peak position to the point where the concentration decays by a factor of e^{-1} relative to the peak concentration. The relationship between length and standard deviation is given by:

$$GLength_y = 2 \times stdDev_y \tag{4.1}$$

Using these parameters, the Gaussian iis defined as:

$$g(y) = C_{peak} \times e^{-\left(\frac{y-y_{peak}}{GLength_y}\right)^2} \tag{4.2}$$

The general shape of the Gaussian function is shown in Figure 4.3.

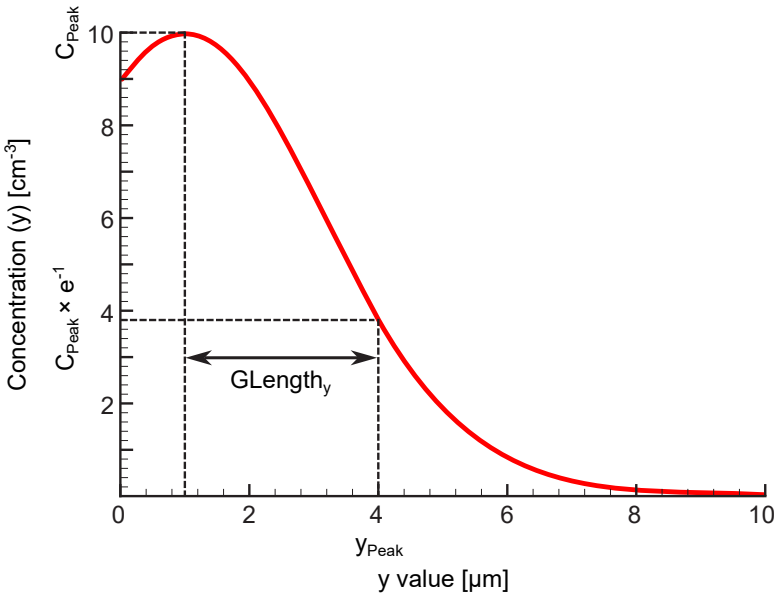


Figure 4.3: Illustration of the general shape of the Gaussian function, showing the peak concentration (C_{Peak}) and its exponential decay as a function of position (y). The function highlights key parameters, such as the peak value and its distribution along the y-axis [45].

In semiconductor fabrication, dose D is a critical parameter that quantifies the total implanted dopant concentration. While the dose is not explicitly included in standard Gaussian formulas, the peak concentration value can be derived from it. From a process simulation perspective, implantation profiles are determined based on dose concentration.

The general definition of the dose is:

$$D = \int_0^\infty g(y)\, dy \tag{4.3}$$

Using Equation 4.2, the dose for a Gaussian doping profile is given by [45]:

$$D = \int_0^\infty C_{peak} \times e^{-\left(\frac{y-y_{peak}}{GLength_y}\right)^2} dy \tag{4.4}$$

Solving this integral, the dose can be expressed as:

$$D = \frac{C_{peak} \times \sqrt{\pi} \times stdDev_y}{\sqrt{2}} \times \left(1 + erf\left[\frac{y_{peak}}{\sqrt{2} \times stdDev_y}\right]\right) \tag{4.5}$$

The Gaussian doping profile is implemented using the "define-gaussian-profile" command.

4.1.4 Contacts

Contacts are interfacial regions where TCAD tools, such as *Sentaurus Device*, apply electrical, thermal, or other boundary conditions. They correspond to the physical contacts present in the actual device. In two-dimensional simulations, contacts appear as edges, whereas in three-dimensional simulations, they are represented as faces. These contact edges and faces are integral to the device geometry [44].

An example of the code used to generate a contact is shown below:

```
(sdegeo:create-rectangle (position bf-r M1-bottom 0) (position bf-l M1-top
0) "Aluminum" "metal")
(sdegeo:create-rectangle (position bf-r M1-bottom 0) (position bf-r 0  0)
"Aluminum" "metal")
(sdegeo:create-rectangle (position bf-l M1-bottom 0) (position bf-l 0  0)
"Aluminum" "metal")

(sdegeo:define-contact-set "HV" 4.0 (color:rgb 0 1 0) "##")
(sdegeo:set-current-contact-set "HV")
(sdegeo:define-2d-contact (list (car (find-edge-id (position bf-r M1-top
0)))) "HV")
```

The code consists of two main parts. The first block defines a geometrical shape, which is later assigned as a contact in the second block. The "create-rectangle" command generates three regions, all named "metal," and filled with aluminium.

The second block consists of three commands that must be executed in the following sequence:

- "define-contact-set" defines a contact set named "HV," along with its assigned colour and pattern.
- "set-current-contact-set" sets "HV" as the active contact.
- "define-2d-contact" assigns the activated two-dimensional contact to a specified physical edge created earlier.

Figure 4.4 illustrates the generated aluminium contact using a TCAD visualisation tool.

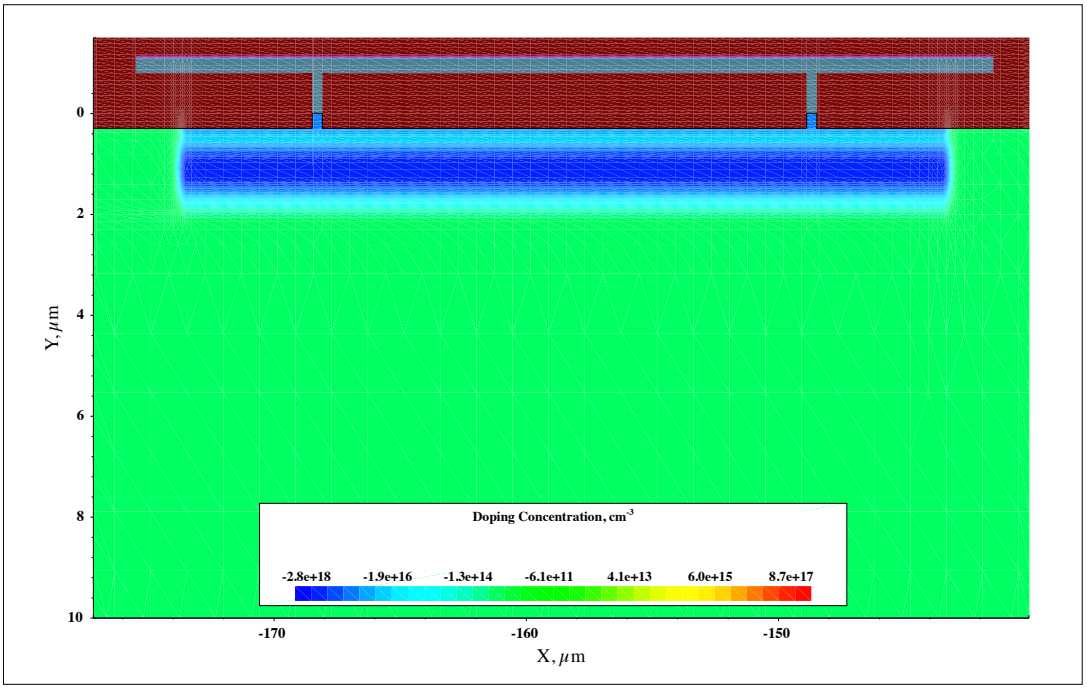


Figure 4.4: Visualisation of the generated aluminium contact "HV" using a TCAD tool. The grey region represents the aluminium metal contact, while the brown area corresponds to silicon dioxide. The main body, shown in green, represents the silicon substrate. The gradient in the silicon indicates the doping concentration, with values ranging from low to high, as represented by the scale on the right.

Placing contacts can be a challenging task in the development of CMOS structures. Modern integrated circuits typically contain multiple metal layers and independently biased contact pads. Figure 4.5 illustrates this concept.

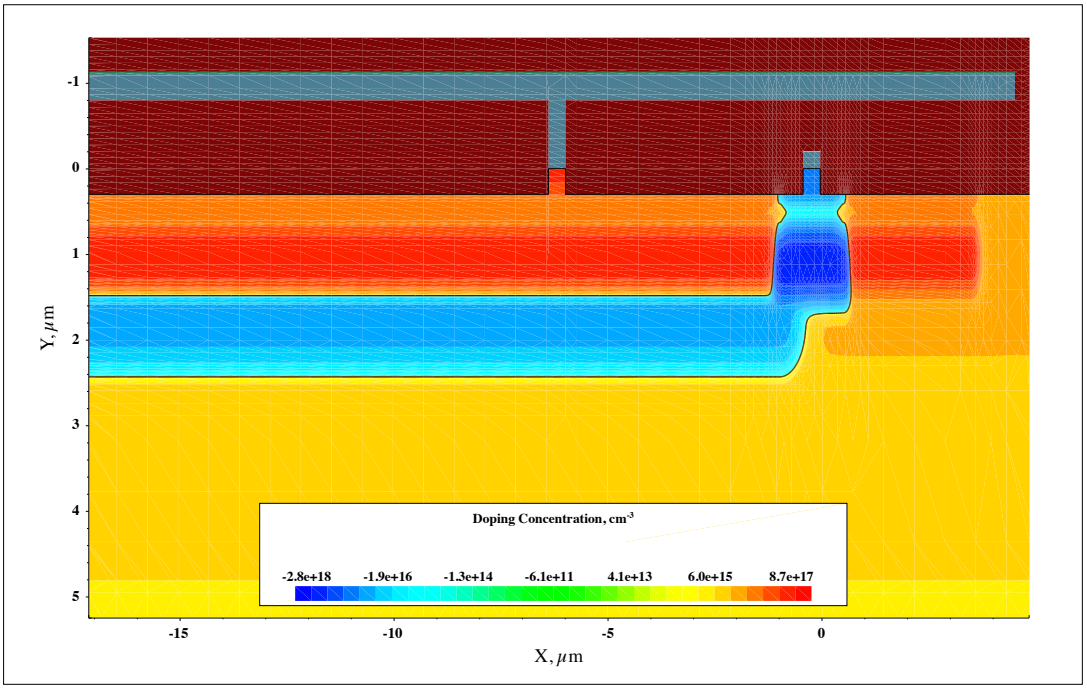


Figure 4.5: Illustration of two separate aluminium contact layers isolated by silicon dioxide, as commonly used in chip designs. The grey regions represent aluminium metal contacts, while the brown areas correspond to silicon dioxide. Each contact layer is independently biased, demonstrating the multi-layer structure typical in chip architectures. The doping concentration gradient in the silicon substrate is also shown, represented by the scale on the right.

4.1.5 Mesh

In addition to defining the geometry of a structure, doping profiles, and contacts, refinement parameters can be specified for different regions of the structure. This refinement information is required to control the mesh generation procedure according to user requirements. A correctly established mesh is vital for further mathematical calculations.

A mesh refinement box allows specific meshing constraints to be set. This typically involves two conditions for the selected area: the mesh spacing must not exceed a defined value, and refinement should stop if the mesh size becomes smaller than a given threshold. The meshing engine can also generate mesh spacings smaller than the requested global minimum to satisfy additional constraints, such as interface resolution, doping concentration variations, or specific doping regions. In these areas, achieving the highest possible mesh precision is crucial.

To summarise, the meshing strategy follows this approach: after defining the main mesh, denser meshing is applied to regions with significant concentration gradients, large electric field variations, or regions with a substantial impact on the device's electrical properties.

The code for defining the global mesh of silicon bulk is shown below:

```
(sedr:define-refinement-size "RefinementDefinition" 2 2 1e-2 1e-2)
```

```
(sdedr:define-refinement-material "Refinement-Silicon"
"RefinementDefinition" "Silicon")
(sdedr:define-refinement-function "RefinementDefinition"
"DopingConcentration" "MaxTransDiff" 0.5)

(sde:build-mesh "snmesh" " " "n@node@-msh")
```

The first step is to use the command "define-refinement-size" to create a refinement definition, where:

- "RefinementDefinition" is a unique name of this mesh, which must be used in every command assigned to this meshing.
- The number sequence "2 2 1e-2 1e-2" corresponds to "dx_{max}, dy_{max} dx_{min}, dy_{min}", which are minimum and maximum required mach spacing in x- and y-axis.

The *Sentaurus Structure Editor* can also obtain the lateral dimension of the refinement box from the specified area, taking the vertical refinement box dimensions from the argument list. For this, commands to define refinement by reference doping window "define-refinement-window", by assigning mesh into specific placement "define-refinement-placement", region "define-refinement-region" can be used.

In the example above, the mesh is assigned to a particular material, requiring the use of the "define-refinement-material" command. Its syntax is as follows:

- "Refinement-Silicon" is the name of the refinement.
- "RefinementDefinition" is the name of the action, connecting the meshing procedure together.
- "Silicon" refers to the material of choice for this mesh.

Figure 4.6 shows that doping concentration in the silicon bulk is not homogeneous due to additional doping regions needed to form pn junctions and other electronic elements. To account for this, denser meshing must be assigned to regions with a high concentration gradient. The "define-refinement-function" command enables additional meshing constraints in a pre-existing refinement area. The syntax of the command is the following:

- "RefinementDefinition" is the name of the action which connects the meshing procedures. Here, additional meshing is added to the existing refinement named "RefinementDefinition", to which a global meshing was already applied.
- "DopingConcentration" refers to the variable of interest. In this example, it is a doping concentration. Still, the same method can also be used for the other quantities, such as changes in the electric field "ElectricField" or potential gradients "ElectrostaticPotential".
- "MaxTransDiff" is a Refine Function, an abbreviation for Maximum Transformed Difference. It evaluates the maximum difference of the transformed values of a variable at the point of the element. If the difference is greater than "Value", which is in this example 0.5 μm, then the element is refined.

As the final step in the meshing process, the command "build-mesh" generates the boundary output and mesh command file while invoking *Sentaurus Mesh*. The result of the entire mesh of silicon material can be seen in Figure 4.7.

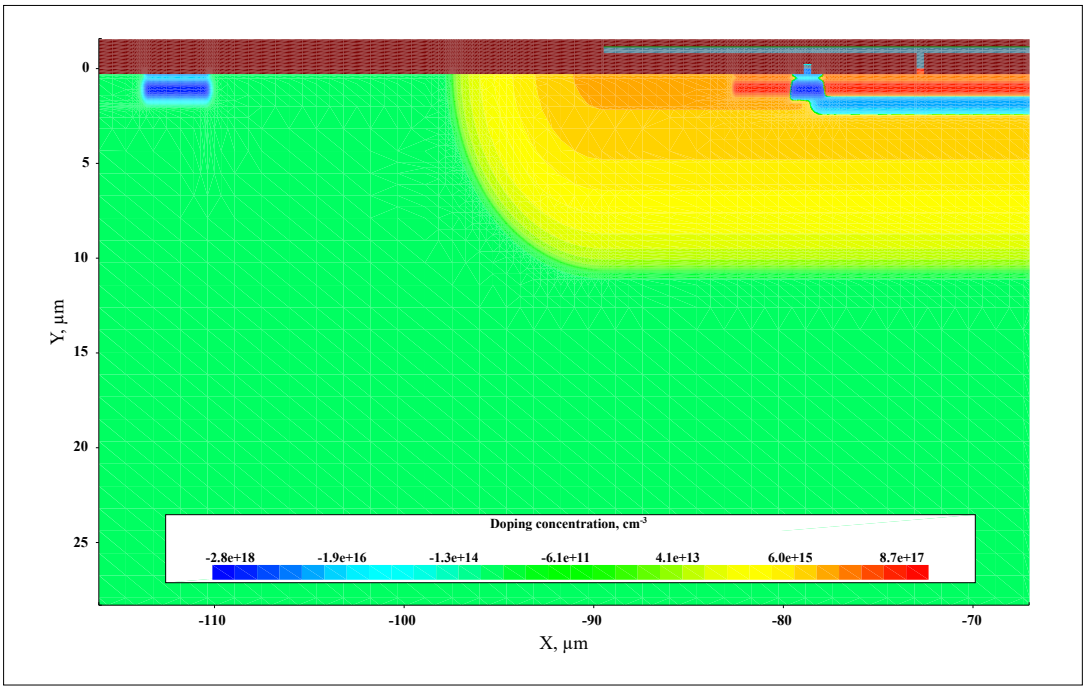


Figure 4.6: Zoomed-in view of the HV-CMOS chip's TCAD simulation, showing various doping regions with distinct concentrations. The non-uniform doping profile is evident, highlighting areas needed for forming *pn* junctions and electronic components. The visualisation excludes the mesh for clarity, focusing on the doping concentration gradient, as indicated by the scale on the right.

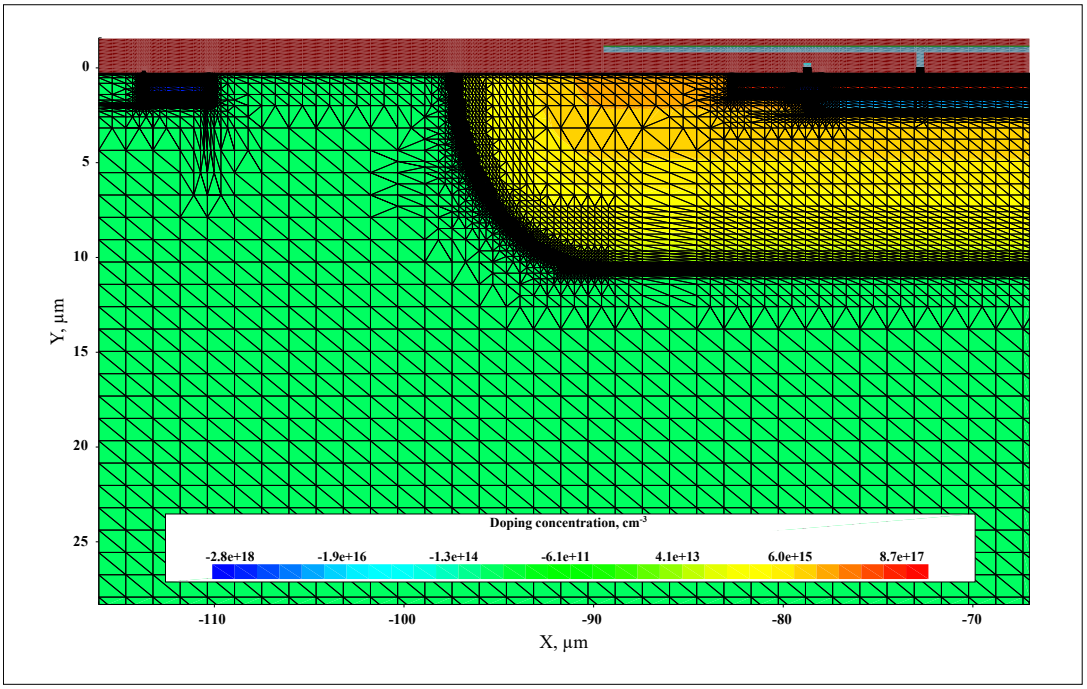


Figure 4.7: Zoomed-in view of the HV-CMOS chip's TCAD simulation, showing distinct doping regions alongside the simulation mesh. The mesh structure highlights the discretisation used for numerical analysis, overlaid on the non-uniform doping profile. This visualisation combines the doping concentration gradient with the mesh framework to provide greater clarity in the simulation process.

4.2 Numerical simulations and physical models

This section focuses on the next stage of detector design, which is numerical semiconductor device simulation using *Sentaurus Device*. Following the device geometry modelling outlined in Section 4.1, a virtual device structure has been generated and stored in two files in TDR format. These files consist of a boundary file (`bnd.tdr`) and a mesh file (`msh.tdr`), which provide information about the device grid. The device grid defines various regions, including boundaries, material types, and the locations of electrical contacts. It also contains the positions of all discrete nodes and their connectivity. Additionally, the files include information about device properties, such as doping profiles, which are stored as data associated with the discrete nodes [42].

The *Sentaurus Device* command file is organised into sections, each containing commands to solve computational problems or keyword statements to define device physics. *Sentaurus Device* keywords are not case-sensitive, and most of them can be abbreviated. However, the syntax is strict, requiring careful adherence to format rules. A typical *Sentaurus Device* command file consists of multiple sections, each performing a distinct function. A standard single-device command file includes mesh definitions, contact specifications, physical models, and solver commands. These sections can appear in any order except in mixed-mode simulations, where specific constraints apply.

Mixed-mode simulations enable the simultaneous solution of not only single-device problems but also configurations involving a combination of devices with circuits or multiple devices interconnected within a circuit. This type of simulation is essential for modelling monolithic sensors and considering the specific requirements of CMOS designs. In TCAD, a CMOS device is described as a silicon bulk with multiple doping regions connected in a specific circuit. For this reason, mixed-mode simulations are extensively used in this work.

Another type of simulation that requires mixed-mode functionality is AC simulation. AC simulation is based on small-signal AC analysis, where the device response to a small sinusoidal signal superimposed on a pre-established DC bias is computed as a function of frequency and the DC operating point [42]. In this approach, a small signal stimulus is applied to the device contacts, enabling the extraction of capacitance at different frequencies.

The command file for the mixed-mode simulation of the HitPix chip (see Chapter 6) consists of several sections, each serving a specific function. These sections are described below, with each explained in detail in the subsequent sections.

1. Device section – defines the simulated devices; this section is specific to mixed-mode simulations.
2. File section – specifies input and output files.
3. Electrode section – defines electrodes, ensuring consistency with those specified in the input grid.
4. Physics section – selects the physical models applied in the device simulation.
5. Plot section – lists all solution variables saved in the output plot files.
6. System section – defines the circuit configuration.
7. Math section – specifies numerical solvers for solving the device equations.
8. Solve section – defines the sequence of solutions obtained by the solver.

4.2.1 Device section

The Device section of the command file defines each physical device included in the simulation. For a multi-device simulation, a command file can contain multiple Device sections. If multiple devices are specified, the command file must include separate definitions for the mesh (File section), contacts (Electrode section), and physical models (Physics section) for each device.

An example of a Device section is shown below:

```
Device HV-CMOS
{
    File {}
    Electrode {}
    Physics {}
}
```

Each device must be assigned a unique identifier following the keyword Device. For example, in this case, the device is named HV-CMOS . For the HitPix chip, a single Device section was used. The Device section is closely linked to the System section (see subsection 4.2.6). While the Device section defines individual devices, the System section is necessary for creating and interconnecting multiple devices within the simulation.

4.2.2 File section

The File section can be divided into two parts: local, which is related to each individual device, and global, which is associated with the overall circuit. Together, they specify the input files that define the device structure and the output files that store the simulation results.

The code for the local File section is shown below. This code is included in the Device section 4.2.1.

```
File
{
    Grid = "@tdr@"
    Current = "@plot@"
    Plot = "@dat@"
}
```

The command **`Grid = "@tdr@"`** is used to read the input file. It automatically retrieves the `msh.tdr` and `bnd.tdr` files generated by *Sentaurus Structure Editor* These essential input files define the mesh and various regions of the device structure, including the contacts. They also contain doping profile data for the device.

The command **`Current = "@plot@"`** generates an output file containing electrical characteristics such as currents, voltages, and charges at the electrodes. The command **`Plot = "@dat@"`**creates a file where updated spatial variables for the structure mesh are stored. This includes information such as electric fields, potentials, and carrier concentrations at every point in the mesh.

Since this local File section is part of the Device section, it includes only device-specific files. However, it cannot contain output files that store the results of circuit simulations, such as AC simulations. These extracted AC results are managed separately in a global File section, which is defined outside the Device statement and has the following syntax:

```
File
{
    Output = "@log@"
    ACExtract = "@acplot@"
}
```

The command **`Output = "@log@"`** creates an output log file, which is automatically generated by *Sentaurus Device* during execution. The command **`ACExtract = "@acplot@"`** is responsible for saving the computed small-signal AC components of the circuit into a file.

In all these commands, the names of the read or generated files are not explicitly specified. *Sentaurus Device* automatically appends the appropriate file name extensions, maintaining compatibility between different Sentaurus tools. However, careful attention must be paid to the file format to ensure seamless integration with other simulation tools. This information is always specified as follows: **"@format@"**.

4.2.3 Electrode section

As a part of the geometry design, the contacts' regions were generated in the *Sentaurus Structure Editor* (see Subsection 4.1.4). Specifying which of these contacts is to be treated as electrodes is necessary for the next simulation steps. The Electrode section defines all the electrodes and their types, which will be used in the *Sentaurus Device* simulation, with their respective boundary conditions and initial biases. Any contacts not defined as electrodes are ignored by *Sentaurus Device*.

Below is an example of the Electrode section used for the HitPix simulation. In this example, all electrodes are ohmic contacts with voltage boundary conditions and an initial zero bias.

```
Electrode
{
    {Name="HV" Voltage=0}
    {Name="NW-INSIDE-PI" Voltage=0}
    {Name="DN-OUTSIDE-PI" Voltage=0}
    {Name="PI" Voltage=0}
}
```

Each electrode is specified by a case-sensitive name **Name="string"** that must exactly match an existing contact name in the structure file. Only the contacts listed in the Electrode section are included in the simulation. Names of the electrodes, which correspond to the names of the metal placements from *Sentaurus Structure Editor*, have the following meaning:

- **"HV"** represents the substrate bias.
- **"NW-INSIDE-PI"** is the electrode of the sensor diode.
- **"DN-OUTSIDE-PI"** corresponds to the electrodes connected to electronics outside the sensor diode.
- **"PI"** is the contact to the isolation *p*-doped region.

The command **Voltage=0** defines a voltage boundary condition with an initial value. One or more boundary conditions must be defined for each electrode, and any value assigned to a boundary condition applies in the initial solution. Other than voltage boundary conditions, the Electrode section supports the definition of current boundary condition, of a floating electrode with a charge boundary condition and an initial charge value, the definition of an electrode as a Schottky contact, establishing the metal-semiconductor work function difference, and definition as a series resistance.

4.2.4 Physics section

The Physics section allows users to select the physical models that will be applied in the device simulation. In the *Sentaurus Device* User Manual [42], two sections (Part II and Part III) are dedicated to descriptions of physical models and the underlying physical laws. These models can be specified either globally for the entire device or individually for specific regions, interfaces, materials, or electrodes.

For the HitPix simulation, the Physics section was defined for the entire sensor and written within the Device section. The code is presented below, where each line is annotated to clarify its function, to link it to the corresponding model from the User Manual [42], and to relate it to the fundamental semiconductor theories discussed in Chapter 2. The selected models were adapted to accurately describe a silicon particle sensor and will serve as the foundation for many simulations conducted in this thesis. Additionally, a specialised Physics section is used to simulate the interaction of a heavy-ion beam and its effects (see Section 4.3).

```
Physics
{
    areafactor = 170
    Temperature = 293
    Fermi
    EffectiveIntrinsicDensity (BandGapNarrowing (OldSlotboom))
    eQCvanDort
    Mobility
        (
        DopingDependence
        eHighFieldSaturation
        hHighFieldSaturation
        Enormal
        CarrierCarrierScattering (ConwellWeisskopf)
        )
    Recombination
        (
        SRH
            (
            DopingDependence
            TempDependence
            ElectricField (Lifetime=Hurkx DensityCorrection=None)
            )
        CDL
        Auger
        eAvalanche (vanOverstraeten Eparallel)
        hAvalanche (vanOverstraeten Eparallel)
        )
}
```

AreaFactor specifies a multiplier for currents and charges. In a two-dimensional simulation, this is achieved by defining the width of the third dimension. By default, this value is set to 1 µm. When **AreaFactor** is specified, currents are multiplied by its value, which in this example is 170 µm. This number is chosen based on the geometry of the actual device.

Temperature specifies the lattice temperature, which is set to 293 *K*.

Moving to the description of semiconductor properties, the first topic to be discussed is the thermal equilibrium state. This refers to a steady-state condition at a given temperature, without any external excitations such as light, pressure, or an electric field. In this state, parameters such as carrier concentration and electron density are determined. **Fermi** activates Fermi statistics, which is essential for accurately modelling high carrier densities in the active regions of a silicon device (see Subsection 2.1.2).

The other fundamental property of a semiconductor is its band structure. Realistic band structures are complex and can be challenging to simulate. In the *Sentaurus Device*, the band structure is simplified to four quantities: the energies of the conduction and valence band edges, and the density-of-states masses for electrons and holes [42]. This simulation approximation has to be activated by the command **EffectiveIntrinsicDensity**.

The phenomenon of bandgap narrowing in heavily doped silicon has been extensively studied both theoretically and experimentally. Theoretical studies indicate that at high impurity concentrations, the density of energy states no longer follows a parabolic energy distribution but instead becomes dependent on impurity concentration. The bandgap is effectively reduced due to the broadening of the impurity band and the formation of band tails at the conduction and valence band edges. This phenomenon is discussed in detail in the work of Slotboom [48]. Bandgap narrowing in silicon has also been characterised in the base region of bipolar transistors, as described in Slotboom's subsequent work [49]. At impurity concentrations greater than approximately 1×10^{17} cm^{-3}, this effect becomes significant. Bandgap narrowing strongly influences the electrical behaviour of silicon devices, particularly affecting minority carrier charge storage and minority carrier current flow in heavily doped regions.

Sentaurus Device supports different bandgap narrowing models, which can be selected using the **BandGapNarrowing** option in the **EffectiveIntrinsicDensity** statement. This simulation employs the **OldSlotboom** model. Bandgap narrowing, ΔE_g^0, for the Slotboom model in *Sentaurus Device*, is defined by the following equation:

$$\Delta E_g^0 = E_{ref} \times \left[\ln\left(\frac{N_{tot}}{N_{ref}}\right) + \sqrt{\left(\ln\left(\frac{N_{tot}}{N_{ref}}\right)\right)^2 + 0.5} \right] \tag{4.6}$$

Here, N_{tot} represents the total doping concentration, including both acceptors and donors. The parameters N_{ref} and E_{ref} are default values for the analytic bandgap narrowing models available in *Sentaurus Device*, with values of 1×10^{17} cm^{-3} and 9×10^{-3} eV, respectively.

Due to the modern design features of MOSFETs, such as the nanometre-scale thickness of the oxide layer and the narrow width of the channel, the quantum-mechanical nature of electrons and holes cannot be neglected. This effect is incorporated in *Sentaurus Device* using the

eQCvanDort option, which stands for the van Dort quantisation model. The most fundamental quantisation effects in MOSFETs include the threshold voltage shift and the reduction of gate capacitance. To account for quantisation effects in a classical device simulation, *Sentaurus Device* introduces a potential-like quantity, Λ, in the classical density formula (see Subsection 2.1.2). For electrons, this is given by:

$$n = N_C \times e^{-\left(\frac{(E_C+\Lambda)-E_F}{k\times T}\right)} \tag{4.7}$$

An analogous quantity is introduced for holes.

Sentaurus Device implements different quantisation models, each of which provides a specific value for Λ. These models vary in terms of physical sophistication, numerical complexity, and robustness. The van Dort model is a numerically robust, computationally efficient, and well-established approach. It is widely used in bulk MOSFET simulations and calculates Λ as a function of the electric field normalised to the semiconductor–insulator interface. This quantum correction is applied to carriers that are attracted towards the interface by the electric field, while for carriers repelled from the interface, the correction is zero.

Mobility is one of the main characteristics of the semiconductor when it comes to carrier transport. More details about the underlying physics can be found in Subsection 2.1.3. *Sentaurus Device* implements multiple mobility models, where, in the simplest case, mobility is treated as a function of the lattice temperature. For undoped materials, the constant mobility model accounts for phonon scattering. In doped materials, where various scattering mechanisms contribute to carrier transport, multiple degradation dependencies can be applied.

For MOSFET-type devices operating under high bias voltage, the following parameters are particularly important:

- Doping-dependent mobility degradation, activated by the option **DopingDependence**, covers carrier mobility degradation with increasing dopant concentration. The default model used by *Sentaurus Device* to simulate doping-dependent mobility in silicon was proposed by Masetti [50]. This model provides empirical relationships that accurately describe the experimentally observed carrier mobility in As-, P-, and B-doped silicon across a wide range of carrier concentrations.
- Carrier–carrier scattering, activated by the option **CarrierCarrierScattering**, accounts for carrier mobility degradation due to the carriers' scattering by charged impurity ions. The default model used by *Sentaurus Device* for carrier–carrier scattering is the Conwell–Weisskopf model [51], with Fletcher transposition into mobility dependence for the high level of injected carriers [52].

  Conwell and Weisskopf's work demonstrated that lattice scattering alone cannot fully explain the temperature dependence of resistivity. To address this, they investigated and formalised the impact of scattering by ionised impurity centres [51]. Building upon this theory, Fletcher examined the effect of high carrier densities on carrier lifetime and mobility at very high operating levels. This means that the density of carriers injected into the body of a semiconductor junction device is comparable to the carrier density in

the device's emitter regions. As a result, hole-electron collisions were found to have a significant impact at high carrier densities. Although secondary in determining overall device characteristics, this effect must be considered for accurate calculations.

The formula for carrier–carrier scattering using the Conwell–Weisskopf model is given as [42]:

$$\mu_{eh} = \frac{D \times \frac{T}{300\ K}^{\frac{3}{2}}}{\sqrt{n \times p}} \times \left[\ln\left(1 + F\left(\frac{T}{300\ K}\right)^2 \times (n \times p)^{-\frac{1}{3}}\right)\right]^{-1} \tag{4.8}$$

Here, F and D are empirical fitting parameters, which for silicon are given as:

- $F = 7.452 \times 10^{13}\ \mathrm{cm}^{-2}$
- $D = 1.04 \times 10^{21}\ \mathrm{cm}^{-1}\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$

- Mobility degradation at interfaces is a critical factor in the channel region of a MOSFET. The high transverse electric field, $F_\perp$, forces carriers to interact strongly with the semiconductor–insulator interface. These carriers experience scattering due to acoustic surface phonons and surface roughness, leading to mobility degradation. To account for this degradation at interfaces, the transverse electric field, $F_\perp$, must be computed. By enabling the **`Enormal`** option within **`Mobility`**, the calculation of the field perpendicular to the semiconductor–insulator interface is activated.

  Assuming mobility degradation occurs at a semiconductor–insulator interface Γ, for a given point $\vec{r}$, *Sentaurus Device* locates the nearest vertex $\vec{r_i}$ on the interface Γ. It approximates the distance of $\vec{r}$ to the interface by the distance to $\vec{r_i}$, and determines the direction $\hat{n}$ normal to the interface at vertex $\vec{r_i}$. From $\hat{n}$, the field perpendicular to the semiconductor–insulator interface is computed as:

$$F_\perp(\vec{r}) = |\,\vec{F}(\vec{r}) \times \hat{n}| \tag{4.9}$$

  The mobility degradation models are sensitive to mesh spacing. It is recommended that the vertical mesh spacing be reduced to 0.1 nm in the silicon at the oxide interface to ensure accurate modelling.

- In high electric fields, the carrier drift velocity is no longer proportional to the electric field; instead, the velocity saturates to a finite speed v_{sat} (see Subsection 2.1.3). High electric field saturation is activated by specifying the **`eHighFieldSaturation`** and **`hHighFieldSaturation`** options in **`Mobility`**.

  The high-field saturation model, specifically the Canali model [53], consists of three submodels: the actual mobility model, the velocity saturation model, and the driving force model. In this work, the drift velocity of electrons and holes in silicon was measured across a wide range of electric fields at temperatures up to 430 K. The experimental data

were fitted with temperature-dependent formulas. According to the Canali model, the carrier mobility for electrons and holes, $\mu(F)$, in a high electric field is given by:

$$\mu(F) = \frac{(\alpha + 1) \times \mu_{low}}{\alpha \left[1 + \left(\frac{(\alpha+1) \times \mu_{low} \times F_{hfs}}{v_{sat}}\right)^{\beta}\right]^{\frac{1}{\beta}}} \tag{4.10}$$

The exponent β is a temperature-dependent coefficient, calculated as:

$$\beta = \beta_0 \times \left(\frac{T}{300\ K}\right)^{\beta_{exp}} \tag{4.11}$$

Several parameters used in the Canali model mobility definition are unique to the simulated material. The default values for silicon are [42]:

$$\beta_{0,e} = 1.109 \qquad \beta_{exp,e} = 0.66 \qquad \alpha_e = 0 \tag{4.12}$$

$$\beta_{0,h} = 1.213 \qquad \beta_{exp,h} = 0.17 \qquad \alpha_h = 0 \tag{4.13}$$

The value of μ_{low} is defined by the low-field mobility, which is determined by the constant mobility model (see Subsection 2.1.3). The value of v_{sat} is given by the following formula:

$$v_{sat} = v_{sat,0} \times \left(\frac{300\ K}{T}\right)^{v_{sat,exp}} \tag{4.14}$$

This model is recommended for silicon but is not suitable for other semiconductor materials. The velocity saturation parameters are defined as follows [42]:

$$v_{sat,0,e} = 1.07 \times 10^7\ cm/s \qquad v_{sat,exp,e} = 0.87 \tag{4.15}$$

$$v_{sat,0,h} = 8.37 \times 10^6\ cm/s \qquad v_{sat,exp,h} = 0.52 \tag{4.16}$$

Sentaurus Device supports five different models for the driving force F_{hfs}. For drift-diffusion simulations, the gradient of the quasi-Fermi potential is chosen as the most appropriate model. This method is particularly advantageous in regions where a small current may be present. In this case, the electric field (for electrons and holes, respectively) replaces the gradient of the quasi-Fermi potential within mesh elements adjacent to a contact:

$$F_{hfs,n} = |\vec{F}| \tag{4.17}$$

Sentaurus Device employs a modular approach to describe carrier mobilities. If multiple mobility models are activated for a given carrier type, the different mobility contributions

(μ_1, μ_2, μ_3, etc.), including bulk, surface mobility, and thin-layer effects, are combined using Matthiessen's rule [42]:

$$\frac{1}{\mu} = \frac{1}{\mu_1} + \frac{1}{\mu_2} + \frac{1}{\mu_3} + \ldots \tag{4.18}$$

The next step is to define the recombination and generation models. In *Sentaurus Device*, the command **`Recombination`** is used for this purpose. The primary recombination mechanism occurs through deep defect levels in the bandgap, known as Shockley–Read–Hall (**`SRH`**) recombination. More details on the underlying physics can be found in Subsection 2.1.3.

Sentaurus Device provides multiple dependencies for this model to enhance accuracy. Additional factors, such as doping concentration, temperature, and high electric fields, can influence carrier lifetime. For the simulation of a particle detector in CMOS technology, the following options were selected:

- SRH doping dependence, activated by the **`DopingDependence`** option, describes the minority-carrier lifetime $\tau(N)$ as a function of impurity concentration. This dependence is derived from experimental data and can be expressed as [54]:

$$\tau(N) = \frac{\tau_0}{1 + \left(\frac{N}{N_0}\right)} \tag{4.19}$$

  Here, $\tau(N)$ represents the doping-dependent minority-carrier lifetime, which is defined separately for electrons in $p-$type silicon and for holes in $n-$type silicon. N is the concentration of acceptors and donors, respectively. The parameters τ_0 and N_0 are empirical coefficients, determined experimentally and adjustable to specific simulation conditions [54, 42]. In this work, the default values were used:

$$\tau_0 = 1 \times 10^{-6}\, s \qquad N_0 = 1 \times 10^{16}\, cm^{-3} \tag{4.20}$$

- SRH temperature dependence, activated by the **`TempDependence`** option, describes the variation of minority-carrier lifetime $\tau(T)$ as a function of temperature. Although this relationship is empirical, studies have shown that carrier lifetime increases with rising temperature. This dependence was initially described and tested by Van Overstraeten [55] and later modelled by Goebel and Hoffmann [56]. To analytically describe the temperature dependence, the following relation is used:

$$\tau(T) = \tau_{300\ K} \times \left(\frac{T}{300\ K}\right)^{\alpha} \tag{4.21}$$

  The parameters for this equation are:

$$\tau_{300\ K} = 180\ \text{ns} \qquad \text{and} \qquad \alpha = 2.1 \tag{4.22}$$

- SRH field enhancement, activated by the **`ElectricField`** option, describes the dependence of recombination lifetimes in regions with strong electric fields. In areas where the electric field exceeds approximately 3×10^5 V/cm, the recombination lifetime decreases. This effect plays a significant role in, for example, the IV characteristics of reverse-biased $p - n$ junctions, which are highly sensitive to defect-assisted tunnelling. It causes electron-hole pair generation before band-to-band tunnelling or avalanche generation sets in. Therefore, including field enhancement in the simulation is recommended for analysing drain reverse leakage and substrate currents in MOS transistors. *Sentaurus Device* provides several models for trap-assisted tunnelling, which define **`Lifetime`** and **`DensityCorrection`** parameters. In this work, the Hurkx trap-assisted tunnelling model for silicon is chosen. This model is based on a complex theory that derives an analytical expression for the field and temperature dependence of Shockley–Read–Hall lifetimes in silicon. Further details on this model can be found in [57].

Recombination in semiconductors is typically well described by the Shockley–Read–Hall (SRH) model, in which electrons and holes are captured by a single defect level. Even when multiple defect levels are considered, they are usually treated as independent, except for the indirect statistical link via band occupation. However, experimental studies have demonstrated direct charge transfer between deep defect levels in silicon. For instance, in n–type silicon, the shallow hydrogenic level of the phosphorus donor acts as an efficient recombination channel due to its coupling with deep levels associated with the vacancy–oxygen (V–O) complex. This charge transfer process has been observed to occur on a timescale shorter than 50 ns. Coupled Defect Level (CDL) recombination, activated by the **`CDL`** option, accounts for the steady-state recombination rate between two coupled defect levels. An important characteristic of this model is an enhanced field effect, which can lead to significantly increased excess currents. Further details on this model can be found in [58].

The second most important recombination model after SRH is radiative Auger recombination, activated by the **`Auger`** option. This is a phonon-assisted band-to-band recombination process that plays a significant role in regions with high doping concentrations. A more detailed description of Auger recombination is provided in Subsection 2.1.4.

The generation of electron-hole pairs due to impact ionisation requires a certain threshold energy and the presence of a sufficiently wide space charge region to allow carrier acceleration. If the width of the space charge region exceeds the mean free path between two ionising impacts, charge multiplication occurs, potentially leading to electrical breakdown. Avalanche multiplication due to impact ionisation is the primary mechanism responsible for the breakdown of p-n diodes. It limits the maximum reverse voltage for most diodes, as well as the collector voltage in bipolar transistors and the drain voltage in field-effect transistors. This multiplication process, known as avalanche generation, is activated by using the keyword **`Avalanche`** in the **`Recombination`** section. The keywords **`eAvalanche`** and **`hAvalanche`** specify separate models or driving forces for the electron and hole ionisation coefficients, respectively [42]. In this work, the selected model is the default van Overstraeten-de Man model for the electron impact ionisation process, where the driving force is defined as a uniform electric field. Further details on this model can be found in [59, 60].

As mentioned earlier, in *Sentaurus Device*, various physical models can be specified for different regions and materials within a device structure. This can be achieved by defining an additional Physics section. For example:

```
Physics (MaterialInterface="Silicon/SiO2")
{
    Recombination (SurfaceSRH)
    Traps (FixedCharge Conc=1e12)
}
```

In this case, the models are applied to all interfaces between silicon and adjacent oxide regions.

In addition to generation-recombination in the bulk of a semiconductor, electrons and holes can also be generated and recombined at surfaces. The command **`Recombination(SurfaceSRH)`** activates the surface component of Shockley–Read–Hall recombination. Under certain conditions, the rate of surface generation-recombination can be significantly higher than that of bulk generation-recombination. In such cases, an additional formula is used, which is structurally equivalent to the bulk expression of SRH generation-recombination but applied to the interface of interest:

$$R_{surf}^{SRH} = \frac{n \times p - n^2}{\frac{1}{s_p} \times \left(n + n_{surf}\right) + \frac{1}{s_n} \times \left(p + p_{surf}\right)} \tag{4.23}$$

Here, s_p and s_n represent the surface recombination velocities, while n_1 and p_1 are the carrier concentrations at the surface [61]. It is important to note that not all models can be defined within interface-specific Physics sections. For instance, mobility models and bandgap narrowing cannot be specified at interfaces.

Another command, **`Traps(FixedCharge)`**, provides information about fixed charges at the insulator interface. In this case, the parameter **`Conc`** specifies the maximum surface charge concentration [42]. Further details on trap models can be found in Subsection 2.2.2. Additionally, the radiation damage model is discussed as part of the heavy ion beam simulation (see Chapter 4.3).

4.2.5 Plot section

After defining the physics models for the device in the Physics section (see Subsection 4.2.4), *Sentaurus Device* requires a specification for device plots. Plot outputs enable the visualisation of internal device quantities, providing insights that are otherwise inaccessible in real experiments. The Plot section defines the data that is saved at the end of the simulation. These results are recorded in the plot file, specified by the command **`Plot = "@dat@"`**, which is defined in the File section (see Subsection 4.2.2).

It is important to note that only data that *Sentaurus Device* can compute—based on the selected physics models—is saved to the plot file. A comprehensive list of available plotting commands for each physics model can be found in Appendix F of the *Sentaurus Device* User Manual [42].

The following example illustrates a typical Plot section, specifying output results for fundamental semiconductor properties:

```
Plot
{
    SpaceCharge
    Potential
    EffectiveIntrinsicDensity IntrinsicDensity eDensity hDensity
    eMobility hMobility
    eVelocity hVelocity
}
```

Not all required datasets for plotting can be defined within the Plot section. Some data, which can only be determined at specific steps of the simulation, must be specified using the **`Plot`** function within the Solve section. This function and its usage will be described in detail in Subsection 4.2.8.

4.2.6 System section

At this stage, the Device section (Subsection 4.2.1), which includes the electrode definition (Subsection 4.2.3) and the complete physics description (Subsection 4.2.4), is fully defined, resulting in a complete sensor model. However, for monolithic sensors, simply biasing the bulk is insufficient. This type of device incorporates circuit transistors within the main bulk, necessitating multiple biasing configurations. As previously mentioned, simulating a system that includes both a device and a circuit is only possible in mixed-mode operation. Therefore, after defining the physical device in the Device section, the System section is required to establish the electrical circuit connections.

The System section employs the SPICE "Simulation Program with Integrated Circuit Emphasis" syntax as its base simulation tool. SPICE is a general-purpose, open-source analogue electronic circuit simulator developed at the Electronics Research Laboratory of the University of California, Berkeley. It is widely used in integrated circuits and board-level design to verify circuit integrity and predict circuit behaviour [62].

The code in the System section defines the netlist, which consists of physical devices, electronic components in a circuit, and the nodes to which they are connected. At this stage of the simulation, the previously designed device is electrically integrated with other components in a circuit. These components can include elementary elements such as resistors and capacitors, voltage and current sources, or predefined compact device models, such as transistors and diodes. These predefined models, known as compact models, are available in SPICE syntax. Utilising the SPICE device definition enables the selection of the required component while only specifying its parameters and connections. By default, and in this design, a circuit node

is electrical; however, it can also be declared as thermal. Each electrical node is associated with a voltage variable.

The system description consists of two steps: first, defining the main physical device, and second, specifying the additional circuit components. Figure 4.8 schematically illustrates the connection between the device and the circuit. The lower part of the figure depicts the *p*- and *n*-doped regions, shown in blue and red, respectively. Each of these regions is assigned a contact name, which is first introduced in the geometry file (see Subsection 4.1.4) and corresponds to the electrode names in the Electrode section (see Subsection 4.2.3). Each contact is associated with a node name, which is connected to a voltage source through a resistor.

The code template from *Sentaurus Device* User Manual [42], which is used to define a physical device and the example from HitPix chip simulation are presented below:

```
System
{
    device-type instance-name (connectivity list) {<attributes>}

    HV-CMOS "sensor" ("HV"=bias "PI"=pi "NW-INSIDE-PI"=nw
    "DN-OUTSIDE-PI"=dn)
}
```

This command line follows the syntax described below:

- **`device-type`** – HV-CMOS – defines the model used, which must match the identifier name specified in the Device section to establish a connection between code sections;
- **`instance-name`** – "sensor" – defines the name assigned to this model, which will be referenced in the Solve section (see Subsection 4.2.8) and serves as an inter-section connecting variable;
- **`(connectivity list)`** – a list of ‘"contact name" = "node name"‘ connections. The contact name refers to the contact defined in the device grid file (see Subsection 4.1.4) and corresponds to the electrode name in the Electrode section (see Subsection 4.2.3). The node name represents the circuit netlist node, which is necessary to establish an electrical connection and define a circuit element. For example, in ‘"HV" = "bias"‘, ‘"HV"‘ is the name of the bias contact, while ‘"bias"‘ is the corresponding node name assigned to it. The elements in this connectivity list are separated by white space only;
- each compact model can include parameters, referred to as **`<attributes>`**, to control its behaviour. These parameters typically remain constant throughout the simulation. In this example, this feature is not utilised.

To define a constant model of an elementary device, the following code is used:

```
System
{
    parameter-set-name instance-name (node0 node1 ...) {<attributes>}
}
```

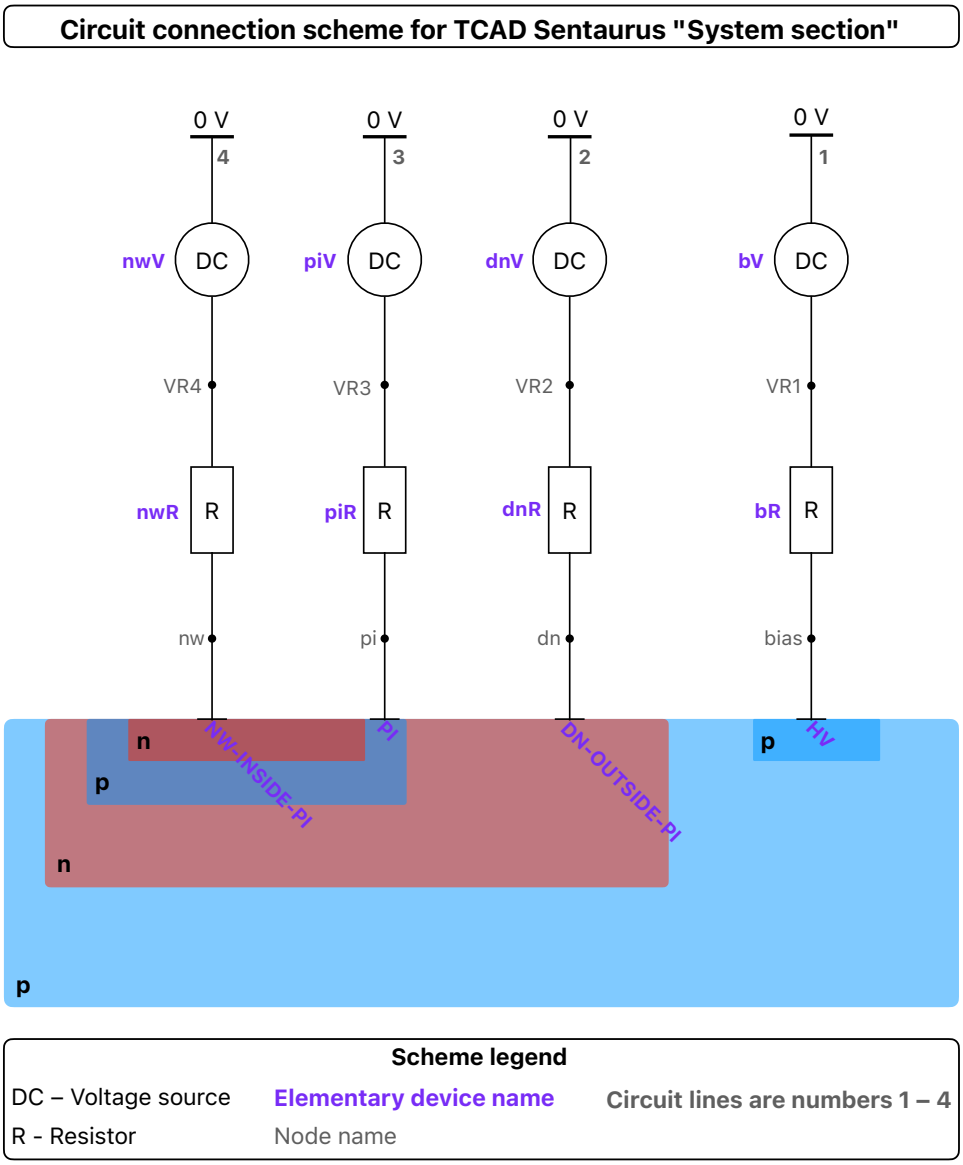


Figure 4.8: Schematic representation of the circuit connection used for the simulation of the HitPix chip, as applied in the System section of the simulation. The lower part of the diagram shows the *p*- and *n*-doped areas, coloured in blue and red, respectively, with assigned contact names. These names are first introduced in the geometry file (subsection 4.1.4) and correspond to the electrode names in the Electrode section (subsection 4.2.3). Each contact is connected through a resistor to a voltage source, as illustrated in the circuit connection scheme.

Using this template, the code for elementary devices, which are used to simulate the scheme illustrated in Figure 4.8, is the following:

```
System
{
    Vsource_pset bv (VR1 0) {dc = 0}
    Vsource_pset dnV (VR2 0) {dc = 0}
    Vsource_pset piV (VR3 0) {dc = 0}
    Vsource_pset nwV (VR4 0) {dc = 0}

    Resistor_pset bR (bais VR1) {resistance = 1e7}
    Resistor_pset dnR (dn dn_VR2) {resistance = 1e5}
    Resistor_pset piR (pi VR3) {resistance = 1e6}
    Resistor_pset nwR (nw VR4) {resistance = 1e5}
}
```

The syntax of this command is as follows:

- **`parameter-set-name`** represents the SPICE name of the specified device. In this simulation, **`Vsource_pset`** defines an independent voltage source, while **`Resistor_pset`** defines a linear resistor.
- Each device has a unique **`instance-name`**, which is freely chosen. For clarity, every device name is shown in Figure 4.2.3 next to its corresponding circuit symbol.
- On the contrary to physical device **`(connectivity list)`** for elemetrary device additional node definition is needed. The list of nodes **`(node0 node1 ...)`**, specifying the terminals between which the device is placed, must be provided with white space separation. If a node is neither zero nor referenced elsewhere, the device is considered floating.
- The command **`<attributes>`** is used to define the parameters of the selected device. In this case, all voltage sources have an initial value of 0 V and operate in DC mode. For linear resistors, their resistance values are specified in ohms [Ω].

After these definitions, the full circuit is set to its initial state and is ready for computation and processing, as described in the Math (Subsection 4.2.7) and Solve (Subsection 4.2.8) sections.

4.2.7 Math section

To successfully conduct a physical experiment, it is essential not only to understand the physics of the device under investigation but also to be proficient in using the measurement equipment. The same principle applies to numerical experiments: to ensure a reliable simulation, both the physics of the device and the mathematical methods employed must be correctly defined. *Sentaurus Device* solves the device equations, which are essentially a set of partial differential equations, self-consistently, iteratively and on the discrete mesh. The solution methods are specified in both the Math and Solve sections (for the Solve section, see Subsection 4.2.8). These two sections are interdependent: the Math section determines "how to solve" by controlling

the solver configuration, while the Solve section specifies "what to solve" by defining the simulation conditions. This section provides an introduction to the computational approach and solvers used in *Sentaurus Device*. It covers linear and non-linear solvers, accuracy control, convergence monitors, additional software-related functionalities, and an overview of relevant *Sentaurus Device* keywords.

Iterative methods are effective in achieving convergence efficiently. However, there is an inherent trade-off between convergence robustness and computational speed. More robust convergence typically requires a greater number of iterations and a finer mesh, leading to longer computation times. Depending on the specific problem, two primary approaches are employed to solve the coupled equations of the drift-diffusion model in semiconductor device simulations: Newton iteration, a fully coupled method, and Gummel iteration, a decoupled method [40].

In the fully coupled Newton iteration, the total system of unknowns is solved simultaneously, resulting in a single system that includes all the equations to be solved, such as the Poisson and Continuity equations (see Subsection 2.1.5). In contrast, the Gummel method solves each equation separately within each iteration. By solving them iteratively, it updates only the primary unknown for each equation: the electric field for the Poisson equation and carrier densities for the Continuity equations. An iteration is considered complete once all independent variables have been updated.

Newton's iteration converges faster than Gummel's, with quadratic convergence compared to the linear convergence of Gummel. However, each Newton iteration is computationally more demanding than a Gummel iteration. While Gummel converges more slowly, it can tolerate poor initial guesses, making it particularly useful when a good starting point is difficult to determine. In such cases, an effective strategy is to refine the initial guess using Gummel and then switch to Newton for faster convergence. In general, Gummel is preferred when the transport problem can be decoupled from the electrostatic problem, such as in low-field conditions where diffusion dominates and effects like band-to-band tunnelling or avalanche multiplication are negligible. In these scenarios, Gummel is more efficient than Newton, as the latter continues to update quantities that change only weakly. The equilibrium solution is often used as the initial guess [40].

The Math section in *Sentaurus Device* consists of a list of keywords, where each keyword represents a specific solver parameter. These parameters, which define the solution algorithms, are device-independent and must only appear in the Math section. Certain physics models, such as beam simulations, require additional function calls within the Math section. This is further described in Section 4.3. An example of the Math section code is shown below:

```
Math
{
    Method = PARDISO
    Number_Of_Threads = maximum

    Iterations = 25
    RhsFactor = 1e30
```

```
    Digits = 6
    Error = 1e8
    NotDamped = 200

    Extrapolate
    RelErrControl
}
```

The code is structured into three sections: configuring the processor's computational capabilities, defining parameters for the chosen iteration method, and setting additional parameters for different simulation modes. In the first section, the keyword **`Method`** specifies the linear solver to be used. Library **`PARDISO`** is Parallel Sparse Direct Solver, which is a high-performance, robust, memory-efficient and easy-to-use software for solving large sparse symmetric and non-symmetric linear systems of equations on shared–memory and distributed-memory architectures [63]. *Sentaurus Device* utilises thread parallelism to accelerate simulations on shared-memory computers, which also extends to the linear solver. For this purpose, the keyword **`Number_Of_Threads`** controls the number of computational threads, where setting it to **`maximum`** assigns the number of threads equivalent to the number of processors available on the execution platform.

The next section defines the parameters that determine whether the Coupled statement converges or diverges. In the following discussion, RHS refers to the right-hand side or the residual of the equations. The keyword **`Iterations`** specifies the maximum number of Newton iterations allowed per bias step. **`RhsFactor`** provides control over the size of the RHS by setting a limit on the increase of the RHS during a single Newton step [42]. During the solving procedure, not only are the limit criteria important, but so is the precision of the calculation. To ensure numerical accuracy, *Sentaurus Device* evaluates an equation variable x such that the computed update Δx after the n-th iteration is sufficiently small. This means that iterations will continue until:

$$\frac{\left|\frac{\Delta x}{x^*}\right|}{\varepsilon_R \times \left|\frac{x}{x^*}\right| + \varepsilon_A} < 1 \tag{4.24}$$

where $\varepsilon_R = 10^{-\texttt{Digits}}$, $\varepsilon_A = \texttt{Error}$, and x^* is a scaling constant equal to 0.1 [40].

Another computational approach that is useful for obtaining an initial solution is damping. In some cases, it can enhance convergence when large bias steps are taken. This approach, named Bank-Rose dumping [64], tries to achieve convergence of the coupled iteration far from the final solution by automatically adjusting the size of the update based on how the RHS changes. Bank–Rose damping becomes activated when the number of iterations exceeds the value specified with **`NotDamped`** in the Math section [42].

The final two keywords, **`Extrapolate`** and **`RelErrControl`**, regulate quasistationary bias ramps used in the Solve section. Enabling **`Extrapolate`** allows the initial guess for a given step to be obtained through extrapolation from the solutions of the previous two steps. Meanwhile, **`RelErrControl`** modifies the error control mechanism during iterations, switching from in-

ternal error parameters to more physically meaningful parameters. Further details on the Quasistationary command are discussed in Subsection 4.2.8.

4.2.8 Solve section

As described in the previous Subsection 4.2.7, which is dedicated to the Math section of TCAD simulation, the solving methods are defined by both the Math and Solve sections. These two sections function co-dependently, where the Math section is responsible for "how to solve" or controls for the used solver, and the Solve section describes "what to solve" or simulation conditions.

The Solve section is the only part of the command file where the order of commands is crucial. It consists of a series of simulation commands executed sequentially according to their arrangement in the command file. Many of these Solve commands are high-level, with lower-level commands as their parameters. An example of these high-level commands is provided below [42]:

- The **`Coupled`** command activates a Newton-like solver to solve a system of equations. This is a fundamental command associated with Newton's iteration method. It is typically used for solving the Poisson equation, continuity equations, various thermal and energy equations, or a combination of the Poisson and continuity equations. This command computes the device's response under steady bias conditions without applying an external voltage, determining the system's initial state.
- The **`Plugin`** command is used to iterate between multiple coupled equations. This command employs the Gummel iteration method. Since the fully coupled Newton method is generally preferred due to its faster convergence and reliability in most cases, the **`Plugin`** approach is rarely necessary.
- The **`Quasistationary`** command ramps a solution from one boundary condition to another. It adjusts a parameter, typically the bias voltage, from one value to another in discrete steps. At each step, the device is solved under steady-state conditions. Additionally, small-signal AC analysis can be performed using the **`ACCoupled`** lower-level command, which is a subcommand of **`Quasistationary`** for small-signal AC analysis.
- The **`Transient`** command is used for transient simulations, where the system is analysed over time. This is particularly useful for simulating time-dependent phenomena, such as the signals generated in a radiation detector when struck by a particle or during charge injection events. Further details are provided in Section 4.3.

The structure of the Solve section is demonstrated below. Following this, the high-level commands used to compute the device response are discussed in detail, considering their order and purpose.

```
Solve
{
    Coupled ( <optional parameters> ){ <equation> }
```

```
    Quasistationary ( <parameter-list> ) { <solve-command> }
    Transient ( <parameter-list> ) { <solve-command> }
}
```

Establishing the initial conditions

The first step is to solve only the Poisson equation:

```
Coupled (iterations = 50) {Poisson}
```

Starting the Solve section with a single equation enhances computational efficiency and reliability. This step provides the initial condition solution, which serves as the starting point for subsequent calculations.

The second step introduces the continuity equations for electrons and holes, incorporating the initial bias conditions defined in the System section:

```
Coupled (iterations = 15) {Poisson Electron Hole}
```

The solution is obtained iteratively: it begins with the Poisson equation, after which the electron and hole carrier continuity equations are introduced. Here, the electron continuity equation is solved fully coupled to the Poisson equation, using the solution from the previous step as the initial guess. The same approach is applied to the holes continuity equation.

Computing the current continuity and transport equations separately under equilibrium conditions improves convergence robustness. This additional step provides a more accurate initial estimate for the electron and hole densities when a nonzero current is present.

Additional initial conditions available in Mixed Mode

The transport equations for the contact and circuit are introduced in mixed-mode simulations. The **`Circuit`** keyword enables the solution of electrical circuit models and nodes, while the **`Contact`** keyword activates the solution of electrical interface conditions at the contacts.

```
Coupled (iterations = 15) {Electron Hole Poisson Contact}
```

By default, the **`Contact`** and **`Circuit`** keywords are not required, as the Poisson equation inherently includes the contact and circuit domains. If only the Poisson equation is solved, no additional equations are introduced. However, if a **`Coupled`** statement includes equations beyond the Poisson equation, the circuit and contact equations are also incorporated into the solution.

Achieving needed condition by ramping up voltage

The **Quasistationary** command gradually transitions a device from one solution to another by modifying its boundary conditions or parameter values. The most commonly adjusted boundary condition is the voltage at the contacts. The simulation proceeds iteratively, with each iteration involving an adjustment to the boundary conditions or parameters, followed by a re-solving of the device equations. The re-solving process at each iteration is governed by the **Quasistationary** command, a widely used method in TCAD simulations for semiconductor device modelling [42]. For example, to ramp the voltage applied to the electrode contacts of a sensor diode to 1.8 V, which is necessary for its proper operation, the following **Quasistationary** command can be used:

```
Solve
{
    Quasistationary
    (
        InitialStep = 1e-6
        MinStep = 1e-9
        MaxStep = 0.01
        Goal
        {
            Parameter = dnV.dc,
            Voltage = 1.8
        }
    )
    {
        Coupled {Poisson Electron Hole Contact Circuit}
    }
}
```

In this case, the goal is to ramp the voltage from its initial value, which is typically 0 V, to 1.8 V. The voltage at the contact changes according to the formula:

$$V = V_0 + t \times (V_1 - V_0) \tag{4.25}$$

where V_0 is the initial voltage, V_1 is the final voltage, which is 1.8 V in this example, and t is a variable that controls the ramping process. The variable t ranges from 0 to 1, and it governs how the voltage transitions from V_0 to V_1 over the course of the simulation. The **<parameter-list>** defines the stepping process, controlling how t changes during each iteration. This iterative process is a core feature in TCAD simulations, where the goal is to find the steady-state solution of the system, assuming that the transient effects are negligible within the time frame of the simulation [65]. It is important to note that the control of the ramping process is expressed in terms of the t-variable and is not directly applied to the contact values themselves. This allows for the simultaneous ramping of multiple contacts, with the same stepping process applied across all relevant contacts. This feature is crucial for modelling complex semiconductor

structures, where different parts of the device may be subject to different boundary conditions but are still solved within a unified framework [66].

The stepping process is controlled using a few key parameters in the **<parameter-list>**:

- **MaxStep** and **MinStep** specify the maximum and minimum step sizes for t during the ramping process.
- **InitialStep** determines the size of the first step.
- The step size can increase or decrease depending on the success of the inner solve command. If the solver converges successfully, the step size will be increased according to the **Increment** factor; if the solver fails, the step size will be reduced based on the **Decrement** factor. The ramping process will terminate if the step size becomes smaller than **MinStep**.

These parameters are described in more detail in the *Sentaurus Device* User Manual [42], which provides explicit instructions for setting up simulations with various solver options and control settings [42].

The **<solve-command>** specifies the solver used during each iteration. Possible solvers include **Coupled**, **Plugin**, **ACCoupled**, or another **Quasistationary** command. These solvers dictate how the device equations are solved during the ramping process. Coupled solvers are frequently used in simulations that involve the simultaneous solution of multiple physical models, such as Poisson's equation, electron-hole transport, and contact circuit models, which are all common in semiconductor device simulations [67].

Small-signal AC analysis

An **ACCoupled** solve section is an extension of the **Coupled** section, with an additional set of parameters that allow for small-signal AC analysis. In general, **ACCoupled** is used in mixed-mode simulations. AC analysis is based on small-signal techniques, where the device's response to small sinusoidal signals, superimposed upon an established DC bias, is computed as a function of frequency and the DC operating point [68]. In this approach, a steady-state solution is first computed to establish the DC operating point. Then, a linear algebraic system is built up, which describes how the solution vector varies in response to small sinusoidal perturbations at the device contacts. The solution to this system provides the real and imaginary parts of the response, characterising how the device behaves under small-signal conditions.

The following example demonstrates how to set up a **QuasiStationary** statement to perform an AC analysis at a single frequency for each DC bias step during a voltage source sweep. The voltage source **bV**, which is attached to the bias contact of the sensor, is used to sweep the DC bias. This procedure is essential for obtaining frequency-dependent responses in semiconductor simulations.

```
Solve
{
    QuasiStationary
    (
        InitialStep = 1e-7
        MinStep = 1e-10
        MaxStep = 1e-2
        Increment = 2
        Decrement = 4
        Goal
        {
            Parameter = bV.dc
            Voltage = @HV@
        }
    )
    {
        ACCoupled
        (
            StartFrequency = 1e6
            EndFrequency = 1e6
            NumberOfPoints = 1
            Iterations = 15
            Node(nw pi dn bias)
            Exclude(nwV piV dnV bV)
            ACMethod = Blocked
            ACSubMethod("sensor") = ParDiSo
        )
        { Poisson Electron Hole Contact Circuit }
        ACPlot
        (
            FilePrefix = "@tdrdat@"
            Time = (Range = (0.0 1.0) Intervals = 3)
            NoOverwrite
        )
     }
}
```

The parameters for the **QuasiStationary** command are the same as those used in the **Coupled** method, which were described earlier. Within an **ACCoupled** section, you must specify the frequencies of interest and the circuit nodes that will be included in the admittance matrix. Additionally, some components may need to be excluded from the analysis to properly describe the AC system [69].

Key parameters for **ACCoupled**:

- **StartFrequency**, **EndFrequency**, **NumberOfPoints**, **Linear**, **Decade** – these parameters define the frequency range and resolution for the AC analysis. If the request is to analyse the system over multiple frequencies, set both **StartFrequency** and **EndFrequency**. The **NumberOfPoints** defines the resolution of the frequency sweep, while **Linear** and **Decade** can be used to adjust how frequencies are distributed.
- **Node** specifies the list of AC nodes considered in the admittance matrix. The list must be non-empty, as it determines where the small-signal perturbations will be applied. For example, in the statement **Node(nw pi dn bias)**, the AC analysis is performed at the specified circuit nodes **nw**, **pi**, **dn**, and **bias**.
- **Exclude** defines a list of circuit components or instances that should not be included in the AC system. Typically, the power supply is excluded to prevent it from dominating the AC analysis. It can also be used to isolate a single device within a larger circuit for analysis [70].
- **ACPlot** invokes the creation of plots of the device's AC response functions. The output file name prefix can be specified here, and the simulation will generate corresponding plot files for each AC node voltage excitation and frequency.

In the File section (see Subsection 4.2.2), the **ACExtract** command can be used to specify a prefix for the extraction file. The extraction file contains the frequency, node voltages, and matrix entries for the admittance matrices, which describe the system's response at different frequencies.

This section has provided a detailed exploration of the process of writing and configuring TCAD code for device simulations, with a focus on the key steps and techniques required for accurate semiconductor device modelling. Through examples and explanations, the setup of simulation parameters, selection of appropriate models, and integration of physical equations into the simulation framework have been discussed. By understanding how these parameters are defined and manipulated, a deeper insight into the behaviour of semiconductor devices under various conditions can be gained, enabling more accurate predictions and optimised designs. With this foundation in TCAD device simulation established, attention will now be turned to the simulation of beam interactions in TCAD in the following Section 4.3.

4.3 Beam simulation and radiation damage model

This section focuses on modelling the interaction of Minimum Ionising Particles (MIPs) and radiation-induced damage in silicon detectors using TCAD simulations. When a heavy ion penetrates a device structure, it loses energy and generates a trail of electron-hole pairs. These additional carriers can produce sufficiently large charges to disrupt the normal operation of the device, potentially causing phenomena such as latch-ups or permanent damage, depending on the device's design and the intensity of the interaction. Synopsys enables the simulation of bulk radiation damage by directly modelling the dynamics of radiation-induced traps. To achieve

accurate results, it is essential to select a set of traps that produce the correct macroscopic behaviour and are reasonably consistent with experimental measurements of trap types and concentrations.

MIP simulations are fundamental to particle physics experiments, particularly in the design and optimisation of tracking silicon detectors. MIPs serve as a standardised model for particles that deposit minimal energy as they traverse a medium, closely replicating the behaviour of charged particles. In tracking detectors, MIP simulations are vital for assessing charge collection efficiency, signal-to-noise ratios, and the impact of radiation damage on long-term detector performance. By accurately modelling MIPs' impact on detector performance, various parameters can be optimised, including detector geometry, material properties, and signal processing techniques, which can enhance spatial resolution and ensure precise tracking of particle trajectories. This makes MIP simulations a key tool for improving the accuracy and reliability of modern silicon tracking detectors.

Incorporating radiation damage models into TCAD simulations is critical for accurately predicting the performance of detectors in high-radiation environments, such as those encountered in particle physics experiments or certain medical applications. Radiation introduces defects, or traps, into the silicon bulk and at interfaces, which significantly affect charge collection efficiency, leakage currents, and overall detector operational lifetime. These defects alter the electrical properties of the detector, leading to signal degradation and reduced spatial resolution. By including detailed radiation damage models, such as trap-based approaches, TCAD simulations enable precise evaluation of these effects. This enables the optimisation of detector designs, the prediction of performance over time, and the assurance of reliability in high-radiation environments [71].

For more information about the interaction of the particle with the matter and radiation damage in silicon, see Subsection 2.2.1 and Subsection 2.2.2, respectively. This section is dedicated to simulations of these processes, including specific modes and codes that are used.

4.3.1 Minimum Ionising Particle (MIP) simplification in TCAD

In the TCAD simulation environment, the **`HeavyIon`** function enables precise modelling of energy deposition and the subsequent carrier generation caused by an MIP traversing the device. It simulates the particle's track as a series of localised energy deposits, allowing detailed analysis of charge generation, recombination, and transport processes. This method helps study the transient and steady-state device responses, contributing to the design and optimisation of radiation-hardened semiconductor devices.

Key parameters for MIP simplification using **`HeavyIon`** function in TCAD include:

- the energy and type of the ion,
- the angle of incidence of the ion,
- the relationship between energy loss, or linear energy transfer (LET), and the number of electron-hole pairs created.

In this thesis, MIP was simulated using the **HeavyIon** function within the *Sentaurus Device*. The simulation employed a linear energy transfer function of $LET_f = 1.282 \times 10^{-5}$ pC / µm corresponding to 80 electron-hole pairs per micrometre. Impact ionisation effects were incorporated using the avalanche model from the University of Bologna, which accounts for both electron and hole driving forces [72]. Below is the TCAD simulation code for modelling an MIP, which was implemented in this work:

```
Physics
{
    HeavyIon
    (
        Direction = (@angle@,1)
        Location = (@x_coordinate@,0)
        Time = 2e-9
        Length = [0 0.001 !(puts @thickness@)! !(puts [expr @thickness@ + 0.001])!]
        Wt_hi = [1.0 1.0 1.0 1.0]
        LET_f = [0 1.281741e-05 1.281741e-05 0]
        Gaussian
        PicoCoulomb
    )
}
```

The parameters used in the **HeavyIon** model are described as follows [38]:

- **Direction** defines a direction vector for the particle's path.
- **Location** is an incident position for the particle.
- **Time** is a moment of particle incidence.
- **Wt_hi** is a characteristic distance of the energy deposition.
- **LET_f** is linear energy transfer (LET) function.
- **Length** is a spatial region where **Wt_hi** and **LET_f** are specified.
- **Gaussian** defines a Gaussian spatial distribution.
- **PicoCoulomb** specifies units for **LET_f**, **Wt_hi**, and **Length** in micrometres.

In the provided code, many beam characteristics are represented as variable parameters, defined using the `@parameter@` syntax, which can later be specified in *Sentaurus Workbench*.

The beam simulation results are illustrated in Figure 4.9.

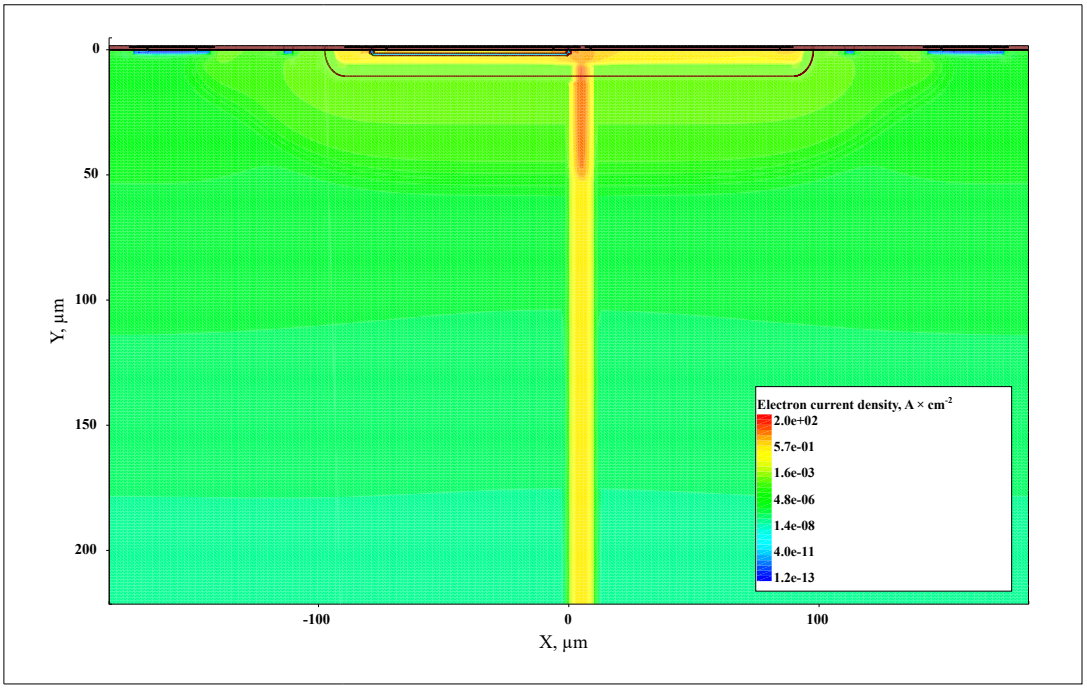


Figure 4.9: Electron current density distribution as a result of MIP traversal. The colour scale represents the current density in units of A $\times$ cm^{-2}.

4.3.2 Traps and radiation damage modelling

Traps play a crucial role in the physics of semiconductor devices. They introduce allowed energy states within the semiconductor bandgap, influencing device behaviour in various ways, such as altering the effective doping, enhancing generation and recombination processes, and increasing leakage current. While some models, such as Shockley–Read–Hall recombination, account for traps implicitly, they do not explicitly model trap occupation or the space charge stored within traps. This section discusses models that consider trap occupation and charge storage, as well as the specification of fixed charges [73].

Sentaurus Device offers several types of traps, including acceptor and donor traps, as well as fixed charges. It supports five types of energy distributions—level, uniform, exponential, Gaussian, and table—and provides various models for capture and emission rates, such as the V-model (with optional field-dependent cross sections), the J-model, and non-local tunnelling. Acceptor traps are uncharged when unoccupied but carry a negative charge equivalent to one electron when fully occupied. Conversely, donor traps are uncharged when unoccupied but carry a positive charge equivalent to one hole when fully occupied. Typically, traps in the upper half of the bandgap are assumed to be acceptors, while those in the lower half are assumed to be donors. However, care must be taken when defining trap types, especially for interface trap states, which often act as amphoteric traps. Both bulk and interface traps are available, making accurate parameterisation essential for modelling radiation damage effects correctly [74].

Simulations were conducted using the three-level "new Perugia" trap model for $p-$type silicon, summarised in Table 4.1. This model, originally derived from the Perugia damage model [75], includes three trap levels and was adapted with parameters from D. Pennicard [76].

Table 4.1: New Perugia bulk damage model for $p-$type silicon, with parameters modified by D. Pennicard

Type	Defect	Energy, eV	σ_e (cm^2)	σ_h (cm^2)	η (cm^{-1})
Acceptor	VV	$E_c - 0.42$	9.50×10^{-15}	9.50×10^{-14}	1.613
Acceptor	VVV	$E_c - 0.46$	5.00×10^{-15}	5.00×10^{-14}	0.900
Donor	C_iO_i	$E_v + 0.36$	3.23×10^{-13}	3.23×10^{-14}	0.900

The trap energy levels in the table are specified relative to the conduction and valence bands rather than the midgap. The two acceptor levels lie slightly above the midgap, contributing to the leakage current, increasing effective $p-$type doping, and trapping excess electrons from the conduction band. The donor level, significantly below the midgap, has minimal impact on leakage current or doping concentration but primarily traps excess holes. The bandgap width is 1.12 eV.

Trap concentrations increase linearly with radiation fluence, following the relationship:

$$N_t = \eta\Phi \tag{4.26}$$

where Φ is the radiation fluence in 1×10^{15} MeV-n_{eq}/cm^2, and η is the trap introduction rate in cm^{-1}.

This trap model is illustrated in Figure 4.10.

The damage model was developed by selecting traps based on direct measurements of trap properties, such as those obtained through Deep Level Transient Spectroscopy (DLTS) [77], and adjusting them to better match macroscopic damage effects observed in detectors. While the model closely reproduces leakage current and effective doping concentration seen in experiments [75], its free carrier trapping behaviour deviates from experimental results. Experimental data show that trapping rates for electrons and holes increase linearly with fluence up to 1×10^{15} 1 MeV-n_{eq}/cm^2, requiring adjustments to the original Perugia model.

In *Sentaurus Device*, this information is specified within the Physics section of the command file (`.cmd`). The following keywords define trap types and distributions:

- **`FixedCharge`** represents traps that are always fully occupied.
- **`Acceptor`** and **`eNeutral`** are uncharged when unoccupied, negatively charged when occupied.
- **`Donor`** and **`hNeutral`** are uncharged when unoccupied, positively charged when occupied.

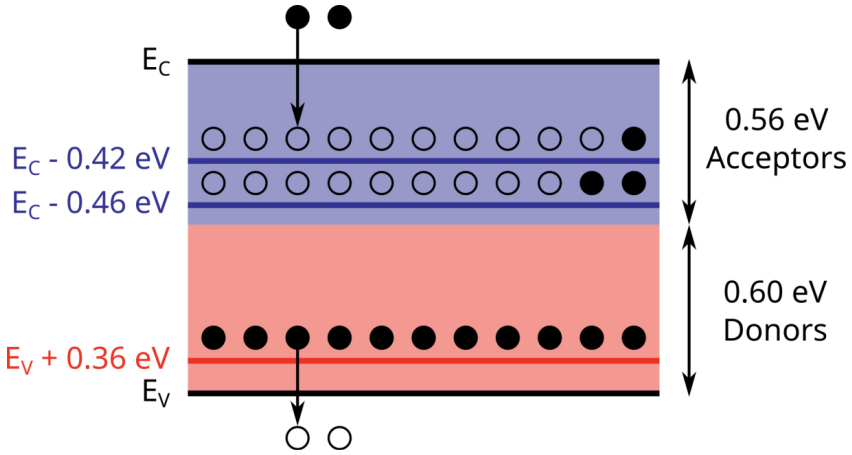


Figure 4.10: Diagram illustrating the three-level trap model for $p-$type silicon. The model highlights energy levels within the bandgap and their associated properties, which are crucial for understanding charge carrier dynamics in the material.

An example of the trap specification in the *Sentaurus Device* command file is shown below:

```
Physics (material = "Silicon")
{
    Traps (
        (Acceptor Level EnergyMid=0.42 fromCondBand Conc=2.3226E15 Randomize=0.29
         eXsection=9.5E-15 hXsection=9.5E-14)

        (Acceptor Level EnergyMid=0.46 fromCondBand Conc=1.8E15 Randomize=0.23
         eXsection=5E-15 hXsection=5E-14)

        (Donor Level EnergyMid=0.36 fromValBand Conc=1.8E15 Randomize=0.31
         eXsection=3.23E-13 hXsection=3.23E-14)
    )
}
```

In this code, the traps and their properties are defined in detail:

- **`Acceptor`** traps are uncharged when unoccupied but carry the charge of one electron when fully occupied, as indicated by the **`Level`** specification **`EnergyMid=0.42 fromCondBand`** that places the trap slightly below the conduction band.
- **`Donor`** traps, specified with **`Level EnergyMid=0.36 fromValBand`**, are uncharged when unoccupied but carry the charge of one hole when fully occupied, with their energy level slightly above the valence band;
- The **`Conc`** keyword sets the trap concentration, which is recalculated for a fluence of 2×10^{15} MeV-n_{eq}/cm^2 based on earlier formula 4.26. For instance, the first acceptor has a concentration of 2.3226×10^{15} cm^{-3}.

- **eXsection** and **hXsection** specify the electron and hole capture cross-sections, critical for accurately modelling carrier trapping.
- the **Randomize** parameter controls the spatial distribution of traps. A value greater than zero allows the distribution to be randomised, with the number serving as the seed for the random number generator. This ensures the reproducibility of results across simulations.

Specifying trap distributions and their associated parameters accurately is critical for modelling radiation damage effects. This capability ensures simulations reflect the influence of traps on the semiconductor's electrical properties under various conditions [42].

4.3.3 Transient analysis and solvers

In TCAD simulations, the **Transient** command is a fundamental tool for modelling heavy ion interactions, as it captures the time-dependent dynamics of a device's response to ionising events. When a heavy ion or Minimum Ionising Particle (MIP) traverses a semiconductor device, it deposits energy along its path, creating electron-hole pairs. These carriers drive dynamic processes such as drift, diffusion, recombination, and eventual collection at the electrodes.

The **Transient** simulation provides detailed temporal resolution of these processes, enabling the analysis of the immediate response of the device, including transient currents and charge collection, as well as the longer-term recovery to steady-state conditions. This approach is crucial for understanding radiation-induced effects, such as charge collection efficiency and Single Event Effects (SEEs), including Single Event Transients (SETs) and Single Event Upsets (SEUs). By utilising the **Transient** command, TCAD enables precise modelling of these complex, time-dependent phenomena, which is essential for designing and evaluating radiation-hardened semiconductor devices [42].

The **Transient** command performs a transient time simulation by iterating between time increments and re-solving the device's state. The simulation must begin with a device whose initial conditions have already been solved. Below is an example of a typical **Transient** command structure:

```
Solve
{
    NewCurrentPrefix = "trans_"

    Transient(
        InitialTime = 0
        FinalTime = 1e-7
        MinStep = 1e-15
        MaxStep = 1e-9
        Increment = 1.5
        Decrement = 2
```

```
    ) {
        Coupled {
            Poisson
            Electron
            Hole
            Contact
            Circuit
        }

        CurrentPlot(
            Time = (range = (0 2e-9) intervals = 40)
            Time = (range = (2e-9 3e-9) intervals = 40)
            Time = (range = (2e-9 1e-8) intervals = 20)
            Time = (range = (1e-8 1e-7) intervals = 20)
        )

        Plot(
            FilePrefix = "beam_@tdrdat@"
            Time = (0; 2e-9; 3e-9; 1e-8; 1e-7)
            NoOverwrite
        )
    }
}
```

In this example, the **InitialTime** and **FinalTime** define the start and end points of the simulation, while **MinStep** and **MaxStep** control the minimum and maximum allowed time steps. The time step is dynamically adjusted during the simulation. If a time step fails to compute the system's status, it is reduced iteratively by the factor specified with the **Decrement** keyword. Conversely, after a successful computation, the time step is increased up to the maximum factor defined by **Increment**. This dynamic adjustment ensures efficient and accurate resolution of the system's transient behaviour.

Plotting points at specific times requires additional effective time steps, as specified in the **CurrentPlot** and **Plot** commands. These effective steps do not affect the advancing time step, meaning subsequent steps can remain as large as the advancing step.

To improve numerical accuracy, additional mathematical commands must be included in the **Math** section of the command file:

```
Math
{
    CheckTransientError
    RecBoxIntegr
    RecBoxIntegr(5e-3 300 5000)
}
```

The **RecBoxIntegr** keyword activates a numerical integration procedure, originally developed for phonon generation, which is adapted here for MIP. This method divides each control volume into smaller rectangular boxes and integrates the generation rate within them, enhancing spatial accuracy. The default parameters can be adjusted to control the precision of the integration. The **CheckTransientError** keyword ensures time-step control. It can be specified globally in the Math section or locally within the **Transient** statement.

Through transient simulations, TCAD allows researchers to replicate various experimental conditions and study the dynamic behaviour of sensors under ionising radiation. This makes it an indispensable tool for sensor simulation and the development of robust, radiation-hardened devices.

4.4 Summary for TCAD simulations and applications

This chapter demonstrated the capabilities of TCAD Sentaurus tools, which are crucial for designing and simulating semiconductor devices. These tools enable a detailed understanding of physical processes, including doping, charge carrier dynamics, and circuit interactions. By integrating modules such as *Sentaurus Process*, *Sentaurus Structure Editor*, and *Sentaurus Device*, the framework provides an efficient way to model and optimise semiconductor devices. The focus of this chapter was to showcase the application of these tools using the HitPix chip as a detailed example. This chip was among the most complex to simulate and, therefore, serves as a demonstration of how TCAD simulations were performed for many of the chips discussed throughout this thesis The sensor's application and its detailed characterisation are presented in Chapter 6.

The simulation flow described in this chapter covers key steps from defining device geometry and material properties to meshing and performing advanced simulations. Figure 4.2 summarises this complete flow. In this work, this structured approach was applied to design and study the functionality of the HitPix chip. For example, simulations of $p-n$ junctions and doping profiles (Subsection 4.1.3) provided critical insights into the formation of depletion regions.

The Physics section of the simulation was carefully built to reflect the chip's design in order to accurately demonstrate its potential functionality (see Subsection 4.2.4). Based on the chip description and its performance requirements, relevant physical models were selected, including those for doping, mobility, recombination, and trap effects. These models were tailored to simulate key processes such as charge carrier dynamics, $p-n$ junction behaviour, and voltage responses. By aligning the Physics section with the chip's design specifications, the simulation provided a reliable representation of its operational characteristics.

To model the physical and electrical properties of the HitPix chip, the developed TCAD simulation flow was implemented. The system and its components were defined, as illustrated in Figure 4.8, and simulations were conducted to explore interactions between doping regions, contacts, and circuits. A three-level trap model for $p-$type silicon (Figure 4.10) was also

implemented, which improved the understanding of charge dynamics and their impact on device performance.

The visualisations of doping gradients (Figures 4.6 and 4.7) highlighted the relationship between doping profiles and device performance. The placement and biasing of contacts (Figure 4.5) were simulated to ensure optimal connectivity and functionality in the final chip design.

As a result, this work has demonstrated how TCAD tools bridge the gap between theoretical design and practical implementation. The methodologies developed in this work provide a strong foundation for future advancements in semiconductor design. The flexibility and precision of TCAD tools make them invaluable for addressing challenges in emerging technologies, such as high-performance detectors and advanced electronic systems. These approaches can be adapted to other devices, enabling continued innovation in semiconductor development.

Part II

Development and investigation of monolithic HV-CMOS sensor technology

5 The HV-CMOS technology

The metal–oxide–semiconductor field-effect transistor, or MOSFET, is a fundamental component in modern integrated circuits (ICs). It serves as the core for high-density microelectronic systems such as microprocessors and semiconductor memories, and it has also become important in power electronics and particle detection applications. The concept of a surface field-effect transistor (FET) was initially proposed in the early 1930s by Lilienfeld and Heil. However, it was not until 1960 that Ligenza and Spitzer succeeded in fabricating a high-quality Si/SiO_2 MOS system using thermal oxidation [2]. This achievement laid the foundation for the widespread use and further refinement of the technology in the decades that followed.

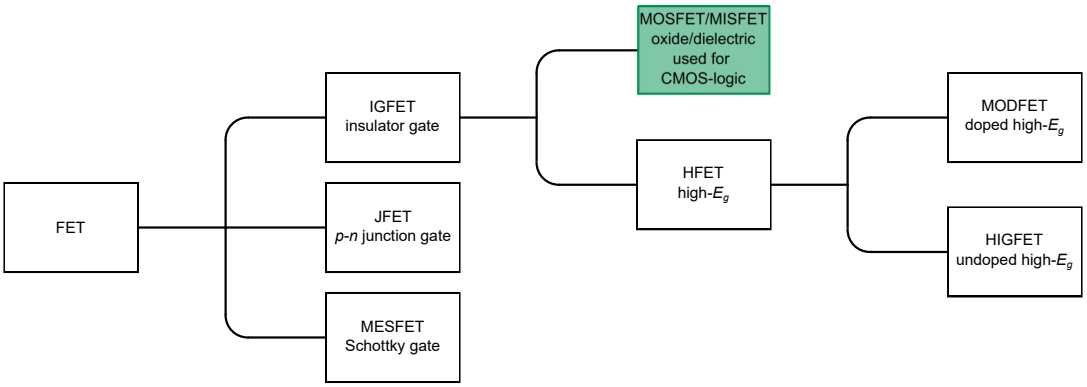


Figure 5.1: Family tree of field-effect transistors (FETs), classified by gate structure and material system [2].

The MOSFET is a central branch of the broader FET family. A classification of FET types is presented in Figure 5.1. At the highest level, FETs are divided into three primary categories: insulated-gate FETs (IGFETs), junction FETs (JFETs), and metal–semiconductor FETs (MESFETs). These types differ in how the gate capacitance is formed: IGFETs employ an insulating layer, typically an oxide; JFETs use a reverse-biased $p - n$ junction; and MESFETs rely on a Schottky barrier. The IGFET family is further subdivided into MOSFETs (or more generally, MISFETs) and heterojunction FETs (HFETs). In MOSFETs, the gate insulator is typically a thermally grown silicon dioxide layer, while in MISFETs, it may be a deposited dielectric. HFETs utilise a high-bandgap semiconductor layer formed via heteroepitaxy to create the insulating barrier. Although MOSFETs can be fabricated using various semiconductors such as Si, Ge, or GaAs, and with different insulator materials including SiO_2, HfO_2, Si_3N_4, and Al_2O_3, the most widely adopted and technologically mature system remains the silicon–silicon dioxide combination. This will be the main focus of the following sections, especially in the context of CMOS logic applications [2].

The fundamental structure of a MOSFET is shown in Figure 5.2. This example illustrates an $n-$channel MOSFET fabricated in a planar $p-$bulk silicon substrate. The corresponding circuit symbol is also shown. In this work, we focus on the $n-$channel variant, in which the mobile carriers in the conducting channel are electrons. A common MOSFET is a four-terminal device, which consists of a $p-$type silicon substrate into which two heavily doped n^+- regions are introduced. These are the source and drain, typically formed by ion implantation. A gate electrode is positioned above the channel region, separated from the silicon surface by a thin oxide layer (usually SiO_2) grown via thermal oxidation. While the illustration shows a generic metal gate, modern CMOS processes typically use heavily doped polysilicon or silicide-polysilicon stacks. In practical ICs, the MOSFET is laterally isolated from neighbouring devices either by thick field oxides or by trench isolation.

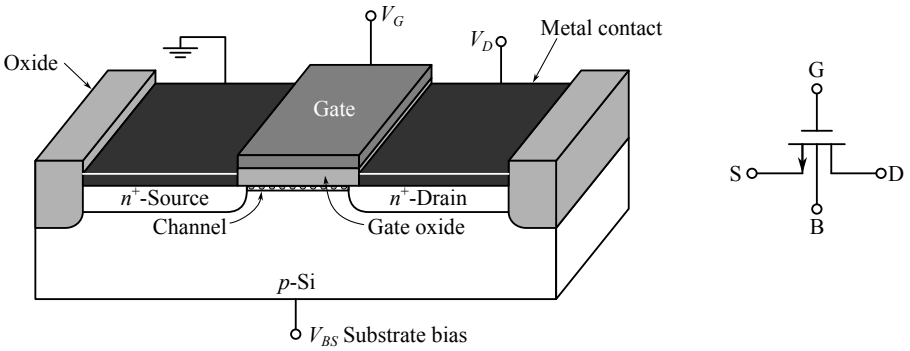


Figure 5.2: Cross-sectional schematic of an $n-$channel MOSFET and corresponding circuit symbol of the $n-$channel MOSFET [2]

The operation of the MOSFET under varying bias conditions is illustrated in Figure 5.3. The source contact is typically used as the reference point. When the gate is held at a low voltage or grounded, the two n^+- regions are separated by the p-type substrate, resulting in two back-to-back $p-n$ junctions that block current flow. When a small drain voltage V_D is applied while the gate is positively biased, the device operates in the *linear region*. However, when a sufficiently positive voltage is applied to the gate $V_G > V_T$, an inversion layer forms beneath the gate oxide, creating a conductive $n-$type channel that connects source and drain. The inversion channel behaves like a resistor, and the drain current increases proportionally with V_D (a). The conductivity of this induced channel is modulated by the gate voltage, enabling control of the drain current. As V_D increases, the channel potential near the drain rises, reducing the local inversion charge density. When the drain voltage reaches a critical value $V_{D\text{sat}}$, the inversion charge at the drain end is nearly depleted, and a *pinch-off* point forms, as shown in (b). While current continues to flow, the channel no longer extends fully to the drain. Beyond this point, as shown in (c), any further increase in V_D causes the pinch-off point to shift towards the source, but the voltage at this point remains fixed at $V_{D\text{sat}}$. As a result, the drain current saturates, becoming largely independent of V_D. The substrate terminal can be left floating, grounded, or biased; the choice of substrate bias V_{BS} affects the threshold voltage and channel behaviour [2]. In our application, which targets particle detection, a high reverse bias is always applied to the bulk in order to enlarge the depleted volume and ensure fast, drift-based charge collection.

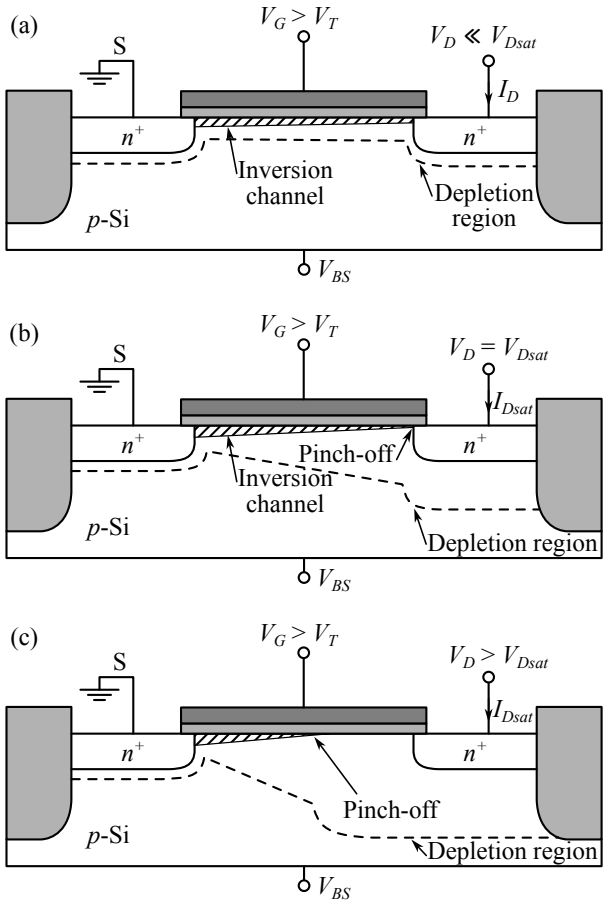


Figure 5.3: Operation regimes of an $n-$channel MOSFET: (a) linear region with low V_D; (b) onset of saturation, where the channel begins to pinch off; and (c) saturation region, where effective channel shortening occurs as $V_D > V_{D\text{sat}}$ [2]

HV-CMOS stands for High-Voltage Complementary Metal-Oxide-Semiconductor. In this context, "high-voltage" refers to supply voltages significantly above those used in standard low-voltage CMOS logic, typically exceeding 5 V. Originally developed for control electronics in automotive and industrial systems, high-voltage CMOS technologies were not initially designed for sensor applications [7, 8]. HV-CMOS predominantly uses MOSFETs, which enable both analogue front-end and digital logic integration. Modern HV-CMOS processes offer mixed-signal capabilities and support both PMOS and NMOS devices on a single substrate, making them highly versatile for circuit designers [8].

One of the defining strengths of detectors based on HV-CMOS is the integration of sensing and signal processing functions within a single silicon die. Unlike traditional sensors that only incorporate diodes as passive elements, HV-CMOS devices embed active electronics, which typically include signal amplification, and in more advanced designs, parts or the full readout chain. In some implementations, the entire signal path from charge generation to digital encoding and serial data transmission is monolithically integrated. Such devices are referred to as *monolithic sensors*, as they merge sensing, analogue, and digital functionality in one

substrate. More about this design approach and its differences from hybrid sensors can be found in Section 3.5.

To accommodate high-voltage operation, commercial HV-CMOS platforms provide specialised device structures. A key feature is the inclusion of “floating logic,” which isolates low-voltage transistors inside a deep nwell, allowing the application of high reverse bias between the $p-$substrate and nwell. The floating logic structure enables the integration of complex CMOS circuits directly within the deep nwell. Signal amplification, threshold discrimination, time measurement, and data sparsification can all be embedded at the pixel level, similarly to what is found in hybrid detectors. All transistors experience only small voltages between their terminals, enabling the use of compact, low-voltage device geometries even when the surrounding substrate is biased to tens or hundrens of volts [7].

Two approaches have been explored to increase the depletion volume in monolithic sensors: High-Voltage CMOS (HV-CMOS) and High-Resistivity CMOS (HR-CMOS). HV-CMOS designs use low-cost substrates and guarding structures to increase the maximum depletion voltage before breakdown. HR-CMOS, on the other hand, employs more expensive high-resistivity substrates to achieve deep depletion at lower voltages. These approaches are not mutually exclusive and are often combined depending on the application. While HV-CMOS relies on circuit-level tricks, HR-CMOS derives its advantage from the material itself. HR-CMOS should not be confused with standard CMOS sensors used, for example, in consumer optics. Like HV-CMOS, HR-CMOS aims for 100% fill-factor and fast drift collection. Some of the developments presented in the following chapters use high-resistivity substrates, while retaining the same basic design of HV-CMOS pixels [8].

Part II of this thesis, titled "Monolithic HV-CMOS sensors", is dedicated to the development and characterisation detectors based on of HV-CMOS technology. It includes two application-driven sensor projects: "HV-CMOS sensors for ion beam monitoring — HitPix", presented in Chapter 6, and "HV-CMOS sensors for future gamma-ray telescopes — AstroPix", discussed in Chapter 7. Although developed for different use cases, both sensors are based on the same HV-CMOS technology, demonstrating its versatility and adaptability across domains. In addition, Chapter 8 presents a more fundamental investigation: the influence of heavy ion irradiation on HV-CMOS sensors, which explores critical aspects of device performance under extreme beam conditions.

6 HV-CMOS sensors for ion beam monitoring — HitPix

This chapter is dedicated to the development of a monolithic active HV-CMOS hit-counting pixel sensor named HitPix. It is designed for radiation therapy with a high-rate ion or proton beam. This detector is proposed as a possible replacement for the existing beam monitoring system at the Heidelberg Ion Beam Therapy Center (HIT). The electronics of this chip were previously developed and covered in [78]; in this work, the main sensor design features are discussed. Nevertheless, a comprehensive system summary is provided, including new parameter calculations. The main focus is on sensor development using TCAD[1] and characterisation of its properties as a semiconductor detector.

6.1 Application description

In 2020, 2.7 million people in the European Union were diagnosed with cancer. Even though the environmental factors that lead to cancer have been reduced over the last decades, and the problem of behavioural risks such as smoking, poor diet, alcohol consumption, and low physical activity is addressed in the media, the challenge of fighting cancer is still high. According to the European Commission, around 450,000 new cases of cancer were diagnosed in Germany in 2022. Figure 6.1 shows the distribution by type of cancer for both men and women in Germany [79].

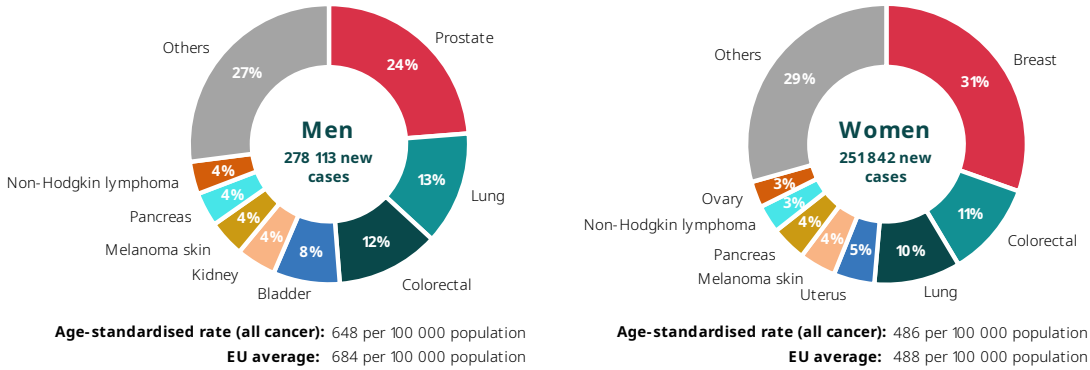


Figure 6.1: Data for cancer types for men and women in Germany, 2023.

[1] Technology Computer-Aided Design by Synopsys: https://www.synopsys.com/manufacturing/tcad.html

When it comes to treating cancer, many aspects have to be considered, for example, the type and stage of cancer and the general health of the patient. Each treatment plan may combine different phases, such as [80]:

- primary treatment, the goal of which is to surgically remove or to kill the cancer cells using radiotherapy or chemotherapy;
- adjuvant treatment, commonly includes chemotherapy, radiotherapy and hormone therapy, is used against cancer cells, which remain after primary treatment, to reduce the chance that cancer will recur;
- neoadjuvant therapy, which is similar to adjuvant but used before the primary treatment to make it easier and more effective;
- palliative treatment, which may help reduce side effects of treatment or signs and symptoms caused by cancer itself. While medications may relieve symptoms such as pain and shortness of breath, the use of surgery, radiotherapy, chemotherapy and hormone therapy is not uncommon at this stage.

Independent of the treatment phase, radiotherapy is a globally acknowledged and approved approach. It damages the cancerous cells, expectedly stopping the growth and multiplication of the cancerous tissue. Traditional radiotherapy uses X-rays; however, significant research has been conducted on the use of hadron therapy since the middle of the 20th century. Even though X-ray treatment has an obvious advantage over protons or ion beams due to its lower price and ease of exploitation, more and more hadron therapy facilities are operating worldwide. Statistics made by the Particle Therapy Co-Operative Group are shown in Figure 6.2 [81, 82].

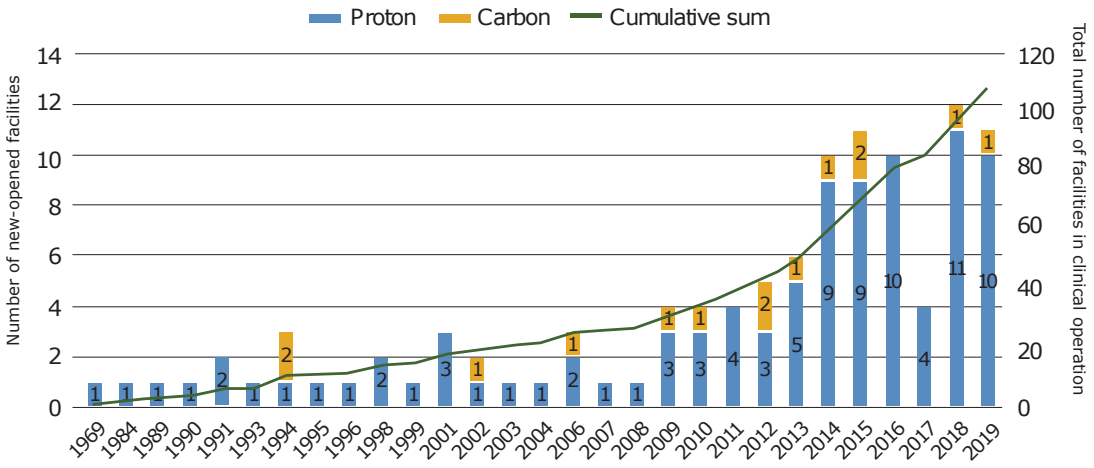


Figure 6.2: Number of particle therapy centres worldwide.

The physics-based advantages of charged particle therapy come from the different depth-dose distributions compared to conventional X-rays. While the photon dose decreases exponentially as a function of depth, hadrons deliver the highest dose at the stopping point, called the Bragg peak, while the dose everywhere else is low. Figure 6.3 shows the difference in dose distribution for X-rays and hadrons [83]. Protons' or heavier ions' energies can be well controlled, allowing for adjustment of the penetration depth and making the tumour treatment even more precise.

For example, by using a set of proton beams of different energies one after another, a spread-out proton peak can be achieved to irradiate the entire treatment region uniformly.

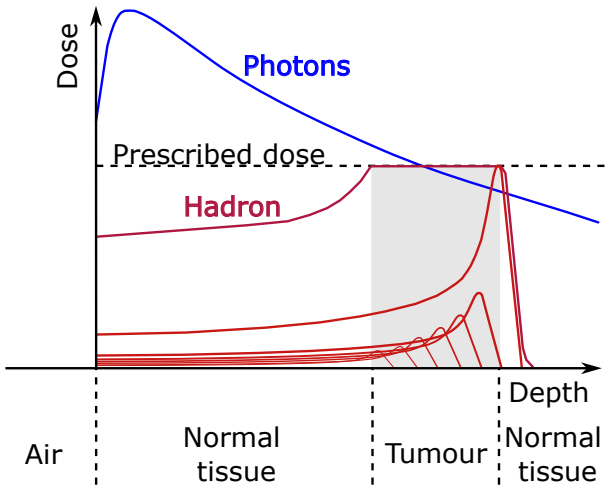


Figure 6.3: Energy deposition as a function of penetration depth for photons and hadrons [83].

Another critical parameter of radiotherapy is the relative biological effectiveness (RBE). It describes the ratio of radiation energy deposited at the tumour site to that in the path of the particle, which is a measure of the benefit versus risk of radiotherapy techniques [84]. A reference RBE of 1.0 is generally used for photons; for protons, it is 1.1, whereas, for carbon ions, most institutions use an RBE value of 3.0 [85]. By having a higher RBE, heavy ions can provide more effective treatment for certain, deep-seated tumours that are otherwise considered "radioresistant" [81].

Despite all the advantages of ion beam therapy, it is still technically challenging. Over the last twenty-five years, more and more different centres have opened worldwide; one was The Heidelberg Ion Beam Therapy Center (HIT) at Heidelberg University Hospital (UKHD) in 2009. It is a therapy facility which consists of a linear accelerator, synchrotron, two horizontal irradiation rooms, one room with a 360-degree rotating ion gantry and an irradiation research room for quality assurance [86]. Figure 6.4 illustrates the schematic plan of the facility, and Figure 6.5 shows patient rooms and the accelerator site.

HIT has several unique characteristics regarding delivering, controlling the beam, and treating patients. Protons, carbon, helium or oxygen ions are extracted from one of the three sources, accelerated in a linear accelerator, and injected into the compact synchrotron after all electrons have been removed. The control of the particle fluence and direction, called fluence-controlled continuous spot scanning, is applied during treatment. The extracted energy is controlled directly by the synchrotron and may be varied from spill to spill to achieve depth variation. The accelerator control system can currently provide 255 different energy settings, 15 intensity steps and 6 spot sizes for each ion species. The beam is produced in spills with lengths of up to 5 s and intensities up to 3.2×10^9 s^{-1} for protons. As the extracted beam intensity is not constant and may vary considerably during extraction, the monitor system measures the number of delivered ions at any time [6]. Currently, the beam monitoring system consists of

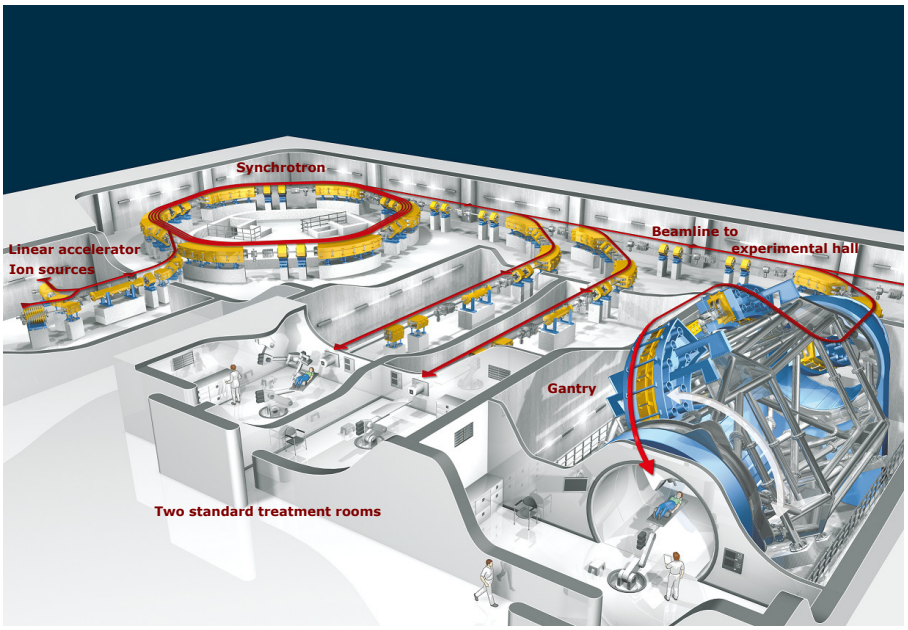


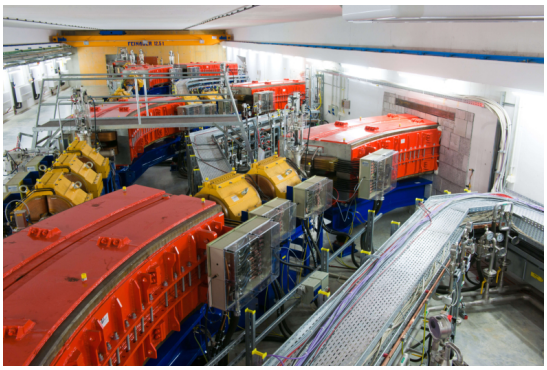
Figure 6.4: The plan of the HIT facility located at Heidelberg, Baden-Württemberg [87].

gas-filled ionisation chambers and multi-wire chambers (MWCs), which provide dose, position, and spot size information and are connected to the therapy control system [88].

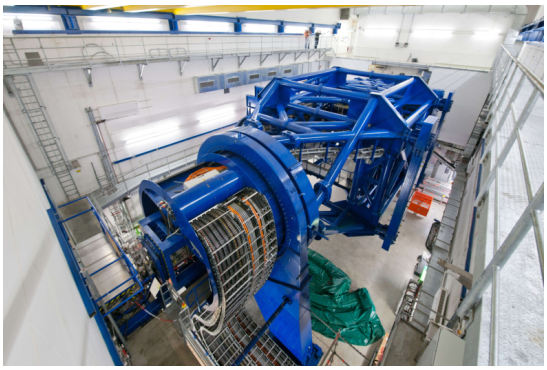
To fully benefit from high-precision ion beam dose delivery, image guidance is mandatory nowadays. Therefore, the development of magnetic resonance imaging (MRI)-guided ion beam therapy is an essential step in tumour therapy. This real-time imaging offers better soft tissue contrasts, can help locate the target volume in moving tissue, and, consequently, adapt the treatment plan during irradiation. However, compared to modern photon-based radiotherapy, MR-based image guidance for ion beam therapy is technically challenging and has yet to be made commercially available.

Gaseous detectors have been in operation for a long time as beam monitoring systems. Although it is a cheap and reliable technology, it has limitations for beam diagnostics. At HIT, each of the two multi-wire plate detectors provides only the projection of the beam spot, and due to the 2 mm wire pitch, the resolution is limited. The signal delay can be up to several hundreds of microseconds [6], which limits a very fast beam adjustment. The strong magnetic field of the MR scanner might influence the movement of the ionised gas in the MWCs, leading to an incorrect determination of beam position. The rapidly changing magnetic field and acoustic noise might also cause the wires to vibrate, leading to additional uncertainties. Prompt response, position, and dose measurement are used to control the irradiation process; they must always be precise and reliable.

Different beam monitors compatible with MRI are under investigation to make real-time imaging possible. One way is to improve an existing gaseous system, while another is to create a new one based on semiconductor detectors. For many years, silicon-based detectors have been used as the preferred option over MWCs in particle tracking for high-energy physics. Their stable performance in a strong magnetic field is standard for HEP application [89, 90].



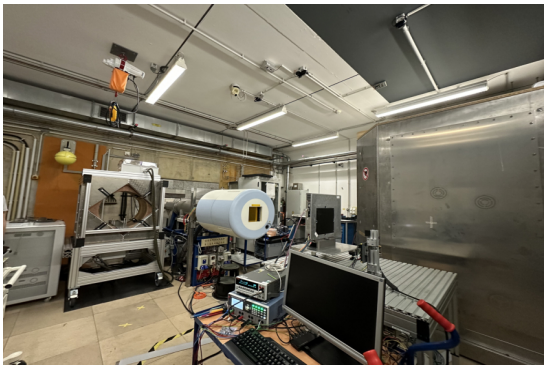
(a) Division of the beam in the direction of treatment rooms [86]



(b) Beam line through Gantry [86]



(c) Rotatable patient room in the Gantry [86]



(d) Experimental hall at HIT

Figure 6.5: Photos of The Heidelberg Ion Beam Therapy Center (HIT) at Heidelberg University Hospital (UKHD).

Therefore, they are expected to operate correctly in the presence of an MR scanner. Reviewing the most commonly used silicon sensors, the problem of high-rate beam detection appears not easily solvable: hybrid pixel detectors have a high material budget, strip detectors are not suitable for counting particles at a high rate, and standard CMOS sensors are not radiation tolerant enough [88].

HV-CMOS sensors, on the other hand, cover these requirements. As a monolithic sensor based on commercial chip production technologies, it implements readout circuits, such as particle counters in every pixel, making it capable of sensing high-intensity beams. Also, HV-CMOS technology is intrinsically more radiation tolerant than other CMOS technologies due to a larger depletion region. This minimises the charge collection time for signal charge. The sensors can be thinned to 50 µm, providing low material use. Finally, commercial technology grants low prices and wide availability, allowing small projects with limited budgets to be realised [88]. Finally, to create a complete beam monitor, several individual sensor chips must be combined, which are discussed in Section 6.2.

6.2 Detector description

Before discussing the sensor chip, it is essential to establish the specifications that development has been facing. Table 6.1 provides an overview of the main functions of the detector based on application requirements. Figure 6.6 shows the schematic 3D rendering of the full size detector design [6].

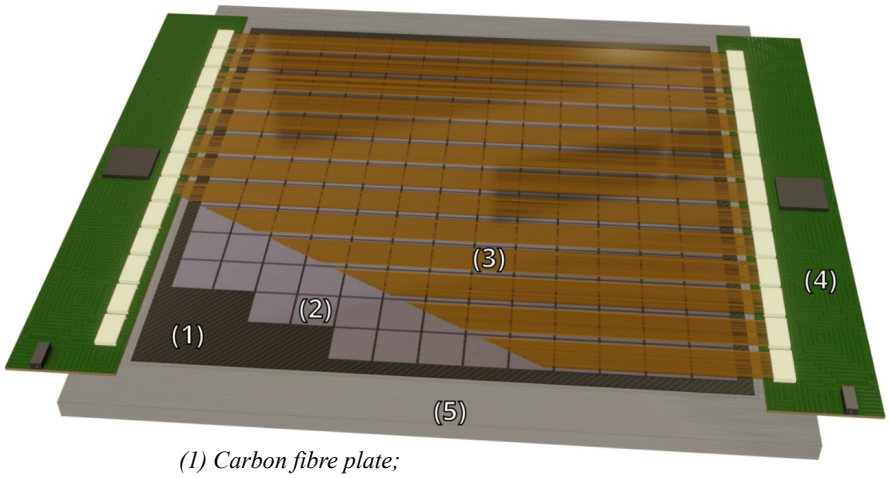
Table 6.1: List of requirements for beam monitor development

Category	Parameter	Solution
Detection concept	Contrary to more common particle identification, the beam monitor application requires hit counting for protons and light ions.	The proposed silicon monolithic HV-CMOS implements an in-pixel readout orientated to count the number of detected particle hits. The result of it is read out in frames.
Beam spot's size and position measurements	The beam position is calculated as the centre of gravity of the induced signal. The required spatial resolution is 200 µm, and the accuracy of the calculated beam spot width shall be better than 10% for a Gaussian shape, which gives us an FWHM resolution of 400 µm.	The size of each pixel is designed to be 200 µm × 200 µm. According to simulations, the gaps between chips, the same as the appearance of dead pixels, will not lead to a violation of the requirements [91].
Dose measurement	The deposited dose should be measured with 0.5% accuracy for radiation treatment planning and controlling.	The counting of particles of known energy directly provides the dose information, and the sensor current can be used to confirm this measurement [88].
Timing and latency as system response measure	The integration time for the beam spot's size and position should be less than 100 µs, with additional latency less than 100 µs. The values should be updated every 1–2 µs. As a safety measure, fast interlock generation is preferred in real-time.	The hit counting electronics provide fast processing of high particle rates, with integrating time in a range of 30 µs. The built-in projection minimises the amount of transmitted data. Together with fast FPGA processing, it gives a latency of less than 100 µs.
	The system's startup time for data recording should be less than a few microseconds.	The startup time for the proposed system is equal to its latency.

Table 6.1 – continued from previous page

Category	Parameter	Solution
Radiation tolerance	The detector should operate for five years before system replacement and withstand the maximum accumulated dose in the central position, corresponding to a particle fluence of 3×10^{14} cm^{-2} with carbon ions and 3×10^{15} cm^{-2} with protons [6].	Silicon sensors have proven to be adequately radiation resistant for HEP. Using HV-CMOS instead of standard CMOS technology and special layout techniques increases radiation tolerance due to a larger depletion region. The fast charge separation and collection improve tolerance to radiation-induced bulk damage [88].
Dimensions and material budget	The overall active area should be 25×25 cm^2 to cover a treatment area of 20×20 cm^2 [6].	It can be implemented by building a sensor matrix from several sensors 2×2 cm^2 each. The matrix consists of 13×13 electronically connected individual chips.
	The thickness of the outer housing shall be smaller than 5 cm and fit in a cylinder of 60 cm diameter [6].	The thickness of the sensors can vary from 100 µm to 700 µm; together with the electrical interconnection and mechanical support, it covers this requirement.
	The material budget of the total detector system, including several sensor layers and mechanical structure, should be less than 2 mm water equivalent [88].	The proposed design consists of 100 µm of silicon for the thinned detectors and 90 µm polyamide as a flex cable with 7 µm of aluminium per side and 200 µm carbon-fibre-reinforced polymer as the thermal and mechanical support.

To successfully operate in a high-rate ion or proton beam, HitPix has specific electronic design features, such as hit-counting pixels and frame-based readout. As mentioned before, contrary to HEP, the beam monitoring application for radiation therapy does not require the identification of individual particles for secondary particle tracking but hit counting for protons and light ions. Those high particle rates can be processed if each pixel is equipped with electronics that count the number of detected particles during a defined period of time [6]. One *hit* is registered when the induced charge of a particle passes through the sensitive volume of the device and its signal reaches over an established threshold, which is programmed with respect to noise. These hit counts are accumulated and are read out periodically. One resulting set of counts is called a *frame* and can be compared to a single picture of an imaging sensor [88].



(1) Carbon fibre plate;
(2) HV-CMOS sensors;
(3) Flex cables with aluminium traces;
(4) Electronics boards with FPGA for data processing;
(5) Stabilising frame.

Figure 6.6: Schematic 3D rendering of the full size proposed detector design [6].

Another essential feature of this design is on-chip projection calculation, which is connected to the device's medical application. Each patient is prescribed a specific dose, which must be administered as precisely as possible. Thus, the prompt interruption of the irradiation program after the dose has been delivered is vital for the safety and success of the treatment. The signal to stop the process must reach the control system in less than a second. One frame can be read out in two ways: either the hits from each individual pixel are recorded, or the data is summed up by on-chip calculation. Even though the first approach is ideal for developing the chip and quality assurance of the beam delivery system, in reality, reading out all the hit values of one frame is time-consuming. Pixel sensors can meet the safety timing restrictions only if data is being preprocessed on-sensor, so it can be transmitted to FPGA in reduced amounts. One solution is that each sensor of the detector generates a *projection* of the counted data in both x- and y-directions. This can be implemented by *adders* within each pixel, which add their own counts to the sum of the previous ones. The adders in the last pixel then hold the sum of all in this direction, giving the projection of the beam profile. This information can be read out quickly for real-time beam monitoring and interrupt signal generation [88].

All these design solutions, used in order to fulfil the application demand, form together the latency of the entire system. Figure 6.7 shows the simplified data flow diagram used in latency calculation.

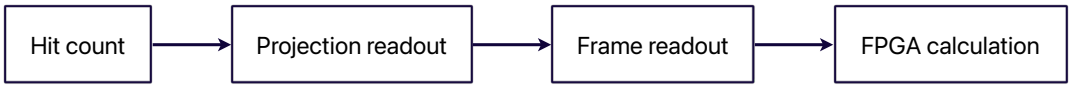


Figure 6.7: Data flow diagram for HitPix chip, including all main contributions to the latency.

The first step is hit count, which includes the zero-moment of the particle hit, charge collection in silicon, and digitisation of this charge into a count signal. Every single pixel has an 8-bit counter, which can compute 256 hits of particles. This full chain is a relatively fast process, which makes its temporal component (t_{hit}) less than 1 µs. After acquiring all counts and performing on-chip calculations, the information must be read out. This is called integration time *($t_{projection}$)*. It is typically 30 µs, and includes the time when the detector is counting hits and the time after which projections of all pixels within one frame are read out [6]. The next step is to transfer the data from the chip to the FPGA. 2 cm × 2 cm chip can have up to 98 × 98 pixels, 200 µm × 200 µm each. Each vertical or horizontal line of pixels gives one projection number. For one-dimensional x- (or y-) projection, the sum from each pixel in a respective column (or row) is calculated; however, in the case of two-dimensional (x- and y-) projections, only each second pixel is summed. Regardless of the mode used, the same amount of data is kept. Transferring this data takes 14 bits *(k)* per column (or row). As a result, the number of cycles of the FPGA *(C_{frame})* to transfer the hit counts of 98 columns *(n)* can be calculated as

$$C_{frame} = n \times k = 98 \times 14\ bit = 1372\ bit \tag{6.1}$$

Communication between the FPGA and the HitPix chip is designed to be up to 200 Mbit/s *(T)*. Therefore, frame readout time *(t_{frame})* can be calculated as

$$t_{frame} = \frac{C_{frame}}{T} = \frac{1372\ bit}{200 \times 10^6\ bit/s} = 6.86 \times 10^{-6}\ s = 6.86\ \mu s \approx 7\ \mu s \tag{6.2}$$

The last part of the data flow is a calculation on FPGA, which is expected to take 3 cycles per bit [92]. The total number of needed cycles *(C_{calc})* can be estimated as

$$C_{calc} = 3 \times k \times n = 3\ \frac{1}{bit} \times 14\ bit \times 98 = 4116 \tag{6.3}$$

With the clock speed *(S)* of the chosen FPGA (Artix 7) equal to 200 MHz, calculation time *(t_{FPGA})* can be determined as

$$t_{FPGA} = \frac{C_{calc}}{S} = \frac{4116}{200 \times 10^6\ Hz} = 20.58 \times 10^{-6}\ s = 21\ \mu s \tag{6.4}$$

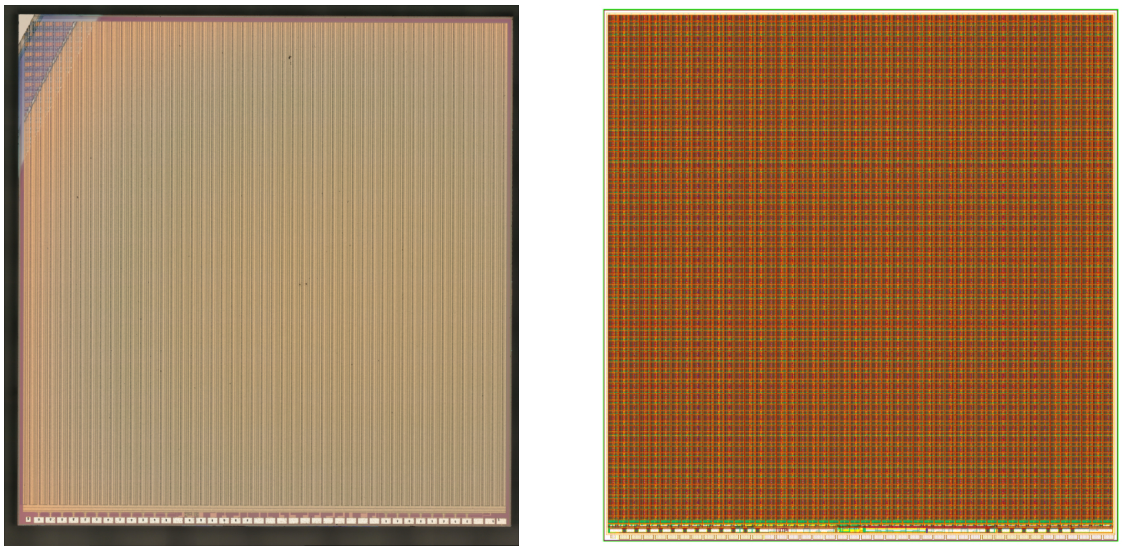
To summarise all this values together, latency *($t_{latency}$)* can be estimated as

$$t_{latency} = t_{hit} + t_{projection} + t_{frame} + t_{FPGA} = 1\ \mu s + 30\ \mu s + 7\ \mu s + 21\ \mu s = \mathbf{59\ \mu s} \tag{6.5}$$

6.3 Chip description

Figure 6.8 demonstrates a chip from the HitPix family. A picture made with a microscope is shown in Figure 6.8a. This exact chip was not part of the module assembly because of its location on the non-functional border of the wafer. However, this allows us to see a non-metalised part of the chip at the top left corner, where the pixel's structure is visible. The layout

of the same chip made in Cadence[2] is shown in Figure 6.8b. The chip consists of 48×48 pixels. The readout periphery and connection pads are located at the bottom of the layout.



(a) HitPix chip under the microscope. **(b)** Layout of the HitPix chip.

Figure 6.8: Picture and layout of HitPix chip, second generation.

The following significant point in the chip description is its electronics. Simplified schematics of the HitPix can be seen in Figure 6.9. Each pixel has the same architecture; it is split into three parts: sensing, analogue, and digital. After summing all hits, data is sent to the chip periphery and then transmitted to the FPGA. Data transferring is illustrated with lines and arrows, while signals are illustrated with solid lines. The sensor part of the chip, which is discussed in more detail in section 6.5, includes the sensor diode and reset circuit. It is biased using a circuit with a high equivalent resistance. This assures that the signal charge flows into the CSA [88].

After collecting the charge, it is passed to the amplifier. Charge-sensitive amplifier (CSA) consists of a folded cascode amplifier with PMOS input and a source follower. The gate of the amplifier's input transistor is AC-coupled to the sensor diode. This coupling is implemented with a PMOS transistor because the amplifier's input node has a different DC voltage than the sensor bias potential [78]. The CSA is AC-coupled to a comparator that compares the CSA output signal with a threshold. The threshold setting is a chip-wide global parameter that can also be adjusted for each comparator. Each pixel also contains a capacitive injection circuit to generate test signals for fast verifications. The output of the CSA of one pixel per column is connected to a test pad in the periphery [88].

The next block is a digital part of the pixel, which is the central driving part of on-chip projection calculations. The comparator output is connected to the digital circuit, which consists of an 8-bit *ripple counter* and a 15-bit *adder*. As previously discussed, the data transferring sequence is as follows: all hits are counted by *counters*, summed up by *adders* from first to last, transferred

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

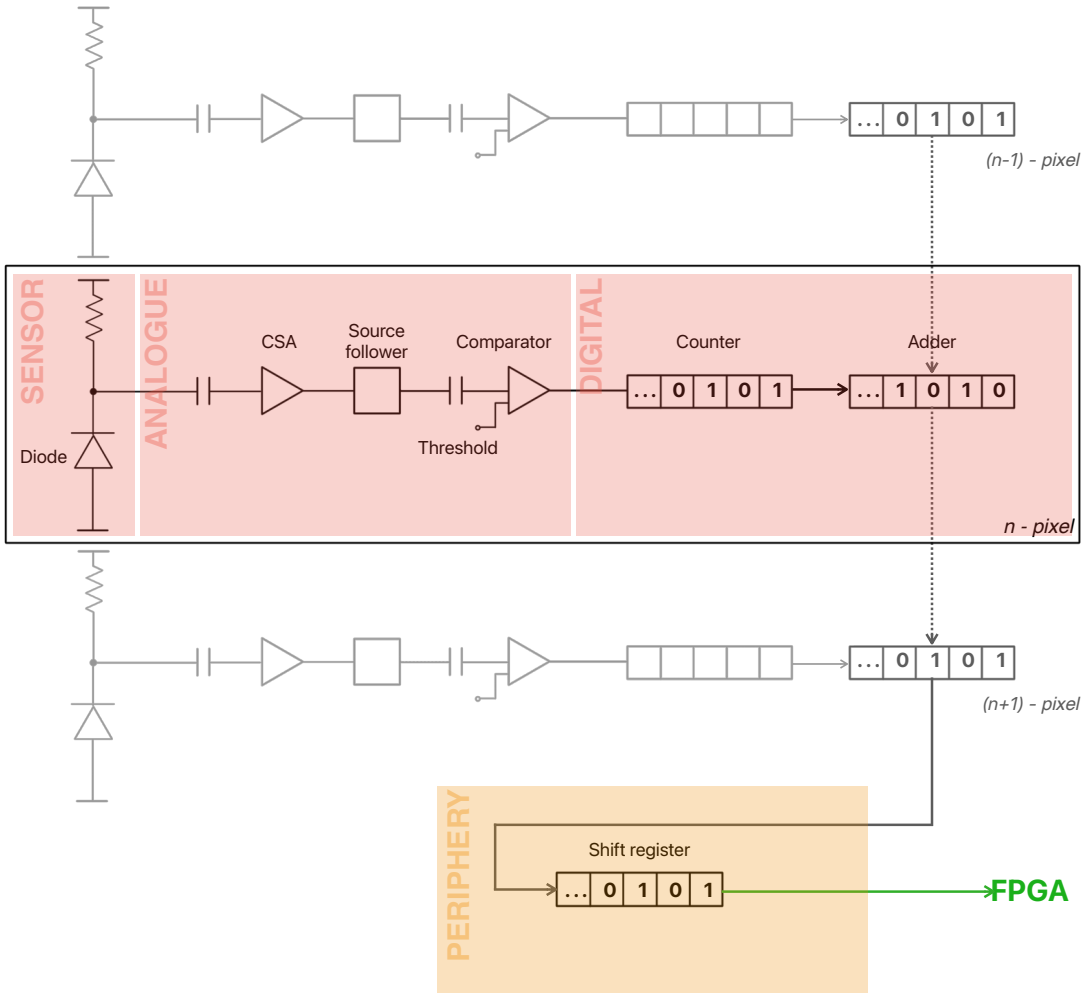


Figure 6.9: Simplified schematics of the pixel electronics, chip periphery and connections with other pixels of the matrix.

to the periphery, and then shifted to the FPGA. From this, it can be concluded that the primary role of the chip periphery is to receive configuration bits for the chip, store readout data, summed values, or counter states, and then shift this data out. The HitPix is prepared for daisy-chain readout of several detectors to reduce readout lines on the interconnecting cable when moving to a multi-chip array. This chain needs to be at least seven detectors long [88].

6.4 Chip design evolution

The HitPix chip family has undergone several iterations of improvements, with each version introducing critical upgrades to the design. The development of the initial version, HitPix1, focused predominantly on the electronic components of the design, as detailed in Section 6.2. This foundational work set the stage for subsequent versions.

With HitPix2, the choice was made towards the NMOS amplifier, replacing the PMOS amplifier used in HitPix1. This modification aimed to test the faster performance of linear NMOS transistors against enclosed ones. Further, the decision to return to PMOS, as it is more radiation tolerant, was made. In parallel, significant attention was devoted to enhancing the sensor's functionality using TCAD Synopsys simulation tools. These simulations provided valuable insights into key aspects of sensor behaviour, including depletion regions (see Section 6.6), electrical breakdown analysis (see Section 6.7) and charge transport and collection (see Section 6.8).

The findings from these simulations were vital in guiding the development of HitPix3. This chip version underwent substrate selection and pixel configuration adjustments, both of which were proposed as the simulation results. A major focus of this iteration was addressing efficiency losses, as explored in Chapter 8. The design improvements aimed to mitigate issues related to charge loss and signal degradation, which were identified as significant challenges in earlier versions.

Finally, the lessons learned throughout this iterative process were all summed up in the design of HitPix4, which was implementation in AMS ah18 technology. This version represents the most advanced and robust solution within the HitPix family. It addresses the limitations of its predecessors while introducing further enhancements to sensor performance and integration. The downside of this chip design is that the used technology does not support the ISO pixel version and can provide only six metalisation layers.

A comparison of the key parameters for all versions of the HitPix chip family is presented in Table 6.2, highlighting the specific versions discussed in this section. The table summarises the evolution of the HitPix chip family so that it is easy to follow the key trends and improvements across versions. Initially, the chips featured smaller dimensions of 5×5 mm^2, which made the production cost lower and gave the opportunity to test the chip concept. Subsequent versions, starting with HitPix2, introduced larger chip sizes, reaching 2×2 cm^2 in HitPix4. The pixel matrix size also expanded significantly, from 24×24 pixels in HitPix1 to 96×96 pixels in HitPix4, while keeping the active area of the chip the same. Another reason for this subsequent increase in the pixel matrix is to test readout approaches and reach the final experiment requirement steadily.

Table 6.2: Comparison of HitPix chip family versions. Highlighted are the versions of the chips discussed in this chapter.

Parameter	**HitPix1 Std** [93]	**HitPix1 ISO** [94]	**HitPix2** [95]	**HitPix3** [95]	**HitPix4** [96]
Company	TSI	TSI	TSI	TSI	AMS
Chip technology	180 nm	180 nm	180 nm	180 nm	180 nm
Amplifier	PMOS	PMOS	NMOS	PMOS	PMOS
Chip size	5×5 mm^2	5×5 mm^2	1×1 cm^2	1×1 cm^2	2×2 cm^2

Table 6.2 – continued from previous page

Parameter	**HitPix1 Std** [93]	**HitPix1 ISO** [94]	**HitPix2** [95]	**HitPix3** [95]	**HitPix4** [96]
Pixel size	$200 \times 200\ \mu m^2$	$200 \times 200\ \mu m^2$	$200 \times 200\ \mu m^2$	$200 \times 200\ \mu m^2$	$200 \times 200\ \mu m^2$
Chip matrix size	24×24 pixels	24×24 pixels	48×48 pixels	48×48 pixels	96×96 pixels
Thickness	100 µm / 800 µm	100 µm / 800 µm	150 µm / 700 µm	150 µm / 700 µm	150 µm / 700 µm
Substrate	10 Ωcm / 300 Ωcm	10 Ωcm / 300 Ωcm	300 Ωcm	370 Ωcm	370 Ωcm
Pixel version	Standard	ISO	Standard	ISO	Standard
Sensitive area of the pixel	70%	100%	70%	100%	70%
Active area of the chip	92.16%	92.16%	92.16%	92.16%	92.16%
Readout	In-pixel counting, column adding, one dimension projection	In-pixel counting, column adding, one dimension projection	In-pixel counting, column adding, one dimension projection	In-pixel counting, column and raw adding, two dimensions projection	In-pixel counting, column and raw adding, two dimensions projection

6.5 Sensor design and TCAD modelling

Before examining the TCAD simulations, it is essential to understand the sensor's design features and layout. As described before, in Section 6.2, each pixel consists of a sensor diode connected to a CSA. The CSA is AC-coupled to a comparator, which compares the CSA output signal with a predefined threshold. The comparator output is connected to the digital circuitry [88]. The schematic is illustrated in Figure 6.9.

The HitPix chip family includes two distinct design approaches: the Standard (STD) design and the ISO (Isolated Shallow nwells) design. While the electronics are largely similar, they differ in the shape of the deep nwells, additional deep pwells, and minor adjustments to shallow pwells. These variations affect the pixel sensitivity and signal isolation and play a significant role in the system's efficiency.

In the Standard design, each pixel layout is divided into three regions: the sensor electrode, the analogue electronics containing the CSA and comparator, and the digital electronics containing *adders* and *counters*. As shown in Figure 6.10, these components are implemented

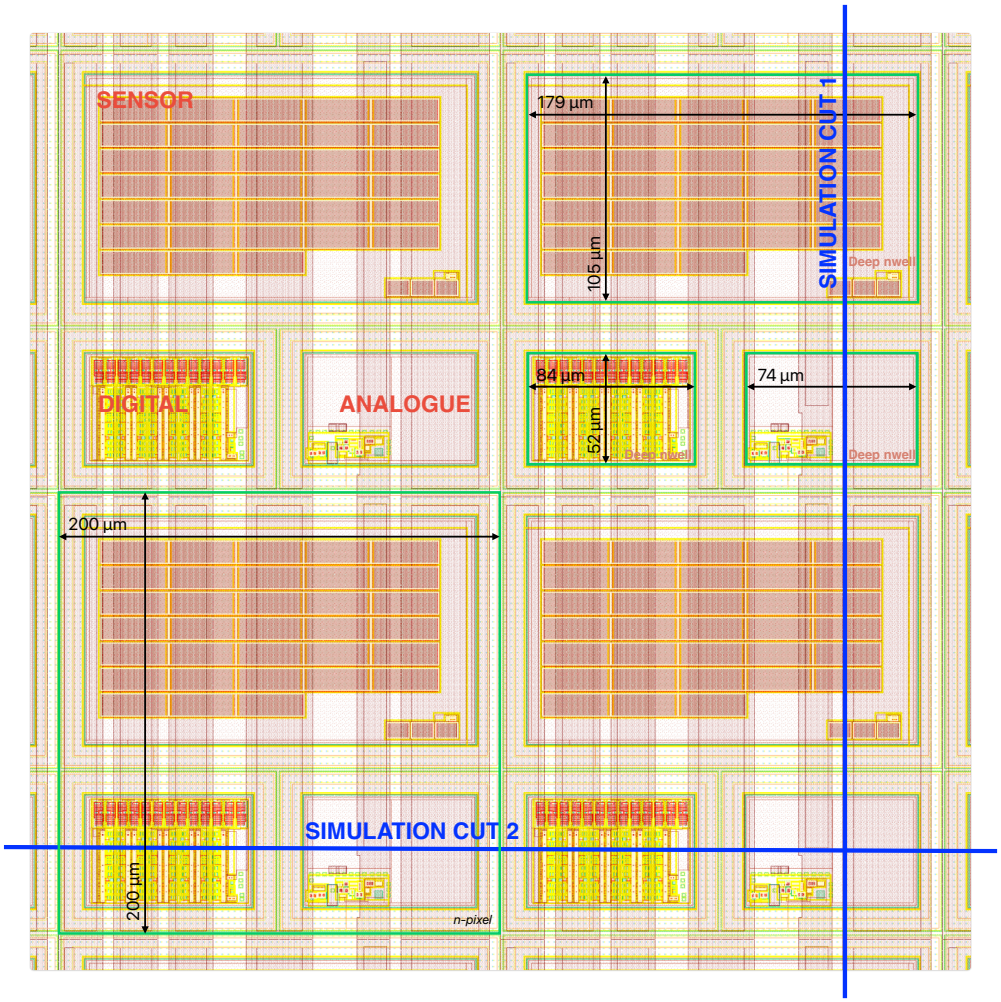
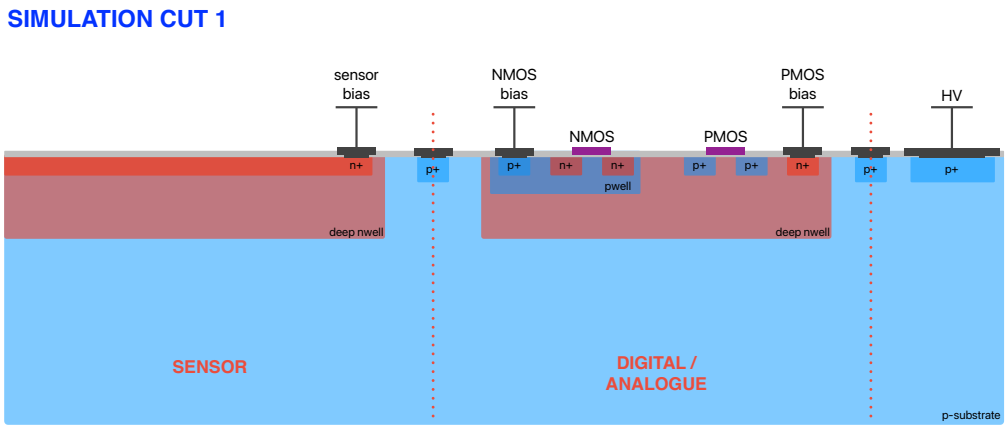


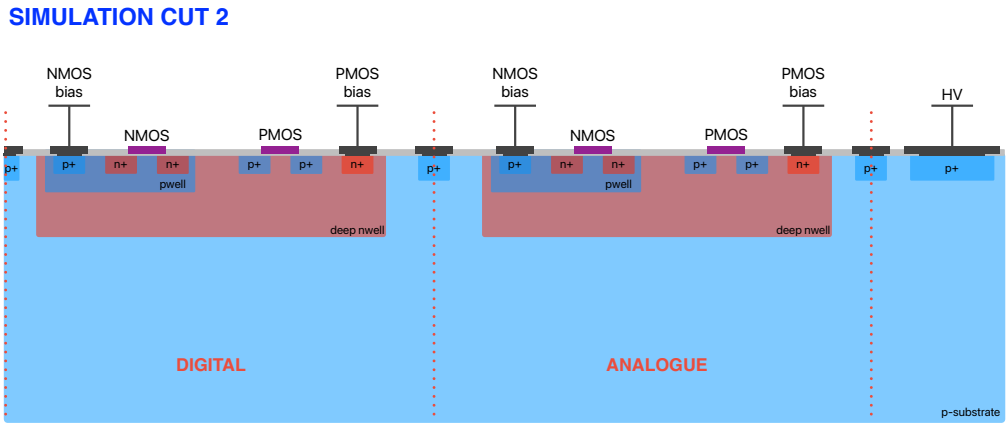
Figure 6.10: Layout of four pixels in the Standard HitPix1 version. This figure highlights the layout of the sensor, analogue, and digital regions, along with dimensions and simulation cuts. Nwells and pixels are separated with p−type of dopings, called p-stop, shown by a green line.

using three separate wells. The analogue and digital electronics are embedded in two deep nwells, which are biased at 1.8 V. The sensor nwells is similarly biased, using a high-resistance circuit to ensure the signal charge flows to the CSA. If a particle ionises the substrate below the electronics wells, the signal charge generated there flows directly to the positive supply without amplification. All nwells are separated with p−type doping, which is called p-stop, to prevent cross-talk. Pixels are also separated by a p-stop, but also by an additional type of p−type doping, which at the same time serves as isolation and biasing grid. Cross-sectional views of the Standard pixel design, presented in Figure 6.11, were used in TCAD simulations to examine the electrical characteristics of these regions.

In contrast, the ISO design uses a single deep nwell that acts as both the sensor electrode and the housing for the nwells and pwells containing the analogue and digital electronics. This arrangement ensures that the entire pixel area is active, achieving full sensitivity at 100%.



(a) Cross-section including the sensor nwells and one electronics nwells.



(b) Cross-section focusing on the electronics-only nwells.

Figure 6.11: Cross-sections of the Standard HitPix1 pixel design, illustrating the arrangement of sensor and electronics regions for TCAD simulations.

The layout of this chip version is shown in Figure 6.12. Additional deep pwells isolate the PMOS transistors from the sensor deep nwells, minimising capacitive crosstalk of signals to the sensor and preventing shorting of the sensor nwells with biasing. In this design, pixels are also separated by both p-stop and biasing $p-$type grid contact.

Cross-sectional views of the ISO pixel design, shown in Figure 6.13, were also used for TCAD simulations. These views highlight the sensor's integration with the electronics and demonstrate the improved isolation enabled by the deep pwells. The ISO design significantly enhances charge collection efficiency by ensuring that all signal charge is directed to the sensor.

To evaluate and further develop the HitPix chip, TCAD simulations were performed across both designs and different substrate resistivities. Due to computational resource limitations, simulating the entire chip was not feasible; instead, specific pixel regions were selected for detailed analysis. For the HitPix Standard design, two areas were chosen: the middle of the sensor matrix and the edge of the sensing region. These regions represent contrasting

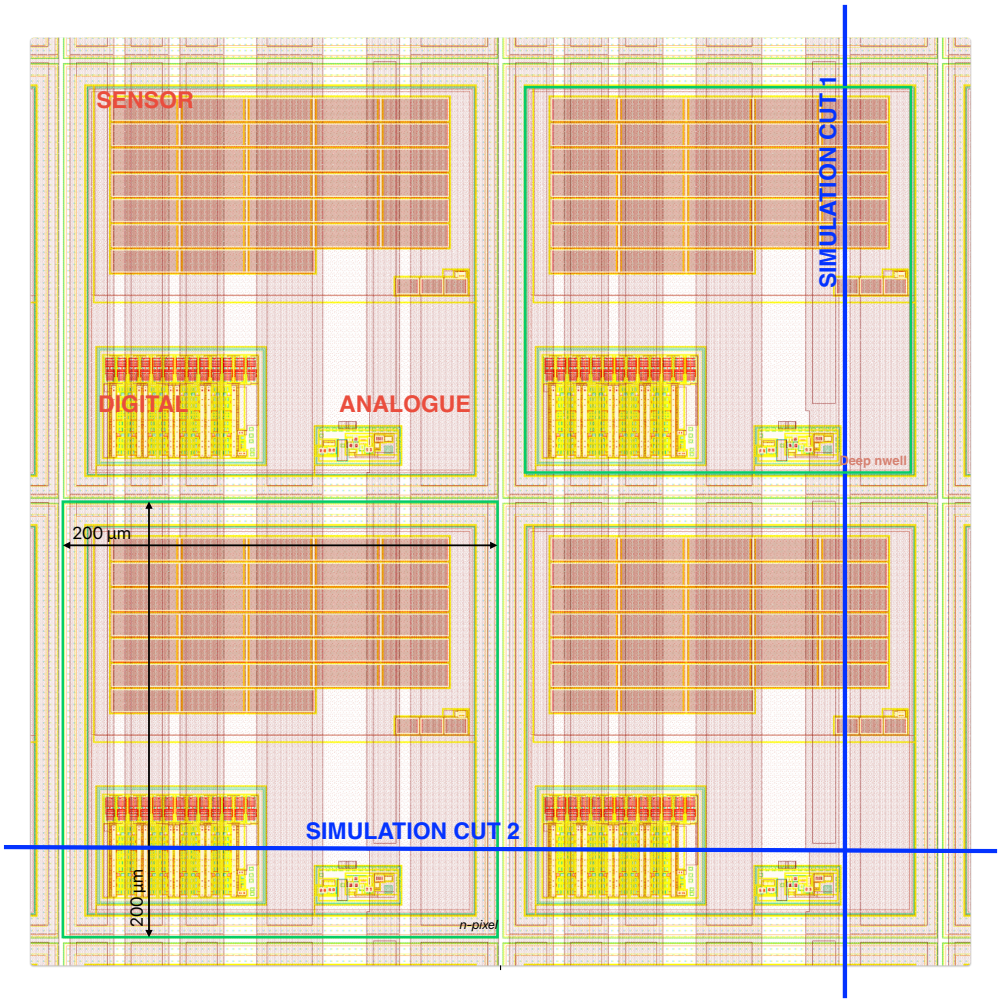
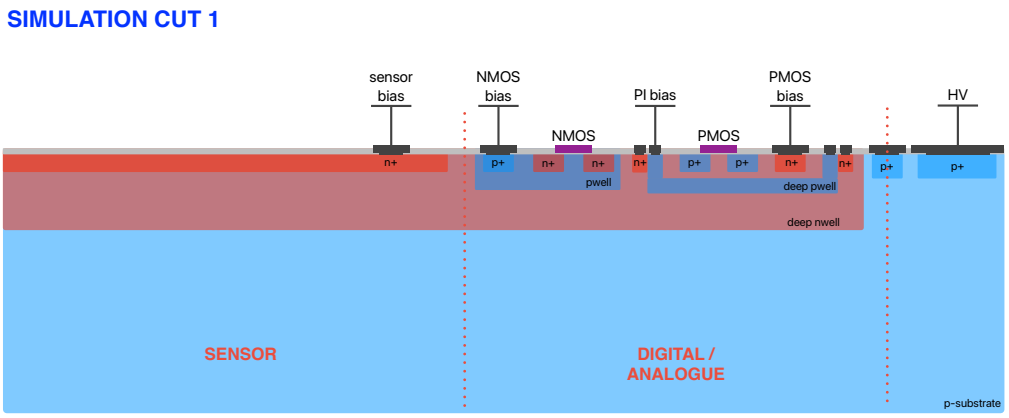


Figure 6.12: Layout of four pixels in the ISO HitPix version. This design employs a single deep nwell for both charge collection and housing electronics, with additional deep pwells for isolation. Pixels are separated with p–type dopings, called p-stop, shown by a green line.

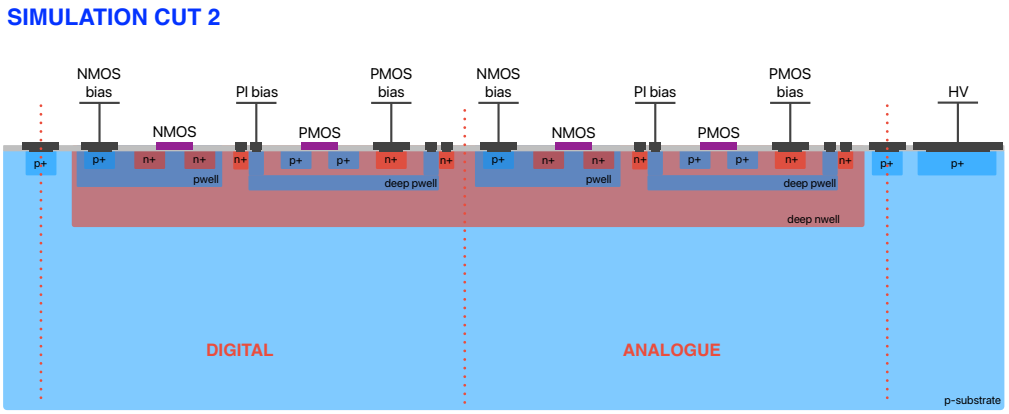
conditions within the chip and can be used together to create a precise reconstruction model of its functionality. To mimic the environment in the middle of the chip, simulations were conducted with backplane biasing. In contrast, for the sensor edge, a more realistic top-biasing approach was applied. Both scenarios are illustrated in Figure 6.14.

Due to the complexity of its structure and computational constraints, the behaviour of the HitPix ISO design was effectively modelled by simulating a single pixel. The primary focus in this version was on the insides of the deep nwell area, where its multi-doping structure makes it the most critical part of the pixel. The TCAD layout of the ISO design is presented in Figure 6.15.

Across all simulations, certain common parameters can be observed due to the production capabilities. These included a sensor and electronics bias voltage of 1.8 V, a sensor thickness of 300 µm, shallow trench isolation (STI) depths of 300 nm within the structure and 800 nm on the



(a) Cross-section showing the sensor area and one electronics nwells.

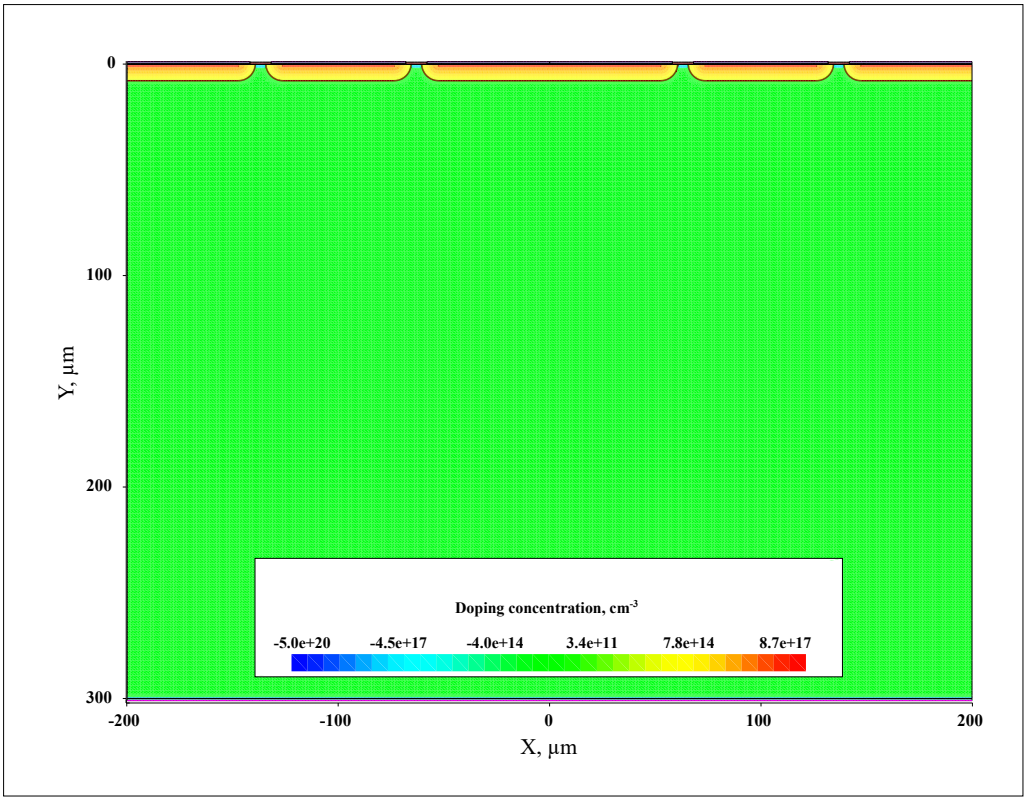


(b) Cross-section focusing on digital and analogue regions.

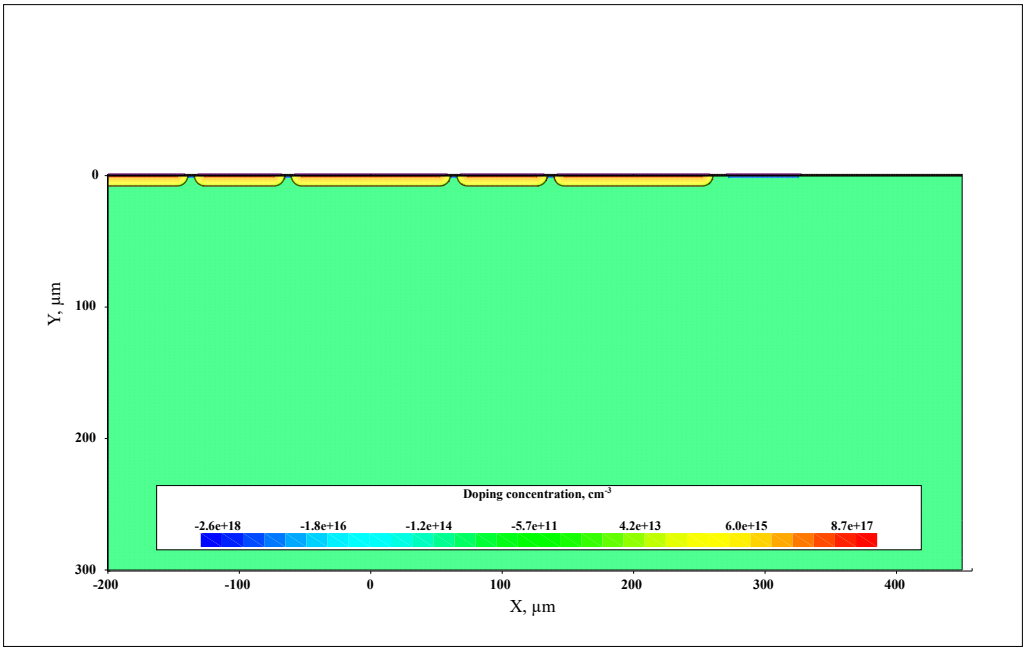
Figure 6.13: Cross-sections of the ISO HitPix pixel design, highlighting the integration of sensor and electronics within a single deep nwells.

surface, and floating p-stop configurations. Additionally, the metal layers were etched to match the designs specified in Cadence layouts. For a detailed description of the TCAD simulation process, refer to Chapter 4, titled "Simulations of Semiconductor Devices — Technology Computer-Aided Design." Special emphasis was placed on the sensor periphery simulation (see Section 6.7), as it plays a crucial role in preventing premature breakdown.

Simulations, which were conducted for all of these selected areas, included electric field distribution and electrostatic potential analysis to assess breakdown behaviour. Substrate resistivity was optimised to balance the depletion region size, breakdown voltage, and charge collection efficiency. Furthermore, charge collection and timing simulations, using simulations of a 200-300 MeV/n proton beam, were conducted to estimate sensor performance and address efficiency losses, as detailed in Chapter 8. The next sections present these simulations, which provide valuable insights for refining the designs and improving overall performance.



(a) TCAD simulation plot of the middle of the sensor matrix, where backplane biasing is used to imitate uniform operating conditions.



(b) TCAD simulation plot of the edge of the sensor, where top-biasing is applied to model realistic boundary conditions.

Figure 6.14: TCAD simulation configurations for the HitPix Standard design. The middle of the chip and the edge of the sensor are examined to create a real-life model of the produced chip.

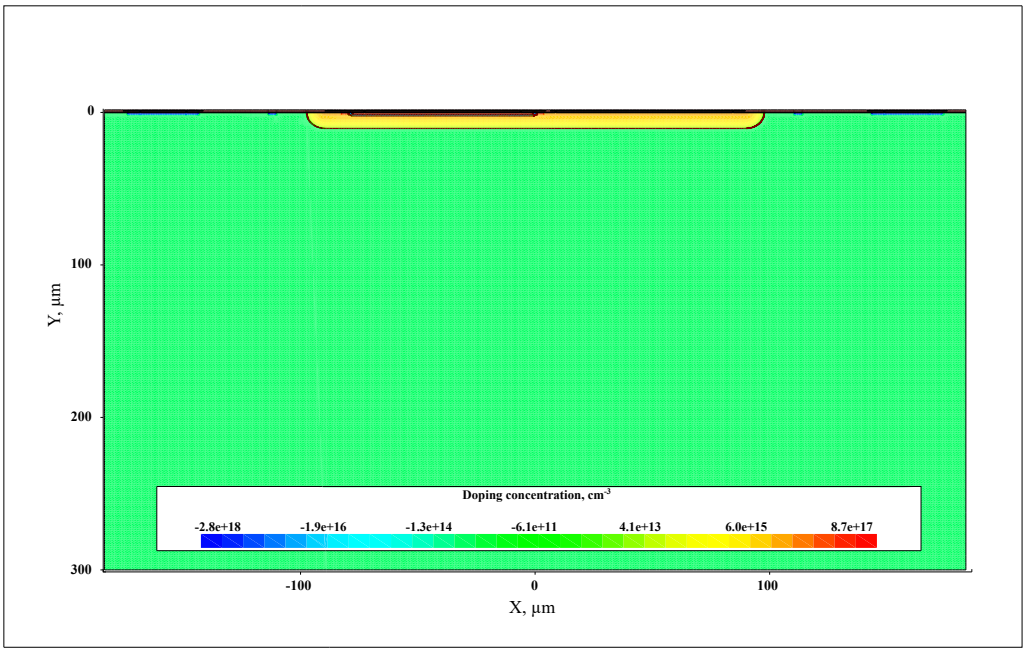


Figure 6.15: TCAD simulation layout for the HitPix ISO design. The focus is on the deep n-well region, which integrates charge collection and electronics, showcasing the complex interactions within the structure.

6.6 Operational conditions and choice of substrate

Silicon sensors operate based on several critical parameters that influence their performance, stability, and reliability. The choice of operational conditions, such as bias voltage and temperature, is crucial in determining sensor efficiency and long-term stability. For the HitPix sensor, these parameters must be carefully optimised to ensure reliable operation, particularly given its medical use.

One of the most important operational parameters is the *bias voltage* (V_{bias}), which determines the depletion region width and directly influences charge collection efficiency (CCE), capacitance, and leakage current. A higher bias voltage expands the depletion region, improving CCE and reducing capacitance, which enhances the signal-to-noise ratio (SNR). However, increasing the bias voltage also results in higher leakage current (I_{leak}), which contributes to additional noise and potential overheating [97].

Temperature is another critical factor in sensor performance, as a leakage current is highly temperature-dependent. In high-radiation environments, active cooling techniques are often required to mitigate the effects of radiation-induced leakage current. Additionally, the *capacitance* (C) of the sensor, which is inversely proportional to the depletion depth, plays a key role in defining the sensor's response time and noise characteristics. Lower capacitance improves SNR, making it particularly beneficial for low-noise applications [97].

Radiation tolerance is another fundamental aspect of sensor operation, particularly in high-energy physics, space and medical applications. Exposure to ionising and non-ionising radiation can cause charge-carrier trapping in the bulk material, increased leakage current and full depletion voltages, as well as long-term degradation [98]. To maintain long-term sensor performance, depletion depth, bias voltage, and temperature control must be carefully optimised.

While several parameters impact sensor operation, some, such as pixel pitch, pixel type (see Section 6.5), and pixel thickness, are fixed by design choices. Consequently, this study focuses on optimising bias voltage by selecting the most suitable sensor substrate while also evaluating power consumption in operational conditions.

6.6.1 Substrate selection and depletion region analysis

Many operational parameters depend on the chosen substrate, including crystal orientation, doping uniformity, and defect density. The most common silicon orientations are (100) and (111), with (100) being preferred for CMOS applications due to its superior interface properties. The Si/SiO_2 interface quality is particularly important in minimising unwanted charge trapping and leakage current.

The effects of resistivity and crystal orientation on leakage current and radiation response characteristics have been extensively studied [99]. Sensors with (111)-oriented substrates exhibit higher leakage currents than those with (100) orientation due to an increased interface trap density at the Si/SiO_2 interface. Additionally, high-resistivity substrates tend to have larger

leakage currents than low-resistivity ones due to their wider depletion regions at the same bias voltage. However, in (100)-oriented substrates, the leakage current in low-resistivity materials can become larger than in high-resistivity materials under high reverse bias conditions [99].

For the HitPix sensor, a bias voltage of -80 V was selected based on prior experience with HV-CMOS designs and validated through simulation. This value provides a stable depletion region while allowing for future adjustments over the sensor's operational lifetime. Additionally, considering the medical application of HitPix, the bias voltage must remain within a safe range for human interaction.

TCAD simulations were performed to evaluate substrates with different resistivities. Figure 6.16 presents a plot comparing the depletion regions of different substrates as a function of bias voltage. The results show that a 20 Ω·cm substrate does not have a sufficient depletion region, leading to low signal efficiency. Thus, the choice is narrowed to either a 300 Ω·cm or 5k Ω·cm substrate.

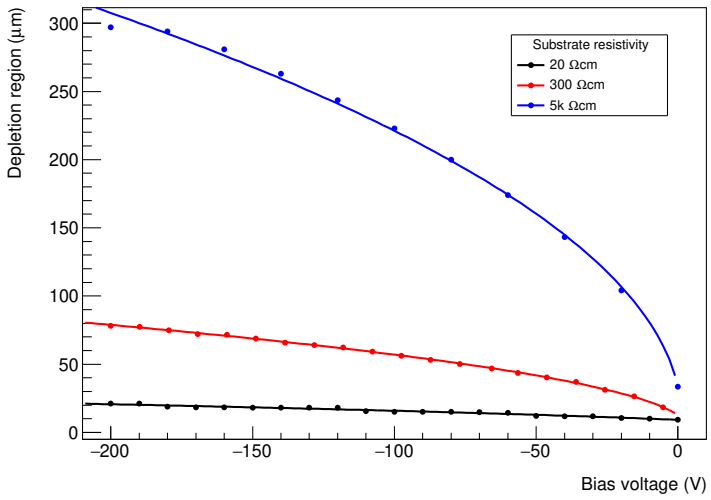
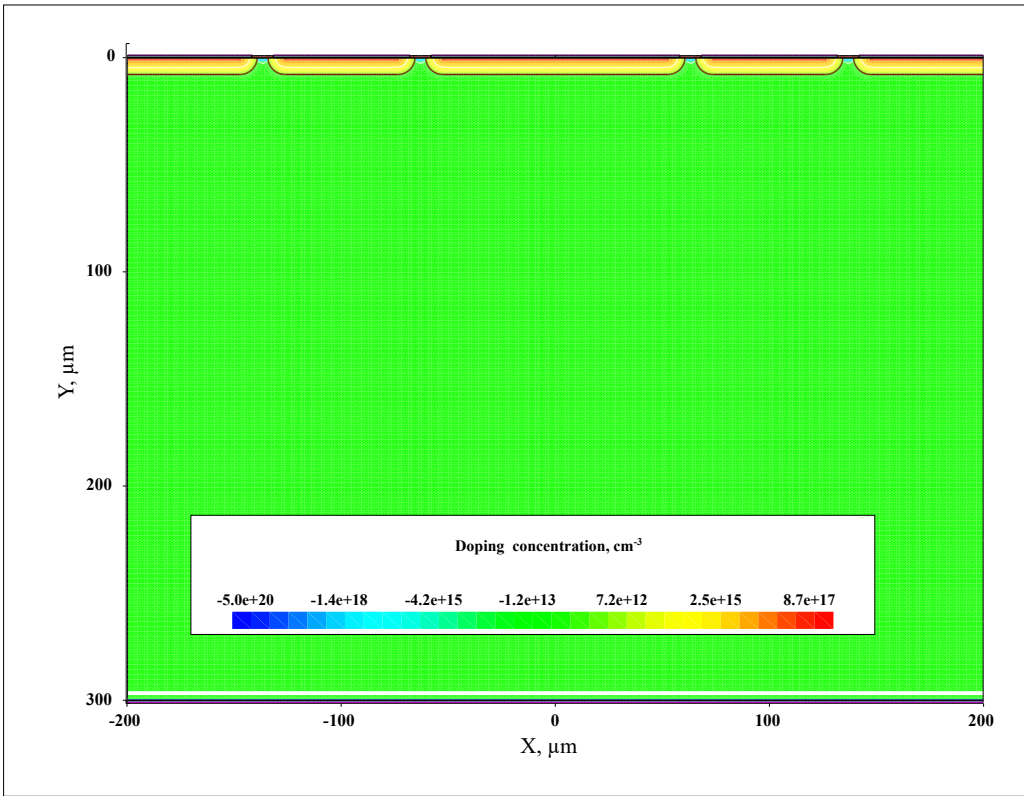


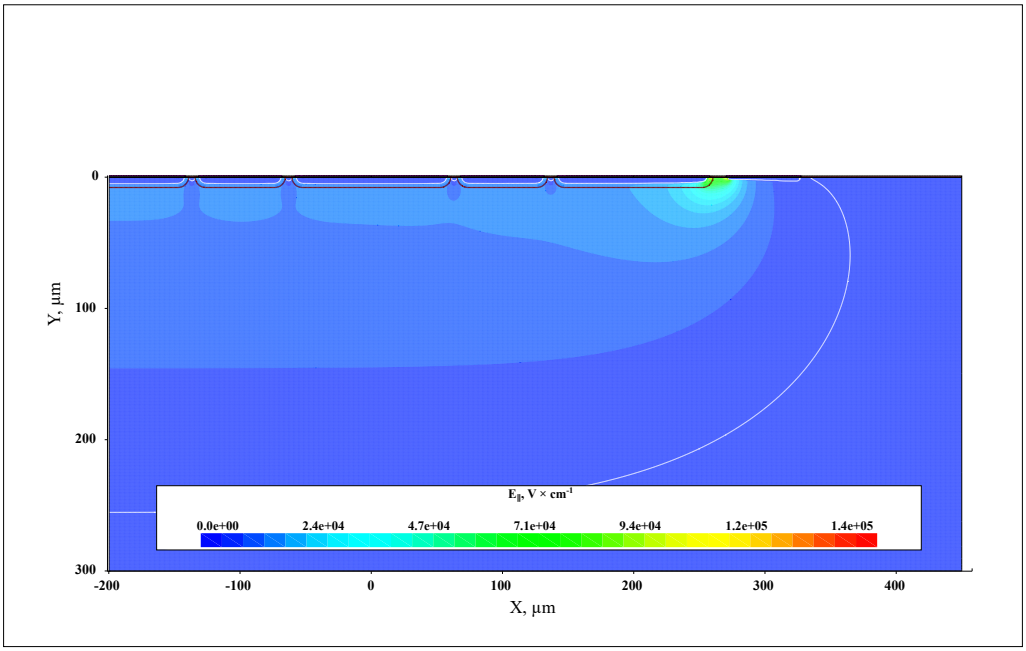
Figure 6.16: Comparison of depletion region depth for different substrate resistivities as a function of bias voltage. The 20 Ω·cm substrate does not provide sufficient depletion depth, making it unsuitable for this application. The 300 Ω·cm and 5k Ω·cm substrates show greater depletion, but further analysis is needed to determine their suitability.

A 5k Ω·cm substrate achieves full depletion at lower voltages due to its lower impurity concentration. Figure 6.17a shows that at -200 V, the depletion region extends fully. However, as can be seen in Figure 6.17b, at -133 V, a high electric field is observed near the sensor edge, increasing the risk of premature breakdown.

In contrast, a 300 Ω·cm resistivity substrate provides a stable and gradually increasing depletion region. Although full depletion is challenging without approaching breakdown voltage, it offers reliable performance over time. Figure 6.18 illustrates the depletion region at -80 V, which remains stable and allows for gradual bias voltage adjustments as the sensor ages.



(a) TCAD simulation of a 5k Ω·cm resistivity substrate at -200 V, showing full depletion. The white region represents the depletion area, and the colour scale indicates doping concentration.



(b) Electric field distribution at the sensor edge for a 5k Ω·cm resistivity substrate at -133 V. A strong electric field (in green) is observed, increasing the risk of breakdown.

Figure 6.17: Comparison of depletion and breakdown behaviour in a 5k Ω·cm resistivity substrate. While it provides a large depletion region at lower voltages, it is more prone to breakdown under operational conditions.

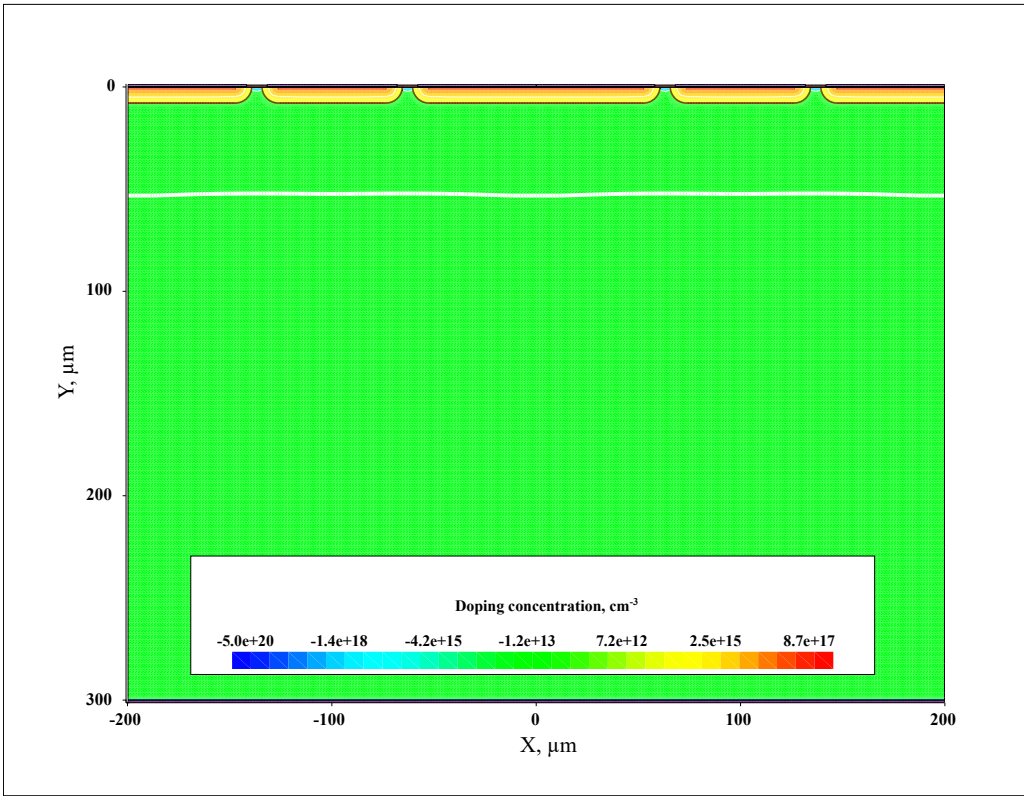


Figure 6.18: TCAD simulation of a 300 Ω·cm resistivity substrate at -80 V, showing a stable depletion region. The white region represents the depletion area, and the colour scale indicates doping concentration. The structure ensures long-term reliability and allows for bias voltage adjustments as the sensor ages.

6.6.2 Power consumption evaluation

Another critical factor to consider in sensor operation is power consumption, which determines the amount of electrical power required for proper functionality. In silicon sensors, power consumption is typically measured in watts (W) and comprises various contributions. Minimising power consumption is essential for the HitPix sensor due to its medical application, stringent material budget constraints, and operation within a particle beam, where cooling options are limited.

The total power consumption P_{total} of the sensor can be broken down into three main components:

$$P_{total} = P_{electronics} + P_{beam-on} + P_{ageing} \tag{6.6}$$

The first component, $P_{electronics}$, accounts for the power required by analogue and digital electronics. This value serves as a baseline estimate and can be determined during the design phase. The current consumption of the charge-sensitive amplifier (CSA) and the comparator in the HitPix sensor is adjustable. A typical current draw for the CSA is 4.7 µA, while the comparator consumes 2.3 µA [88]. Given the nominal main voltage of 1.8 V and an amplifier voltage of 1.0 V, the total per-pixel power consumption is calculated as:

$$P^{\,pixel}_{\,electronics} = (1.8\ V \times 2.3\ \mu A) + (1.0\ V \times 4.7\ \mu A) = 8.84\ \mu W \tag{6.7}$$

For the full sensor matrix of HitPix2 or HitPix3, which is 48 × 48 pixels, this value scales accordingly:

$$P^{\,sensor}_{\,electronics} = 8.84\ \mu W \times 48 \times 48 = 20.4\,mW \tag{6.8}$$

The second contribution, $P_{beam-on}$, arises from the additional leakage current generated when the sensor operates under a particle beam. The relationship between substrate resistivity and leakage current is not straightforward, as it is influenced by factors such as mechanical defects, bulk damage, and Si/SiO_2 interface properties. However, under ideal conditions, higher resistivity leads to larger depletion regions, which may increase space charge volume and, consequently, leakage current.

When exposed to high-intensity particle beams, the sensor experiences an increase in leakage current. This effect is particularly relevant in irradiation applications, such as those at the Heidelberg Ion Beam Therapy Center (HIT). At HIT, particle rates reach up to 8×10^7 particles per second for carbon ions and 3.2×10^9 particles per second for protons. Such intensities are sufficient to induce several orders of magnitude changes in the sensor's leakage current. To facilitate comparison across different beam energy settings, leakage currents must be rescaled, as described in [100].

The estimated rescaled leakage current (I_{leak}) during proton beam operation is:

$$I_{leak} = -5\ \mu A \tag{6.9}$$

Considering a bias voltage of -80 V, the power consumption associated with beam-induced leakage current is given by:

$$P_{beam-on} = (-5\ \mu A) \times (-80\ V) = 0.4\ mW \tag{6.10}$$

The third contribution, P_{ageing}, accounts for the increase in leakage current due to radiation damage over prolonged operation. Over time, irradiation leads to bulk defects, charge trapping, and an increase in leakage current. To estimate the long-term effects, TCAD simulations were conducted for different substrates to assess leakage current growth under radiation exposure. Figure 6.19 presents the simulated leakage current for substrates with resistivities of 20 Ω·cm, 300 Ω·cm, and 5k Ω·cm.

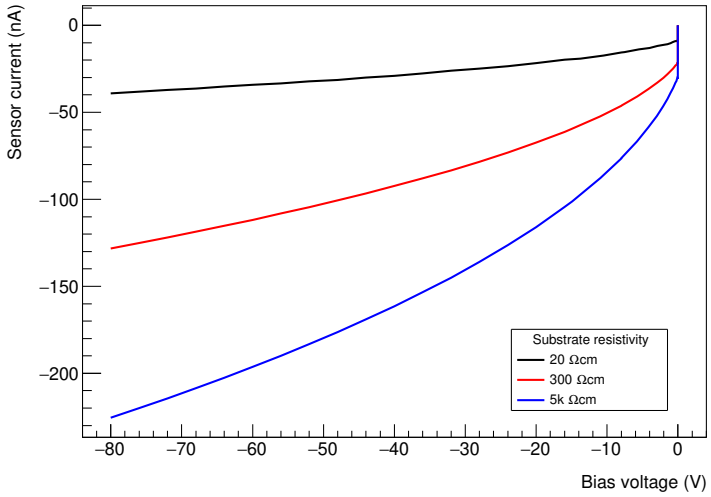


Figure 6.19: Simulated leakage current as a function of bias voltage for substrates with different resistivities: 20 Ω·cm, 300 Ω·cm, and 5k Ω·cm. The data illustrates that high-resistivity substrates exhibit higher leakage currents due to their larger depletion widths.

The TCAD simulations estimate that for a 300 Ω·cm substrate, the radiation-induced leakage current per pixel after prolonged operation under fluence is:

$$I_{leak}^{ageing} = -130\ nA \tag{6.11}$$

Thus, the per-pixel power consumption due to radiation damage is:

$$P_{ageing}^{pixel} = (-130\ nA) \times (-80\ V) = 10.4\ \mu W \tag{6.12}$$

For the full sensor matrix:

$$P_{ageing}^{sensor} = 10.4\ \mu W \times 48 \times 48 = 23.96\ mW \tag{6.13}$$

Summing all components, the total power consumption of the HitPix sensor after years of operation in a proton beam is estimated as:

$$P_{total} = P_{electronics} + P_{beam-on} + P_{aging} = 20.4\ mW + 0.4\ mW + 23.96\ mW = 44.76\ mW \quad (6.14)$$

Resume: the results indicate that a 300 Ω·cm substrate offers the most stable depletion region while maintaining power consumption within an acceptable range. In contrast, high-resistivity substrates such as 5k Ω·cm provide larger depletion regions but are associated with increased leakage current and a higher risk of premature breakdown. Radiation-induced ageing significantly impacts long-term power consumption, which highlights the need to select a substrate that ensures both reliability and operational stability under extended irradiation. The analysis of power consumption in the HitPix sensor demonstrates the importance of balancing depletion efficiency, leakage current, and long-term power stability. The final calculated power consumption estimation for HitPix2 or HitPix3, after considering all contributions, is approximately 44.76mW. This value is well within acceptable limits for the intended application, providing confidence in the long-term viability of the selected operational conditions.

6.7 Electric fields and breakdown simulations

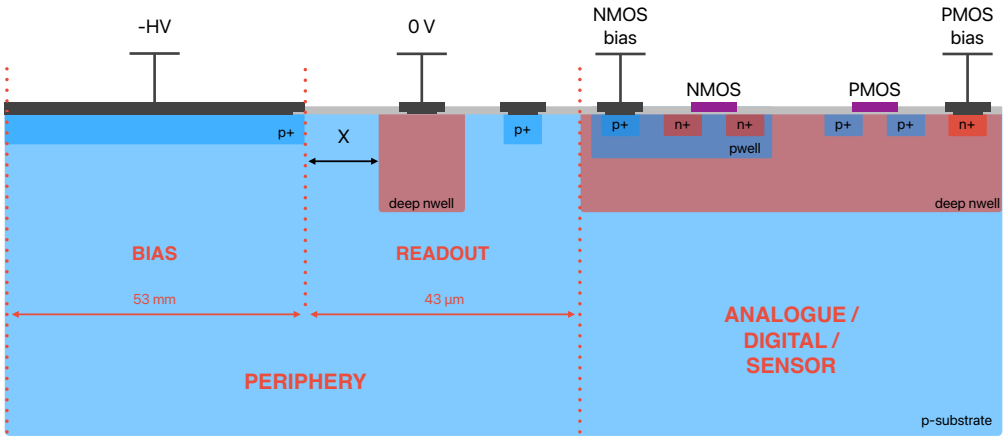
Understanding the distribution of electric fields within the sensor is crucial for identifying areas prone to impact ionisation or premature breakdown. Sensor designs can be optimised by analysing these regions to mitigate these effects and enhance device performance.

When a silicon sensor is irradiated, it undergoes radiation-induced damage, which can lead to the formation of bulk defects, interface traps, and charge build-up in the silicon lattice. These defects significantly affect the breakdown voltage, which is the voltage at which a sharp increase in leakage current occurs due to avalanche breakdown. Two main effects of irradiation are surface damage at the Si/SiO_2 interface and the generation of deep-level defects that act as charge generation centres, increasing leakage current and lowering the breakdown voltage threshold. These phenomena lead to an observable reduction in the post-irradiation maximum bias voltage, confirming the degradation of breakdown voltage due to radiation-induced defects.

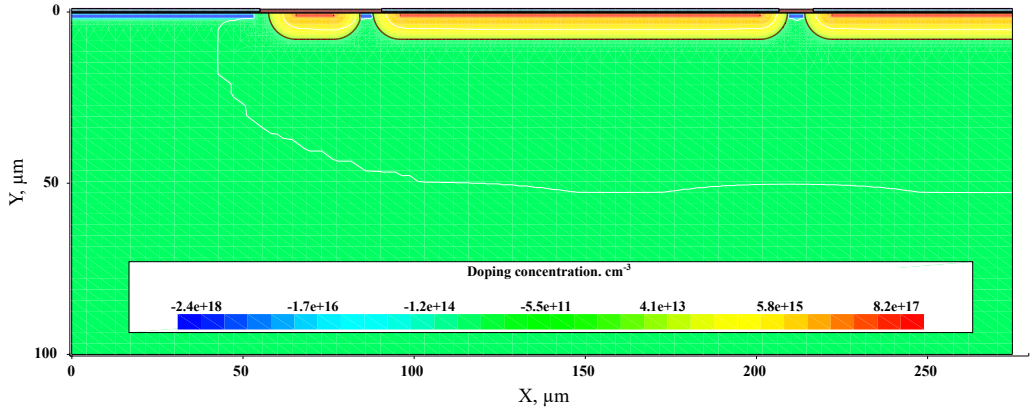
The HitPix1 sensor exhibited breakdown at approximately -60 V. Design improvements in HitPix2 allowed breakdown to reach an expected -120 V, where the simulated value was at -170 V. However, post-irradiation tests revealed that radiation exposure significantly affected its stability. The most sensitive part of the sensor to electrical breakdown is the chip periphery, where critical structures such as the readout nwell and bias pwell are located. Readout configuration is schematically illustrated in Figure 6.9. The space between the high-voltage (HV) contact and the nwell containing the readout electronics is particularly vulnerable to breakdown.

6.7.1 Breakdown optimisation through design adjustments

In the HitPix2 version, the original spacing between the bias pwell and the readout nwell was designed to be 12.5 µm. A cross-sectional view of the periphery is shown in Figure 6.20a, while Figure 6.20b presents a corresponding TCAD simulation with doping concentrations. The simulation was performed with a bias voltage of -80 V, resulting in a depletion region extending to 52.5 µm.



(a) Schematic cross-section of the periphery in the Standard HitPix design, showing the spacing between the bias pwell and readout nwell (marked by X).



(b) TCAD simulation of the same periphery region, illustrating the doping concentration distribution at a bias voltage of -80 V. The depletion region extends to 52.5 µm.

Figure 6.20: Cross-section and TCAD simulation of the periphery region in HitPix Standard, highlighting the spacing between the bias pwell and readout nwell.

To mitigate breakdown risks, simulations were conducted to explore different spacing distances and their impact on breakdown voltage. The objective was to determine the minimal distance between the pwell and nwell that still provides a stable breakdown voltage.

The first step involved simulating various spacing distances and assessing the corresponding improvements in breakdown voltage. Figure 6.21 presents the simulated current-voltage

(IV) characteristics for different spacing distances. These results, while idealised, provide a useful trend indicating that beyond a spacing of 15 µm, further increases in distance do not significantly improve breakdown voltage.

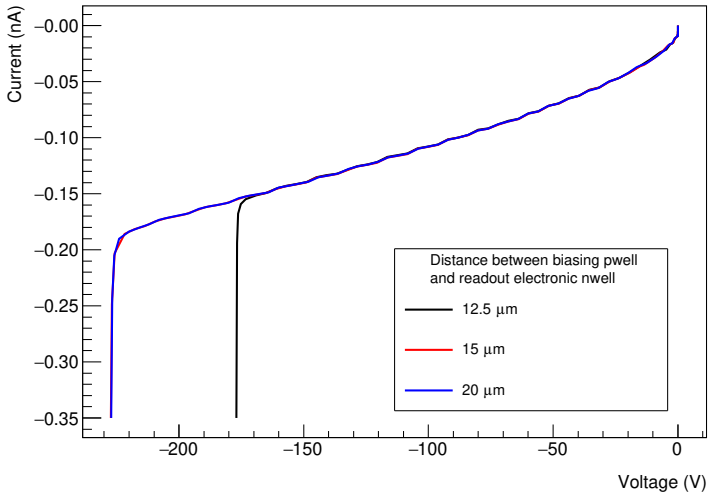


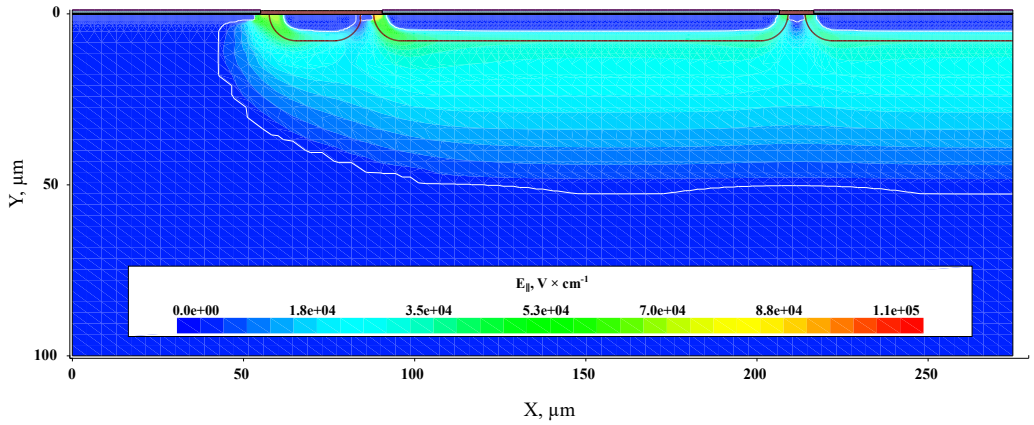
Figure 6.21: Simulated IV characteristics for different distances between the bias pwell and readout nwell. The results indicate that a spacing of 15 µm or greater stabilises breakdown voltage without significant further improvement. Note that the red curve is completely overlapped by the blue curve.

6.7.2 Electric field simulation

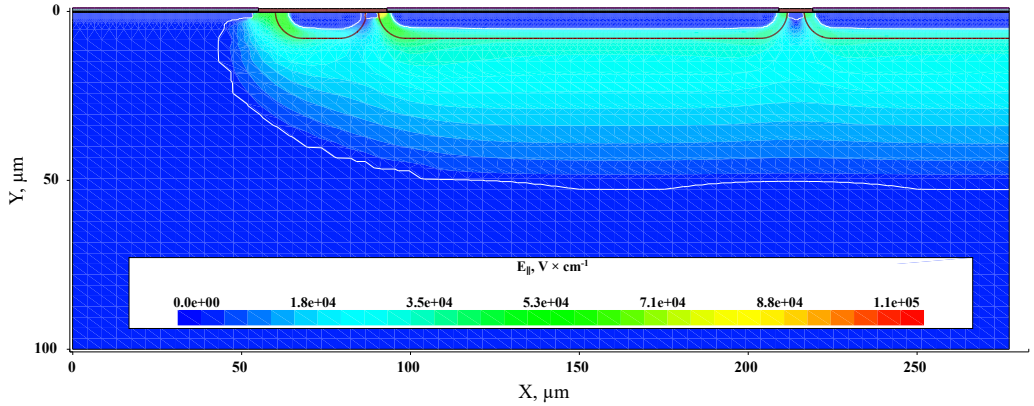
Electric field simulations were performed to visualise field distribution and identify regions prone to high electric field intensities, which increase the likelihood of breakdown. As shown in Figure 6.22, the electric field intensity is significantly higher for a spacing of 12.5 µm compared to 15 µm and 20 µm. The maximum electric field values observed in these regions are:

- 12.5 µm spacing: 7.5×10^{-4} V · cm^{-1}
- 15 µm spacing: 4.9×10^{-4} V · cm^{-1}
- 20 µm spacing: 3.9×10^{-4} V · cm^{-1}

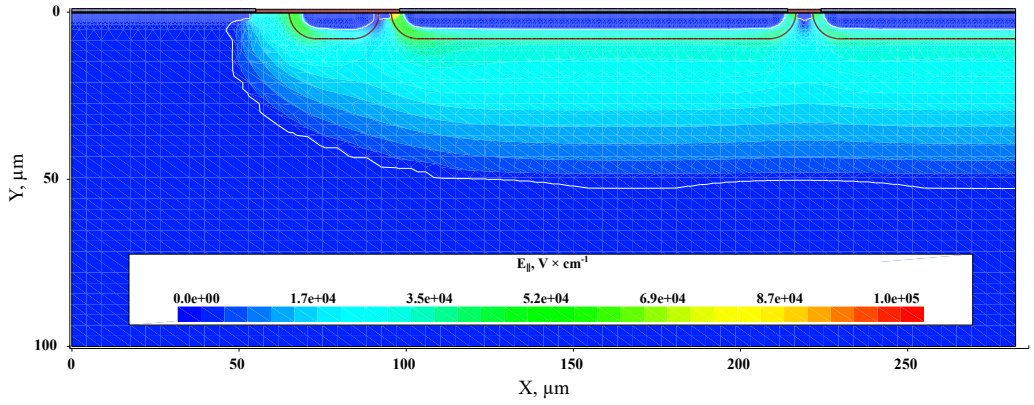
At larger spacings, the electric field becomes more uniformly distributed and lower in intensity, significantly reducing the risk of breakdown. This trend suggests that increasing the spacing to at least 15 µm is crucial for maintaining field stability and preventing premature breakdown.



(a) Electric field distribution for a spacing of 12.5 µm. The field intensity is high, increasing the risk of breakdown.



(b) Electric field distribution for a spacing of 15 µm. The field intensity is lower and more uniformly distributed.



(c) Electric field distribution for a spacing of 20 µm. The field intensity remains stable, with no significant difference from 15 µm.

Figure 6.22: TCAD simulations of electric field distributions for different periphery spacing distances. A reduction in peak electric field intensity is observed for increased spacing, minimising breakdown risk.

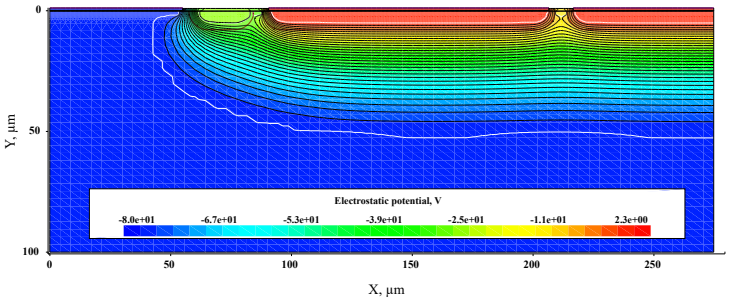
6.7.3 Electrostatic potential simulations

Another crucial aspect of sensor optimisation is electrostatic potential distribution, which directly influences charge transport and collection efficiency. A well-distributed potential ensures that charge carriers drift smoothly towards the collection electrode, minimising losses and improving signal fidelity [101]. In an optimised design, the electrostatic potential should be as uniform as possible, preventing large voltage drops, particularly under high-voltage (HV) operation. A more even potential distribution helps to stabilise the sensor response and reduce unwanted electric field variations.

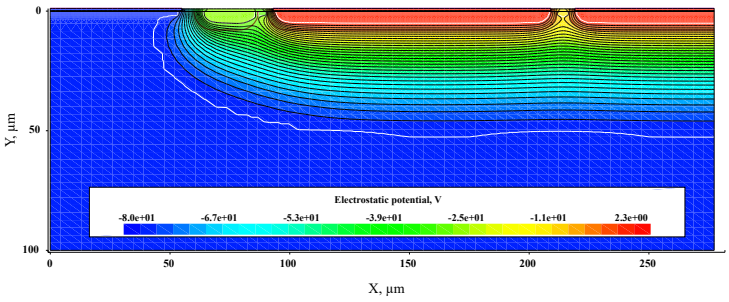
To evaluate the impact of different periphery spacing distances, TCAD simulations were performed to assess the potential distribution. Figure 6.23 presents the electrostatic potential maps for three different spacing values: 12.5 µm, 15 µm, and 20 µm. In each case, a zoomed-in version of the potential distribution is also provided for better visualisation.

The electrostatic potential analysis demonstrates a clear improvement in voltage distribution as the spacing between the bias pwell and readout nwell increases. For the 12.5 µm spacing, noticeable voltage drops are present in the HV region, leading to a less uniform potential. This can introduce charge transport and collection instabilities, negatively impacting sensor performance. A significant improvement is observed when increasing the spacing to 15 µm. The potential becomes considerably more uniform, reducing fluctuations that could affect signal quality. Further increasing the spacing to 20 µm brings the potential distribution close to an ideal state, where voltage gradients are minimal and charge collection is highly efficient.

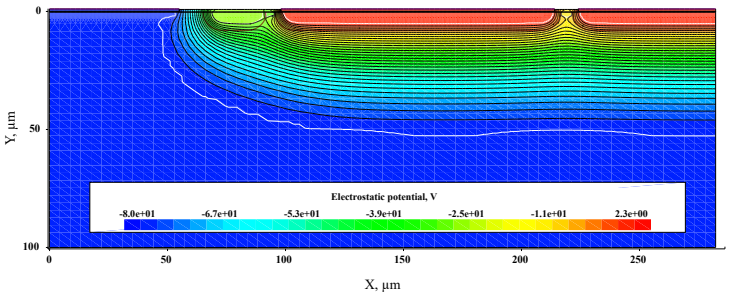
However, the most notable improvement occurs between 12.5 µm and 15 µm, indicating that 15 µm is the *minimum optimal spacing* for balancing breakdown voltage stability and potential uniformity. If the layout design allows, a spacing of 15 µm to 20 µm is recommended to ensure robust performance, particularly in irradiated environments where breakdown effects may be exacerbated.



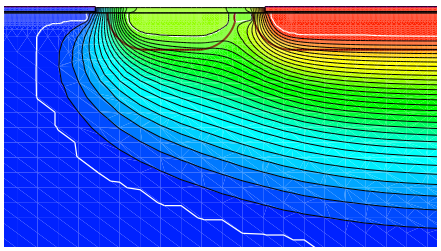
(a) 12.5 µm



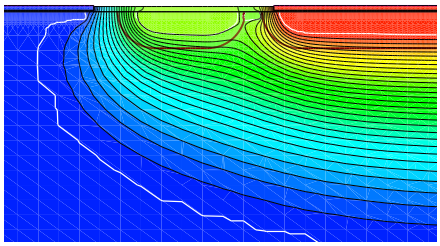
(b) 15 µm



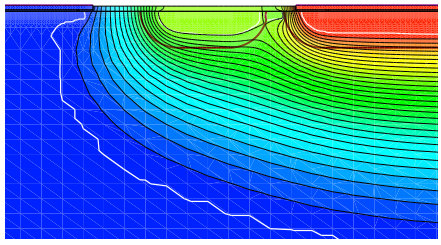
(c) 20 µm



(d) Zoomed 12.5 µm



(e) Zoomed 15 µm



(f) Zoomed 20 µm

Figure 6.23: TCAD simulations of electrostatic potential distributions for different periphery spacings: (a) – (c) show the full potential distribution for 12.5 µm, 15 µm, and 20 µm; (d) – (f) present zoomed-in views of critical areas for each spacing, highlighting improvements in potential uniformity with increasing distance.

6.7.4 Conclusion and recommendation

The simulations conducted in this study highlight the impact of periphery spacing on electric field distribution and electrostatic potential uniformity, both of which are critical for the stability and performance of the HitPix sensor. The results of these simulations were instrumental in refining the design of later HitPix versions, contributing to improved robustness against breakdown, enhanced charge collection efficiency, and operational stability, particularly in high-radiation environments.

The electric field analysis revealed that a spacing of 12.5 µm leads to a higher electric field near the high-voltage (HV) contact and readout nwell, increasing the likelihood of premature breakdown. Increasing the spacing to 15 µm significantly reduces the maximum value of the electric field and ensures a more stable voltage distribution. Beyond this point, further increases in spacing, such as 20 µm, offer diminishing improvements. The results indicate that 15 µm is the minimum optimal spacing for balancing breakdown stability and design constraints.

The electrostatic potential simulations demonstrate the potential distribution, the uniformity of which enhances charge transport and collection efficiency. At 12.5 µm, sharp voltage drops are observed, while increasing the spacing to 15 µm leads to a significantly more uniform potential. At 20 µm, the potential distribution reaches an almost ideal state, ensuring minimal voltage fluctuations under HV operation.

These findings confirm that optimising periphery spacing is essential for minimising breakdown risks and enhancing sensor reliability. The most significant improvement occurs when increasing the spacing from 12.5 µm to 15 µm, making this the *recommended minimum spacing*. If design constraints allow, extending the spacing to 20 µm provides additional stability without significant drawbacks.

6.8 Charge collection and timing using beam simulations

One of the significant advantages of TCAD simulations is their ability to model charge transport and collection under close to realistic beam conditions. When a high-energy particle traverses a silicon sensor, it generates electron-hole pairs, which then move under the influence of the internal electric field. Understanding how these charge carriers propagate and are collected at the electrodes is crucial for optimising sensor performance. TCAD simulations provide a detailed visualisation of charge distribution dynamics, allowing for an in-depth analysis of charge collection efficiency (CCE) and signal formation.

CCE is a key metric used to evaluate the performance of semiconductor sensors. It quantifies how effectively the charge carriers generated within the sensor are collected at the sensor electrodes. CCE is mathematically defined as:

$$\mathrm{CCE} = \frac{Q_{\mathrm{collected}}}{Q_{\mathrm{generated}}} \times 100\% \tag{6.15}$$

where $Q_{\text{generated}}$ is the total charge created due to radiation interaction, and $Q_{\text{collected}}$ represents the charge contributing to the signal at the electrodes. In most cases, this refers to the induced charge from carrier motion, as governed by Ramo's theorem, rather than only the charge that physically reaches the electrode.

The charge generation process in TCAD is modelled using various physical mechanisms, including radiation-induced charge generation, which accounts for particle interactions within the semiconductor material. For this analysis, a beam simulation model is used together with radiation damage models, as described in Section 4.3. The following calculations are performed on a sensor that has undergone radiation damage equivalent to years of operation within a high-energy beam environment.

To estimate the generated charge, the spatial and temporal distribution of charge carriers must be integrated:

$$Q_{\text{generated}} = \int G(x,t)\, dx\, dt \tag{6.16}$$

where $G(x,t)$ represents the charge generation rate as a function of position and time.

A key parameter used for beam simulation and charge generation is the Linear Energy Transfer (LET). The LET data was obtained from Brookhaven National Laboratory [102], as shown in Figure 6.24.

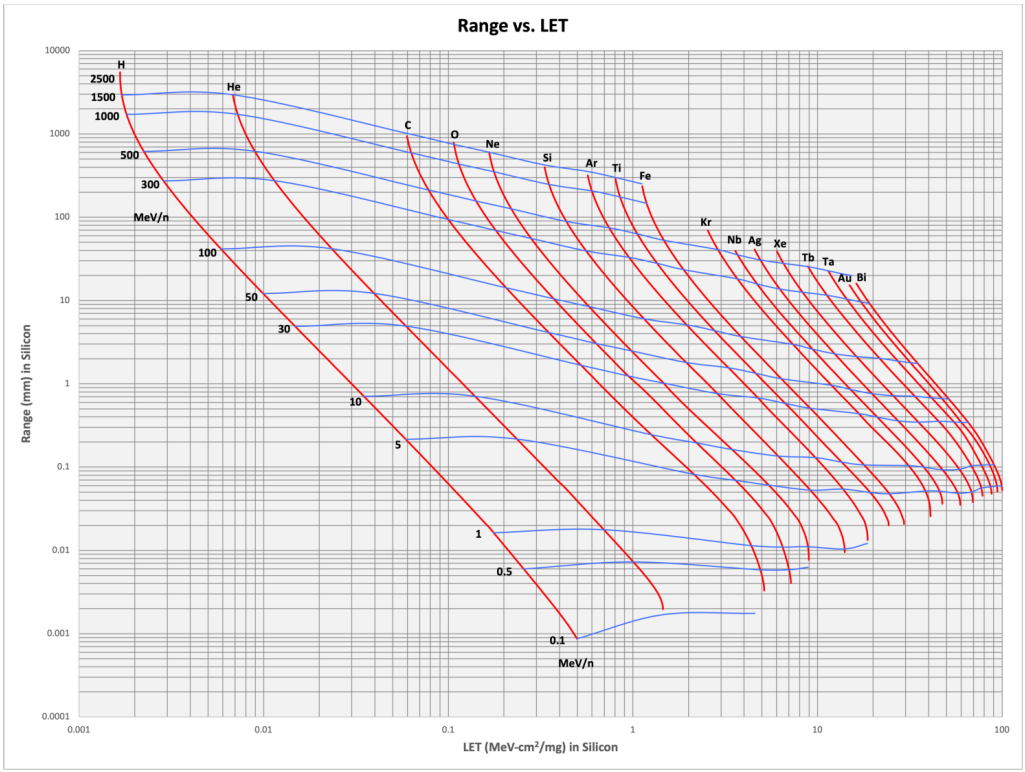


Figure 6.24: LET range plots for silicon [102]. The stopping power of different ions is represented as a function of energy, which determines the charge deposition in the material.

The electronic stopping power, also known as LET, is defined as the energy lost by the ion per unit path length due to interactions with atomic electrons. It is commonly expressed in units of MeV/(mg/cm^2) or MeV-cm^2/mg. This value determines the amount of charge deposited per centimetre, divided by the density of the material through electron-hole pair creation. A LET of 97 MeV-cm^2/mg corresponds to a charge deposition of approximately 1 pC/μm in silicon [103].

For this study, a proton beam of 200-300 MeV/n was selected due to the sensor's intended medical application. The beam duration was set to 1 ns. Based on available LET data, the stopping power for this proton beam in silicon is approximately:

$$\text{LET} = 0.0026 \text{ MeV-cm}^2/\text{mg} \tag{6.17}$$

This LET value can be converted into charge deposition per micrometre:

$$\text{LET in pC}/\mu\text{m} = \frac{\text{LET MeV-cm}^2/\text{mg}}{\text{Conversion Factor MeV-cm}^2/\text{mg per pC}/\mu\text{m}} \tag{6.18}$$

Substituting the values:

$$\text{LET} = \frac{0.0026 \text{ MeV-cm}^2/\text{mg}}{97 \text{ MeV-cm}^2/\text{mg per pC}/\mu\text{m}} \tag{6.19}$$

$$\text{LET} = 2.68 \times 10^{-5} \text{ pC}/\mu\text{m} \approx 2.67 \text{ pC}/\mu\text{m} \tag{6.20}$$

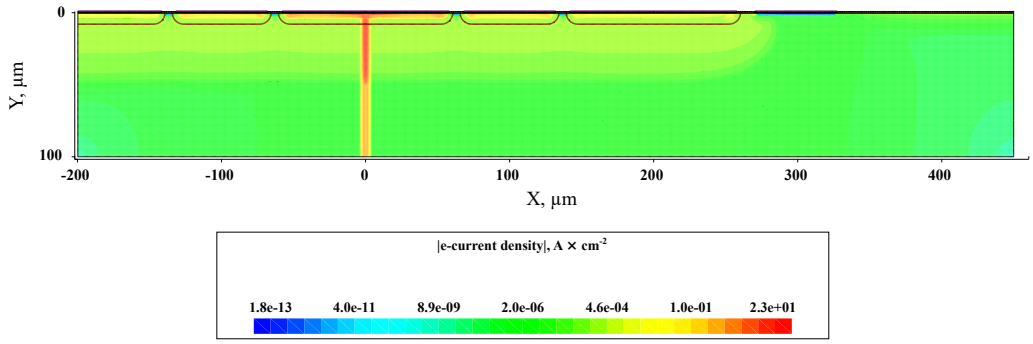
For a sensor with a thickness of 300 μm, the total charge generated is given by:

$$Q_{\text{generated}} = \text{LET} \times \text{Sensor Thickness} = 2.67 \text{ pC}/\mu\text{m} \times 300\mu\text{m} \approx 8 \text{ fC} \tag{6.21}$$

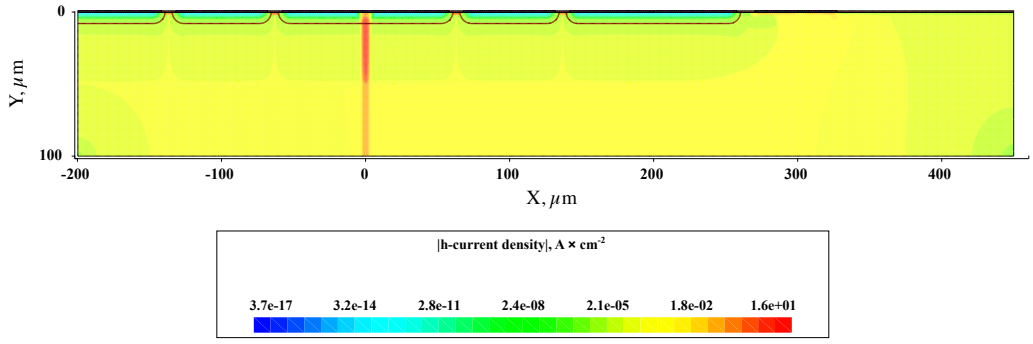
6.8.1 Charge carrier dynamics in TCAD simulations

Charge transport can be visualised in TCAD through different components of current density. The electron current density is the dominant contribution in reverse-biased $p-$type sensors, while the hole current density plays a minor role. The total current density is determined as the sum of both. Figure 6.25 illustrates these three current density components for the standard HitPix sensor when the particle passes through the collection node. Even though the primary focus is on electron current density, the total current density is used for CCE calculations. It is also important to note that current density values are expressed in A/cm², which requires appropriate scaling based on the contact pad size.

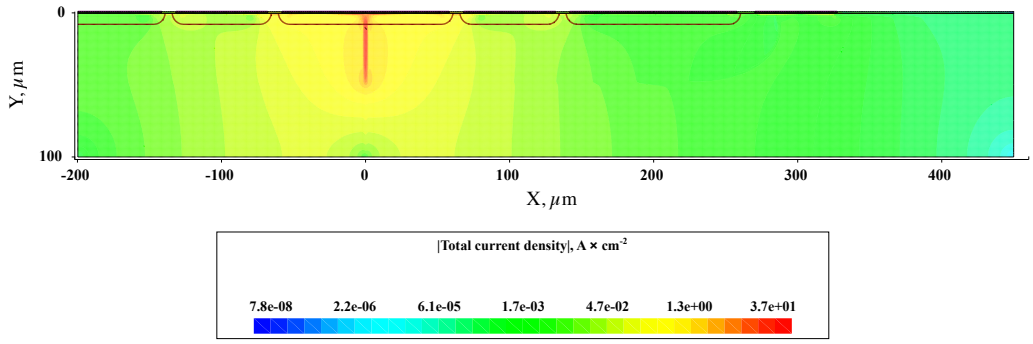
Charge collection efficiency is not always uniform across the sensor and also varies depending on the particle impact location. Simulations were performed for both the HitPix Standard and HitPix ISO designs, considering two different cases: when the beam strikes the collection node and when it impacts the electronics region.



(a) Electron current density at the collection node.



(b) Hole current density at the collection node.

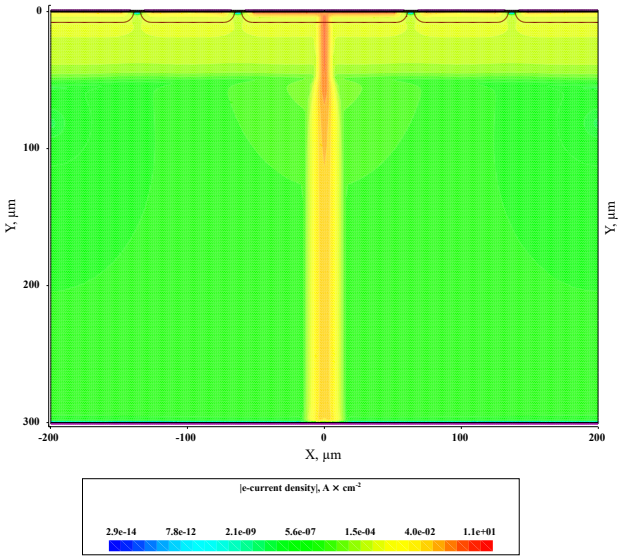


(c) Total current density at the collection node

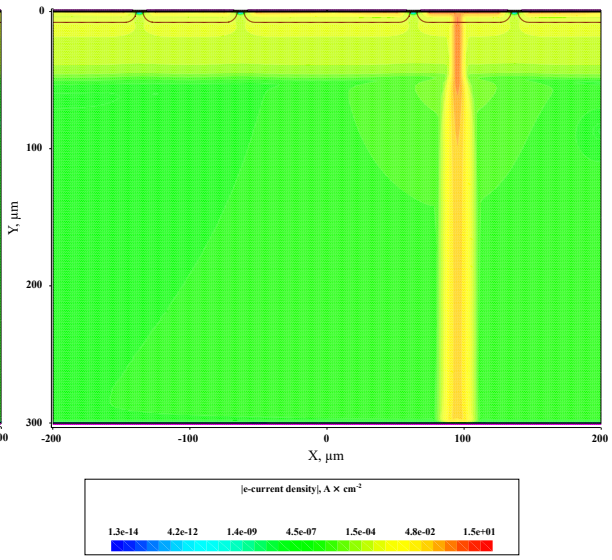
Figure 6.25: TCAD simulations of current density components for the HitPix Standard at the moment of the beam. The electron component dominates charge collection, while the total current density is used for efficiency calculations.

The charge collected directly from the collection node is the most efficient, as the generated carriers are quickly transported to the electrode. However, when the particle beam hits the electronics area, a significant fraction of the charge is lost due to diffusion effects. Figure 6.26

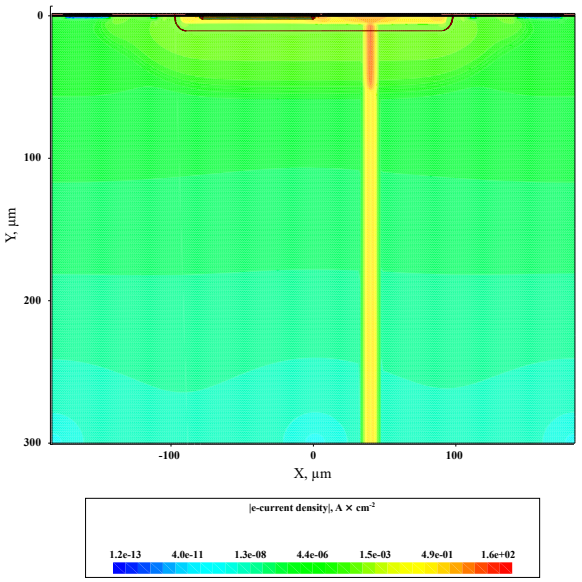
compares the electron current density for both sensor designs in these scenarios. When the beam impacts the collection node, both designs efficiently transport charge to the electrode. For HitPix ISO configuration, charge collection remains efficient even when the beam interacts with the electronics region, as the shared nwell helps to direct the charge towards the collection node. In contrast, in the standard configuration, the charge generated in the electronics area takes longer to reach the electrode, leading to higher signal losses.



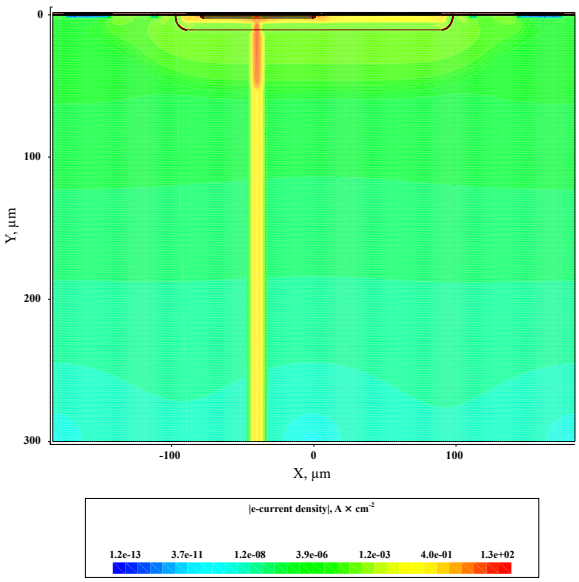
(a) Standard HitPix, sensor region.



(b) Standard HitPix, electronics region.



(c) ISO HitPix, sensor region.



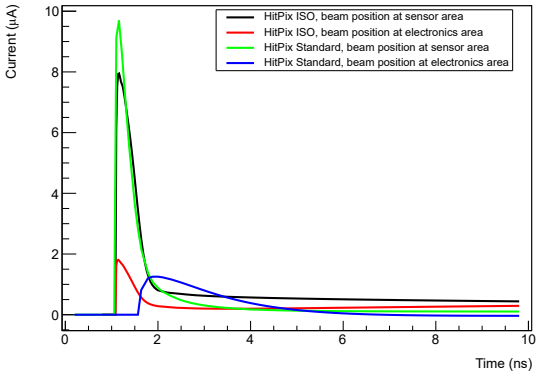
(d) ISO HitPix, electronics region.

Figure 6.26: Electron current density distribution in different sensor regions for the HitPix Standard and ISO designs at the moment of the beam. The ISO design maintains efficient charge collection even when the beam interacts with the electronics area, reducing signal loss.

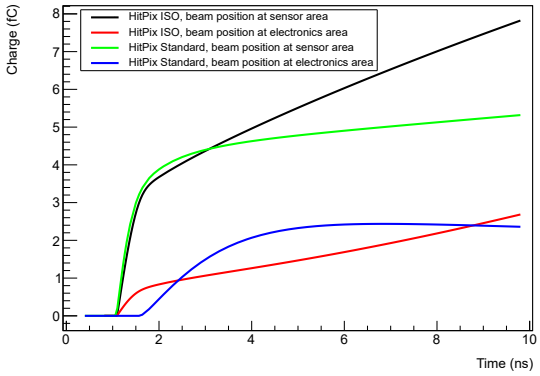
6.8.2 Transient simulation

TCAD simulations allow for transient analysis, which provides critical insights into the charge collection dynamics over time. These simulations not only reveal the signal strength but also provide an estimate of the charge collection speed. The transient response is particularly important in evaluating the sensor's performance under high-energy particle interactions, as it directly influences signal integrity and detection efficiency. More details about transient simulation can be found in Subsection 4.3.3.

Figure 6.27a presents the simulated current over time, $I(t)$, for both the HitPix Standard and ISO designs, with the beam injected at t = 1ns. Each design is tested in two cases: when the beam directly impacts the collection node and when it interacts with the electronics area. By integrating the current response over time, the total collected charge can be determined, as shown in Figure 6.27b.



(a) Current response over time, $I(t)$, for HitPix Standard and ISO designs under different impact locations.



(b) Integrated charge collection over time.

Figure 6.27: Transient charge collection analysis for HitPix Standard and ISO configurations. The current response determines the signal strength, while the integrated charge evaluates the total charge collection. In this simulation, the beam was injected at $t = 1$ ns.

Charge transport in silicon sensors is primarily governed by two mechanisms: drift and diffusion. In a reverse-biased $p-$type substrate, electrons move under the influence of the electric field, reaching the collection node within a few nanoseconds. This drift-dominated process ensures efficient charge transport from the depletion region. However, the collection of the charge carriers generated outside the depletion region relies on diffusion, which is a much slower process and results in charge losses, particularly in regions far from the collection node.

Charge collection behaviour differs significantly between the HitPix Standard and ISO designs. In the Standard configuration, when the beam strikes the collection node, a strong initial signal is observed. However, the total charge collected remains lower due to the limited diffusion

collection of the separated nwells. In contrast, when the beam impacts the electronics area, charge transport is significantly delayed, and only a small fraction reaches the electrode.

In the ISO design, the initial signal is slightly weaker when the beam hits the collection node. However, over time, more charge is collected due to the common nwell structure, which enhances charge transport efficiency. When the beam interacts with the electronics area, the ISO configuration enables immediate charge transport due to the shared nwell, preventing charge loss. From the charge-over-time plots, it is evident that during the first nanosecond, the signal slope for both Standard and ISO designs under direct collection is similar, suggesting that initial charge collection is dominated by direct drift. However, after 2 ns, the ISO design collects significantly more charge due to its extended collection efficiency.

6.8.3 Charge collection efficiency calculation

The total charge collected is determined by integrating the transient current at the collection electrode over time:

$$Q_{\text{collected}} = \int_0^t I_{\text{collected}}(t)\, dt \tag{6.22}$$

where $I_{\text{collected}}(t)$ is the transient current response.

Using TCAD simulation data and numerical integration, the collected charge in each configuration is determined as:

$$Q_{\text{collected, Standard, Sensor}} = 5.1 \times 10^{-15}\ \text{C} = 5.1\ \text{fC} \tag{6.23}$$
$$Q_{\text{collected, Standard, Electronics}} = 2.3 \times 10^{-15}\ \text{C} = 2.3\ \text{fC} \tag{6.24}$$
$$Q_{\text{collected, ISO, Sensor}} = 7.8 \times 10^{-15}\ \text{C} = 7.8\ \text{fC} \tag{6.25}$$
$$Q_{\text{collected, ISO, Electronics}} = 2.7 \times 10^{-15}\ \text{C} = 2.7\ \text{fC} \tag{6.26}$$

The charge collection efficiency (CCE) is then calculated as:

$$\text{CCE}_{\text{Standard, Sensor}} = \frac{5.1\ \text{fC}}{8\ \text{fC}} \times 100\% = 63.8\% \tag{6.27}$$
$$\text{CCE}_{\text{Standard, Electronics}} = \frac{2.3\ \text{fC}}{8\ \text{fC}} \times 100\% = 28.8\% \tag{6.28}$$
$$\text{CCE}_{\text{ISO, Sensor}} = \frac{7.8\ \text{fC}}{8\ \text{fC}} \times 100\% = 97.5\% \tag{6.29}$$
$$\text{CCE}_{\text{ISO, Electronics}} = \frac{2.7\ \text{fC}}{8\ \text{fC}} \times 100\% = 33.8\% \tag{6.30}$$

Based on the calculation of the collected charge, the charge collection efficiency was determined for both the Standard and ISO sensor configurations. The results show that the ISO sensor achieves significantly higher charge collection efficiency, with $CCE_{\text{ISO, Sensor}} = 97.5\%$ compared to $CCE_{\text{Standard, Sensor}} = 63.8\%$. Similarly, the charge collection efficiency in the electronics region is improved in the ISO configuration, reaching $CCE_{\text{ISO, Electronics}} = 33.8\%$ as opposed to $CCE_{\text{Standard, Electronics}} = 28.8\%$. These findings indicate that the ISO sensor architecture enhances charge transport and minimises losses, resulting in improved signal integrity and overall detector performance.

6.8.4 Final observations

The TCAD simulations conducted in this study provide crucial insights into the charge transport and collection mechanisms within the HitPix sensor under particle beam irradiation. These simulations allow for a detailed evaluation of charge carrier dynamics, revealing key performance differences between the Standard and ISO sensor architectures.

The charge collection efficiency (CCE) analysis demonstrates that the ISO design significantly outperforms the Standard configuration. When the particle beam impacts the sensor region, the ISO design achieves near-complete charge collection, whereas the Standard design experiences notable losses. This difference is primarily attributed to the common nwell in the ISO structure, which facilitates more efficient charge drift and collection. The transient current analysis further confirms that while both designs exhibit similar initial collection rates, the ISO sensor continues to collect charge over an extended period due to its enhanced depletion region and improved charge transport.

A more pronounced distinction arises when the beam interacts with the electronics region. In the Standard configuration, charge collection occurs predominantly through diffusion, leading to slower and less efficient transport to the electrode. In contrast, the ISO design again benefits from a unified nwell structure that ensures more direct charge transport, minimising losses and improving overall efficiency.

The findings from these simulations validate the advantages of the HitPix ISO architecture in optimising charge collection and signal integrity. Future work should focus on further refining the sensor design, including adjustments to the depletion region, doping profiles, and periphery spacing, to enhance performance in high-radiation environments. The continued use of TCAD simulations will remain essential in guiding these design optimisations, ensuring the development of more robust and efficient pixel sensors for demanding experimental conditions.

A final example of TCAD simulation results is presented in Figure 6.28, illustrating the electron current density in the HitPix Standard sensor at different timestamps. This visualisation provides essential insights into the temporal evolution of charge distribution and transport following beam interaction. While these simulations are computationally demanding, they offer valuable information on electric field behaviour, signal formation, and charge collection dynamics, all of which contribute to optimising sensor design.

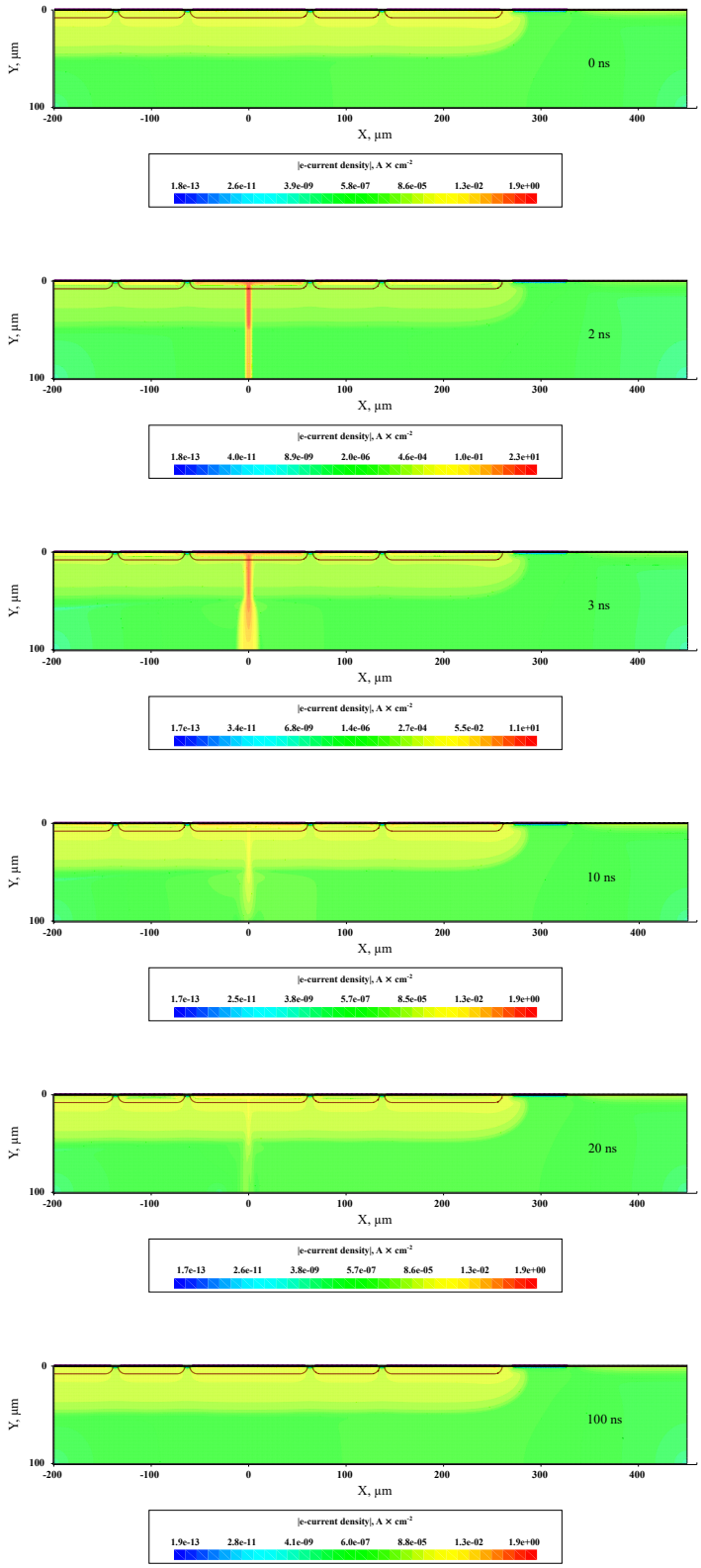


Figure 6.28: Electron current density in the HitPix Standard sensor during beam interaction at different timestamps. The simulation provides insights into charge transport behaviour and signal formation.

6.9 Summary for HitPix development

This chapter presents the development of HitPix, a monolithic active HV-CMOS hit-counting pixel sensor designed for beam monitoring in ion beam therapy. HitPix is proposed as a replacement for the current beam monitoring system at the Heidelberg Ion Beam Therapy Center (HIT), offering improved accuracy, real-time monitoring, and radiation hardness.

The development of HitPix followed an iterative approach, where each version introduced key modifications to enhance performance. HitPix1 focused on the electronics design, while later versions involved significant refinements in the sensor structure. In HitPix2, an NMOS amplifier was tested for improved speed, but PMOS was reinstated due to superior radiation tolerance. HitPix3 incorporated insights from TCAD simulations provided by this chapter, leading to substrate selection, pixel adjustments, and efficiency improvements. Finally, HitPix4 represents the most advanced version, addressing the challenges of its predecessors and the efficiency loss problem, which is analysed in Chapter 8.

TCAD simulations and sensor performance

TCAD simulations played a crucial role in optimising sensor efficiency, charge collection, and radiation tolerance. Key findings include:

- **Substrate resistivity effects:** Even though 5k Ω·cm substrate achieves full depletion at lower voltages with this specific HitPix architecture, it increases breakdown risks. A 300 Ω·cm substrate provides greater long-term stability, making it the preferred choice.
- **Power consumption considerations:** The final power estimate for HitPix3 is 44.76 mW, ensuring efficient operation within medical constraints while accounting for long-term radiation-induced ageing.
- **Electric field distribution:** Simulations showed that periphery spacing impacts breakdown stability. A spacing of 15 μm was identified as the optimal balance between breakdown prevention and efficient space usage.
- **Charge Collection Efficiency (CCE):** The ISO sensor architecture outperforms the standard configuration, achieving near-complete charge collection of 97.5% due to an improved depletion region and a more efficient charge drift mechanism.

Future outlook

The results of this study confirm that HitPix is a viable candidate for high-precision beam monitoring in radiation therapy. Future work will focus on prototype validation, integration into clinical environments, and further design optimisations to enhance its performance in high-radiation conditions. The continued use of TCAD simulations will play a crucial role in refining the sensor, ensuring its long-term reliability and effectiveness in medical applications.

7 HV-CMOS sensors for future gamma-ray telescopes — AstroPix

This chapter focuses on the development of monolithic CMOS active pixel silicon sensors, AstroPix, as a novel technology for future gamma-ray telescopes. The primary design was led by Nicolas Striebig, Ivan Perić and Richard Leys, as detailed in [104]. Currently, AstroPix has undergone four design iterations, from AstroPix_v1 to AstroPix_v4. This work discusses the transition between versions AstroPix_v2 and AstroPix_v3. At this stage of development, as a part of this thesis, the sensing section of the chip was simulated and refined. The results of these design modifications were presented as part of the AstroPix development at the TREDI 2023 [105] and TWEPP 2023 [106] conferences. The improved design has been fully integrated into AstroPix_v3.

7.1 Application description

Multimessenger astrophysics is a rapidly evolving field of science that offers unparalleled access to extreme processes shaping the universe. Gamma-ray observations have been central to the multimessenger initiative and will continue to play a crucial role as the field advances. NASA's Fermi mission has demonstrated that many of the extreme processes responsible for generating gravitational waves, neutrinos, and cosmic rays also emit gamma rays. These multimessenger sources are particularly bright in the under-explored MeV band. To address this, a mission proposal was submitted to the 2021 NASA Astrophysics Medium Explorer (MIDEX), aiming to explore these critical energies and fully exploit the potential of multimessenger astrophysics [107].

The All-Sky Medium Energy Gamma-ray Observatory eXplorer (AMEGO-X), as part of this project, is designed to detect and characterise gamma rays originating from extreme astrophysical accelerators and explosions. The primary scientific objectives include [107]:

- investigating supermassive black holes and their connections to neutrinos and cosmic rays;
- studying binary neutron star mergers and the relativistic jets they produce;
- exploring cosmic ray acceleration sources, including galactic supernova remnants;
- continuously monitoring other astrophysical events and sources across the entire sky in this crucial energy range.

AMEGO-X will operate within the 100 keV to 1 GeV energy range, offering an increase in sensitivity compared to previous gamma-ray observatories. During its three-year baseline mission, AMEGO-X will survey nearly the entire sky every two orbits, progressively constructing a highly sensitive all-sky map of gamma-ray sources and emissions [107].

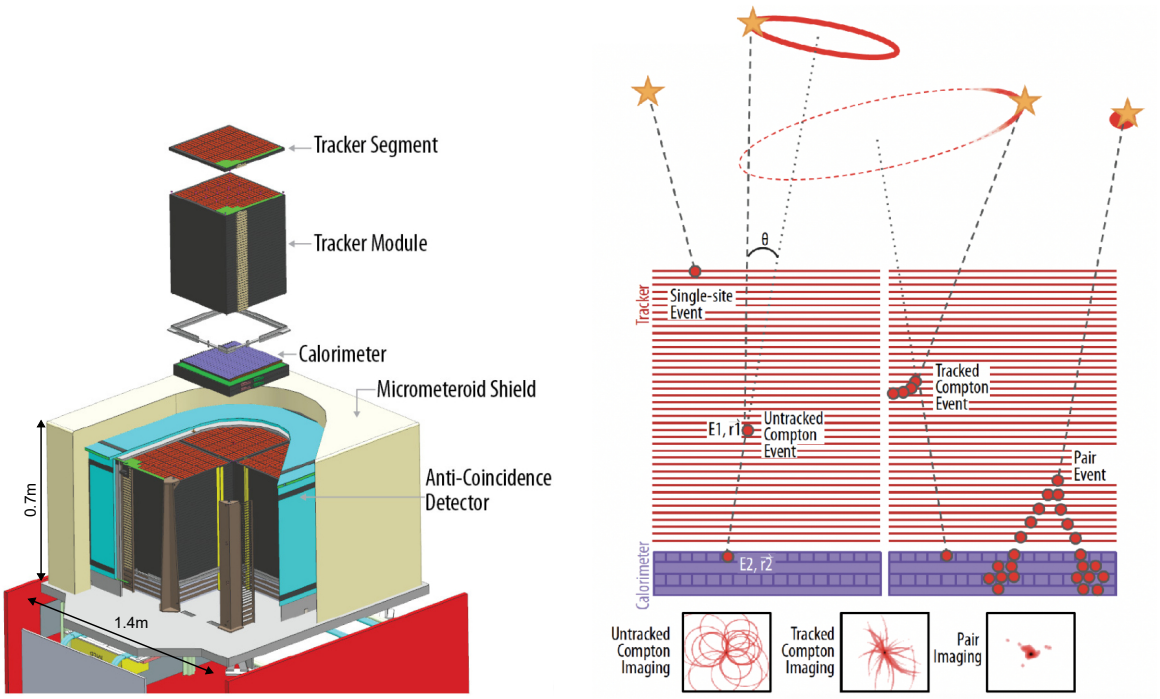
The AMEGO-X gamma-ray telescope (GRT) is its primary scientific instrument, designed to enable imaging, spectroscopy, and polarimetry of gamma-ray sources. The GRT consists of two primary subsystems: the gamma-ray detector (GRD) and the anti-coincidence detector (ACD). Both components are protected by a micrometeoroid shield (MMS) [108], as shown in Figure 7.1a. The GRD is composed of a tracker built with 40 layers of silicon Complementary Metal-Oxide-Semiconductor (CMOS) Monolithic Active Pixel Sensors (MAPS) and a calorimeter incorporating four layers of caesium iodide (CsI) scintillator bars. Together, these elements facilitate the precise characterisation of gamma rays over the 100 keV to 1 GeV energy range [107].

AMEGO-X employs three primary gamma-ray interaction mechanisms, as depicted in Figure 7.1b:

- **Compton scattering** (100 keV – 10 MeV) – this occurs when an incident photon interacts with an orbital electron, treating it as a free particle due to the electron's low binding energy relative to the photon's energy [68]. The measured position and energy of a Compton scattering event, combined with subsequent absorption, constrain the gamma-ray's initial trajectory to a circle in the sky [107]. These are classified as:
  - **untracked Compton events** – only the scattered photon is detected, limiting localisation to a circular uncertainty region;
  - **tracked Compton events** – if the recoil electron's trajectory is also measured, additional kinematic constraints refine the photon's path to an arc, significantly improving background rejection and angular resolution;

  Compton scattering is also inherently polarisation-sensitive; a linearly polarised source produces a sinusoidal scattering angle distribution in the instrument.
- **Pair production** (>10 MeV) – at higher energies, gamma rays convert into electron-positron pairs. Their ionisation tracks are recorded in the tracker, while the total energy is determined by electromagnetic showers in the calorimeter.
- **Photoelectric absorption** (<100 keV) – in this process, photons are absorbed in a single detector pixel. While these events lack imaging capability, their combined signals help in locating transient sources. AMEGO-X enhances low-energy gamma-ray burst (GRB) detection by employing rapid (<100 s) readouts, capturing emissions down to 25 keV.

The following section 7.2 will provide a detailed description of AstroPix, which serves as a key component of the AMEGO-X gamma-ray telescope (GRT). It covers all existing versions of the chip and their development.



(a) Exploded view of the AMEGO-X gamma-ray telescope, showing the tracker, built from AstroPix chips, and calorimeter subsystems [109].

(b) Detection principles of AMEGO-X: Compton scattering, pair production, and photoelectric absorption are utilised to characterise gamma-ray interactions [109].

Figure 7.1: The AMEGO-X gamma-ray telescope and its detection techniques.

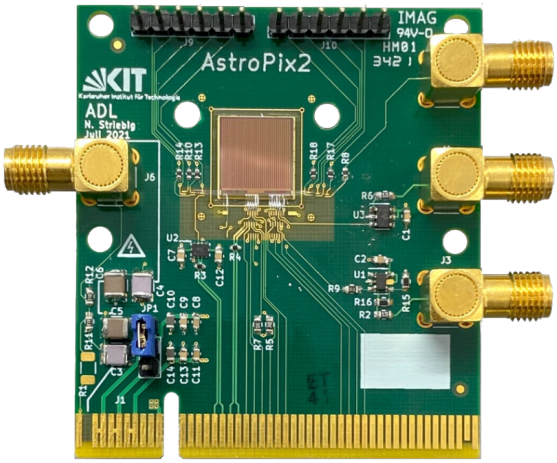
7.2 Chip description

Space-based gamma-ray telescopes, such as the Fermi Large Area Telescope, have traditionally relied on single-sided silicon strip detectors to track secondary charged particles produced by primary gamma rays with high resolution. However, for keV-MeV telescopes, precise event reconstruction requires two-dimensional position information within a single detector, especially in the Compton regime. AstroPix, a monolithic CMOS active pixel silicon sensor, was developed as an innovative approach for future gamma-ray telescopes [104]. This technology belongs to the HV-CMOS chip family.

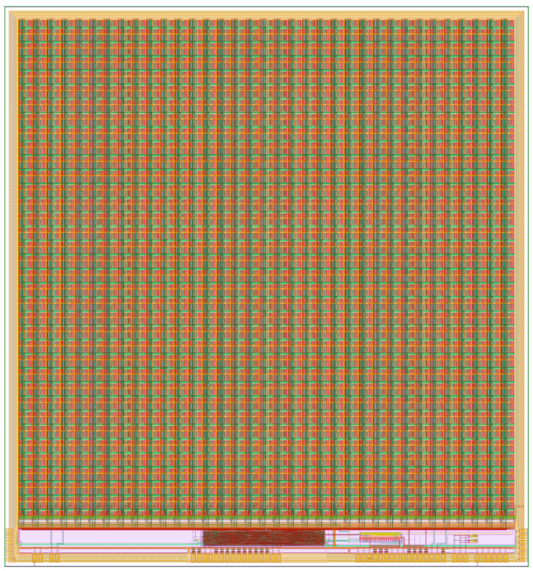
Each AstroPix sensor features individual pixels that can range from 10 μm^2 to 1 mm^2, offering fine control over position resolution while balancing power consumption and data rate. The integration of in-pixel CMOS circuitry eliminates the need for an external readout ASIC, as charge collection and digitisation occur directly on-chip. The CMOS architecture incorporates charge collection, signal amplification, and a comparator for signal processing. A deep nwell structure isolates the pixel electronics from the bulk substrate, and a high-voltage bias is applied at each pixel, ensuring rapid and efficient charge collection. AstroPix is classified as a monolithic active pixel sensor, where each pixel is individually depleted using high voltages. The comparator output for each pixel is transmitted to the periphery, where digital logic pro-

cesses the data. This design avoids the need for bump bonding each pixel, minimising passive material in the active detector region. Unlike traditional silicon strip detectors, AstroPix's architecture reduces capacitance and noise, enhancing signal fidelity [104].

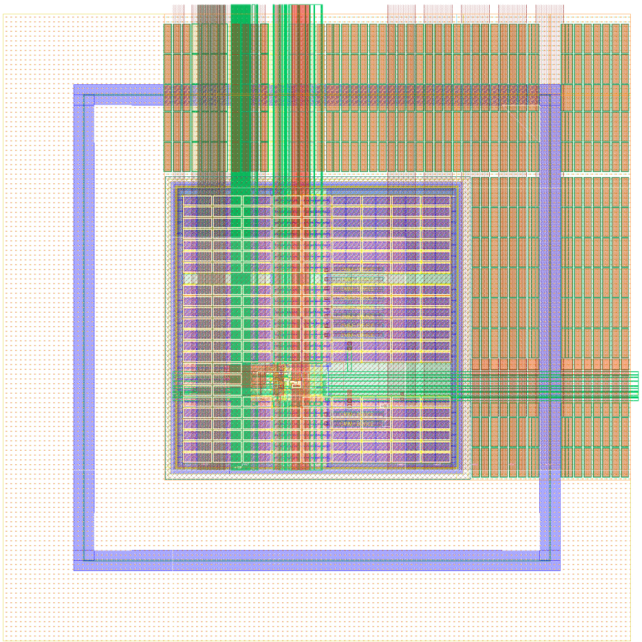
AstroPix's data collection strategy focuses on optimising power efficiency and data bandwidth. The comparator outputs from each row and column are OR'ed and sent to the periphery, making the output signal similar to the output of a silicon strip-based detector. This enables efficient event reconstruction using row/column pairing. The current implementation employs SPI readout via an FPGA-based DAQ system, which is discussed in the work of collaboration [104].



(a) AstroPix_v2 mounted on a carrier board for testing [105].



(b) AstroPix_v2 full chip layout [105].



(c) Cadence layout of a single pixel of AstroPix_v2, including metallisation layers.

Figure 7.2: Overview of AstroPix_v2 chip design and testing setup.

7.3 Evolution of chip designs

The development of the AstroPix sensor follows a structured, iterative process aimed at optimising its performance for space-based applications. Derived from ATLASPix [100], initially designed for the ATLAS detector at the Large Hadron Collider, AstroPix was adapted for gamma-ray detection with a focus on improving its performance, lowering the noise, and optimising power consumption. AstroPix has undergone multiple design iterations, each improving upon the previous version in terms of charge collection, power efficiency, and readout architecture.

AstroPix_v1, submitted in April 2021 and received in December 2021, was the first adaptation of ATLASPix for gamma-ray applications. It featured an 18×18 pixel matrix with a pitch of 175×175 µm^2, integrated onto a 5×5 mm^2 chip. The choice of at least 500 µm-thick wafer extended the dynamic range without sacrificing spatial resolution. AstroPix aims to achieve an energy resolution with a FWHM (Full Width at Half Maximum) of 5 keV in the 25–122 keV range and 5.6 keV at 622 keV, while maintaining a pixel dynamic range of 25–700 keV [110]. Initial testing of AstroPix_v1 focused on analogue data, as the digital readout was unavailable due to a design flaw. While digital bitstream data was considered, untuned pixel responses led to large variations (up to $\sigma/\mu = 60\%$) in the Gaussian-fit distribution, making analogue measurements the primary basis for characterisation. A critical issue identified during testing was comparator oscillations caused by parasitic capacitance due to missing shielding, leading to unwanted feedback and instability. These oscillations were confirmed through simulations and were addressed in subsequent iterations [104].

AstroPix_v2, introduced several improvements to enhance charge collection and enable higher depletion voltages. The pixel size was increased to 250×250 µm^2, reducing the crosstalk while improving charge collection efficiency. The implementation of guard rings allowed for depletion voltages up to 190 V, increasing the depleted sensor depth. Analogue power consumption was optimised, reducing it from 14.7 mW/cm^2 to 3.4 mW/cm^2 by eliminating unused digital-to-analogue converters. This version also introduced digital readout functionality, although initial tests still focused primarily on analogue signal processing. Additionally, global timestamping was enabled, allowing digital signals to be held in a buffer during readout to improve synchronisation with subsequent triggers. Further studies were conducted on leakage current control and depletion characteristics to refine the chip's operational parameters. The study of the depletion depths, impact of metalisation design and guard ring configurations between the pixels is further detailed in Sections 7.5 and 7.6 as part of the TCAD simulation study.

AstroPix_v3, submitted in July 2022 and received in February 2023, brought significant design enhancements, particularly in depletion depth and substrate resistivity. The sensor size was increased to 2×2 cm^2, featuring a 35×35 pixel matrix with a pixel pitch of 500×500 µm^2, while the collection node is 300×300 µm^2. A key advancement was the introduction of a high-resistivity (10k Ω·cm) substrate, improving depletion depth and charge collection uniformity. Additionally, the guard ring design was refined to remove conductive pwells, which had previously contributed to charge leakage into neighbouring pixels. Tests with different substrate resistivities (20 Ω·cm, 370 Ω·cm, and 18k Ω·cm) revealed that lower-resistivity sensors

exhibited early breakdown at 240 V. In contrast, higher-resistivity versions remained stable beyond 400 V. Simulations of charge collection efficiency and depletion depth across different biasing conditions were conducted, validating the choice of materials for AstroPix_v3. These depletion studies for HV-CMOS sensors are provided as part of HitPix sensor development in Subsection 6.6.1. The power consumption of AstroPix_v3, measured at 3.06 mW/cm^2 for the digital part and 1.06 mW/cm^2 for the analogue part, exceeded the total available budget of 1.5 mW/cm^2 required for the satellite tracker. This posed a challenge, as the final system would contain 61,000 chips, requiring further power-saving modifications to meet mission constraints [106].

AstroPix_v4, submitted in May 2023, focused on power optimisation, event timing precision and readout architecture. Although a detailed discussion of this version is beyond the scope of this work, key improvements include optimisations in timestamp generation and data acquisition. The removal of the 200 MHz clock and an LVDS receiver significantly reduced digital power consumption to below 0.25 mW/cm^2. Additionally, the implementation of per-pixel Flash-Time-to-Digital Converters (TDCs), controlled by a global delay-locked loop (DLL), enabled a time-of-arrival (ToA) and time-over-threshold (ToT) resolution of 3.125 ns. Simulations indicate that eliminating high-speed clocks and receivers reduces digital power consumption for ToT and ToA measurements to 700 µW, including the contributions of the CMOS phase-locked loop (PLL) and global delay-locked loop (DLL), with a safety margin of factor 2. Furthermore, the updated readout scheme enhances rate capability and resolves identification issues observed in AstroPix_v3. The increased ToA resolution of 3.125 ns is particularly beneficial for the integration of AstroPix into the Electron-Ion Collider (EIC) barrel calorimeter ePIC, where precise timing and high-rate performance are critical. More details about these improvements were presented at the Topical Workshop on Electronics for Particle Physics in 2023 [106].

The progressive enhancements from AstroPix_v1 to AstroPix_v4 demonstrate a well-defined development path, addressing key design challenges while preserving the advantages of monolithic CMOS active pixel sensors. Each version systematically refined aspects such as charge collection efficiency, digital power management, and depletion characteristics, ensuring the sensor meets the high-performance requirements of future gamma-ray space missions. The next sections, from 7.4 to 7.6, present a characterisation of the sensor's performance and explore geometric adjustments aimed at improving breakdown voltage and sensing capabilities of the chip through TCAD simulations.

7.4 Sensor design and TCAD modelling

A central objective of this research was to develop a sensor capable of operating under extreme high-voltage conditions while ensuring full depletion of the sensitive volume. In addition, the impact of surface metallisation on breakdown behaviour and the effectiveness of various guard ring configurations were studied. These investigations were essential steps in the development from AstroPix_v2 to AstroPix_v3, and the insights gained were implemented in the latter version. Simulations for AstroPix focused on two key aspects: optimising surface structures,

particularly metal layers and p-stop design, and achieving complete depletion of the sensor. The pixel layout and the geometrical arrangement of key features were examined and refined using TCAD tools. A schematic cross-section of the sensor is shown in Figure 7.3, which highlights the critical geometrical regions subject to this optimisation process.

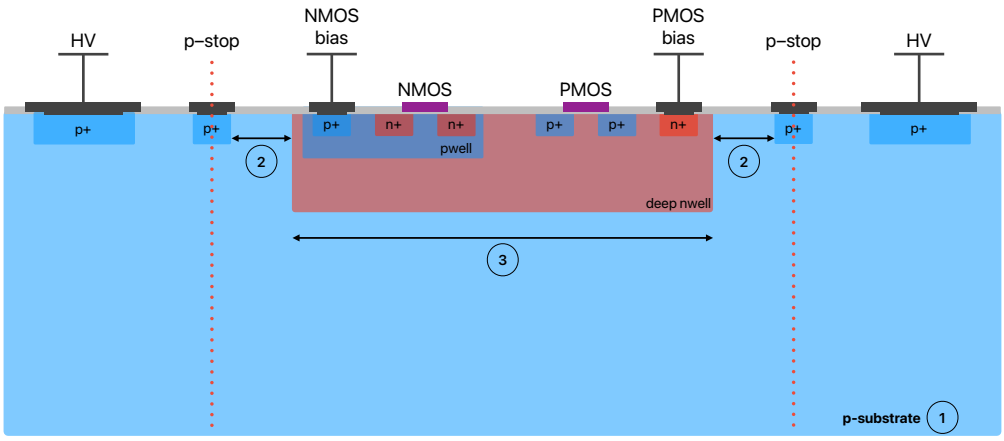


Figure 7.3: Schematic cross-section of the AstroPix sensor. The diagram illustrates the primary components of the pixel structure, including the high-resistivity substrate, p-stop implants, deep nwell for the electronics node, and metallisation layers. The numbers (1–3) correspond to the main geometrical parameters studied in TCAD: (1) substrate type and resistivity, (2) spacing between the electronics node and the p-stop, and (3) the size of the deep nwell. These parameters strongly influence breakdown voltage and the extent of sensor depletion. The full pixel pitch is defined from the centre of one p-stop to the centre of the adjacent p-stop.

The substrate selection followed a similar strategy as that for the HitPix chip, which can be found in Section 6.6, where several resistivity values were evaluated to determine suitability for high-voltage operation. For AstroPix, a broader range of substrate resistivities was explored: 20 Ω·cm, 370 Ω·cm, 10k Ω·cm, 15k Ω·cm, and 18k Ω·cm. These variants were simulated to understand their effect on the depletion region and breakdown behaviour.

The layout of the pixel cell was another critical area of study. Various configurations of the distance between the deep nwell (the electronics node) and the surrounding p-stop were simulated. Three primary spacings were investigated: 25 µm, 55 µm, 75 µm and 100 µm. This distance also determines the full pixel pitch. Additionally, two sizes of the deep nwell were tested: 300 µm and 440 µm. These parameters play a significant role in defining the electric field distribution and potential charge-sharing effects between adjacent pixels.

Simulations were conducted at bias voltages of -250 V and -500 V. While -250 V represents the nominal operational voltage, the ability to withstand up to -500 V is highly beneficial for ensuring stable and efficient performance as a full-depleted sensor in the long term.

The TCAD simulations provided the following key outputs:

- Electric field distribution, which is essential for assessing breakdown risks and optimising implant profiles.

- Extent of the depletion region, which is directly tied to the chosen substrate resistivity and applied bias.
- Electron current density profiles, which are valuable for understanding charge transport dynamics within the pixel volume.

These detailed simulation results were instrumental in finalising the design of AstroPix_v3, ensuring improved robustness, depletion depth, and high-voltage tolerance.

7.5 Choice of substrate and depletion region analysis

One of the primary objectives in the design of the AstroPix sensor is to achieve full depletion of the sensitive volume, preferably at an operating voltage of -250 V. At the same time, the sensor must be capable of withstanding bias voltages up to -500 V, as required by the intended experimental environment. The simulations explored a variety of free parameters, such as pixel pitch, the size of the electronics node (deep nwell), and substrate doping. To isolate the influence of the substrate, this analysis focuses on substrate resistivity, keeping the geometry, specifically the pixel pitch and nwell size, fixed. For all results shown here, the deep nwell size was set to 440 µm.

Figure 7.4 illustrates the simulated depletion region for each of the five substrate types under a -250 V bias. As expected, increasing the resistivity of the substrate leads to a deeper depletion region.

The quantitative outcomes of the simulations are summarised in Table 7.1. Substrates with resistivity values of 5k Ω·cm and above result in depletion depths of over 200 µm, which is sufficient for full-depleted operation.

Table 7.1: Summary of simulated dependences for substrate resistivity, doping concentration, and depletion thickness at -250 V.

Substrate resistivity	Doping concentration	Depletion thickness
20 Ω·cm	6.73×10^{14} cm^{-3}	19.7 µm
370 Ω·cm	4.41×10^{13} cm^{-3}	90.7 µm
5k Ω·cm	6.65×10^{12} cm^{-3}	220 µm
10k Ω·cm	1.46×10^{12} cm^{-3}	425 µm
18k Ω·cm	8.84×10^{11} cm^{-3}	520 µm

Although depletion depth is a crucial parameter, the ability to sustain high voltages is equally important. To explore this, further simulations were performed at -500 V for the three most promising substrates: 5k Ω·cm, 10k Ω·cm, and 18k Ω·cm. Figure 7.5 shows the depletion regions and electric field distributions at this higher voltage.

These results show that the 5k Ω·cm substrate remains partially depleted, while the 18k Ω·cm substrate, although achieving full depletion, exhibits significant asymmetry and irregularity in the electric field. This may lead to non-uniform charge collection. The 10k Ω·cm substrate offers a balanced compromise, achieving sufficient depletion while maintaining field uniformity. To quantify these fields, electric field profiles were extracted beneath the collection node for each case. These are plotted in Figure 7.6. The peak fields remain well below the avalanche breakdown threshold for silicon, which is approximately 3×10^5 V×cm^{-1}.

According to the van Overstraeten–de Man model [60], avalanche generation becomes relevant for electric fields between 1.75×10^5 V×cm^{-1} and 6×10^5 V×cm^{-1}. While these effects can be harnessed for gain, this requires intentional design, such as specialised doping or geometry.

The electric fields observed here are within safe limits. Among the tested configurations, the 10k Ω·cm substrate consistently provides optimal balance between depletion depth, electric field symmetry, and voltage tolerance, and is therefore selected for further design steps, including guard ring optimisation.

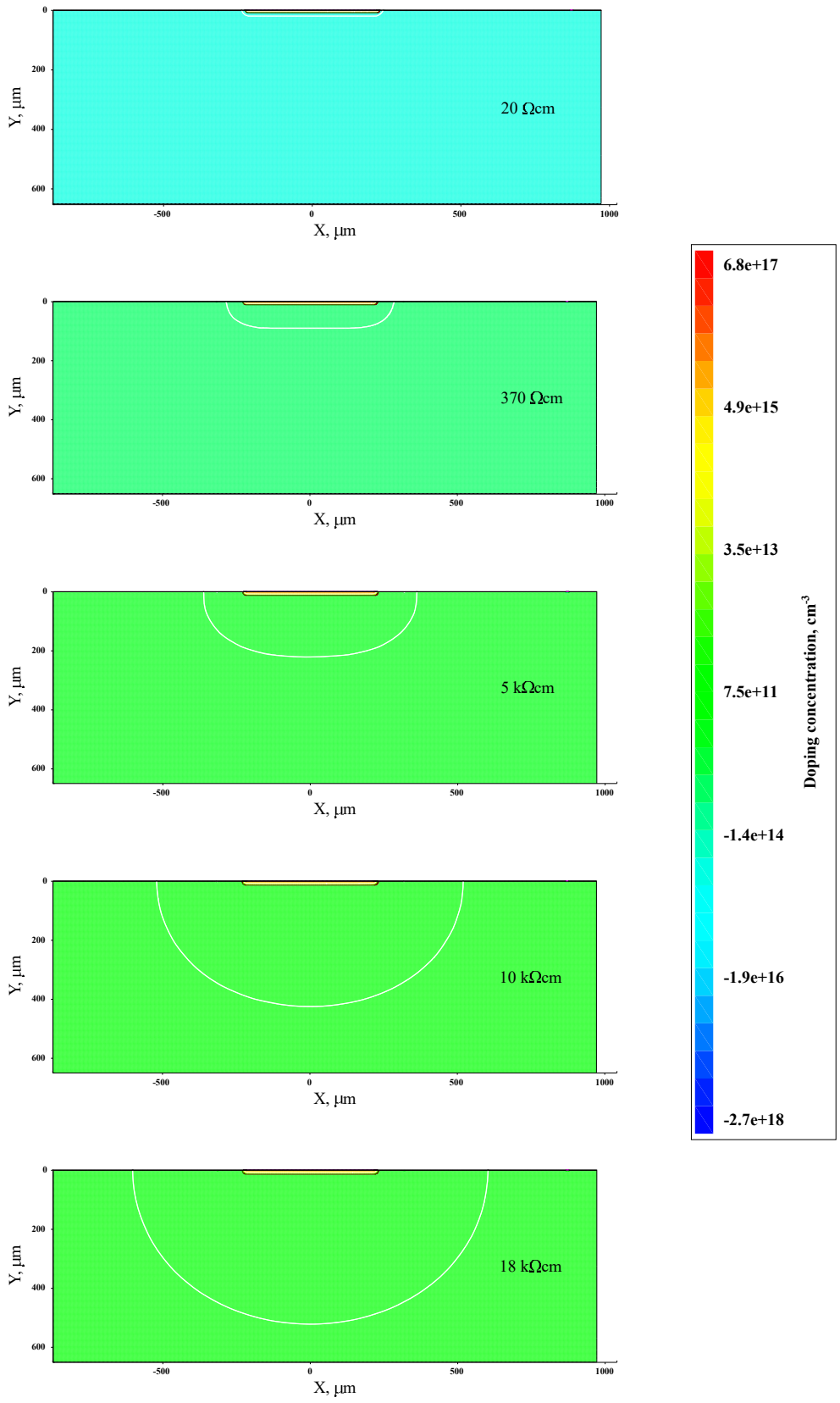
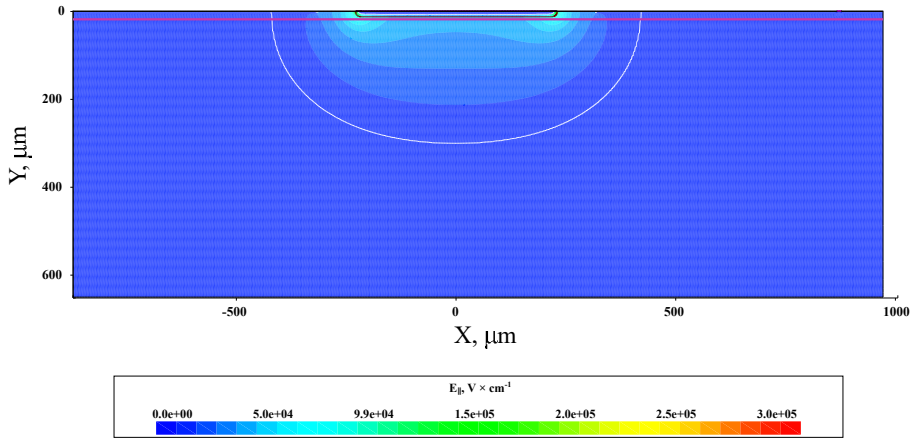
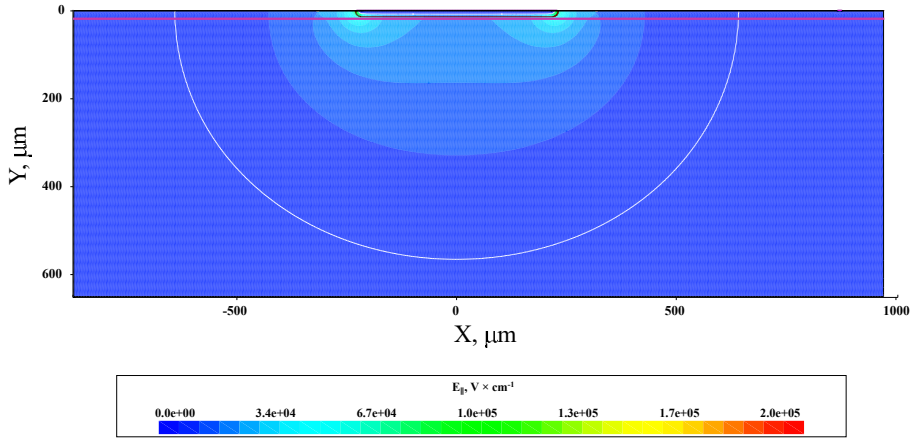


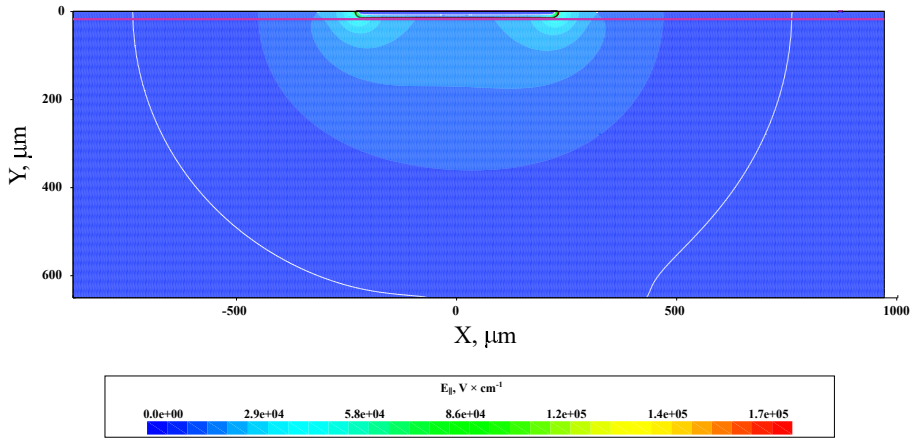
Figure 7.4: Simulated depletion regions at -250 V for substrates with different resistivities: 20 Ω·cm, 370 Ω·cm, 5k Ω·cm, 10k Ω·cm, and 18k Ω·cm. Higher resistivity results in a wider and deeper depleted volume. The pixel layout and nwell geometry are fixed in all cases.



(a) 5k Ω·cm substrate.



(b) 10k Ω·cm substrate.



(c) 18k Ω·cm substrate.

Figure 7.5: Simulated depletion regions and electric field distributions at -500 V for three high-resistivity substrates. The depth and symmetry of the depleted volume vary with resistivity, indicating differing charge collection behaviour.

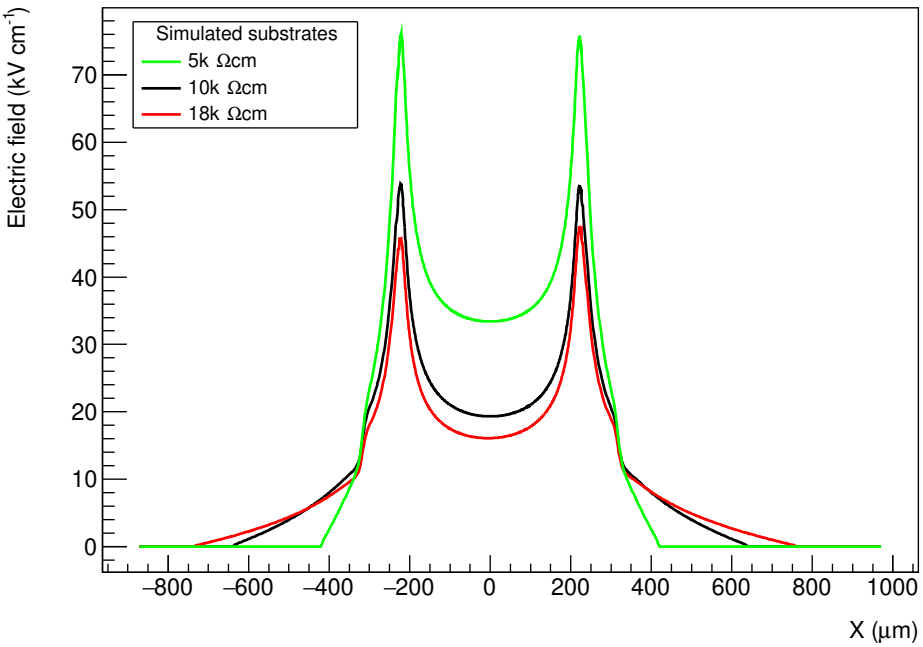


Figure 7.6: Electric field distribution beneath the nwell collection node at -500 V. The highest peak is seen in the 5k substrate case, with a value of 7.6×10^4 V×cm^{-1}. All designs remain below the avalanche threshold.

Comparison of deep nwell sizes

With the substrate choice established, the next design consideration concerns the pixel pitch. For that, two sizes of the collection node were explored based on experimental constraints: 300 µm and 440 µm. Table 7.2 summarises the key outcomes for the simulated depletion regions for these configurations.

Table 7.2: Depletion depth and length for 300 µm and 440 µm pixel sizes at two applied voltages.

Collection node size	300 µm	300 µm	440 µm	440 µm
Applied voltage	-250 V	-500 V	-250 V	-500 V
Depletion depth (y-dimension)	400 µm	522 µm	425 µm	565 µm
Depletion length (x-dimension)	440 µm	555 µm	520 µm	640 µm

The depletion depth in the y-dimension is similar for both designs and approaches 500 µm. However, in the x-dimension, the 440 µm pitch yields a depletion region larger than the pixel pitch itself. This enables charge sharing and potentially improves spatial resolution. Conversely, the 300 µm design is more suited for discrete, well-isolated detection. The electric field at the nwell boundaries is approximately 2.7×10^4 V×cm^{-1} in both cases, indicating that both pitch designs are electrically robust. The final choice will depend on the application's sensitivity to charge sharing and spatial resolution.

7.6 Breakdown optimisation and enhancement of design stability

Following the results presented in the previous sections, the 10k Ω·cm substrate was selected as the optimal choice. This decision was based on a balanced trade-off between depletion region size and electric field distribution beneath the collection node. A critical requirement for the sensor design is to maintain breakdown stability, particularly in this high-field region, which governs the charge collection performance and defines the maximum tolerable bias voltage.

A significant part of this design study focused on the role of surface metalisation and its impact on the electric field configuration and breakdown characteristics. This investigation was guided by both manufacturer recommendations and extensive TCAD simulations. Different metalisation designs surrounding the main collection node are presented and evaluated in Subsection 7.6.1.

The design and function of p-stops were another key topic in this iteration of the sensor. Guard rings and isolation structures are essential to control the onset of electrical breakdown,

especially at high bias voltages. This is particularly important in high-energy physics environments, where radiation tolerance and interpixel isolation are crucial. More background on the general concepts of isolation and guard ring techniques can be found in Section 3.6.

One of the main challenges in n-in-p type detectors is the presence of surface electron accumulation due to positive charges trapped in the silicon dioxide. This phenomenon intensifies with radiation exposure and requires effective isolation strategies. Three approaches are commonly used: p-spray, which is a uniform shallow $p-$type layer covering the entire surface; p-stop, which uses individually patterned $p-$type implants between pixels; and a hybrid method combining both. Studies show that careful optimisation of p-stop parameters is necessary to balance pixel isolation with high breakdown voltage and low leakage current. Improper calibration may compromise charge collection efficiency or result in premature breakdown [111].

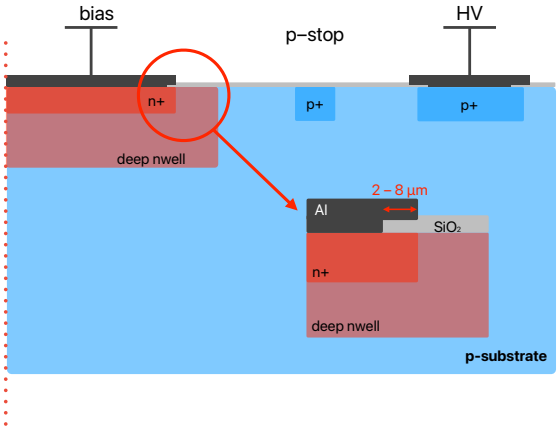
In the case of AstroPix, the sensor incorporates p-stops with fixed implant dimensions, which are 2 µm wide and 2 µm deep, using a Gaussian doping profile. The p-stop depth is defined as the point at which the doping concentration equals the bulk substrate doping. Figure 7.3 shows the area of interest where the p-stop is located with respect to the collection node.

While the geometry and doping concentration are fixed, the position of the p-stop relative to the edge of the deep nwell remains a tunable parameter. In this study, four placement distances were simulated and compared: 25, 35, 95, and 100 µm. The impact of each configuration on electric field strength and breakdown behaviour is analysed in Subsection 7.6.3.

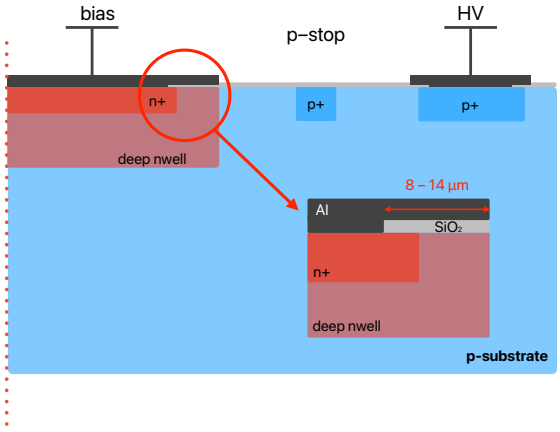
7.6.1 Metallisation

One of the most critical aspects in sensor design is the treatment of the surface region of the chip. This interface comprises a variety of materials, including silicon, silicon dioxide, and metals, which interact differently under electric fields, temperature variations, and radiation exposure. Moreover, surface effects near the edges of the active structures are often more pronounced and can significantly influence breakdown behaviour. In the AstroPix sensor, the first interface above the silicon bulk is the oxide layer, followed by successive metallisation layers. This study focuses on the lowest metallisation layer, that is, the one in closest proximity to the silicon surface, as it plays a key role in determining electric field intensity near the junctions.

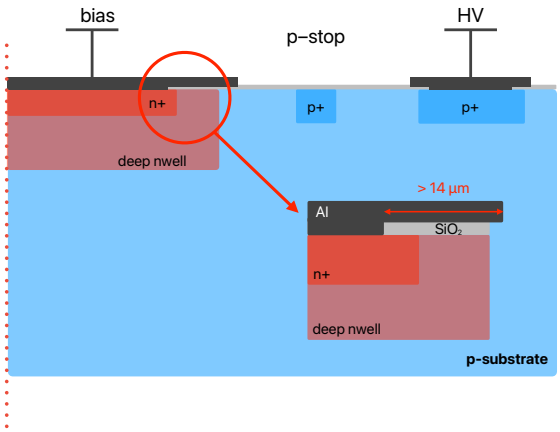
One approach to mitigate high electric fields at the surface, as recommended by the chip manufacturer, is to increase the lateral extent of the metal layer covering the oxide. This strategy reduces field concentration near critical junctions by introducing field plates. In this context, the *free pad length* is defined as the distance from the metal-silicon ohmic contact to the edge of the metal pad. As a part of this thesis, a series of TCAD simulations was performed with steady increases in the metal extension, using 2 µm steps. For all simulations, a bias voltage of -250 V was applied, and the substrate resistivity was set to 10k Ω·cm. Based on the resulting field configurations, three distinct regimes of metal coverage were identified. These are illustrated in Figure 7.7, and will be referred to as Zones 1, 2, and 3.



(a) Zone 1: metal coverage over the undepleted region of the nwell. The free pad length in this configuration ranges from approximately 2 µm to 8 µm.



(b) Zone 2: metal coverage extending over the depleted area of the nwell, up to the junction with the $p-$type region. The free pad length is between 8 µm and 14 µm.



(c) Zone 3: Metal coverage extending beyond the edge of the nwell and over the adjacent $p-$type region, without overlapping the p-stop. The free pad length exceeds 14 µm.

Figure 7.7: Three metalisation variations based on free pad length and alignment relative to the nwell and $p-n$ junction. All simulations assume a substrate resistivity of 10k Ω·cm and a bias of -250 V.

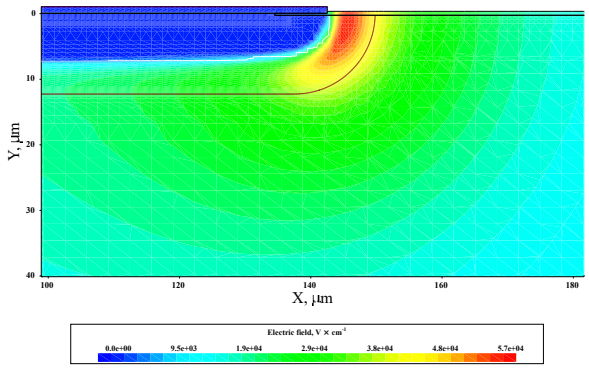
Zone 1 represents cases where the metal overlaps only the undepleted region of the nwell. In these cases, the extent of the depletion region inside the nwell depends on the doping profiles of the shallow and deep nwell structures, as well as the voltage applied to the collection node, which is 1.8 V in this case. Zone 2 corresponds to the metalisation reaching into the depleted region of the nwell but not beyond the junction with the surrounding $p-$type material. Zone 3

covers scenarios where the metal extends beyond the nwell boundary onto the p–type silicon, but still avoids encroaching upon the p-stop region.

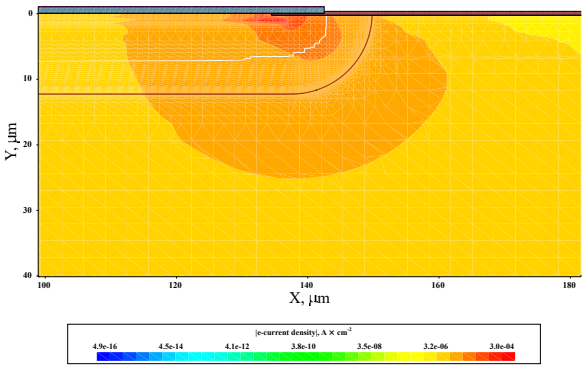
The primary outputs from these simulations, which will guide the final choice of metallisation layout, are the spatial distribution of the total electric field (in $\mathrm{V{\times}cm^{-1}}$) and the electron current density (in $\mathrm{A{\times}cm^{-2}}$). These results will be analysed and compared across the three zones to evaluate the design trade-offs between field suppression and layout constraints.

Zone 1: Metalisation above the undepleted region of the nwell

In the first configuration, the metal layer extends over the undepleted region of the deep nwell, without reaching into the depletion zone or the $p-n$ junction. Figure 7.8 illustrates the electric field distribution and electron current density for a representative free pad length of 6 µm.



(a) Electric field distribution under the metal pad (length: 6 µm). The maximum field strength reaches approximately 5.7×10^4 $\mathrm{V{\times}cm^{-1}}$, concentrated in the depleted part of the nwell.



(b) Electron current density map. The distribution is smooth, and the peak current density near the $p-n$ junction is approximately 1.6×10^{-5} $\mathrm{A{\times}cm^{-2}}$.

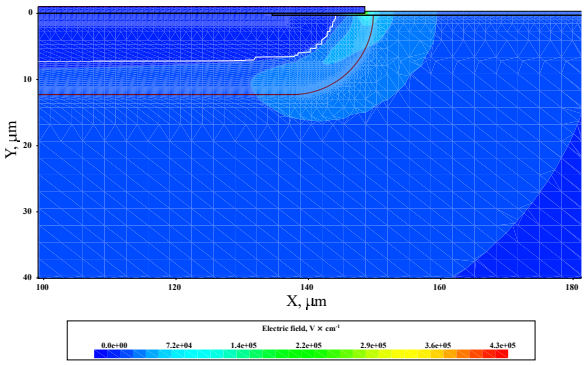
Figure 7.8: Simulation results for Zone 1 configuration with a 6 µm metal extension above the undepleted region of the nwell. No critical field concentrations or surface leakage are observed.

The simulation shows that the region at the metal–oxide–semiconductor interface remains electrically quiet, with low field strength. This helps reduce the risk of breakdown initiation at the surface. The highest electric field in this configuration is located within the depletion region of the nwell and reaches a peak value of approximately 5.7×10^4 $\mathrm{V{\times}cm^{-1}}$. Importantly, this peak is not located at the $p-n$ junction itself, but rather concentrated within the n–type region, away from critical boundaries.

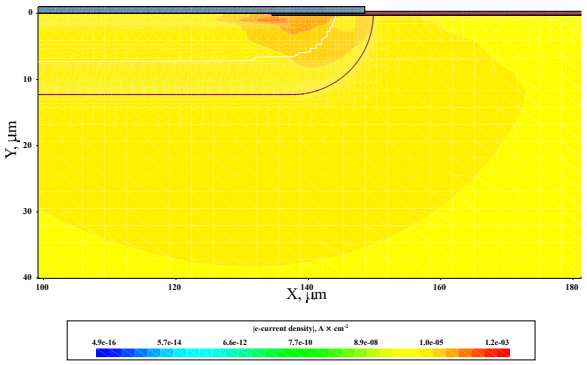
The electron current density in the junction region remains stable and uniform. The measured value is approximately 1.6×10^{-5} $\mathrm{A{\times}cm^{-2}}$, with no indication of localised surface leakage or dangerous accumulation zones. Overall, this design variant appears stable and well-suited for high-voltage operation. It effectively avoids surface field intensification and exhibits no adverse current effects under the simulated bias conditions.

Zone 2: Metalisation over the depleted region of the nwell

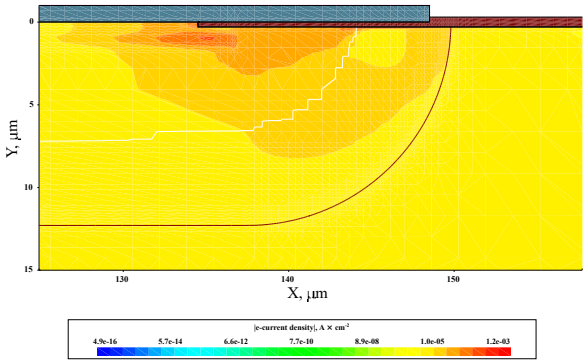
The second configuration involves extending the metal pad over the depleted portion of the nwell, up to the junction with the adjacent p−type region. In this case, a free pad length of 12 µm was selected for illustration. The resulting electric field distribution and electron current density maps are shown in Figure 7.9.



(a) Electric field distribution for 12 µm metal extension. Extremely high fields up to 3.8×10^5 V×cm^{-1} appear in the oxide.



(b) Electron current density distribution across the region of interest. A significant increase is observed at the oxide–silicon interface.



(c) Zoomed-in view of the high-density region at the surface. Electron current density peaks at approximately 6.1×10^{-4} A×cm^{-2}.

Figure 7.9: Simulation results for Zone 2 configuration with a 12 µm free metal pad length. The metal layer extends over the depleted nwell and approaches the p−type region. Elevated fields and surface currents are observed.

This configuration presents a more critical situation. The highest electric field is no longer confined to the bulk silicon but instead emerges at the oxide–semiconductor interface directly beneath the extended metal. The peak field values in this region range from approximately 3×10^5 to 8×10^5 V×cm^{-1}, which is within the values where surface breakdown becomes a serious concern. Corner regions are particularly vulnerable in semiconductor devices, as these are natural sites for defect formation. After irradiation, additional trapped charges, particularly holes, can accumulate in the oxide, further increasing the local field and the likelihood of

breakdown. For this reason, the oxide region under the metal extension becomes an area of high risk for long-term stability, particularly in environments with high radiation levels.

Within the silicon, the highest field remains in the depleted $n-$doped region, reaching values between 7×10^4 V×cm^{-1} and 1×10^5 V×cm^{-1}. These values increase as the metal extends closer to the $p - n$ junction, confirming that both metal geometry and placement significantly affect field concentration.

The electron current density map reveals a sharp increase in surface current at the silicon–oxide interface beneath the metal. This can be clearly seen in the zoomed view in Figure 7.9c, where the peak current reaches 6.1×10^{-4} A×cm^{-2}. The accumulation of current at this location suggests enhanced carrier transport at the surface, likely driven by the high field in the oxide and the close proximity of the metal. This behaviour is consistently observed across all simulations within the Zone 2 range and must be carefully considered when selecting a safe and stable metallisation length.

Zone 3: Metalisation beyond the nwell, over $p-$type silicon

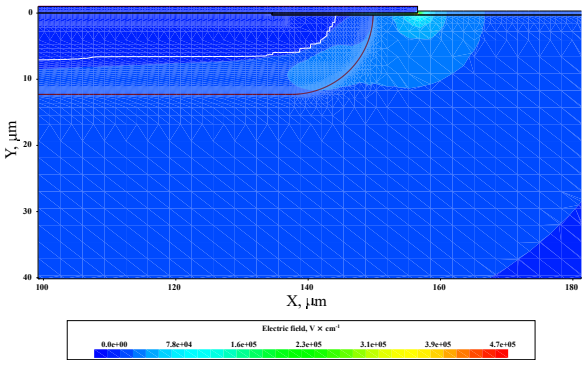
The third configuration places the metal layer entirely outside the nwell region, extending beyond the junction and over the $p-$doped silicon area. In this case, a free pad length of 20 µm was used as a representative example. The simulation results are shown in Figure 7.10.

This layout exhibits the most critical behaviour among the three zones. Similar to Zone 2, the highest electric field values occur within the oxide layer at the surface, reaching levels on the order of 10^5 V×cm^{-1}. These values stabilise slightly after crossing the $p - n$ junction, with typical field magnitudes around 4.4×10^5 V×cm^{-1}. In the $p-$doped silicon underneath the metal, the electric field ranges between 6.8×10^4 V×cm^{-1} and 1.2×10^5 V×cm^{-1}, which approaches the threshold for avalanche breakdown. While not yet at critical levels, this range signals a significant risk, particularly under radiation exposure or temperature cycling.

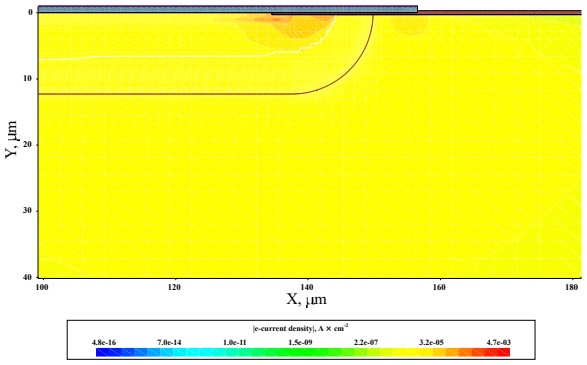
The most concerning feature in this simulation is the formation of a high-density electron channel near the silicon surface, directly beneath the metal layer. This effect mirrors what was seen in Zone 2 but is even more pronounced. The zoomed current density plot in Figure 7.10c shows a uniform surface layer where the electron current density reaches 3.4×10^{-3} A×cm^{-2}. This conductive path spans approximately 5 µm from the edge of the deep nwell and indicates a sustained current transport channel along the oxide–silicon interface. Such a configuration is not only susceptible to surface breakdown but may also compromise charge isolation between adjacent pixels. Although the peak fields in the silicon remain below the hard breakdown threshold, the presence of a long and active surface channel makes this layout unfavourable for robust and radiation-hardened detector applications.

After reviewing the individual behaviours of Zones 1,2, and 3, it is essential to consolidate the findings and explore the shared physical effects at the metal–oxide–semiconductor interface.

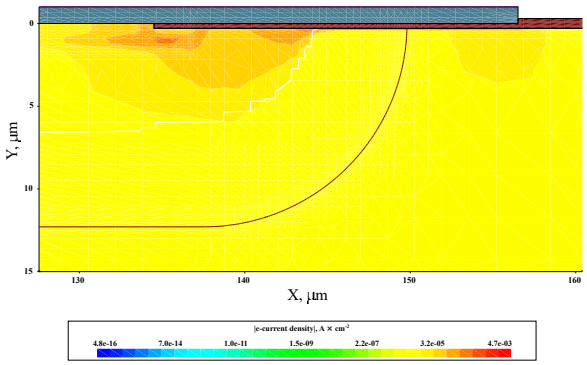
The use of shallow trench isolation (STI) structures is one method to mitigate surface conduction and reduce interpixel leakage. However, as illustrated in this study, STI alone is not always sufficient. In the specific example presented, a 100 nm gap remained between STI boundaries,



(a) Electric field distribution for a 20 µm metal extension beyond the nwell. Field stabilises in the oxide around 4.4×10^5 V×cm^{-1}.



(b) Electron current density across the simulation region. The current concentration appears to be near the oxide–semiconductor interface.



(c) Zoomed-in view of the surface current channel beneath the metal. Electron current density peaks at approximately 3.4×10^{-3} A×cm^{-2}.

Figure 7.10: Simulation results for Zone 3 configuration with metal extended 20 µm beyond the nwell and onto the p–type substrate. High surface fields and current channels are evident.

allowing for the formation of a continuous electron channel along the silicon–oxide interface. The primary purpose of STI is to electrically isolate active regions from one another, and its effectiveness is typically assessed by measuring the leakage current between adjacent pixels. In high-voltage detectors, surface isolation becomes especially critical, as even modest surface fields can lead to sustained leakage through inversion layers or defect-assisted conduction paths.

Across all simulated designs, the vertical depletion depth remains constant at approximately 400 µm, as determined by the substrate resistivity and applied bias. The metal layer layout does not significantly affect the depth of the depleted volume. However, the lateral spread of the depletion region is influenced by the metallisation pattern.

- For a 6 µm metal extension (Zone 1), the lateral depletion region extends about 290 µm from the edge of the deep nwell.
- For a 12 µm extension (Zone 2), this distance remains similar, at approximately 290 µm.

- For a 20 µm extension (Zone 3), the lateral depletion extends slightly further, reaching 295 µm.

Although the variation is modest, this increase in lateral depletion width corresponds to the expanded surface field distribution introduced by the longer metal pad. As the metallisation extends further towards or beyond the $p - n$ junction, the lateral electric field gradients also increase, affecting both charge collection symmetry and breakdown susceptibility.

To fully understand the challenges associated with these configurations, it is necessary to consider the underlying physics of the metal–insulator–semiconductor (MIS) interface. This region, often overlooked in simplified models, governs many critical aspects of sensor performance, including surface leakage, field inversion, and long-term stability under irradiation. Further discussion of these effects and physical models will follow in the next Subsection.

7.6.2 Metal–insulator–semiconductor (MIS) physics and its implications for sensor design

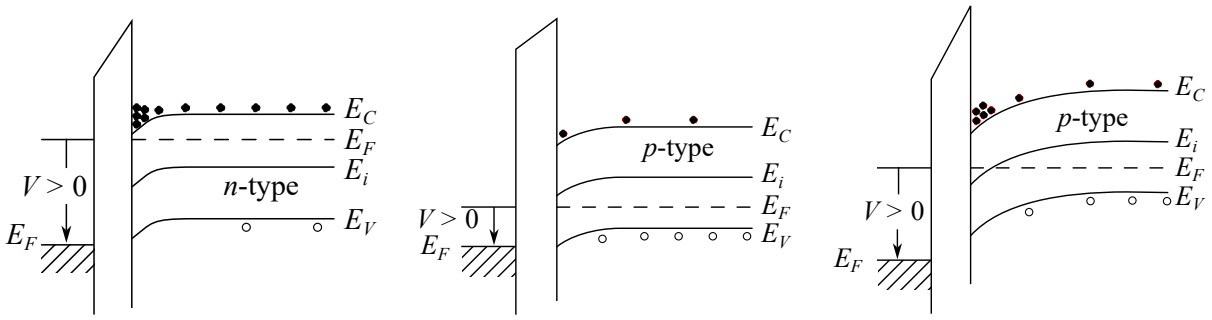
The metal–insulator–semiconductor (MIS) capacitor is a fundamental structure used to study semiconductor surface physics. Since many challenges related to device variability, reliability, and stability originate at the surface, MIS capacitors offer essential insight into surface effects across a wide range of semiconductor technologies, including most silicon-based planar devices and integrated circuits [2]. The metal–oxide–silicon (MOS) capacitor is the most widely studied form of MIS structure, particularly due to its role in the MOSFET. Its development dates back to 1959 when Moll, Pfann, and Garrett proposed it as a voltage-controlled varistor. The first successful MIS device with thermally grown SiO_2 on a silicon surface was fabricated in 1960 by Ligenza and Spitzer, and this directly led to the development of the MOSFET by Kahng and Atalla. A detailed treatment of the MOS system can be found in Nicollian and Brews [112].

An ideal MIS capacitor is used to illustrate basic surface states: accumulation, depletion, and inversion. In a p–type semiconductor, the application of a negative gate bias bends the energy bands upward, increasing hole concentration near the surface, which is called the accumulation regime. As the voltage becomes more positive, holes are repelled, leading to depletion. Further increase in positive bias inverts the surface, generating more electrons than holes, which is the inversion regime. These three cases are foundational to understanding surface behaviour under bias [2].

Figure 7.11 shows ideal energy band diagrams for MIS capacitors under different regimes. The first image illustrates accumulation in an n–type substrate (Zone 1 or Zone 2 scenario), while the latter two show depletion and inversion in a p–type substrate (Zone 3 context).

Under strong positive bias, deep inversion occurs where electrons accumulate near the surface of a p–type semiconductor, forming a conductive channel. This state is critical to understand, especially in radiation environments where surface charge trapping becomes significant.

Figure 7.12 shows an idealised MIS energy-band diagram under strong inversion, along with the resulting electric field and charge distribution. For surface charge neutrality, the total



(a) $n-$type in accumulation **(b)** $p-$type in depletion **(c)** $p-$type in inversion

Figure 7.11: Energy band diagrams for ideal MIS capacitors under different biasing conditions, illustrating accumulation, depletion, and inversion at the semiconductor surface [2].

charge on the gate balances the combined contributions of inversion electrons and ionised acceptors in the depletion region:

$$Q_M = -\left(Q_n + q \times N_A \times W_D\right) = -Q_s \tag{7.1}$$

Here, Q_M is the charge per unit area on the metal; Q_n represents the electron charge per unit area near the surface of the inversion region; $q \times N_A \times W_D$ corresponds to the ionised acceptor charge per unit area in the space-charge region, where W_D is the depletion width; and Q_s denotes the total charge per unit area in the semiconductor. The applied voltage appears across both the oxide and the semiconductor [2].

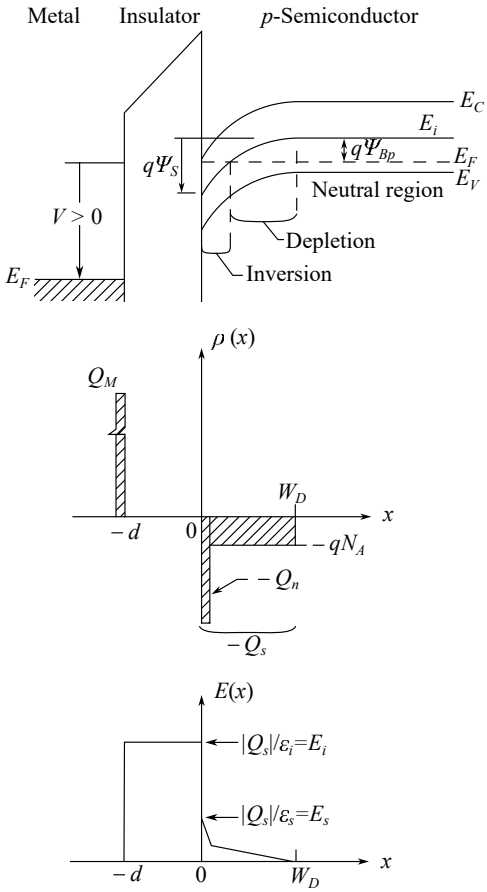


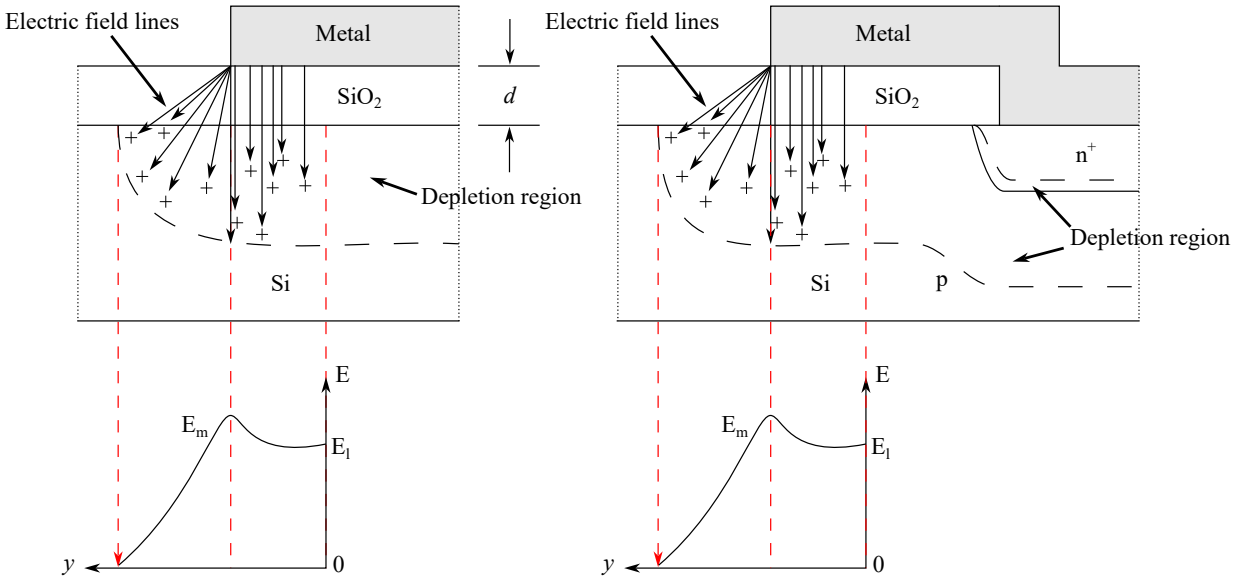
Figure 7.12: Band diagram, charge, and electric field profiles for an ideal MIS capacitor in strong inversion [2].

Breakdown in deep depletion

The breakdown behaviour of MIS capacitors in the deep-depletion regime has been studied in detail. Goetzberger and Nicollian first reported that avalanche breakdown could occur near the edge of the depletion region for oxide thicknesses around 0.1 µm [19]. Such a breakdown becomes especially relevant for fast ramp rates, where thermal generation cannot supply minority carriers fast enough, leading to nonequilibrium. This deep depletion or inversion is depicted in the band diagram of Figure 7.11c.

Two specific breakdown mechanisms are shown in Figure 7.13. Figure 7.13a illustrates electric field distribution and edge breakdown in a standard MIS capacitor, while Figure 7.13b shows avalanche onset in a metal-overlapped planar diode. In both cases, peak fields appear at the oxide–semiconductor interface or near electrode peripheries [2].

Edge effects become dominant when the depletion width is large and the oxide thickness is small. Maintaining a ratio of $d/W_{\max} > 0.3$ (oxide thickness to maximum depletion width) is recommended to minimise field peaks. Interestingly, at low doping levels, MOS capacitors may have lower breakdown voltages than equivalent $p - n$ junctions due to enhanced surface field concentration [2].



(a) MOS capacitor: edge breakdown near the oxide interface.

(b) Metal-overlapped planar diode: breakdown localised at edge.

Figure 7.13: Breakdown phenomena in deep-depleted MOS capacitors and related structures [113].

Dielectric breakdown and percolation theory

Another failure mechanism relevant to the Zone 2 and Zone 3 configurations is dielectric breakdown, especially in thick oxides. Under high bias, tunnelling currents or energetic carriers create localised damage in the oxide. Over time, this leads to defect accumulation. Once a critical density of defects is reached, a percolation path forms between the gate and the substrate, leading to catastrophic failure [2].

Figure 7.14 illustrates this breakdown via percolation theory. When the density of randomly generated and pre-existing defects becomes sufficient to form a continuous conductive path, dielectric breakdown occurs. This phenomenon is a key reliability concern, especially for long-term biasing in radiation environments. The total stress time before breakdown is known as the time-to-breakdown (D), and it is a critical metric in time-dependent dielectric breakdown (TDDB) analysis [2].

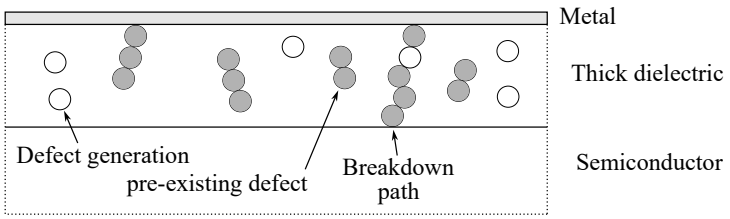


Figure 7.14: Breakdown via percolation theory. Defect accumulation in the dielectric forms a conduction path between gate and substrate [2].

From all observations, it is clear that Zone 2 and Zone 3 show critical drawbacks. Both cases result in high electric fields at the surface, increasing the risk of inversion or deep depletion, and exhibit signs of possible breakdown, either through avalanche or dielectric failure. In contrast, Zone 1 shows low surface fields, no formation of surface electron channels, and stable electron current density. This makes it the most reliable and robust option for this sensor design.

7.6.3 Influence of p-stop on electric field distribution

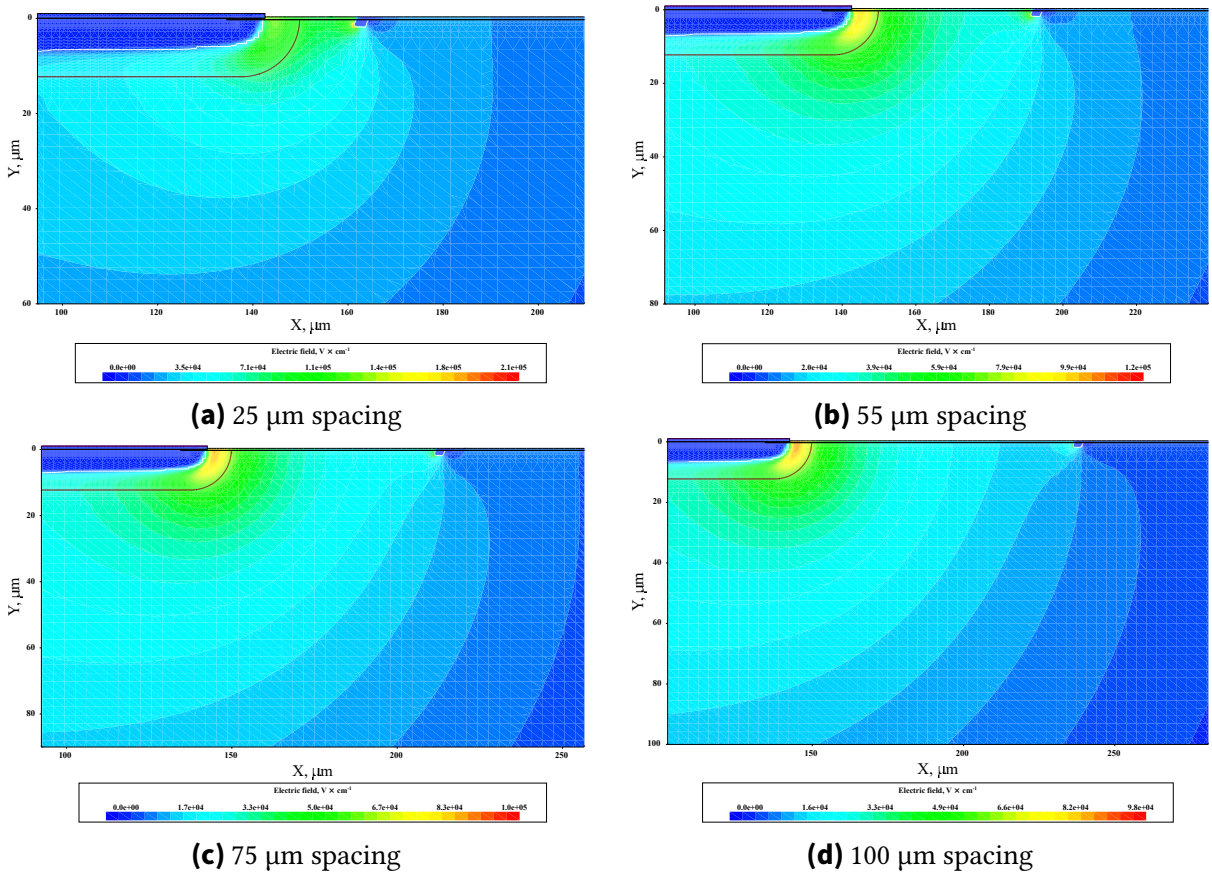
The p-stop plays a key role in providing pixel-to-pixel isolation in n-in-p type sensors. This isolation is necessary to suppress the accumulation of surface electrons, which results from positive fixed charges in the silicon dioxide. These positive oxide charges attract mobile electrons to the silicon surface, forming an inversion layer. This effect becomes more pronounced with increasing radiation fluence [111]. For a broader overview of isolation techniques, including p–spray and combined methods, see Section 3.6.

While p-stops are essential for charge isolation, they also affect the local electric field distribution, and therefore must be carefully optimised. Incorrect placement can lead to field crowding and increase the risk of breakdown. To maintain high breakdown voltages and interpixel resistance, the distance between the p-stop and the main collection node (deep nwell) must be selected with care. The technique, often used to mitigate breakdown, involves the application of floating guard rings, which reduce field crowding at the implant edges. These rings deplete gradually with increasing reverse bias, effectively reshaping the field distribution into a more uniform profile, similar to a parallel-plane junction [114]. However, in this study, we focus on floating p-stop spacing without guard rings.

To investigate the optimal placement, TCAD simulations were performed for four configurations with p-stop-to-nwell distances of 25 µm, 55 µm, 75 µm, and 100 µm. The applied bias voltage was -500 V, and a substrate resistivity of 10k Ω·cm was used in all cases. Figure 7.15 shows the resulting electric field maps for each configuration.

The electric field strength at the inner edge of the p-stop is the main criterion of interest. For the 25 µm spacing, the peak field reaches approximately 2.1×10^5 V×cm^{-1}, which is already close to the breakdown limit of a silicon $p-$junction. However, at 55 µm spacing, this value is significantly reduced to around 1.1×10^5 V×cm^{-1}, which is already within a safer operational range. The smallest distance that consistently results in a stable electric field profile is 60 µm, where the field value drops to around 1.0×10^5 V×cm^{-1}. Further increasing the distance improves the situation slightly: the field reduces to 8.8×10^4 V×cm^{-1} at 75 µm, and 6.8×10^4 V×cm^{-1} at 100 µm.

This behaviour shows that the electric field strength at the p-stop edge is not significantly affected by the metal pad length, but rather by the spacing from the nwell. Once the p-stop is placed further than 60 µm from the collection node, the local field stabilises, and the risk of breakdown is significantly reduced. Since the overall sensor layout also benefits from a large pixel pitch for charge-sharing purposes, and because the high-field zone remains suppressed beyond 60 µm, a spacing of 100 µm was chosen as the final design value.



(a) 25 µm spacing

(b) 55 µm spacing

(c) 75 µm spacing

(d) 100 µm spacing

Figure 7.15: Electric field distribution between the collection node (deep nwell) and the p-stop for various spacings. Simulated with a -500 V bias and 10k Ω·cm substrate.

7.7 Outcome and recommendation

This design study focused on improving the depletion, breakdown stability, and charge collection uniformity of the sensor implemented in AstroPix_v3. It addressed the limitations of AstroPix_v2, including shallow depletion, high leakage currents, and electric field instabilities at high voltage, as described in Section 7.3.

The simulation results led to the following outcomes:

A substrate resistivity of 10k Ω·cm was selected to achieve nearly full depletion at the nominal bias of -250 V and ensure operational robustness at -500 V. Among the simulated values (20 Ω·cm, 370 Ω·cm, 5k Ω·cm, 10k Ω·cm, and 18k Ω·cm), this option provided the best compromise between uniform depletion depth and electric field stability. Lower-resistivity substrates failed to fully deplete or introduced breakdown risk; higher resistivities introduced non-uniformities in the depleted volume. The full simulation results and analysis are provided in Section 7.5.

The pixel pitch of 500 µm, with a 300 µm-wide deep nwell collection node, was shown to successfully meet the spatial resolution and charge-sharing requirements of the AMEGO-X

telescope. The telescope itself is described in Section 7.1. TCAD simulations confirmed that the lateral depletion depth for this configuration sufficiently overlaps adjacent pixels, supporting high-resolution event localisation in the Compton scattering regime. This is discussed in detail in Section 7.5.

Surface breakdown was a key design concern. Three configurations of metal coverage were simulated and compared in Subsection 7.6.1. Placing the metal pad directly above the undepleted nwell (Zone 1) provided the most stable electric field distribution and prevented surface current formation. Configurations overlapping depleted silicon or extending above $p-$type regions (Zones 2 and 3) led to strong electric field enhancement near the silicon–oxide interface and formation of surface electron channels, both of which are critical for dielectric breakdown, particularly in irradiated detectors. Based on these observations, Zone 1 metalisation was adopted in AstroPix_v3.

The p-stop design was optimised by simulating electric field profiles for various spacings between the deep nwell and p-stop, as presented in Subsection 7.6.3. A distance of 100 µm was selected as the most reliable option. It minimised electric field peaks near the p-stop corners, which otherwise approached breakdown thresholds at shorter spacings. This configuration also aligned with the large pixel pitch and maintained sufficient inter-pixel isolation for radiation environments.

In summary, the following sensor configuration was selected and implemented in AstroPix_v3:

- substrate resistivity: 10k Ω·cm;
- deep nwell size: 300 µm;
- pixel pitch: 500 µm;
- metal placement: Zone 1 configuration (above undepleted nwell);
- p-stop spacing: 100 µm from deep nwell.

These design improvements directly support the goals of the AMEGO-X gamma-ray telescope by ensuring a robust, low-leakage, and highly depleted sensor. The validated TCAD-driven layout offers stable operation under high voltage, with minimal surface breakdown risk and high inter-pixel isolation, which is essential for both event reconstruction and radiation hardness in space applications.

8 Investigation of heavy-ion beam influence on HV-CMOS sensors

This chapter investigates the issue of particle detection efficiency loss in HV-CMOS sensors under high-intensity heavy-ion beams. The problem was first identified during the development of a monolithic active HV-CMOS hit-counting pixel sensor named HitPix, described in detail in Chapter 6. Designed for use in radiation therapy with high-rate ion or proton beams, this detector is intended as a candidate for replacing the current beam monitoring system at the Heidelberg Ion Beam Therapy Centre (HIT). In such medical applications, the accuracy and stability of the sensor response are of critical importance. During the beam test campaigns conducted to validate the HitPix sensor family, a recurring effect was observed: a sudden and prolonged drop of the voltage signal from the in-pixel amplifier. Initial observations prompted further analysis, leading to the hypothesis that a form of Single Event Effect (SEE), such as Single Event Latch-up (SEL), could be responsible.

To understand the origin and mechanism of this behaviour, a comparative study was undertaken using five sensor designs: **HitPix2**, **HitPix3**, **AstroPix** (more work on this chip can be found in Chapter 7), **ATLASpix**, and **MuPix**. These devices differ in their front-end electronics, pixel layout, and well structures, providing a systematic basis to isolate the key architectural parameters driving the drop effect. Complementary to experimental beam data, advanced TCAD simulations with full CMOS layout extraction were performed, further supporting the SEL hypothesis. This chapter presents the measurement results, identifies correlations between chip architecture and drop occurrence, and proposes design solutions to mitigate the observed behaviour.

8.1 Efficiency loss under heavy-ion beam irradiation

During the development of the HitPix family of HV-CMOS sensors, beam tests were conducted to assess the chip's suitability for high-radiation environments. However, repeated measurements revealed a recurring issue: a sudden and sustained drop in both the output of the in-pixel amplifier and the overall signal baseline during ion irradiation. This led to a smaller value in the recorded voltage signal from the in-pixel amplifier, such that subsequent signals often failed to reach the comparator threshold. This phenomenon, referred to as *baseline drops* or *voltage drops*, was consistently observed across different HitPix versions and under a range of ion species and beam conditions.

Early tests were conducted at the Heidelberg Ion-Beam Therapy Center (HIT), a facility capable of delivering a broad spectrum of ion species, including protons, carbon, helium, and oxygen, which are extracted from the sources, accelerated in a linear accelerator, and injected into the compact synchrotron after all electrons have been removed. This system offers 255 energy settings, 15 intensity steps, and 6 spot sizes per ion type [6, 88]. More about the HIT facility can be found in Section 6.1.

To ensure compatibility with this environment, both HitPix1 and HitPix2 sensors, including variants with standard (STD) and isolated (ISO) pixel topologies, were investigated with high-energy ion beams. In each case, signal efficiency degradation was observed during operation. Figure 8.1 illustrates a representative case recorded under carbon beam irradiation. The sensor output exhibits irregular but repeatable drops in baseline voltage, which significantly compromise event detection efficiency.

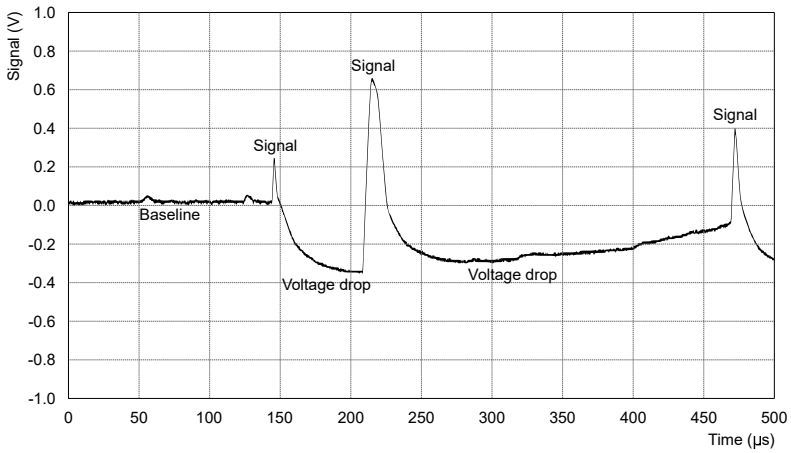


Figure 8.1: Example of the observed voltage drop effect in a HitPix2 chip during carbon beam exposure at HIT. Beam parameters: carbon ions, 430.1 MeV/u, intensity 2×10^6 particles/s, beam width 3.4 mm. The observed drop in signal occurs despite a constant amplifier output rate of approximately 0.035 MHz. Image adapted from [95].

To rule out artefacts specific to the HIT environment, additional tests were performed at the Ruđer Bošković Institute's microbeam facility [115] in Zagreb in March 2023, using low-energy proton beams [116]. The setup provides a finely focused proton beam with a spot size of 10×10 µm^2, allowing sub-pixel precision irradiation. The beam energy was tunable between 1, 2, and 3 MeV, corresponding to stopping ranges of 16, 47, and 92 µm in silicon, respectively. Beam intensities of 1–2 kHz were used throughout the study.

The test station consisted of a HitPix chip mounted on a carrier board and connected to the Generic Configuration and Control Setup board (GECCO) [8], positioned on a precision mechanical stage. Beam alignment was achieved using a dedicated beam nozzle, ensuring accurate irradiation of targeted sensor regions. Samples included HitPix2 STD, HitPix1 STD, and HitPix1 ISO. The high spatial resolution and controlled energy deposition made this setup ideal for investigating localised charge effects and their influence on the observed voltage drops.

The behaviour was again reproduced: baseline drops occurred irrespective of the applied bias voltage but showed dependence on beam energy and pixel location. Key observations included:

- HitPix2 STD exhibited signal drops within a well-defined rectangular region of the sensing diode (200 µm × 100 µm size), with amplitude and duration modulated by beam position. At a lower bias (e.g. −3 V), drops occurred in every second signal; at a higher bias (−50 V), nearly every signal was affected.
- HitPix1 showed greater resilience, particularly in the ISO topology. At 1 MeV, protons failed to induce a signal; at 2 MeV, drops were observed only in the STD cell but not in ISO, and with much lower frequency compared to HitPix2 STD. Signal amplitude from the HitPix1 amplifier was approximately four times larger than that of HitPix 2 for equivalent irradiation conditions.

These findings indicated that not only the beam parameters but also the analogue front-end design and pixel topology strongly influence susceptibility to the effect. In summary, ISO topologies and earlier amplifier versions, as in HitPix1, proved more tolerant to signal degradation during this test campaign.

The origin of the signal drop was initially hypothesised to relate to charge build-up between neighbouring nwell regions, which can form parasitic conduction paths or localised fields under prolonged ion exposure. While p-stop structures are traditionally used to mitigate inter-well accumulation, this mechanism typically manifests after extended operational lifetimes and was considered unlikely to explain the immediate response seen in these tests. A similar effect in long-term detectors is discussed in the work of CMS Collaboration [111], which provides further background on inter-well charge effects under radiation.

To explore the phenomenon in more detail, TCAD simulations were conducted throughout the first half of 2023. These included breakdown characterisation, charge collection dynamics, and timing simulations under proton beams in the 200–300 MeV/n range. These simulations are presented in Chapter 6, which provides a detailed investigation of the HitPix chip family. While helpful in estimating sensor performance, such simulations lacked the granularity to capture transistor-level effects, such as localised current spikes or shorting events, that might cause sudden baseline collapse.

More advanced simulations were thus initiated, including full CMOS layout extraction. These pointed toward a plausible cause: a Single Event Effect (SEE), more specifically, Single Event Latch-up (SEL). In this scenario, the heavy ion induces a parasitic thyristor structure between the $p-$ and $n-$type regions of the CMOS, triggering a high-current state that cannot be cleared without power cycling. This theory aligns well with the observed persistent signal drop and its dependence on beam energy and localisation.

By early 2024, the voltage drop effect had been identified as a recurring issue across multiple HitPix versions and was consistently observed under different ion species and test environments. Beam tests conducted at HIT from January to June 2024 included an additional HV-CMOS sensor design. This allowed for evaluation of whether the observed effect was specific to the HitPix architecture or more generally related to HV-CMOS technologies. The

following section outlines the theory of SEE in CMOS circuits, followed in the next sections by test results obtained from different HV-CMOS sensor structures.

8.2 Influence of single-event effects (SEE) on CMOS architecture

The continued downscaling of CMOS transistor geometries has introduced new challenges in terms of radiation susceptibility, particularly for single-event effects (SEE). While advanced CMOS technologies generally show improved tolerance to total ionising dose (TID), especially the damage induced by the accumulation of charge in the oxide layers, they remain vulnerable to SEEs [117]. SEE phenomena are a significant concern in space electronics and defence systems, where heavy ions and other high-energy particles can penetrate microelectronic structures [118]. A highly energetic ion can generate a large number of electron–hole pairs, potentially leading to both temporary effects, such as single-event latch-up, snapback, functional interrupts, or multiple upsets, and permanent damage, including single-event burnout or gate rupture. These interactions may result in logic errors, transient circuit disturbances, potentially compromising the device's normal operation [117], or complete device failure. For a SEE to occur, the energy deposited by the ion must exceed a specific threshold characteristic of the affected device [119].

Among the different categories of SEE, a potentially destructive one is the single-event latch-up (SEL). This occurs when the charge induced by a heavy ion triggers a parasitic feedback loop between transistors in a CMOS circuit. This latch-up state can lead to permanent damage unless the power is removed to break the loop [118, 120]. The trigger conditions for SEL include not only particle species but also energy, strike location, and angle of incidence. Particles known to initiate latch-up include high-energy photons, alpha particles, cosmic rays, heavy ions, and protons [118, 120]. For heavy ions, both direct ionisation and nuclear interactions (e.g. silicon fission) can contribute to the charge deposition necessary to trigger latch-up [119].

Latch-up can also occur due to minority carrier diffusion into well regions or from interactions with parasitic structures that are not fully isolated. These mechanisms depend not only on the properties of the incident particle but also on detailed aspects of the layout geometry, doping profiles, and isolation schemes. As such, evaluating latch-up sensitivity requires both physical analysis and simulation, particularly when investigating HV-CMOS sensors intended for use in high radiation environments [121, 117, 120].

A comprehensive overview of latch-up mechanisms and mitigation strategies can be found in Voldman's textbook "Latchup in Semiconductor Technology" [120], which provides an in-depth analysis of thyristor behaviour, charge transport, and failure conditions in modern CMOS circuits.

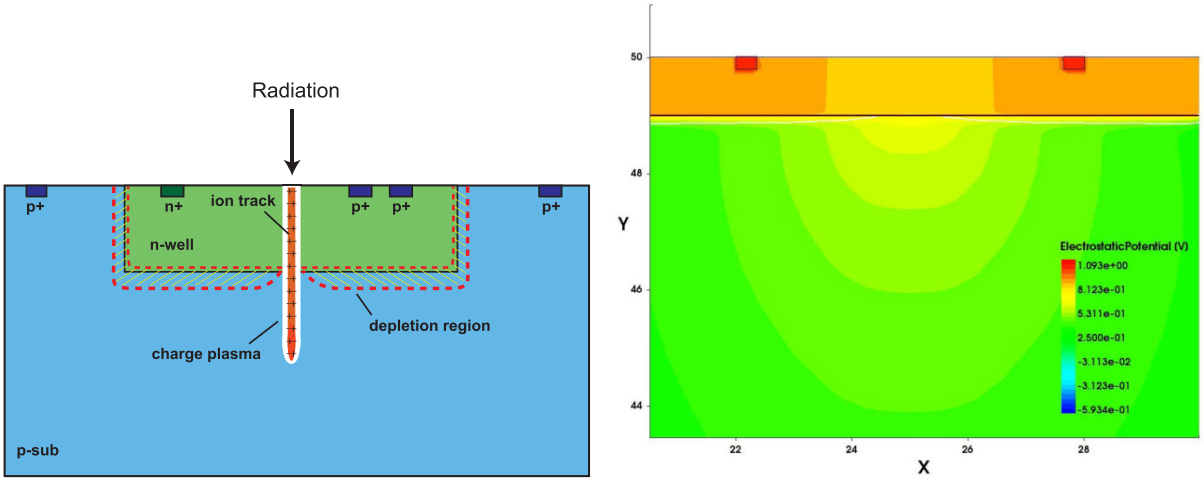
8.2.1 The funnelling effect

An important mechanism that contributes to the onset of single-event effects in CMOS devices is the *funnelling effect* [117]. This phenomenon occurs when a highly energetic ionising particle traverses the space charge region of a semiconductor device, typically across a $p-n$ junction. Crucially, the electron-hole pairs generated by this radiation must have a density higher than the local doping concentration within that space charge region. As the particle deposits this energy along its path, it generates a dense track of electron–hole pairs [121]. This track behaves as a highly conductive channel with low electrical resistance, which is often referred to as a "plasma wire" [117].

The funnelling effect significantly alters the electric field distribution and charge transport mechanisms within a semiconductor device [117]. In an idle state, the electric field is confined to the depletion region of the $p-n$ junction. However, the injection of high carrier density from the radiation track causes the electric field to spread into previously neutral regions of the substrate. Under these conditions, the transport of charge in quasi-neutral regions transitions from pure diffusion to drift-dominated behaviour. This leads to the collapse of the space charge region; on the other hand, this plasma-like track behaves as an extension of the depletion region induced by the sudden charge concentration. This enables rapid and efficient charge collection well beyond the static depletion width. As a result, charge may be collected from a substantially larger volume than expected, effectively increasing the sensitive region of the device and enhancing the likelihood of triggering a parasitic conduction path [121].

Figure 8.2 illustrates the visualisation of the funnelling effect and its impact on the electric field distribution. Figure 8.2a shows a schematic diagram of the ion track collapsing the space charge region due to excess carrier generation. Figure 8.2b presents the simulated potential distribution in a similar structure at 0.1 ns after the particle strike, obtained using Sentaurus TCAD. The potential of the nwell visibly extends into the p–type substrate along the ion path, indicating that the $p-n$ junction barrier has effectively vanished in that region. As a result, the carriers within the track experience an almost linear potential gradient and are quickly collected via drift. Once the excess carrier density falls below the doping level, which typically occurs within a few nanoseconds, the original depletion region begins to recover, and the remaining carriers are collected by slower diffusion.

This phenomenon has significant consequences for the current response of the device, and it is particularly critical in bulk CMOS technologies. In the absence of funnelling, excess carriers travel by diffusion, resulting in slower, longer-lasting current pulses that may spread to multiple circuit nodes. In contrast, funnelling causes most of the generated charge to be collected by the impacted node, producing short, high-current pulses that typically decay in less than 1 ns. These rapid current transients are capable of forward biasing internal junctions and activating parasitic bipolar transistors, for example, between nwell and p–substrate [117]. If the funnelling effect causes sufficient carrier injection into these parasitic structures, a latch-up event may be initiated through positive feedback between the vertical *pnp* and lateral *npn* transistors [121]. Simulations have shown that funnelling currents not only peak higher and decay faster than diffusion-dominated currents, but also result in several times greater total charge collection.



(a) Schematic illustration of the funnelling mechanism. The space charge region collapses along the ion track as the density of generated carriers exceeds the local doping concentration [117].

(b) Sentaurus TCAD simulation showing the 2D potential distribution 0.1 ns after particle impact. The potential drop along the $p-n$ junction (brown line) indicates the funnelling effect with the distortion of the potential in the structure [117].

Figure 8.2: Visualisation of the funnelling effect and its impact on electric field distribution [117].

8.2.2 Simulation of the funnelling effect for the HV-CMOS designs

Several approaches exist for modelling single-event effects in microelectronics. A widely adopted method relies on compact models, where the impact of radiation is emulated by injecting a parasitic current pulse into a standard circuit simulation environment (e.g. SPICE). This technique uses empirical current waveforms to replicate the transient behaviour caused by an ion strike. While this method enables the fast and convenient simulation of large circuits, its accuracy depends heavily on the assumed current shape and amplitude, and it does not capture underlying physical mechanisms, such as field redistribution or charge transport [117].

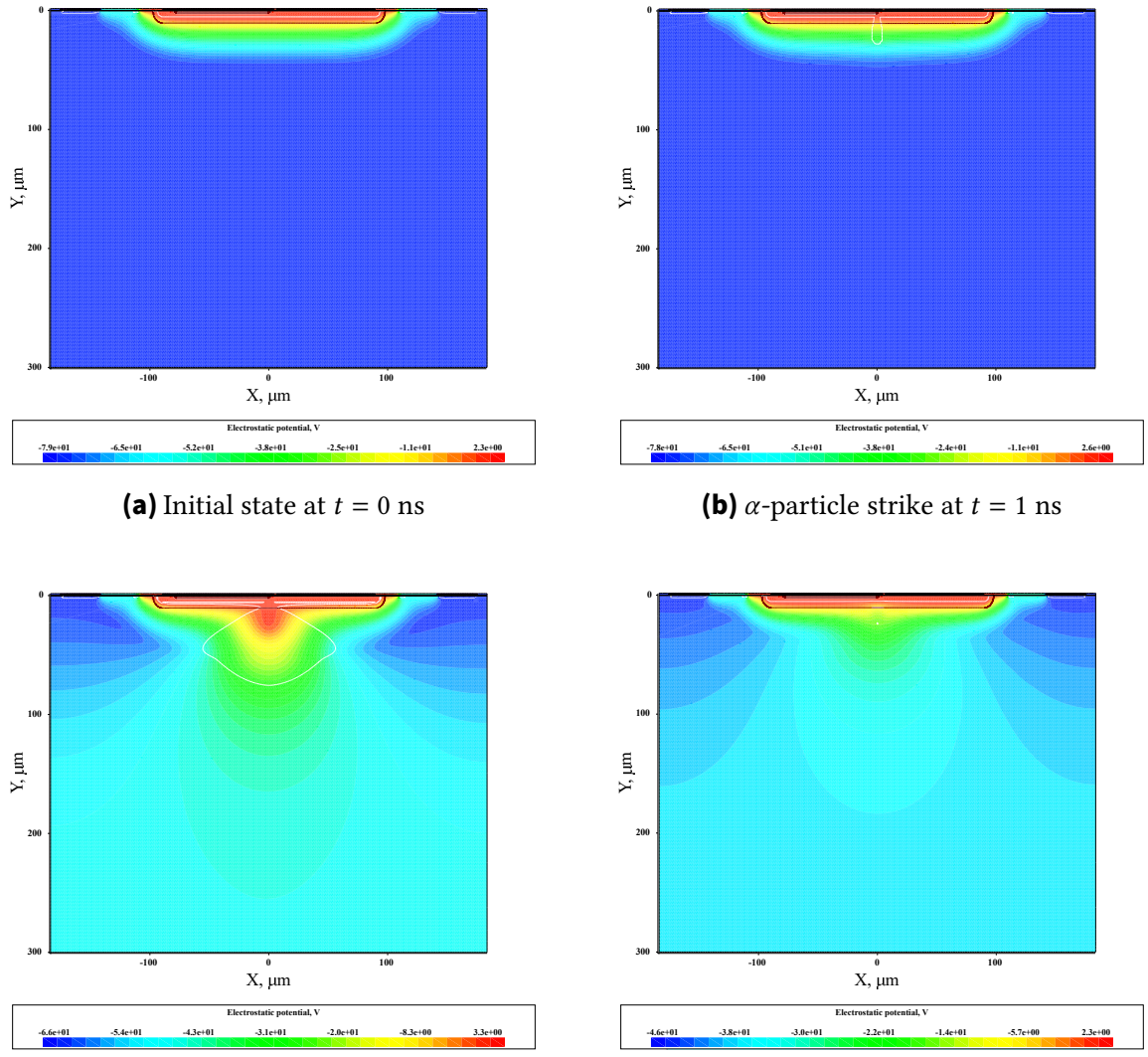
An alternative, more physics-based approach involves the use of the Generalised Lumped Devices (GLDs) model. In this framework, the semiconductor substrate is described as a network of virtual devices that model the generation, propagation, recombination, and collection of radiation-induced carriers. The funnelling effect is modelled explicitly by introducing a resistance element, R_{funnel}, between the neutral $p-$ and $n-$type regions of the device. Under normal conditions, this resistance is high, acting as an open circuit. However, when the excess carrier density exceeds the doping level, which is the condition for the funnelling effect, R_{funnel} drops to a low value, effectively shorting the junction and allowing charge to move freely between regions. This reflects the physical collapse of the junction potential and allows accurate modelling of high-current transients without relying on fitted parameters.

In the work of Rossi and Sallese [122], a GLD-based simulation methodology was proposed for modelling soft errors caused by ionising radiation, fully compatible with standard SPICE tools, such as TCAD *Sentaurus Device*. This approach enables the simulation of charge transport

from alpha particles or other ion species within the substrate, using only technological and geometrical parameters of the structure without empirical fitting. The model has been successfully applied to simulate the current collected by nwell, which represents the source/drain of NMOS transistors, in a $p-$type substrate, impacted by alpha particles of 1–10 MeV energy. This method provides both improved accuracy and physical insight compared to conventional current-pulse injection schemes.

To support the theoretical description of the funnelling effect, dedicated TCAD simulations were performed for a 5 MeV α-particle traversing the ISO pixel structure of the HitPix chip. Unlike minimum ionising particles (MIPs), which produce only weak perturbations, α-particles are high-LET projectiles that deposit charge densely along their track and are therefore well suited to demonstrate funnelling. The evolution of the electrostatic potential over time and carrier density around the impact region is shown in Figure 8.3.

At the initial time step, $t = 0$, before the particle strikes, the electrostatic potential and charge distribution are undisturbed. At the moment of impact, $t = 1$ ns, no immediate change in potential is observed. The white line, which is visible inside the deep nwell, marks the extension of the depletion region, where the excess electron–hole pairs are generated along the ion path. The situation evolves within nanoseconds. By $t = 2$ ns, the electrostatic potential in the region has visibly deformed, matching the localised increase in charge density. The rise in electrostatic potential extends deep into the bulk. By $t = 10$ ns, the electrostatic potential inside the deep nwell is largely restored, and the depletion regions are returning to their normal state. This distortion is consistent with the collapse of the depletion region and the onset of drift-dominated carrier transport, which are characteristics of the funnelling effect described in the literature [117].



(a) Initial state at $t = 0$ ns

(b) α-particle strike at $t = 1$ ns

(c) Transient response at $t = 2$ ns

(d) Relaxation at $t = 10$ ns

Figure 8.3: Sentaurus TCAD simulation of a 5 MeV α-particle traversing the ISO cell of the HitPix sensor. The sequence shows the temporal evolution of the electrostatic potential and charge density: (a) initial state before the strike, (b) ion impact and track formation, (c) transient potential deformation within nanoseconds, and (d) recovery of the depletion region after several nanoseconds. The results visualise the funnelling effect, characterised by a temporary collapse of the space charge region, as described in [117].

8.2.3 Susceptibility to single-event latch-up (SEL)

In modern ICs, SEL is less dependent on the absolute charge collected at a single node and more closely tied to the total current generated within a well or substrate region. The most sensitive regions are typically reverse-biased $p - n$ junctions, for example, nwell/p–substrate junctions that are poorly contacted and located near the source diffusion regions. If the current transient is sufficient to drop the local potential beneath the source contact, it may forward-bias a nearby

parasitic diode, initiating a latch-up event. This mechanism closely mirrors the well-known conditions for electrical latch-up [118].

The sensitivity of a device to single-event latch-up is strongly influenced by its technology process and functionality. These factors determine the physical layout, doping profiles, and parasitic structures within the device, all of which affect how charge is collected and whether latch-up mechanisms can be triggered.

Technology process. The specific semiconductor technology used in manufacturing has a statistically significant impact on SEL reliability. Comparative studies have shown that CMOS technologies are generally the most susceptible to SEL, particularly in space environments where high-energy particles are prevalent [118]. BiCMOS processes tend to offer improved SEL resistance relative to CMOS, but still fall short of the robustness observed in bipolar devices. Among the studied categories, bipolar processes consistently exhibit the highest resistance to SEL initiation. For example, at LET values exceeding 75 MeV·cm^2/mg, the average survival probability against SEL was reported as 58% for CMOS, 79% for BiCMOS, and 91% for bipolar processes [118]. This trend reflects the role of physical structure and isolation schemes in determining SEL vulnerability. It is also noted that misidentification of other failure mechanisms, such as single-event burnout or dielectric rupture, can occasionally introduce outliers into survival analyses, especially when reported without detailed test data.

Device functionality. The intended function of the device also influences its SEL sensitivity. Microcontrollers, analogue-to-digital converters (ADCs), and processors show relatively low reliability under SEL conditions. This is partly attributed to high circuit density. In contrast, operational amplifiers show the highest SEL resistance, likely due to their frequent implementation in bipolar technologies. Regulators and static RAM (SRAM) modules tend to exhibit average reliability. However, because functional categories may encompass a range of technologies, the correlation between function and SEL risk is more complex than with the process node alone [118].

Technology scaling. The evolution of technology nodes has a non-linear influence on SEL reliability. In CMOS processes, older technologies with 2-10 µm-sized nodes exhibit high SEL resistance. As feature sizes shrink, reliability initially degrades, reaching a minimum around the 0.8 µm node. However, further scaling into the deep submicron regime (e.g. 180 nm or 65 nm) introduces certain forms of "natural hardening", which can improve SEL immunity. Nevertheless, this trend does not continue indefinitely. At the 30 nm node and below, SEL reliability drops again [118].

Understanding how process technology, circuit topology, and scaling trends interact is essential for assessing radiation tolerance in HV-CMOS designs. In the case of the HitPix family, such insights are particularly relevant given their implementation in bulk CMOS processes and exposure to a high-radiation environment during operation.

8.3 Investigation of SEE influence on HV-CMOS

The previously discussed SEE mechanisms raised the question of whether efficiency loss and transient failures observed in HitPix sensors are a general issue affecting HV-CMOS technologies, or whether they are specific to a particular circuit architecture or layout. To explore this, a broader set of HV-CMOS sensors was tested in beam campaigns at the HIT facility conducted from January to June 2024. These included devices with different readout architectures, amplifier types, and process origins.

As described earlier in Chapter 6, the HIT facility is capable of delivering a broad spectrum of ion species, including protons, carbon, helium, and oxygen. The facility provides 255 discrete energy settings, 15 intensity steps, and 6 spot sizes for each ion type. Such flexibility is ideal for medical applications, where precise tumour treatment planning requires fine control of beam energy, intensity, and spatial distribution. However, for detector testing, this wide parameter range can complicate direct comparisons between runs.

In practice, the operator selects both the beam energy and intensity. The intensity defines the dose rate, which during therapy also gives us the treatment time, while the energy determines the penetration depth. For each energy–intensity combination, a range of beam foci is available. In this context, the "focus" refers to the full width at half maximum (FWHM) of the beam spot at the isocentre, in other words, the effective spot diameter. For example, a nominally fixed focus setting F1 will result in a larger spot size at a low beam energy Setting 1 compared with a smaller spot size at a high energy Setting 255. In medical treatment, this relationship is exploited to shape the treatment field precisely in both space and time. In our case, the exact beam focus is less critical, since any available focus size at the chosen settings fully covers the active area of the chip, which is always positioned at the isocentre. Instead, the intensity per active area is a more relevant parameter for our measurements.

For the tests, beam intensity settings 5 and 10, corresponding to particle fluxes of 4.0×10^8 particles/s and 3.2×10^9 particles/s for protons, and 1.0×10^7 particles/s and 5.0×10^8 particles/s for carbon ions, respectively. Focus settings 1, 4, and 6 were employed in the initial phase, later restricted to setting 4 for consistency. Beam energies were initially set to configurations 1 and 255, corresponding to 48.12 MeV/u and 221.06 MeV/u for protons, and 88.83 MeV/u and 430.10 MeV/u for carbon ions. To improve beam stability, the lower-energy setting was later changed to configuration 10, giving 58.72 MeV/u for protons and 108.53 MeV/u for carbon ions. To investigate the behaviour of the chip under varying conditions, multiple run sets were compared, considering both beam settings and beam type. In general, direct comparison of the results proved challenging, even with the high available statistics, due to the large number of varying beam parameters and their non-linear interdependencies. After normalising selected beam parameters, the analysis was reduced to several key observations relevant to the chip's performance.

Initial investigations focused on signal loss in HitPix chips under proton and carbon beams at HIT facility. Given the recurring voltage drop phenomena observed in these chips, it was decided to test a sensor with a fundamentally different architecture, such as the MuPix chip. Unlike HitPix2, which uses an NMOS-based amplifier and includes both analogue and digital

circuitry inside the pixel, MuPix employs a PMOS amplifier in pixel, with the comparator placed in the periphery. This separation reduces in-pixel complexity and alters current paths and sensitive volumes. Testing revealed no voltage drops or signal degradation in MuPix under the same irradiation conditions that affected HitPix2. This result suggested that the observed effect may be architecture-dependent and motivated a more systematic study across different HV-CMOS designs.

Table 8.1: Overview of tested HV-CMOS sensor samples

Chip name	**Pixel structure**	**Foundry**	**Pixel size**	**Substrate resistivity**	**Substrate thickness**
HitPix2 STD	NMOS amplifier, CMOS comparator	TSI	$200 \times 200\ \mu m^2$	370 Ω·cm	700 µm
HitPix3 ISO	PMOS amplifier, CMOS comparator	TSI	$200 \times 200\ \mu m^2$	370 Ω·cm	700 µm
MuPix8	PMOS amplifier, peripheral CMOS comparator	AMS	$80 \times 81\ \mu m^2$	200 Ω·cm	700 µm
AstroPix_v2	NMOS amplifier, CMOS comparator	TSI	$250 \times 250\ \mu m^2$	370 Ω·cm 6–8 k Ω·cm	700 µm
AstroPix_v3	NMOS amplifier, CMOS comparator	TSI	$300 \times 300\ \mu m^2$ (node) $500 \times 500\ \mu m^2$ (pitch)	370 Ω·cm	700 µm
ATLASpix1_Simple	PMOS amplifier, NMOS comparator	AMS	$130 \times 40\ \mu m^2$	80 Ω·cm 200 Ω·cm	70 µm 200 µm 700 µm
ATLASpix1_ISO	PMOS amplifier, CMOS comparator	AMS			
ATLASpix1_M2	PMOS amplifier, NMOS comparator	AMS	$60 \times 50\ \mu m^2$		

To investigate the potential for single-event latch-up (SEL) and related transient effects, additional HV-CMOS sensors were evaluated using proton and heavy-ion beams. These devices varied in pixel design, substrate resistivity, and amplifier topology (PMOS vs NMOS), as summarised in Table 8.1. Over 100 measurement runs were performed across this diverse set of chips. Each test run was recorded and analysed to extract average signal amplitude, drop frequency, and temporal characteristics such as width and depth of the voltage drop. A full data analysis chain was implemented and applied consistently across all sensor types. The full results of this test campaign are discussed in the following sections, where the full

analysis chain is shown, and also the correlation between sensor architecture and SEE-induced efficiency loss is analysed in detail.

8.3.1 Analysis procedure of the ion beam tests results

To illustrate the data analysis chain used in the investigation of the efficiency loss for HV-CMOS sensors, one representative test run is described in detail. Run 20 corresponds to a measurement performed on AstroPix_v3 using a substrate with 370 Ω·cm resistivity. The sensor was irradiated with a proton beam at HIT under the following conditions: Setting 255, beam intensity level 10, focus setting 4, five spills, and a bias voltage of −50 V.

Step 1: Data acquisition. In this test run, 446 individual waveforms were recorded. Figure 8.4a displays a subset of the raw data, illustrating the characteristic noise and variations in baseline amplitude typical of analogue readout systems under beam conditions.

Step 2: Data smoothing. To reduce noise and improve peak identification, a moving average filter was applied to each waveform. The resulting smoothed signals are shown in Figure 8.4b. During this step, signal and drop markers were introduced into the waveforms for visual verification and to support the subsequent classification.

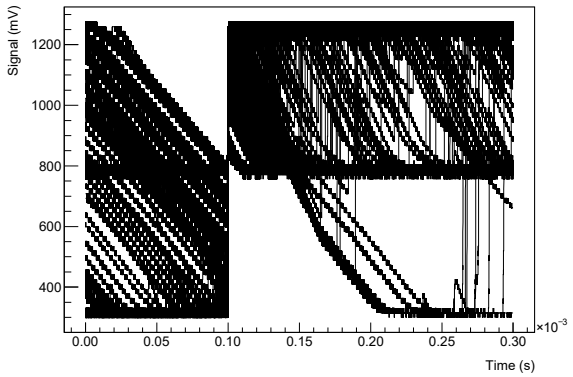
Step 3: Baseline and threshold definition. A set of representative smoothed waveforms was used to define amplitude thresholds. For this run, the signal baseline was determined to be 805 mV. Signals were defined by amplitudes exceeding the 830 mV threshold, whereas drops were defined as voltage values falling below the 760 mV threshold. These thresholds were kept constant throughout the analysis of this run.

Step 4: Peak identification. To identify local maxima and minima more precisely, numerical differentiation was performed using the discrete difference between adjacent data points. Figure 8.4c shows the differentiated and subsequently smoothed waveforms. Combined with the original data, this enabled accurate localisation of signal peaks. This requirement was introduced to suppress artefacts from baseline fluctuations and saturated traces, which otherwise produced broad extrema. Events with a sufficiently rapid decline were therefore classified as valid events, whereas flat or slowly varying features were rejected as non-physical or saturated responses.

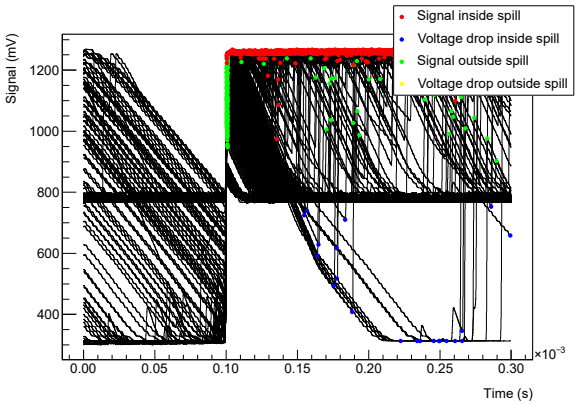
Following waveform smoothing and differentiation, the next steps of the analysis focus on extracting quantitative information from each waveform and identifying characteristic patterns associated with signal and voltage drops.

From the processed data, several key parameters were computed:

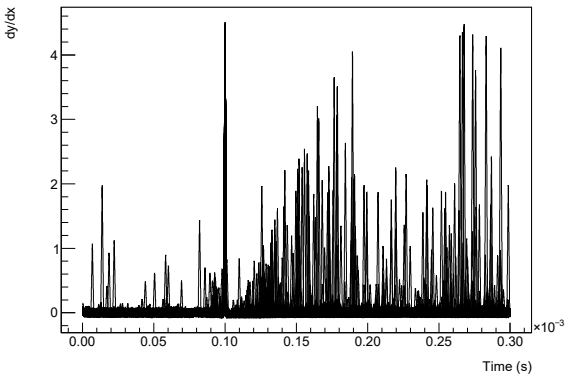
- maximum and minimum amplitude values of each waveform;
- duration of signal peaks and voltage drop length;
- number of identified signals and drops per waveform;
- ratio of signal to drop occurrences;



(a) Raw waveforms showing signal and noise distribution before processing.



(b) Smoothed waveforms after application of a moving average filter. Markers indicating potential drops and signals are also shown.



(c) Differentiated waveforms, smoothed post-derivation to identify peaks.

Figure 8.4: Analysis steps for Run 20 of the beam test on the AstroPix_v3 chip: (a) raw waveform data, (b) smoothed signal using a moving average filter, and (c) derivative-based peak detection after smoothing.

- calculated baseline for comparison with manually defined thresholds.

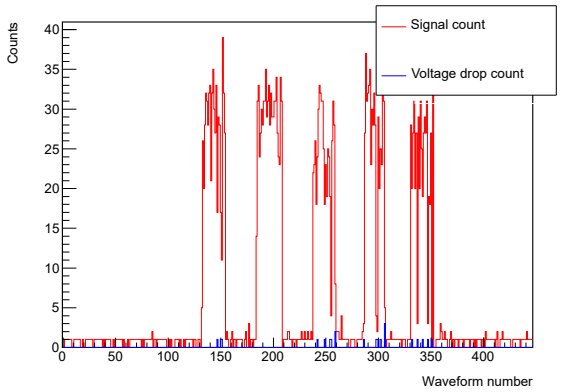
Each waveform was classified using the defined thresholds and differentiated slopes, allowing identification of valid signal peaks and drop events. Markers were then superimposed on the smoothed waveforms to cross-verify the detection logic.

Figure 8.5a presents a histogram of the number of detected signals (in red) and drops (in blue) plotted against the waveform index. Together, they provide a visible indication of the beam spills. Notably, signal counts far exceed drop counts throughout the run, indicating that drops represent a rare but significant anomaly.

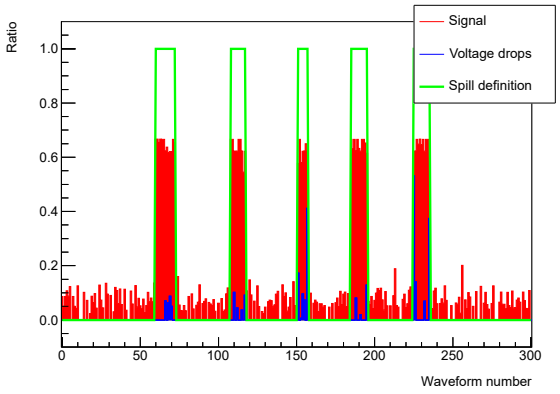
The ratios of signals and drops per waveform are shown in Figure 8.5b. This plot further emphasises the spill structure and allows for the identification of spill boundaries. These

boundaries were later used to distinguish between effects occurring inside the beam exposure window and those occurring in between spills.

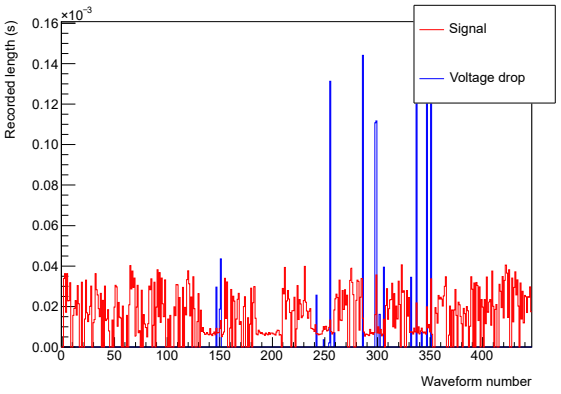
Figure 8.5c displays the average duration of signal and drop across all waveforms. Two key observations emerge: signal durations are relatively uniform across the run, and voltage drop events, when present, are both longer in duration and less frequent than signals.



(a) Histogram of detected signals (red) and voltage drops (blue) per waveform for Run 20. A clear spill structure is visible, with drops that are non-uniformly distributed.



(b) Ratio of signals and drops per waveform. Used to define the beam spill intervals for further classification.



(c) Average duration of signals and drops over all waveforms. Drops are significantly longer than signals and less frequent.

Figure 8.5: Waveform-level signal statistics from Run 20. These plots support quantitative classification of voltage drops and beam-induced signals.

An example of the summary table compiled from the analysis is shown in Figure 8.6. It provides a compact overview of drop statistics across Runs 20–27 for AstroPix_v3 chips under various conditions, allowing direct comparison between configurations and beam settings.

Beam (Type of the beam)	Protons				Carbon			
Run (Run number)	22	23	21	20	26	27	24	25
Energy, beam settings	10	10	255	255	10	10	255	255
Energy, [MeV/u]	58,72	58,72	221,06	221,06	108,53	108,53	430,1	430,1
Intensity	5	10	5	10	5	10	5	10
Intensity, particle/sec (Beam settings)	4E+08	3,2E+09	4E+08	3,2E+09	1E+07	8E+07	1E+07	8E+07
Fucus	4	4	4	4	4	4	4	4
Focus, mm	28,5	28,5	12,6	12,6	12,5	12,5	9,8	9,8
Signals in spills (Ratio and count of signals and drops)	2562	1077	3345	2866	3337	1783	2509	1003
Drops spills	5	14	25	21	43	43	22	21
Drops per signals	0,20 %	1,30 %	0,75 %	0,73 %	1,27 %	2,41 %	0,88 %	2,09 %
Time to drop, µs (Time for one full drop)		104	100	120	100	90	120	80
1 point is	5	ns	It is	0,005	µs			
Av. signals length, points	1554,78	2343,62	1413,67	1451,98	2125,11	1872,68	1788,52	1832,75
Av. drop length, points	6078,00	6995,29	7822,76	11552,00	5208,86	5298,98	8403,23	7575,29
Tot. signals length, points (Average and total lengths of the signals and drops)	3983345,00	2524084,00	4728713,00	4161378,00	7176513,00	3338990,00	4487396,00	1838247,00
Tot. drops length, points	30390,00	97934,00	195569,00	242592,00	223981,00	227856,00	184871,00	159081,00
Av. signals length, µs	7,77	11,72	7,07	7,26	10,63	9,36	8,94	9,16
Av. drop length, µs	30,39	34,98	39,11	57,76	26,04	26,49	42,02	37,88
Tot. signals length, µs	19.916,73	12.620,42	23.643,57	20.806,89	35.882,57	16.694,95	22.436,98	9.191,24
Tot. drops length, µs	151,95	489,67	977,85	1.212,96	1.119,91	1.139,28	924,36	795,41
Comment (Observation)	Good signals with drop length increasing to the last spills	Longer drops then for Intensity 5	Long uniform drops	More drops to the end of the spills	Defined spills, strong signal	Less signal with higher intensity	Defined spills, less drops then for energy 10	Worse signal

Figure 8.6: Example result table for Runs 20–27 on AstroPix_v3. It summarises drop and signal statistics under different proton beam conditions.

8.3.2 Performance of the AstroPix chip family

The AstroPix chip family consists of several successive generations, each building on the design of its predecessor. All versions employ individually depleted pixels with areas ranging from 10 µm^2 to 1 mm^2. In-pixel CMOS circuitry integrates charge collection, signal amplification, and discrimination directly on the sensor, eliminating the need for an external readout ASIC. As a result, both analogue and digital signal processing are performed on-chip, reducing passive material in the active region and simplifying detector assembly. A deep nwell structure electrically isolates the pixel electronics from the bulk substrate. High-voltage biasing applied at each pixel ensures rapid and efficient charge collection, while the comparator output is routed to the periphery for further digital logic processing [104]. Additional details on the chip design and fabrication are presented in Chapter 7.

In the present work, tests were performed on **AstroPix_v2** sensors with substrate resistivities of 370 Ω·cm and 5k Ω·cm, and on **AstroPix_v3** sensors with 370 Ω·cm resistivity. The AstroPix_v2 chip features a pixel size of 250 × 250 µm^2, allowing higher depletion voltages, and implements additional guard rings to reduce crosstalk between neighbouring pixels. AstroPix_v3 employs a 500 × 500 µm^2 pixel pitch with a 300 × 300 µm^2 collection node. For this version, the guard ring layout was redesigned to eliminate conductive pwells, which had previously contributed to charge sharing between pixels.

The layout of the AstroPix_v2 amplifier and comparator architecture is shown in Figure 8.7. For AstroPix_v3, the front-end architecture remained unchanged; only the substrate properties and overall chip layout were modified.

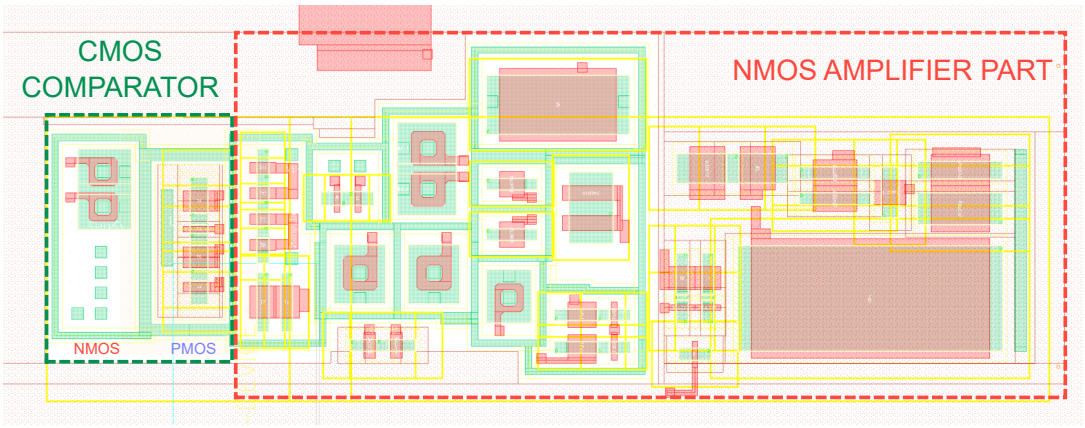
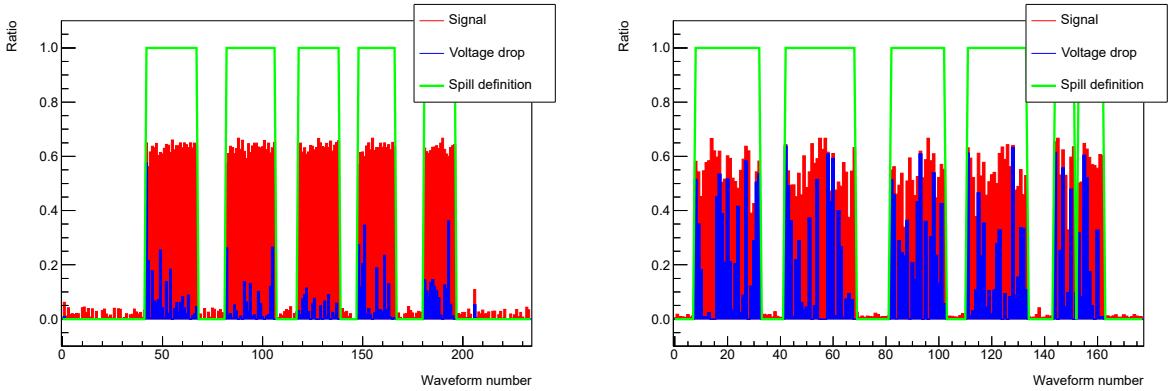


Figure 8.7: Schematic layout of the AstroPix_v2 amplifier and comparator architecture. The design is identical for AstroPix_v3, with changes limited to substrate resistivity and chip dimensions.

In order to test the chips for SEE-induced efficiency loss, a total of 27 runs were conducted using both carbon and proton beams at the HIT facility. The first comparison focused on AstroPix_v2 sensors with different substrate resistivities: 370 Ω·cm versus 5k Ω·cm. The ratio of drops to signals was consistently higher for the 5k Ω·cm substrate (20–60%) compared to the 370 Ω·cm substrate (10–20%). A representative example is shown in Figure 8.8. Run 9 corresponds to the 370 Ω·cm substrate, measured with a carbon beam, energy setting 255 (430.1 MeV/u), intensity level 10 (8×10^7 particles/s), and focus setting 6 (FWHM is 20 mm). Run 19 corresponds to the 5k Ω·cm substrate, recorded under the same beam energy and intensity conditions, but with focus setting 4 (FWHM is 9.8 mm).



(a) Run 9: 370 Ω·cm substrate

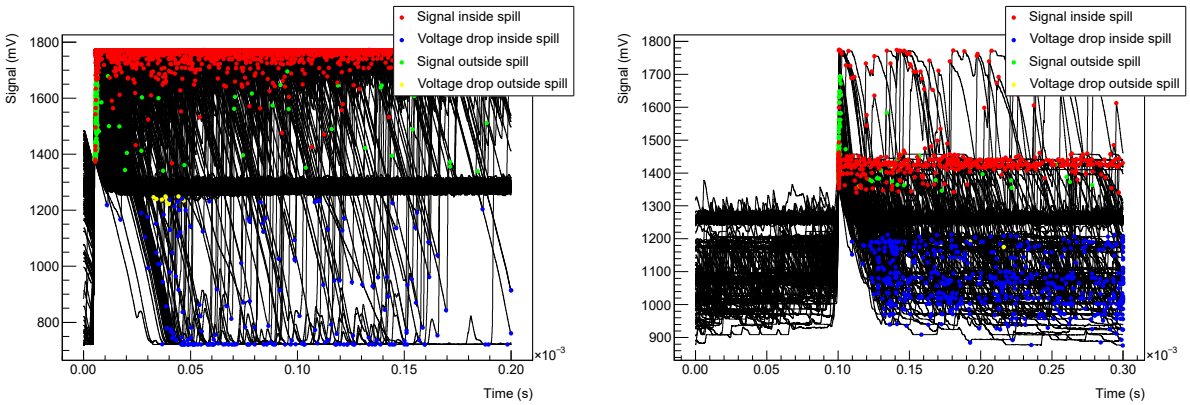
(b) Run 19: 5k Ω·cm substrate

Figure 8.8: Comparison of the ratio of signals to drops for AstroPix_v2 sensors with different substrate resistivities under identical carbon beam conditions.

Another notable observation concerns the limited effectiveness of the high-resistivity substrate in mitigating voltage drops. While the increased depletion volume of this substrate resulted in a higher number of recorded signals, it also led to a greater occurrence of drops. These drops

were typically longer in duration, with slower recovery to the baseline. In many cases, the recorded signals did not reach the maximum amplitude of 1700 mV (baseline at 1280 mV), but instead plateaued around 1450 mV. This effect was observed for the high-resistivity substrate irrespective of the beam type. For this chip, low-energy signals appeared consistently across measurements; however, full-amplitude signals were less frequent for low-energy carbon and proton beams, with signals at ~1400 mV being far more common than those reaching 1700 mV.

Figure 8.9 illustrates this phenomenon for the complete dataset. The left plot corresponds to AstroPix_v2 with a 370 Ω·cm substrate (Run 7), and the right plot to AstroPix_v2 with a 5k Ω·cm substrate (Run 15). Both measurements were performed under identical beam settings: proton beam, energy setting 255 (corresponding to 221.06 MeV/u), intensity level 10 (3.2×10^9 particles/s).

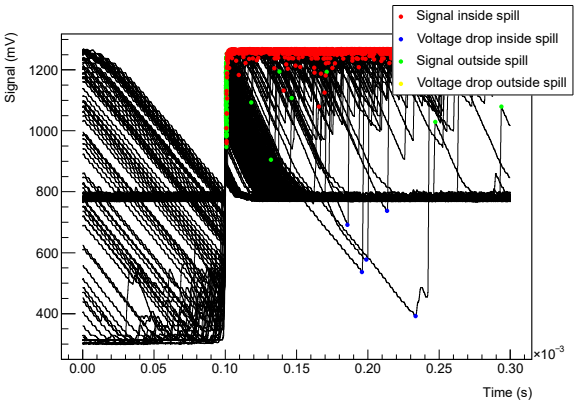


(a) Run 7: 370 Ω·cm substrate

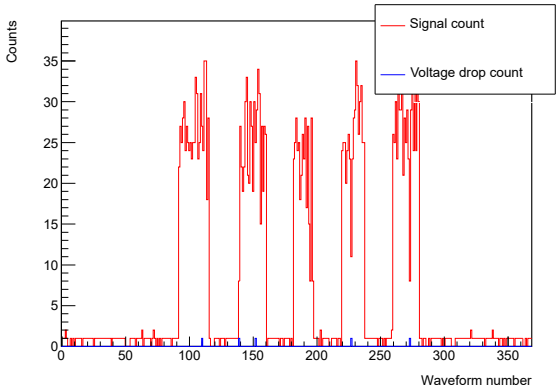
(b) Run 15: 5k Ω·cm substrate

Figure 8.9: Comparison of recorded waveforms for AstroPix_v2 sensors with different substrate resistivities under identical proton beam conditions, used to assess the effectiveness of a high-resistivity substrate in mitigating voltage drops.

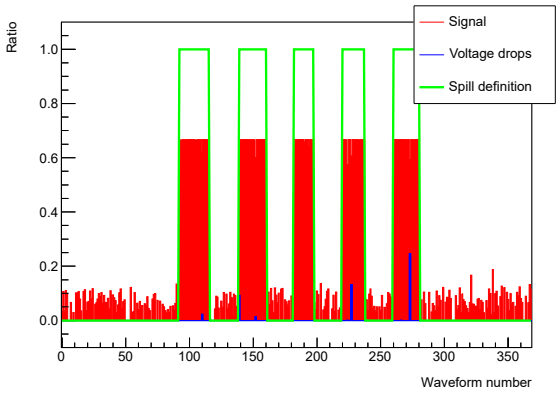
Another parameter investigated, which showed a distinctly positive effect on drop recovery, was the pixel pitch. For AstroPix_v3, the main modification concerned the sensor geometry. While the size of the collection node remained almost unchanged compared to AstroPix_v2, the distance between neighbouring pixels was increased. In AstroPix_v2, the pixel pitch matched the 250×250 µm^2 collection node, whereas in AstroPix_v3 the collection node measured 300×300 µm^2 within a pixel pitch of 500×500 µm^2. This geometric adjustment led to a considerable improvement in performance. AstroPix_v3 achieved the best results in terms of drop behaviour, with the drop-to-signal ratio reduced to only 2–3%. As an example, Run 22 is presented, using a proton beam with energy setting 10 (58.72 MeV/u) and intensity 5, corresponding to 4×10^8 particles/s. The complete set of recorded waveforms for this run, showing distinct signal shapes and individual long drops, is provided in Figure 8.10a. Figure 8.10b illustrates the number of identified signals and drops, as well as visible beam spills, demonstrating a stable signal count throughout the run. The ratio of drops to signals, presented in Figure 8.10c, shows that the length of the drops increases over time. Finally, Figure 8.10d summarises the calculated lengths of each recorded event.



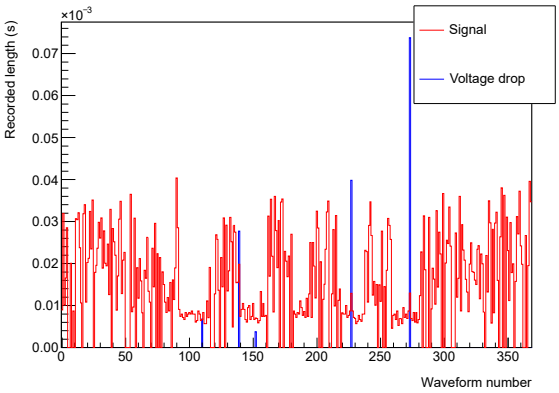
(a) Recorded and smoothed waveforms for Run 22



(b) Number of identified signals and drops



(c) Ratio of drops to signals over time



(d) Calculated average lengths of signals and drops

Figure 8.10: Performance of AstroPix_v3 with increased pixel pitch, illustrated for Run 22 (proton beam, 58.72 MeV/u, 4×10^8 particles/s). The increased spacing between pixels reduces the drop-to-signal ratio and improves drop recovery.

In summary, the beam test campaign demonstrated that substrate resistivity, pixel geometry, and beam conditions significantly influence the signal stability and drop behaviour of the AstroPix chip family. While high-resistivity substrates increase the number of recorded events, they also lead to more frequent and prolonged voltage drops with reduced peak amplitudes. The most substantial improvement was achieved with the increased pixel pitch of AstroPix_v3, which reduced the drop-to-signal ratio to 2–3% and enhanced recovery characteristics.

8.3.3 Performance of the ATLASpix1 chip family

ATLASpix1 has been designed as a demonstrator chip for the outermost pixel layer of the ATLAS ITk. It is implemented in 180 nm HV-CMOS technology and was fabricated on substrates with resistivities between 20 Ω·cm and 1k Ω·cm, with some samples thinned to 70–200 µm. The layout of one chip consists of three fully independent matrices, each approximately 3.5 mm

wide and 18.4–19.5 mm long, implementing a fully monolithic readout capable of high hit rates, zero suppression, and providing additional hit timing information [8].

The overall chip layout is shown in Figure 8.11. All three matrices have bond pads on the top and bottom edges. The **ATLASpix1_Simple** and **ATLASpix1_IsoSimple** matrices share most characteristics, while **ATLASpix1_M2** differs considerably in geometry and readout scheme. ATLASpix1 was developed by the KIT ASIC and Detector Laboratory in collaboration with the Universities of Heidelberg and Geneva, under the leadership of Prof. Perić [8]. More details on the design, implementation, and characterisation of the ATLASpix1 chip can be found in the PhD thesis of Dr Ehrler [8].

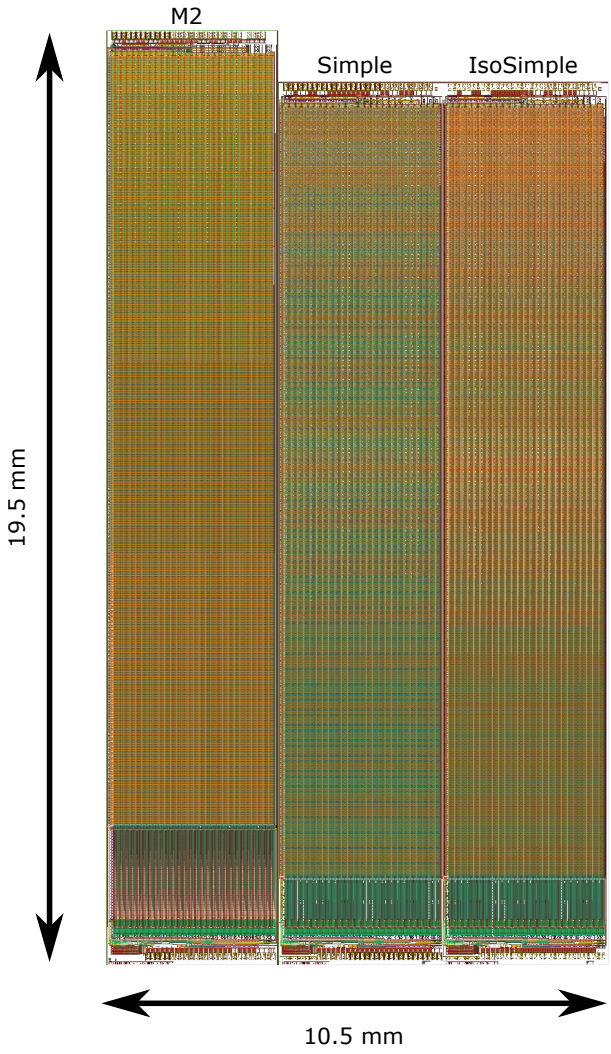


Figure 8.11: ATLASpix1 chip layout. Three independent matrices with bond pads on the top and bottom edges. ATLASpix1_Simple and ATLASpix1_IsoSimple share most features; ATLASpix1_M2 has a different geometry and readout concept [8].

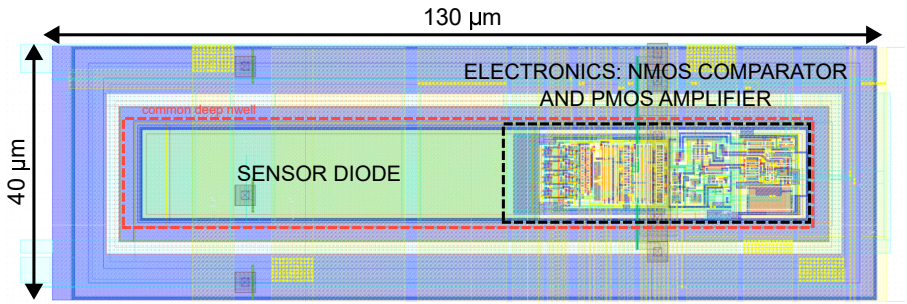
All three matrices use the same analogue front-end, consisting of a charge-sensitive amplifier and a comparator inside each pixel. The sensor diode is formed by the $p-$substrate and deep nwell, which also houses the front-end electronics. Digitised signals are transmitted to the

digital periphery containing End-of-Column (EoC) and Readout (RO) blocks, where the pixel address, primary time stamp, and a secondary time stamp for Time-over-Threshold (ToT) are generated [8].

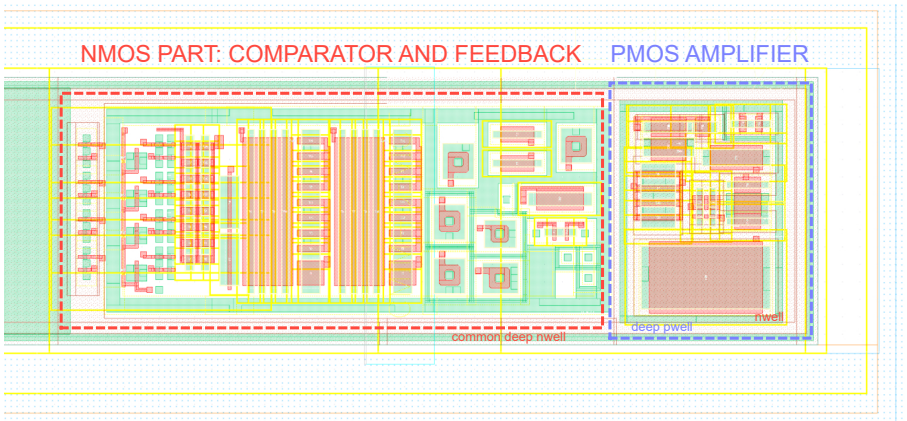
The pixel matrix of ATLASpix1_Simple and ATLASpix1_IsoSimple consists of 25 × 400 pixels, each 130 × 40 µm^2 in size, optimised for high spatial resolution in one direction. They share identical pixel geometry and readout schemes. The difference lies in the comparator implementation [8]:

- ATLASpix1_Simple uses an NMOS comparator. NMOS transistors are isolated from the substrate by the deep nwell. This design is straightforward but has higher power consumption and lower radiation tolerance.
- ATLASpix1_IsoSimple includes a deep pwell, enabling the use of both NMOS and PMOS transistors in a CMOS comparator. This reduces power consumption and improves radiation tolerance.

Figures 8.12 and 8.13 show the pixel and electronics layouts for both matrices. The pixel outline is defined by a guard ring. The deep nwell forms the sensor diode together with the $p-$substrate. The amplifier and comparator are located on the right side, with the line driver transmitting digitised signals to the periphery.

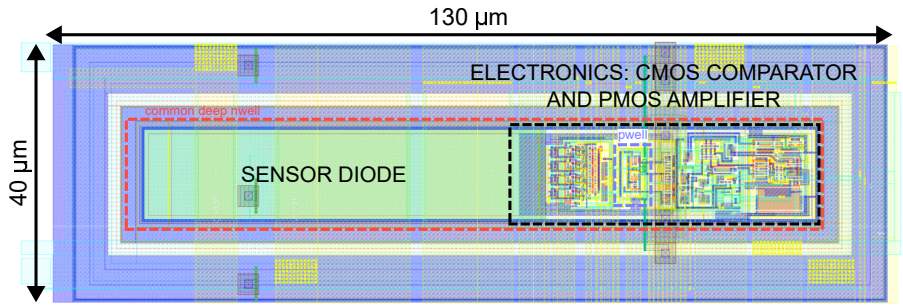


(a) Pixel layout

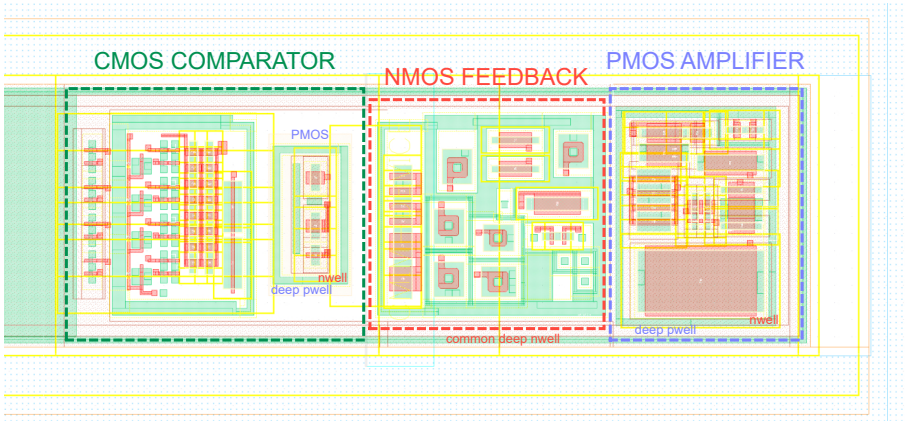


(b) Electronics layout

Figure 8.12: ATLASpix1_Simple pixel and electronics layout.



(a) Pixel layout



(b) Electronics layout

Figure 8.13: ATLASpix1_IsoSimple pixel and electronics layout. A deep pwell enables CMOS comparator implementation.

The ATLASpix1_M2 matrix differs substantially in design. It uses a nearly square 60×50 µm^2 pixel, reducing pixel area by 43% compared to other two desings. The matrix has 56 columns $\times$ 320 rows (17 920 pixels in total), which is about 80% more than the other designs usings the same active area. The front-end remains similar to ATLASpix1_Simple but includes two line drivers for transmitting the digitised voltage signal [8]. The pixel layout for ATLASpix1_M2 is shown in Figure 8.14.

Table 8.2 summarises the ATLASpix1 samples evaluated in this work. Each sample contains all three design variants (ATLASpix1_Simple, ATLASpix1_IsoSimple, and ATLASpix1_M2), although in several cases, not all variants were operational or available for testing. The tested samples had varying substrate resistivities and sensor thicknesses. The ATLASpix campaign was particularly valuable due to the high variability in design and fabrication parameters, enabling direct comparison across configurations. In total, 40 beam runs were performed. As in previous campaigns, both proton and carbon beams were used with varying energies and intensities, selected to ensure that the results from all chips are directly comparable.

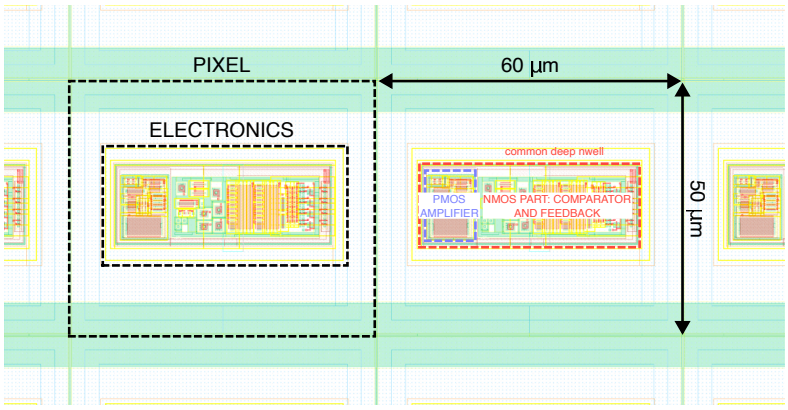


Figure 8.14: ATLASpix1_M2 pixel layout. The smaller, nearly square pixel, where the front-end remains similar to ATLASpix1_Simple.

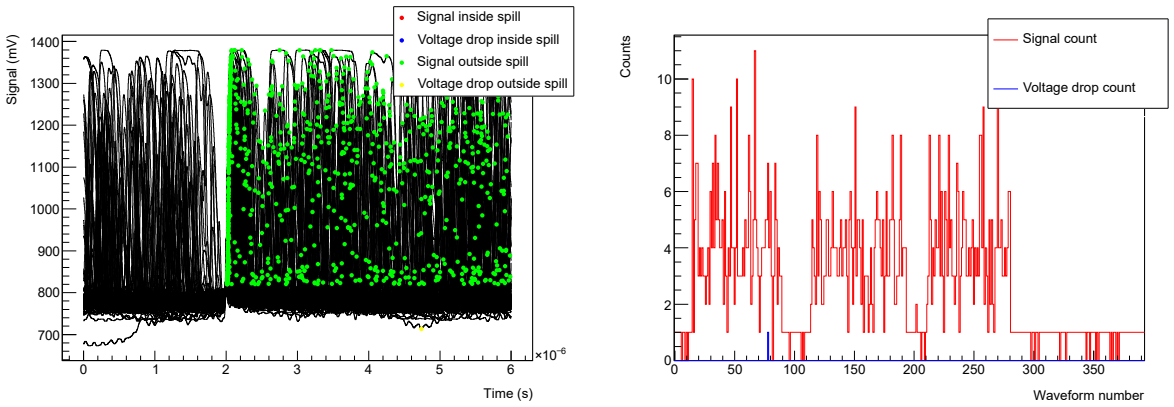
Table 8.2: Overview of ATLASpix1 samples tested

Sample	Resistivity	Thickness	ATLASpix1_ Simple	ATLASpix1_ IsoSimple	ATLASpix1_ M2
A200A	200 Ω·cm	700 µm	tested	tested	not tested
A200	200 Ω·cm	undefined	tested	tested	tested
BS2	200 Ω·cm	200 µm	tested	tested	not tested
AT	80 Ω·cm	700 µm	tested	tested	not tested
An3	80 Ω·cm	80 µm	not tested	not tested	tested

Runs were grouped to compare different design variants, substrate resistivities, and sensor thicknesses. The first set of tests focused on the main architectural difference between the designs, namely the additional pwell required for the CMOS comparator implementation. For this purpose, ATLASpix1_Simple and ATLASpix1_IsoSimple were compared. The main finding is that, for proton beams, both ATLASpix1_Simple and ATLASpix1_IsoSimple exhibited no voltage drops. From this result, it can be concluded that, from the perspective of voltage drop behaviour, switching from an NMOS comparator to a CMOS comparator has no significant influence due to the effective isolation provided by the additional p-well. However, what fundamentally distinguishes these chips from, for example, AstroPix is the use of a PMOS amplifier instead of an NMOS amplifier.

An example is shown in Figure 8.15 for sample A200A. Run 30 corresponds to the ATLASpix1_Simple design with a substrate resistivity of 200 Ω·cm and a thickness of 700 µm, tested with a proton beam at energy setting 255, focus F4, and intensity setting 10. The left panel presents all recorded waveforms, while the right panel shows the number of signals and drops per waveform. One event initially classified as a voltage drop was later identified as

noise occurring outside the beam spill. Overall, the signal count is lower and noisier compared to AstroPix devices, and spill boundaries are more difficult to clearly distinguish.



(a) Recorded and smoothed waveforms

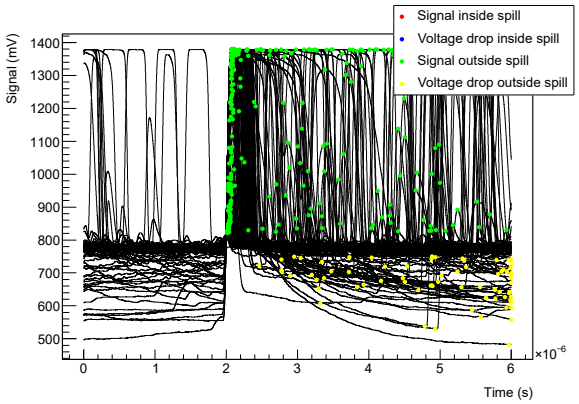
(b) Number of signals and drops per waveform

Figure 8.15: Run 30, sample A200A, ATLASpix1_Simple, proton beam at energy setting 255, focus F4, and intensity setting 10

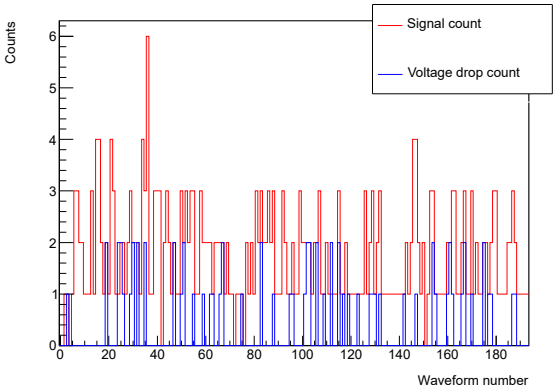
For carbon beams, voltage drops are clearly observed. Figure 8.16 shows run 34 for the same chip, ATLASpix1_Simple from sample A200A, tested with a carbon beam at energy setting 255, focus F4, and intensity 10. The complete set of recorded waveforms is presented in Figure 8.16a. Drop identification, both within and outside the spill, is complicated by poorly defined spill boundaries and the difficulty of distinguishing drops and signals from noise. After applying additional filtering, Figures 8.16b and 8.16c reveal a low signal count and the presence of long drops from which the signal does not recover. This behaviour explains the lower-than-expected counts.

For the ATLASpix1_Simple design, further analysis was carried out to evaluate the influence of different design parameters. First, a comparison with the ATLASpix1_M2 matrix showed that the latter, with its smaller pixel pitch, exhibited significantly higher noise, consistent with observations for AstroPix_v3 where a larger pixel pitch improved the quality of recorded data. Next, comparing sample A200A with a 200 Ω·cm substrate to sample AT with an 80 Ω·cm substrate showed that both devices exhibited no drops for proton beams and drops for carbon beams; however, for the 80 Ω·cm sample, the signals were slower and did not recover between drops, leading to the dominance of long drops over valid signals. Finally, comparing sample A200A with a thickness of 700 µm to sample BS2 with a thickness of 200 µm revealed no significant performance changes and, importantly, no deterioration.

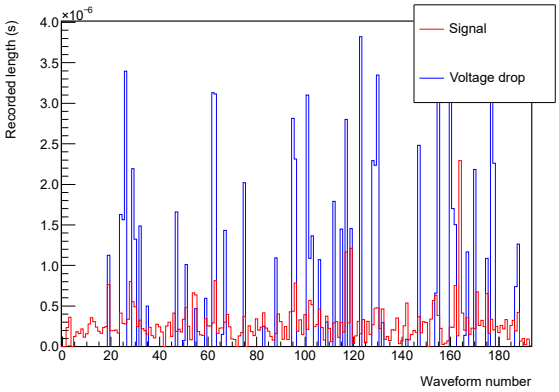
In summary, the ATLASpix1 test campaign demonstrated stable performance across all tested design variants under proton irradiation, while voltage drops appeared predominantly under the carbon beam. The introduction of a CMOS comparator in the IsoSimple design did not negatively affect drop behaviour, confirming the effectiveness of the additional pwell isolation. From this, it can be concluded that from the perspective of voltage drop behaviour, switching from an NMOS comparator to a CMOS comparator has no significant influence due to the effective isolation provided by the additional pwell. The main architectural difference between



(a) Recorded and smoothed waveforms



(b) Number of signals and drops per waveform



(c) Average lengths of signals and drops.

Figure 8.16: Run 34, sample A200A, ATLASpix1_Simple, carbon beam at energy setting 255, focus F4, and intensity setting 10

the previously tested AstroPix, which exhibited voltage drops for both proton and carbon beams, and ATLASpix lies in the use of a PMOS amplifier instead of an NMOS amplifier. Design choices such as the smaller pixel pitch of ATLASpix1_M2 and the lower substrate resistivity, as in sample AT with an 80 Ω·cm substrate, affected noise levels, signal recovery, and drop characteristics such as length, but not their occurrence. These findings were consistent with the conclusions drawn from the AstroPix testing campaign. Variations in sensor thickness had minimal impact on performance.

8.3.4 Performance of MuPix8 chip

MuPix8 was developed for the Mu3e experiment at the Paul Scherrer Institute (PSI) in Switzerland [123]. The Mu3e detector is designed to measure the rare decay $\mu^+ \rightarrow e^+e^-e^+$ with unprecedented precision in a clean, low-energy environment. This decay is of particular interest in the search for physics beyond the Standard Model. The experiment is using high-intensity beams at comparatively low energy; it investigates the theory that new particles

will be produced, rather as virtual particles, due to their mass is expected to be larger than the provided energy [8]. MuPix8 is the seventh prototype in a development line aimed at equipping Mu3e experiment with HV-CMOS sensors. The design was led by Prof. Perić, with contributions from ASIC designers at Karlsruhe Institute of Technology and Heidelberg University, as well as other Mu3e collaboration members.

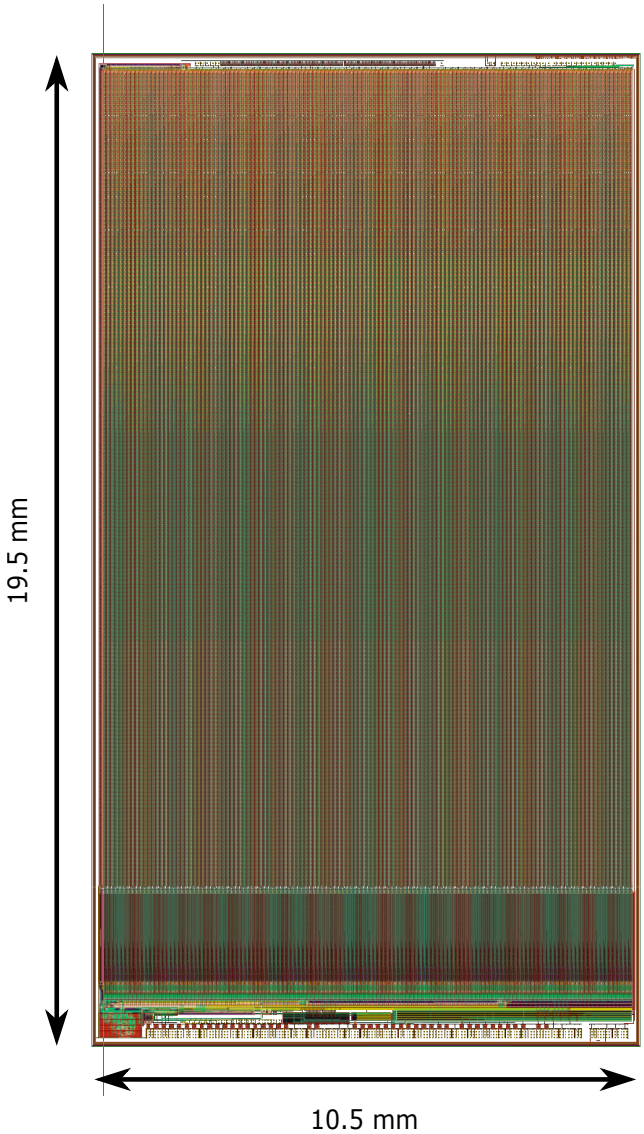


Figure 8.17: Layout of the MuPix8 chip

The Mu3e tracking requirements led to a square pixel geometry, in contrast to the rectangular pixels optimised for ATLAS ITk. The MuPix8 chip, shown in Figure 8.17, measures 10.5 mm in width and 19.5 mm in height, occupying approximately half of the reticle area. The periphery, located along the bottom edge, accounts for 15% of the total chip area, which is comparable to ATLASpix1_M2 but considerably larger than in the other ATLASpix1 matrices. This increased periphery size is necessary because part of the readout circuitry, precisely comparators, is located there. Each pixel measures 80×81 µm^2, corresponding to an area of 6480 µm^2, which

is 24% larger than the ATLASpix1_Simple pixel. The matrix contains 25 600 pixels arranged in 128 columns and 200 rows, split into three independently read out regions. The MuPix8 design is closely related to the ATLASpix1 chip family. Both were implemented and manufactured in a joint submission using the AMS aH18 process with a minimum feature size of 180 nm, sharing many design concepts, particularly in the front-end electronics, bias generation, clock generation, and off-chip data readout. All pixels share an identical layout, shown in Figure 8.18, and employ the same front-end as ATLASpix1. In MuPix8, however, the analogue output of the in-pixel PMOS amplifier is transmitted to the periphery instead of a digitised signal. The comparator, implemented in CMOS, resides in the periphery, marking a key architectural difference from all other tested devices.

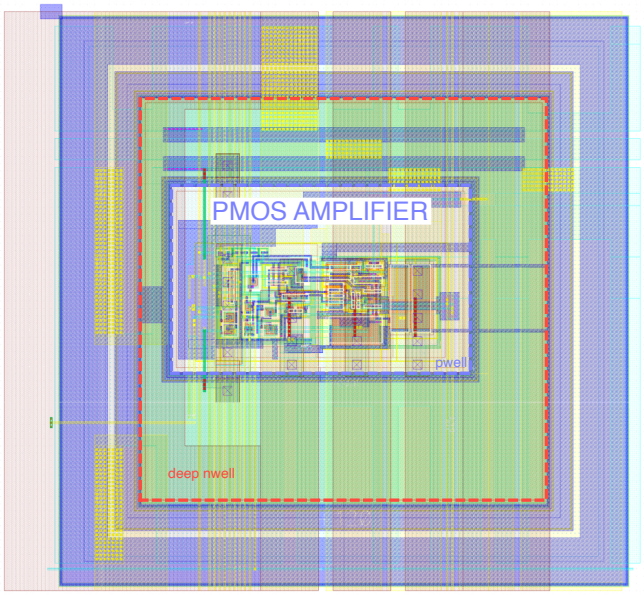
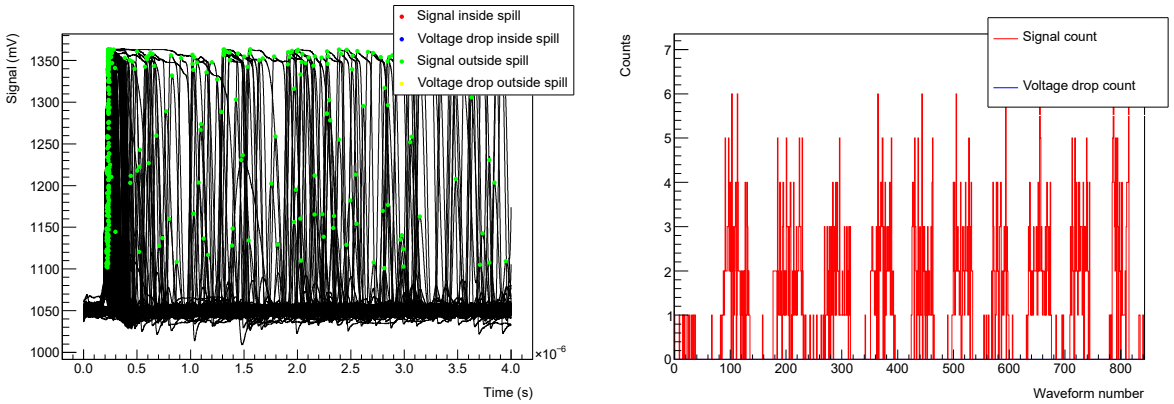


Figure 8.18: Pixel layout of MuPix8

Two MuPix8 variants were tested in this campaign: **MuPix N1** and **MuPix M80**. In total, 35 runs were performed using both proton and carbon beams at intensity setting 10, with focus 6, beam energy settings of 1 and 255, and ten spills per run. This is consistent with previous campaigns, except for the increased number of spills to improve statistics. The range of beam settings was selected to allow direct comparison with other tested chips and to study the influence of beam parameters.

Across all tested beam conditions, for both proton and carbon beams, neither MuPix N1 nor MuPix M80 exhibited voltage drops. Representative results from the MuPix8 test campaign with a carbon beam at energy setting 255 and intensity 10 are shown in Figure 8.19. The left panel shows an example of recorded and smoothed waveforms for MuPix N1 (run 5), and the right panel presents the number of signals and drops per waveform for MuPix M80 (run 102). This stability was independent of particle type, beam energy, focus setting, or intensity. The recorded signals were noticeably noisier than those of the larger-pitch ATLASpix matrices, a trend similar to that observed when comparing ATLASpix1_Simple and ATLASpix1_M2. Nevertheless, the spill structure was well-defined, and signal counting was reliable. These



(a) Example recorded and smoothed waveforms for MuPix N1 (run 5)

(b) Number of signals and drops per waveform for MuPix M80 (run 102)

Figure 8.19: Representative results from the MuPix8 test campaign under carbon beam, energy setting 255, intensity 10. No voltage drops were observed for either chip variant.

results indicate that the MuPix8 analogue transmission architecture, combined with its PMOS front-end and peripheral CMOS comparator, can operate robustly under the tested high-intensity irradiation conditions without demonstrating voltage drops.

In summary, MuPix8 demonstrated stable operation under all tested proton and carbon beam conditions, with no voltage drops observed for either variant. The analogue architecture, featuring an in-pixel PMOS amplifier and peripheral CMOS comparator, maintained reliable performance across varying energies, intensities, and focus settings. While the recorded signals were noisier than those from larger-pitch ATLASpix matrices, the spill structure remained well-defined and signal counting consistent, confirming the robustness of the MuPix8 design for high-intensity irradiation environments.

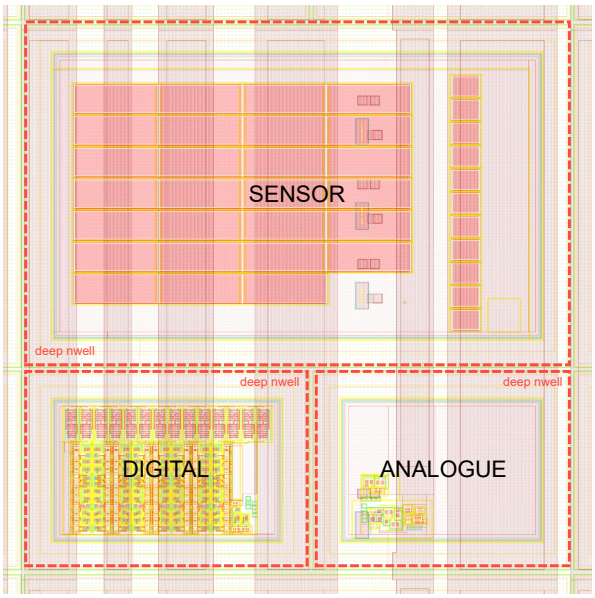
8.3.5 Performance of the HitPix chip family

The final part of this investigation into the possible causes of efficiency loss under heavy-ion irradiation focuses on the HitPix chip family, which originally motivated this study. Although voltage drop phenomena had been observed during earlier test campaigns, they were not previously systematised or consistently analysed. HitPix is a monolithic active HV-CMOS hit-counting pixel sensor, developed for use in high-rate ion and proton beams, with the aim of replacing the current beam-monitoring system at the Heidelberg Ion Beam Therapy Centre (HIT). Further details on the application, chip architecture, characterisation, and simulation results are provided in Chapter 6.

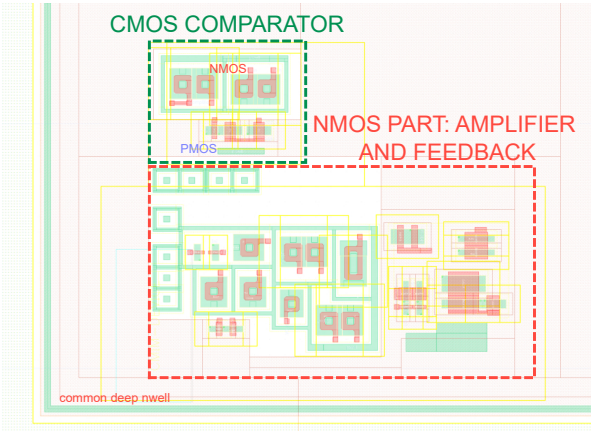
The HitPix chip family includes two distinct design approaches: the Standard (STD) design and the ISO (Isolated Shallow nwells) design. While the electronics are largely similar, they differ in the shape of the deep nwells, additional deep pwells, and minor adjustments to shallow pwells. In the Standard design, each pixel layout is divided into three regions: the sensor electrode, the analogue electronics containing the CSA and comparator, and the digital electronics. These

components are implemented using three separate deep nwells. On the contrary, the ISO design employs a single deep nwell that serves both as the collection electrode and as the host for the analogue and digital wells, making the entire pixel area active and, in principle, 100% sensitive.

Two chip versions were measured: **HitPix2** and **HitPix3**. HitPix2 uses an in-pixel NMOS amplifier, CMOS comparator and the STD pixel topology; HitPix3 implements a PMOS amplifier, CMOS comparator and the ISO pixel topology. Pixel layouts and zoom-ins of the analogue front-end are shown in Figures 8.20 and 8.21. Additional design details, including all chip versions and TCAD studies, are provided in Sections 6.3–6.5.



(a) HitPix2 pixel layout: STD topology

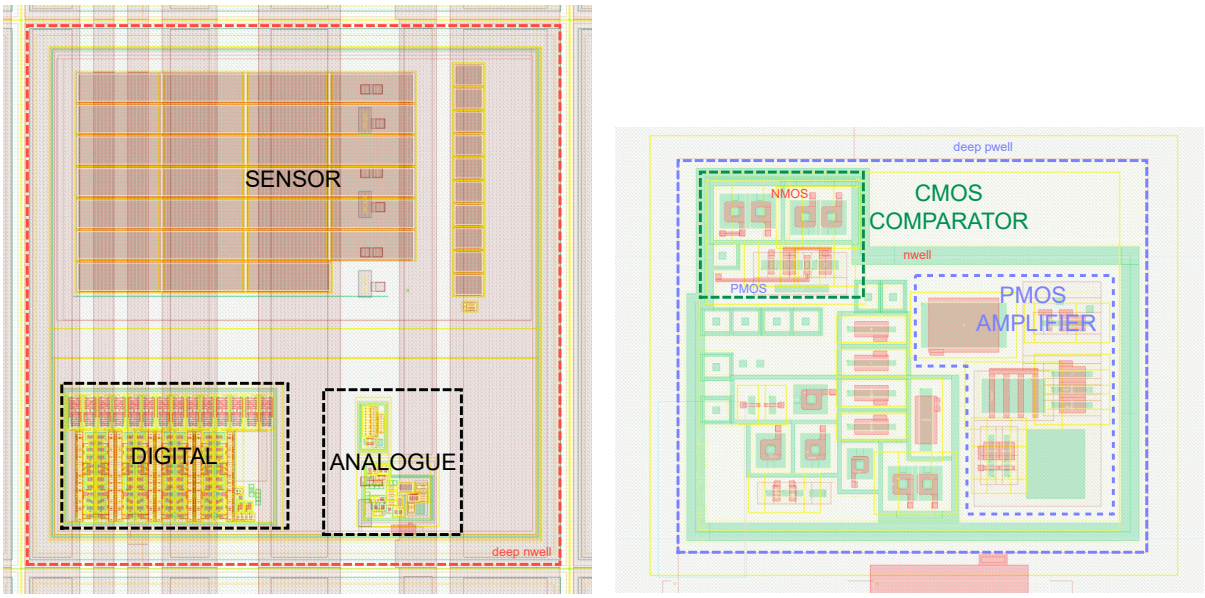


(b) Zoom-in of the analogue front-end

Figure 8.20: HitPix2 overview: standard pixel topology with an in-pixel NMOS amplifier and CMOS comparator.

Beam tests were conducted using both proton and carbon beams. For further comparison with the other chips, the range of beam settings was chosen to allow direct comparison and to study the influence of the beam parameters on drop behaviour. In contrast to earlier campaigns, a broader range of intensity settings was employed (settings 3, 7, and 10), all with focus setting 1, but together with the standard beam energy settings of 1 and 255. Each run consisted of three spills to ensure reliable statistics and to allow a detailed analysis of HitPix performance. Unless stated otherwise, all results for HitPix were obtained at a sensor bias voltage of −80 V, which is considered the nominal operating point and matches the conditions used in the simulations presented in Section 8.2.2.

The ion beam measurements were recorded in AC mode, which does not employ a static, global baseline. Outside the spill, the baseline returns to approximately 0, as expected in AC-coupled readout. However, during the spill, the sustained hit rate and the coupling network led to a downward shift of the baseline. As a result, applying a fixed baseline threshold would introduce a bias in separating true drops from signal events. To address this, the analysis focused on the



(a) HitPix3 pixel layout: ISO topology

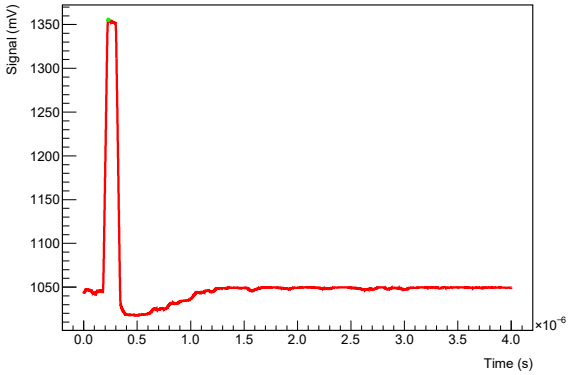
(b) Zoom-in of the analogue front-end

Figure 8.21: HitPix3 overview: ISO pixel topology with an in-pixel PMOS amplifier and CMOS comparator.

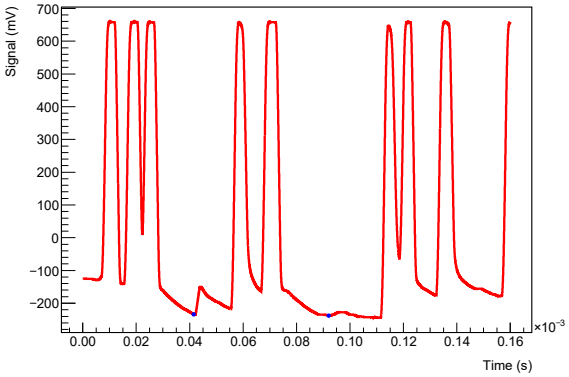
shape and dynamics of individual waveforms, rather than relying on an absolute baseline level. A clean MuPix waveform was used as a reference to define the expected signal shape. HitPix waveforms recorded both inside and outside the spill were then used to develop classification criteria for the statistical evaluation of signal and drop events.

Figure 8.22 illustrates the procedure. Figure 8.22a shows a representative single signal waveform from MuPix (run 1), used as a reference shape. Figure 8.22b shows an example waveform from HitPix3 recorded inside the spill during run 5 (carbon beam, energy setting 255, intensity settings 3). The baseline is visibly pushed down, and some signals are missing or distorted. A voltage drop candidate is marked in the plot. Figure 8.22c shows a waveform from HitPix3 recorded outside the spill during the same run. This serves as a reference to distinguish real signals from noise, since no baseline shift is present in this region.

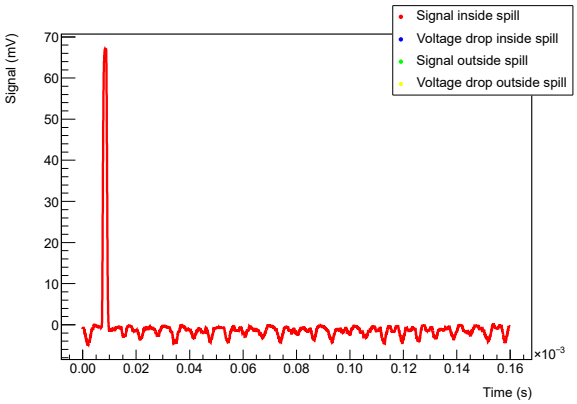
After analysing all HitPix2 and HitPix3 runs, the key beam settings and outcome metrics were compiled in Table 8.3 to enable a direct comparison between the two chip generations. Figure 8.23 illustrates the number of identified signals and drops per waveform for HitPix2 (run 2) and HitPix3 (run 6) under identical carbon beam conditions (energy setting 255, intensity 7). While the signal counts are comparable, the number of drops differs significantly. This result, supported by the full statistics in Table 8.3, reflects the consistent trend observed across all tested beam settings: HitPix3, featuring a PMOS amplifier and ISO topology, systematically shows a reduced drop rate compared to HitPix2, which uses an NMOS amplifier and standard pixel design.



(a) Reference signal waveform (MuPix, run 1).



(b) HitPix3 waveform inside the spill (run 5).



(c) HitPix3 waveform outside the spill (run 5).

Figure 8.22: Additional analysis step for AC-coupled HitPix data. (a) Reference waveform from MuPix used to define the expected signal shape. (b) Example waveform inside the spill for HitPix3, showing baseline sinking and a marked voltage-drop candidate. (c) Example waveform outside the spill for HitPix3, used as a reference to distinguish real signals from noise in the absence of baseline displacement.

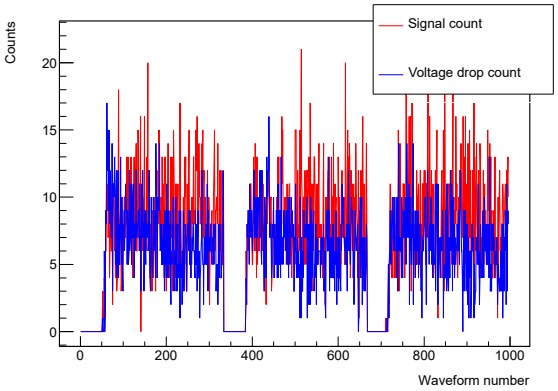
Table 8.3: Summary of beam parameters and results for HitPix2 and HitPix3 evaluation

Parameter	**HitPix2**				**HitPix3**			
Beam	**Carbon**		**Proton**		**Carbon**		**Proton**	
Run number	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**	**Run 6**	**Run 8**	**Run 9**
Energy setting	255	255	255	255	255	255	255	255
Energy, MeV/u	430.10	430.10	221.06	221.06	430.10	430.10	221.06	221.06
Intensity setting	3	7	3	7	3	7	3	7

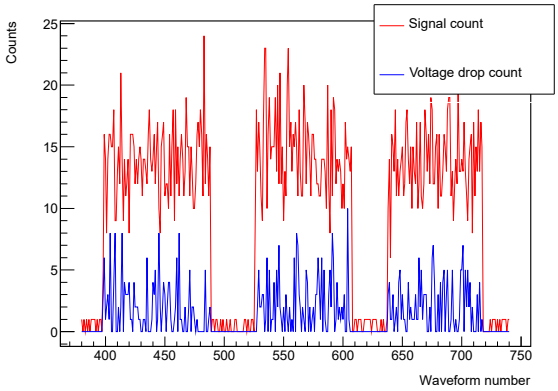
Table 8.3 – continued from previous page

Parameter	**HitPix2**				**HitPix3**			
Beam	**Carbon**		**Proton**		**Carbon**		**Proton**	
Run number	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**	**Run 6**	**Run 8**	**Run 9**
Intensity, particles/s	5×10^6	3×10^7	2×10^8	8×10^8	5×10^6	3×10^7	2×10^8	8×10^8
Focus	F1	F1	F1	F1	F1	F1	F1	F1
FWHM, mm	3.4	3.4	8.1	8.1	3.4	3.4	8.1	8.1
Number of spills	2.5	3	3	3	3	3	3	3
Signals in spills	5014	7137	12375	22263	1692	3532	6579	6055
Drops in spills	1602	5654	631	3255	156	521	67	897
Drops per signals	31.95%	79.22%	5.10%	14.62%	9.22%	14.75%	1.02%	14.81%
Avg. signal length,µs	17.52	10.40	1.76	4.36	4.37	4.22	2.06	1.63
Avg. drop length, µs	88.33	44.80	43.54	28.02	68.45	22.20	6.27	3.44
Total signal length, ms	87.86	74.25	21.72	97.08	7.40	14.89	13.54	9.89
Total drop length, ms	141.51	253.28	27.48	91.21	10.68	11.57	0.42	3.08

As a result of this first quantified voltage drop analysis for HitPix, conducted across multiple beam conditions, it was possible to identify systematic patterns in drop behaviour. Under the most demanding irradiation scenarios, HitPix3 consistently shows a reduced fraction of drops relative to signals, down to approximately ∼15%, whereas HitPix2 exhibits drop rates of up to ∼70% under identical conditions. The application of the ISO topology, combined with the use of a PMOS amplifier and additional deep pwell isolation, clearly improves signal integrity. Similar improvements from switching to a PMOS amplifier were previously observed in the ATLASpix1 and AstroPix chips. In addition to the significantly lower drop rate, a qualitative enhancement in signal morphology is also evident for HitPix3. From a design perspective, increasing the physical distance between the in-pixel amplifier and comparator is expected to further reduce proton-induced perturbations. Nonetheless, carbon ions remain the dominant source of drop-inducing events. Among all tested configurations within the HitPix chip family, HitPix3 demonstrates the most stable and robust response. These findings offer a concrete



(a) HitPix2, run 2: carbon beam, energy setting 255, intensity 7.



(b) HitPix3, run 6: carbon beam, energy setting 255, intensity 7.

Figure 8.23: Signals and drops per waveform for HitPix2 and HitPix3 under identical carbon beam conditions (energy setting 255, intensity 7). While the number of recorded signals is similar, HitPix3 shows a significantly lower number of voltage drops.

foundation for the next iteration of HitPix design and support the refinement of mitigation strategies against single-event-induced baseline collapse.

8.4 Summary and suggestions

This chapter addressed the issue of efficiency loss in monolithic HV-CMOS sensors exposed to heavy-ion beams, a phenomenon first observed during the development of the HitPix sensor family. Through a series of systematic measurements and waveform analyses under varying ion beam conditions at the HIT facility, the root cause of the problem was investigated and potential solutions explored.

During the beam test campaigns carried out to validate the HitPix sensor family, a recurring effect was observed: a sudden and prolonged drop in output voltage from the in-pixel amplifier, occurring exclusively during ion irradiation. This phenomenon, referred to throughout this work as *voltage drops* or *baseline drops*, was reproducibly detected across multiple HitPix variants and under different ion species, energies, and intensities. The consistency of the phenomenon motivated a deeper investigation, including full CMOS layout TCAD simulations, which suggested a possible origin in Single Event Effects (SEE), more specifically Single Event Latch-up (SEL). The voltage drop behaviour, manifesting itself as missing or distorted signal waveforms, was consistently observed across different sensors and beam settings. However, the frequency and severity of these drops were found to depend strongly on the sensor architecture.

AstroPix campaign revealed that substrate resistivity, pixel geometry, and beam parameters significantly affect signal stability. While high-resistivity substrates increase the number of recorded events, they also lead to more frequent and prolonged voltage drops with reduced peak amplitudes. The enlarged pixel pitch introduced in AstroPix_v3 reduced the drop-to-signal ratio to 2–3% and shortened recovery times, representing a key architectural improvement.

ATLASpix1 test series demonstrated robust performance across all design variants of this chip under proton beams. Voltage drops occurred primarily under carbon irradiation. The introduction of a CMOS comparator in the IsoSimple design did not negatively affect drop behaviour, confirming the effectiveness of the additional pwell isolation. From this, it can be concluded that from the perspective of voltage drop behaviour, switching from an NMOS comparator to a CMOS comparator has no significant influence due to the effective isolation provided by the additional pwell. The main architectural difference between the previously tested AstroPix, which exhibited voltage drops for both proton and carbon beams, and ATLASpix lies in the use of a PMOS amplifier instead of an NMOS amplifier. Factors such as smaller pixel pitch and lower substrate resistivity affected noise levels and recovery time but not the overall incidence of drops.

MuPix8 operated reliably across all tested beam conditions, with no voltage drops detected in either carbon or proton runs. Its analogue front-end, comprising an in-pixel PMOS amplifier and a peripheral CMOS comparator, maintained consistent performance, even at the highest intensities. Although signal-to-noise performance was lower due to the smaller pixel pitch, the absence of baseline shifts or signal loss affirmed the robustness of the design.

HitPix campaign enabled the first quantified study of voltage drops in this sensor family. Across all tested conditions, HitPix3 consistently demonstrated a significantly lower drop rate, typically below 15%, compared to HitPix2, which exhibited drop rates as high as 70%. The

transition to ISO pixel topology, adoption of a PMOS amplifier, and inclusion of additional deep pwell isolation markedly improved signal integrity. These improvements are consistent with trends observed in the ATLASpix1 and AstroPix series. Beyond the reduction in drop frequency, HitPix3 also exhibited improved signal shape and recovery behaviour.

These findings provide a solid foundation for the next generation of HitPix chips and offer practical design guidance for high-rate ion beam applications. In particular, the following suggestions are made:

- Further optimisation of the in-pixel layout should aim to maximise the physical separation between amplifier and comparator to mitigate proton-induced perturbations.
- Continued use of ISO topologies and PMOS amplifier instead of NMOS is strongly recommended, as they demonstrably improve resilience against substrate transients and voltage drops.

The quantitative analysis presented in this chapter offers concrete benchmarks and design strategies for future radiation-tolerant HV-CMOS sensors in environments such as ion beam therapy. Figure 8.24 shows the latest prototype of the HitPix detector, submitted to the Heidelberg Ion Beam Therapy Centre at the time of writing.

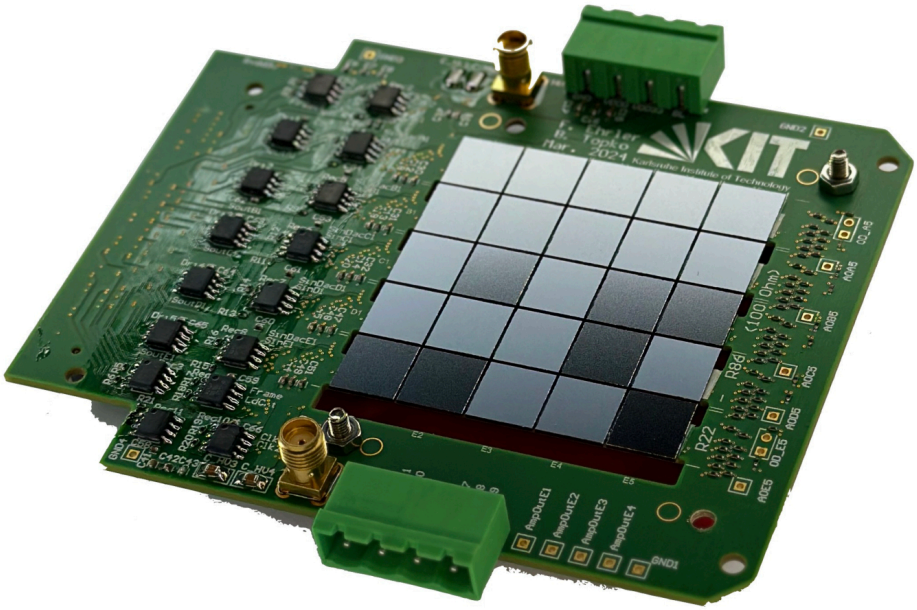


Figure 8.24: Assembled prototype of the HitPix detector, completed at the end of 2024.

9 Summary for HV-CMOS sensor technology development

High-Voltage Complementary Metal-Oxide-Semiconductor (HV-CMOS) technology as a silicon sensor concept was first proposed by Prof. Dr Ivan Perić in 2007 [7], targeting applications in high-energy physics. In this approach, both the sensing volume and the readout electronics are integrated monolithically on a single piece of silicon, fabricated using a standard commercial CMOS process. The concept has attracted increasing interest due to its favourable combination of cost-efficiency, fast timing, high spatial resolution, and radiation tolerance. These properties make HV-CMOS a competitive alternative to hybrid designs, especially for large-area tracking systems where reduced material budget and cost are critical. Over the past decade, the scope of HV-CMOS technology has broadened beyond collider experiments. Its robust performance in harsh radiation environments and flexibility in design have led to growing interest in space-based and medical imaging applications.

Chapter 6 presented the development of the *HitPix* sensor, a monolithic HV-CMOS pixel device designed for beam monitoring in ion beam therapy. HitPix is proposed as a next-generation replacement for the current system at the Heidelberg Ion Beam Therapy Centre (HIT), which relies on gas-filled ionisation and multi-wire chambers. It aims to deliver real-time monitoring with enhanced resolution and improved radiation tolerance.

The development of HitPix followed an iterative strategy, with each version targeting specific limitations of its predecessors. The initial prototype, HitPix1, primarily focused on the front-end electronics, laying the foundation for basic signal processing and pixel-level architecture. In HitPix2, NMOS amplifiers were introduced to increase readout speed; however, due to superior radiation tolerance, PMOS transistors were reintroduced in later versions. HitPix3 incorporated key insights from TCAD simulation studies, particularly those concerning substrate resistivity, charge collection efficiency, and pixel geometry, leading to substantial improvements in signal response. The latest version, HitPix4, represents the most advanced iteration, addressing previously observed efficiency losses, a topic examined in depth in Chapter 8.

Chapter 7 focused on *AstroPix*, a monolithic CMOS active pixel sensor developed for future gamma-ray telescopes. The AstroPix design originated from the ATLASPix sensor [100], initially intended for the ATLAS upgrade at the LHC. AstroPix adapts this architecture for space applications by prioritising low-noise operation, reduced power consumption, and optimised charge collection under low-energy photon interactions. Since its inception, AstroPix has gone through four design generations, from AstroPix_v1 to AstroPix_v4, each iteration implementing incremental improvements in analogue frontend, pixel layout, and data acquisition schemes.

Although HitPix and AstroPix were developed for very different applications, medical beam monitoring versus astrophysical photon detection, they share the same technological foundation. This illustrates the versatility of the HV-CMOS approach. At the same time, it highlights the importance of targeted pre-development and pre-production steps tailored to each use case. For monolithic designs in particular, successful sensor development depends not only on electronic performance, but also on the selection and optimisation of the underlying silicon substrate.

TCAD simulation-driven design

Physics-based modelling and numerical simulation have proven to be indispensable tools in this context. Technology Computer-Aided Design or TCAD simulations, in particular, were critical for exploring sensor response under different operating conditions, optimising charge-collection dynamics, and evaluating radiation-induced effects on the sensor. These simulations help predict how the sensor will behave prior to fabrication, making it easier to choose a suitable design and reduce the number of prototypes needed.

In the case of *HitPix*, the TCAD-guided development highlighted several key aspects relevant for both performance and long-term operation, showing that effective sensor design always requires a careful balance between competing factors such as depletion depth, breakdown stability, and layout constraints. Although high-resistivity substrates around 5k Ω·cm offer advantages in depletion behaviour, they also increase the risk of premature breakdown, especially within the tight layout constraints of the HitPix geometry. A substrate of 300 Ω·cm was identified as a more stable option, striking a balance between efficient depletion and breakdown resilience. Simulations also emphasised the impact of the periphery design, particularly in terms of electric field distribution and stability; an optimised periphery spacing of 15 µm was selected to suppress early breakdown events. Additionally, power consumption was a critical design parameter for HitPix3, where the final estimated value of 44.76 mW met the operational constraints of medical environments while considering long-term radiation ageing. Most notably, charge-collection simulations demonstrated that the ISO sensor architecture achieved a collection efficiency of up to 97.5%, significantly outperforming the standard layout due to improved depletion characteristics and optimised charge drift paths.

TCAD simulations for *AstroPix*, precisely AstroPix_v3, guided several key design decisions. A substrate resistivity of 10k Ω·cm was selected as the optimal compromise, offering near-full depletion at −250 V and stable operation up to −500 V without introducing breakdown risk or depletion non-uniformities. The chosen pixel geometry, featuring a 500 µm pitch and a 300 µm-wide deep nwell collection node, met the spatial resolution requirement of the AMEGO-X telescope. Surface breakdown was addressed by evaluating metal coverage schemes; placing the pad above the undepleted nwell yielded the most stable electric field distribution and avoided surface currents. For inter-pixel isolation, a p−stop spacing of 100 µm was found to minimise electric field peaks while remaining compatible with the large pixel pitch and expected radiation environment.

Investigation of heavy-ion beam influence on HV-CMOS sensors

In addition to application-driven sensor development, **Chapter 8** presented a fundamental investigation into the behaviour of monolithic HV-CMOS sensors under heavy ion irradiation. This work, initially triggered by observations during HitPix beam test campaigns, uncovered a phenomenon of sudden and prolonged decrease of signal amplitude and baseline, referred to as *voltage drops* or *baseline drops*, occurring during high-intensity ion beam exposure. The effect was reproducibly observed across several chip variants and beam conditions, prompting a dedicated study to identify its origin.

A comparative analysis was conducted using multiple sensor architectures, including HitPix2, HitPix3, AstroPix_v3, ATLASpix1, and MuPix8, each featuring different front-end electronics and pixel layouts. TCAD simulations with full CMOS layout extraction were employed to model possible *Single Event Effects (SEE)*, such as *Single Event Latch-up (SEL)*, that could explain the observed signal disruptions.

The results demonstrated that signal stability is strongly dependent on substrate resistivity, pixel pitch, and front-end topology. In *AstroPix*, high-resistivity substrates increased the number of collected events but also led to more frequent voltage drops. *ATLASpix*, featuring a PMOS amplifier and CMOS comparator, exhibited no voltage drops under proton irradiation, with the ISO topology providing effective pwell isolation. *MuPix8*, using an in-pixel PMOS amplifier and a peripheral CMOS comparator, showed no voltage drops under any tested conditions, validating the robustness of its architecture and the importance of physical separation of the front-end building blocks. *HitPix3*, which incorporated PMOS amplifiers and deep pwell isolation, consistently outperformed *HitPix2*, achieving drop rates below 15% compared to over 70% in earlier designs.

Design strategies for future HV-CMOS sensors

These findings offer critical guidance for the design of future HV-CMOS sensors intended for high-rate hadron and ion beam environments. Key strategies include maintaining PMOS-based front-ends with effective well isolation, preserving ISO topologies, and optimising in-pixel layout to separate sensitive analogue components and reduce substrate-induced perturbations. Together, these measures provide a path toward radiation-tolerant, monolithic detector systems suited for both medical and high-energy physics applications.

As HV-CMOS technology continues to expand into new fields, its success will rely not only on circuit-level design and process compatibility, but also on a deep physical understanding of semiconductor behaviour under application-specific constraints. The examples of HitPix, AstroPix, and the comparative irradiation study presented in this thesis demonstrate how simulation-driven sensor design, grounded in the physics of silicon devices, underpins the evolution of this technology toward increasingly demanding environments.

Part III

Bi-CMOS technology for high-speed monolithic sensors for particle physics

10 Motivation for technology application

Monolithic pixel sensors have become a key component in modern high-energy physics (HEP) experiments due to their fine spatial resolution, low material budget, and high integration potential. By integrating the sensor diode and readout electronics on the same silicon die, monolithic designs reduce interconnect complexity and enable large-area, low-mass tracking detectors. The most widely used implementations are based on CMOS technology, which offers micrometre-scale resolution, low power consumption, and high scalability. However, limitations remain in terms of time resolution, signal-to-noise ratio, and radiation tolerance, where hybrid solutions still outperform [7].

In this context, the feasibility of a monolithic pixel sensor based on BiCMOS technology has been investigated. The motivation behind this work is to assess the suitability of SiGe BiCMOS processes for charged-particle detection and to explore how the inclusion of SiGe HBTs can enhance analogue performance beyond what is achievable with CMOS alone. Key objectives include improving the signal-to-noise ratio, enhancing timing performance, and achieving fine spatial resolution within a compact pixel cell.

10.1 The BiCMOS technology

BiCMOS stands for Bipolar Complementary Metal-Oxide-Semiconductor, a technology introduced in the early 1980s that combines the high-speed switching characteristics of bipolar junction transistors (BJTs) with the low-power, high-integration logic capabilities of CMOS transistors. The first demonstration of BiCMOS occurred in 1983, and by 1985 the technology had progressed from research to production use [124]. This integration enabled designers to leverage the strengths of both device types within a unified platform, combining the speed of BJTs with the scalability of CMOS.

Initial BiCMOS processes incorporated conventional silicon BJTs, but a major leap forward came with the development of silicon-germanium (SiGe) heterojunction bipolar transistors (HBTs) in the late 1980s. The theoretical foundations for heterojunction transistors were laid much earlier, in 1948, by Shockley [16]. After his transistor patent, which envisioned a silicon emitter and germanium base, Kroemer formalised HBT theory in 1957 [125]. This type of transistor makes use of a bandgap-engineered emitter-base junction to improve current gain and high-frequency behaviour.

When implemented using silicon-germanium (SiGe) alloys, these devices offer both performance benefits and compatibility with standard silicon processing. HBTs leverage the advantageous bandgap engineering enabled by heterojunctions, which in turn are made possible

through advances in epitaxial growth techniques. This section outlines the essential physical principles and structural considerations underlying HBTs and their integration in BiCMOS processes. For further in-depth reading, the monography by Cressler and Niu [126] and the textbook by Sze and Ng [2] are highly recommended.

10.1.1 Heterojunctions

A heterojunction is formed at the interface between two dissimilar semiconductor materials. Unlike a homojunction, where the same semiconductor is doped differently, a heterojunction introduces a bandgap discontinuity (ΔE_g) that offers additional design flexibility. If the lattice constants of the two materials are well matched, epitaxial growth can produce sharp, defect-free interfaces with negligible interface trap density, allowing for efficient carrier transport [2].

Depending on the doping types of the materials, heterojunctions are classified as:

- Isotype heterojunctions, where both semiconductors are of the same conductivity type (n–n or p–p).
- Anisotype heterojunctions, where the conductivity types differ (p–n).

Anisotype heterojunctions are particularly valuable in device applications and are extensively used in photodetectors, light-emitting diodes, solar cells, and laser diodes. When grown in thin layers on the order of 10 nm, heterostructures can form quantum wells or superlattices, enabling a wide range of optoelectronic and high-speed electronic applications [2].

A critical parameter in heterojunction devices is the injection ratio of the diffusion currents J_n/J_p, where J_n and J_p are the electron and hole diffusion currents, respectively. For an anisotype heterojunction where the p–type semiconductor has the smaller bandgap, the injection ratio is approximately given by:

$$\frac{J_n}{J_p} \approx \left(\frac{N_A}{N_D}\right) \times \exp\left(\frac{\Delta E_g}{kT}\right) \tag{10.1}$$

Here, N_A and N_D are doping concentrations of p–and n–type semiconductors, ΔE_g is the bandgap difference, k is Boltzmann's constant, and T is temperature. The exponential dependence on ΔE_g enables significant enhancement of injection efficiency, which is directly linked to the current gain of bipolar transistors.

This principle is illustrated in Figure 10.1, showing the energy band diagrams of a heterojunction between a wide-bandgap n–type emitter and a narrow-bandgap p–type base. In the abrupt junction (solid lines), a conduction band offset ΔE_c introduces a barrier to electron flow. This barrier can be effectively mitigated by grading the material composition across the depletion region, resulting in a smoother transition and enhanced carrier injection (dashed lines).

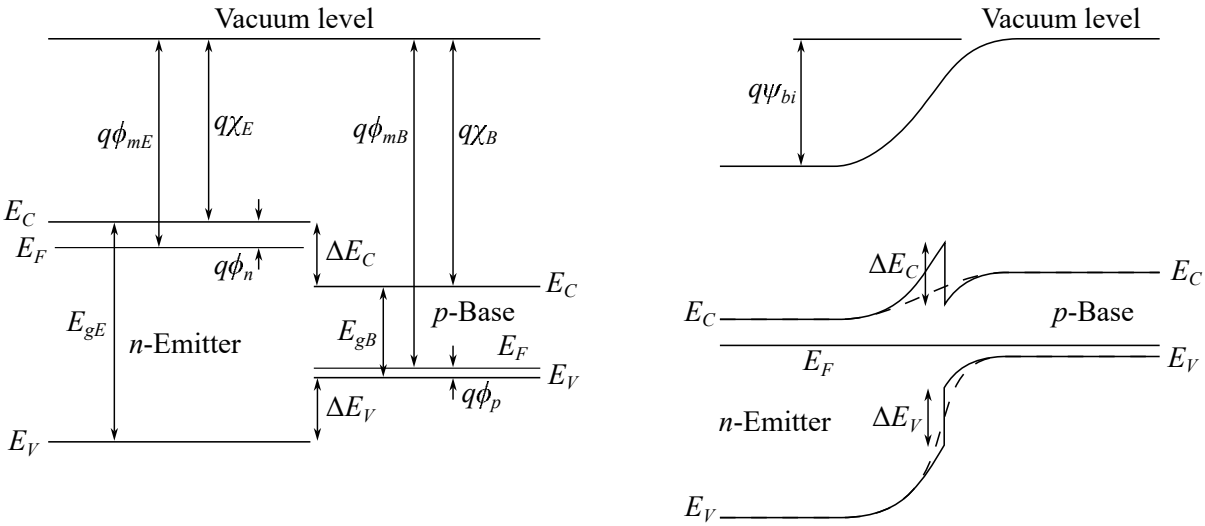


Figure 10.1: Energy band diagrams of a heterojunction between a larger bandgap $n-$type emitter and a smaller bandgap $p-$type base. (left) Before contact. (right) After junction formation. The abrupt junction introduces a conduction band barrier (ΔE_c), which is smoothed out in a graded junction (dashed lines), enhancing electron injection [2].

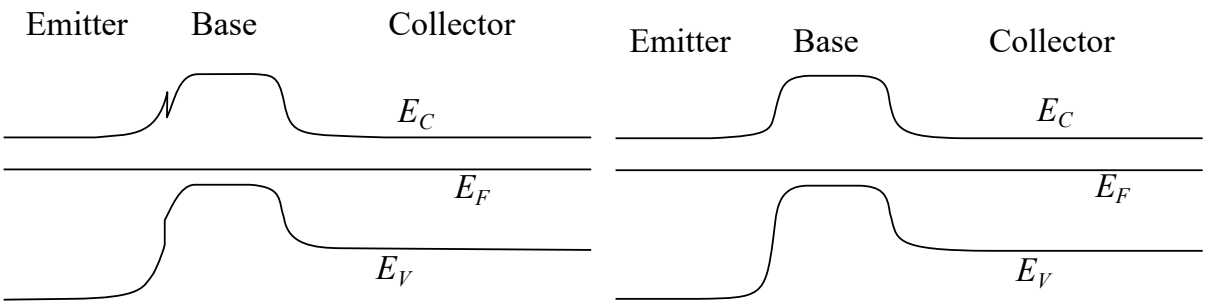
10.1.2 Heterojunction Bipolar Transistors (HBT)

The fundamental advantage of an HBT lies in its improved emitter injection efficiency. In a standard *n-p-n* BJT, the current gain depends on the ratio of electron injection from the emitter to hole injection from the base. By using a heterojunction with a wider bandgap emitter, the base current is suppressed due to the barrier for hole injection, while electron injection remains efficient. This results in a significant gain enhancement. Interestingly, in contrast to conventional BJTs, HBTs allow for a reversal of the doping profile: the base can be doped more heavily than the emitter while still maintaining a reasonable gain. This also enables several practical benefits:

- reduced base resistance, enhancing high-frequency performance,
- reduced high-injection effects,
- lower base-emitter capacitance due to reduced bandgap narrowing.

Figure 10.2 shows the energy band diagrams of two types of HBT emitter-base transitions: an abrupt heterojunction and a graded heterojunction. The graded profile, enabled by compositionally varied epitaxy, removes sharp potential barriers and supports improved carrier transport.

While the earliest and fastest HBTs were fabricated on III–V compound semiconductors, such as GaAs and InP, enabling high-frequency, low-capacitance designs, silicon-based HBTs employing SiGe epitaxial layers offer compatibility with CMOS fabrication. This enables full monolithic integration in BiCMOS technology.



(a) Abrupt HBT structure.

(b) Graded HBT structure.

Figure 10.2: Energy band diagrams illustrating the emitter-base heterojunction of (a) an abrupt HBT and (b) a graded HBT. The graded profile smooths out the band discontinuities, facilitating efficient carrier injection [2].

The three most common material systems used for HBTs include:

- GaAs-based: emitter/base = InAlAs/InGaAs.
- InP-based: emitter/base = InP/InGaAs.
- Si-based: emitter/base = Si/SiGe.

The III–V systems are typically fabricated using molecular beam epitaxy (MBE) or metal-organic chemical vapour deposition (MOCVD) to control the layer thickness and composition with atomic precision [2]. SiGe HBTs, on the other hand, can be realised through chemical vapour deposition (CVD) and are compatible with high-volume silicon wafer processing. This makes SiGe HBTs highly attractive for low-power, high-speed applications in both analogue and digital domains, particularly relevant for particle detectors, which benefit from fast front-end amplification with minimal power consumption.

10.1.3 Bandgap engineering

Although silicon has long been the dominant semiconductor in microelectronics, it presents several intrinsic limitations from a device physics perspective. Carrier mobilities in silicon are relatively modest, and the maximum drift velocity under high electric fields is limited to approximately 1×10^7 cm/s. As transistor speed is ultimately limited by carrier transport under operating bias conditions, silicon-based devices are fundamentally slower than their III–V counterparts [126]. Moreover, silicon is an indirect-bandgap material, making it poorly suited for optical emission. As a result, applications such as diode lasers and light-emitting devices are dominated by direct-bandgap materials like GaAs and InP, which also offer higher saturation velocities and superior carrier transport properties. From a design standpoint, one of the major advantages of III–V semiconductors lies in their tunability. Through epitaxial growth, the composition of these materials can be tailored at the atomic level to modify their bandgaps, transport properties, and emission wavelengths. This process, known as bandgap engineering, enables many of the performance advantages associated with III–V devices.

Despite their clear performance advantages, III–V semiconductors face practical limitations that restrict their use in large-scale integration. They lack a thermally grown oxide, such as SiO_2 for Si, which complicates device isolation and gate formation. In addition, they are mechanically fragile, show higher defect densities, are available only on smaller wafers, and conduct heat less effectively. As a result, although III–V devices are well established in high-frequency and optoelectronic applications, they remain impractical for mainstream CMOS-like integration [126].

However, the intrinsic limitations of silicon become significant in radio-frequency (RF) and microwave applications, where high transistor speed and gain are essential. In these regimes, silicon's lower carrier mobilities and saturation velocities restrict achievable performance, even if the overall platform remains scalable and cost-efficient. The introduction of SiGe HBTs addresses this gap by enhancing carrier transport and enabling higher-frequency operation, all within a process compatible with standard silicon fabrication.

The idea of using silicon-germanium alloys to engineer silicon's electronic properties dates back to the 1950s, with early conceptual work by Shockley. However, the practical implementation of defect-free SiGe films proved elusive until the mid-1980s. The challenge lies in the lattice mismatch between silicon and germanium, which is approximately 4.2%, inducing strain in the epitaxial layers [126]. When SiGe is grown on a silicon substrate, the film experiences compressive strain to match the substrate's lattice constant. This pseudomorphic growth enables the formation of thin, coherent layers that are stable and free of dislocations, provided the layer thickness and germanium content remain below a critical limit. Layers that fall within this stability region can undergo further processing, like thermal annealing or implantation, without defect formation. However, layers exceeding the critical thickness or grown at elevated temperatures can relax, resulting in dislocations and other defects that degrade device performance.

The successful integration of strained SiGe layers within standard CMOS processes has enabled the fabrication of high-speed, low-noise HBTs in a silicon-compatible platform. This development marks a key achievement in silicon bandgap engineering, allowing BiCMOS technologies to extend into domains traditionally dominated by III–V devices.

10.1.4 Properties of SiGe alloys

Silicon and germanium are both group IV semiconductors, crystallising in the diamond cubic structure. They are fully miscible, forming chemically stable silicon-germanium $Si_{1-x}Ge_x$ alloys across the entire compositional range. These alloys retain the diamond crystal lattice and exhibit a lattice constant that, to first order, follows Vegard's law [126]:

$$a_{Si_{1-x}Ge_x} = a_{Si} + x \times (a_{Ge} - a_{Si}) \tag{10.2}$$

Experimental measurements reveal a small deviation from this linearity, which can be modelled with a second-order fit [126]:

$$a_{\mathrm{Si}_{1-x}\mathrm{Ge}_x} = \left(0.002733 \times x^2 + 0.01992 \times x + 0.5431\right)\ \mathrm{nm} \tag{10.3}$$

The ability to precisely engineer alloy composition and strain makes SiGe an effective platform for bandgap engineering within the silicon ecosystem.

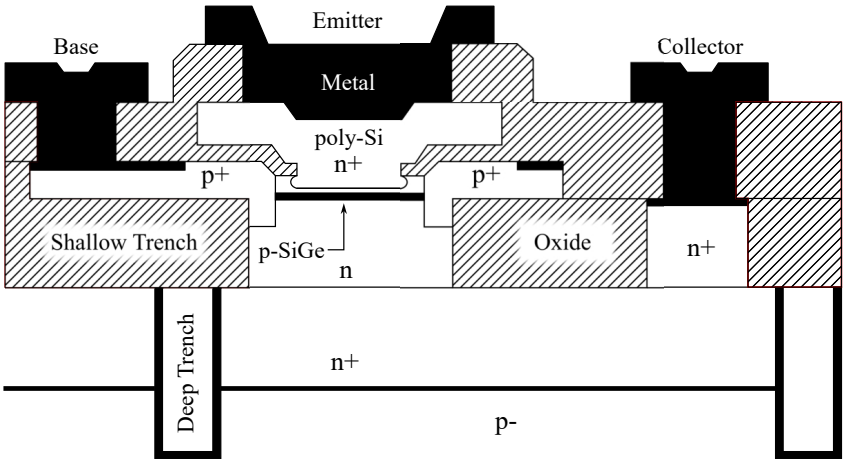
10.1.5 The SiGe HBT

The first functional SiGe HBT was demonstrated in December 1987, exactly 40 years after the invention of the bipolar transistor. Although the initial device exhibited significant leakage, it successfully demonstrated the core concept. A major milestone followed in 1990, when IBM reported a SiGe HBT with a cut-off frequency f_T of 75 GHz. It was roughly double the performance of the best silicon BJTs at the time. This breakthrough attracted worldwide attention and established SiGe HBTs as a serious candidate for high-frequency electronics. [127]

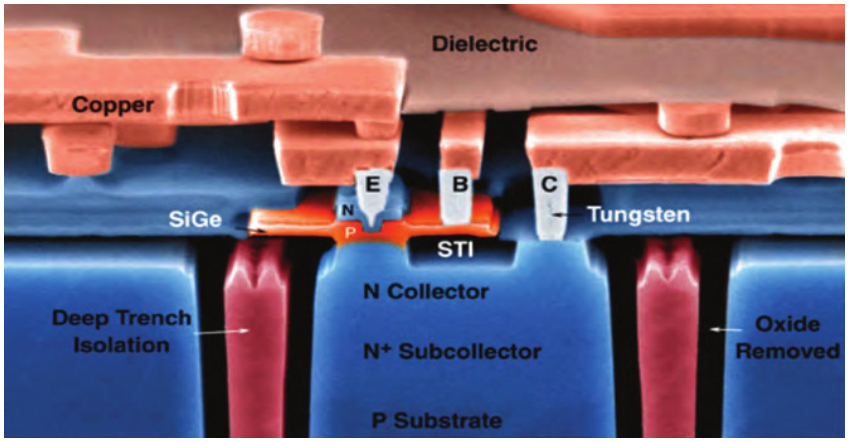
Introducing germanium into silicon modifies both the structural and electronic properties of the material. Because germanium has a larger lattice constant and a smaller bandgap than silicon (0.66 eV compared to 1.12 eV), adding germanium to silicon reduces the bandgap of the material. This makes it a suitable candidate for bandgap engineering. The resulting SiGe alloy has a smaller bandgap than pure silicon, and when grown epitaxially on a silicon substrate, compressive strain leads to further bandgap narrowing. For typical compositions, the reduction is around 75 meV per 10% Ge content [126]. This bandgap offset ΔE_g occurs primarily in the valence band, improving hole injection and making strained SiGe ideal for use as the base in a heterojunction bipolar transistor. A typical device structure consists of an $n-$type Si emitter, a $p-$type SiGe base, and an $n-$type Si collector, forming a double heterojunction. For brevity, the term "SiGe HBT" is used, although this technically refers to a double-heterojunction configuration. The SiGe HBT represents the first successful realisation of bandgap-engineered transistors in silicon. A schematic cross-section and a decorated scanning electron micrograph of a SiGe HBT are shown in Figure 10.3.

Although fabrication schemes vary, most SiGe HBTs share several core processing modules [126]:

- n^+ subcollector on a p^- substrate, often patterned to enable CMOS integration;
- lightly doped $n-$type collector epitaxy, 0.4–0.6 µm thick at $5 \times 10^{15}\mathrm{cm}^{-3}$;
- deep polysilicon-filled trenches for subcollector isolation (0.8–1.2 µm wide, 7–10 µm deep);
- oxide-filled Shallow Trench Isolation (STI) or LOCal Oxidation of Silicon (LOCOS) about 0.4–0.6 µm, planarised using chemical-mechanical-polishing (CMP);
- implanted collector sinker (reach-through) connecting to the subcollector;



(a) Schematic cross section of a first-generation SiGe HBT, drawn through the first metal. Not to scale [126].



(b) Decorated scanning electron micrograph of a SiGe HBT cross section. The epitaxial SiGe base layer is highlighted in orange. IBM Corporation [127].

Figure 10.3: Device-level structure and implementation of a first-generation SiGe HBT.

- composite SiGe epitaxial stack: Si buffer (10–20 nm), boron-doped SiGe or SiGeC active layer (70–100 nm, $1–3 \times 10^{13}\text{cm}^{-2}$ integral boron dose), and Si cap (10–30nm);
- emitter-base self-alignment schemes with spacers (0.1–0.3 µm), adapted from Si BJT processes;
- selectively implanted collector (SIC) for high-J_C operation and breakdown tuning (e.g. $0.5–1 \times 10^{17}\text{cm}^{-3}$ at the base junction, graded toward the subcollector);
- polysilicon extrinsic base contacts over the STI region, with self-aligned base implants to reduce sheet resistance;
- silicided extrinsic base;

- heavily doped polysilicon emitter (150–200 nm, $> 5 \times 10^{20} \mathrm{cm}^{-3}$), implanted or in-situ doped;
- multilevel back-end-of-the-line (BEOL) metallisation (Al- or Cu-based), typically 3–6 levels.

10.1.6 The SiGe HBT BiCMOS

A key advantage of SiGe HBTs is their seamless integration with standard silicon CMOS, enabling the development of monolithic SiGe HBT BiCMOS technologies. This co-integration is not a minor feature but a critical enabler for the long-term success of SiGe in commercial applications. Unlike III–V technologies, which typically require separate processing for analogue and digital functions, SiGe allows the analogue/RF performance of the HBT to be combined directly with the logic and integration density of CMOS on a single die. This fundamental compatibility distinguishes SiGe from compound semiconductor platforms and forms the basis for scalable, cost-effective system-on-a-chip (SoC) solutions [126].

In practice, SiGe HBTs are introduced as modular extensions, often referred to as “adders”, to a baseline CMOS process. These HBTs, which can be tailored for different breakdown voltages, are added where analogue or high-speed performance is required, without altering the digital CMOS core. The additional process complexity is modest: current SiGe BiCMOS platforms typically incur a mask count increase of only ~20% compared to standard digital CMOS. This trade-off is generally acceptable given the performance gains [126]. This integration strategy has become the industry standard. Modern SiGe HBT BiCMOS technologies are widely used in both wired and wireless domains, enabling the design of compact, high-performance ICs with mixed-signal capabilities. As a result, monolithic BiCMOS is now widely recognised as the natural evolutionary path for SiGe HBTs, one that combines the strengths of two complementary device types into a unified, scalable platform. An example of a monolithic SiGe BiCMOS technology platform is shown in Figure 10.4.

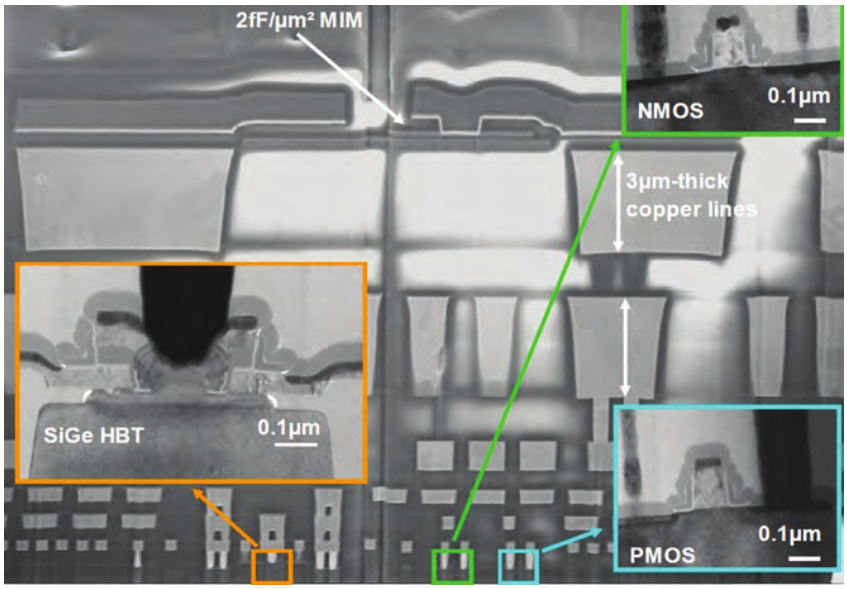


Figure 10.4: Cross-sectional scanning electron micrograph of a modern SiGe BiCMOS platform. The image shows co-integrated SiGe HBTs, n-channel and p-channel CMOS devices, and backend-of-line (BEOL) metallisation layers with passive components. ST Microelectronics [127].

10.2 BeMiPix concept and project motivation

BeMiPix, short for **Bipolar experimental Miniature Pixel**, is a fully monolithic silicon sensor developed in the SG13G2 130 nm SiGe BiCMOS process from IHP[1]. It is designed as a high-resolution, radiation-tolerant sensor for future HEP applications. The chip integrates SiGe HBT-based front-end electronics directly into each pixel, allowing fast signal amplification and shaping with low noise. The pixel cell also includes CMOS components for biasing and logic, realised within a deep-nwell architecture.

The sensor belongs to the class of Monolithic Active Pixel Sensors (MAPS), where the part of the readout chain is located inside the sensitive area. Each pixel has a compact 10 µm × 10 µm nwell collection node within a 100 µm × 100 µm cell. The front-end amplifier is directly integrated at the sensor node, reducing input capacitance and improving timing. The sensor diode is electrically isolated using p-stop structures. A photograph of the chip glued to the ASIC carrier is shown in Figure 10.5. The BeMiPix chip is located in the upper left quadrant.

The sensor was designed, implemented, and characterised within the ADL group at IPE, KIT. The development spanned the entire design flow, from device-level sensor layout and front-end circuit design to testing and validation under various conditions.

[1] https://www.ihp-microelectronics.com/

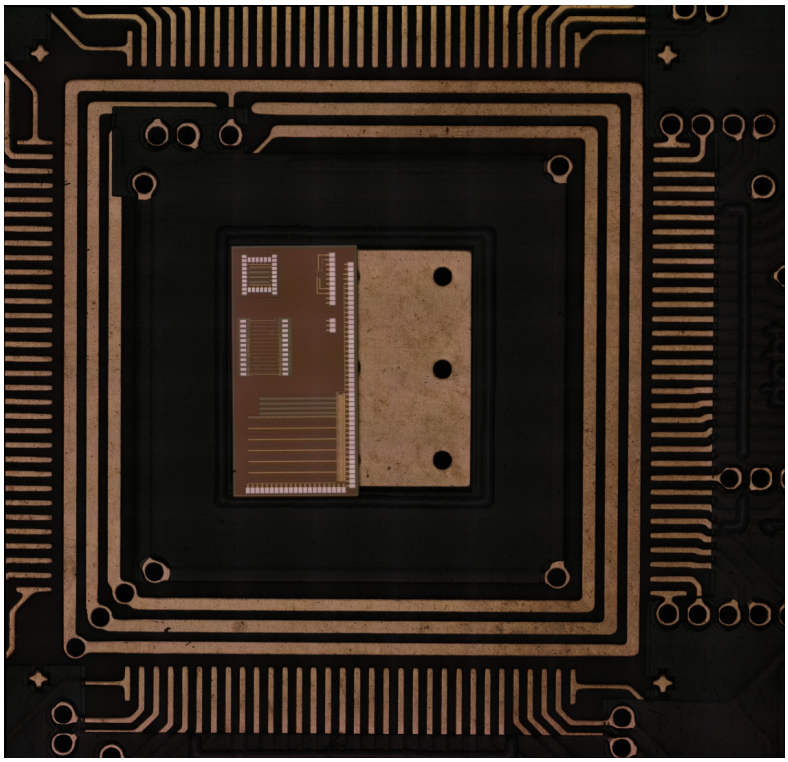


Figure 10.5: Photograph of the BeMiPix chip mounted on an ASIC carrier board. The chip is seen in the upper left quadrant. Wire bonding is used to connect the pads to the PCB traces.

Advantages of SiGe BiCMOS for sensor design

SiGe BiCMOS technology combines high-performance analogue functionality with the logic density of CMOS, offering a number of advantages, some of which are also relevant for particle sensor applications:

- **High speed and frequency performance:** SiGe HBTs achieve significantly higher transition frequencies (f_T) and maximum oscillation frequencies (f_{MAX}) than silicon BJTs, enabling operation in the mm-wave and sub-THz range.
- **High current drive and gain:** The heterojunction emitter-base interface increases carrier injection efficiency, resulting in higher gain and the ability to drive capacitive loads with fast signal edges.
- **Integration density and power efficiency:** By integrating both bipolar and CMOS devices on the same die, BiCMOS enables compact designs that combine fast analogue processing with low-power digital logic.
- **Process compatibility:** SiGe layers are incorporated into existing CMOS flows with relatively few modifications, making commercial fabrication feasible at 130 nm, 90 nm, and even 55 nm nodes [128].

SiGe BiCMOS for radiation-tolerant high-speed electronics

SiGe BiCMOS technology has shown excellent performance in environments that are both radiation-rich and cryogenic. SiGe devices have demonstrated a resilience to radiation effects and excellent operation at cryogenic temperatures. This has drawn significant interest from the aerospace and high-energy physics communities.

In space applications, SiGe circuits have demonstrated strong tolerance to total ionising dose (TID) and resistance to single-event upsets (SEUs), often outperforming conventional deep-submicron CMOS in terms of radiation hardness [127]. In HEP experiments, such as those at the Large Hadron Collider (LHC), front-end ASICs must operate in radiation fields that exceed fluences of 10^{15} n_{eq}/cm^2. They must also handle very fast detector pulses with low noise. SiGe HBT front-ends are well suited to this task: they provide fast signal amplification and shaping, while the integrated CMOS logic allows for on-chip trigger processing and data handling. This reduces the need for external components, saves power, and improves signal fidelity. Importantly, these circuits continue to function even after exposure to extremely high levels of radiation. For example, recent monolithic pixel sensors fabricated in a SiGe BiCMOS process remained operational at fluences up to 1×10^{16} n_{eq}/cm^2, with only modest degradation in timing and no significant change in signal-to-noise ratio up to about 6×10^{14} n_{eq}/cm^2 [129].

A practical example is a prototype monolithic sensor developed for the FASER experiment [130]. This chip was designed in the IHP SG13G2 130 nm SiGe BiCMOS process and features a matrix of hexagonal pixels, each containing a SiGe HBT-based analogue front-end and a CMOS discriminator. Timing resolution was measured with a ^{90}Sr source: the sensor achieved ~140 ps at 7 µA and down to 45 ps at 150 µA amplifier bias current, showing the trade-off between power and timing. These results demonstrate that sub-100 ps timing can be reached without internal gain, using only standard sensor structures and BiCMOS electronics.

Beyond particle physics, SiGe BiCMOS is also used commercially in 5G transceivers, satellite communications, automotive radar (LIDAR), and high-speed optical links. These systems benefit from the combination of analogue speed, digital integration, and robustness. This dual capability of fast analogue and scalable logic makes SiGe BiCMOS a promising platform for the next generation of radiation-tolerant ASICs.

Scope and structure of the BeMiPix study

Chapter 11 describes both the design and characterisation of the BeMiPix sensor in detail. The first part focuses on the development process. Subsection 11.1.1 introduces the physical structure of the sensor, including special features such as the selection of epitaxial layers and the layout of the doped regions. This is followed by Subsection 11.1.2, which presents the analogue readout chain embedded within each pixel. In Subsection 11.1.3, the influence of pixel pitch and layout on the breakdown voltage is analysed through TCAD simulations, with a focus on maintaining full depletion while avoiding premature breakdown. Finally, Subsection 11.1.4 discusses simulated charge collection from MIP-like tracks, providing a baseline for experimental comparison with source measurements.

The second part of the chapter reports the results of sensor characterisation. Section 11.4 presents measurements of dark current, X-ray response, and breakdown voltage. Section 11.5 shows temperature-dependent IV characteristics recorded in a climate chamber, illustrating thermal effects on leakage and breakdown. Section 11.7 discusses charge collection measurements using a ^{90}Sr source, focusing on depletion depth, signal amplitude, and timing response in comparison with TCAD predictions. Finally, Section 11.8 describes laser-based tests designed to investigate deeper depletion regions and to assess the timing resolution of the front-end.

11 Pixel sensor with miniature collection node — BeMiPix

This chapter presents the development and characterisation of the first monolithic sensor prototype designed using the SG13G2 130 nm SiGe BiCMOS technology from the Leibniz Institute for High Performance Microelectronics (IHP)[1]. SiGe BiCMOS combines standard silicon CMOS processes with high-performance SiGe heterojunction bipolar transistors (HBTs), offering a platform for compact, low-power, and ultra-fast integrated sensors. This technology has already shown strong potential in the development of advanced pixel detectors. The full design of the prototype, including both the analogue frontend and the sensor layout, was developed in-house within the ADL group at IPE, KIT.

This chapter provides a detailed overview of the design approach, the sensor architecture, and the electronics integration. Additionally, a comprehensive characterisation of the chip is presented, including results from laser-based scans, radioactive source tests, and current-voltage (IV) measurements. Together, these results establish a foundation for further optimisation of BiCMOS-based pixel sensors in future detector systems.

11.1 Sensor design and TCAD simulation

BeMiPix, short for **Bipolar experimental Miniature Pixel**, is a novel monolithic silicon sensor designed for high-resolution charged-particle detection. It combines a compact pixel geometry with integrated analogue readout electronics in a fully monolithic architecture. The sensor is developed in the IHP SG13G2 SiGe BiCMOS process.

The sensor belongs to the class of Monolithic Active Pixel Sensors (MAPS), with the entire readout chain placed inside the active area. Each pixel integrates a small nwell collection node that also hosts the front-end electronics, electrically isolated by a p-stop. This compact and vertically integrated approach reduces parasitic capacitance and improves noise performance.

The design concept was shaped by several constraints and performance goals. A range of structural options was considered in early simulations, including variations in epitaxial thickness, substrate resistivity, and pixel pitch. Although a variety of pixel geometries were explored, the 10 µm × 10 µm collection node proved to be the most promising option. It

[1] https://www.ihp-microelectronics.com/

provides a balance between charge sharing, spatial resolution, and compatibility with the IHP technology node.

A complete optimisation and validation workflow was carried out using the Technology Computer-Aided Design (TCAD) tool by SYNOPSYS[2]. These simulations guided the development of the sensor geometry and also provided predictions of electric field distributions, breakdown behaviour, and charge collection performance. The resulting layout achieves full lateral depletion and supports efficient charge sharing at a bias voltage below the breakdown limit.

The subsections that follow provide a detailed view of each aspect of the design process:

- In Subsection 11.1.1, the physical structure of the sensor is introduced. This includes the choice of epitaxial layers and the design of doped regions.
- Subsection 11.1.2 focuses on the readout electronics embedded in the pixel. The analogue front-end, implemented with a SiGe HBT amplifier and biasing components, is described along with its signal path and output stage.
- In Subsection 11.1.3, the influence of pixel pitch and geometry on sensor breakdown is analysed. TCAD simulations of various pixel spacings are used to optimise for electrical breakdown while maintaining full depletion and charge sharing.
- Finally, Subsection 11.1.4 presents the simulation of charge collection and inter-pixel charge sharing using MIP-like particle tracks. These results are key for understanding the signal formation dynamics and serve as a reference for later ^{90}Sr measurements.

11.1.1 Development of the structure

The Technology Computer-Aided Design (TCAD) tool by SYNOPSYS was used for simulating the BeMiPix sensor structure. This tool allows for the modelling and optimisation of semiconductor devices and process technologies. two primary TCAD modules were used: the *Sentaurus Structure Editor* (SDE) for geometry generation and meshing, and the *Sentaurus Device* (SDEVICE) simulator for electrical and charge transport simulations.

The SDE tool was used to create the sensor's physical geometry, define doping profiles, and generate the simulation mesh. This includes specifying key structures, such as bulk and epitaxial layers, nwell and pwell implants, and p-stop regions. The tool also enables local refinement of the mesh to ensure high resolution where abrupt doping gradients or fine features are present. The SDEVICE simulator was then used to perform two types of electrical simulations: quasi-stationary and transient. In the quasi-stationary mode, electrostatic fields and potential distributions are calculated under steady-state bias conditions. In the transient mode, dynamic phenomena are modelled, such as the charge deposition from traversing particles and the subsequent drift and diffusion of carriers. Further details on how TCAD simulation models were developed and validated can be found in Chapter 4, "Simulations of semiconductor devices — Technology Computer-Aided Design".

[2] https://www.synopsys.com/manufacturing/tcad.html

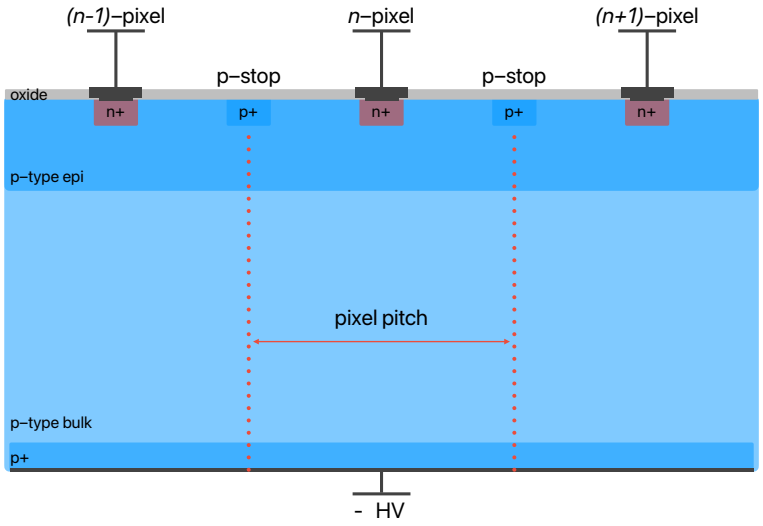
The sensor is fabricated using a high-resistivity epitaxial $p-$type silicon layer that was grown using chemical vapour deposition (CVD). It has a resistivity around 5k Ω·cm and a typical doping concentration in the range of 1×10^{12} to 1×10^{13} cm^{-3}. The goal of the epitaxy is not just higher resistivity but also better material quality at the surface, which helps reduce leakage current and improve uniformity in charge collection, making it well-suited for high-performance, low-noise pixel sensors. The layer is grown on top of $p-$type silicon bulk with a resistivity of 50 Ω·cm, which is a standard value for electronics-grade silicon. The bulk doping is approximately 3×10^{14} cm^{-3}.

To study the influence of the epitaxial layer, several sensor variants were simulated. These included structures with epitaxial layers of 20 µm and 30 µm thicknesses, as well as a reference design fabricated directly on the 50 Ω·cm substrate without any epitaxial layer. The comparison of these variants allows for a deeper understanding of how epitaxial thickness and material quality affect the electrical performance, breakdown characteristics and charge collection of the device. These studies are discussed throughout this thesis.

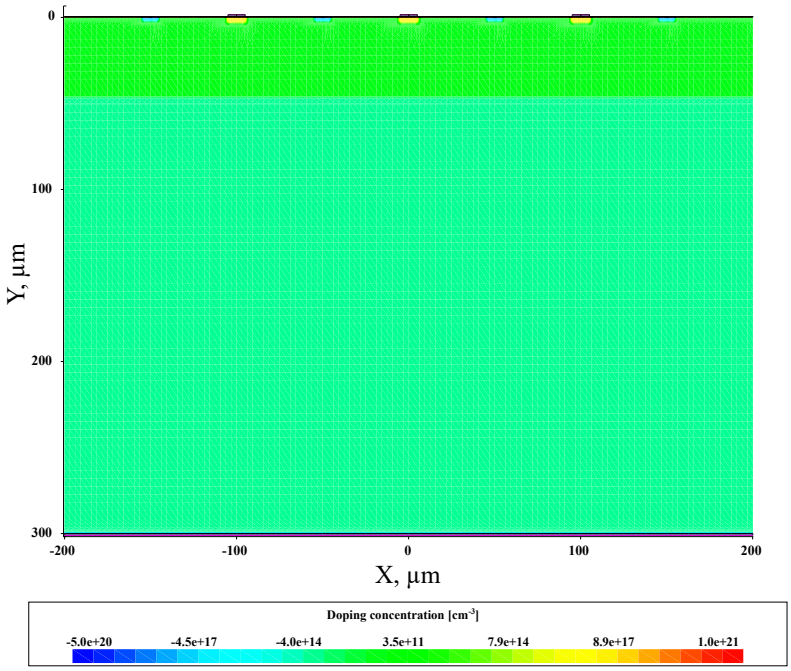
All doped regions, including nwell and pwell implants, were defined using Gaussian profiles based on manufacturer data. Doping values and depth distributions were selected to match measurements from Secondary Ion Mass Spectrometry (SIMS) analysis provided by IHP. For the purposes of the simulation, the high-voltage bias was applied to the backplane contact, which is located on the bottom of the simulated structure. In the actual device, this biasing is applied from the top via the metal layer, but the simplified back-contact configuration is commonly used in TCAD studies to reduce simulation complexity.

A critical step in TCAD device simulation is the generation of a high-quality mesh. The simulation domain is divided into small triangular or tetrahedral elements, within which Poisson's equation and the carrier continuity equations for electrons and holes are numerically solved. To ensure accuracy, a doping-dependent mesh refinement was used: finer mesh elements were introduced in regions with steep doping gradients, such as the p-stop, nwell collection node, and backplane implant areas.

Figure 11.1a shows the cross-section of the sensor structure. The structure includes the $p-$type bulk, the epitaxial layer, p-stop isolation, and three neighbouring pixels with their respective nwell collection nodes. The $p - n$ junction is formed between these collection nodes and the $p-$type or bulk. The oxide, passivation layers, and the high-voltage contact are also visible. Figure 11.1b shows the corresponding simulated doping profile. The colour scale spans from 3.5×10^{11} to 1×10^{21} cm^{-3}, and highlights the difference between the lightly doped epi and the heavily doped contacts. The nwell collection node, in yellow, sits at the top surface and defines the active sensing region.



(a) Cross-sectional schematic of the BeMiPix sensor, showing the pixel layout and vertical structure. The central and neighbouring pixels are visible, with labelled doping regions and key features such as the nwell collection node, p-stop, and epitaxial layer.



(b) Example TCAD simulation output displaying the doping profile. The nwell implants forming the collection nodes are highlighted in yellow at the top surface. The distinction between the low-doped epitaxial region and the more heavily doped bulk is clearly visible.

Figure 11.1: Geometry and doping structure of the BeMiPix sensor used for TCAD simulations. These figures illustrate both the design layout and the doping gradients relevant for device operation and breakdown analysis.

11.1.2 Readout electronic design

The analogue readout chain of the BeMiPix chip is designed using SiGe BiCMOS technology, which combines high-performance bipolar transistors with CMOS functionality. A simplified schematic of the front-end circuit is shown in Figure 11.2.

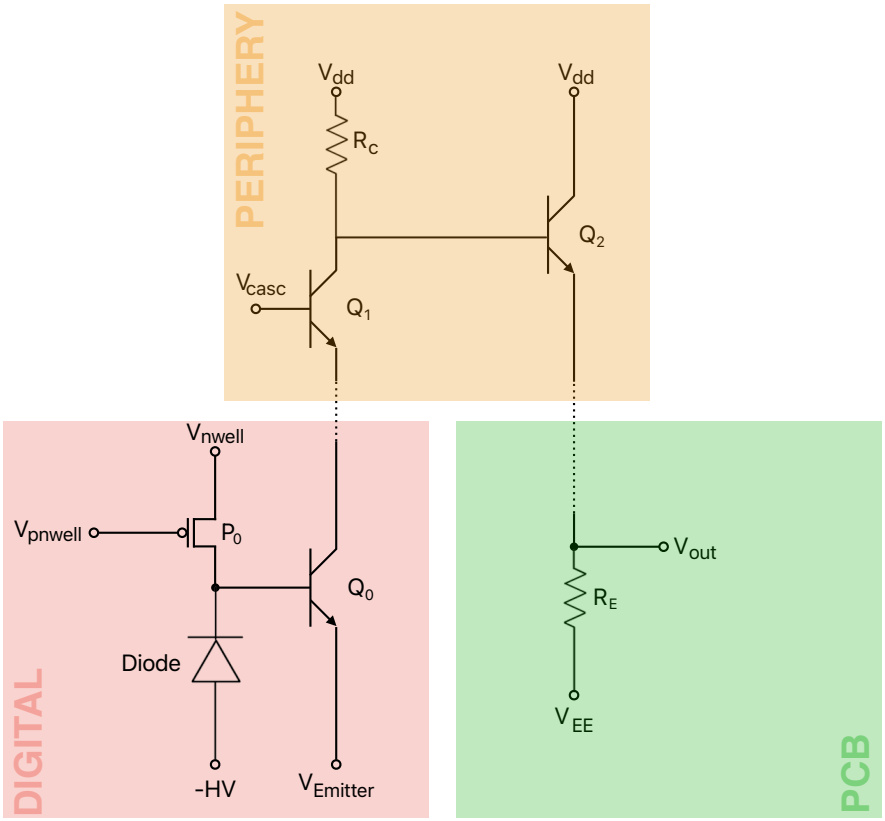


Figure 11.2: Simplified schematic of the BeMiPix analogue readout chain. Each pixel integrates a charge-sensitive diode, a PMOS bias transistor (P_0), and a SiGe HBT transconductance amplifier (Q_0). A common-base stage (Q_1, R_C) at the chip periphery amplifies the current signal. The system is read out via an analogue output stage, where all emitters are summed at V_{EE} and buffered off-chip.

The pixel features a 10 µm × 10 µm nwell collection node, where charge generated in the silicon substrate is collected. This collection node is part of a reverse-biased $p - n$ junction diode, which is electrically connected to a PMOS transistor labelled P_0. The transistor has a width of 1 µm and a channel length of 400 nm. The source of P_0 is connected to the node potential V_{nwell}, and the gate to V_{pnwell}, both of which are routed to external bond pads. This configuration enables tuning of the diode biasing conditions, such as the discharge time constant of the detector capacitance, directly from outside the chip.

When ionising radiation generates charge in the sensor, the potential of the nwell node drops due to the discharge of the detector capacitance. This voltage drop is slowly compensated as

the capacitance recharges through the bias PMOS, generating a transient voltage signal at the diode node. The analogue signal is then fed into an in-pixel voltage-to-current converter, implemented using a SiGe heterojunction bipolar transistor (HBT), labelled Q_0. This is an $n-p-n$ transistor configured to operate as a transconductance amplifier. When the input is voltage-driven, the relevant performance metric is the transconductance g_m, defined as the derivative of the collector current I_C with respect to the base-emitter voltage V_{BE}.

The collector current of the transistor, under appropriate boundary conditions, can be described by the expression:

$$I_{nC} = \left(\frac{A_E \times q \times D_n \times n_{p0}}{L_n} \right) \times \frac{1}{\sinh\left(\frac{W}{L_n}\right)} \times \exp\left(\frac{qV_{BE}}{kT}\right) \tag{11.1}$$

where

- A_E is the emitter-base junction area, cm^2;
- q is the elementary charge, equal 1.6×10^{-19} C;
- D_n is the diffusion coefficient for electrons, cm^2/s;
- n_{p0} is minority carriers in thermal equilibrium, here the electron concentration in a $p-$type semiconductor, cm^{-3};
- L_n is the electron diffusion length, cm;
- W is the width of the neutral base region, cm;
- V_{BE} is the base-emitter voltage, V;
- kT/q is the thermal voltage, approximately 0.0259 V at room temperature 300 K.

From this, the transconductance can be derived as:

$$g_m \equiv \frac{dI_C}{dV_{BE}} = \frac{q}{kT} \times I_C \tag{11.2}$$

indicating that g_m is directly proportional to the collector current I_C.

At the chip periphery, the signal current from the in-pixel HBT is delivered to a common-base amplifier stage formed by transistor Q_1 and load resistor R_C. This architecture provides a low-impedance signal path between the pixel and the peripheral electronics. The voltage amplification at this stage is given by $g_m \times R_C$. A low-impedance output is beneficial because it reduces capacitive crosstalk and enables high-speed signal transmission over long interconnects with relatively high parasitic capacitances, all while maintaining low power consumption.

The emitter of the in-pixel HBT (Q_0) is routed to external bond pads. This enables fine adjustment of the pixel's operating point and facilitates biasing of the analogue amplifier chain. Since the BeMiPix prototype does not include a digital readout, analogue signals must be transmitted off-chip and observed via an oscilloscope. This is achieved through a single-ended

analogue output. The output node V_{out} requires an external pull-down resistor on the PCB to establish the correct bias current. In the design, all pixel emitters are connected and summed into a shared low-impedance node V_{EE}, stabilised using a decoupling capacitor C_E to suppress common mode noise and enhance signal fidelity.

11.1.3 Optimising pixel pitch for breakdown behaviour and depletion

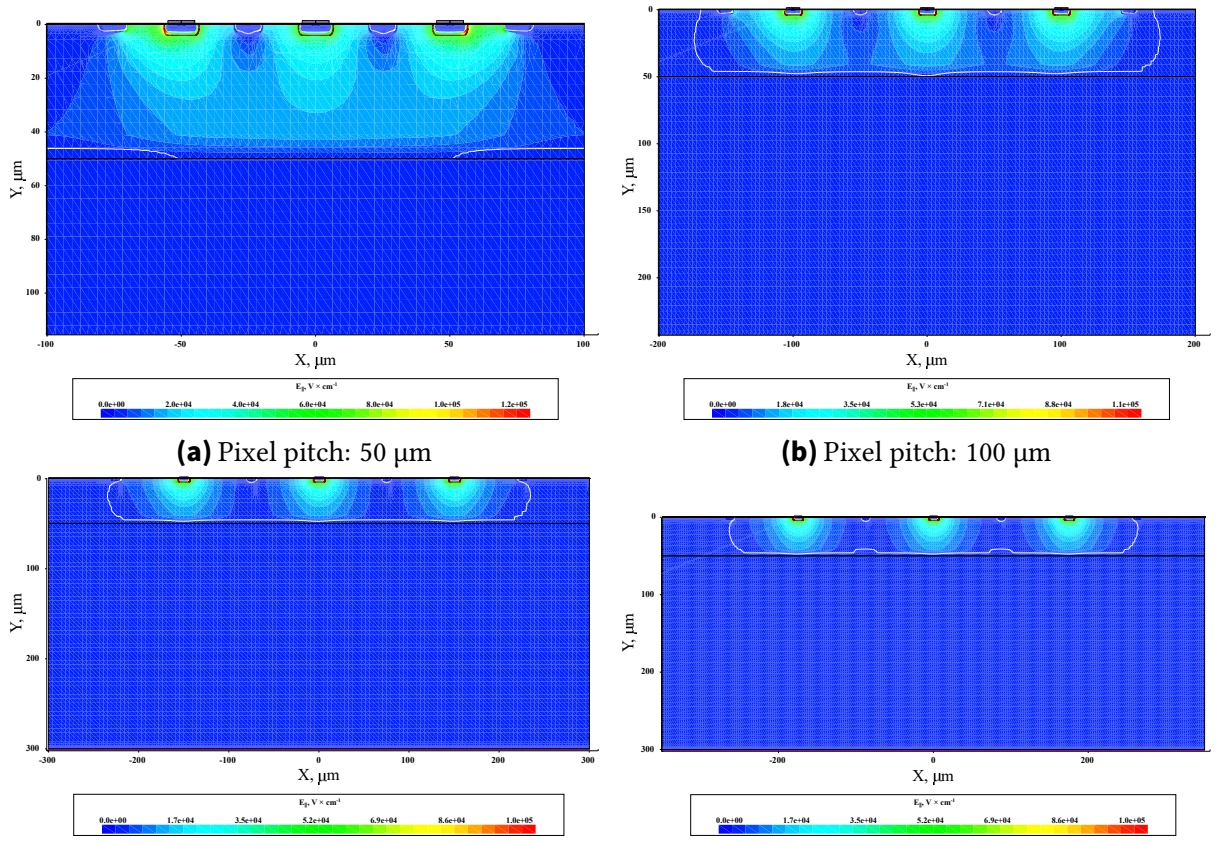
The breakdown behaviour of the BeMiPix sensor was studied through TCAD simulations for a range of pixel geometries and substrate configurations. The objective was to identify a pixel pitch that supports lateral depletion across the entire sensor while avoiding premature breakdown effects near the pixel edges.

Simulations were conducted using $p-$type epitaxial substrates of 50 µm, 30 µm, and 20 µm thickness. For each configuration, the pixel pitch was varied between 50 µm and 200 µm in 10 µm steps. From these results, four representative pitch sizes, 50 µm, 100 µm, 150 µm, and 200 µm, were selected for visual comparison and are shown in Figure 11.3. In these images, the colour scale represents the electric field at an applied bias of −100 V, and the white lines indicate the extent of the depletion region.

At small pixel pitches, from 50 µm until 100 µm, the electric field is strongly enhanced between nwell regions and p-stop structures. This creates localised field peaks that increase the risk of premature breakdown before full depletion is achieved. At larger pitches, from 150 µm until 200 µm, the field remains well-behaved, but full lateral depletion across the pixel, even at −100 V becomes difficult to reach. This limits charge sharing and weakens spatial resolution.

Intermediate pitches, particularly between 100 µm and 150 µm, provided the most favourable trade-off. At these values, the depletion regions from neighbouring pixels merge fully at around −100 V, forming a common depletion zone without introducing excessive electric field distortions near the collection nodes. Among these options, 100 µm was selected as the optimal pitch. This configuration not only ensures uniform depletion and charge sharing but also matches the layout constraints of IHP SG13G2 technology, allowing a practical integration of front-end electronics without risking edge breakdown between the nwell and p-stop implants. The 100 µm pitch also supports regular matrix arrangements, such as 4×4 or 3×3, which are advantageous for both characterisation and ASIC routing. For comparison, a larger pitch, like 150 µm or more, would allow only 2×2 pixel blocks in the same chip area.

In a later simulation stage, a reference design using only the 50 Ω·cm bulk substrate without an epitaxial layer was also introduced. This configuration was used to investigate the impact of substrate quality on the breakdown behaviour and overall device functionality. Experimental measurements from these test structures using radioactive sources and laser injection are presented in Sections 11.7 and 11.8, respectively.



(a) Pixel pitch: 50 µm

(b) Pixel pitch: 100 µm

(c) Pixel pitch: 150 µm

(d) Pixel pitch: 200 µm

Figure 11.3: Simulated electric field and depletion region (white line) at $V = -100$ V for sensors with different pixel pitches. Although not all substrate thicknesses are visible in each panel, the focus here is on the behaviour within the epitaxial region.

11.1.4 Charge collection and charge sharing simulations

This study was conducted with two objectives: to investigate the possibility of charge sharing between adjacent pixels and to simulate charge collection dynamics in preparation for experimental validation using a ^{90}Sr beta source.

Details of the simulation method, including models for radiation damage, are presented in Section 4.3. Briefly, charge deposition was modelled using the `HeavyIon` function within Synopsys Sentaurus TCAD. The simulation corresponds to the passage of a Minimum Ionising Particle (MIP), implemented with a linear energy transfer (LET) of 1.282×10^{-5} pC/µm, which is approximately 80 $e-h$ pairs per micrometre of silicon traversed [131]. The avalanche multiplication process was activated using the University of Bologna impact ionisation model, and bulk radiation damage was incorporated using the Hamburg Penta Trap model. Surface effects were included via fixed oxide charge densities of $N_{ox} = 1 \times 10^{11}$ cm^{-2} for unirradiated and $N_{ox} = 3 \times 10^{12}$ cm^{-2} for irradiated silicon [132].

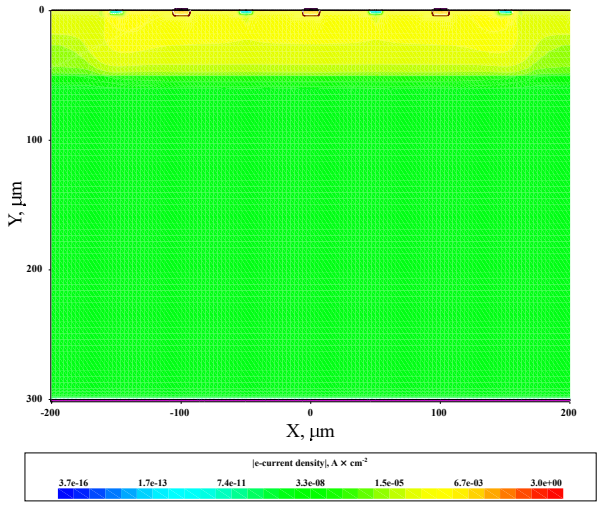
Figure 11.4 shows the simulated electron current density at several time points following the MIP impact. The simulated structure features a 50 µm epitaxial layer on top of a p–type

low-resistance substrate. The applied bias voltage is −75 V, which was later established as the operational limit for this chip through experimental measurements.

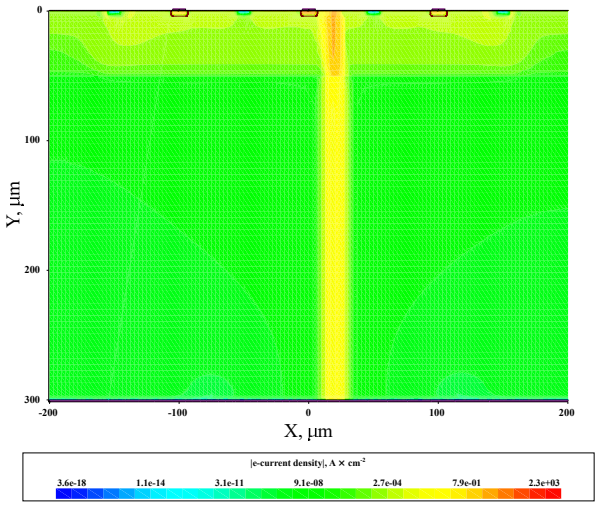
At $t = 0$ ns, before the MIP impact, the electron density is distributed uniformly across the active region. At $t = 1$ ns, a sharp localised current appears at the beam entry point, marking the moment of energy deposition. By $t = 2$ ns, most of the charge has been collected through drift processes. At $t = 3$ ns, drift collection is largely complete, and diffusion becomes the dominant transport mechanism. By $t = 8$ ns, only residual diffusion currents remain.

In addition to verifying the expected timing behaviour, these simulations provide visual evidence of charge sharing. Figure 11.5 presents zoomed-in images at $t = 2$ ns and $t = 3$ ns, showing the lateral spread of the electron current. Notably, charge carriers can be seen reaching neighbouring collection nodes, indicating the existence of a common depletion region and confirming inter-pixel charge sharing.

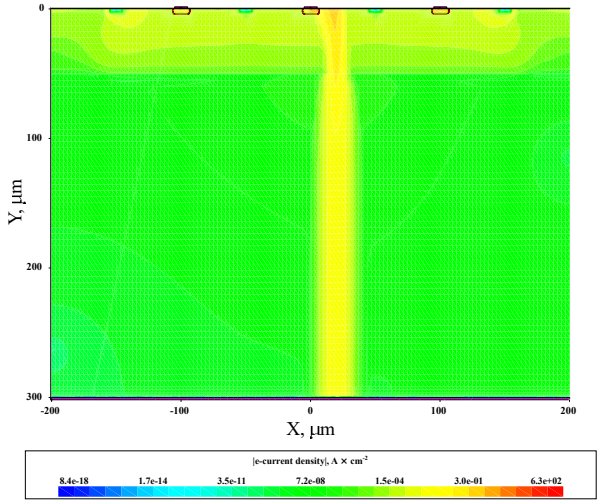
These results are in agreement with the outcomes of the ^{90}Sr characterisation measurements presented in Section 11.7. The consistency between simulation and measurement strengthens the confidence in the predictive value of the TCAD model used in this study.



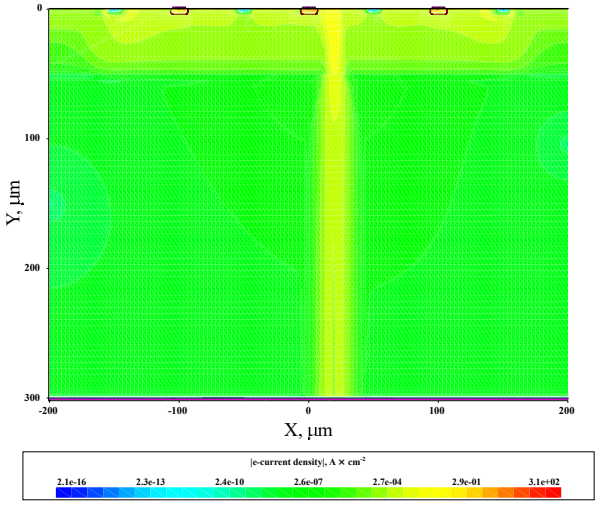
(a) Time = 0 ns (pre-impact)



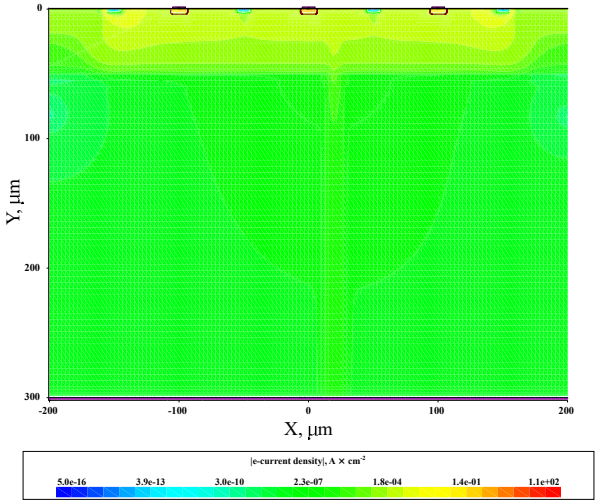
(b) Time = 1 ns (MIP impact)



(c) Time = 2 ns (drift collection)

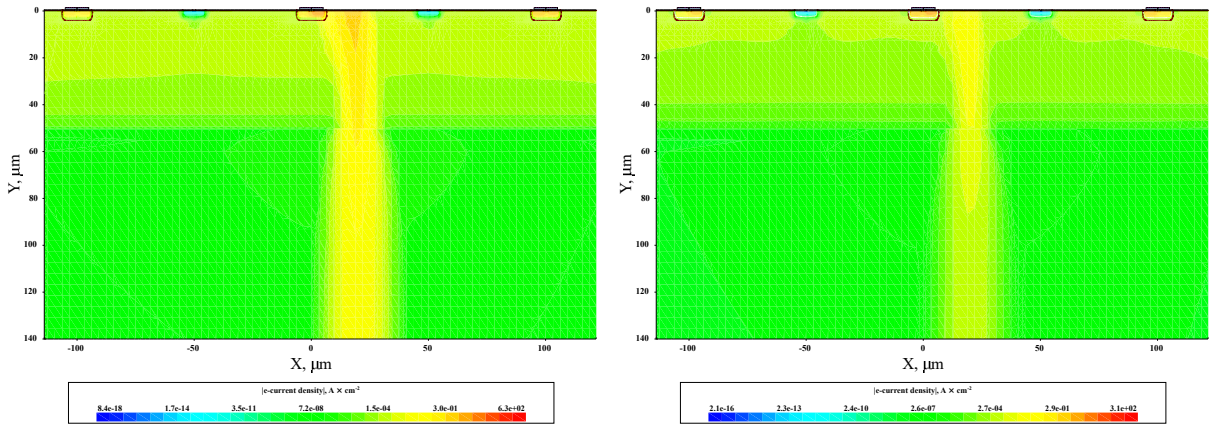


(d) Time = 3 ns (transition to diffusion)



(e) Time = 8 ns (residual collection)

Figure 11.4: Time evolution of the electron current density following a MIP impact at $V = -75$ V. The simulated structure corresponds to a 100 µm pitch sensor with a 50 µm epitaxial layer. The current density scale reflects charge movement toward the nwell collection nodes.



(a) Time = 2 ns (zoomed)

(b) Time = 3 ns (zoomed)

Figure 11.5: Zoomed view of the electron current density during charge sharing. Carriers are visible reaching multiple collection nodes, demonstrating inter-pixel connectivity and confirming the presence of lateral charge sharing.

11.2 Layout and production

The BeMiPix chip was designed using the Cadence[3] layout package, with the primary goal of minimising the size of the active pixel while maintaining compatibility with analogue front-end integration. Each pixel incorporates a 10×10 μm^2 nwell collection node, which also serves as the placement area for the in-pixel electronics. The node is surrounded by two concentric polysilicon rings and isolated by a p-stop structure.

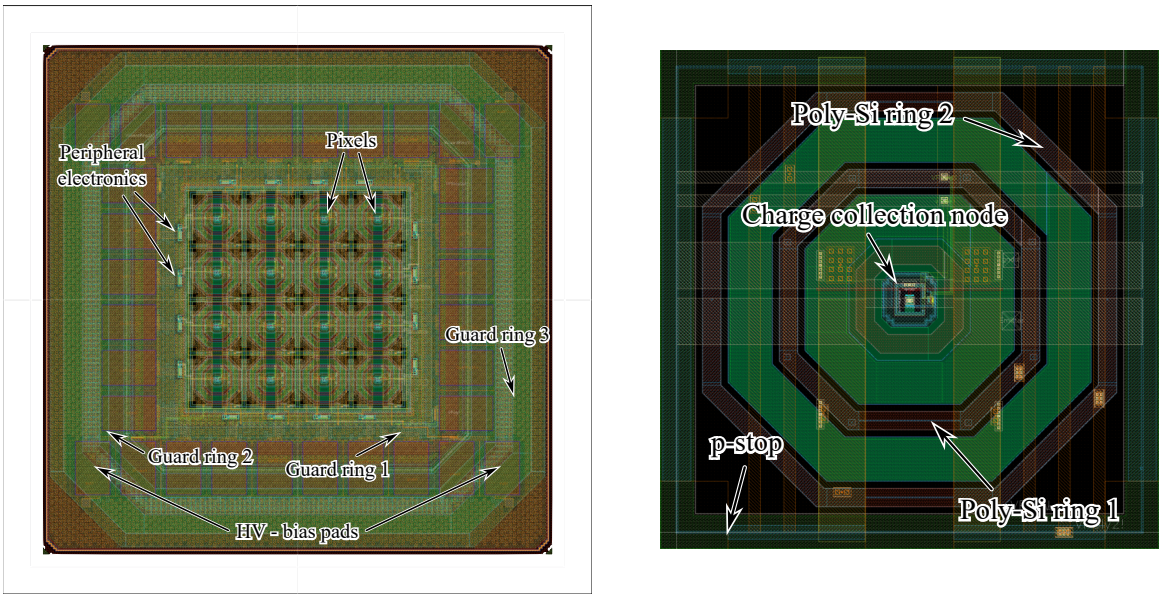
The selected pixel pitch is 100 μm, which was optimised to ensure full lateral depletion without introducing premature breakdown, as discussed in Section 11.1.3. The pixel matrix is composed of 4×4 pixels, resulting in a total chip area of 1×1 mm^2. Each pixel is separated by a p-stop implant.

Three guarding rings enclose the matrix. The innermost ring is connected to ground and supports peripheral electronics, while the two outer rings are accessible via individual pads and allow voltage tuning to control the electric field distribution during operation. In all standard measurements, these rings were left floating. Substrate bias is provided via four dedicated pads located at the corners of the die.

The final layout includes a total of 32 pads: 10 along the top edge, 10 at the bottom, and 6 on each vertical side. The pad pitch is 85 μm, defined by a 65 μm × 100 μm pad area (TopMetal2) with 20 μm spacing. The passivation opening measures 95.8 μm × 60.8 μm. In total, the chip includes 16 output channels. The nwell collection node and the surrounding poly rings and p-stop are all routed to probing pads for characterisation.

An overview of the full chip layout is shown in Figure 11.6a, and a zoomed view of a single pixel layout is shown in Figure 11.6b.

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



(a) Full chip layout with 4×4 pixel matrix, pads, guard rings, and routing

(b) Pixel layout showing the 10×10 µm^2 nwell collection node, polysilicon rings, and p-stop implant

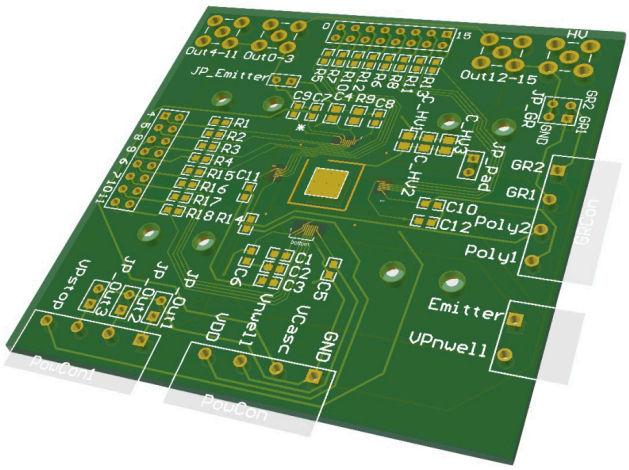
Figure 11.6: BeMiPix chip layout: full matrix view (a) and single pixel detail (b)

11.3 Characterisation and evaluation

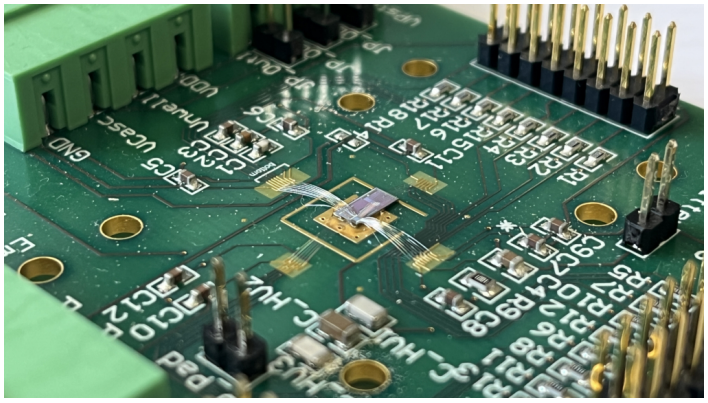
The printed circuit board (PCB) developed for BeMiPix is shown in Figure 11.7a. It provides both mechanical support and electrical interfacing for the chip during characterisation. The chip is glued onto a central metallic pad, which is thermally connected to the backside metal plane of the PCB through vias, enabling passive cooling. The chip requires several supply voltages for operation. These include the main supply voltage V_{DD}, the cascode bias V_{Casc}, and additional bias voltages for the internal wells: V_{nwell} and V_{pnwell}. The sensor itself is biased using a negative high-voltage supply V_{HV}. Optional bias lines are available for the guard ring structures and surrounding implants, including V_{pstop}, V_{GR1}, V_{GR2}, V_{Poly1}, and V_{Poly2}, which may be either externally set or left floating during operation. All supply lines are wire bonded from the chip to the PCB and decoupled via on-board capacitors before reaching external power connectors.

The analogue signal output of each pixel is buffered by an emitter follower stage. The output current can be set either by direct connection to a power source on the PCB or by selecting pull-down resistors to ground. All 16 output channels are routed to dedicated pins for voltage readout. Each channel includes a pull-down resistor to initialise the line to a known state during reset.

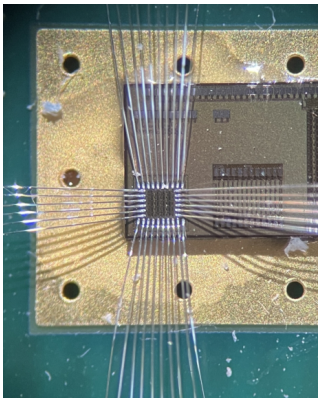
The chip was wire bonded to the PCB with aluminium wires, using the Hesse Bondjet BJ653 wedge bonder. Details of the wire bonding process and other interconnection technologies can be found in Chapter 13. Figure 11.7b shows the assembled board, and Figure 11.7c provides a detailed view of the wire bonds connecting the chip to the pad frame.



(a) 3D rendering of the BeMiPix PCB



(b) Assembled chip wire bonded to the PCB



(c) Close-up of the aluminium wire bonds

Figure 11.7: BeMiPix chip mounted on a custom-designed PCB with wire bonding to external power and signal lines

An overview of the fabricated and tested samples is provided in Table 11.1. Variants include designs with and without epitaxial layers, different substrate resistivities, and varying production runs from 2021 to 2024.

Table 11.1: Overview of sensor wafer types and their substrate properties

Sample name	Substrate type	Resistivity	Doping concentration	Epitaxial thickness	Special feature
s1.1	Standard	50 Ω·cm	3×10^{14} cm^{-3}	0	Produced in 2021
s1.2					
s31					
s7	Epitaxial	∼5 k Ω·cm	1×10^{12} cm^{-3} – 1×10^{13} cm^{-3}	50 µm	Epitaxial layer on standard substrate
s34					
w7				30 µm	Epitaxial layer on standard substrate; produced 2024
w8					
w6.1				20 µm	
w6.2					

The following sections detail the electrical and functional characterisation of these devices:

- Section 11.4 presents dark current measurements, X-ray response, and breakdown voltage analysis.
- Section 11.5 reports temperature-dependent IV curves obtained in a climate chamber.
- Section 11.7 shows experimental results with ^{90}Sr, assessing charge collection, depletion depth, and rise time characteristic against TCAD predictions.
- Section 11.8 describes tests, which were performed with a laser setup to investigate deeper depletion regions, and the time resolution of the chip.

11.4 Sensor current and breakdown investigation

High leakage currents contribute to the overall power consumption of the detector and can introduce noise that degrades signal quality, and even lead to thermal runaway and chip destruction. The breakdown voltage of the sensor defines the safe operational range, limiting the maximum reverse bias that can be applied without causing irreversible damage or an uncontrolled increase in current.

11.4.1 Dark current measurements

Dark current measurements were performed for all available BeMiPix samples. The sensors were tested in an X-ray shielding housing to ensure that no external illumination or radiation affected the results. The current was recorded as a function of the applied bias voltage up to the breakdown point. Characterisation was carried out on sensors implemented on both standard substrates and epitaxial layers of varying thickness. Results for the standard substrate and the 50 µm epitaxial version were presented in [133], while newer measurements for 30 µm and 20 µm epitaxial layers were added during the most recent evaluation campaign.

The measured breakdown voltage for all tested sensors falls within the range of −70 V to −90 V. At an operational bias of −50 V, the dark current typically ranges from 4 nA to 10 nA for both substrate types. One sample, w6.2, exhibited a significantly elevated current, approximately two orders of magnitude higher, which is attributed to a suspected production or wiring defect. An overview of the dark current and breakdown voltages for the measured samples is presented in Table 11.2.

Table 11.2: Dark current and breakdown voltage of samples with different substrates and epitaxial layer thicknesses

Sample name	Substrate type	Resistivity	Epitaxial thickness	Dark current	Breakdown voltage
s1.1	Standard	50 Ω·cm	0	7.12×10^{-9} A	-80 V
s1.2				9.86×10^{-9} A	-78 V
s7	Epitaxial	~5 k Ω·cm	50 µm	5.01×10^{-9} A	-77 V
w7			30 µm	5.02×10^{-9} A	-86 V
w8				4.80×10^{-9} A	-88 V
w6.1			20 µm	4.17×10^{-9} A	-76 V
w6.2				4.81×10^{-7} A	-79 V

11.4.2 Sensor current evaluation with X-ray photons

To investigate the sensor's performance, it was exposed to a steady X-ray beam. The setup is shown in Figure 11.8. The PCB was mounted on a movable stage, positioned perpendicular to the X-ray tube window, allowing uniform exposure of the sensor area.

X-ray irradiation measurements were performed using the tungsten X-ray tube operated at the irradiation facility of the Institute of Experimental Particle Physics (ETP)[4], Karlsruhe [8, 134].

[4] https://www.etp.kit.edu/english/irradiation_center.php

The tube is enclosed in a protective housing made of an aluminium–lead sandwich to shield the environment and ensure operator safety. It produces a broad X-ray spectrum, which can be converted into a monochromatic beam by targeting a single-element material that emits characteristic X-ray radiation. A vanadium filter is used to modulate the energy spectrum of the emitted X-rays. The acceleration voltage during these measurements was 60 kV. The X-ray tube current is adjustable between 2 and 60 mA.

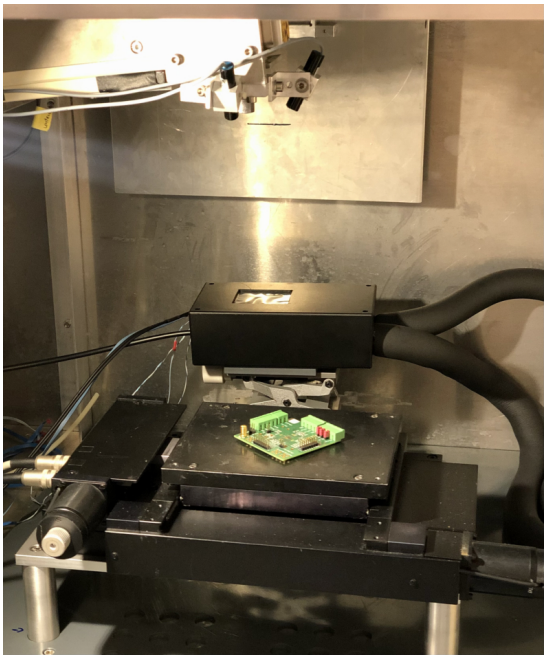


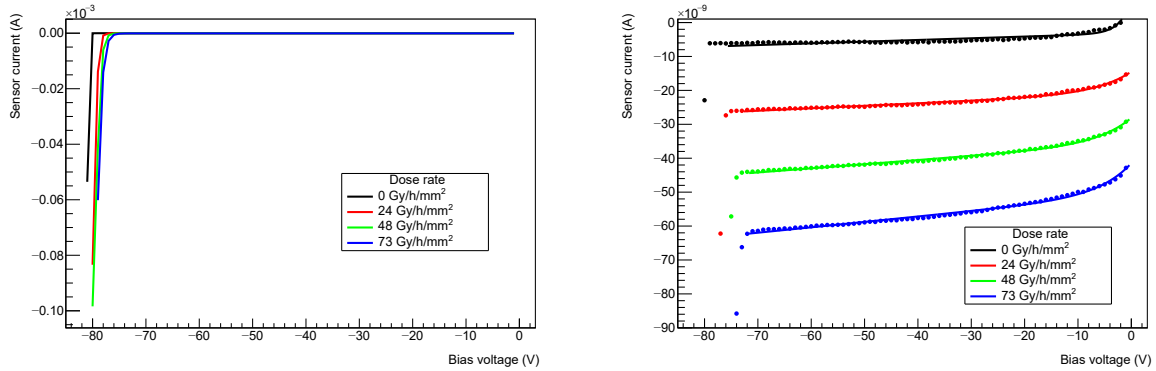
Figure 11.8: X-ray measurement setup at the ETP. The BeMiPix PCB is positioned perpendicular to the tube window for uniform exposure.

Figures 11.9 – 11.12 show the IV characteristics of four BeMiPix sensors fabricated on different substrates. Each chip was tested at an operating voltage up to V_{max} = 60 V, under varying X-ray tube currents of 0, 10, 20, and 30 mA. These correspond to dose rates of 0 Gy/h/mm^2, 24 Gy/h/mm^2, 48 Gy/h/mm^2, and 73 Gy/h/mm^2 respectively, as calculated for a beam spot of 6 mm diameter, at a sample distance of 300 mm [135].

The left-hand plots 11.9a, 11.10a, 11.11a, and 11.12a present the full IV curves for each substrate. The current rises sharply after the breakdown, obscuring the differences between dose rates in the main plots. However, small shifts in the breakdown voltage are observable, with higher X-ray current slightly increasing the breakdown point.

The corresponding right-hand plots, shown on Figures 11.9b, 11.10b, 11.11b and 11.12b, provide a zoomed view of the pre-breakdown region. In this view, the effect of X-ray irradiation on sensor leakage current becomes distinguishable. For all substrate types, the leakage current increases with dose rate, indicating dose-dependent bulk or interface activation effects.

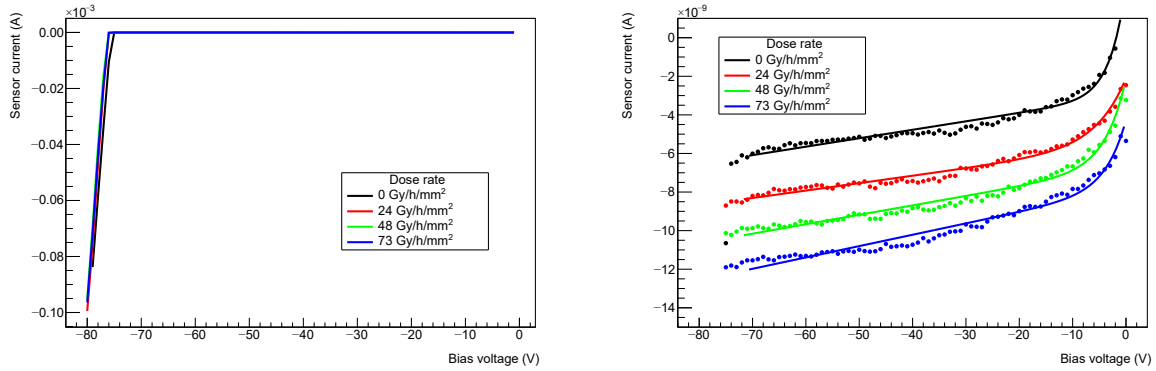
A summary of measured leakage currents and breakdown voltages is presented in Table 11.3. Leakage current values were calculated as the average current in the range between −45 V and −55 V. Breakdown voltage was defined as the voltage at which the current exceeded



(a) Full IV characteristic showing breakdown behaviour

(b) Zoom on the pre-breakdown region

Figure 11.9: IV characteristics at different X-ray dose rates for Sample s1.2 with standard substrate

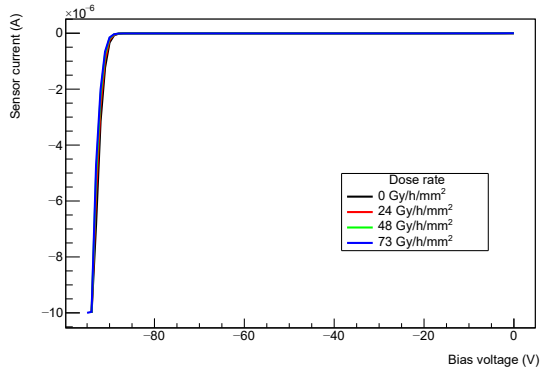


(a) Full IV characteristic showing breakdown behaviour

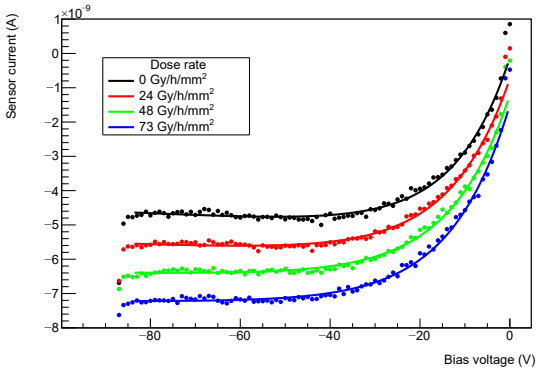
(b) Zoom on the pre-breakdown region

Figure 11.10: IV characteristics at different X-ray dose rates for Sample s7 with 50 µm epitaxial layer on standard substrate

150% of the average leakage current measured at −50 V. The table summarises the sensor leakage current and breakdown voltage for all samples under increasing X-ray dose rate. As expected, all sensors exhibit an increase in leakage current with increasing dose rate. The highest currents are observed for the Sample s1.2 without an epitaxial layer, ranging from approximately 5.77 nA to 58.9 nA across the tested range. In contrast, samples featuring an epitaxial layer on top of the standard substrate show lower leakage currents, with values improving as the epitaxial layer thickness decreases. The lowest and most stable current is recorded for the Sample w6.1 with a 20 µm epitaxial layer, remaining close to 4 nA at all dose levels. Interestingly, the thinner the epitaxial layer, the less variation is observed in the sensor current with increasing irradiation. For w6.1, the current remains nearly constant, suggesting that the sensor becomes fully depleted more rapidly and that the active volume is less affected by X-ray-induced generation. In contrast to the leakage current, breakdown

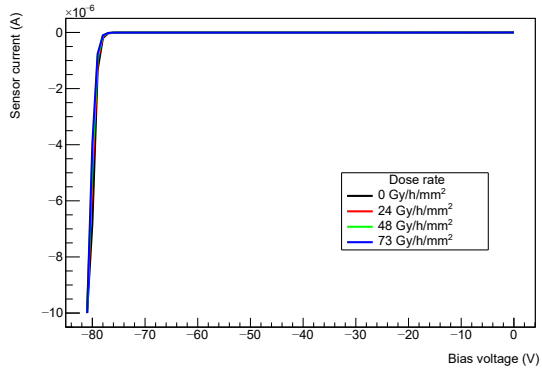


(a) Full IV characteristic showing breakdown behaviour

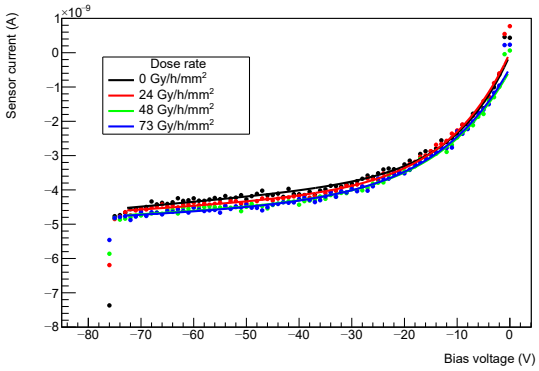


(b) Zoom on the pre-breakdown region

Figure 11.11: IV characteristics at different X-ray dose rates for Sample w8 with 30 µm epitaxial layer on standard substrate



(a) Full IV characteristic showing breakdown behaviour



(b) Zoom on the pre-breakdown region

Figure 11.12: IV characteristics at different X-ray dose rates for Sample w6.1 with 20 µm epitaxial layer on standard substrate

voltages show little variation across dose rates and appear to be more dependent on the sample structure than on irradiation. The measured breakdown values remain in the range of −70 V to −88 V, with the lowest observed in the standard substrate and the highest in Sample w8. This behaviour can be attributed to the depleted volume and the intrinsic properties of the epitaxial layer. Epitaxial silicon typically exhibits fewer defects and a more uniform doping profile, which leads to lower generation currents. The improved performance of the 20 µm sample is consistent with faster depletion and reduced X-ray sensitivity of the active region. This hypothesis is further explored in Section 11.7, where the sensor response to ^{90}Sr is analysed in more detail.

Table 11.3: Sensor current and breakdown voltage under increasing X-ray dose rate for standard and epitaxial substrates

Substrate type		**Standard**		**Epitaxial**					
Resistivity		50 Ω·cm		~5 kΩ·cm					
Epitaxial thickness		0		50 µm		30 µm		20 µm	
Sample name		s1.2		s7		w8		w6.1	
Results		**I, nA**	**V_{BR}, V**	**I, nA**	**V_{BR}, V**	**I, nA**	**V_{BR}, V**	**I, nA**	**V_{BR}, V**
X-ray dose rate, Gy/h/mm	0	5.77	-80	5.20	-75	4.79	-88	4.17	-76
	24	24.67	-77	7.61	-76	5.63	-88	4.36	-77
	48	41.86	-76	9.30	-76	6.34	-88	4.50	-77
	73	58.97	-75	11.04	-76	7.18	-88	4.51	-77

11.5 Temperature dependence measurements

This section presents current–voltage (IV) measurements performed under controlled temperature conditions using a climate chamber. Measurements were conducted with the sensor biased and all on-chip electronics powered, thereby emulating realistic operational conditions of the complete detector. Temperature variation plays a critical role in the behaviour of semiconductor devices, affecting both electrical characteristics, such as leakage current and breakdown voltage, and mechanical integrity. Although mechanical effects are not directly investigated in this study, they are important to acknowledge, particularly for future sensor integration and packaging.

From a mechanical perspective, rapid temperature changes can lead to thermal shock, inducing stress within the sensor assembly. This is especially relevant in hybrid or flip chip systems, where materials with different coefficients of thermal expansion (CTE) are joined. For instance, silicon has a relatively low CTE of $2.6–3.3 \times 10^{-6}$ K^{-1}, while metals such as aluminium exhibit much higher values about 23×10^{-6} K^{-1}. This mismatch can lead to delamination, wire bond failure, or micro-cracking in the die. Thin gold or aluminium bonding wires are particularly vulnerable to fatigue from repeated thermal cycling. Similarly, stresses at chip interfaces may accumulate over time, degrading reliability.

In the case of the SiGe BiCMOS technology used in this work, additional considerations arise due to the inclusion of SiGe layers. These layers introduce further CTE mismatches and mechanical complexity. The thermal-mechanical properties of SiGe depend strongly on the germanium content, which typically ranges from 10% to 30% in HBTs and sensor structures. These differences can lead to dislocation formation at the Si/SiGe interface, delamination during rapid thermal shifts, or mechanical instability in 3D or multilayer structures [126].

Although the primary aim of these measurements is to assess electrical performance, these thermal–mechanical considerations provide essential context for interpreting the observed

leakage behaviour and ensuring long-term sensor robustness under fluctuating environmental conditions.

11.5.1 Thermal generation of charge carriers and leakage current

Temperature has a significant impact on leakage currents in semiconductor devices, primarily through the thermally assisted generation of charge carriers. At elevated temperatures, increased lattice vibrations and bandgap narrowing lead to the enhanced generation of electron–hole pairs. As a result, the resistivity of the material decreases and its conductivity rises. This thermal generation increases the sensor's leakage current, which may contribute to higher noise levels and potentially mask low-energy events, particularly critical in applications involving long integration times or low signal thresholds.

The temperature dependence of leakage current in a $p-n$ junction can be understood via the ideal diode equation, also known as the Schockley equation, which describes the total current density J as the sum of hole and electron contributions:

$$J = J_p + J_n = J_0 \left(\exp\left(\frac{qV}{kT}\right) - 1 \right) \tag{11.3}$$

where J_p and J_n represent the hole and electron current components, respectively, in units of A/cm^2. The saturation current density J_0 is defined as:

$$J_0 \equiv \frac{qD_p p_{n0}}{L_p} + \frac{qD_n n_{p0}}{L_n} \equiv \frac{qD_p n_i^2}{L_p N_D} + \frac{qD_n n_i^2}{L_n N_A} \tag{11.4}$$

with the following definitions:

- D_p, D_n — diffusion coefficients for holes and electrons, cm^2/s;
- L_p, L_n — diffusion lengths for holes and electrons, cm;
- p_{n0} — equilibrium minority (holes) concentration in an $n-$type semiconductor, cm^{-3};
- n_{p0} — equilibrium minority (electrons) concentration in a $p-$type semiconductor, cm^{-3};
- n_i — intrinsic carrier concentration, cm^{-3};
- N_D, N_A — donor and acceptor impurity concentrations, cm^{-3}.

Equation 11.4 is strongly temperature-dependent. To examine this, we focus on the first term, as the second behaves similarly and can be neglected in the case of a one-sided abrupt p^+-n junction whereN_D, $p_{n0} \gg n_{p0}$. The key quantities, such as n_i, $D_{p,n}$, p_{n0}, n_{p0}, and $L_{p,n}$, are all temperature-dependent. Given that $L_p \equiv \sqrt{D_p \tau_p}$ and assuming $D_p/\tau_p \propto T^\gamma$, where γ is a constant, the saturation current density scales approximately as:

$$J_0 \approx \frac{qD_p p_{n0}}{L_p} \approx q\sqrt{\frac{D_p}{\tau_p}}\frac{n_i^2}{N_D} \propto T^{\gamma/2}\left[T^3 \exp\left(-\frac{E_g}{kT}\right)\right] \propto T^{(3+\gamma/2)} \exp\left(-\frac{E_g}{kT}\right) \tag{11.5}$$

where E_g is the semiconductor bandgap energy. The polynomial pre-factor $T^{(3+\gamma/2)}$ has a much weaker influence than the exponential term. Thus, the dominant temperature dependence is governed by the exponential $\exp(-E_g/kT)$ term. As a result, the reverse-bias current, where $|J_R| \approx J_0$, is expected to increase with temperature approximately as $\exp(-E_g/kT)$. In the forward direction, the current follows $J_F \approx J_0 \exp(qV/kT)$ and thus scales as $\exp[-(E_g - qV)/kT]$ [68]. In reverse-biased $p-n$ junctions and depleted pixel sensors, the leakage current is thus dominated by thermally generated carriers, particularly those originating in high-field regions or deep depletion zones.

Silicon–germanium alloys exhibit a narrower bandgap than pure silicon, typically ranging from $E_g \approx 0.9$ to 1.05 eV depending on the Ge content (10–30%). This reduction in bandgap leads to a higher intrinsic carrier concentration, which in turn results in increased leakage current under identical bias and thermal conditions. While this effect is material-dependent, it is difficult to isolate the contribution of SiGe to leakage current in complex sensor assemblies such as the BeMiPix chip.

More generally, the inclusion of SiGe HBTs within the BiCMOS process may alter the temperature sensitivity of the device. SiGe HBTs are known to perform reliably even at cryogenic temperatures, where conventional Si BJTs typically fail, due to the beneficial effects of bandgap engineering on low-temperature carrier transport [126]. In this study, we do not investigate cryogenic behaviour, but rather focus on an extended operating range from -40 °C to +30 °C. This range is relevant both for expected usage environments and for evaluating robustness under thermal cycling conditions.

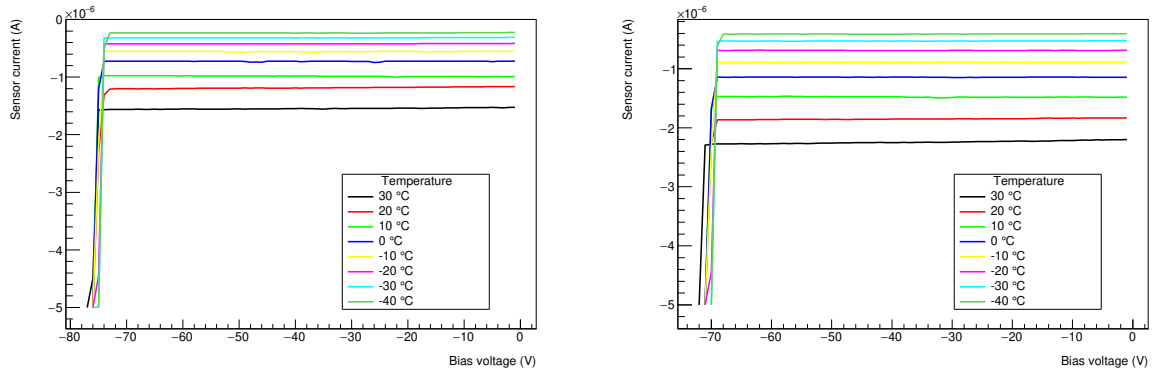
Although we cannot isolate the material-level contributions of SiGe versus Si, what can be investigated experimentally is the effect of the epitaxial layer. A comparative analysis of sensors with and without high-resistivity epitaxial silicon allows us to assess the structural impact on temperature-dependent leakage behaviour. Our measurements indicate that the presence of an epitaxial layer does not significantly increase the absolute leakage current.

11.5.2 Temperature-dependent IV characteristics

Temperature-dependent IV measurements were performed in a climate chamber for two BeMiPix samples: one fabricated on a standard substrate (Sample s31), and one on a substrate with a 50 µm epitaxial layer (Sample s34). The measurements were taken in steps from -40 °C to +30 °C, with the full detector powered on, including sensor bias and active front-end, thus closely emulating real operating conditions.

Figure 11.13 shows the IV curves of both samples across the temperature range. The overall behaviour is as expected: the leakage current increases with temperature due to thermally generated charge carriers.

To visualise the leakage trend more clearly, the average sensor current at each temperature is plotted in Figure 11.14. The sample with the epitaxial layer consistently shows higher leakage current across the entire temperature range and exhibits a steeper increase with temperature.



(a) IV curve of Sample s31 (standard substrate) **(b)** IV curve of Sample s34 (50 µm epitaxial layer)

Figure 11.13: Temperature-dependent IV characteristics for BeMiPix samples with standard and epitaxial substrates. Measurements were performed with full electronics active

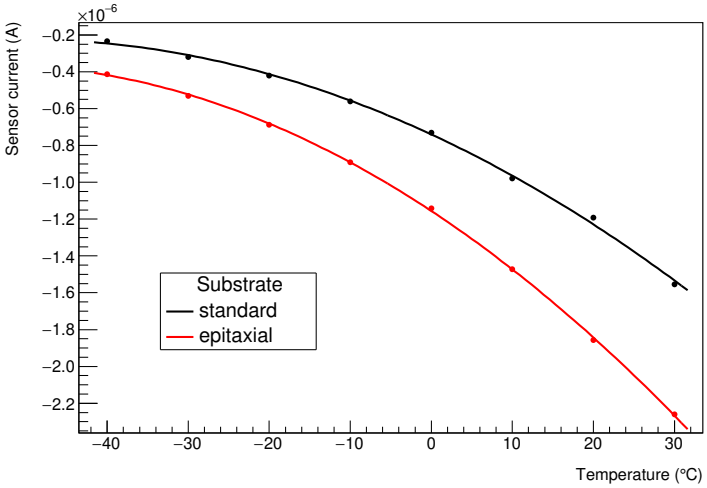


Figure 11.14: Temperature dependence of average leakage current for Samples s31 (standard) and s34 (epitaxial)

Result: The temperature-dependent IV analysis reveals two main findings. First, both samples follow the expected exponential increase in leakage current with rising temperature, consistent with thermally generated carrier behaviour. Second, the sample with a 50 µm epitaxial layer exhibits both higher leakage current and greater temperature sensitivity. This suggests that, under operational conditions, epitaxial structures may require more careful thermal management, especially in systems with prolonged exposure to elevated temperatures. The results also confirm that powering the full chip introduces additional leakage compared to passive sensor measurements, further reinforcing the importance of realistic operating conditions in performance evaluation.

11.6 Waveform analysis procedure

Before presenting the results of the source measurements in Section 11.7 and laser characterisation in Section 11.8, it is important to outline the general waveform analysis procedure applied to both datasets. Although the specific analysis goals differ between measurement types, the processing of waveform data follows a common methodology. In both cases, the detector response is recorded in the form of digitised waveforms, which are processed offline using the same framework. The first stage involves unpacking the waveform files and extracting key signal characteristics. A representative waveform, illustrating all relevant parameters used in the analysis, is shown in Figure 11.15.

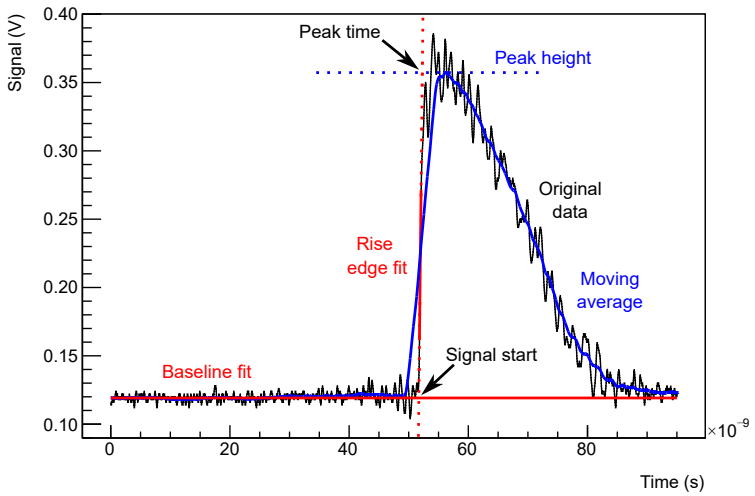


Figure 11.15: Example waveform showing the analysis procedure. The original signal (black) is smoothed with a moving average (blue). The baseline is estimated from the first third of the waveform (red line). A linear fit around the inflexion point defines the rising edge, with the signal start time given by its intersection with the baseline. The peak is taken at 95% of the signal height. The rise time is defined as the time between the signal start and the peak time reference.

For each measurement point, multiple waveforms are recorded to get a statistical average. In every waveform (black), the baseline voltage is first calculated using the first third of the recorded data points (red). To reduce noise, a moving average fit (blue) is applied to the original data. The smoothed waveform is then used to determine the signal height, which is defined as the difference between the peak value of the smoothed signal, referred to as the peak height, and the value of the baseline signal. In the example shown, this corresponds to a signal height of 0.24 V, calculated from a peak height value of 0.36 V and a baseline of 0.12 V.

The next step in the analysis is finding the inflexion point in the smoothed waveform. It is done by differentiating the smoothed waveform and identifying the maximum of this derivative. The x-coordinate of this maximum corresponds to the inflexion point of the smoothed waveform. The x-coordinate of the inflexion point is then used to locate the corresponding region on the unsmoothed original waveform. A linear fit is performed to a small region around this

inflexion point to determine the rising edge (red). The signal start time is defined as the intersection between this linear fit and the previously determined baseline. The signal peak time is determined as the point at which the waveform reaches 95% of the maximum amplitude. This threshold is chosen to improve robustness against saturation effects in the amplifier.

In time resolution studies, the extracted signal start time serves as the timing reference. The rise time is defined as the time between the signal start and the peak time. All measured quantities are averaged over the full set of waveforms recorded at a given measurement point. This waveform-based analysis method is applied consistently across all data in this study.

11.7 Source measurements with Iron-55 and Strontium-90

This section focuses on evaluating the signal response of the BeMiPix sensor to well-defined energy deposits generated by radioactive sources. The output of the analogue front-end is measured at the V_{out} node using an oscilloscope. Known signal shapes are induced by two types of radiation sources: soft X-rays from ^{55}Fe and electrons from ^{90}Sr. These measurements serve to characterise the sensor response across different substrate types and operating conditions.

The detection principle is based on charge collection from the substrate. Upon energy deposition, electron–hole pairs are generated in the sensitive volume. An electric field separates the charge carriers, and the collected electrons induce a voltage signal at the input of the front-end amplifier. The HBT-based amplifier converts this input voltage into an output current signal through its transconductance, functioning as a transimpedance stage. The voltage output is then monitored by an oscilloscope. For each triggered event, the signal waveform is recorded, and the peak voltage is extracted and filled into a histogram. This allows reconstruction of the most probable signal amplitude for each measurement condition.

Measurements were performed using four sensor types, differing by substrate structure: standard substrate with 50 Ω·cm resistivity, and three sensors with epitaxial layers of 50 µm, 30 µm, and 20 µm thickness grown on the same base. The objective is to compare charge collection performance and signal shape characteristics across these configurations.

The sensor supply voltages and bias settings for these measurements were as follows:

- V_{nwell} = 1.2 V
- V_{pnwell} = 0.99 V
- V_{DD} = 1.2 V
- V_{casc} = 1.2 V
- $V_{emitter}$ = 0.14 V

The setup is shown in Figure 11.16. The front-end signal is read out via coaxial connection and analysed directly on a high-bandwidth oscilloscope. Stable operation required fine-tuning of the bias voltages. The chip was sensitive to environmental fluctuations and often required adjustment before each data run.

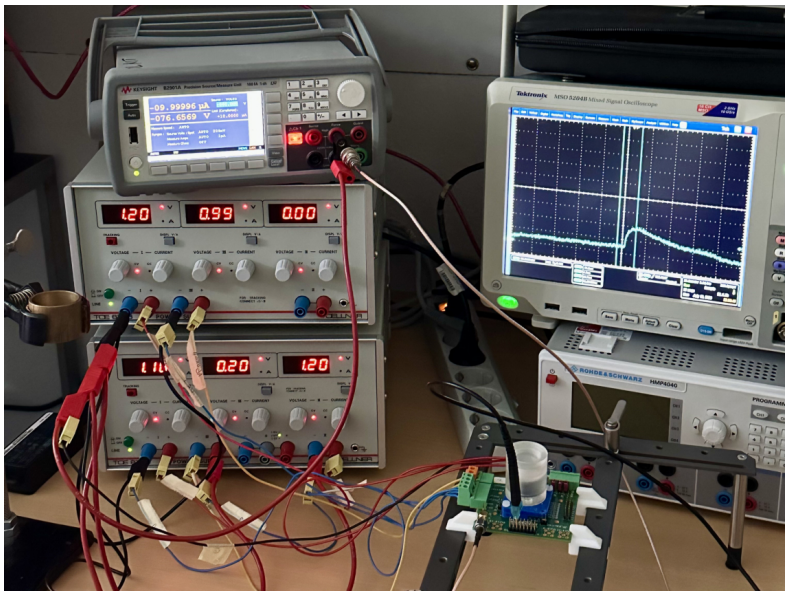


Figure 11.16: Photograph of the setup for radioactive source measurements. The signal is read out via the V_{out} line and observed using an oscilloscope. The power supplies set all bias voltages for analogue front-end operation.

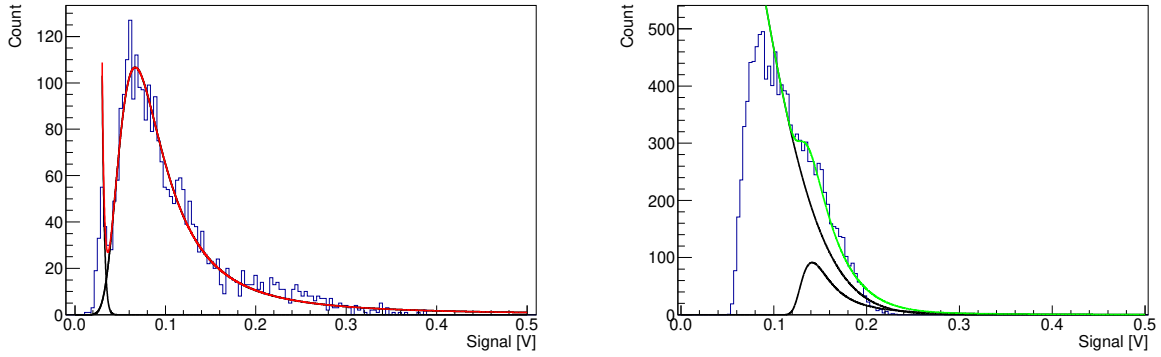
11.7.1 Charge collection performance and depletion depth evaluation

The following measurements focus on the sensor's performance when irradiated with ^{90}Sr. ^{90}Sr is a β^- emitter with decay energy of approximately 0.546 MeV distributed to an electron, an antineutrino, and the yttrium isotope ^{90}Y. These emitted electrons deposit charge across the full sensor thickness, making the isotope an ideal probe for assessing depletion depth and charge collection efficiency.

As described in Section 11.6, waveforms are recorded and processed to extract peak amplitudes. These amplitudes are then histogrammed to form a spectrum. Figure 11.17 shows representative fits for two sensor configurations.

In the ideal case, shown in Figure 11.17a, the noise part is clearly visible and well-separated from the signal peak. The fit proceeds in three stages: first, the noise peak is fit using a Gaussian function and subtracted from the spectrum (black line left); then, a Landau–Gaussian convolution is fitted to the remaining distribution (black line right); finally, both components are combined into a full fit across the data range (red line). For lower signal-to-noise conditions, shown as an example in Figure 11.17b, the noise peak merges with the signal tail, but the same fitting procedure remains valid. The green line in this case represents the combined fit function.

The most probable value (MPV) of the Landau distribution is used as a measure of charge collection performance. It corresponds to the maximum of the Landau–Gaussian peak. Figure 11.18 shows the MPV as a function of applied reverse bias for substrates with epitaxial layers of 20 µm, 30 µm, and 50 µm thickness.



(a) Fitted histogram for Sample w6, $V_{bias} = -80$ V **(b)** Fitted histogram for Sample s7, $V_{bias} = -50$ V

Figure 11.17: Signal histograms from ^{90}Sr measurements with fitted Landau–Gaussian convolutions

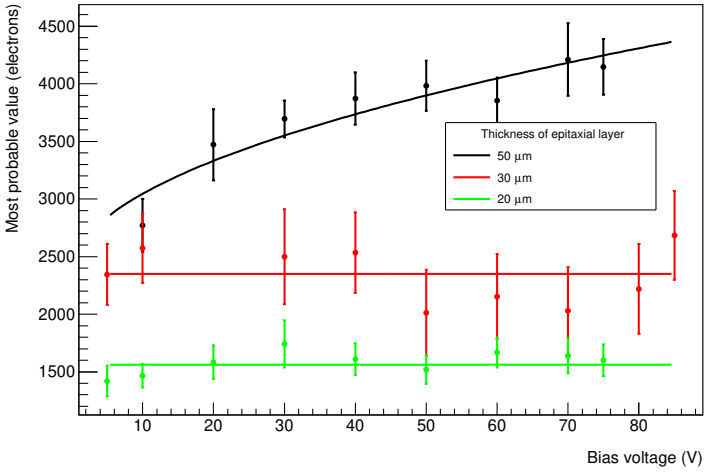
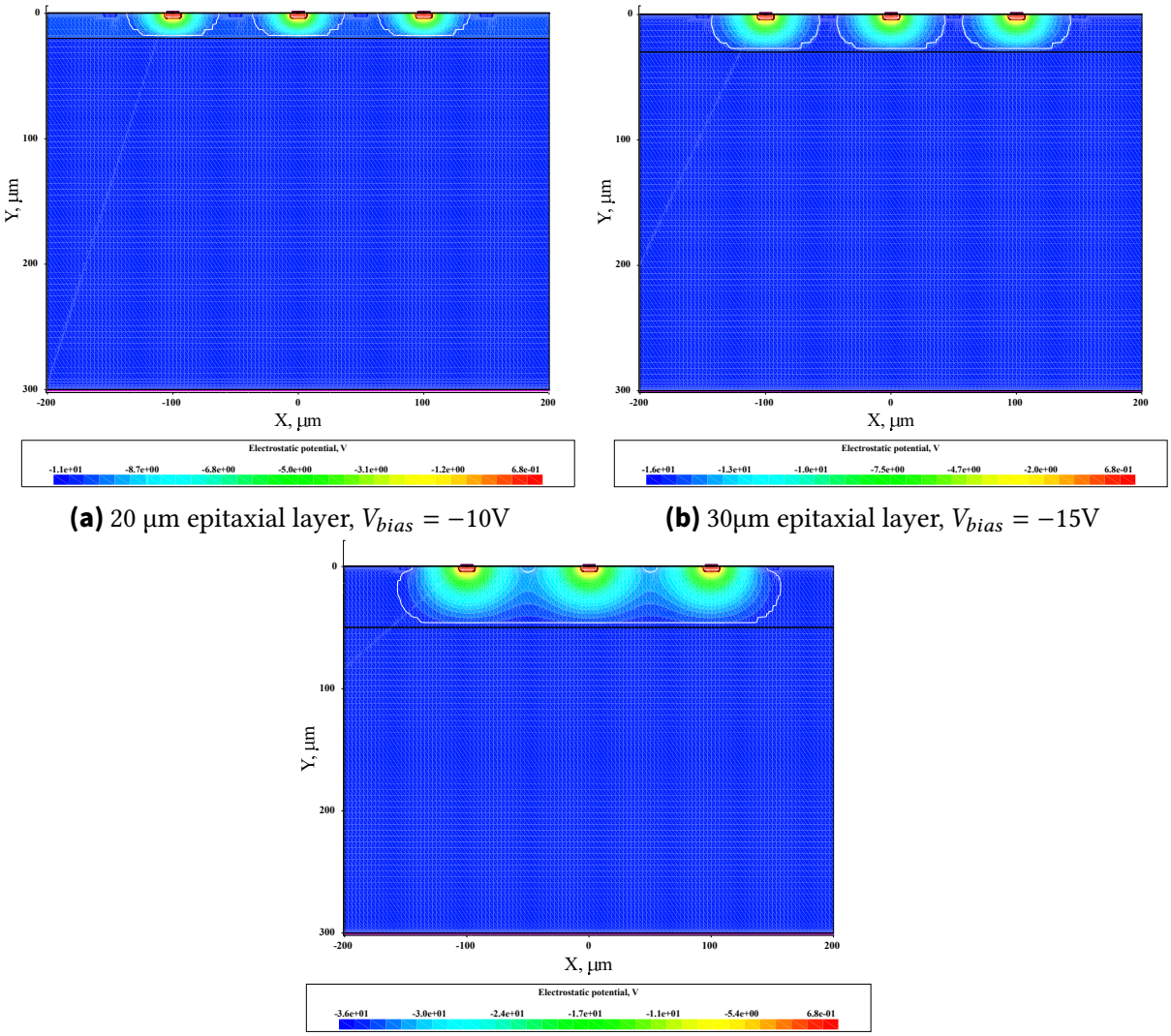


Figure 11.18: MPV versus bias voltage for different substrate configurations

From the MPV curves, the full depletion voltage of each sample can be concluded. These measurements are cross-validated with TCAD simulations, where the electrostatic potential distribution and depletion region are determined. Figure 11.19 illustrates the simulated electrostatic potential distribution for sensors with 20 µm, 30 µm, and 50 µm epitaxial layers at voltages matching the voltages for full depleted state based on MVP saturation.

The simulation confirms that for the 50 µm-thick epitaxial layer, the depletion regions merge at -35 V. For the 30 µm case, full depletion in the vertical direction is achieved by -15 V; further bias extension leads primarily to lateral growth. For the 20 µm epitaxial layer, the critical depletion voltage is approximately -10 V. These results are consistent with the MPV saturation behaviour observed in the measurements.



(a) 20 µm epitaxial layer, $V_{bias} = -10$V

(b) 30µm epitaxial layer, $V_{bias} = -15$V

(c) 50µm epitaxial layer, $V_{bias} = -35$ V

Figure 11.19: Simulated electrostatic potential maps showing depletion progress in sensors with different epitaxial thicknesses. The black line marks the epi–bulk interface.

11.7.2 Hit rate evaluation

The hit rate measured with the ^{90}Sr source for each substrate is presented in Figure 11.20. It is important to note that the values across different substrates are not directly comparable, as the source position was adjusted individually for each set of measurements. Therefore, only the trend of the hit rate as a function of bias voltage should be considered for each substrate individually.

For all samples, the hit rate exhibits a square-root-like dependence on the applied voltage. This behaviour can be attributed to the geometry of the depletion region growth in these sensors. Initially, the depleted volume grows spherically from the small collection node. However, due to the limited thickness of the epitaxial layer, full vertical depletion is quickly reached. Beyond

that point, the depletion region expands primarily laterally, resulting in a more cylindrical shape. This also explains the early onset of full depletion in depth and the continued growth in the lateral direction with increasing voltage.

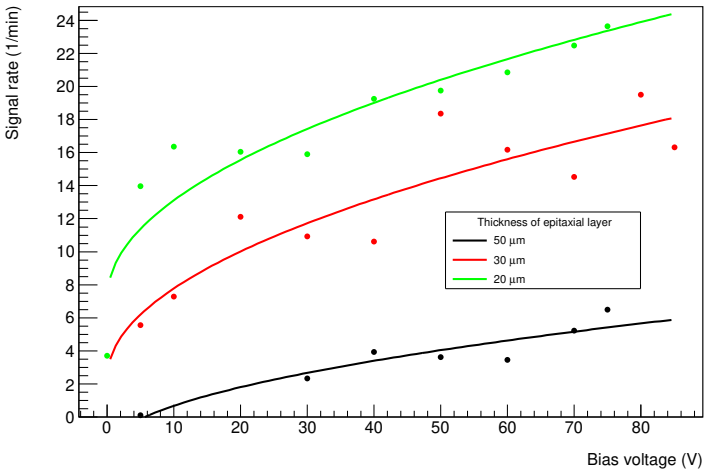


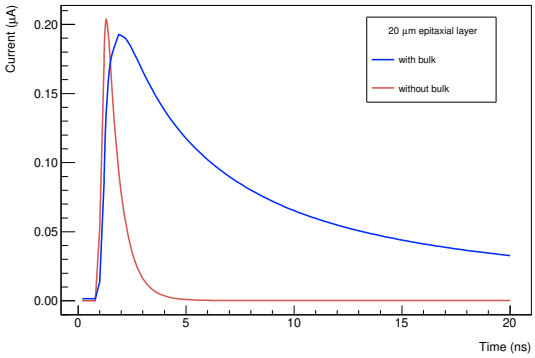
Figure 11.20: Hit rate as a function of bias voltage for different epitaxial layer thicknesses under ^{90}Sr irradiation. Values are normalised per measurement set; trends within each substrate group are to be taken into account, but absolute comparisons between substrates are not.

11.7.3 Rise time characterisation

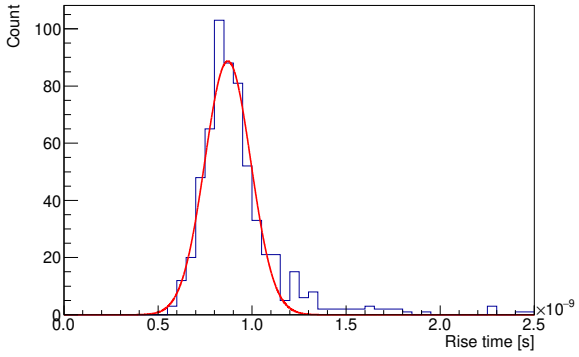
An important parameter that limits the time resolution of the sensor is the rise time of the front-end electronics. In this subsection, the measured rise time values obtained from ^{90}Sr source tests are compared with TCAD simulations of the transient response for different substrate thicknesses. While the measurement procedure follows the same waveform analysis approach described in Section 11.6, additional precision was achieved by increasing the number of recorded points per voltage step. In the simulation framework, the rise behaviour is studied through current–time I(t) plots resulting from a simulated MIP event. These simulations are discussed in more detail in Section 11.1.4. Here, two configurations were used per substrate type: one with only the epitaxial layer, and another including the bulk substrate. This comparison helps to analyse the difference between the fast drift component and the slower diffusion contribution to charge collection. All tests and simulations were performed in the voltage range from 0 to −75 V.

Figures 11.21 to 11.23 show representative comparisons between simulation results (left) and experimental rise time histograms (right). Each simulation plot presents two curves: one for the epitaxial layer only (red), and one including the 50 Ω·cm bulk substrate (blue). The corresponding experimental results were obtained for three sensor configurations: 20 µm (Sample w6.1), 30 µm (Sample w8), and 50 µm (Sample s7).

In the case of the 20 µm-thick epitaxial sensor on low-resistivity bulk, shown in Figure 11.21, the recorded current reaches around 0.19 µA, with the version excluding the bulk showing

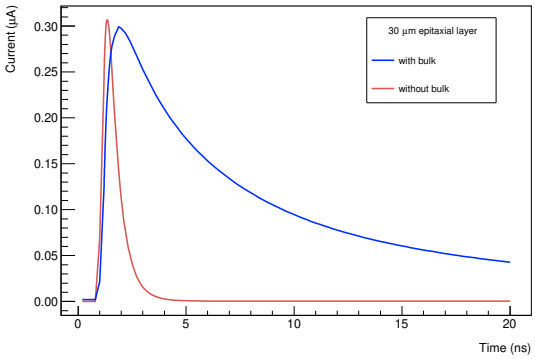


(a) Simulated current–time behaviour, 20 µm epitaxial layer (with and without bulk)

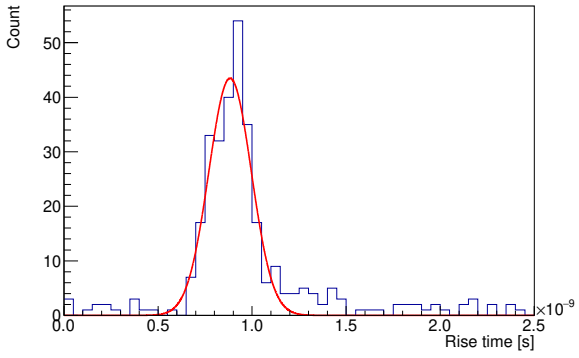


(b) Measured rise time distribution, Sample w6.1 (20 µm), −40 V

Figure 11.21: Comparison of simulated and measured rise time characteristics for the 20 µm epitaxial substrate



(a) Simulated current–time behaviour, 30 µm epitaxial layer (with and without bulk)

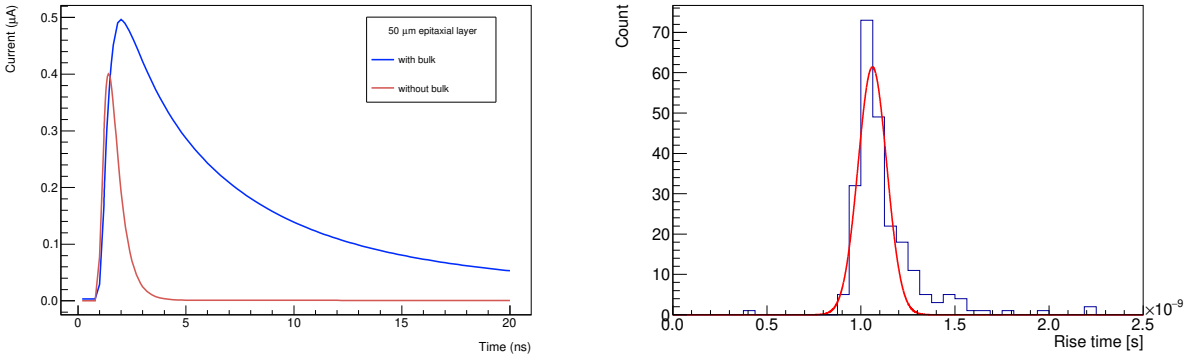


(b) Measured rise time distribution, Sample w8 (30 µm), −60 V

Figure 11.22: Comparison of simulated and measured rise time characteristics for the 30 µm epitaxial substrate

slightly faster signal formation. For the 30 µm-thick epitaxial sensor, shown in Figure 11.22, the current increases to approximately 0.3 µA, with the rising edge of both curves overlapping more closely. In the 50 µm-thick epitaxial sensor, shown in Figure 11.23, the inclusion of the bulk noticeably increases both the current amplitude and rise slope, reaching 0.5 µA. These observations confirm the expected contribution of diffusion-based charge collection in thicker substrates.

The rise time values extracted from both the simulations and the waveform histograms are summarised in Table 11.4. The experimental distributions, shown on the right side of each figure, confirm the simulated trends and demonstrate a strong correlation between substrate thickness and signal formation time.



(a) Simulated current–time behaviour, 50 µm epitaxial layer (with and without bulk)

(b) Measured rise time distribution, Sample s7 (50 µm), −50 V

Figure 11.23: Comparison of simulated and measured rise time characteristics for the 50 µm epitaxial substrate

Table 11.4: Comparison of simulated and experimental rise times for different epitaxial thicknesses

Sample name	Substrate thickness	Simulated rise time	Experimental rise time
w6.1	20 µm	0.83 ns	0.870 ns
w8	30 µm	0.84 ns	0.877 ns
s7	50 µm	1.04 ns	1.087 ns

The overall agreement between the simulation and experimental data validates the model and supports the hypothesis that faster charge collection is primarily driven by drift in thinner substrates. The influence of diffusion contributions becomes more evident at 50 µm, where the slower component increases the signal amplitude and the rise time.

11.7.4 Iron-55 calibration and charge response validation

To calibrate the response of the readout chain and estimate the number of collected electrons, measurements were carried out using X-rays from an Iron-55 source. ^{55}Fe decays by electron capture to ^{55}Mn, resulting in emission of Auger electrons (5.19 keV, ~60% probability), and characteristic X-rays: Kα_1 (5.89875 keV, 16.2%), Kα_2 (5.88765 keV, 8.2%), and Kβ (6.49045keV, 2.85%). Since the Kα_1 and Kα_2 lines are nearly identical in energy, the emission is commonly treated as monoenergetic 5.9 keV X-rays (combined ~28%).

When a 5.9 keV X-ray photon is fully absorbed in silicon, it generates approximately 1639 electron-hole pairs. Although the energy of the X-ray is well defined, the resulting signal exhibits a Gaussian distribution due to fluctuations in both charge generation and charge collection, as well as electronic noise from the front-end. A Gaussian fit is therefore applied to the signal histogram [8]. An example of such a histogram is shown in Figure 11.24 for Sample s7. Identical measurements and analyses were performed for Samples w8 and w6.1.

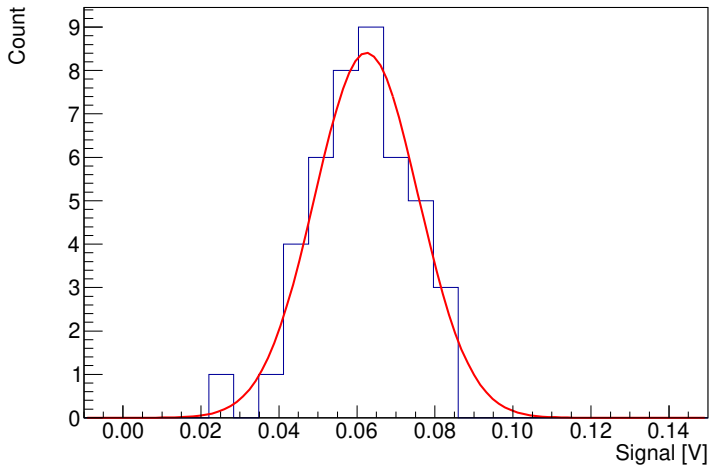


Figure 11.24: Histogram of ^{55}Fe X-ray response for Sample s7. A Gaussian fit is applied to determine the amplifier output.

The energy deposited by the 5.9 keV X-rays corresponds to:

$$\text{AmpIn}\left(^{55}\text{Fe}\right) = 1639\ e\text{-}h \text{ pairs} \tag{11.6}$$

For Sample S7, the corresponding amplifier output signal is:

$$\text{AmpOut}\left(^{55}\text{Fe}\right) = 0.06076\ \text{V} \tag{11.7}$$

Using the information about the signal from ^{55}Fe through the amplifier, we can estimate the number of electron-hole pairs generated in the ^{90}Sr measurement. For Sample s7 at 50 V, the signal MPV was:

$$\text{AmpOut}\left(^{90}\text{Sr}\right) = 0.1468\ \text{V} \tag{11.8}$$

which corresponds to:

$$\text{AmpIn}\left(^{90}\text{Sr}\right) = \frac{0.1468\ V}{0.06076\ V} \times 1639\ e\text{-}h \approx 3964\ e\text{-}h \text{ pairs} \tag{11.9}$$

For comparison, a minimum ionising particle (MIP) generates approximately 80 *e-h* pairs per micron in silicon. Thus, a 50 µm epitaxial layer corresponds to:

$$80 \frac{e\text{-}h}{\mu m} \times 50\ \mu m = 4000\ e\text{-}h \tag{11.10}$$

The result confirms a good agreement between the ^{90}Sr signal and the expected MIP response.

The same calculation was repeated for Samples w8 (30 µm) and w6.1 (20 µm), with results summarised in Table 11.5.

Table 11.5: Comparison of input/output amplitudes from ^{55}Fe and ^{90}Sr sources with MIP reference values

Sample name	Epitaxial thickness, µm	Iron-55		Strontium-90		MIP reference
		AmpIn	AmpOut, V	AmpIn	AmpOut, V	
s7	50	1639 *e-h*	0.061	3964 *e*	0.147	4000 *e-h*
w8	30	1639 *e-h*	0.034	2419 *e*	0.050	2400 *e-h*
w6.1	20	1639 *e-h*	0.036	1557 *e*	0.034	1600 *e-h*

In addition to calibration, ^{55}Fe measurements were used to verify the absence of charge multiplication at high voltages. Table 11.6 shows the fitted mean and standard deviation for different bias voltages. The gradual, nearly linear increase in signal amplitude excludes the onset of impact ionisation in the accessible voltage range.

Table 11.6: MPV and standard deviation of ^{55}Fe signal for different bias voltages

Bias voltage, V	MPV, V	Standard deviation, V
30	0.0583	0.0116
50	0.0612	0.0114
70	0.0621	0.0130
Average	0.0608	0.0124

11.8 Laser scanning measurements for time and spatial resolutions

To characterise the time and spatial resolution of the sensor, a series of focused laser measurements was performed. Two sensor variants were tested: one with a standard low-resistivity substrate (Sample s31), and one with a 50 µm epitaxial layer on top of the substrate (Sample s34). The primary focus of these measurements was on Sample s31, as it provides insight into the baseline performance of the substrate design used across all BeMiPix chips.

The setup is illustrated in Figure 11.25. A pulsed laser beam is directed onto the sensor through a motorised head, which enables precise movement in the x-, y-, and z-directions. The laser is focused to a spot size of approximately 1×1 µm^2 at a working distance of around 2 cm from the sensor surface. The entire system is enclosed within an aluminium shielding box to ensure eye safety and to reduce external light and electromagnetic interference. However, this enclosure also introduces mild thermal elevation during operation.

For each measurement, a specific pixel is selected for analogue readout. The output waveform is recorded by a digital oscilloscope and read out via computer for storage and offline analysis. Pulse heights are also monitored in real time through an online fitting routine to ensure the consistency and reliability of the scan.

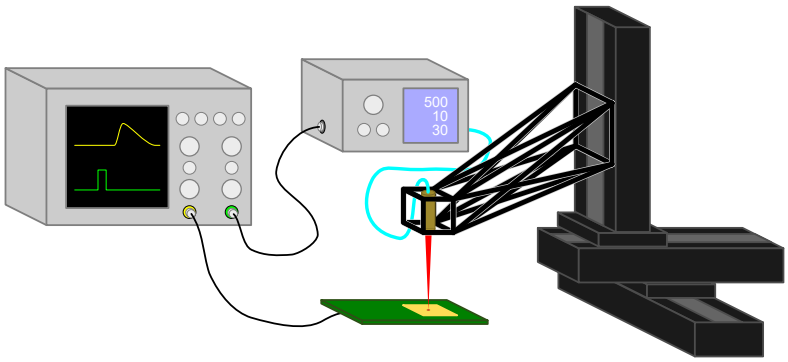


Figure 11.25: Schematic of the laser scanning setup. The pulsed laser beam is directed onto the chip through a movable head with sub-micron focus, while the analogue signal from a selected pixel is read out and digitised via an oscilloscope. The chip was illumitated through the back side.

11.8.1 Measurement procedure

The laser scanning measurement was carried out in two primary stages: (1) laser alignment and focusing, and (2) scanning to extract the spatial response map of the pixel. The main objective was to determine the lateral extension of the depletion region by analysing the sensitivity area of an individual pixel under pulsed laser excitation.

Step 1: Focus optimisation. Initially, the laser head was manually positioned to locate the approximate pixel position in both x- and y-directions by scanning the chip at full laser power and intentionally defocusing the beam. In this configuration, the scanning spot is quite broad to facilitate a coarse y-direction alignment. Once the signal region was identified, a fine scan in the x–z plane was performed at reduced power to find the point of best focus. It is defined as the z-position where the sensitive region in the y-direction is minimal. This point, corresponding to the minimal lateral spread of the laser spot, was found at a z-distance of 400 µm (see Figure 11.26).

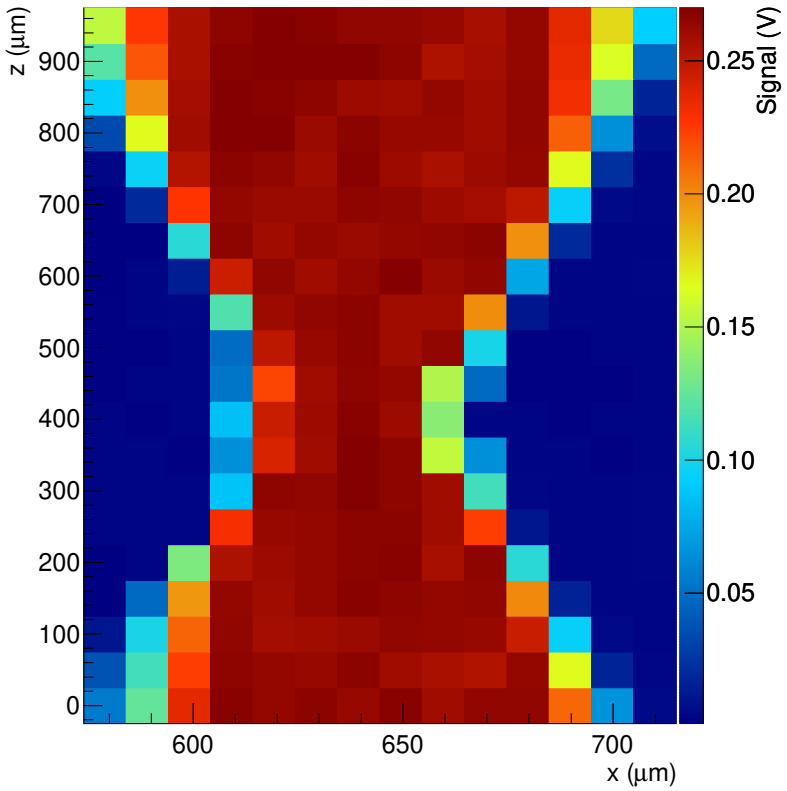
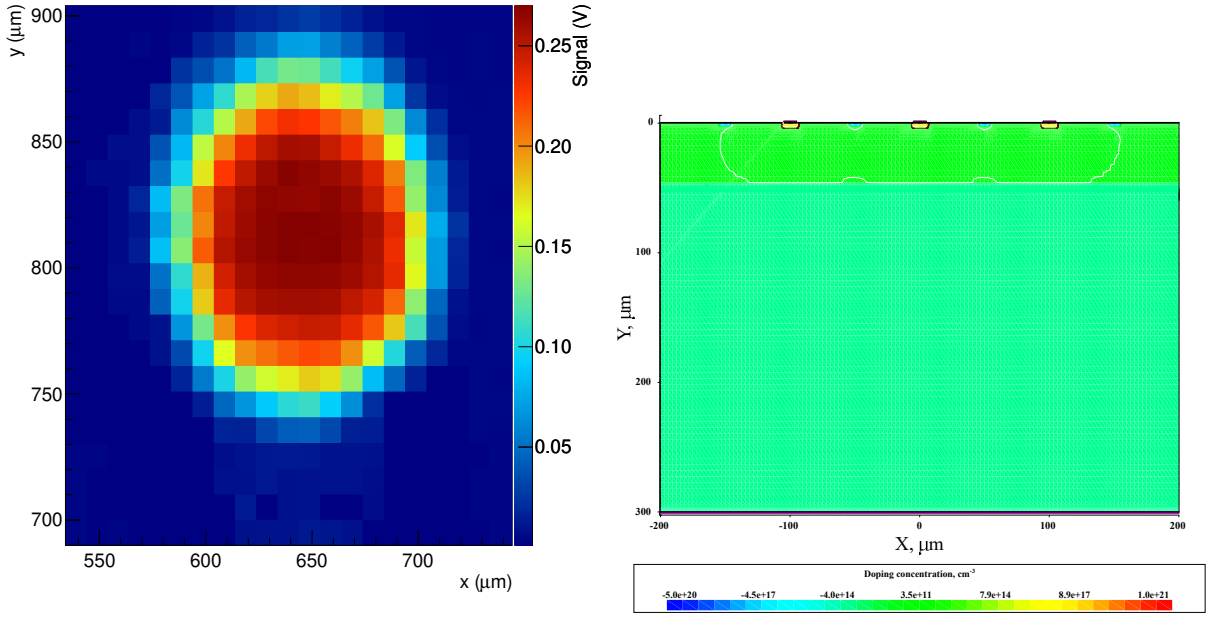


Figure 11.26: Laser beam width in the ZX-plane. The narrowest profile is observed at a z-position of 400 µm, corresponding to the optimal focus.

Step 2: Pixel area scan. With the laser focused, a 2D scan was carried out across the pixel in the x–y plane. The scan began with large step sizes to identify the active region and was followed by a fine scan (typically 10 µm steps) for accurate mapping. At each position, 20 waveforms were acquired from the analogue output using an oscilloscope, with the pulse amplitude being fitted in real-time for quick verification. The oscilloscope was operated in free-trigger auto mode.

An example of the measured sensitivity map is shown in Figure 11.27a, where the active response area of a single pixel is clearly visible. This measurement was conducted on a 50 µm epitaxial layer sample at −30 V. To validate the experimental result, a corresponding TCAD

simulation of the depletion region under the same bias condition is shown in Figure 11.27b. The strong agreement between measured and simulated data confirms the reliability of the model and supports its use for predicting behaviour at untested voltages or geometries. Both results will be summarised later in Table 11.7.



(a) Measured response map at −30 V with 10µm laser steps

(b) Simulated depletion region for 50 µm epitaxial sensor at −30 V

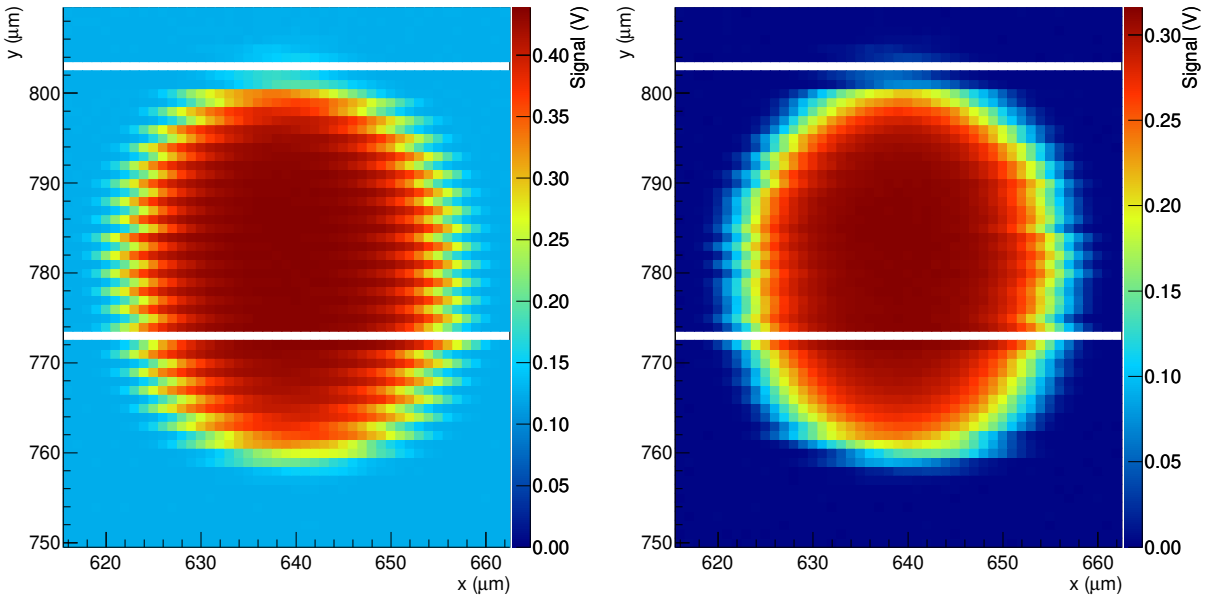
Figure 11.27: Comparison of experimental laser scan and TCAD simulation of depletion region

In addition to spatial resolution studies, the laser setup can also be used for timing measurements. By synchronising the laser driver trigger with the oscilloscope, the delay between the laser pulse and the pixel's analogue response can be extracted. Offline waveform processing followed the same procedure as described in Section 11.6.

11.8.2 Sensitivity analysis and depletion characterisation

To investigate the spatial sensitivity of the sensor, a fine-pitch two-dimensional laser scan was performed. The result of this scan is shown in Figure 11.28a, where the signal amplitude is mapped over the scanned area. Each bin corresponds to a $1 \times 1\ \mu m^2$ region, and the colour scale indicates the recorded signal height in volts. This uncorrected map provides an initial indication of the sensitive area.

To fix for systematic scanning artefacts caused by scanning one row from left to right and the next row from right to left, an x-axis correction was applied row by row. For each row, a one-dimensional histogram of the signal was generated, and a Gaussian function was fitted to extract the mean values. These values were then plotted against the row index to reveal a



(a) Initial 2D signal amplitude map obtained from laser scan (uncorrected)

(b) Corrected amplitude map after stage movement compensation and normalisation

Figure 11.28: Signal maps of the sensor under laser stimulation

systematic zig-zag displacement, as illustrated in Figure 11.29. To smooth the data and improve fit robustness, the arithmetic mean between each pair of adjacent points was calculated. A linear fit (shown in red) was then applied to represent the global trend. The final correction for each row was obtained as the difference between this linear fit and the local Gaussian mean. Applying this correction aligned the signal peaks and allowed a more accurate estimation of the sensor's sensitive region.

Figure 11.28b shows the corrected amplitude map after this procedure. From this corrected image, the sensitive area of the pixel was extracted using a threshold of 0.2 V. This method was applied consistently across other datasets, including the one shown in Figure 11.27a. As a result, a comparison of depletion region radii was conducted for various substrates and bias voltages. The extracted experimental radii are listed in Table 11.7 alongside the corresponding TCAD simulation results. For the epitaxial substrate sample at −30 V bias, the sensitive area corresponds to a depletion diameter of approximately 110 µm, which is in close agreement with the simulation result of 98 µm.

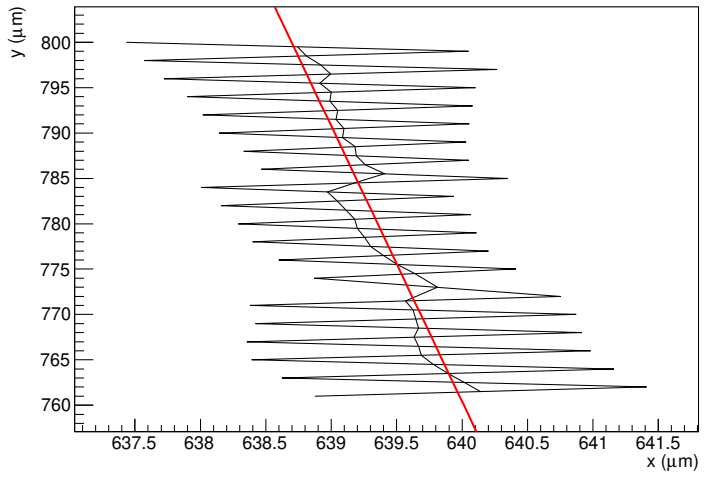
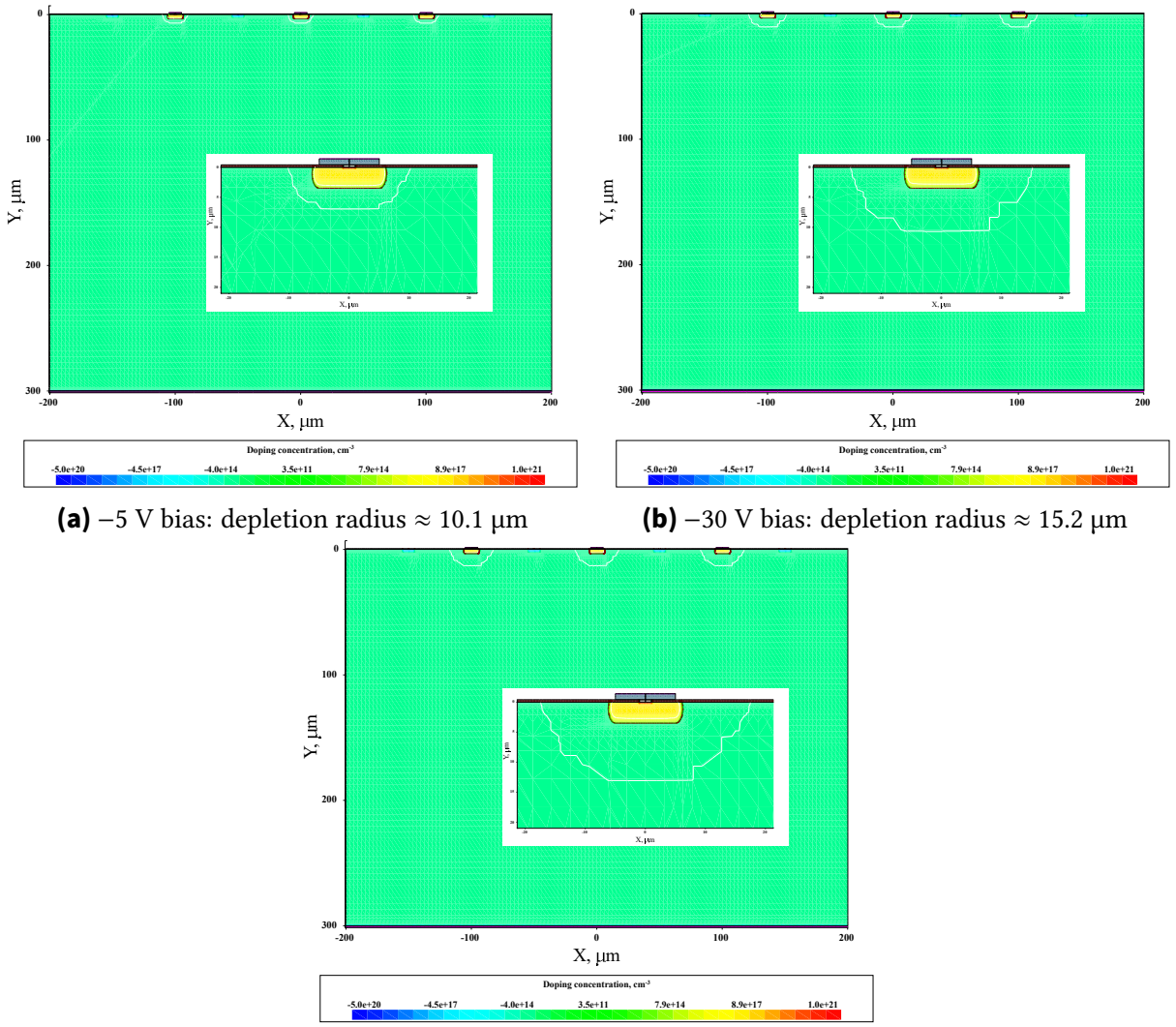


Figure 11.29: Zig-zag correction for x-stage movement. The red line indicates the linear fit to the smoothed Gaussian means.

Table 11.7: Measured and simulated depletion region radii under various bias voltages for different substrates

Sample name	Substrate type	Applied bias voltage, V	Radius of depletion region (experimental), µm	Radius of depletion region (simulated), µm
s31	Standard	−5 V	(15.96 ± 3.1) µm	10.10 µm
		−30 V	(18.71 ± 3.1) µm	15.18 µm
		−50 V	(19.54 ± 3.1) µm	17.47 µm
s34	Epitaxial	−30 V	(55.00 ± 3.1) µm	52.00 µm

The corresponding TCAD simulations are presented in Figure 11.30, showing the simulated depletion regions and doping distributions around the pixel for the standard substrate (Sample s31) at three bias voltages: −5 V, −30 V, and −50 V. These results visually confirm the consistency between the simulated and measured depletion region sizes.



(a) −5 V bias: depletion radius ≈ 10.1 µm

(b) −30 V bias: depletion radius ≈ 15.2 µm

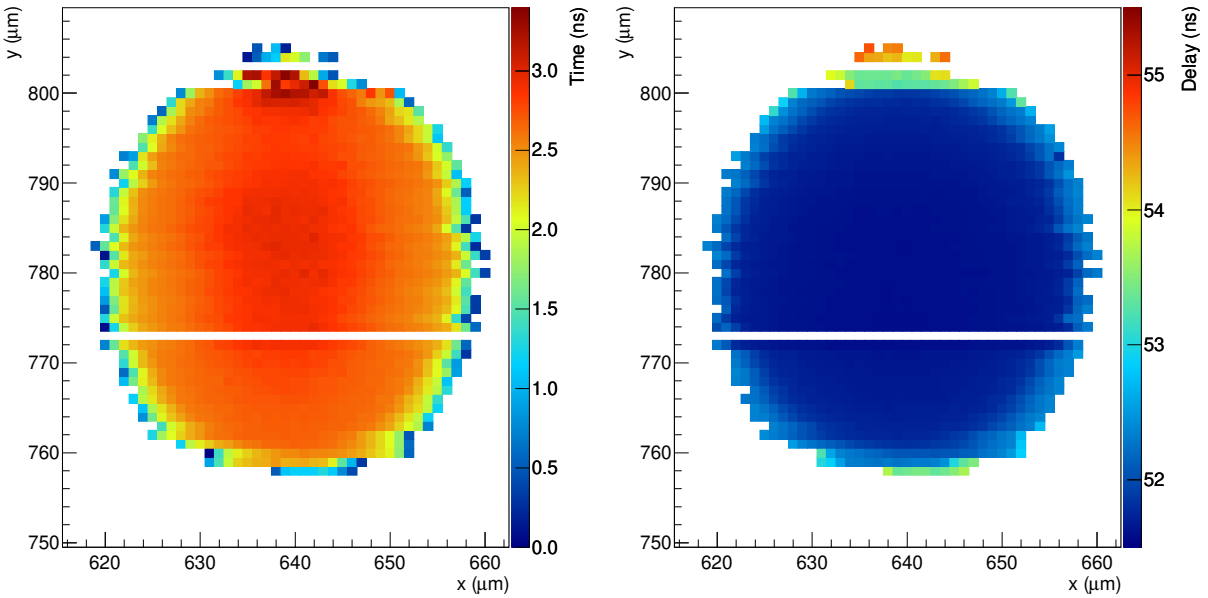
(c) −50 V bias: depletion radius ≈ 17.5 µm

Figure 11.30: TCAD-simulated depletion regions for Sample s31 at various bias voltages, with doping distributions shown in colour scale

11.8.3 Time resolution analysis

To characterise the timing performance of the sensor, a laser scan was conducted using a standard substrate (Sample s31) with a resistivity of 50 Ω·cm and a reverse bias of −30 V. The resulting three-dimensional distributions of rise time and response time across the pixel area are shown in Figure 11.31.

The rise time remains relatively constant within the sensitive area, which includes both the central 10×10 µm^2 collection node and the surrounding depletion region. At the centre, the rise time is approximately 2.5 ns. However, as the position moves radially outward, the signal amplitude decreases, as shown on Figure 11.28b, resulting in shorter apparent rise times. This artefact arises because the weaker signals reach their peak earlier in the signal window,



(a) Rise time map over the scanned sensor area

(b) Signal response time (delay) across the pixel region

Figure 11.31: Rise and response time maps for a standard substrate (Sample s31) at −30 V bias

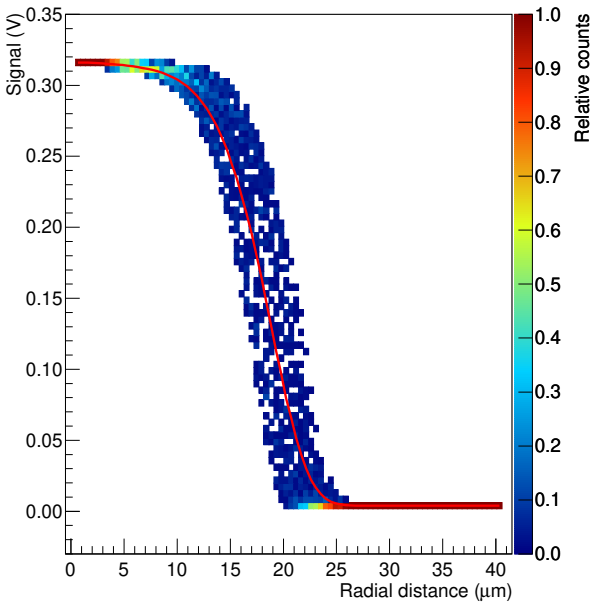
not due to faster carrier dynamics. At the pixel edge, where the signal approaches zero, the extracted rise time becomes dominated by measurement noise and systematic uncertainty, which in this setup is estimated at approximately 3 µm spatially. It is worth noting that for sensors with high-resistivity epitaxial substrates, rise time measurements were performed using a ^{90}Sr source and are discussed in Subsection 11.7.3. In those cases, the measured rise time was typically around 1 ns, suggesting that the observed timing differences here are more indicative of the electronic response rather than the substrate properties.

The delay map in Figure 11.31b provides further insight into charge transport mechanisms within the sensor. The measured delay time, defined as the time between the laser pulse and the onset of the recorded signal, is approximately 52 ns at the centre of the pixel. This includes contributions from charge drift in the substrate, the intrinsic delay of the laser trigger, and the front-end electronics. For comparison, TCAD simulations of the same sensor structure estimated a total delay of around 40 ns, indicating good qualitative agreement.

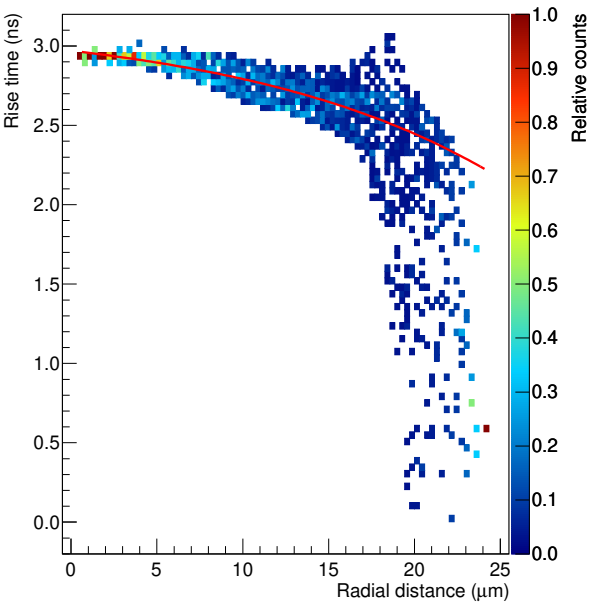
Additionally, to investigate radial symmetry and spatial dependence of the timing performance, the rise time, signal amplitude, and delay were each projected onto the radial distance from the pixel centre. This was done by dividing the area around the pixel centre into concentric rings and creating a histogram for each ring. The results are shown in Figure 11.32.

Figure 11.32a shows a clear drop in signal amplitude beyond a radial distance of approximately 20 µm, in agreement with earlier depletion region estimates. In Figure 11.32b, the rise time becomes increasingly scattered at the sensor periphery, where the signal-to-noise ratio de-

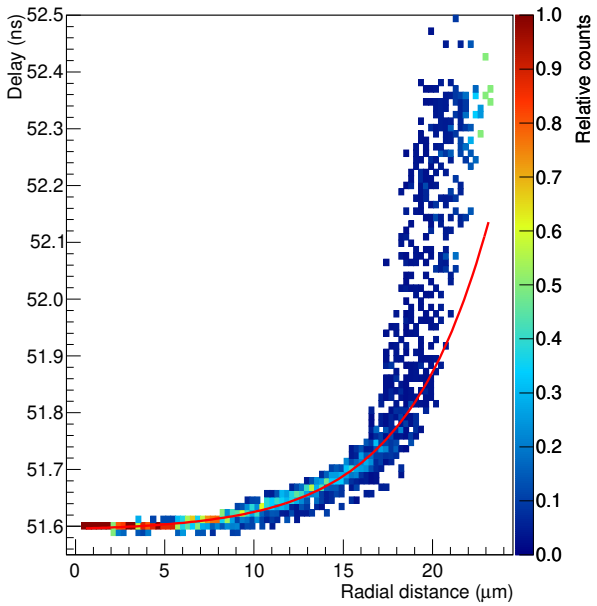
grades. Finally, Figure 11.32c reveals a sharp increase in delay near the pixel edge, further supporting the interpretation that charge collection efficiency and transport time deteriorate beyond the depleted region. These radial plots reinforce the observation that within the central depleted area, timing characteristics remain stable, whereas beyond the edge, both amplitude and timing degrade due to weaker fields and incomplete charge collection.



(a) Signal amplitude as a function of radial distance



(b) Rise time versus radial distance



(c) Response time (delay) as a function of radial position

Figure 11.32: Radial dependence of signal characteristics derived from concentric ring analysis

12 Summary of BeMiPix development in SiGe BiCMOS technology

BeMiPix, short for **Bipolar experimental Miniature Pixel**, represents a new class of monolithic silicon sensors developed for high-resolution charged-particle detection. The sensor leverages the advantages of the IHP SG13G2 130 nm SiGe BiCMOS process[1], which allows the integration of high-performance SiGe Heterojunction Bipolar Transistors (HBTs) alongside standard CMOS logic in a single pixel cell. This integration supports compact, high-speed analogue front-ends within the pixel area, enabling a fully monolithic architecture with improved noise and timing characteristics.

The BeMiPix sensor is based on a Monolithic Active Pixel Sensor (MAPS) design, in which both the sensing diode and analogue readout circuitry are located within the active pixel matrix. Each pixel includes a 10 µm × 10 µm nwell collection node and is electrically isolated by a p-stop. This approach minimises parasitic capacitance and enhances signal integrity, particularly under low-noise, high-rate conditions.

The development cycle included sensor structure simulations, front-end design, layout integration, and a full characterisation campaign across multiple fabricated variants.

Design and simulation insights

The sensor structure and pixel geometry were optimised using the Sentaurus TCAD and presented in Section 11.1. Simulations were carried out on various substrate types, including standard 50 Ω·cm bulk and epitaxial layers of 20 µm, 30 µm, and 50 µm thickness. These simulations guided design choices for charge collection efficiency, breakdown voltage, and depletion behaviour. A 100 µm pixel pitch was selected as optimal, balancing charge sharing and layout constraints while avoiding premature edge breakdown. Charge collection simulations with MIP-like particles confirmed good lateral depletion and efficient signal formation dynamics.

The analogue front-end was implemented using a SiGe HBT-based transconductance amplifier integrated within the collection node. The signal was processed by a common-base amplification stage, allowing for fast, low-noise external readout. Layout integration in SG13G2 required careful matching of pixel pitch, front-end circuitry, and isolation structures.

[1] https://www.ihp-microelectronics.com/

Experimental characterisation

Following the completion of the design and simulation phase, several sensor variants were fabricated and tested. An overview of the fabricated and characterised samples is provided in Table 11.1, which serves as a reference for the experimental results described in the subsequent sections. Fabricated sensor samples were tested for leakage current, breakdown behaviour, temperature dependence, X-ray and particle source response, and time resolution.

Section 11.4 presents dark current measurements, X-ray response, and breakdown voltage analysis. The measured breakdown voltage for all tested sensors falls within the range of −70 V to −90 V, which is typical for silicon sensors in standard operating conditions. At an operational bias of −50 V, the dark current ranges from 4 nA to 10 nA depending on the substrate type. Exposure to a steady X-ray beam was used to investigate sensor current performance. The highest dark currents were observed in the sensor fabricated without an epitaxial layer (Sample s1.2), ranging from approximately 5.77 nA to 58.9 nA, whereas samples with epitaxial layers showed improved leakage characteristics. The best performance was recorded for the 20 µm epitaxial sample (w6.1), which maintained a stable current near 4 nA across all dose rate levels. This is a significantly better result than typical silicon sensors, where dark currents often reach the mA/cm^2 range.

Section 11.5 presents the results of temperature-dependent IV measurements performed in a climate chamber, aimed at evaluating the thermal behaviour of the BeMiPix sensor under controlled environmental conditions. Two sensor samples, one fabricated with a 50 µm epitaxial layer and the other without, were tested across a temperature range, recreating typical operating conditions. The leakage current shows an exponential increase with temperature in both cases, consistent with the expected behaviour of thermally generated charge carriers in silicon.

Notably, the sample with the 50 µm epitaxial layer displayed both a higher absolute leakage current and greater temperature sensitivity compared to the sensor without. This suggests that thicker epitaxial layers may require more effective thermal management, especially in environments where sustained elevated temperatures are anticipated.

To evaluate the signal response and charge collection properties of the BeMiPix sensor, measurements were performed using radioactive sources ^{90}Sr and ^{55}Fe and presented in Section 11.7.

The ^{90}Sr measurements were used to study the sensor's charge collection dynamics under realistic MIP-like conditions. The analogue output voltage was recorded directly at the output node using an oscilloscope, allowing for the precise extraction of signal waveforms. Measurements were performed on multiple samples differing in substrate configuration: a standard 50 Ω·cm bulk substrate and three epitaxial structures with 50 µm, 30 µm, and 20 µm layer thicknesses. The rise time values showed excellent agreement with TCAD simulations, confirming the validity of the drift-dominated collection model. Thinner epitaxial layers demonstrated faster rise times, with the 20 µm sample reaching 0.87ns and the 50µm sample about 1.09 ns, closely matching simulated values of 0.83 ns and 1.04 ns, respectively. These results confirm that signal speed is primarily limited by drift path length and validate the choice of epitaxial thickness for optimised timing.

In addition to β^- particles, the ^{55}Fe X-ray source provided a well-defined energy reference for calibrating the system. The 5.9 keV X-rays produce approximately 1639 electron-hole pairs in silicon, allowing a direct comparison between measured signal amplitudes and expected charge collection. The calibration was essential for converting the analogue voltage output into an estimate of the collected charge, providing a quantitative foundation for comparing measurements with TCAD predictions and evaluating sensor performance under different substrate and biasing conditions. These measurements were used to verify that the system operated without charge multiplication at higher bias voltages.

Together, the source measurements confirmed the effectiveness of the BeMiPix architecture in terms of fast signal formation, high charge collection efficiency, and well-matched simulation-to-experiment behaviour. The results validate both the sensor and the front-end design, highlighting the potential of SiGe BiCMOS monolithic sensors for high-rate particle tracking applications.

Laser-based characterisation, presented in Section 11.8, was used to investigate both the spatial and temporal response of the BeMiPix sensor. The scanned signal amplitude map allowed precise identification of the sensitive area, including the central 10×10 µm^2 collection node and its surrounding depletion region. To improve spatial accuracy, an x-axis correction was applied to compensate for systematic stage movement artefacts, resulting in a corrected signal map from which the depletion diameter was extracted. These measurements were repeated across different bias voltages and compared with TCAD simulations, confirming good agreement between experiment and simulation. The estimated depletion radius ranged from 10 to 55 µm, depending on substrate type and voltage, and matched simulation results within a few microns. In addition to spatial mapping, laser timing measurements provided detailed insight into the dynamic behaviour of the sensor. Rise time and response delay maps were generated for a standard substrate sample under −30 V bias. Within the central region, the rise time was uniform at approximately 2.5 ns, while edge effects and signal weakening led to apparent shortening of the rise time. The measured signal delay was about 52 ns, which includes drift, laser trigger latency, and electronic response. This value is consistent with TCAD simulations predicting approximately 40 ns. Radial projections of amplitude, rise time, and delay confirmed the uniform response within the depleted area and highlighted degradation at the periphery.

Overall, the laser characterisation clearly defined the size of the active charge collection area and confirmed the stable timing behaviour of the sensor. While simulations do not always perfectly reflect experimental reality, in this case, the close agreement between TCAD results and measurements underscores the reliability of both approaches and their value in guiding sensor development. These measurements, taken with high spatial precision, support the simulation results and serve as a reference for comparison with measurements using radioactive sources.

Conclusion

BeMiPix demonstrates the successful application of SiGe BiCMOS technology in a fully monolithic pixel sensor. The measured performance, in terms of leakage current, timing resolution, and depletion control, is promising, particularly for fast and low-noise applications. However, the complexity of design and production in this technology remains high, with significant effort required for layout, isolation, and process adaptation. Overall, the project confirms the functionality of SiGe HBT BiCMOS as a platform for next-generation high-speed, radiation-tolerant particle detectors. This technology combines the advantages of fast analogue amplification with dense CMOS logic in a single process, making it suitable for demanding environments such as future collider experiments or space-based instrumentation. The success of BeMiPix reinforces the potential of BiCMOS MAPS to complement and possibly replace hybrid sensors in specific applications, provided further optimisation and scalability efforts are pursued.

Part IV

Semiconductor detectors production for high-energy physics experiments

13 High-density interconnection techniques

This chapter explores high-density interconnection techniques for sensor assembly, focusing on gold bonding due to its widespread application in fine-pitch interconnections. Section 13.1 introduces key bonding technologies, including wire bonding and flip chip bonding, before detailing two projects that exemplify their application. The first, presented in Section 13.2, involves ultra-thin gold wire bonding for small pad-size SiGe devices, addressing the challenge of adapting the bonding process to a pad size of 38 µm × 42 µm. The devices were developed by Université de Genève. The second project, detailed in Section 13.3, focuses on gold bump bonding for The Pixelated Silicon Drift Detector (PixDD), overcoming challenges such as highly asymmetrical bump placement, the handling of fragile X-ray coating layers as thin as 0.5 µm, and the need for exceptionally strong interconnects for space applications. This chapter provides a comprehensive analysis of these bonding processes, highlighting their technological advancements and their relevance in modern sensor integration.

13.1 Introduction to high-density interconnection techniques for sensor assembly

High-density interconnection techniques play a fundamental role in modern sensor integration, establishing both electrical and mechanical connections between silicon sensor elements and their readout electronics. The choice of interconnection method significantly impacts the overall performance, reliability, and integration density of the assembled system. The most typical interconnection techniques in microelectronics packaging include wire bonding, flip chip bonding, and tape automated bonding (TAB). Each method presents unique advantages and trade-offs, influencing their suitability for specific applications.

Wire bonding has been the dominant interconnection method in semiconductor packaging since its introduction in the 1960s. It remains a cost-effective, flexible, and mature technology, with over 80% of semiconductor packages in the industry still relying on wire bonding [136]. Over the decades, wire bonding has evolved to accommodate the increasing demands of miniaturisation and high-performance applications, such as power electronics, image sensors, and high-performance computing [137]. The process involves forming fine metal wires—typically gold, copper, or aluminium—between the bond pads of an integrated circuit and a substrate or lead frame.

In contrast, flip chip bonding, introduced by IBM in the 1960s, emerged as an alternative to traditional wire-based interconnections. The introduction of the Controlled Collapse

Chip Connection (C4) technology enabled direct electrical connections between a die and a substrate through solder bumps [138]. Flip chip technology initially found applications in high-performance computing, but advancements in organic substrates and packaging materials, particularly Ajinomoto build-up films, facilitated its widespread use from the 1990s onward [139]. By the early 2000s, flip chip technology gained prominence in high-frequency and high-pin-count devices, offering superior electrical performance and interconnect density compared to wire bonding [140].

Tape automated bonding (TAB) was developed in the 1960s as an alternative to wire bonding, aiming to enhance interconnection density and reliability. TAB employs a flexible polymer tape, typically polyimide, with pre-patterned copper leads that establish electrical contact between the semiconductor die and an external circuit. The method gained prominence in the 1970s and 1980s, particularly in applications requiring high-speed and high-density interconnections, such as LCD driver circuits, military electronics, and high-performance computing. Unlike wire bonding, which sequentially connects each pad, TAB enables simultaneous bonding of multiple leads, reducing assembly time and improving electrical performance by minimising lead inductance. By the 1990s, major semiconductor manufacturers, including IBM and Hitachi, had adopted TAB for high-reliability chip packaging. However, despite its advantages, TAB gradually declined in popularity in favour of flip chip bonding in the late 1990s and early 2000s, as flip chip technology offered superior scalability, cost efficiency, and compatibility with modern surface-mount assembly techniques. TAB remains relevant in specialised applications, particularly in flexible electronics and display driver interconnections [141].

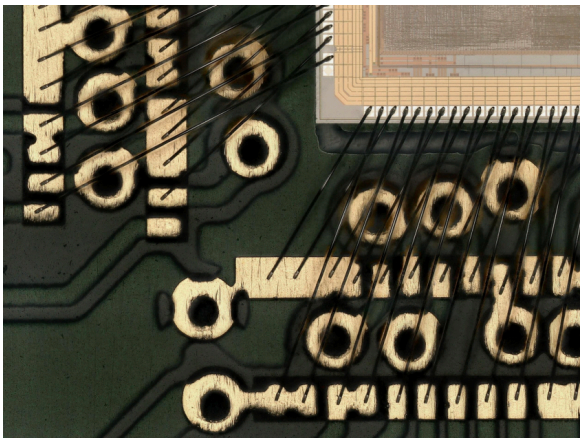
Table 13.1 summarises the primary differences between wire bonding, flip chip bonding, and TAB, highlighting their advantages and trade-offs. Figure 13.1 presents microscope images illustrating the key interconnection technologies examined in this study.

Table 13.1: Comparison of wire bonding, flip chip bonding, and TAB for semiconductor interconnections.

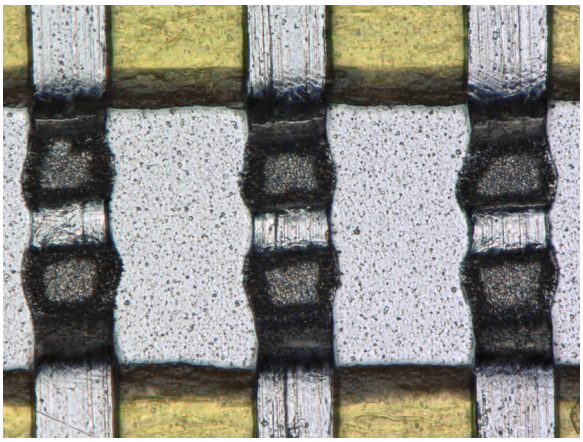
Attribute	Wire bonding	Flip chip bonding	Tape automated bonding (TAB)
Bond pitch	Typically 50–100 µm bond pad pitch with peripheral pads only.	Area-array bump layout allows much higher connection density around 150 µm bump pitch, and <50 µm for advanced applications.	It provides fine-pitch interconnections enabled by flexible polymer tape. Possible to achieve high-density layouts similar to flip chip but with added mechanical flexibility.

Table 13.1 – continued from previous page

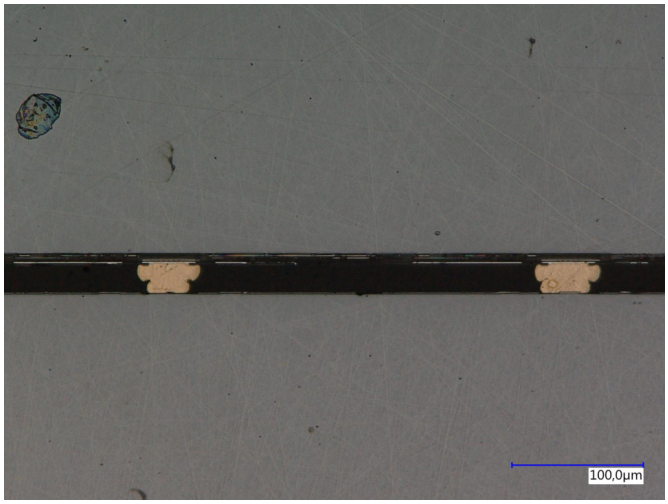
Attribute	Wire bonding	Flip chip bonding	Tape automated bonding (TAB)
Electrical performance	Longer signal path of mm-scale wires gives higher parasitic inductance and capacitance, which may limit high-frequency performance and cause more noise. This technology is suitable for up to moderate speeds.	Short interconnects with µm-scale bumps give low parasitic inductance, excellent high-frequency and high-speed performance with minimal signal distortion.	Reduced lead inductance compared to wire bonding; simultaneous bonding of multiple leads enhances electrical performance, but not as compact as flip chip.
Thermal management	Heat mainly conducts from the die's backside. May require thermal enhancements such as heat spreaders or thermal vias.	Direct heat dissipation through bumps into the substrate and upwards through a heatsink. Underfill spreads thermal expansion stress.	Heat dissipation primarily depends on the polymer tape and supporting structures; thermal performance is typically lower than that of flip chip.
Reliability	Highly mature process; robust if properly encapsulated. Wire breakage and bond lift-off are possible under mechanical stress.	Strong mechanical support after underfill. No long wires to vibrate. Solder bumps can fatigue under thermal cycling if not underfilled.	Flexible interconnections provide mechanical compliance but may be susceptible to degradation in harsh environments.
Manufacturing and cost	Low cost for low to medium density of input/output (I/O) connection packages, where standard leadframe substrates keep package cost low. Sequential bonding adds time as the I/O connection count increases.	Higher initial cost due to wafer bumping, substrate, and precision placement. Batch bonding of all bumps at once is efficient for high-volume, high-density I/O connection applications.	It requires pre-patterned tape production; it is cost-effective for specialised applications but less scalable than flip chip.



(a) Wire bonds used in the PixDD project, demonstrating fine-pitch gold wire interconnections.



(b) TAB interconnects fabricated by the AVT group (Aufbau- und Verbindungstechnik - Electronic Packaging and Interconnects) at IPE, as part of the CBM project.



(c) Cross-section of bonds after a flip chip process, showing solder bump attachment and interface structure.

Figure 13.1: Microscope images of interconnection technologies used in high-density packaging.

13.1.1 Wire bonding

Wire bonding is a widely used method for creating electrical interconnections in semiconductor packaging. It involves the use of thin metal wires, typically made of gold, aluminium, or copper, which are ultrasonically or thermosonically bonded between a chip's bond pad and a substrate or printed circuit board (PCB) pad [142]. This process plays a crucial role in microelectronics assembly due to its cost-effectiveness, reliability, and adaptability across various applications.

A crucial prerequisite for wire bonding is the die attachment process, where the semiconductor chip is mounted onto a substrate. The three primary die bonding techniques include polymer,

solder, and glass die bonding. Among these, polymer die bonding is the most common, employing adhesives such as epoxy, polyimide, or silicone to secure the die in place. The adhesive properties, including its mechanical, electrical, and thermal characteristics, can be modified by adding filler materials like gold, silver, alumina (Al_2O_3), or boron nitride. This process is typically automated, involving substrate preparation, adhesive dispensing, chip placement, and curing, making it an attractive option due to its flexibility, low process temperature, and cost-effectiveness [143].

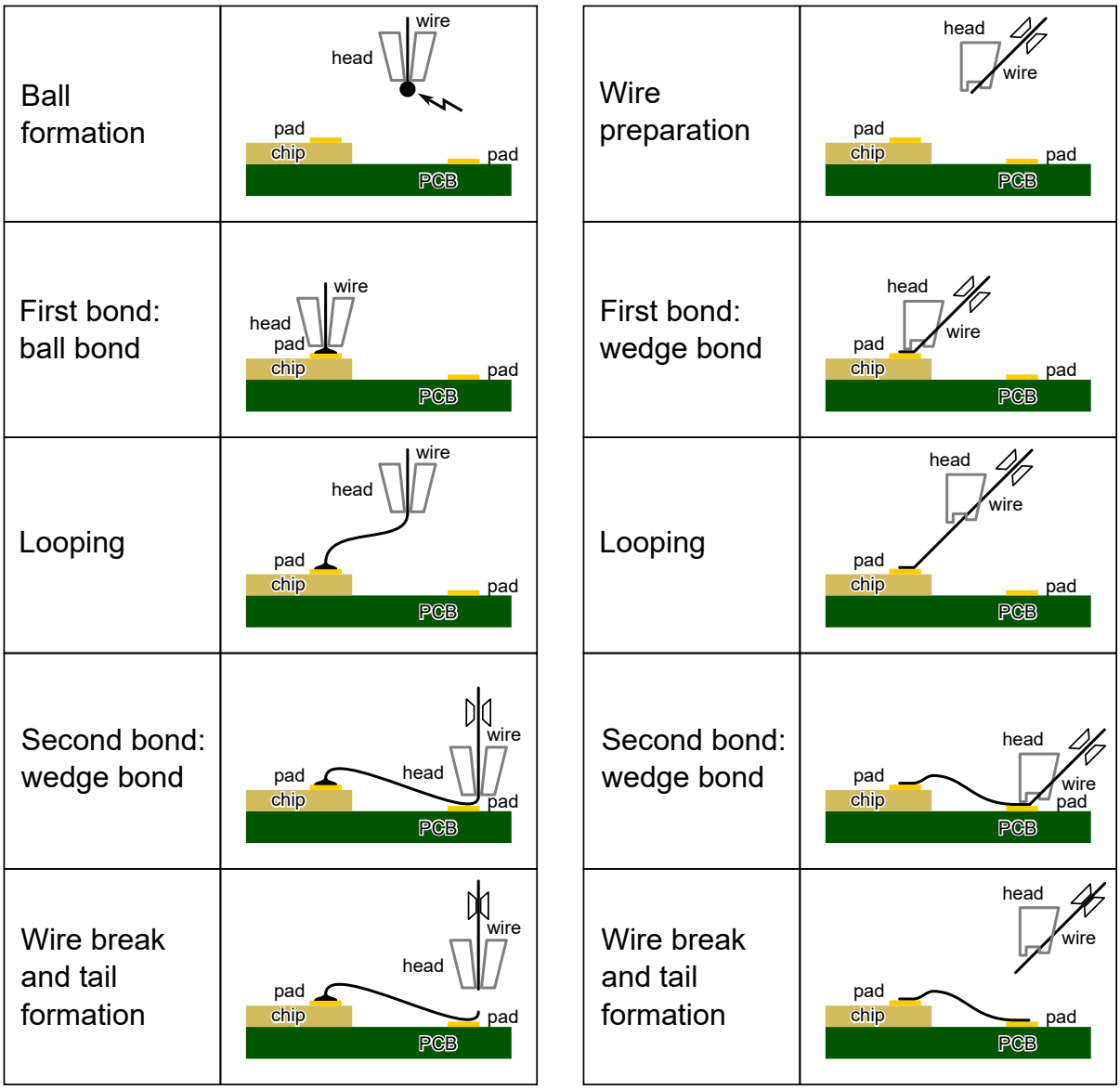
For applications requiring superior thermal performance and mechanical stability, solder die bonding is an alternative solution. This method uses solder alloys such as AuSi (eutectic), AuGe, PbSn, AuSn, and PbAgIn, providing excellent thermal conductivity and high-temperature stability. However, the process requires significantly higher bonding temperatures, typically ranging between 200 °C and 400 °C. Glass die bonding, on the other hand, uses a glass frit material that is screen-printed or dispensed onto the substrate before being cured at around 400 °C. Although this method is commonly employed in ceramic packaging, its requirement for high-temperature processing restricts its wider application [143].

Following die attachment, wire bonding is performed by connecting the chip's bond pads to the corresponding substrate pads using fine wires. Two primary bonding sequences are distinguished: ball–wedge bonding and wedge–wedge bonding. These sequences are implemented using one of three bonding processes: ultrasonic (US), thermosonic (TS), and thermocompression (TC) bonding.

Ball–wedge bonding, shown in Figure 13.2a, is the most commonly used method, particularly for gold wires. The process begins with forming a spherical ball at the end of the bonding wire through an electronic flame-off discharge. The first bond, known as the ball bond, is established by pressing the ball onto the chip's bonding pad while applying heat (thermocompression bonding) or a combination of heat and ultrasonic energy (thermosonic bonding). The wire is then looped to the second bonding site on the substrate, where a wedge bond is formed. This technique enables high-speed bonding, with modern automated bonders achieving up to 20 cycles per second. In fine-pitch applications, gold wires as thin as 15 µm are used, allowing wire bonding with pad pitches as small as 35 µm.

Wedge–wedge bonding, commonly used with aluminium wires, differs in that both bonds in the cycle are wedge bonds. This process is shown in Figure 13.2b. Unlike ball–wedge bonding, this method requires the substrate to be rotated during bonding since the bonding direction must align with the tool's centerline. Wedge–wedge bonding can be carried out at room temperature using ultrasonic energy alone or with heat in thermosonic or thermocompression bonding. This approach is beneficial for fine-pitch bonding, laboratory prototyping, and microsystems packaging, where flexibility in substrate choice and wire material is required.

Wire bonding offers several advantages, including cost-effectiveness, high reliability, and adaptability to different chip sizes and applications. As a mature and well-established process, it remains the most widely used interconnection method in semiconductor manufacturing. However, challenges such as susceptibility to mechanical stress and process limitations for high-frequency applications persist. Table 13.2 summarises the key characteristics of different wire bonding techniques.



(a) Stud ball bumping or ball–wedge bonding. The process involves forming a ball at the end of the wire, bonding it to the chip, looping the wire, and creating a wedge bond on the substrate.

(b) Wedge–wedge bonding. The wire is bonded in a sequential manner using ultrasonic energy. This technique is commonly used with aluminium wires and requires substrate rotation.

Figure 13.2: Schematics of (a) ball–wedge and (b) wedge–wedge bonding cycles; only the tips of the bonding tools and a schematic of the wire clamp are shown for clarity.

Table 13.2: Comparison of wire bonding techniques [144].

Type	Method of bonding	Material of wire	Temperature	Shape of wire end	Bond force (typical)
Thermo-compression	Melting (softening), pressure	Au	300–500 °C	Ball/wedge	15–25 mN
Ultrasonic	Pressure, US vibration	Al, Au	25 °C	Wedge/wedge	5–25 mN
Thermosonic	Elevated temperature, pressure, US vibration	Au	100–150 °C	Ball/wedge	5–25 mN

13.1.2 Flip chip bonding

Flip chip bonding is an advanced packaging technique that enables direct electrical and mechanical connection between a semiconductor die and a substrate. Unlike conventional wire bonding, which requires bonding wires to be attached peripherally, flip chip bonding employs an array of conductive bumps on the chip's bond pads. These bumps allow the die to be flipped upside-down and placed directly onto a matching set of contact pads on the substrate. The interconnection is then secured using various attachment methods such as solder reflow, thermocompression, or conductive adhesives [145].

One of the major advantages of flip chip bonding is its ability to provide shorter interconnect paths, significantly reducing parasitic effects such as capacitance and inductance. This results in superior electrical performance, making the method ideal for high-frequency and high-speed applications. Additionally, the area-array nature of flip chip bonding allows for much higher interconnect densities compared to wire bonding, making it particularly suited for applications requiring a large number of I/O connections in a compact space [142].

The flip chip bonding process begins with the fabrication of conductive bumps on the bond pads of the semiconductor die. These bumps, commonly made from solder, gold, or conductive polymers, are deposited using electroplating, screen printing, electroless plating, evaporation, or wire bonding-based bumping methods. The choice of bumping technique depends on the required interconnect pitch, reliability considerations, and application requirements. Once the bumps are formed, the die is flipped and aligned with the corresponding contact pads on the substrate, ensuring precise positioning to achieve reliable electrical connections.

After alignment, the bonding process is carried out using one of several methods. Solder reflow bonding is the most widely used approach, where the assembly is heated to melt the

solder bumps, forming robust metallic interconnections upon cooling. Alternatively, thermocompression bonding relies on applying heat and mechanical pressure to create solid-state metallic bonds, often preferred for gold bump interconnections. Another approach, conductive adhesive bonding, employs anisotropic conductive films (ACF) or pastes, eliminating the need for high-temperature processes while still providing reliable electrical contact.

To enhance mechanical stability and long-term reliability, an underfill encapsulant is applied between the chip and the substrate. This polymeric material, typically an epoxy resin, fills the gaps around the bumps, mitigating mechanical stress and reducing thermal expansion mismatch between the two components. The underfill is then cured to a solid state, reinforcing the structural integrity of the assembly and improving resistance to thermal cycling and mechanical vibrations.

Compared to wire bonding, flip chip bonding offers several key advantages [146]. The use of an area-array connection allows for significantly higher interconnect densities, making it ideal for applications requiring a large number of I/O connections. The shorter interconnect paths improve electrical performance by minimising parasitic effects, enabling high-frequency and high-speed operation. The direct chip-to-substrate contact also enhances thermal dissipation, making flip chip bonding particularly suitable for high-power applications. The absence of bond wires also results in a more compact form factor, which is beneficial for modern miniaturised electronic packages. Furthermore, flip chip bonding enables advanced three-dimensional integration, allowing multiple semiconductor dies to be stacked vertically, further enhancing integration density and performance.

Flip chip bonding has become the preferred choice in applications where high-density interconnects and optimal electrical performance are critical. High-energy physics detectors, imaging sensors, and high-speed computing components all leverage flip chip bonding to achieve nearly 100% active area coverage while minimising dead space [147]. For example, modern pixel detectors in particle physics experiments integrate readout ASICs directly onto sensor arrays using flip chip bonding, ensuring optimal signal integrity and minimising interconnect lengths. This technology has also become indispensable in high-performance computing and mobile devices, where compact, high-speed, and thermally efficient designs are essential.

The flip chip bonding process is illustrated in Figure 13.3. The image shows the sequential steps of depositing solder bumps onto the contact pads of a semiconductor die, aligning the die with the substrate, and performing a reflow process to establish electrical and mechanical connections. The final step involves applying an underfill encapsulant, which enhances the mechanical robustness of the assembly by distributing stress and protecting the interconnects from environmental factors.

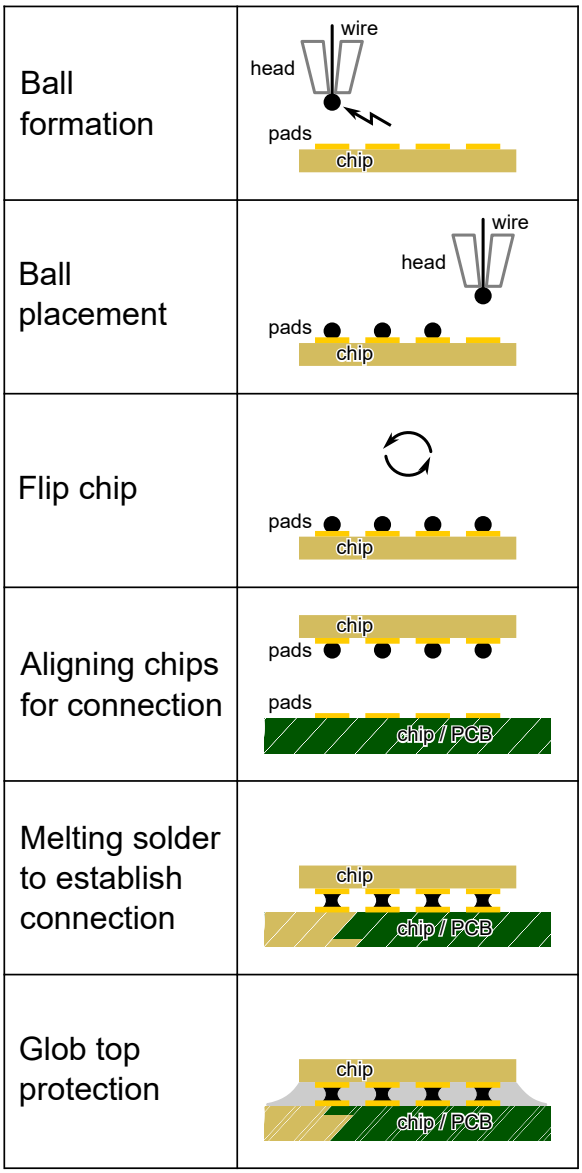


Figure 13.3: Illustration of the flip chip bonding process using solder bumps. The chip is flipped and precisely aligned with the substrate pads, followed by a reflow process that melts the solder and establishes interconnections. A polymeric underfill is applied to enhance mechanical stability and reliability.

13.1.3 Tape Automated Bonding (TAB)

Although this work primarily focuses on wire bonding and flip chip bonding, Tape Automated Bonding (TAB) deserves mentioning as an alternative interconnection technique that has played a significant role in semiconductor packaging and is used as an interconnection technique for CBM module assembly, which is discussed in Chapter 14. While TAB has largely been replaced in mainstream semiconductor applications, it remains relevant in specialised fields such as flexible electronics, Display Driver ICs (DDICs), and certain automotive and

aerospace applications. Its inherent flexibility makes it ideal for environments where rigid substrates may be impractical. Furthermore, developments in anisotropic conductive film (ACF) bonding have enabled TAB to maintain its relevance in niche applications that require fine-pitch, large-area interconnects.

TAB involves mounting a semiconductor die onto a flexible polymer tape, typically made of polyimide. The bonding sites on the die, often in the form of a solder bump, are precisely aligned and connected to fine conductive traces embedded in the tape. These conductors serve as the interface between the die and its external circuitry, enabling electrical connections either to a package or directly to an application circuit. In certain cases, the tape itself integrates the complete circuit, eliminating the need for additional interconnect layers.

The tape used in TAB is generally single-sided, with copper traces that are either electrodeposited or laminated using adhesives. The metal patterns are defined using photolithography and etching to ensure high precision and fine-pitch capabilities. Standard polyimide tape sizes range from widths of 35 mm, 45 mm, and 70 mm, with thicknesses varying between 50 µm and 100 µm. The tape is supplied in roll form [148].

To facilitate bonding, perforations are made at the edges of the tape, while openings are created in the middle to accommodate the semiconductor die. A copper foil is laminated onto the tape, forming a cantilevered lead frame through photolithography and etching. Once the chip is correctly positioned, bonding is achieved through either thermosonic or thermocompression bonding. In single-point thermosonic bonding, each lead is bonded individually by applying heat, force, and ultrasonic energy, forming intermetallic connections between the gold bump and the lead. Although reliable, this method is time-consuming. In contrast, thermocompression bonding uses a specially designed bonding tool to bond multiple leads simultaneously. This process, which does not rely on ultrasonic energy, allows for significantly higher throughput and is favoured for high-volume manufacturing.

Following bonding, the die, leads, and sections of the tape are encapsulated with a protective coating that enhances mechanical and chemical resistance. The encapsulated assembly undergoes electrical testing, after which the functional part of the tape is separated from the frame and prepared for final integration into the application circuit [148].

TAB offers several advantages over conventional wire bonding, particularly for applications requiring fine-pitch, high-density interconnections. The ability to bond multiple leads simultaneously enhances manufacturing efficiency and ensures consistent electrical performance. Additionally, the mechanical flexibility of TAB-based circuits makes them well-suited for applications involving flexible electronics, display technology, and environments subjected to shock and vibration [144].

However, the decline of TAB in favour of flip chip bonding highlights some of its limitations. The need for specialised tape-processing infrastructure and challenges in scaling for high-volume production have restricted its widespread adoption in modern semiconductor packaging. Nevertheless, in specialised applications where its unique properties provide distinct advantages, TAB remains a viable interconnection solution.

Figure 13.4 illustrates the TAB process. The scheme outlines the sequential steps, including tape preparation, lead formation, chip attachment, and encapsulation.

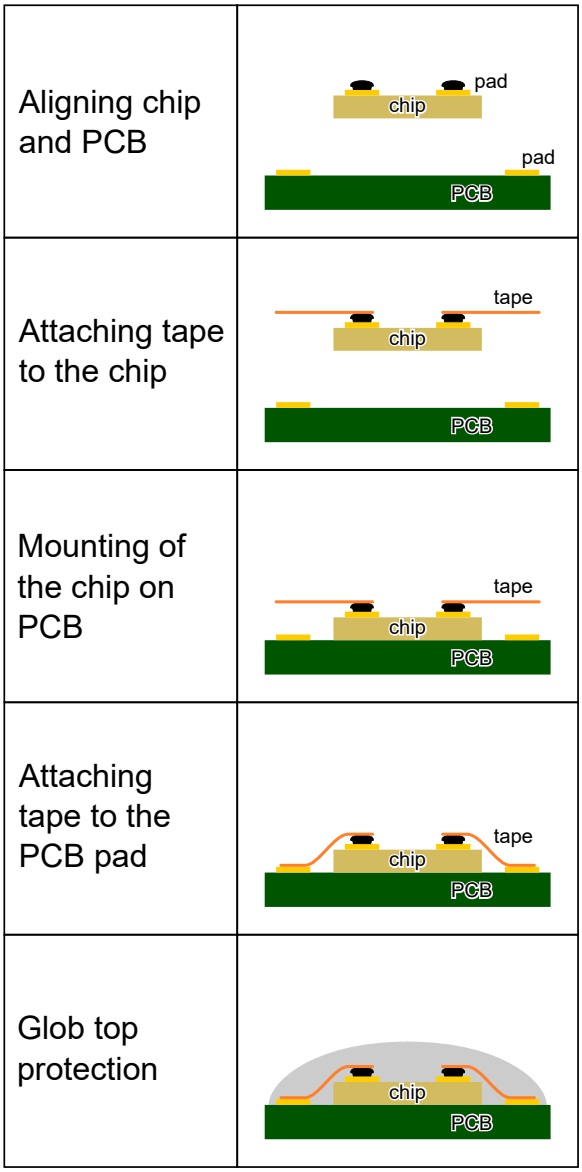


Figure 13.4: Illustration of the Tape Automated Bonding (TAB) process. A semiconductor die is mounted onto a flexible polyimide tape containing pre-patterned copper leads. The die's bond pads are precisely aligned with the tape's leads, and bonding is achieved through thermosonic or thermocompression methods. After bonding, encapsulation is applied to provide mechanical and chemical protection, followed by electrical testing and final assembly.

13.1.4 Gold bonding at IPE KIT — infrastructure and machines

The Institute for Data Processing and Electronics (IPE) at KIT has developed a sophisticated infrastructure for high-density interconnections. The cleanroom facility at IPE is equipped

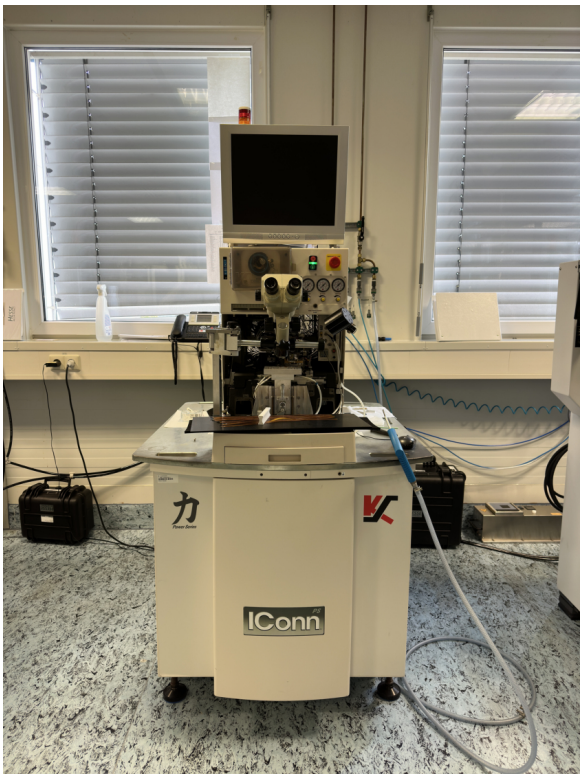
with advanced tools to handle fragile detector components with high precision and to carry out complex multi-step assembly processes. Key machinery includes state-of-the-art bonders and alignment systems that ensure accurate placement and reliable electrical connections. Figure 13.5 shows the cleanroom bonding laboratory at IPE, including the three main machines dedicated to different types of bonding.

The primary equipment for bonding at IPE includes:

- ICCON Gold Stud Bonder – a dedicated stud bump bonding machine used to deposit gold bumps (studs) onto chips or sensor pads. This tool uses a modified wire bonding process to form gold stud bumps: a gold ball is first bonded to the pad, and then the wire is shortened, leaving a hemispherical gold bump on the pad [149]. These gold bumps can also serve as interconnect bases for subsequent flip-chip bonding. The ICCON bonder allows fine control over the bump size, height, and pitch, enabling prototyping of bump-bonded assemblies without needing full wafer bumping facilities [150]. An example application is the bump fabrication for the PixDD project, discussed later (see Section 13.3).
- Finetech Fineplacer Femto Die Bonder – a high-precision flip chip bonder capable of sub-micron alignment accuracy for die placement. This automatic machine supports a range of bonding techniques, including thermocompression bonding of gold-to-gold bump interfaces, as well as solder reflow and adhesive bonding. Enclosed in a controlled environment, the Fineplacer Femto can accurately align a chip with respect to a sensor or substrate and apply defined pressure and temperature profiles to form robust bonds. At IPE, this tool is used for flip chip bonding, particularly gold-gold thermocompression bonding, where both chip and substrate have gold metallisation or studs. Its ±0.5 µm placement capability and controlled bonding forces are crucial for assembling hybrid pixel detectors and other high-density chip-sensor systems [151].
- Hesse Bondjet BJ653 Wedge/Ball Wire Bonder – an ultrasonic wire bonder used primarily for wire bonding (aluminium or gold wire) but also can be used for TAB. The Bondjet BJ653 features interchangeable bonding heads that support wedge-wedge bonding and ball-wedge bonding, together with TAB bonding, accommodating fine wire diameters as well as ribbon bonds. This versatility allows the same machine to perform traditional aluminium wedge bonds, as used in the CBM STS module assembly, or gold ball bonding for finer pitches. The open work area and semi-automatic operation make it ideal for prototype and small-scale production bonding with excellent repeatability [152]. In the KIT facility, multiple Bondjet BJ653 bonders are employed to interconnect sensors and front-end electronics, underscoring their role as the main machine for module assembly. Notably, in the CBM STS project, the BJ653 was used with aluminium wire for thousands of bonds per module, owing to the aluminium bond pad surfaces and the proven reliability of wedge bonding in that application. More about CBM STS Assembly can be found in Chapter 14.

In the cleanroom at IPE, a variety of projects are undertaken. Two projects that focused extensively on gold bonding techniques include gold wire bonding for Université de Genève (see Section 13.2) and gold bump bonding for The Pixelated Silicon Drift Detector (PixDD) (see

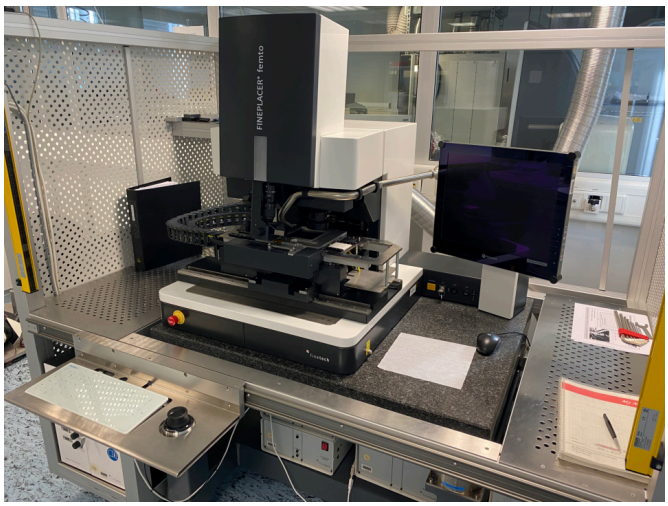
Section 13.3). Gold bonding encompasses a range of interconnecting processes that use gold as the principal material, including gold wire bonding, gold stud bump bonding, and gold-gold thermocompression flip chip bonding. Compared to aluminium-based bonding, gold bonding enables finer pitch interconnections. It exhibits superior long-term bond stability, although it comes at a higher material cost and often necessitates heat or ultrasound to achieve optimal bonding. These attributes make gold bonding preferred for high-density, high-reliability interconnections at IPE. Another key project, which is conducted at IPE, is the CBM STS module assembly. This project is covered in Section 14.4, with emphasis on module production, but it's important to mention that the IPE AVT team performed aluminium wire bonding and TAB bonding as a part of this project.



(a) ICCON gold stud bump bonder. This system is used for gold bump deposition to enable fine-pitch flip chip bonding.



(b) Hesse Bondjet BJ653 wire bonder. This system supports gold and aluminium wire bonding as well as TAB bonding.



(c) Finetech Fineplacer Femto flip chip bonder. This system achieves sub-micron alignment accuracy for high-density flip chip bonding applications.

Figure 13.5: Cleanroom bonding infrastructure at IPE, showing the ICCON gold stud bump bonder (a), Hesse Bondjet BJ653 wire bonder (b), and Finetech Fineplacer Femto flip chip bonder (c). These machines provide IPE with comprehensive capabilities in gold-based interconnection techniques, from fine wire bonds to flip chip assemblies.

13.2 Ultra-thin gold wire bonding for small pad-size SiGe devices — UniGe

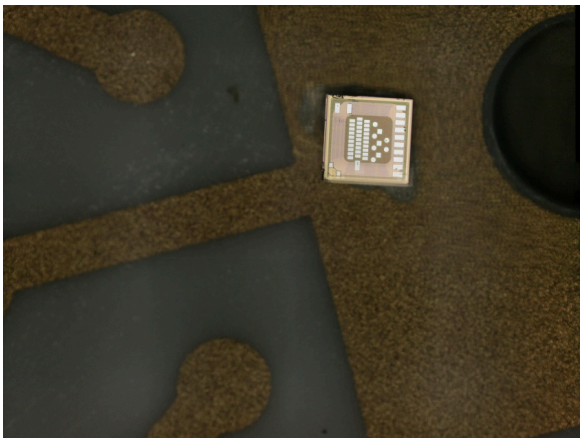
Gold wire bonding is a critical interconnection technique at IPE KIT, particularly for ultra-fine pitch bond applications. One key project involved the wire bonding of the chips developed by Université de Genève as part of the TANGERINE initiative for next-generation silicon detectors. This project presented unique challenges due to the extremely small pad sizes and the requirement for precise fine-tuning of the bonding parameters to ensure reliable interconnections.

The project involved bonding of three chips made in SiGe HBT (silicon–germanium Heterojunction Bipolar Transistors) technology with pad dimensions of 30×40 µm^2, which is significantly smaller than conventional bond pads. Two of these samples measured 1×1 mm^2, while the third was slightly larger at 1.5×1.5 mm^2. Achieving high-quality bonds on these compact pads required optimising bond force, ultrasonic power, and stage temperature, which is typically set between 100–150 °C for thermosonic bonding. Using the ICCON Gold Stud Bonder, the bonding process was adapted to accommodate 23 µm diameter gold wire. The bond placement accuracy was enhanced through automated pattern recognition and precision motion control.

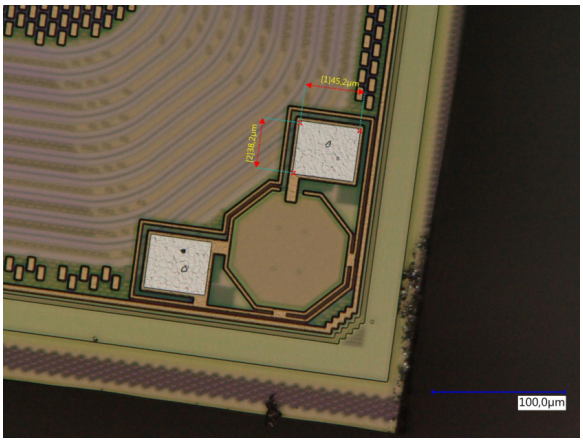
Initially, the Stand-Off-Stitch Bond (SSB) technique was chosen for this process. Unlike conventional wire bonding, which follows a forward bond approach with a ball bump bond on one side and a stitch on the other, SSB utilises an initial bump bond followed by a secondary bond on top of the bump [153]. A stand-off stitch bonding involves a ball bump placed in the exact location as a stitch before bonding the primary wire interconnect. When the primary wire is bonded, the termination of the wire is stitched on top of the previously placed ball bump. This approach was crucial for ensuring stable connections, given the small pad size [154]. The optimisation of the SSB process was carried out in three stages: bump bond optimisation, bump shaping, and stitch bond optimisation.

Initially, the bonding process was tested using smaller chips with different pad configurations on the surface to fine-tune bump formation parameters. Figure 13.6 illustrates one of these chip samples (13.6a) with measurement of pad sizes (13.6b), highlighting the precise dimensions necessary for fine-pitch bonding. The formation of the free air ball (FAB), which plays a key role in determining the bond quality, was optimised by adjusting bond force, ultrasonic power, bond time, and electronic flame-off (EFO) current [155]. For this application, the FAB size was set to 1.9 mils, which is equal to approximately 48 µm, with a bond force of 25 g, an ultrasonic current of 105 mA and a bond time of 5 ms. The electronic flame-off current was set at 30 mA to ensure consistent ball formation.

Following the optimisation of the bump formation, the focus shifted to refining the wire looping process. Loop shaping is a crucial step, as it prevents wire shorting and provides the necessary clearance between adjacent bonds. For this project, the BGA looping technique was employed. This method involves a programmable kink near the second bond, ensuring adequate standoff over ground or power rings. Several variations of the BGA loop, including



(a) Illustration of the small chip mounted on the PCB, demonstrating the bonding area. The chip size is 1×1 mm^2.



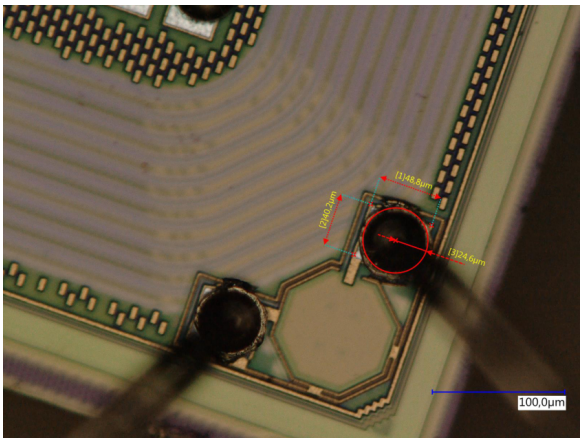
(b) Measurement of pad sizes on the chip, highlighting the precise dimensions necessary for fine-pitch bonding.

Figure 13.6: Images of the chips used in the wire bonding process.

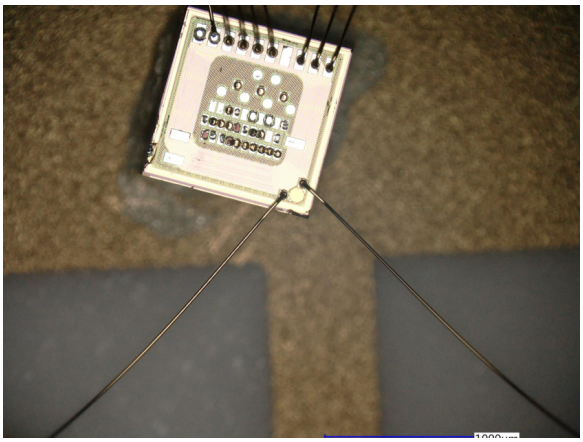
multi-bend and spider loops, were tested to determine the optimal configuration for minimising stress and improving reliability [156].

Once the bonding parameters were finalised, the wire bonding process was performed on the three chip samples mounted on ceramic substrates. The final bonded samples were examined during optical inspection to verify bond integrity, and the bonds were evaluated based on their mechanical strength. Electrical performance was checked by the Université de Genève group upon delivery. Figure 13.7 presents the results, showcasing the final bonded structures and the uniformity of the wire loops.

In summary, the ultra-thin gold wire bonding process developed at IPE KIT successfully enabled the reliable interconnection of the chip with extremely fine pad geometries. The process achieved secure and repeatable connections through systematically optimising bonding parameters and loop-shaping techniques, making it a viable approach for high-density microelectronics applications. This project underscored the importance of precise control over bond formation and looping in achieving high-quality wire bonds for advanced semiconductor devices.



(a) Microscopic view of the wire and ball bond on the sensor, with precise dimensional measurements.



(b) Final bonded chip showing the wire loops and the connection integrity after optimisation. This image illustrates bond adjustment outcomes in the middle of the chip.



(c) Microscope 3D-image of the final bonded assembly, illustrating the precise interconnections between the chip and PCB.

Figure 13.7: Final bonded samples demonstrating the success of the optimised wire bonding process.

13.3 Gold bump bonding for The Pixelated Silicon Drift Detector — PixDD

The PixDD project, short for Pixelated Silicon Drift Detector, is focused on developing an advanced two-dimensional multi-pixel X-ray sensor system using Silicon Drift Detector (SDD) technology. It is designed to operate in the 0.5–15 keV energy band and aims to achieve good spectral resolution at room temperature or with minimal cooling (FWHM of 150 eV at 6 keV at 0 °C). To realise this goal, a custom Application-Specific Integrated Circuit (ASIC), known as RIGEL, has been developed. The sensor and ASIC are interconnected via gold bump bonding. This project is a collaborative effort involving several institutions: the Italian National Institute for Astrophysics (INAF) is responsible for overall coordination and device testing; the National Institute for Nuclear Physics (INFN) is handling sensor design; the Bruno Kessler Foundation (FBK) is in charge of sensor production; the Polytechnic University of Milan and the University of Pavia are responsible for ASIC development; and the Karlsruhe Institute of Technology (KIT) leads the critical ASIC-to-sensor gold bump bonding integration [157].

The PixDD sensor employs lateral depletion technology, which is characteristic of Silicon Drift Detectors, significantly reducing the size of the collecting anode and, therefore, lowering the detector capacitance. This low capacitance is vital for performing soft X-ray spectroscopy with Fano-limited energy resolution. Fabricated on a 450 µm-thick $n-$type $\langle 100 \rangle$-silicon substrate with approximately 9k Ω·cm resistivity, each pixel measures 300 µm per side, and the PixDD detector comprises 256 pixels arranged in a 16×16 matrix, as illustrated in Figure 13.8 [157].

The pixel anodes consist of a circular n^+-type implant with a diameter of 40 µm, covered by a similarly sized aluminium pad. A 28 µm circular opening in the sensor's SiO_2 passivation layer is prepared for gold-stud bump bonding connections with the RIGEL ASIC. Two p^+-type cathodes surround the anodes, with an inner low-voltage cathode (CLV) and an outer high-voltage cathode (CHV). The gap between these cathodes is specifically designed at 60 µm to maximise the punch-through threshold voltage, which is crucial for sensor stability. An additional $n-$spray implant manages the oxide charge within this sensitive region, which is further protected from environmental influences by a positive field plate connected to the CLV [157].

On the opposite side of the bonding pads, a thin, shallow p^+-type implant covers the sensor array, acting as an entrance window for incoming X-ray photons. This window implant is biased at -70 V to ensure full depletion, while CLV and CHV are biased at -32 V and -39 V, respectively. Through specialised production techniques at FBK, sensor leakage currents are exceptionally low, typically below a few hundred pA cm^{-2}, allowing PixDD pixels to exhibit sub-picoampere leakage currents even at room temperature. Combined with an anode capacitance below 35 fF, this allows for Fano-limited energy resolutions without active cooling. However, this result also depends on the front-end ASIC and the coupling between the two devices [157].

The integration of the Pixelated Silicon Drift Detector (PixDD) sensor with the RIGEL ASIC was achieved using a specialised gold bump bonding process. This method was chosen due to several unique requirements and constraints associated with the project, including the sensor's

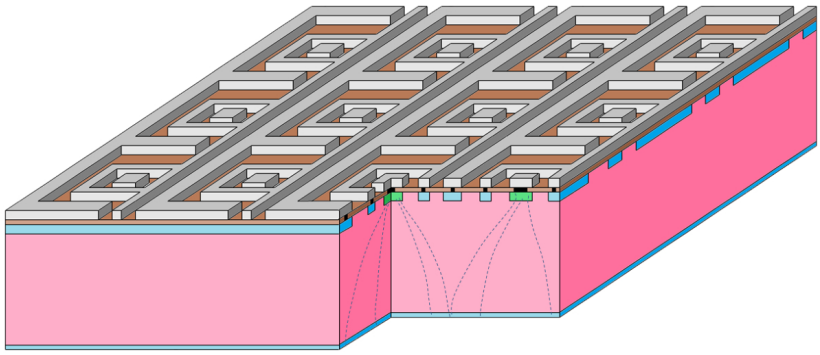


Figure 13.8: Layout of the PixDD sensor illustrating the 16 × 16 pixel arrangement [157].

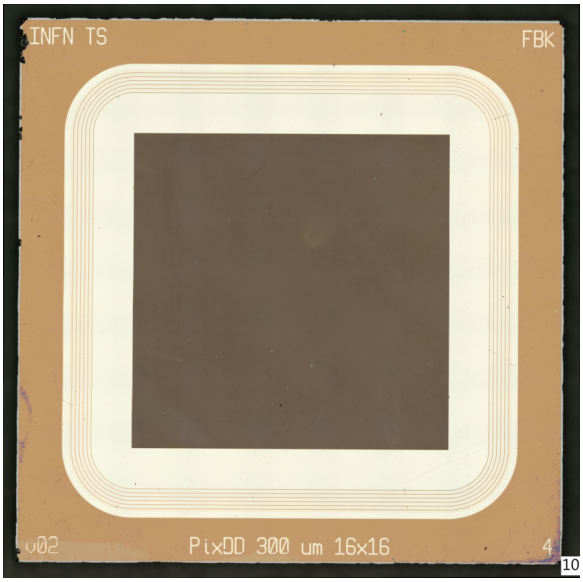
fragile detection window, asymmetrical chip layouts, and stringent mechanical reliability requirements.

One of the principal challenges was the presence of an extremely delicate, 0.5 µm thick X-ray-sensitive window, which is located on the opposite side of the bonding pads. During bump formation and subsequent flip chip bonding, any friction or pressure could irreversibly damage this window. Special precautions were thus necessary. To address this issue, the sensor was carefully positioned on a ceramic substrate, secured by vacuum suction with an intermediate polished cleanroom-compatible paper, eliminating the risk of direct frictional damage.

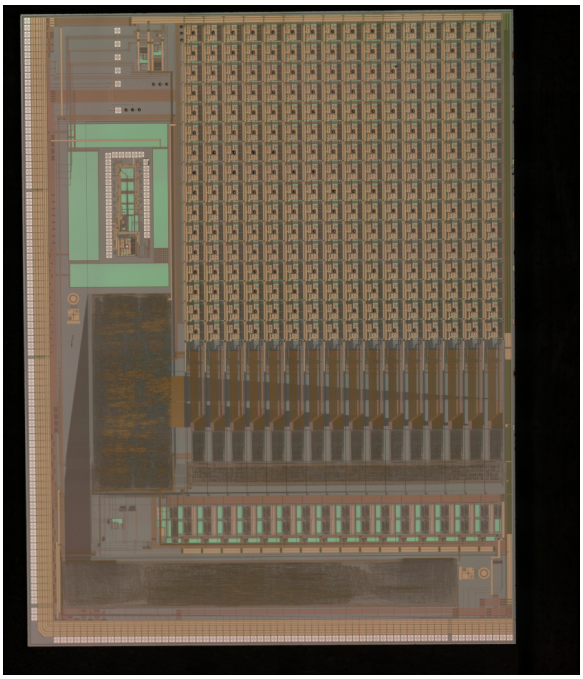
Another significant complexity arose from the highly asymmetrical geometry between the ASIC and sensor, which were not only different in size but also offset relative to each other. This asymmetry poses alignment difficulties during flip chip bonding and makes it complicated to distribute the bonding forces uniformly, which is critical for achieving the high mechanical reliability essential in space applications. Figure 13.9 illustrates the delicate sensor window and provides an overview of the geometric differences between the sensor and ASIC.

Gold bump bonding involves creating small gold studs directly on the bonding pads of the chip or sensor. For the ICCON Gold Stud Bonder at IPE, this approach requires modifying the wire bonding technique, mentioned in the Subsection 13.2, into a bump bonding procedure. In this process, a gold ball bond is made on the pad, but unlike normal wire bonding, no second bond is made; instead, the wire is cut after the first bond, leaving a gold stud bump on the pad. By repeating this on all required pads (either on the chip, the sensor, or both), an array of gold bumps is formed. These bumps can then be aligned and pressed to the corresponding pads on the mating component to form electrical connections, typically via thermocompression. The ICCON bonder allows precise control of bump dimensions by adjusting wire diameter, bonding force, ultrasonic energy, and wire length, producing bumps as small as 30 µm in diameter with pitches between 50–100 µm. The key advantage of this technique is that it avoids complex lithographic processes and wafer-level deposition, making it ideal for rapid prototyping and cost-sensitive small-scale projects.

The PixDD sensor and RIGEL ASIC both received gold bumps on their respective bonding pads. Figure 13.10 provides images of the bumped sensor and detailed views of these gold studs. Optimisation of the gold bumping process included adjustments to bond force to 20 g,



(a) The sensitive X-ray window of the PixDD sensor, which requires special protection during bump bonding.



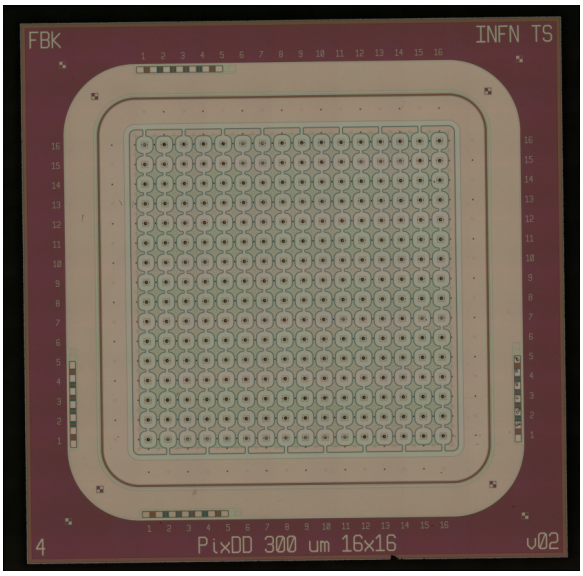
(b) The RIGEL ASIC illustrating geometrical asymmetry with respect to the sensor, complicating alignment during flip chip bonding.

Figure 13.9: Sensor window protection and ASIC geometry challenges.

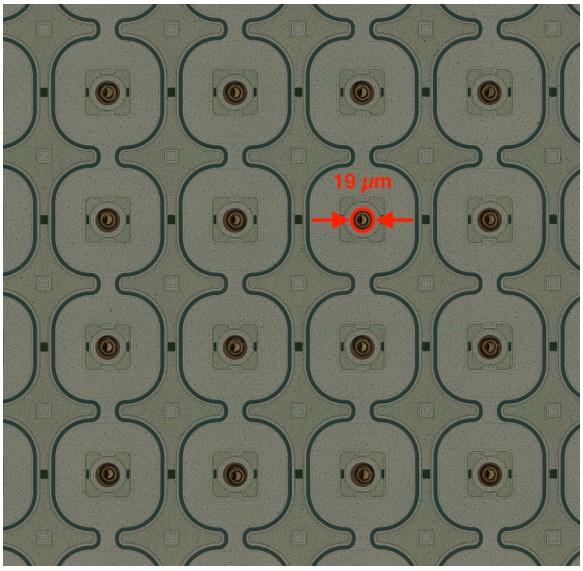
ultrasonic current to 70 mA, bond time to 5 ms, free air ball (FAB) size to 1.42 mils, and electronic flame-off (EFO) current to 30 mA, all using 18 µm diameter gold wire. This ensured consistent high-quality bump formation tailored specifically for this challenging application.

Following bump formation, flip chip bonding was performed using the Finetech Fineplacer Femto Die Bonder. Its sub-micron alignment accuracy was critical for addressing the sensor's and ASIC's geometric asymmetry. In the flip chip bonding step, thermocompression bonding was utilised. Under precisely controlled temperature of 185 °C and force conditions, gold studs on both sides diffused and fused, forming a robust gold-gold metallic bond. This method provided excellent electrical performance and mechanical reliability, achieving alignment accuracies better than 1 µm and a controlled gap of approximately 23 µm between the sensor and ASIC after bonding. Figure 13.11 illustrates the successfully bonded PixDD sensor and ASIC assembly.

Despite significant challenges, including protecting the delicate sensor window, handling asymmetric alignment, and ensuring uniform mechanical stress distribution, the bonding process was successfully optimised to meet all design specifications. The gold bump bonding, combined with precise flip chip thermocompression, has been proven to be effective in meeting the stringent demands of the PixDD sensor project, demonstrating reliability, high mechanical strength, and excellent electrical connectivity suitable for space-based detector applications.



(a) PixDD sensor showing the array of gold bumps on the bonding pads side.



(b) Close-up view detailing the dimensions and uniformity of the gold bumps.

Figure 13.10: Gold bumps produced by ICCON Gold Stud Bonder on the PixDD sensor.

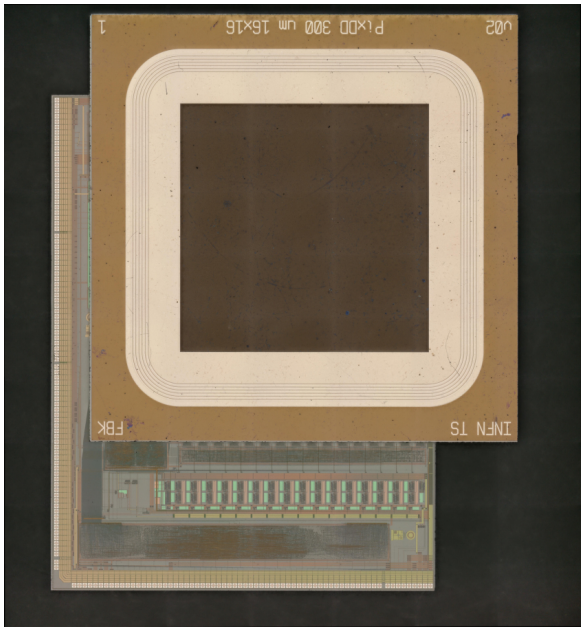


Figure 13.11: Completed PixDD sensor assembly after gold stud bump and flip chip bonding.

14 Compressed Baryonic Matter experiment at the future Facility for Antiproton and Ion Research

14.1 Introduction to the Facility for Antiproton and Ion Research (FAIR)

The Facility for Antiproton and Ion Research (FAIR) is a state-of-the-art international research centre currently being constructed in Germany, Darmstadt, at the Gesellschaft für Schwerionenforschung GmbH (GSI). FAIR is designed to advance the understanding of matter under extreme conditions by replicating environments similar to those found in neutron stars or during the early moments of the universe. This creates a base for the deep research of nuclear physics, the behaviour of matter at high densities, and the fundamental properties of the strong force [158, 159]. Construction of FAIR is progressing at a rapid pace, with the facility expected to become fully operational by 2028. The project is a major upgrade to the existing GSI facilities, combining modern infrastructure and advanced technologies. Figure 14.1 shows an aerial view of the construction site as of March 2024, highlighting the scope of the project, its scale and complexity. Figure 14.2 presents a schematic diagram of the planned facility, including both the existing and future infrastructure. Together, these images demonstrate the ambition and scale of the existing and future FAIR projects.

The scientific programme at FAIR is diverse, covering areas such as hadron and nuclear physics, atomic and plasma physics, materials research, and radiation biophysics. FAIR also contributes to applied sciences, including novel medical treatments and space science. Upgraded FAIR will be able to provide particle beams of chemical elements and their ions, as well as beams of protons and antiprotons. The facility's accelerator system will deliver high-intensity primary beams, enabling a wide range of experiments. Protons with kinetic energies of up to 29 GeV, uranium beams with energies of up to 11 GeV, and beams of nuclei with $N = Z$ and energies of up to 15 GeV will be produced using the SIS100 synchrotron. The accelerator tunnel has also been designed to accommodate a second synchrotron, the SIS300, offering the potential for future expansion [162].



Figure 14.1: Aerial view of the FAIR construction site, March 2024. The image highlights the integration of new infrastructure with existing GSI facilities to support future scientific research [160].

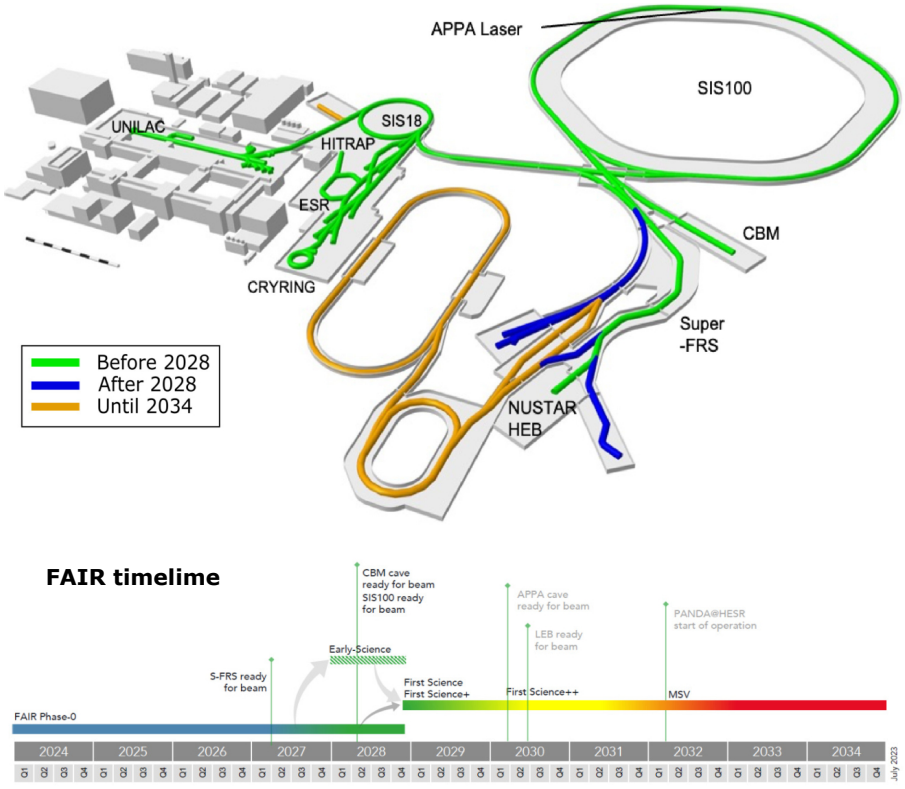


Figure 14.2: Schematic of the planned FAIR facility, including the development timeline and its integration with existing GSI infrastructure [161].

14.2 The Compressed Baryonic Matter (CBM) experiment

The Compressed Baryonic Matter (CBM) experiment is one of the flagship projects of the FAIR facility. Its primary objective is to explore the properties of dense nuclear matter, particularly at high baryon densities and moderate temperatures. The experiment aims to discover phenomena such as the transition to quark-gluon plasma and the critical point of the QCD phase diagram.

To achieve these goals, CBM employs an array of detectors designed to handle the massive data rates generated during high-energy heavy-ion collisions. Each detector is tailored to perform specific tasks, such as tracking particles, identifying their types, or measuring their energy. Together, these systems allow CBM to study rare processes of dense baryonic matter [163, 159].

Figure 14.3 provides a visual representation of the CBM experimental setup, including its integration with the High-Angle Di-Electron Spectrometer (HADES). The setup features multiple specialised detectors, each contributing to a specific aspect of the experiment.

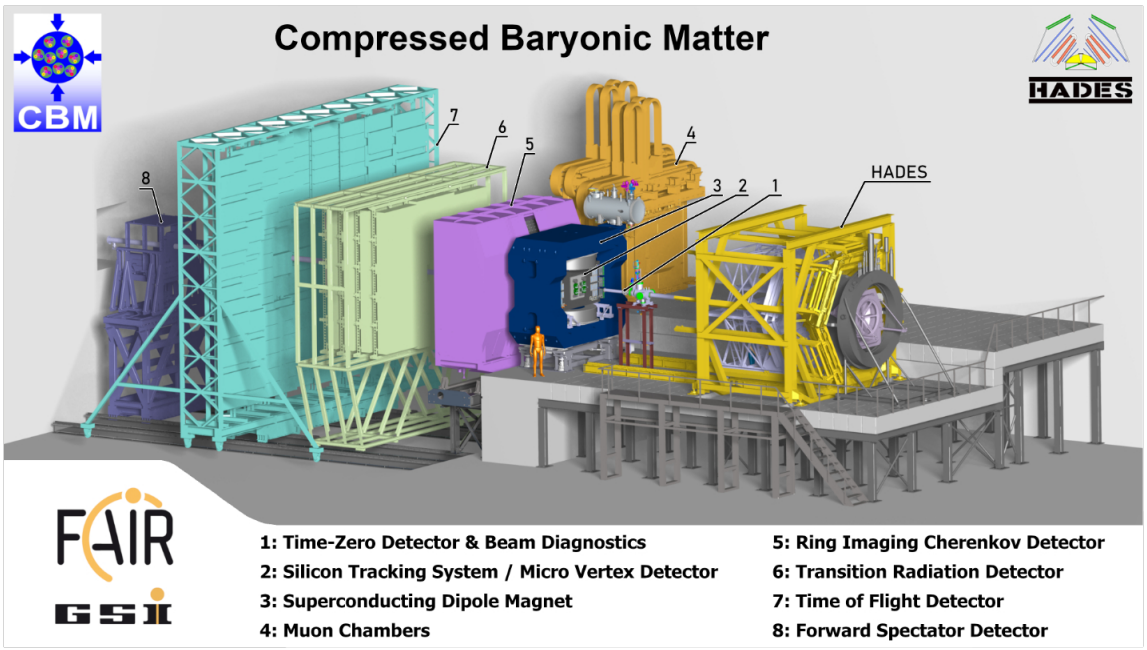


Figure 14.3: Illustration of the CBM experimental setup, including HADES (High-Angle Di-Electron Spectrometer), at FAIR. The setup includes a range of advanced detectors, such as the Silicon Tracking System (STS), Ring Imaging Cherenkov Detector (RICH), and Muon Chambers. Each detector plays a unique role in capturing and analysing data from high-energy collisions [164].

The CBM experimental setup includes several unique detectors, where the subsystem is optimised and integrated with the high-intensity particle beams provided by FAIR:

- the **Time-Zero Detector (T0)** provides beam diagnostics and precise event timing;

- the **Silicon Tracking System (STS)** together with the **Micro Vertex Detector (MVD)** serve as the core detector, offering high-precision tracking of charged particles with lightweight silicon sensors and advanced readout electronics;
- the **Ring Imaging Cherenkov Detector (RICH)** identifies particles by measuring their velocities through Cherenkov radiation;
- the **Transition Radiation Detector (TRD)** and the **Time of Flight (ToF) Detector** work together to provide complementary particle identification capabilities;
- the **Muon Chambers (MUCH)** enable the detection of rare muon events;
- the **Superconducting Dipole Magnet (MAGNET)** generates the magnetic field required for momentum analysis of charged particles;
- the **Forward Spectator Detector (FSD)** and the **Neutron Calorimeter (NCAL)** identifies fragments from collisions and monitors reaction centrality.

Central to the CBM experiment is the Silicon Tracking System (STS), which serves as the backbone of the detector array. The STS is responsible for reconstructing the trajectories of charged particles with high precision, providing the foundational data required for all subsequent analyses. This system uses lightweight silicon microstrip sensors, which are optimised for minimal material interference and high spatial resolution. These sensors are paired with advanced read-out electronics, ensuring efficient and accurate data acquisition. The STS design balances both durability and performance, making it a crucial component of the CBM experiment [158, 159].

14.3 The Silicon Tracking System (STS)

The Silicon Tracking System (STS) is a critical component of the CBM experiment, designed to reconstruct the trajectories of charged particles with exceptional precision. This capability is central to the experiment's goal of exploring the properties of dense nuclear matter. The construction of the STS requires a combination of advanced engineering techniques and precise assembly processes [165].

The STS must meet stringent requirements to operate effectively. It is designed for free-streaming readout at beam-target interaction rates of up to 10 MHz. It also should manage approximately 700 tracks in aperture per interaction, requiring high granularity. Moreover, the application demands for a low material budget, approximately 2–8% X_0. Its spatial resolution should be less than 30 µm, with timing accuracy under 5 ns, and be capable of recording signals with an amplitude of 15 fC/5 *bit* in free-streaming mode [166]. The full design of the STS, including its integration within the CBM dipole magnet, is shown in Figure 14.4.

The final stages of construction of the STS began in 2023 and involved the collaboration of multiple project groups. The detector's core part is double-sided silicon microstrip sensors, which are critical for tracing the trajectories of particles produced during high-energy collisions (see Subsection 14.4.2). Each sensor undergoes rigorous testing to assess its electrical properties,

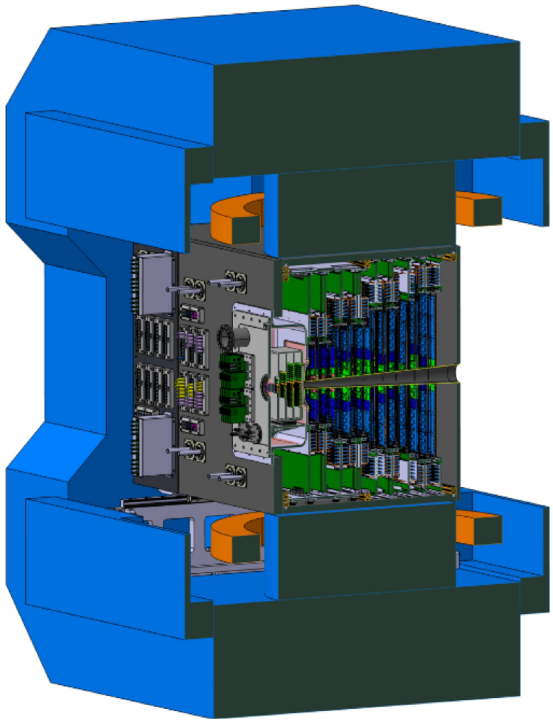


Figure 14.4: Design of the Silicon Tracking System (STS), including its placement inside the CBM dipole magnet (blue). The STS, along with the Micro Vertex Detector (MVD), forms the central tracking system of the CBM experiment [165].

such as noise levels and signal clarity, as well as its mechanical robustness. Sensors that meet the performance standards are assembled into modules. This process requires precise alignment to ensure the detector achieves its required resolution. Details on module assembly can be found in Section 14.4.

Once the modules are assembled, they are mounted onto structures called ladders. These ladders provide mechanical support for the modules while keeping the material budget low, ensuring minimal interference with particle trajectories. The ladders are then installed into the STS detector frames, known as half-units, which form the complete STS system. This process demands high precision to guarantee perfect alignment of all components. Figure 14.5 illustrates one of the first STS detector ladders, which comprises 10 assembled modules.

Despite significant progress, the construction of the STS has faced numerous challenges. Ensuring precise alignment of sensors and modules throughout the assembly process is one of the major difficulties. This has been addressed by using advanced alignment tools and metrology techniques to achieve the required accuracy (see Subsection 14.4.6). Another significant challenge is managing the heat generated by the readout electronics. To prevent thermal stress or misalignment, a custom cooling system has been developed, efficiently dissipating heat while maintaining the system's stability. This aspect of the design was developed at GSI and is detailed in related PhD theses [168, 169].

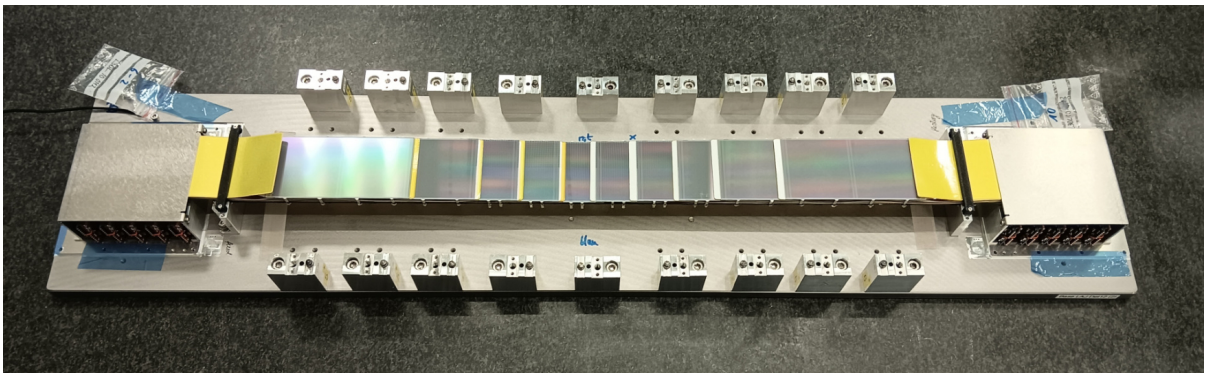


Figure 14.5: One of the first fully assembled STS detector ladders. This ladder consists of 10 modules, illustrating the precise integration required for successful operation. Modules consist of sensors, microcables with shielding and Front End Boards [167].

The assembly of STS modules is one of the key scientific tasks undertaken at the Institute for Data Processing and Electronics (IPE). Coordination among teams and consistent quality control have been essential to overcoming the complexities introduced by the involvement of multiple production sites. Regular collaboration and thorough documentation have been instrumental in ensuring a smooth assembly and integration process [166, 159].

14.4 Assembly of STS modules at KIT — infrastructure and workflow

The Karlsruhe Institute of Technology (KIT) plays an important role in the CBM experiment as one of the key production centres for assembling the Silicon Tracking System detector modules. KIT is responsible for producing approximately half of the STS modules. The assembly takes place in the cleanroom facility at KIT's Institute for Data Processing and Electronics (IPE), which is equipped with advanced tools to handle fragile components with precision and to fulfil complicated, multiple-step production. Key machinery includes wire bonders and vacuum pick-up systems, designed to ensure accurate alignment and reliable connections [158, 170]. Figure 14.6 shows the Wire Bonder Bondjet BJ653, the centrepiece of the bonding process. This bonder is capable of making robust wire bonds with high precision, which is crucial for assembling the Front-End Boards (FEBs) and connecting electronic components.

The assembly process involves multiple stages, each requiring great attention to detail. Electrical and mechanical tests are performed at various points to ensure that each module meets strict quality standards before integration [171]. The main steps in the production process are outlined below and described in detail:

1. **Cleaning and preparation.** The process begins with cleaning and preparing the silicon sensors and FEBs. This involves placing the printed circuit boards (PCBs) in an ultrasonic bath to remove residues, such as solder flux, from the assembly process. This step ensures that no contaminants interfere with the bonding process or the electrical connections.

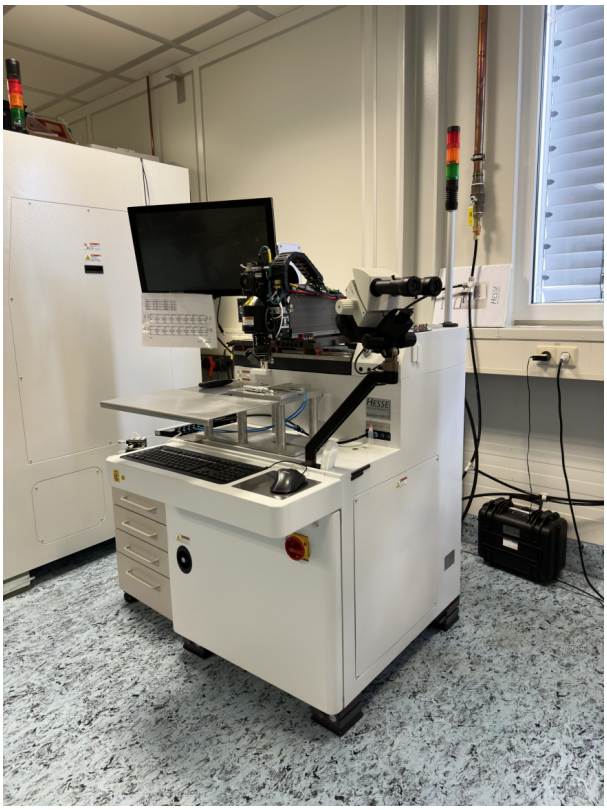


Figure 14.6: The Wire Bonder Bondjet BJ653, located in the clean room of IPE, is used for wire bonding during the assembly of the STS modules.

2. **Die and wire bonding of LDOs to the FEBs.** Low-dropout regulators (LDOs) are die and wire-bonded onto the FEBs. After an initial electrical test to confirm their functionality, the wire bonds are protected using a glob-top material. This protective coating prevents mechanical damage and shields the bonds from environmental influences (see Subsection 14.4.1).
3. **TAB of microcables to ASIC chips.** The microcables are attached to the ASIC chips using the Tape Automated Bonding (TAB) process. This builds the Chip-Cables (CC), which then provide the connection between the sensors and the FEBs. Quality control at this stage includes optical inspections to verify alignment and electrical tests using a PogoPin setup to ensure continuity and signal integrity (see Subsection 14.4.4).
4. **Die and wire bonding of ASICs to the FEBs.** The ASIC chips, now bonded to the microcables, are die and wire-bonded to the FEBs. The completed assemblies undergo another round of testing to confirm the integrity of the connections and ensure proper communication between the ASICs and the FEBs (see Subsection 14.4.5).
5. **TAB of cables to the silicon sensor.** The cables with bonded ASICs and FEBs are attached to the silicon sensors using the TAB process. This step integrates the core components of the STS module, creating a functional data acquisition system (see Subsection 14.4.5).

6. **Gluing of shielding layers and spacers.** To protect the module from electromagnetic interference and ensure mechanical robustness, shielding layers and spacers are glued in place. This step is fundamental for maintaining data integrity and preserving the physical structure of the module during operation.
7. **Full module test.** Once assembly is completed, the module undergoes a comprehensive set of tests, including electrical performance checks, mechanical stability assessments, and data transfer validation. These tests ensure that the module meets the performance criteria necessary for operation in the CBM experiment.
8. **Storage and preparation for transport.** The final step involves securely storing the completed modules and preparing them for transport to the integration site. Proper packaging and handling procedures are followed to prevent any damage during transit and to maintain the modules' integrity.

14.4.1 Preparation and die and wire bonding of the LDOs to the FEBs

The assembly process begins with the cleaning and preparation of the silicon sensors and FEBs. The PCBs are cleaned in an ultrasonic bath to remove residues, such as solder flux, that may remain from the assembly process. This thorough cleaning step ensures that the subsequent bonding steps are not compromised by contamination, maintaining optimal electrical performance.

The FEB8 boards incorporate four low-noise regulators supplied by SCL India, which generate stable 1.2 V and 1.8 V power for the ASICs and other electronic components. Following the cleaning process, these LDOs are bonded to the FEBs. This step involves two bonding methods: die bonding, where the LDO chip is attached to the board using adhesive, and wire bonding, which creates fine wire connections between the LDO and other components on the FEB. Custom-designed, low-noise LDOs are employed to stabilise the voltage supplied to the readout electronics, thereby ensuring reliable operation of the entire system. Figure 14.7 shows an LDO wire bonded onto an FEB. The precision required in this process can be seen in this illustration.

To operate one silicon sensor module, two FEB8 boards are required, one for each side of the sensor. Each FEB8 board is equipped with eight STS-XYTER (Silicon Tracking System-X-Y-Time-Energy Read-out) ASICs, which handle the readout and digitisation of the signals. Each ASIC provides up to five Low-Voltage Differential Signalling (LVDS) data signals (UpLinks), each capable of transmitting 10 Mbits/s per link. This data rate ensures that the system can handle the high interaction rates characteristic of the CBM experiment.

To maintain signal integrity, the isolation of the LVDS signals between the readout board (ROB) and the sensor bias potential is achieved using high-voltage ceramic coupling capacitors. These capacitors ensure that the FEB operates at the required bias potential while the data acquisition (DAQ) system remains grounded. Data signals are accessed through Zero Insertion Force (ZIF) connectors and shielded flat cables, which further reduce noise and interference. The silicon sensors are biased by connecting a high-voltage source between the n-side and p-side

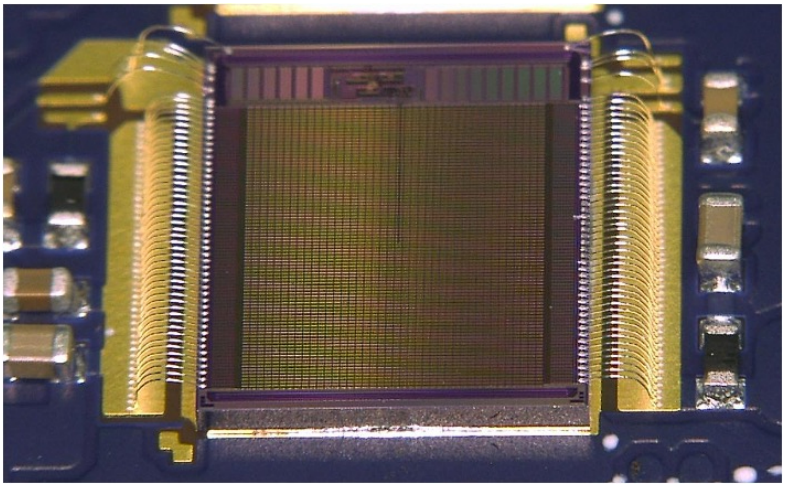


Figure 14.7: Low-dropout regulator (LDO) wire-bonded onto the Front-End Board (FEB). The LDO ensures a stable voltage supply for the ASICs, supporting consistent readout performance by minimising voltage fluctuations. The wire bonds are carefully aligned and secured to avoid mechanical stress and ensure reliable operation during data acquisition [159].

grounds of the sensor to ensure the correct operating conditions [172]. Figure 14.8 presents a complete FEB8 board with wire-bonded LDOs, demonstrating the layout and placement of key components.

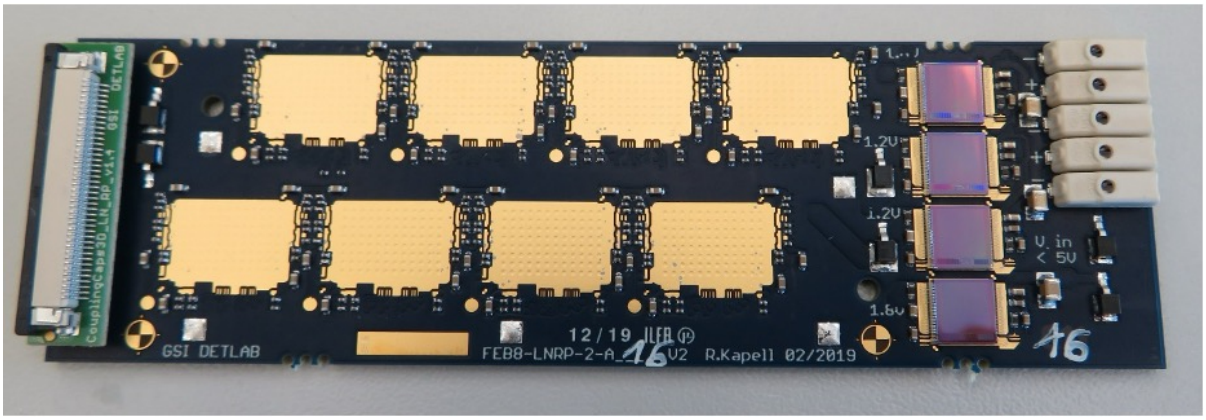


Figure 14.8: Front-End Board (FEB8) with wire-bonded low-dropout regulators (LDOs) on the right. In the middle of FEB, eight metal plates are prepared for ASIC bonding.

14.4.2 Silicon microstrip sensor

The STS employs hybrid strip sensors, which, as a design strategy, exist in various configurations depending on the segmentation strategy and signal routing approach. In all cases, a high-resistivity silicon sensor is bump-bonded or wire-bonded to a readout ASIC that amplifies and digitises the signal induced on each strip.

Hybrid strip sensors can be implemented as single-sided or double-sided designs. In single-sided strip detectors, only one surface of the silicon sensor is segmented with parallel strip electrodes. This side is usually operated at ground potential, while the high bias voltage is applied to the unsegmented backside. These detectors provide precise position information in one dimension and are widely used in the inner tracking regions, where particle density is high and material budget must be minimised.

To provide two-dimensional tracking information, double-sided strip detectors incorporate orthogonal or stereo-angled strip patterns on both surfaces of the sensor. This allows each particle hit to be localised in two dimensions using a single sensor element. However, the complexity of this approach is significantly higher. The strip segmentation requires isolation techniques such as p-stop or p-spray doping, and routing of signals from both sides poses mechanical and electrical challenges.

Several layout options exist for readout of double-sided detectors. In the most straightforward approach, the strips on each side are read out independently at opposite edges of the sensor. More advanced layouts introduce a second metal routing layer on the backside of the sensor to direct the signals to a common edge, thus reducing the number of external readout connections. However, these more complex layouts come with several disadvantages, including increased fabrication complexity due to additional metal and via layers, higher parasitic coupling and crosstalk between signal and routing lines, and a strong dependence of electrical performance on the thickness of the isolation layers [37, 35].

As a workaround, some systems employ stereo modules composed of two single-sided sensors mounted one under the other with a small relative rotation, typically $\sim 6^\circ$, as in CMS [37]. This configuration achieves similar performance to double-sided detectors but at the cost of increased material and overall module thickness. As a result, it is not suitable for all applications.

Double-sided detectors are elegant in that they combine two-dimensional resolution with relatively low channel count. However, their use is limited by high fabrication cost, increased risk of mechanical damage, and susceptibility to ghost hits at high occupancy. The STS employs double-sided, AC-coupled silicon microstrip sensors with stereo angles of -7.5° on the p-side and thicknesses between 280 µm and 320 µm. These sensors are available in lengths of 2 cm, 4 cm, 6 cm, and 12 cm, and they are connected to microcables with lengths from 15 cm to 47 cm. Figures 14.9 and 14.10 present photographs of the sensors and their layout. In contrast, the MuCh system utilises trapezoidal 3-foil gas electron multipliers (GEM) as a detecting element [173].

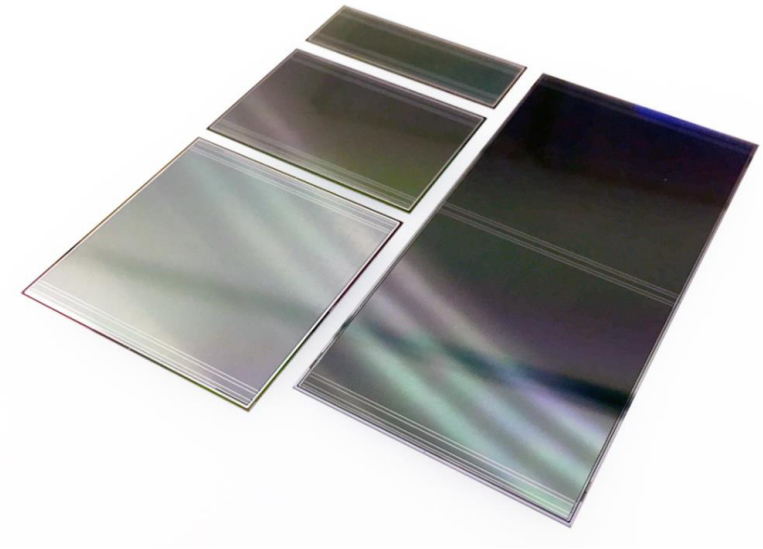


Figure 14.9: Photograph of the silicon microstrip sensors [173]. The sensors are available in lengths of 2 cm, 4 cm, 6 cm, and 12 cm to meet the requirements of ladder assembly.

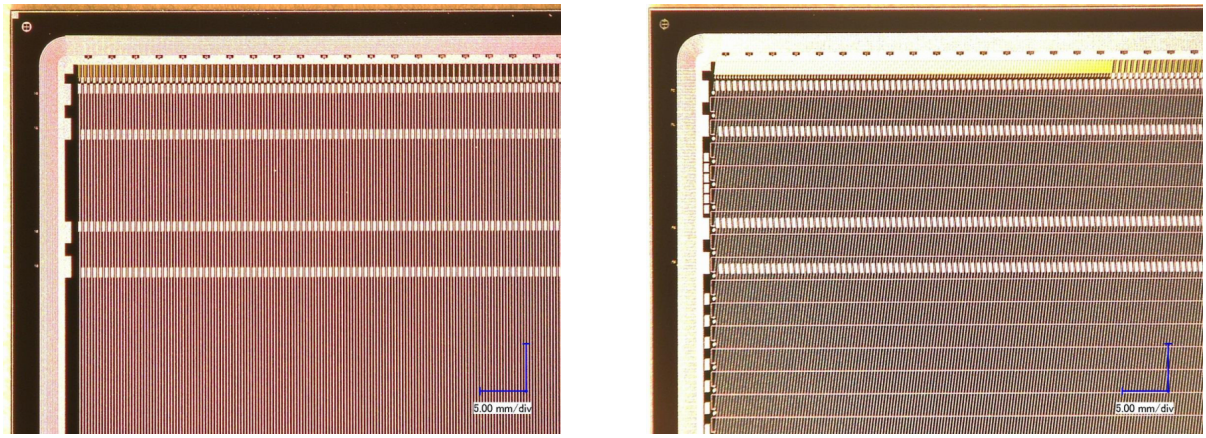


Figure 14.10: Layout of the front and back sides of the silicon microstrip sensor [173]. The figure illustrates the alternating p-side and n-side microstrips and their interconnections, showing the stereo angle of -7.5° used for precise tracking. The sensor's double-sided configuration enhances spatial resolution, which is crucial for tracking charged particles.

14.4.3 Silicon Tracking System-X-Y-Time-Energy (STS-XYTER) readout ASIC

The ASIC used for the Silicon Tracking System (STS) and the Muon Chamber (MuCh) readout is designed to meet strict performance requirements from both the sensor and back-end data acquisition system. These requirements demonstrate the design's adaptability in supporting the unique demands of both detector systems. The universal STS-XYTER ASIC is designed to operate under demanding conditions. For the STS system, a total capacitance is expected to be up to 40 pF with an average hit rate of 250 kHz per channel. For MuCh, these values are up to

50 pF and below 100 kHz, respectively. The STS has a finer channel pitch of 58 µm, while the MuCh system operates with a pitch of 116 µm. Both systems consume less than 10 mW per channel. The ASIC's dynamic range is required from 0 to 15 fC for STS, which corresponds to about 5 times the most probable charge from a MIP in 300 µm silicon (with a typical signal of 4 fC), and from 1 to 100 fC for MuCh. The time resolution for both systems is below 10 ns. The STS ASICs are needed to operate at approximately -10 °C, while the MuCh ASICs, on the other hand, are required at around -60 °C. The STS ASICs have to accept both positive and negative input polarity, which is a prerequisite for reading out signals from both sides of the sensor; the MuCh ASICs are designed for negative polarity signals only. These operational differences highlight the adaptability of the STS/MuCh-XYTER2 (SMX2) ASIC to meet the various demands of the CBM experiment [174].

The SMX2 ASIC is fabricated using the UMC 180 nm CMOS process, occupying an area of 10 mm × 6.75 mm. The key features of the chip include [174]:

- **128 analogue front-end (AFE) channels** with 14-bit time and 5-bit amplitude measurement circuit. The channels can be configured for the gain and peaking times of the shaper to accommodate both detector systems. ELT (Enclosed Layout Transistor) NMOS transistors are used to enhance immunity to Total Ionising Dose (TID).
- **Two additional test channels** are located at the sides of the ASIC. These channels provide buffered outputs of internal signals for diagnostic purposes.
- **Custom-made design**, which includes full-adjustable configuration registers, biasing blocks, calibration circuits, and a band-gap reference voltage.
- **Digital back-end**, which is clocked at 160 MHz, providing access to configuration registers, free-streaming digital readout, timestamp-wise sorting of acquired hits and successive serialisation with 320 Mbps per link or a maximum of 47 Mhits/s, together with multiple diagnostic features.

Over the last few years, the ASIC design underwent significant upgrades to improve performance and reliability. Enhancements include redesigned input amplifiers, improved shaping circuits, and updated testability features. The improved pad layout was tailored to support specific requirements in the STS module assembly [174]. The SMX2 ASIC's architecture is illustrated in Figure 14.11, including the layout of its key components and their interconnections in blocks.

The production yield for the ASICs is impressively high, ranging between 98.5% and 99.0%, with a chip-cable yield of 96%. The CBM experiment requires 14 016 ASICs for the STS and 4 968 for the MuCh detector. The production process, involving 100 wafers, each containing 360 dies, includes wafer-level testing to ensure reliability and performance [165].

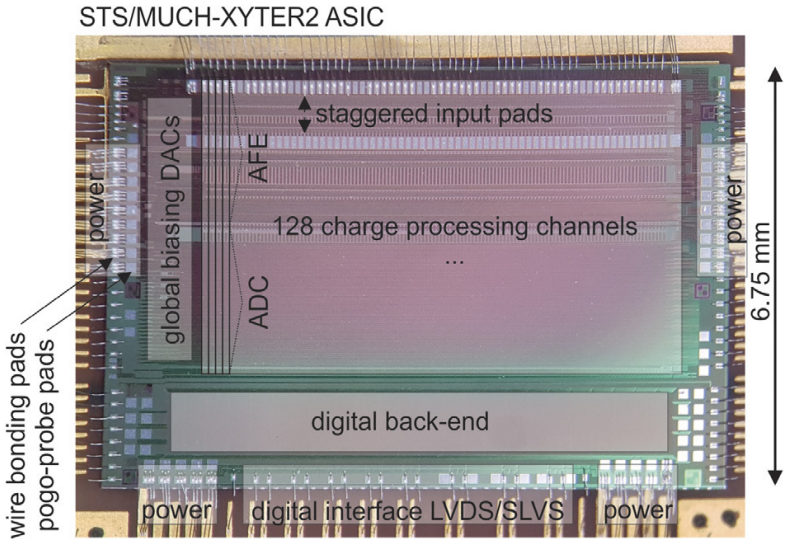
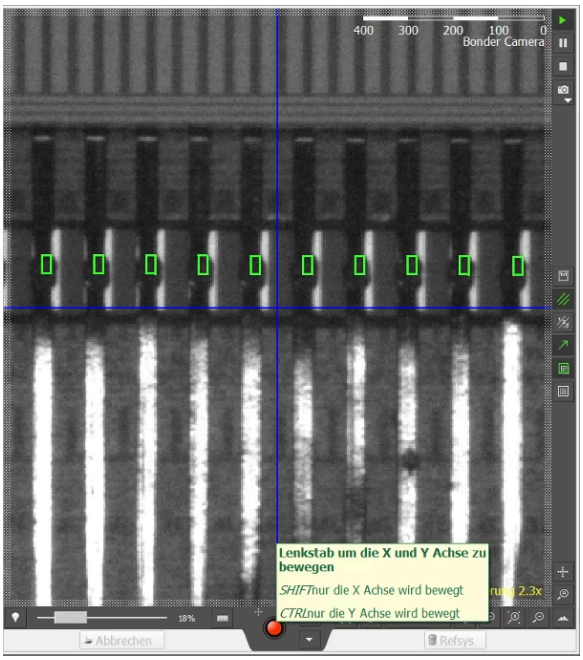


Figure 14.11: Simplified diagram of the STS/MuCh-XYTER2 ASIC (SMX2), a multichannel integrated circuit designed for the STS and MuCh detectors [174]. It shows the layout of 128 analogue front-end (AFE) channels for time and amplitude measurement circuits and the digital back-end responsible for timestamp-wise sorting and serialisation at 320 Mbps per link.

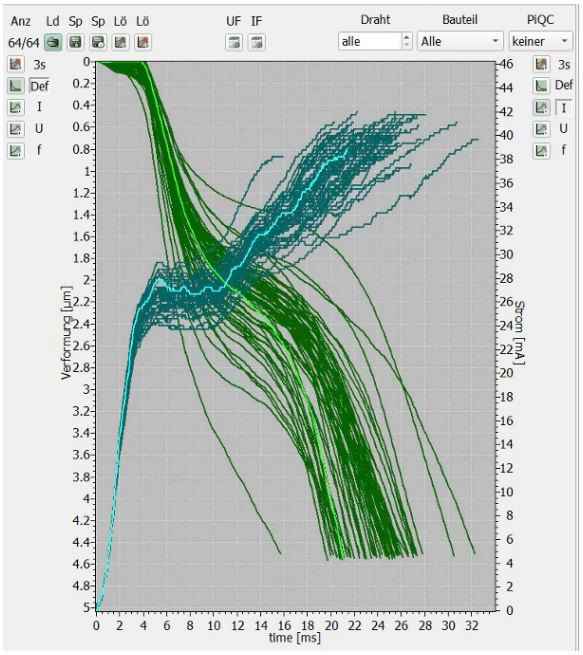
14.4.4 Tape automated bonding of microcables to ASIC

In parallel with the wire bonding of LDOs to FEBs, the STS-XYTER ASICs are tape automated bonded to aluminium microcables. More about TAB can be found in Subsection 13.1.3. This bonding technique ensures robust electrical connections between the readout chips and the microcables, which are essential for efficient data transfer. The TAB process of microcables to ASICs is executed in two phases. In the first bonding phase, the ultrasound intensity is set to 13%, with a bonding force of 18 cN. In the second phase, the ultrasound intensity is increased to 15%, maintaining the same bonding force. These parameters were optimised through rigorous testing to ensure the highest bond quality under operational conditions [170].

Figure 14.12 illustrates this process. On the left 14.12a, the bonder camera view displays the precise alignment of the microcables with the bond pads on the ASIC. This visualisation is crucial for ensuring the accuracy of the bonding process. Each bond pad is carefully aligned with the corresponding segment of the microcable, and adjustments can be made along the x- and y-axis as needed to achieve the required precision. The right panel 14.12b presents the process curve, which illustrates the dynamic parameters of the bonding process over time. Key parameters such as deformation, force, and ultrasonic energy are continuously monitored to ensure consistent bond quality. The curves provide insights into the interaction between the bonder and the materials, showing the progression of deformation in milliseconds. This data is essential for confirming that the bonds meet stringent standards for mechanical strength and electrical connectivity.



(a) Bonder camera view displaying the precise alignment of the microcables with the bond pads. The image shows the accurate positioning to ensure optimal bonding. Adjustments along the x- and y-axis were made to guarantee that each bond pad aligns with the corresponding microcable segment.



(b) Process curve indicating dynamic bonding parameters such as force, ultrasonic energy, and deformation over time. The graph illustrates how these parameters were adjusted during different phases of the TAB process to achieve reliable bonds.

Figure 14.12: TAB process of microcables to ASICs: (a) is the camera view during alignment, displaying how the bond pads and cables are positioned, (b) is the process curve shows the dynamic parameters measured during bonding. This step ensures that the bonds are both mechanically strong and electrically reliable.

14.4.5 Chip-cable to FEB and sensor assemblies

The assembly process for attaching chip-cables to the FEB and sensor involves a carefully coordinated sequence of steps to ensure precise positioning and functionality of electrical connections. Four chip-cables are attached in parallel to the FEB using a custom-designed "Chip-Cable to FEB Mounter" (CC2FM), which is a specialised tool that facilitates accurate placement and bonding. The STS-XYTER chips are positioned on a mechanical tray that defines their exact placement on the FEB. Once aligned, the chips are picked up, and the FEB, pre-coated with conductive glue, is placed onto the mechanical jig. The dies are then lowered onto the FEB to create a conductive bond. The CC2FM is mounted on a heater to cure the conductive glue, ensuring that the bond is both secure and electrically reliable.

Following this, wire bond connections between the STS-XYTER chips and the FEB are established using the Bondjet BJ653 wire bonder. Once the wire bonds are in place and tested, they are protected by the application of glob-top material to shield them from mechanical

stress, prevent electromagnetic interference and ensure the integrity of the signals during data transmission. After completing these steps, the same sequence is performed for the second row. As the first FEB is ready, the same procedure is repeated for the second FEB, which is used for the reverse side of the module.

In the next stage of the assembly, the loose ends of the chip-cables are bonded to the n-side and p-side of the silicon sensor. The connectivity of the bonds is thoroughly checked through electrical tests to ensure that the connections are free of defects. Once the verification is complete, the TAB bonds are reinforced with epoxy glue to provide final mechanical protection and long-term durability [170].

14.4.6 Quality assurance and storage

Quality assurance (QA) is an integral part of the assembly process to ensure the reliability and performance of each component before integration. QA tests for the STS-XYTER ASICs are conducted at GSI using a systematic routine before the chips are selected for module assembly. Two types of tests are performed:

- **Wafer-level testing** is conducted using a probe station to validate the performance of ASICs at the wafer stage.
- **PogoPin** setup simultaneously tests ASIC for electrical connectivity and functionality.

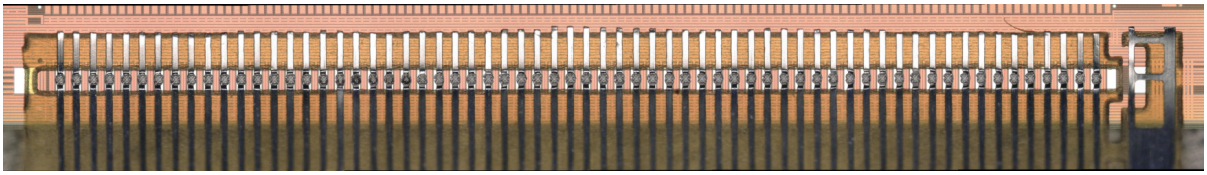
In 2023, significant progress was made in preparing for the final STS module production at GSI's Detector Laboratory (DTL) cleanroom. Additionally, KIT was equipped with new hardware and software tools to monitor the production and assembly processes and to evaluate the module performance. The testing station includes a power supply and GBTx emulator boards, which are connected to the ASIC test socket on one side and to the PC via Ethernet on the other side. A PostgreSQL database was implemented for comprehensive data logging, accessible through a web interface. Communication with the GBTx emulator boards is implemented using IP Bus Drivers, and the testing software is provided by GSI. Currently, two testing stations (at GSI and KIT) are connected to this central database, and a fourth station is planned for wafer-level testing of XYTER (SMX2) ASICs at GSI. To streamline data tracking, QR code-based standards for labelling microcable sets and modules were adopted. Figure 14.13 shows the quality assurance testing setup in the IPE KIT cleanroom [166].

Quality assurance is critical throughout the assembly process. Mechanical stress tests are conducted to simulate operational conditions and verify the robustness of the modules, typically performed at GSI towards the end of the assembly [158, 159]. Electrical tests, in contrast, are performed at each step of the assembly to ensure that connections are functioning as intended, while optical inspections identify potential defects or misalignments. The optical inspection of the bonding process and manual pull test is illustrated in Figure 14.14. The microscope image 14.14a shows the bond pads and their alignment with the microcables. The pull test results 14.14b provide an evaluation of the bonding quality, where the remains of the bond feet indicate successful adhesion and strength. This dual-step inspection approach, in the

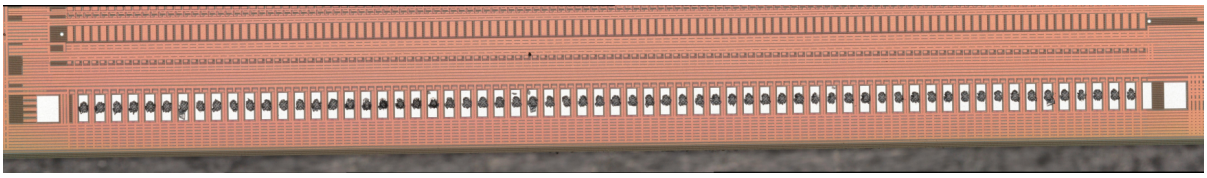


Figure 14.13: Quality assurance testing setup in the IPE KIT cleanroom, used for verifying FEB assemblies and ensuring they meet performance standards before integration [166].

bonding process adjustment stage, ensures that the bonded connections are robust enough to withstand operational conditions without compromising performance.



(a) Microscope image showing the optical inspection of the TAB process. Precise bonding is essential for ensuring electrical conductivity and mechanical stability.

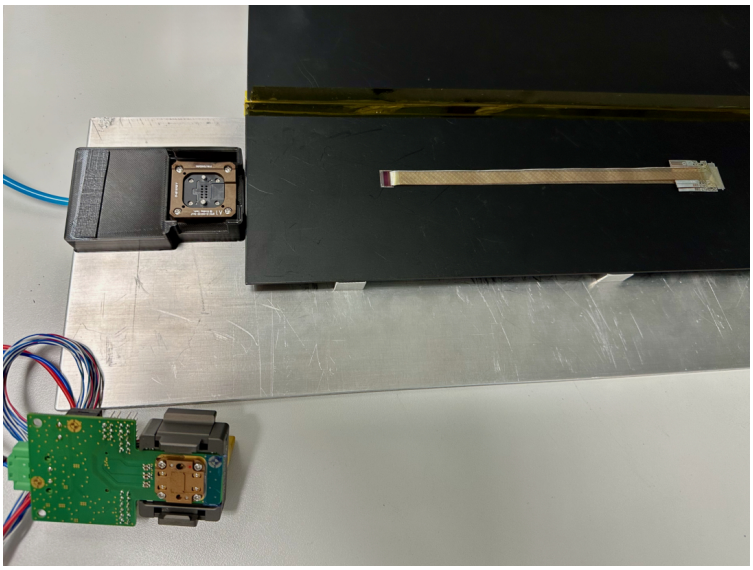


(b) Remains of the bond feet observed under the microscope after a manual pull test. This test assesses the mechanical robustness of the bonds to ensure they meet operational stress requirements.

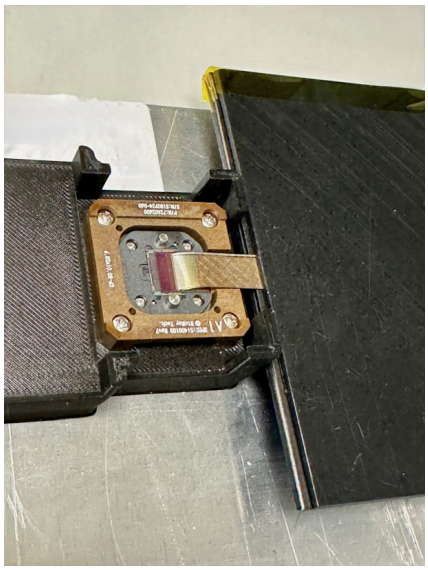
Figure 14.14: Optical inspection of the microcable bonds to the ASIC: (a) shows alignment and integrity of the wire bonds as visualised through a microscope, (b) shows a post-test image of bond feet integrity after the manual pull test to validate adhesion and mechanical resistance.

Electrical connectivity tests, a vital part of the QA process, are performed using a PogoPin setup, as shown in Figure 14.15. This tool enables early detection of faulty connections, allowing for corrections before the full integration of components. The tests are performed at

multiple stages, such as the second control of ASICs after delivery and the test of chip-cable connection after TAB.



(a) Overview of the PogoPin setup, displaying chip-cable and both parts of the testing device.



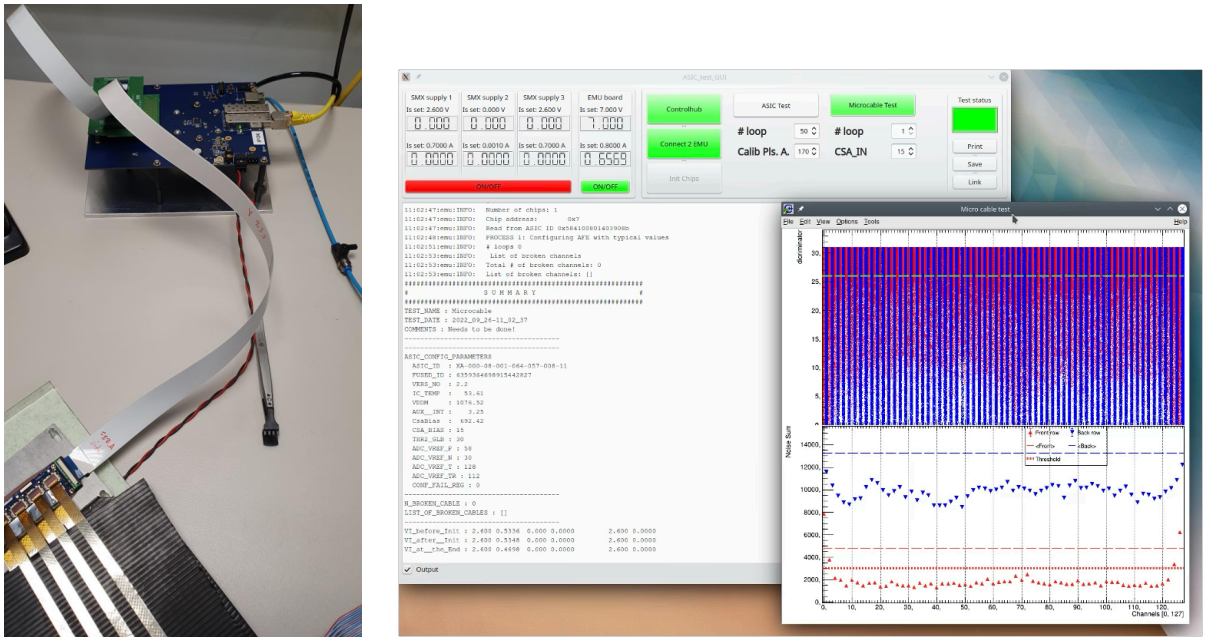
(b) Close-up of the PogoPin interface, showing alignment and contacts.

Figure 14.15: Electrical connectivity testing of the chip-cable assembly using the PogoPin interface: (a) shows the setup with the mounted module; (b) displays the precise contact points.

To date, over 5000 ASIC tests have been performed across three different stations, with more than 3000 microcable tests recorded for assembly preparation. Each ASIC undergoes calibration, often more than once, and the results, including calibration plots, are accessible via the web database. The calibration process of ASICs involves tuning the Analog-to-Digital Converter (ADC) gain and verifying threshold uniformity, along with assessing noise performance. The comprehensive testing procedures ensure the robust performance of each module. The parallel operation at GSI and KIT enables efficient data exchange and regular experience exchange [167].

The die and wire bonding steps are critical for ensuring stable connections and the overall performance of the FEBs. Multiple rounds of optical inspections and electrical tests, such as PogoPin continuity checks, are performed to verify the integrity of the connections and the functionality of the board. These tests guarantee that each FEB meets the required performance specifications before being integrated into the full module assembly. Figure 14.16 illustrates the setup and interface used for testing the Front-End Board (FEB) with bonded chip-cables. The left panel 14.16a shows the testing station where the FEB is connected to the EMU board using a flat cable. The setup includes a power supply and a data acquisition system, enabling real-time analysis of the chip-cable to FEB assembly's electrical performance. The right panel 14.16b presents a screenshot of the GUI used to display the results of the tests. The GUI provides essential data, such as signal quality, broken channels, and calibration metrics,

ensuring that any deviations from expected values can be identified and promptly corrected. This interface facilitates the analysis of collected data and supports the efficient documentation of test outcomes.



(a) Testing of FEB with bonded chip-cables.

(b) Screenshot of the testing GUI displaying results.

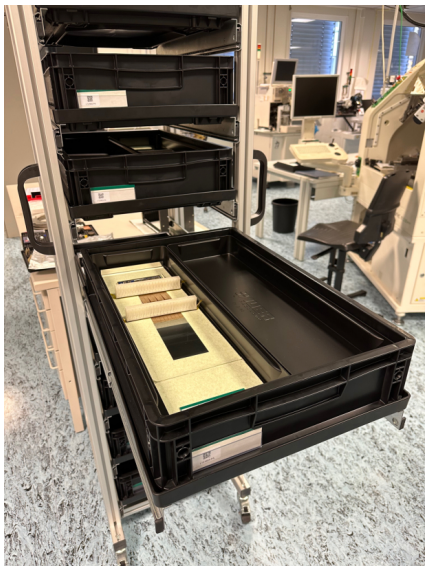
Figure 14.16: Testing setup for FEB chip-cable assemblies and results interface.

During the production process, FEBs are tested twice: first, after mounting the first row of chip-cables and again after the second row is mounted. This dual-stage testing guarantees that the FEB is fully functional before proceeding with sensor bonding. In some instances where chips need replacement, all surrounding components are sealed with glob-top to prevent damage, and the defective chip is replaced. This procedure has been performed successfully multiple times, significantly reducing material waste. After FEB A and B have been produced with appropriate chip-cables, the sensor bonding process is also monitored separately for both sides. This also makes our production process 100% accurate [167].

Figure 14.17 shows the final setup for testing fully assembled modules and their storage configuration before being transported to GSI. The left panel 14.17a illustrates the full module connected to the test bench, where key performance metrics such as electrical connectivity, mechanical stability, and data transfer are validated. The right panel 14.17b presents the storage trays used for safe handling before transport. Modules that pass the final tests are stored securely before being transported to GSI for integration. Proper packaging ensures that the modules remain undamaged during transit, maintaining their performance integrity.



(a) Full module connected to the test station for performance verification.



(b) Storage containers designed for safe transport of completed modules.

Figure 14.17: Final module testing and storage setup: (a) shows the performance test setup, and (b) shows the secure storage trays used to organise and protect the modules before shipment.

14.5 Progress and challenges

Significant advancements have been made in the production and assembly of STS modules for the CBM experiment, reflecting a coordinated effort across all contributing institutions. The Karlsruhe Institute of Technology (KIT) has played a crucial role in developing and refining assembly procedures, resulting in the successful construction of a substantial number of detector modules. The workflow has steadily improved through iterative optimisation, enabling the integration of complex components such as wire bonded LDOs, ASICs, and sensor connections via TAB.

Progress

- **Production milestones:** a consistent assembly rate has been achieved, with modules meeting stringent performance criteria. Quality assurance (QA) testing confirms exceptional yield rates, with ASIC production exceeding 98.5% and chip-cable integration reaching 96%.
- **Enhanced infrastructure:** the cleanroom facility at KIT has been fitted with precision tools, including several Bondjet BJ653, facilitating high-quality wire bonding and robust TAB of microcables to ASICs and to sensors. A dedicated testing station with GBTx emulation boards and a PogoPin setup has ensured consistent QA throughout the production process.

- **Systematic QA and data logging:** comprehensive data logging has been implemented using a PostgreSQL-based system that centralises results from multiple testing stations. This data enables the identification of trends and ensures full traceability of all components.

The final phase of module assembly involves completing the remaining detector ladders, followed by large-scale system integration and calibration using simulated collision data. Continued efforts in QA and collaborative refinement of the production workflow will be essential to ensure the timely delivery of fully functional STS modules to the CBM integration site. The progress achieved so far underscores the importance of precision engineering, rigorous testing, and collaborative expertise. Overcoming the challenges encountered has strengthened the assembly framework, contributing to the overall robustness and readiness of the STS for deployment in the upcoming heavy-ion experiments at FAIR.

15 IPE KIT detector packaging for HEP summary

Part IV presents how high-density interconnection techniques can be adapted both to fragile prototype sensors and to large-scale detector module production. The chapters include case studies where different bonding methods were applied, showing how process optimisation and the available infrastructure enabled reliable assemblies. From single-chip prototypes with ultra-fine pads to the production of hundreds of tracking modules, the focus is on practical solutions that ensure stable performance and reproducible yields in real detector applications.

Summary of key discussions in Chapter 13 and Chapter 14

Chapter 13, "High-density interconnection techniques", presented state-of-the-art methods for fine-pitch sensor integration, highlighting both wire bonding and flip chip bonding as introduced in Section 13.1. Chapter 13 contains two projects demonstrating these techniques in practice. The first work case, described in Section 13.2, focuses on ultra-thin gold wire bonding for silicon–germanium chips with extremely small pads size of 38 µm × 42 µm. This project was done for the chips developed by Université de Genève as part of the TANGERINE initiative for next-generation silicon detectors. This project is a representable example of how careful process optimisation enables reliable connections on such fine pad geometries. The second case, described in Section 13.3, concentrates on gold bump bonding for the Pixelated Silicon Drift Detector (PixDD), where gold stud bumps and flip chip thermocompression techniques were used to connect a delicate X-ray sensor to its ASIC. This project faced challenges like a fragile 0.5 µm thick sensor window and asymmetrical chip layout, which were overcome by tailored handling and precise alignment during bonding.

Chapter 14, "Compressed Baryonic Matter experiment at the future Facility for Antiproton and Ion Research", shifted the focus to a large-scale application – the CBM experiment – outlining its scientific goals and complex detector system. Chapter 14 details the design and assembly of the Silicon Tracking System (STS) modules. It is a key system of the CBM experiment, and it is used as the primary component for particle tracking. All steps of the asssambly are covered in Section 14.3. In particular, Section 14.4 described KIT's contributions to STS construction, which is the assembly of hundreds of silicon sensor modules using both state-of-the-art interconnection techniques and fine-tuned testing procedures.

Technologies covered by IPE cleanroom

The projects in Part IV were possible for realisation due to specialised fabrication technologies available in the IPE KIT cleanroom. Different bonding techniques were a common feature across these projects, including fine gold wire bonding, gold stud bump deposition, precise flip chip bonding, and aluminium tape automated bonding (TAB), all supported by IPE's infrastructure and discussed in Section 13.1. The cleanroom is equipped with dedicated machinery for these tasks, which is described in Subsection 13.1.4 and presented in Figure 13.5, including a Hesse Bondjet BJ653 Wedge/Ball Wire Bonder, which supports bump bonding, wire bonding, and TAB; an ICCON Gold Stud Bonder, utilised for creating gold bumps on pads and fine wire bonds; and a Finetech Fineplacer Femto Die Bonder, used for sub-micron alignment and thermocompression bonding. These capabilities enabled the ultra-fine pitch bonding in the UniGe and PixDD projects and were directly applicable to the CBM STS module production. For instance, as presented in Section 14.4, the STS modules used aluminium wire bonding on the Bondjet bonder, as well as TAB to attach flexible readout cables, processes that were carried out in the IPE cleanroom. All the interconnection technologies highlighted, from gold ball bonding and stud bumping to aluminium wire bonding and TAB, were covered by the IPE cleanroom's machinary park, underscoring its comprehensive role in sensor assembly and packaging for these experiments.

Projects and their outcomes

Each project presented in this thesis achieved important outcomes for semiconductor detector development.

- In the UniGe SiGe chip bonding project in Section 13.2, the fine-pitch gold wire bonding process was refined to successfully connect chips with extremely small pad sizes. The outcome was a reliable bonding with uniform wire loop profiles. This validated the use of gold wire bonding with optimised force and ultrasonic power for next-generation high-density readout chips.
- In the PixDD detector project in Section 13.3, the integration of a pixelated silicon drift sensor with its ASIC via gold stud bumps was accomplished. Despite the fragile sensor window and asymmetry between the sensor and ASIC, the gold bump bonding approach yielded a robust sensor-ASIC assembly. The bonded PixDD module satisfied stringent mechanical integrity standards. Furthermore, it maintained alignment accuracy within 1 µm following the flip chip attachment, demonstrating its suitability for space-based applications.
- Finally, the CBM STS assembly project established the construction workflow, which in the future can cover a large number of modules using the above techniques. All assembled STS modules passed through strict performance tests, with high yield in production – for example, ASIC bonding yields above 98.5% and chip-to-microcable integration yields around 96%. The assembly workflow at KIT matured to a consistent production rate, delivering modules that satisfy the experiment's requirements. These

outcomes demonstrate that the interconnection and packaging strategies discussed in this thesis not only work in isolated case studies but also scale up effectively to full detector systems.

Conclusion

Part IV has demonstrated how cutting-edge semiconductor detector fabrication techniques directly enable advances in high-energy physics experiments and has addressed the challenge of achieving reliable interconnections under very different boundary conditions. Chapter 13 established the foundation in high-density interconnection technologies, from ultra-fine gold bonding to flip chip assembly, and Chapter 14 showed their application in building a sophisticated detector for a real experiment. The IPE cleanroom's machinery park proved essential for both the small-scale prototyping of novel sensors and the large-scale production of detector modules. These projects successfully demonstrated the developed methods through consistent fine-pitch bonding quality, innovative bump bonding, and high-yield mass production of STS modules. At the same time, the CBM experience highlights the importance of precision engineering, quality control, and collaboration when scaling these techniques to an experiment-wide level. Together, these chapters affirm that mastering semiconductor detector production, from individual bond connections to fully instrumented tracking systems, is crucial for realising next-generation physics experiments. The progress achieved, together with insights gained from overcoming technical challenges, provides a solid basis for future detector developments and their successful application in upcoming research at FAIR and other facilities.

Part V

Conclusion

Summary of the conducted research

The development of silicon-based sensors remains central to the advancement of experimental particle physics. As beam energies and experimental constraints continue to increase, the demands on detector systems grow accordingly, requiring fine spatial and temporal resolution, high radiation tolerance, low material budget, and scalable integration. In parallel, interest in implementing solid-state detectors, especially silicon-based sensors, is growing in accelerator-based applications in medical physics, particularly in hadron therapy. Although not yet widely adopted in clinical environments, such detectors are actively being explored as promising candidates for future beam monitoring systems. Their potential lies in offering improved spatial resolution, fast timing, and compatibility with magnetic resonance imaging.

This thesis has addressed these challenges by focusing on the design, simulation, and experimental validation of monolithic detectors implemented in commercial CMOS and BiCMOS processes. Particular emphasis was placed on two technologies: High-Voltage CMOS (HV-CMOS), which is more and more adopted in high-energy physics applications, and silicon-germanium BiCMOS (SiGe BiCMOS), investigated here through the development of a novel sensor concept.

The work combined numerical simulation with empirical testing to demonstrate how these technologies can be optimised for distinct application domains, while sharing a unified development framework. Physics-based modelling and numerical simulation have proven to be indispensable tools in this context. Technology Computer-Aided Design (TCAD) simulations, in particular, were used to investigate sensor behaviour under different operating conditions, optimise charge-collection dynamics, and evaluate the effects of radiation damage. These simulations enabled the prediction of key performance parameters prior to fabrication, reducing the number of design iterations and accelerating the development cycle. Additionally, experimental testing was carried out to validate the simulation results, reveal unforeseen effects, and confirm that the devices perform reliably under real-world operating conditions.

The HV-CMOS technology was studied in Part II, demonstrating its adaptability across diverse application domains. Two representative chip families were presented: *AstroPix*, targeting astrophysical gamma-ray detection, and *HitPix*, designed for real-time ion beam monitoring in medical physics. Both designs are based on the same commercial process, but optimised for different experimental conditions. Additionally, the work addressed performance limitations observed under high-intensity ion irradiation, particularly particle detection efficiency loss, and proposed strategies for mitigating this effect through design improvements.

The second major technology studied in this thesis is SiGe BiCMOS, presented in Part III. The developed chip, named *BeMiPix*, underwent the full development cycle: from TCAD

simulation to laboratory testing and characterisation. This work demonstrated the suitability of the 130 nm IHP SG13G2 SiGe BiCMOS technology for monolithic sensor design, confirming its potential for applications requiring high temporal resolution.

Sensor development, however, does not end at design, fabrication and testing. Equally critical is the challenge of integration, such as establishing reliable, high-density connections between the sensor and its readout electronics. Part IV presents two case studies that illustrate the importance of interconnection techniques in modern detector assemblies: one focuses on ultra-thin gold wire bonding for SiGe-based chips, developed in collaboration with the Université de Genève as part of the TANGERINE initiative; the other examines gold stud bump bonding for the Pixelated Silicon Drift Detector (PixDD). The thesis also broadened its focus to large-scale detector implementation, using the Compressed Baryonic Matter experiment (CBM) as a representative example. This experiment relies on the Silicon Tracking System (STS) as its core tracking component, with its modules assembled at IPE.

The conclusions of each part of this thesis are described in full detail in the respective chapters. The following section presents the main outcomes of the research.

Main results for HV-CMOS development

The HV-CMOS studies presented in this thesis span both application-driven chip development and a broader investigation into radiation-induced signal instabilities. Two main chip families — *HitPix* and *AstroPix* — were simulated and characterised to meet the distinct requirements of medical beam monitoring and astrophysical gamma-ray detection, respectively.

In the case of *HitPix*, TCAD-guided development highlighted the critical balance required between depletion depth, breakdown resilience, and compact layout constraints:

- **Substrate resistivity:** High-resistivity substrates (around 5k Ω·cm) supported deeper depletion but increased susceptibility to premature breakdown, especially in compact pixel layouts. A resistivity of 300 Ω·cm was selected as a balanced solution, ensuring sufficient depletion with greater breakdown resilience.
- **Periphery layout:** Electric field optimisation in the periphery was critical. An optimised periphery spacing of 15 µm was selected to suppress early breakdown events.
- **Power consumption:** For *HitPix3*, designed for use in medical environments, the power consumption estimation is 44.76 mW, ensuring efficient operation for long-term radiation-induced ageing.
- **Charge-collection efficiency:** Simulations showed that the ISO sensor architecture, employing a dedicated isolation pwell, achieved up to 97.5% charge-collection efficiency. This outperformed the standard layout due to enhanced field uniformity and optimised charge-drift paths.

For *AstroPix*, with its space-based telescope application, the development concentrated on achieving uniform depletion for the much larger pixel pitch and inter-pixel isolation. TCAD simulations for AstroPix_v3 guided by the following choices:

- **Substrate resistivity:** 10k Ω·cm substrate was selected, allowing near-complete depletion at −250 V and stable operation up to −500 V without introducing breakdown risks.
- **Pixel geometry:** A large 500 µm pixel pitch and a 300 µm-wide deep nwell collection node were chosen to satisfy the spatial resolution requirement for the AMEGO-X telescope.
- **Surface breakdown control:** Different metalisation schemes were evaluated, with simulations showing that placing the pad over an undepleted nwell provided the most uniform electric field distribution and eliminated surface leakage paths.
- **Inter-pixel isolation:** A p-stop spacing of 100 µm was found to minimise electric field peaks, ensuring isolation while maintaining compatibility with the large pixel geometry and expected radiation environment.

Beyond chip-specific design, a broader and more fundamental issue was explored: the loss of particle detection efficiency in HV-CMOS sensors exposed to high-intensity heavy-ion beams. Initially observed during *HitPix* beam test campaigns, this phenomenon, described as *voltage drops* or *baseline drops*, manifested as sudden, sustained suppression of the signal baseline. TCAD simulations with full CMOS layout extraction were employed to model possible *Single Event Effects (SEE)*, such as *Single Event Latch-up (SEL)*, that explain the observed signal disruptions. A systematic analysis across multiple HV-CMOS chip variants was conducted to understand the root cause. The results demonstrated that signal stability is influenced by a combination of substrate resistivity, pixel layout, and front-end design.

- **AstroPix:** High-resistivity substrates increased the number of collected events but also resulted in more frequent voltage drops, highlighting a trade-off between depletion depth and signal stability.
- **ATLASpix:** Equipped with a PMOS amplifier and CMOS comparator, this chip exhibited no voltage drops under proton irradiation. The ISO topology provided effective pwell isolation, contributing to stable signal behaviour.
- **MuPix8:** Featuring an in-pixel PMOS amplifier and a peripheral CMOS comparator, the chip showed no voltage drops under any tested conditions. This confirms the robustness of its architecture and the effectiveness of physically separating front-end building blocks.
- **HitPix3 vs HitPix2:** *HitPix3*, designed with PMOS amplifiers and deep pwell isolation, significantly outperformed its predecessor. Voltage drop rates were reduced to below 15%, compared to over 70% in *HitPix2*, demonstrating the impact of improved layout and well isolation.

Design strategies for future HV-CMOS sensors

The combined insights from simulation and experiment provide clear guidance for future HV-CMOS design targeting high-rate ion or hadron environments:

- **Front-end design:** Use of PMOS-based front-ends with strong well isolation and ISO topologies is essential to suppress single-event effects and preserve signal integrity under irradiation.
- **In-pixel layout:** Sensitive analogue and digital domains should be physically separated within the pixel to minimise substrate-induced perturbations, especially in high-rate radiation environments.
- **Substrate selection:** The choice of substrate resistivity must balance depletion depth with transient signal stability, considering the specific charge-generation characteristics of the intended application.

These findings demonstrate that continued development of HV-CMOS technology requires a deep physical understanding of silicon behaviour under operating conditions specific to each experimental context.

Main results for BiCMOS development

BeMiPix, short for **Bipolar Experimental Miniature Pixel**, represents a new class of fully monolithic silicon sensors developed for high-resolution charged-particle detection. The sensor was implemented using the IHP SG13G2 130 nm SiGe BiCMOS process, which enables the co-integration of high-performance SiGe Heterojunction Bipolar Transistors (HBTs) and standard CMOS logic within each pixel cell. This architectural combination enables the design of compact, fast analogue front-ends with low noise, which are embedded directly inside the pixel matrix.

Extensive simulations and measurements were carried out to evaluate the performance of the BeMiPix sensor; the main results are summarised below.

- **Leakage current and breakdown:** Electrical characterisation revealed excellent leakage current performance and robust breakdown behaviour. Breakdown voltages for all tested sensors ranged from −70 V to −90 V. At an operating bias of −50 V, dark currents remained low, typically between 4 and 10 nA per mm^2 depending on substrate type. The best result was observed for the 20 µm epitaxial sample, which maintained a stable leakage current of approximately 4 nA across all dose rates. In contrast, the non-epitaxial sample reached up to 58.9 nA. These values are significantly lower than those of conventional silicon sensors, which often exhibit leakage currents in the mA/cm^2 range, confirming the effectiveness of the BiCMOS process and layout optimisation.
- **Timing performance:** The sensor's timing response was evaluated using ^{90}Sr source measurements. Rise times ranged from 0.87 ns (20 µm epitaxial sample) to 1.09 ns (50 µm epitaxial sample), closely matching TCAD simulation results of 0.83 ns and 1.04 ns, respectively. These results validate the front-end design and confirm that drift path length is the dominant factor affecting signal speed.

- **Signal delay (laser measurements):** Laser-based measurements showed a total signal delay of approximately 52 ns, including contributions from carrier drift, laser trigger latency, and electronic response. This measurement is in good agreement with the simulated delay of 40 ns, excluding laser trigger delay.

Insights to BiCMOS technology for particle detection

Overall, the BeMiPix architecture demonstrated excellent agreement between design predictions and experimental observations. The results confirm the sensor's capability for fast detection, efficient charge collection, and consistent timing behaviour across a wide range of test conditions. This validation, achieved through both source-based and laser-based measurements, highlights the feasibility of SiGe BiCMOS MAPS as a platform for high-performance monolithic detector development.

The successful implementation of SiGe BiCMOS technology in the BeMiPix prototype confirms its strong potential for next-generation pixel sensors, particularly in applications demanding high speed, radiation tolerance, and compact integration. With further optimisation and scaling, BiCMOS MAPS could complement, or in some cases replace, conventional hybrid sensors. Future work should focus on extending this platform to larger pixel matrices, developing radiation-hard design strategies, and implementing on-chip data processing architectures to leverage the advantages of this promising technology fully.

Main results for interconnection and module assembly

The final part of this thesis focused on interconnection strategies and detector module assembly, with an emphasis on applying high-density bonding techniques to both research prototypes and large-scale detector systems. These studies demonstrated that bonding methods such as gold wire bonding and bump bonding can be successfully implemented across a variety of sensor types and geometries, resulting in mechanically robust and electrically stable assemblies.

The first case involved the assembly of SiGe-based monolithic chips developed by the Université de Genève. These chips feature extremely small pad geometries, only 38 µm × 42 µm, making the interconnection process particularly demanding. A fine-pitch gold wire bonding process was developed and optimised, with careful tuning of ultrasonic energy, bonding force, and loop height. The resulting assemblies exhibited uniform bond profiles and excellent mechanical strength, validating the applicability of gold wire bonding to next-generation readout electronics with reduced pad sizes and high integration density.

In the second project, a hybrid assembly was realised for the Pixelated Silicon Drift Detector (PixDD), integrating the sensor with its readout ASIC via gold stud bump bonding and thermocompression flip chip process. This sensor features a fragile 0.5 µm-thick entrance window and an asymmetrical chip layout, both of which posed significant challenges during handling and bonding. By implementing precise alignment strategies and tailored bonding procedures,

the final assembly achieved mechanical robustness, with post-bonding inspection confirming an alignment accuracy better than 1 μm.

The third example addressed the transition from small-scale assembly to module production, using the STS project of the CBM experiment as a representative case. In this context, the Institute for Data Processing and Electronics (IPE) established and refined a full assembly workflow for double-sided strip sensor modules, combining wire bonding, Tape Automated Bonding (TAB) interconnection, and electrical testing. The production process achieved high reliability and throughput, with ASIC bonding yields exceeding 98.5% and chip-to-microcable integration yields around 96%. The production infrastructure matured into a reproducible, scalable process capable of supporting the construction of hundreds of modules, all passing rigorous performance validation required by the CBM experiment. This project confirmed that the interconnection and packaging strategies developed and tested in this thesis are not limited to isolated case studies but are transferable to large-scale detector construction.

Final thoughts

This thesis demonstrates the necessity of addressing all stages of detector development, from physical modelling and process optimisation to experimental validation and system-level integration, in order to realise the next generation of silicon-based sensors for particle physics and related fields. Each part of the work contributes to a different aspect of this goal, from simulation-guided design and radiation effect studies in HV-CMOS to prototyping and SiGe BiCMOS technology testing, as well as addressing the packaging challenges of high-density interconnection and large-scale detector assembly. The integrated approach adopted here illustrates how progress in semiconductor detectors relies not only on advances in circuit or process technology, but also on the close cooperation between design, characterisation, and manufacturing.

The results presented across the HitPix, AstroPix, BeMiPix, and CBM STS projects provide concrete evidence that monolithic and hybrid silicon sensors, when combined with tailored packaging strategies and careful validation procedures, can meet the increasingly complex demands of modern experiments. Moreover, the demonstrated compatibility of commercial CMOS and BiCMOS technologies with harsh radiation environments, fast timing requirements, and fine-granularity architectures opens a clear path toward scalable and application-specific detector solutions.

Looking forward, the evolution of detector systems will likely involve even tighter integration of sensing, processing, and communication within compact silicon platforms. Future developments may leverage emerging techniques such as three-dimensional integration, in-pixel artificial intelligence, and radiation-hard materials. However, their success will continue to depend on the fundamental principles explored in this work: simulation, characterisation, and adaptable, scalable assembly techniques.

Acknowledgements

This dissertation would not have been possible without the support of many people, to whom I would like to express my deepest gratitude.

First of all, I would like to thank Prof. Dr Marc Weber for seeing potential in me, for inviting me to join the Institute for Data Processing and Electronics (IPE), and for providing me with the opportunity to conduct my doctoral research under his supervision. I would also like to thank Prof. Dr Ivan Perić and Prof. Dr Steven Worm for taking on the role of co-referees of my dissertation and for their valuable evaluation of the thesis.

I am also grateful to the three IPE groups and the people within them, where I had the opportunity to work during my PhD: Aufbau- und Verbindungstechnik (AVT), Prozessdaten-verarbeitung (PDV), and ASIC- und Detektorlabor (ADL). I would like to thank Dr Thomas Blank, head of the AVT group, for presenting me with my first working challenges and for leading the CBM sensor assembly at IPE with clarity and dedication. I am especially grateful to Dr Andreas Kopmann, who has been a consistently supportive group leader during my time in the PDV group, always available with advice, patience, and encouragement at every stage of my work. And I am deeply grateful to Prof. Dr Ivan Perić, who gave me a place in the ADL group and challenged me with the most fascinating scientific tasks. I am especially thankful for the trust he placed in me, as well as for his personal support during the most difficult moments of my doctoral journey.

My sincere gratitude also goes to my advisors and collaborators, whose time and insight made this work valuable. Dr Alexander Dierlamm and the Institute for Experimental Particle Physics (ETP), for the engaging projects we worked on together; and Dr Felix Ehrler for his support during experiments and for reading and commenting on my manuscript. I also extend my thanks to the entire STS team at GSI, with whom I shared both challenging projects and great conferences.

I am deeply grateful to all my colleagues at IPE for their ongoing support and the friendly atmosphere they have created. I would especially like to thank Lars for helping me during my first steps in a new country; and Armen, for all the conversations and for creating a feeling of home. My thanks also go to Benjamin Leyrer, Alexander Bacher, and Jennifer Derschang, as well as the administration team at IPE, for their ever-reliable support with both infrastructure and bureaucracy.

To the young scientists, former and current PhD students, thank you for the heart you bring into everyday work. Special thanks go to Meghana, Luis, Patrick, and my two officemates, Roberto and Alexander.

In the hours beyond research, I would like to thank my friends, who brought another perspective to my life. Simon, Luca, Angeline, and Eleonora, thank you for the fun times, for lending your ears and shoulders, and, of course, for your advice. Olena, thank you for being a bridge between two cultures and for not letting me forget Russian. Lena, thank you for sharing the experience of moving to Germany with me. And to my oldest friend, Lera, thank you for being there through every stage of life, for somehow knowing my next move before I did, and for patiently listening to me talk about work, no matter how confusing the work stories were.

None of this would have been possible without my family. To my parents, Viktor and Nailia, who gave me a sense of direction and purpose, and did everything possible to help me achieve it. To my sister, Maria, who has become more than a sibling, but a true friend. Although oceans separate us, your support has always been with me. I would not have made it through all of this without you. To my brother-in-law, David, whose positivity is truly inspiring and whose contribution to my English skills is invaluable.

And last, but by no means least, to my best friend, my soulmate, and my husband, Felix. You have been my greatest support in life and in work, brought happiness into my days, and believed in me even when I struggled to believe in myself. I am endlessly grateful to have you in my life. I am especially happy that you took the PostDoc position at IPE, the choice that brought us together.

This has been a long and challenging journey. It covered many places along the way, starting in Volzhsk, a small town in the Mari El Republic, continuing through high school in Saint Petersburg, university in Moscow, work in Geneva, and finally settling in Karlsruhe. But it was also a personal journey: a path towards making science and physics not just my profession, but a part of my life. All the right and wrong choices, the disappointments and excitements, the challenges of education and the rewards of scientific fulfilment have brought me here, to the writing of this thesis.

To everyone who walked alongside me on this path: thank you.

Bibliography

[1] M. Castells, *The Rise of the Network Society: The Information Age – Economy, Society and Culture, Volume I.* Oxford: Wiley-Blackwell, 2000. [Online]. Available: http://dx.doi.org/10.1002/9781444319514

[2] S. M. Sze and Y. Li, *Physics of Semiconductor Devices*, 4th ed. John Wiley & Sons, 2021.

[3] F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics: Basics and Applications*, 3rd ed. Cham, Switzerland: Springer Cham, 2024. [Online]. Available: https://doi.org/10.1007/978-3-031-59720-6

[4] S. Möller, *Accelerator Technology: Applications in Science, Medicine, and Industry*. Cham, Switzerland: Springer Cham, 2020. [Online]. Available: https://doi.org/10.1007/978-3-030-62308-1

[5] ECFA Detector R&D Roadmap Process Group, "The 2021 ECFA Detector Research and Development Roadmap," Geneva, 2020. [Online]. Available: https://cds.cern.ch/record/2784893

[6] A. Dierlamm *et al.*, "A beam monitor for ion beam therapy based on HV-CMOS pixel detectors," *Instruments*, vol. 7, no. 1, 2023. [Online]. Available: https://doi.org/10.3390/instruments7010009

[7] I. Perić, "A novel monolithic pixel detector implemented in high-voltage CMOS technology," in *2007 IEEE Nuclear Science Symposium and Medical Imaging Conference*, vol. 2, 2007. [Online]. Available: https://doi.org/10.1109/NSSMIC.2007.4437188

[8] F. M. Ehrler, "Characterization of monolithic HV-CMOS pixel sensors for particle physics experiments," Ph.D. dissertation, Karlsruher Institut für Technologie (KIT), 2021. [Online]. Available: https://publikationen.bibliothek.kit.edu/1000133748

[9] L. D. Landau, A. I. Akhiezer, and E. M. Lifshitz, *General Physics. Chapter VI – The Theory of Symmetry*. Oxford, United Kingdom: Pergamon, 1967. [Online]. Available: https://doi.org/10.1016/B978-0-08-009106-8.50009-1

[10] J. L. Moll, *Physics of Semiconductor Devices*. New York: McGraw-Hill, 1970. [Online]. Available: https://archive.org/details/physicsofsemicon00moll

[11] M. Grundmann, *The Physics of Semiconductors: An Introduction Including Nanophysics and Applications*, 4th ed. Cham, Switzerland: Springer, 2021. [Online]. Available: https://doi.org/10.1007/978-3-030-51569-0

[12] L. Huldt *et al.*, "The temperature dependence of Band-to-Band Auger Recombination in silicon," *Applied Physics Letters*, vol. 35, no. 10, November 1979. [Online]. Available: https://doi.org/10.1063/1.90974

[13] C. Sah, R. N. Noyce, and W. Shockley, "Carrier generation and recombination in p-n junctions and p-n junction characteristics," *Proceedings of the IRE*, vol. 45, no. 9, 1957. [Online]. Available: https://doi.org/10.1109/JRPROC.1957.278528

[14] R. N. Hall, "Electron-hole recombination in germanium," *Phys. Rev.*, vol. 87, no. 2, July 1952. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.87.387

[15] W. Shockley and W. T. Read, "Statistics of the recombinations of holes and electrons," *Phys. Rev.*, vol. 87, no. 5, September 1952. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.87.835

[16] W. Shockley, *Electrons and Holes in Semiconductors: With Applications to Transistor Electronics*, 2nd ed. New York: Van Nostrand, 1950.

[17] J. Jean *et al.*, "Pathways for solar photovoltaics," *Energy Environ. Sci.*, vol. 8, February 2015. [Online]. Available: https://doi.org/10.1039/C4EE04073B

[18] A. Rusu, O. Pietrareanu, and C. Bulucea, "Reversible breakdown voltage collapse in silicon gate-controlled diodes," *Solid-State Electronics*, vol. 23, no. 5, 1980. [Online]. Available: https://doi.org/10.1016/0038-1101(80)90084-2

[19] A. Rusu and C. Bulucea, "Deep-depletion breakdown voltage of silicon-dioxide/silicon MOS capacitors," *IEEE Transactions on Electron Devices*, vol. 26, no. 3, 1979. [Online]. Available: https://doi.org/10.1109/T-ED.1979.19404

[20] E. V. Trifonova *et al.*, "Influence of formation conditions of Pt Schottky barriers on the energy resolution of particle detectors based on VPE GaAs," in *18th International Workshop on Radiation Imaging Detectors (IWORID 2017)*, Kraków, Poland, 2017. [Online]. Available: https://indico.cern.ch/event/598230/contributions/2608852/attachments/1494517/2324812/PI_A18_Trifonova.pdf

[21] L. C. T. Cao *et al.*, "Boron doping in next-generation materials for semiconductor device," in *Characteristics and Applications of Boron*. Rijeka: IntechOpen, 2022. [Online]. Available: https://doi.org/10.5772/intechopen.106450

[22] R. Eber, "Untersuchung neuartiger Sensorkonzepte und Entwicklung eines effektiven Modells der Strahlenschädigung für die Simulation hochbestrahlter Silizium-Teilchendetektoren," Ph.D. dissertation, Karlsruhe Institute of Technology (KIT), 2013. [Online]. Available: https://publish.etp.kit.edu/record/20930

[23] H. Bethe, "Zur Theorie des Durchgangs Schneller Korpuskularstrahlen durch Materie," *Annalen der Physik*, vol. 397, no. 3, 1930. [Online]. Available: https://doi.org/10.1002/andp.19303970303

[24] F. Bloch, "Zur Bremsung Rasch bewegter Teilchen beim Durchgang durch Materie," *Annalen Phys.*, vol. 408, 1933. [Online]. Available: https://doi.org/10.1002/andp.19334080303

[25] K. Nakamura *et al.*, "Review of particle physics," *Journal of Physics G: Nuclear and Particle Physics*, vol. 37, no. 7A, July 2010. [Online]. Available: https://dx.doi.org/10.1088/0954-3899/37/7A/075021

[26] C. W. Fabjan and H. Schopper, *Particle Physics Reference Library: Detectors for Particles and Radiation*, 1st ed. Cham, Switzerland: Springer, 2020. [Online]. Available: https://doi.org/10.1007/978-3-030-35318-6

[27] T. Poehlsen *et al.*, "Charge losses in segmented silicon sensors at the Si-SiO_2 interface," *Nucl. Instrum. Meth. A*, vol. 700, 2013. [Online]. Available: https://doi.org/10.1016/j.nima.2012.10.063

[28] M. Moll, "Displacement damage in silicon detectors for high energy physics," *IEEE Transactions on Nuclear Science*, vol. 65, no. 8, 2018. [Online]. Available: https://doi.org/10.1109/TNS.2018.2819506

[29] R. Feigelson, *Crystal Growth Through the Ages*, December 2015. [Online]. Available: https://doi.org/10.1016/B978-0-444-56369-9.00001-0

[30] P. Rudolph, Ed., *Handbook of Crystal Growth, Second Edition, Volume II A–II B: Bulk Crystal Growth*, 2nd ed. Amsterdam, Netherlands: Elsevier, November 2014. [Online]. Available: https://doi.org/10.1016/C2013-0-09791-5

[31] U. Hilleringmann, *Silicon Semiconductor Technology: Processing and Integration of Microelectronic Devices*, 1st ed. Wiesbaden, Germany: Springer Vieweg Wiesbaden, 2023. [Online]. Available: https://doi.org/10.1007/978-3-658-41041-4

[32] MicroChemicals GmbH, "Silicon ingot production: Czochralski and float-zone methods," MicroChemicals GmbH, Ulm, Germany, Tech. Rep., 2023. [Online]. Available: https://www.microchemicals.com/dokumente/application_notes/czochralski_floatzone_silicon_ingot_production.pdf

[33] STMicroelectronics, "Introduction to semiconductor technology," STMicroelectronics, Geneva, Switzerland, Tech. Rep., 2020. [Online]. Available: https://www.st.com/resource/en/application_note/cd00003986-introduction-to-semiconductor-technology-stmicroelectronics.pdf

[34] J. Zhang *et al.*, "Fabrication and measurement of silicon diode and microstrip detectors," Research Institute for High Energy Physics, University of Helsinki, Helsinki, Finland, Tech. Rep., October 1994. [Online]. Available: https://cds.cern.ch/record/260592/files/P00022058.pdf

[35] T. Bergauer, "Silicon detectors in high energy physics," Lecture slides, VO Teilchendetektoren SS 2017, 2017. [Online]. Available: https://www.oeaw.ac.at/fileadmin/Institute/HEPHY/PDF/ausbildung/teilchendetektoren/VO-4-2018-SemiconductorDetectors.pdf

[36] E. Trifonova *et al.*, "High-pressure high-temperature single-crystal diamond type IIa characterization for particle detectors," *physica status solidi (a)*, vol. 217, February 2020. [Online]. Available: https://doi.org/10.1002/pssa.201900888

[37] S. Seidel, "Silicon strip and pixel detectors for particle physics experiments," *Physics Reports*, vol. 828, 2019. [Online]. Available: https://doi.org/10.1016/j.physrep.2019.09.003

[38] A. Velyka, "Concept and development of Enhanced Lateral Drift (ELAD) sensors," Ph.D. dissertation, Universität Hamburg, 2019, dESY-THESIS-2020-014. [Online]. Available: https://bib-pubdb1.desy.de/record/440957

[39] G. Lutz, *Semiconductor Radiation Detectors: Device Physics*, 1st ed. Berlin; Heidelberg: Springer Berlin Heidelberg, 2007. [Online]. Available: https://doi.org/10.1007/978-3-540-71679-2

[40] G. B. Beneventi, "Lecture presentation in Technology Computer Aided Design (TCAD) Laboratory," Advanced Research Center on Electronic Systems (ARCES) University of Bologna, Italy, 2015.

[41] T. Ma *et al.*, "TCAD: present state and future challenges," in *2010 International Electron Devices Meeting*, 2010. [Online]. Available: https://doi.org/10.1109/IEDM.2010.5703367

[42] *Sentaurus Device User Guide*, Synopsys, Inc, March 2016, version L-2016.03.

[43] *Sentaurus Process User Guide*, Synopsys, Inc, March 2016, version L-2016.03.

[44] *Sentaurus Structure Editor User Guide*, Synopsys, Inc, March 2016, version L-2016.03.

[45] *Sentaurus Mesh User Guide*, Synopsys, Inc, March 2016, version L-2016.03.

[46] *Sentaurus Workbench User Guide*, Synopsys, Inc, March 2016, version L-2016.03.

[47] *Sentaurus Visual User Guide*, Synopsys, Inc, March 2016, version L-2016.03.

[48] J. W. Slotboom and H. C. de Graaff, "Measurements of bandgap narrowing in Si bipolar transistors," *Solid-State Electronics*, vol. 19, no. 10, 1976. [Online]. Available: https://doi.org/10.1016/0038-1101(76)90043-5

[49] J. W. Slotboom, "The pn-product in silicon," *Solid-State Electronics*, vol. 20, no. 4, 1977. [Online]. Available: https://doi.org/10.1016/0038-1101(77)90108-3

[50] G. Masetti, M. Severi, and S. Solmi, "Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon," *IEEE Transactions on Electron Devices*, vol. 30, no. 7, 1983. [Online]. Available: https://doi.org/10.1109/T-ED.1983.21207

[51] E. Conwell and V. F. Weisskopf, "Theory of impurity scattering in semiconductors," *Phys. Rev.*, vol. 77, no. 3, February 1950. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.77.388

[52] N. H. Fletcher, "The high current limit for semiconductor junction devices," *Proceedings of the IRE*, vol. 45, no. 6, 1957. [Online]. Available: https://doi.org/10.1109/JRPROC.1957.278485

[53] C. Canali *et al.*, "Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature," *IEEE Transactions on Electron Devices*, vol. 22, no. 11, 1975. [Online]. Available: https://doi.org/10.1109/T-ED.1975.18267

[54] D. J. Roulston, N. D. Arora, and S. Chamberlain, "Modeling and measurement of minority-carrier lifetime versus doping in diffused layers of n^+-p silicon diodes," *IEEE Transactions on Electron Devices*, vol. 29, no. 2, 1982. [Online]. Available: https://doi.org/10.1109/T-ED.1982.20697

[55] M. S. Tyagi and R. V. Overstraeten, "Minority carrier recombination in heavily-doped silicon," *Solid-State Electronics*, vol. 26, no. 6, 1983. [Online]. Available: https://doi.org/10.1016/0038-1101(83)90174-0

[56] H. Goebel and K. Hoffmann, "Full dynamic power diode model including temperature behavior for use in circuit simulators," in *Proceedings of the 4th International Symposium on Power Semiconductor Devices and ICs*, 1992. [Online]. Available: https://doi.org/10.1109/ISPSD.1992.991249

[57] A. Schenk, "A model for the field and temperature dependence of Shockley-Read-Hall lifetimes in silicon," *Solid-State Electronics*, vol. 35, no. 11, 1992. [Online]. Available: https://doi.org/10.1016/0038-1101(92)90184-E

[58] A. Schenk and U. Krumbein, "Coupled defect-level recombination: theory and application to anomalous diode characteristics," *Journal of Applied Physics*, vol. 78, no. 5, September 1995. [Online]. Available: https://doi.org/10.1063/1.360007

[59] J. L. Moll and R. V. Overstraeten, "Charge multiplication in silicon p-n junctions," *Solid-State Electronics*, vol. 6, no. 2, 1963. [Online]. Available: https://doi.org/10.1016/0038-1101(63)90009-1

[60] R. V. Overstraeten and H. D. Man, "Measurement of the ionization rates in diffused silicon p-n junctions," *Solid-State Electronics*, vol. 13, no. 5, 1970. [Online]. Available: https://doi.org/10.1016/0038-1101(70)90139-5

[61] A. Cuevas *et al.*, "Surface recombination velocity of highly doped n-type silicon," *Journal of Applied Physics*, vol. 80, no. 6, September 1996. [Online]. Available: https://doi.org/10.1063/1.363250

[62] L. W. Nagel and D. O. Pederson, "SPICE (Simulation Program with Integrated Circuit Emphasis)," Tech. Rep. UCB/ERL M382, April 1973. [Online]. Available: http://www2.eecs.berkeley.edu/Pubs/TechRpts/1973/22871.html

[63] O. Schenk *et al.*, "PARDISO: a high-performance serial and parallel sparse linear solver in semiconductor device simulation," *Future Generation Computer Systems*, vol. 18, March 2000. [Online]. Available: https://doi.org/10.1016/S0167-739X(00)00076-5

[64] R. E. Bank and D. J. Rose, "Global approximate Newton methods," *Numerische Mathematik*, vol. 37, 1981. [Online]. Available: http://eudml.org/doc/132731

[65] M. Lundström, *Essential Semiconductor Physics*. World Scientific, 2025. [Online]. Available: https://doi.org/10.1142/14454

[66] M. S. Shur, *Introduction to Electronic Devices*, 1st ed. John Wiley & Sons, 1995. [Online]. Available: https://www.wiley.com/en-ie/Introduction+to+Electronic+Devices-p-9780471103486

[67] J. Lorenz, "TCAD: from process roots at IISB to industrial standard tools," Presentation, IISB Annual Symposium, Erlangen, Germany, October 2023, presented at the IISB Annual Symposium 2023.

[68] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. John Wiley & Sons, 2006. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/0470068329

[69] S. E. Laux, "Techniques for small-signal analysis of semiconductor devices," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 4, no. 4, 1985. [Online]. Available: https://doi.org/10.1109/TCAD.1985.1270145

[70] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, ser. Applied Mathematical Sciences. Springer, 1991, vol. 15. [Online]. Available: https://doi.org/10.1007/978-1-4612-3172-1

[71] G. Jain, "Radiation damage modelling: TCAD simulation," in *Proceedings of the 27th International Workshop on Vertex Detectors*. SISSA, 2019. [Online]. Available: https://cds.cern.ch/record/2697098

[72] D. Passeri and A. Morozzi, "Deliverable report on topic TCAD Radiation Damage Model," 2019. [Online]. Available: https://cds.cern.ch/record/2705944/files/AIDA-2020-D7_4.pdf

[73] Synopsys, Inc, "TCAD Sentaurus Tutorial – Sentaurus Device 13. Special focus: traps." [Online]. Available: https://ghzphy.github.io/Sentaurus_Training/sd/sd_13.html

[74] D. Passeri, "Modelling radiation damage in TCAD," Presentation at the 26th International Workshop on Vertex Detectors, 2017. [Online]. Available: https://indico.cern.ch/event/627245/contributions/2676988/attachments/1523246/2380569/Passeri_VERTEX2017.pdf

[75] M. Petasecca *et al.*, "Numerical simulation of radiation damage effects in p-type and n-type FZ silicon detectors," *IEEE Transactions on Nuclear Science*, vol. 53, no. 5, 2006. [Online]. Available: https://doi.org/10.1109/TNS.2006.881910

[76] D. Pennicard *et al.*, "Simulations of radiation-damaged 3D detectors for the Super-LHC," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 592, no. 1, 2008. [Online]. Available: https://doi.org/10.1016/j.nima.2008.03.100

[77] M. Ahmed *et al.*, "Deep-level transient spectroscopy studies of silicon detectors after 24 GeV proton irradiation and 1 MeV neutron irradiation," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 457, no. 3, 2001. [Online]. Available: https://doi.org/10.1016/S0168-9002(00)00788-9

[78] A. Weber, "Development of integrated circuits and smart sensors for particle detection in physics experiments and particle therapy," Ph.D. dissertation, Universität Heidelberg, 2021, 54.12.01; LK 01. [Online]. Available: https://archiv.ub.uni-heidelberg.de/volltextserver/30650/

[79] OECD and European Observatory on Health Systems and Policies, "Germany: country health profile 2023," 2023. [Online]. Available: https://doi.org/10.1787/21dd4679-en

[80] Mayo Clinic Press, "Cancer treatment," 2022. [Online]. Available: https://www.mayoclinic.org/tests-procedures/cancer-treatment/about/pac-20393344

[81] M. Chalmers, "How particle physics can be therapeutic," *Physics World*, vol. 16, no. 8, August 2003. [Online]. Available: https://dx.doi.org/10.1088/2058-7058/16/8/35

[82] H. Youngyih, "Current status of proton therapy techniques for lung cancer," *Radiat Oncol J*, vol. 37, no. April, 2019. [Online]. Available: https://doi.org/10.3857/roj.2019.00633

[83] T. J. Royce and J. A. Efstathiou, "Proton therapy for prostate cancer: a review of the rationale, evidence, and current state," *Urologic Oncology: Seminars and Original Investigations*, vol. 37, no. 9, 2019. [Online]. Available: https://doi.org/10.1016/j.urolonc.2018.11.012

[84] W. Schlegel, C. P. Karger, and O. Jäkel, *Medizinische Physik: Grundlagen – Bildgebung – Therapie – Technik*. Berlin; Heidelberg: Springer Spektrum, 2018. [Online]. Available: https://doi.org/10.1007/978-3-662-54801-1

[85] A. Lühr *et al.*, "Relative biological effectiveness in proton beam therapy – current knowledge and future challenges," *Clinical and Translational Radiation Oncology*, vol. 9, February 2018. [Online]. Available: https://doi.org/10.1016/j.ctro.2018.01.006

[86] "Heidelberg University Hospital," Website, Universitätsklinikum Heidelberg, 2025. [Online]. Available: https://www.klinikum.uni-heidelberg.de/

[87] Heidelberg University Hospital, "Heidelberg Ion Beam Therapy Center (HIT) – patient and research information brochure," Brochure, Corporate Communication and Press Office, Heidelberg University Hospital, Im Neuenheimer Feld 450, 69120 Heidelberg, Germany, October 2012. [Online]. Available: https://www.klinikum.uni-heidelberg.de/fileadmin/hit/dokumente/121019KV_SS_HITImage_engl_web_ID17763.pdf

[88] A. Weber, F. Ehrler, R. Schimassek, and I. Perić, "High-voltage CMOS Active Pixel sensor chip with counting electronics for beam monitoring," *IEEE Transactions on Nuclear Science*, vol. 69, no. 6, 2022. [Online]. Available: https://doi.org/10.1109/TNS.2022.3173807

[89] A. L. Heggelund *et al.*, "Overview of the ATLAS ITk pixel detector," *Journal of Instrumentation*, vol. 18, no. 02, February 2023. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/18/02/C02014

[90] The CMS Collaboration, "The CMS experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 08, August 2008. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/3/08/S08004

[91] M. Pittermann, "Evaluation of HV-CMOS sensors in a beam monitoring system for ion therapy," Master's thesis, Karlsruhe Institute of Technology (KIT), 2022, report Number: ETP-KA/2022-11. [Online]. Available: https://publish.etp.kit.edu/record/22127

[92] P. Trujillo, "Division algorithms in FPGA," 2024. [Online]. Available: https://www.controlpaths.com/2022/06/20/division-algorithms-fpga/

[93] F. Ehrler, "HEiKA: feasibility study of HV-CMOS sensors as beam monitoring devices in hadron therapy," Presentation at KIT – Karlsruhe Institute of Technology, ASIC and Detector Laboratory, Institute for Data Processing and Electronics (IPE), October 2021.

[94] E. Trifonova *et al.*, "A Beam monitor for ion beam therapy based on HV-CMOS pixel detectors," Presentation at KIT – Karlsruhe Institute of Technology, ASIC and Detector Laboratory, Institute for Data Processing and Electronics, October 2023.

[95] B. Topko, "HitPix project status: beam monitoring system and sensor development," Presentation at the HW Meeting, KIT – Karlsruhe Institute of Technology, September 2023.

[96] A. Dierlamm, "HitPix project status: news and design considerations for HitPix3 and HitPix4," Presentation at KIT – Karlsruhe Institute of Technology, Karlsruhe Institute of Technology (KIT), ASIC and Detector Laboratory, November 2024.

[97] Ametek: advanced measurement technology, "Introduction to charged particle detectors," 2020. [Online]. Available: https://www.ortec-online.com/-/media/ametekortec/other/introduction-charged-particle-detectors.pdf?la=en

[98] O. Krasel *et al.*, "Measurement of trapping time constants in proton-irradiated silicon pad detectors," in *2003 IEEE Nuclear Science Symposium. Conference Record (IEEE Cat. No.03CH37515)*, vol. 1, 2003. [Online]. Available: https://doi.org/10.1109/NSSMIC.2003.1352079

[99] K. Park *et al.*, "Effects of the resistivity and crystal orientation of the silicon pin detector on the dark current and radiation response characteristics," in *IEEE Nuclear Science Symposium Conference Record*, vol. 2, December 2006. [Online]. Available: https://doi.org/10.1109/NSSMIC.2006.356031

[100] R. Schimassek, "Development and characterisation of integrated sensors for particle physics," Ph.D. dissertation, 2021. [Online]. Available: https://doi.org/10.5445/IR/1000141412

[101] H. Spieler, *Semiconductor Detector Systems*. Oxford University Press, 2005. [Online]. Available: https://www-physics.lbl.gov/~spieler/misc_stuff/text/sent_02jul05/text_pdf/Semiconductor_Detector_Systems.pdf

[102] Brookhaven National Laboratory (BNL), "LET and range plots," Accessed: January 2025, nSRL User Guide, Brookhaven National Laboratory. [Online]. Available: https://www.bnl.gov/nsrl/userguide/let-range-plots.php

[103] M. A. Xapsos *et al.*, "Proton testing guidelines," NASA Goddard Space Flight Center, Tech. Rep. [Online]. Available: https://radhome.gsfc.nasa.gov/radhome/papers/Proton_Testing_Guidlines_paper.pdf

[104] A. L. Steinhebel *et al.*, "AstroPix: novel monolithic active pixel silicon sensors for future gamma-ray telescopes," in *Space Telescopes and Instrumentation 2022: Ultraviolet to Gamma Ray*, vol. 12181, International Society for Optics and Photonics. SPIE, 2022. [Online]. Available: https://doi.org/10.1117/12.2630405

[105] N. Striebig, R. Leys, and I. Perić, "AstroPix: a novel HV-CMOS pixel sensor for space-based experiments," Presentation at TREDI 2023, March 2023. [Online]. Available: https://indico.cern.ch/event/1223972/contributions/5262042/attachments/2603045/4495158/AstroPix_Striebig.pdf

[106] N. Striebig *et al.*, "AstroPix4 — a novel HV-CMOS sensor developed for space based experiments," *Journal of Instrumentation*, vol. 19, no. 04, April 2024. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/19/04/C04010

[107] R. Caputo *et al.*, "All-sky medium energy gamma-ray observatory explorer mission concept," *Journal of Astronomical Telescopes, Instruments, and Systems*, vol. 8, no. 4, 2022. [Online]. Available: https://doi.org/10.1117/1.JATIS.8.4.044003

[108] A. L. Steinhebel *et al.*, "A-STEP for AstroPix: development and test of a space-based payload using novel pixelated silicon for gamma-ray measurement," *PoS ICRC2023*, 2023. [Online]. Available: https://doi.org/10.22323/1.444.0579

[109] I. Martinez-Castellanos *et al.*, "Improving the low-energy transient sensitivity of AMEGO-X using single-site events," *The Astrophysical Journal*, vol. 934, no. 2, July 2022. [Online]. Available: https://dx.doi.org/10.3847/1538-4357/ac7ab2

[110] I. Brewer *et al.*, "Developing the future of gamma-ray astrophysics with monolithic silicon pixels," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1019, 2021. [Online]. Available: https://doi.org/10.1016/j.nima.2021.165795

[111] M. Printz, "P-stop isolation study of irradiated n-in-p type silicon strip sensors for harsh radiation environments," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 831, 2016. [Online]. Available: https://doi.org/10.1016/j.nima.2016.05.103

[112] E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology*. Wiley-Interscience, 2002. [Online]. Available: https://www.wiley-vch.de/de/fachgebiete/ingenieurwesen/mos-metal-oxide-semiconductor-physics-and-technology-978-0-471-43079-7

[113] F. Conti and M. Conti, "Surface breakdown in silicon planar diodes equipped with field plate," *Solid-State Electronics*, vol. 15, no. 1, 1972. [Online]. Available: https://doi.org/10.1016/0038-1101(72)90070-6

[114] G. A. Beck *et al.*, "Radiation-tolerant breakdown protection of silicon detectors using multiple floating guard rings," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 396, no. 1, 1997. [Online]. Available: https://doi.org/10.1016/S0168-9002(97)00749-3

[115] Ruđer Bošković Institute, "Laboratory for ion beam interactions, division of experimental physics," Webpage, Accessed: May 2025. [Online]. Available: https://www.irb.hr/eng/Divisions/Division-of-Experimental-Physics/Laboratory-for-Ion-Beam-Interactions

[116] E. Trifonova *et al.*, "Microbeam test – first results," Presentation, Karlsruhe Institute of Technology (KIT), 2023.

[117] C. Rossi, A. Chatel, and J.-M. Sallese, "Modeling funneling effect with generalized devices for SPICE simulation of soft errors," *IEEE Transactions on Electron Devices*, vol. 68, no. 6, 2021. [Online]. Available: https://doi.org/10.1109/TED.2021.3076028

[118] D. Chen, "Investigation of heavy ion-induced Single-Event Latchup data using survival analysis," *Microelectronics Reliability*, vol. 127, 2021. [Online]. Available: https://doi.org/10.1016/j.microrel.2021.114413

[119] H. Mebrahtu *et al.*, "Heavy ion radiation damage simulations for CMOS image sensors," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5578, October 2004. [Online]. Available: https://doi.org/10.1117/12.567495

[120] S. H. Voldman, *Latchup in Semiconductor Technology*, 1st ed. Weinheim, Germany: John Wiley & Sons / Wiley–VCH, 2007.

[121] T. Aoki, "Dynamics of heavy-ion-induced latchup in CMOS structures," *IEEE Transactions on Electron Devices*, vol. 35, no. 11, 1988. [Online]. Available: https://doi.org/10.1109/16.7401

[122] C. Rossi and J.-M. Sallese, "SPICE simulation of radiation induced charges and currents in silicon substrate," in *2020 IEEE Latin America Electron Devices Conference (LAEDC)*, 2020. [Online]. Available: https://doi.org/10.1109/LAEDC49063.2020.9073366

[123] "Mu3e experiment," Webpage, Paul Scherrer Institute, 2025. [Online]. Available: https://www.psi.ch/en/mu3e

[124] AnySilicon, "Introduction to silicon germanium (SiGe) technology," Accessed: February 2025. [Online]. Available: https://anysilicon.com/introduction-to-silicon-germanium-sige-technology/

[125] H. Kroemer, "Theory of a wide-gap emitter for transistors," *Proceedings of the IRE*, vol. 45, no. 11, 1957. [Online]. Available: https://doi.org/10.1109/JRPROC.1957.278348

[126] J. Cressler and G. Niu, *Silicon-Germanium Heterojunction Bipolar Transistors*. Artech Hause, 2003.

[127] J. D. Cressler, "The silicon-germanium heterojunction bipolar transistor (SiGe HBT)," Accessed: February 2025, overview article on SiGe HBT technology, published by IEEE Electron Devices Society (EDS). [Online]. Available: https://eds.ieee.org/images/files/About/The_Silicon-Germaniun_Heterojunction_Bipolar_Transistor__-_John_Cressler.pdf

[128] IHP Microelectronics, "SiGe BiCMOS ad silicon photonics technologies," Accessed: February 2025. [Online]. Available: https://www.ihp-microelectronics.com/services/research-and-prototyping-service/mpw-prototyping-service/sigec-bicmos-technologies

[129] M. Milanesio *et al.*, "Radiation tolerance of SiGe BiCMOS monolithic silicon pixel detectors without internal gain layer," *Journal of Instrumentation*, vol. 19, no. 01, January 2024. [Online]. Available: http://dx.doi.org/10.1088/1748-0221/19/01/P01014

[130] L. Paolozzi *et al.*, "Time resolution and power consumption of a monolithic silicon pixel prototype in SiGe BiCMOS technology," *JINST*, vol. 15, no. 11, 2020. [Online]. Available: https://doi.org/10.1088/1748-0221/15/11/P11025

[131] G.-F. D. Betta *et al.*, "Simulation of 3D pixel sensors cells," 2016. [Online]. Available: https://cds.cern.ch/record/2229974

[132] J. Schwandt *et al.*, "A new model for the TCAD simulation of the silicon damage by high fluence proton irradiation," 2019. [Online]. Available: https://arxiv.org/abs/1904.10234

[133] E. Trifonova *et al.*, "Novel high-speed monolithic silicon detector for particle physics," *Journal of Instrumentation*, vol. 18, no. 05, May 2023. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/18/05/C05011

[134] J.-O. Müller-Gosewisch, "Investigation of radiation damage in silicon sensors for the Phase-2 Upgrade of the CMS Outer Tracker," Ph.D. dissertation, Karlsruhe Institute of Technology (KIT), 2021. [Online]. Available: https://publish.etp.kit.edu/record/22097

[135] M. Guthoff *et al.*, "Geant4 simulation of a filtered X-ray source for radiation damage studies," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 675, 2012. [Online]. Available: https://doi.org/10.1016/j.nima.2012.01.029

[136] H. Zhou *et al.*, "Research progress on bonding wire for microelectronic packaging," *Micromachines (Basel)*, vol. 14, no. 2, February 2023. [Online]. Available: https://doi.org/10.3390/mi14020432

[137] Altium Resources, "Wire bonding: modern applications, technology trends, and cost considerations," Accessed: December 2024. [Online]. Available: https://resources.altium.com/p/wire-bonding-modern-applications-technology-trends-and-cost-considerations

[138] Semiconductor Digest, "C4 makes way for electroplated bumps," 2001. [Online]. Available: https://sst.semiconductor-digest.com/2001/03/c4-makes-way-for-electroplated-bumps/

[139] The Ajinomoto Group Global, "Ajinomoto build-up film (ABF) innovation," Accessed: December 2024. [Online]. Available: https://www.ajinomoto.com/innovation/our_innovation/buildupfilm

[140] Semiconductor Digest, "Technology comparisons and the economics of flip chip packaging," 2001. [Online]. Available: https://sst.semiconductor-digest.com/2001/03/technology-comparisons-and-the-economics-of-flip-chip-packaging/

[141] V. L. Maggs, "Issues in new product development: the introduction of Tape Automated Bonding technology," Master of Science Thesis, Massachusetts Institute of Technology, May 1990, submitted to the Department of Mechanical Engineering and Sloan School of Management. [Online]. Available: https://hdl.handle.net/1721.1/158228

[142] PCB Technologies, "Flip chip vs. wire bonding technology," 2024. [Online]. Available: https://www.pcb-technologies.com/article/flip-chip-vs-wire-bonding-technology/

[143] O. Brand, "1.15 - packaging," in *Comprehensive Microsystems*. Oxford: Elsevier, 2008. [Online]. Available: https://doi.org/10.1016/B978-044452190-3.00019-7

[144] Budapest University of Technology and Economics, "Wire bonding: microelectronics course material," Accessed: December 2024. [Online]. Available: https://www.ett.bme.hu/meca/Courses/TEP/3_1.html

[145] M. Schuettler and T. Stieglitz, "4: Microassembly and micropackaging of implantable systems," in *Implantable Sensor Systems for Medical Applications*. Woodhead Publishing, 2013. [Online]. Available: https://doi.org/10.1533/9780857096289.1.108

[146] P. Elenius and L. Levine, "Comparing flip-chip and wire-bond interconnection technologies," *Chip Scale Review*, August 2000. [Online]. Available: https://processsolutionsconsulting.com/wp-content/uploads/2024/07/Comparing-Flip-Chip-and-Wire-Bond-Interconnection-Technologies.pdf

[147] Fraunhofer IZM, "Fine-pitch bumping for pixel detectors," Online, Accessed: December 2024. [Online]. Available: https://www.izm.fraunhofer.de/en/abteilungen/wafer-level-system-integration/leistungsangebot/fine-pitch-bumping-for-pixel-detectors.html

[148] EE Semi, "Tape Automated Bonding (TAB) in semiconductor packaging," 2006. [Online]. Available: https://www.eesemi.com/tab.htm

[149] Semiconductor Digest, "Step 2: stud bump bonding," 2005. [Online]. Available: https://sst.semiconductor-digest.com/2005/02/step-2-stud-bump-bonding/

[150] Kulicke & Soffa, "Ball bonder equipment," Accessed: December 2024. [Online]. Available: https://www.kns.com/Products/Equipment/Ball-Bonder

[151] Finetech USA, "FINEPLACER® femto 2: advanced flip-chip bonder," Accessed: December 2024. [Online]. Available: https://finetechusa.com/products/fineplacer-femto-2/

[152] Hesse Mechatronics, "Bondjet BJ653: fully automated heavy and thin wire bonder," Accessed: December 2024. [Online]. Available: https://www.hesse-mechatronics.com/en/bondjet-bj653/

[153] H. Xu *et al.*, "Wire bonding advances for multi-chip and system in package devices," *IMAPSource Proceedings*, vol. 2018, no. 1, October 2018. [Online]. Available: https://doi.org/10.4071/2380-4505-2018.1.000583

[154] Semiconductor Packaging News, "Improved wire bond reliability with auxiliary wires," February 2021. [Online]. Available: https://www.semiconductorpackagingnews.com/uploads/1/Improved_Wire_Bond_Reliability_With_Auxiliary_Wires_Final_02042021_Small.pdf

[155] R. Kisiel and Z. Szczepański, "Trends in assembling of advanced IC packages," *Journal of Telecommunications and Information Technology*, March 2005. [Online]. Available: https://doi.org/10.26636/jtit.2005.1.291

[156] B. Chylak *et al.*, "Ultra-low-loop wire bonds: interconnect solutions for multi-stacked packages," *Semiconductor Digest*, vol. 15, January 2006. [Online]. Available: https://sst.semiconductor-digest.com/2006/01/ultra-low-loop-wire-bonds/

[157] Y. Evangelista *et al.*, "Characterization of a novel Pixelated Silicon Drift Detector (PixDD) for high-throughput X-ray astrophysics," *Journal of Instrumentation*, vol. 13, no. 09, September 2018. [Online]. Available: http://dx.doi.org/10.1088/1748-0221/13/09/P09011

[158] E. Trifonova, T. Blank, and B. Leyrer, "Module assembly and operation testing at KIT," *CBM Retreat*, 2023, internal report.

[159] E. Trifonova *et al.*, "Setting up STS module production at KIT," *CBM Progress Report*, 2020. [Online]. Available: https://repository.gsi.de/record/237432

[160] J. Blaurock, "CBM Collaboration Meeting - FAIR Project Status," Presentation at the 44th CBM Collaboration Meeting, Czech Technical University in Prague, FAIR / GSI, September 2024.

[161] T. Galatyuk, "Welcome address: 44th CBM Collaboration Meeting," Presentation at the 44th CBM Collaboration Meeting, Czech Technical University in Prague, GSI / TU Darmstadt, September 2024.

[162] P. Pfistner, "Development, characterization, and analysis of silicon microstrip detector modules for the cbm silicon tracking system," Ph.D. dissertation, 2021, 54.12.02; LK 01. [Online]. Available: https://doi.org/10.5445/IR/1000140789

[163] J. Stroth, "The future at GSI and FAIR," *Nuclear Physics A*, vol. 755, 2005, proceedings of the 10th International Conference on the Structure of Baryons. [Online]. Available: https://doi.org/10.1016/j.nuclphysa.2005.03.104

[164] CBM Collaboration, "Silicon Tracking System (STS)," 2024, GSI Helmholtz Centre for Heavy Ion Research. [Online]. Available: https://www.cbm.gsi.de/detectors/sts

[165] M. Teklishyn, "Silicon Tracking System in construction," Presentation at the 44th CBM Collaboration Meeting, September 2024. [Online]. Available: https://indico.gsi.de/event/19534/contributions/82119/attachments/48602/70598/sts_in_construction_teklishyn_17SEP2024.pdf

[166] E. Trifonova and T. Blank, "STS module assembly at KIT," *CBM Collaboration Meeting*, 2021, internal report.

[167] J. M. Heuser and the CBM STS working group, "Silicon Tracking System – Summary," *CBM Progress Report*, 2023. [Online]. Available: https://repository.gsi.de/record/352779

[168] S. Mehta, "Investigation of thermal and structural integrity of Modules and Ladders of Silicon Tracking System of CBM Experiment," Ph.D. dissertation, Eberhard Karls Universität Tübingen, Tübingen, Germany, 2024. [Online]. Available: http://dx.doi.org/10.15496/publikation-104689

[169] K. Agarwal, "Thermal management of the Silicon Tracking System of the CBM Experiment at FAIR," Ph.D. dissertation, Eberhard Karls Universität Tübingen, Tübingen, Germany, 2024. [Online]. Available: http://dx.doi.org/10.15496/publikation-102494

[170] E. Trifonova, "Update on STS module assembly at KIT," 2022, internal report.

[171] T. Blank and E. Trifonova, "Module Production: steering committee meeting," 2022, internal report.

[172] R. Kapell, "Front-end electronics boards," Presentation at EDR STS Modules & Ladders, FAIR, GSI, December 2020.

[173] P. Pfistner, "High-density interconnection technologies for the CBM Silicon Tracking System," Presentation at the DPG 2019 Conference, Karlsruhe Institute of Technology (KIT), 2019.

[174] K. Kasinski *et al.*, "Characterization of the STS/MUCH-XYTER2, a 128-channel time and amplitude measurement IC for gas and silicon microstrip sensors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 908, 2018. [Online]. Available: https://doi.org/10.1016/j.nima.2018.08.076