Karlsruher Institut für Technologie

Entwicklung und Charakterisierung von Integrierten Sensoren für die Teilchenphysik

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von

M.Sc. Rudolf Schimassek

aus Reutlingen

Tag der mündlichen Prüfung:
03.12.2021

Referent: Prof. Dr. rer. nat. Thomas Müller
Korreferent: Prof. Dr. rer. nat. Ivan Perić

Karlsruhe, den 16. Dezember 2021

Deutsche Zusammenfassung

Siliziumsensoren werden in der Teilchenphysik für die Spurrekonstruktion in Beschleunigerexperimenten eingesetzt. Für künftige Experimente werden große Spurdetektoren benötigt. Für derem Realisierung sind neue Technologien wie monolithische Hochspannungs-CMOS-Pixelsensoren (CMOS – komplementäre Metall-Oxid-Halbleiter), die in kommerziellen Prozessen hergestellt werden [Per07] notwendig. Aufgrund der Integration von aktiven Schaltungselementen (wie Verstärkern und Komparatoren) in die Pixel werden diese Sensoren auch als monolithische, aktive Pixelsensoren (MAPS) bezeichnet. Die Verwendung einer hohen Depletionsspannung (im Vergleich zur Versorgungsspannung) wird im Präfix der Bezeichnung HV-MAPS kenntlich gemacht. Mit dieser Technologie können große Sensorflächen relativ konstengünstig realisiert werden [Ehr21]. Im Vergleich zu den bisher häufig genutzten hybriden Pixelsensoren ist kein Auslesechip notwendig, was auch zum Wegfall des Bump-Bondprozesses führt und Kosten und Komplexität des Aufbaus senkt.

Dieser monolithische Sensortyp wird unter anderem am KIT ASIC- und Detektorlabor (KIT-ADL) entwickelt. Die Entwicklung beinhaltet die Konzeption und Umsetzung der Schaltungen auf dem Sensor-ASIC (anwendungsspezifischer integrierter Schaltkreis), sowie die Testkampagnen der produzierten Sensoren. Für letztere sind angepasste Testumgebungen notwendig, die ebenfalls in der Gruppe entwickelt werden. Die Ergebnisse der Untersuchungen werden für die Entwicklung der nächsten Generation der Sensoren genutzt.

In dieser Arbeit wurden Beiträge zu allen dieser Schritte in der Entwicklung eines Prototypen für die Spezifikationen für das ATLAS Inner-Tracker-Upgrade [ATL17] geleistet. Der entwickelte monolithische Sensor ist die dritte Generation für diese Anwendung, der ATLASPix3. Die Beiträge sind im Einzelnen:

- Die **Konzeption der Systemebene** von monolithischen Sensoren, für die ein Simulations-Framework „ReadOut Modelling Environment“ (ROME) entwickelt und für den ATLASPix3 angewendet wurde.
- Das „GEneric Configuration and COntrol“ (GECCO) Testsystem wurde aufbauend auf bestehender Hardware [Ehr21] als schnell anzupassendes **Messsystem** entworfen. Darauf wurde sowohl ein Aufbau für einen einzelnen Sensor, als auch ein Strahlteleskop-Aufbau mit vier Sensorlagen konzipiert und umgesetzt.
- Außerdem wurde ein **Ausgleichsstrom-Messsystem** (TCT – transient current technique) aufgebaut für Messungen, die sonst – außerhalb von Teststrahlmessungen – nicht durchführbar sind.

- Der ATLASPix3-Sensor wurde mit diesen Messsystemen charakterisiert. Dabei wurden **Messungen** im **Labor** und an **Teststrahl-Einrichtungen** wie dem Deutschen Elektronen-Synchrotron (DESY) oder dem Heidelberger Ionenthreapiezentrum (HIT) durchgeführt.
- Für die nötigen **Analysen**, um aus den gewonnenen Messdaten Schlüsse auf die Eigenschaften des Sensors zu ziehen, wurden Skripte und Programme entwickelt und bestehende Analyse-Frameworks erweitert.
- Die **Verwendung des Messsystems** mit dem ATLASPix3 **an anderen Universitäten** in Deutschland, Großbritannien, Italien und China zeigt zwei Dinge: Zum einen, dass der ATLASPix3 als Kandidat für einen Einsatz in Experimenten gehandelt wird, und zum anderen, dass das GECCO-System attraktiv für erste Studien ist. Damit sorgen der Sensor und das Messsystem zusammen für eine weitere Verbreitung der monolithischen HV-CMOS-Sensoren.

Im Folgenden wird auf die Punkte genauer eingegangen.

Systemebenenkonzeption von Sensoren – ROME

Bei der Entwickung von monolithischen Pixelsensoren muss neben der elektrischen Funktionsfähigkeit auch sichergestellt werden, dass die detektierten Signale ausgelesen werden können. Bisher wurde dies durch statische Berechnungen der Datenrate und Simulationen des Entwurfs auf Hardware-Ebene abgeschätzt. Allerdings ist die Abschätzung der Fluktuationen um die Mittelwerte für die statisch berechneten Raten schwierig und die vollständige Simulation auf Hardware-Ebene sehr aufwändig. Diese Lücke wird durch ROME (Kapitel 5) geschlossen: ROME bildet die Auslesestruktur des Sensors ab und testet diese mit Signalfolgen, die beispielsweise in Monte-Carlo-Prozessen generiert werden können. Ein wichtiger Aspekt von ROME ist, dass die Beschreibung nicht direkt in einer Hardware-Beschreibungssprache geschrieben wird, sondern als XML-Baumstruktur. Dadurch ist die Simulation möglich bevor ein vollständiger Entwurf in einer Hardware-Beschreibungssprache existiert und unabhängig von Implementierungsproblemen derselben.

Damit können Abschätzungen getätigt und Optimierungen durchgeführt werden, bevor der eigentliche Hardware-Entwurf gestartet wird. Die abstrahierte Beschreibung ermöglicht es außerdem noch unerfahrenen Entwicklern Beiträge zu leisten. So wurde es erreicht, dass solche Untersuchungen auch im Rahmen von Bachelorarbeiten, deren Umfang dafür normalerweise nicht ausreicht, durchführbar werden. An dieser und den in Kapitel 5 beschriebenen Beispielen zeigt sich, dass die zu implementierenden Systeme durch kleine Änderungen der Strukturen stark beeinflusst werden können und die zuvor genutzten statischen Ratenberechnungen der Komplexität nicht gerecht werden können.

Messsystementwicklung – GECCO

Für die Charakterisierung von Siliziumsensoren gab es bereits vor der Entwicklung des GECCO-Systems verschiedene Messsysteme. Dazu gehören unter anderem das Caribou-System, das Basil-System und das YARR-System. Jedes dieser Systeme ist auf bestimmte Anwendungen hin ausgelegt.

Beim Caribou-System handelt es sich um ein vom Brookhaven National Lab, CERN und DESY entwickeltes System, das auf Open-Source-Komponenten aufbaut [LBC+19]. Allerdings hat die Variabilität des Systems zur Folge, dass die Komponenten teuer sind. So kostet das verwendete FPGA-Board des Caribou-Systems etwa zehn mal so viel wie das Board für das GECCO-System. Das GECCO-System ist außerdem auf sehr hardware-nahe Testszenarien ausgerichtet, die so direkt nicht mit dem Caribou-System möglich sind. Die

Integration der Software in den eingebetteten Prozessor des System-on-Chip ohne grafische Oberfläche erschwert die Fehlersuche zusätzlich.

Basil [SiL21] ist ein modulares Datennahme- und Systemtest-Framework. Dieses wird beispielsweise in der Auslese der ATLAS FEI4(A/B) Pixel-Auslesechips verwendet. Es zielt auf größere Systeme ab und ist im Zuge dessen ebenfalls nicht mit einer grafischen Nutzerschnittstelle ausgestattet. Entsprechend ist die unterstützte Hardware für Basil ebenfalls für den Test einzelner oder Gruppen weniger Sensoren überdimensioniert und entsprechend teurer.

YARR [Hei17] ermöglicht die Verschiebung der Komplexität der Auslese von der Firmware in die Software durch die Anbindung des FPGAs an den Rechner über den PCIe-Bus. Die hierfür verfügbare Hardware erschwert die hardware-nahe Fehlersuche und Inbetriebnahme von Sensoren und basiert wie die anderen Systeme auf höherpreisigen Systemen.

Das GECCO-System (siehe Kapitel 4.1) ist für den kostengünstigen und schnellen Aufbau von kleineren Sensor-Messsystemen gedacht. Um Probleme bei der ersten Inbetriebnahme von Sensoren schnell beheben zu können, ist das zentrale Element, die GECCO-Platine, mit Testpunkten für sämtliche Signale ausgestattet, die es ermöglichen Prüfspitzen zu befestigen.

In der Soft- und Firmware wurden Strukturen geschaffen, die die Implementierung der Kommunikation und deren Konfiguration trennen, sodass bereits validierter Code genutzt werden kann, um die Entwicklungszeit zu reduzieren. Auch wurde die Datenübertragung um einen UDP-Datenkanal erweitert, der über Gigabit-Ethernet mehr Datendurchsatz ermöglicht, was beispielsweise für Teststrahl-Messungen eine signifikante Effizienzerhöhung bedeutet.

Aus der Konfiguration der Software für den zu testenden Sensor wird ebenfalls die grafische Nutzeroberfläche automatisch generiert, sodass diese fehleranfällige Aufgabe ebenfalls wegfällt. Die Konfigurationsstrukturen ermöglichen ebenfalls eine effizientere Verwaltung von Sensorkonfigurations- und Messdaten, womit die Durchführung von Messkampagnen erleichtert wird.

Das Messsystem wird auch von Arbeitsgruppen an anderen Universitäten in mehreren Ländern für den ATLASPix3-Sensor und Quad-Module aus ATLASPix3-Sensoren [RGS+21] und für acht verschiedene Sensoren in dieser Form am KIT-ADL erfolgreich eingesetzt. Für zwei weitere Sensoren wird die selbe Hardware benutzt, die Soft- und Firmware basieren aber noch auf älteren Projekten. Insgesamt wurden bisher 65 Systeme in mehreren Chargen produziert.

Ausgleichsstrom-Messsystem

Das Ausgleichsstrom-Messsystem (transient current technique – TCT, Kapitel 4.2) wurde aus einzelnen Komponenten aufgebaut, da die erhältlichen Systeme – wie Beispielsweise das Compact TCT der Firma Alibava Systems – nicht alle Anforderungen erfüllen konnten. Für die kleinen Pixel der monolithischen HV-CMOS-Sensoren mit $25 - 200\ \mu m^2$ Kantenlänge sind Positionsgenauigkeit und Laserstrahlausdehnung von 10 µm oder besser erforderlich, um Subpixel-Auflösung zu erreichen. Weiterhin sind präzise und kurze Lichtpulse nötig, um die Zeitauflösung der Sensoren vermessen zu können.

Der Aufbau ist nicht auf ein spezielles Messsystem zugeschnitten, sondern erlaubt die Verwendung von beliebigen Systemen, solange sie in die Einhausung des Systems passen. Aus diesem Grund ist die Steuersoftware für das System aus Lasertreiber und den drei Verfahrtischen zur Positionierung des Lasers ein eigenständiges Projekt, das in andere Software-Projekte integriert werden kann und in die Software für das GECCO-System integriert ist.

ATLASPix3-Messungen und -Analysen

Die Anforderungen an einen Sensor für das ATLAS Inner-Tracker-Upgrade werden durch den entsprechenden Bericht [ATL17] festgelegt. Diese beinhalten Grenzwerte unter anderem zu Leistungsaufnahme, Homogenität der Matrix, Strahlungshärte und Randbedingungen für die Art der Auslese. Diese Eigenschaften wurden in Labormessungen und Teststrahlmessungen untersucht (Kapitel 6).

Die Sensordiode wurde durch deren Strom-Spannungs-Kennlinie und Messungen mit Elektronen aus ^{90}Sr-Zerfällen vermessen. Die dafür nötige Kalibrierung des Verstärkers erfolgte mit Röntgenstrahlung aus ^{55}Fe-Zerfällen. Mit Testsignalinjektionen, bei denen gezielt Ladung im Pixel erzeugt wird, wurde die Homogenität der Nachweiseffizienz und Kompensationsschaltung für die Schwellenanpassung untersucht, sowie Signalverzögerungen und die Amplitudeninformation (gemessen als Zeit über der Schwelle) über die Matrix vermessen. Für die Amplitudeninformation ist eine Offline-Kalibrierung vorgesehen. Die im Zuge dieser Kalibrierung gemessenen Unterschiede zwischen den Pixeln (von mehr als einem Faktor zwei) bestätigen deren Notwendigkeit. Außerdem konnten die, aus den ROME-Simulationen erwarteten, Überlastungseffekte der Auslese in Teststrahlmessungen beobachtet werden.

Zusätzlich wurde das Verhalten in Ionenstrahlen am Heidelberger Ionentherapiezentrum (HIT) untersucht, um die Entwicklung von Sensoren zur direkten Strahlbeobachtung vorzubereiten. Dabei wurde sowohl das Signal in Abhängigkeit der Teilchensorte und -Energie vermessen als auch der Einfluss inhomogener Bestrahlung untersucht. Die Ergebnisse dieser Untersuchungen werden in die Entwicklung der nächsten Prototypen für den Strahlmonitorsensor einfließen.

Die zur Analyse der aufgenommenen Daten notwendigen Skripte wurden in C++ unter Zurhilfenahme des Datenanalyse-Frameworks ROOT [BR97] geschrieben, um die Skalierbarkeit der Analysen und deren Wiederverwendbarkeit sicherzustellen. Die zur Aufbereitung der Teststrahldaten entwickelten Skripte haben die Grundstrukturen für die Analyse der Teststrahlmessungen mit den Strahlmonitorsensoren (zum Beispiel in [Web21]) gelegt.

Der Strahlteleskopaufbau wurde für eine Teststrahlmessung am DESY in Betrieb genommen und die Analysekette für die Daten mit dem Corryvreckan-Framework [DDH$^+$21] aufgebaut. Der Grund für diese Entwicklung ist, dass mit diesem System, in dem die Versorgungsspannungen und Signalleitungen geteilt werden, die Funktionsfähigkeit des ATLASPix3-Sensors in Systemen mit mehreren gekoppelten Sensoren gezeigt werden kann. Aufgrund der Einfachheit und des kompakten Aufbaus wurden Aufbau und Analysekette später auch für die Spurrekonstruktion von Sekundärteilchen, die bei der Bestrahlung mit Kohlenstoffionen am HIT entstehen, im Rahmen einer Masterarbeit [Kla22], die unter der Betreuung des Autors durchgeführt wurde, genutzt.

Karlsruhe Institute of Technology

Development and Characterisation of Integrated Sensors for Particle Physics

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von

M.Sc. Rudolf Schimassek

aus Reutlingen

Tag der mündlichen Prüfung:
03.12.2021

Hauptreferent: Prof. Dr. rer. nat. Thomas Müller
Korreferent: Prof. Dr. rer. nat. Ivan Perić

Karlsruhe, den 16. Dezember 2021

Eidesstattliche Versicherung gemäß § 13 Absatz 2 Ziffer 3 der Promotionsordnung des Karlsruher Instituts für Technologie (KIT) für die KIT-Fakultät für Physik:

1. Bei der eingereichten Dissertation zu dem Thema „Entwicklung und Charakterisierung von Integrierten Sensoren für die Teilchenphysik“ handelt es sich um meine eigenständig erbrachte Leistung.
2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.
3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In-oder Auslands als Bestandteil einer Prüfungs-oder Qualifikationsleistung vorgelegt.
4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.
5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Karlsruhe, den 16. Dezember 2021

.....

(**Rudolf Schimassek**)

Contents

Deutsche Zusammenfassung iii

1 Introduction 1

2 Particle Measurement – Applications for Tracking Detectors 5
2.1 Detector Technologies for Tracking . . . 5
2.2 The Large Hadron Collider and the High Luminosity Upgrade . . . 7
2.2.1 Structure of the ATLAS Detector . . . 9
2.2.2 The Inner Detector . . . 11
2.2.3 The Inner Tracker Upgrade . . . 12
2.3 Treatment Monitoring in Ion Irradiation Therapy . . . 14
2.3.1 Irradiation Therapy . . . 14
2.3.2 Direct Beam Monitoring . . . 15
2.3.3 Secondary Particle Tracking . . . 16

3 Silicon Detector Technology 17
3.1 Fundamentals of Semiconductors . . . 17
3.1.1 Theory of Semiconductors . . . 17
3.1.2 Doped Semiconductors . . . 20
3.1.3 PN Junction . . . 21
3.1.4 Metal-Oxide-Semiconductor Field-Effect Transistor . . . 24
3.1.5 Production of Semiconductor Circuits . . . 26
3.2 Interaction of Particles with Silicon Detectors . . . 28
3.2.1 Interaction with Photons . . . 28
3.2.2 Interaction with Heavy Particles . . . 29
3.2.3 Interaction with Light Particles . . . 30
3.2.4 Charge Generation Fluctuation . . . 30
3.2.5 Particle Scattering . . . 31
3.3 Radiation Damage in Semiconductors . . . 32
3.3.1 Bulk Damage . . . 32
3.3.2 Surface Damage . . . 33
3.4 Silicon-based Particle Detector Technologies . . . 34
3.4.1 Sensor Technologies . . . 35
3.4.2 Noise Sources in Silicon Sensors . . . 37
3.5 Monolithic High-Voltage CMOS Sensors . . . 38

4 Measurement Infrastructure 41
4.1 Generic Configuration and Control (GECCO) System . . . 41
4.1.1 Hardware . . . 43
4.1.2 Firmware . . . 45
4.1.3 Software . . . 47
4.1.3.1 Configuration-Data Container Structures . . . 49

4.1.3.2 Configuration Abstraction Layer . . . 51
4.1.3.3 Digital Readout . . . 52
4.1.3.4 Measurement Libraries . . . 52
4.1.3.5 Measurement Device Integration . . . 53
4.1.3.6 User Interface . . . 53
4.1.4 Readout via Network . . . 57
4.2 Laser Measurement System . . . 59
4.2.1 Hardware . . . 61
4.2.2 Software . . . 64

5 Readout Simulation 67
5.1 Structure of the Simulation . . . 68
5.1.1 Detector Representation . . . 69
5.1.2 Data Management . . . 71
5.1.3 Event Generation . . . 72
5.1.4 Simulation Engine . . . 73
5.1.5 User-Input to the Simulation . . . 74
5.2 Comparability of the Simulations to Beam Tests . . . 75
5.3 Grouped Readout Studies . . . 76
5.4 On-Chip Signal Sorting . . . 80
5.5 Development of the ATLASPix3 Readout Structure . . . 82
5.5.1 State Machine Optimisation . . . 85
5.5.2 Buffer Dimensioning . . . 87
5.5.3 Data Word Format Considerations . . . 92
5.6 Insights from the Simulations . . . 95

6 The ATLASPix3 Integrated Sensor 97
6.1 ATLASPix3 Design . . . 98
6.1.1 Matrix and Column Structure . . . 99
6.1.2 Pixel . . . 101
6.1.3 Readout Buffers . . . 103
6.1.4 Configuration . . . 107
6.1.5 Clocking Scheme . . . 109
6.1.6 Hit-Driven Readout . . . 110
6.1.7 Triggered Readout . . . 111
6.1.8 Bias Block . . . 112
6.1.9 Regulators . . . 113
6.1.10 Sensor Layout . . . 113
6.1.11 Design Modifications for ATLASPix3.1 . . . 114
6.2 The ATLASPix3 Measurement Setup . . . 115
6.3 Commissioning Measurements . . . 117
6.3.1 Power Consumption . . . 117
6.3.2 On-Chip Regulators . . . 118
6.3.3 Depletion Voltage . . . 122
6.4 Threshold Tuning . . . 125
6.4.1 Measurement Methods and Circuit Characterisation . . . 125
6.4.2 Matrix Tuning . . . 128
6.4.3 Noise Estimation . . . 131
6.5 Time Resolution Measurements . . . 131
6.5.1 Row Dependence of Signal Timing . . . 133
6.5.2 Matrix Timing Optimisation . . . 134
6.5.3 Time-over-Threshold Calibration . . . 137

6.6 Measurements with Particle Sources . . . 138
6.6.1 Beta Particles from Strontium-90 . . . 139
6.6.2 X-rays from Iron-55 . . . 145
6.6.3 Ion Beam Measurements . . . 147
6.7 Measurements on Irradiated Sensors . . . 156
6.7.1 Homogeneous Proton Irradiation . . . 156
6.7.2 Inhomogeneous Irradiation with Protons and Carbon Ions . . . 160
6.8 The ATLASPix3 Beam Telescope . . . 173
6.9 Telescope Measurements . . . 176

7 Conclusions **183**

Appendix **187**
A Collection of alternatives to GECCO . . . 187
A.1 Caribou Readout System . . . 187
A.2 Basil . . . 188
A.3 YARR . . . 188
B Configuration of ATLASPix3 . . . 188
C ATLASPix3 Quad-Module . . . 189
D Threshold Tuning of a proton-irradiated ATLASPix3 . . . 190
E Matrix Timing on ATLASPix3 . . . 192
F Amplifier Saturation on ATLASPix3 . . . 192
G Further Leakage Current Measurements on ATLASPix3 . . . 193
H Threshold Dispersion on a non-irradiated ATLASPix3 . . . 194
I ATLASPix3.1 Leakage Current . . . 194
J Timestamp Error Effects . . . 195

Abbreviation Index **197**

Publications **199**

Bibliography **203**

List of Figures

2.1 Principle of particle identification with a tracker in a magnetic field 6
2.2 Structure of LHC overlayed on an aerial view of the region 8
2.3 Roadmap towards HL-LHC 8
2.4 ATLAS detector structure 10
2.5 Structure of the inner detector of ATLAS 11
2.6 Structure of the ATLAS ITk upgrade 12
2.7 Signal rate simulation for inclined modules in ATLAS ITk upgrade 13
2.8 Energy deposition profiles of different particles in matter 15

3.1 Band structure of silicon 18
3.2 Brillouin zone of silicon 19
3.3 Fermi distribution for different temperatures 20
3.4 Band model of the pn junction 22
3.5 Operation modes of a MOSFET 24
3.6 IV characteristics of an n-channel MOSFET 25
3.7 Top view of the structure of a linear MOSFET 26
3.8 Czochralski process for growing mono-crystal silicon 26
3.9 Zone melting scheme to convert poly-silicon into a mono-crystal 27
3.10 Wafer type identification marker positions 27
3.11 cross section for photon interaction with carbon 29
3.12 Visualisation of the stopping power of copper for muons 30
3.13 Interaction processes of electrons with matter 31
3.14 Types of lattice defects 32
3.15 MOSFET implementation shapes 34
3.16 Hybrid silicon detector types 35
3.17 Charged particle interaction with silicon sensors 36
3.18 Structure of a silicon 3D pixel sensor 36
3.19 Signal chain in a monolithic HVCMOS sensor 38
3.20 Well structure of HVCMOS sensors 39
3.21 Simplified schematics of the CSA 39
3.22 Time-walk effect visualisation 40

4.1 GECCO Logo 43
4.2 Signal Diagram of the GECCO board 43
4.3 Photographs of the GECCO board 45
4.4 GECCO Firmware Module Structure 46
4.5 GECCO Software Structure 48
4.6 User Interface of the GECCO software 55
4.7 Tests section of the GECCO software user interface 56
4.8 Data processing structure of the UDP reaodut software 57
4.9 UDP Readout User Interface 58
4.10 UDP Readout Live Data Visualisation 59

4.11 Light absorption in semiconductors . . . 60
4.12 Microfocus of the laser setup . . . 62
4.13 Flanges on the Laser Setup box . . . 62
4.14 Hardware inside the box of the laser setup . . . 63
4.15 The Laser Setup Control GUI . . . 65

5.1 ROME Logo . . . 68
5.2 ROME Processing Structure . . . 69
5.3 Analogue Characteristics for Signal Generation . . . 72
5.4 Example State Machine Visualisation from ROME . . . 73
5.5 Comparison of pixel efficiency and particle efficiency . . . 76
5.6 Possible group shapes for 12 pixels . . . 77
5.7 Readout delay distribution for grouped readout . . . 78
5.8 Influence of the group shape on the readout speed . . . 79
5.9 Zoom-in on the readout delay distribution of grouped readout . . . 79
5.10 Readout procedure comparison . . . 81
5.11 Readout Performance depending on grade of time sorting . . . 82
5.12 ROME architecture for the ATLASPix3 simulations . . . 84
5.13 Aurora first-in-first-out buffer (FIFO) fill state for state machine optimisation 86
5.14 State machines for the simulation of the reaodut architecture of ATLASPix3 87
5.15 Encoder FIFO fill state over time . . . 89
5.16 Column buffer content over time per column . . . 89
5.17 Column buffer dependence on triggered fraction and trigger signal delay . . 90
5.18 Influence of the trigger table on the readout delay . . . 91
5.19 Trigger Table Fill State in final design . . . 92
5.20 Particle efficiency dependency on data word size . . . 93
5.21 Lost triggers in ATLASPix3 readout over trigger rate . . . 94
5.22 Influence of the data word on the readout performance of ATLASPix3 . . . 94

6.1 Layout of ATLASPix3 . . . 99
6.2 Readout Architecture of ATLASPix3 . . . 100
6.3 Pixel to hit buffer connection layout on ATLASPix3 . . . 101
6.4 Schematic of the Pixel Amplifier . . . 102
6.5 Schematics of the in-pixel comparator of ATLASPix3 . . . 102
6.6 ATLASPix3 Pixel Layout . . . 103
6.7 Simplified Schematic of the hit buffer of ATLASPix3 . . . 104
6.8 Implementation of the priority chain on ATLASPix3 . . . 105
6.9 Simplified schematics of the content addressable memory of ATLASPix3 . . 106
6.10 Schematic of a shift register bit for configuration . . . 109
6.11 Clocking Scheme of ATLASPix3 . . . 110
6.12 Readout state machine for hit-driven readout on ATLASPix3 . . . 111
6.13 Readout state machine flow chart for triggered readout of ATLASPix3 . . . 112
6.14 Layout of the Readout Structures on ATLASPix3 . . . 114
6.15 Layout of the periphery of ATLASPix3 . . . 115
6.16 Layout of the ATLASPix3 chip carrier for the GECCO measurement setup 116
6.17 Photograph of the ATLASPix3 measurement setup . . . 116
6.18 VDDA Regulator on ATLASPix3.1 . . . 120
6.19 Characteristics of the VSSA regulator on ATLASPix3 . . . 121
6.20 V_{minus} regulator output over the control DAC value . . . 121
6.21 IV Curve of the Sensor Diodes of ATLASPix3 . . . 122
6.22 Well geometry of ATLASPix3 . . . 123
6.23 Temperature dependence of the leakage current on ATLASPix3 . . . 123

6.24 Breakdown voltage shift with temperature on a non-irradiated ATLASPix3 124
6.25 IV characteristics of the HitPix detector ASIC 124
6.26 Detection Threshold determination with an SCurve 126
6.27 TDAC Effect on Detection Threshold on ATLASPix3 126
6.28 Effect of VNDAC on the TDAC step size 127
6.29 V_{minus} dispersion of the TDAC cross-over point 127
6.30 Threshold tuning result on ATLASPix3 129
6.31 TDAC setting distribution after threshold tuning 130
6.32 Noise distribution after tuning on ATLASPix3 130
6.33 Noise Correlation with row number on ATLASPix3 131
6.34 Pixel Timing Key-Hole Measurement 132
6.35 Row Dependence of the Signal Delay 133
6.36 Delay Range for Timing Optimisation of the Matrix with TDACs 135
6.37 Timing Tuning Result Comparison 135
6.38 Result TDAC Distribution for Timing Tuning 136
6.39 Pixel Delay Comparison for Threshold and Timing Tuning 136
6.40 ToT Dependence on Signal Size 138
6.41 ToT Distribution for a fixed signal size for the whole matrix 138
6.42 Strontium measurement signal vetoing 140
6.43 Signal height and length correlation for ^{90}Sr electrons on ATLASPix3 . . . 140
6.44 Strontium-90 signal height distribution with background 141
6.45 Signal Height distribution from ^{90}Sr without background 141
6.46 Baseline histogrammed for the amplifier output on ATLASPix3 142
6.47 Signal length distribution for ^{90}Sr signals on ATLASPix3 142
6.48 Signal height distribution on ATLASPix3 for ^{90}Sr at 20 V depletion 143
6.49 Hitmap of ^{90}Sr on ATLASPix3 144
6.50 ToT distribution for ^{90}Sr on ATLASPix3 144
6.51 ToT dispersion with depletion voltage for ^{90}Sr on ATLASPix3 145
6.52 ^{55}Fe Signal Height and Length Correlation 146
6.53 ^{55}Fe Signal Height Histogram 146
6.54 Test signal injection size correlation on ATLASPix3 147
6.55 Characteristics for the test signal injections on ATLASPix3 147
6.56 Leakage current measurement on ATLASPix3 during a beam test 149
6.57 Beam induces leakage current for different beam energies 150
6.58 Leakage current on ATLASPix3 for different beam intensities 151
6.59 Sketch of the shielding used for the beam at HIT 151
6.60 Clustersize for a carbon beam at HIT on ATLASPix3 152
6.61 Clustersize dependency on the particle energy for a proton beam on ATLASPix3 152
6.62 Clustersize dependency on particle energy for a carbon ion beam 153
6.63 Cluster shape for a carbon beam after the copper target in front of ATLASPix3 153
6.64 Carbon beam hitmap on ATLASPix3 154
6.65 Trigger buffer limited beam profile on ATLASPix3 155
6.66 Column readout over time on ATLASPix3 155
6.67 Leakage Current IV curves for proton irradiated ATLASPix3 157
6.68 Noise dispersion on a proton irradiated ATLASPix3 158
6.69 SCurves on ATLASPix3 after proton irradiation 159
6.70 Threshold and noise distribution on a tuned ATLASPix3 after irradiation . 160
6.71 Hitmap of an iron-55 source on a proton-irradiated ATLASPix3 161
6.72 Irradiation spot shape for ion irradiation at HIT 163
6.73 Sensor leakage current on ion irradiated ATLASPix3 integrated sensors at HIT 164
6.74 Sensor breakdown for ion irradiated samples of ATLASPix3 164
6.75 Noise hit rate on an irradiated ATLASPix3 165

6.76 Noise map for AP3.H1 in depletion scans . . . 166
6.77 Threshold dispersion for temperature on ion irradiated ATLASPix3 167
6.78 Leakage current dispersion compared to pixel noise . . . 167
6.79 Detection efficiency for ^{90}Sr signals on ATLASPix3 after ion irradiation . . 168
6.80 Detection efficiency for laser signals on ATLASPix3 after ion irradiation . . 170
6.81 Detection efficiency recovery after irradiation . . . 171
6.82 Threshold scan for pixel recovery after irradiation . . . 172
6.83 Fraction of working pixels per threshold setting on AP3.H5 . . . 173
6.84 Photograph of the ATLASPix3 beam telescope . . . 174
6.85 Layout of the telescope base board for the ATLASPix3 telescope . . . 174
6.86 Readout Scheme of the Telescope System . . . 176
6.87 Beam profile measured at DESY with ATLASPix3 . . . 177
6.88 Clustersize on ATLASPix3 for the electron beam at DESY . . . 178
6.89 ToT correlation to the secondary timestamp . . . 178
6.90 Charge calibration for the electron beam at DESY . . . 179
6.91 Correlation between the telescope layers at DESY beam test . . . 180
6.92 Time correlation of the data taken at DESY . . . 180
6.93 DUT residuals in DESY beam test . . . 181
6.94 Time-walk measured with the telescope at DESY beam test . . . 181

C.1 Photograph of the ATLASPix3 quad-module . . . 190
D.2 Noise correlation with the row index on a proton-irradiated ATLASPix3 . . 191
D.3 Threshold map of a proton-irradiated ATLASPix3 integrated sensor . . . 191
D.4 Noise map of a proton-irradiated ATLASPix3 sample . . . 192
E.5 Test injection timing on ATLASPix3 . . . 192
F.6 Saturated amplifier for a ^{90}Sr electron signal on ATLASPix3 . . . 193
G.7 Lakage current on AP3.H5 . . . 194
H.8 Threshold dispersion over temperature on a non-irradiated ATLASPix3 . . 194
I.9 IV-Curve of ATLASPix3.1 . . . 195
J.10 Decoding error in ATLASPix3 data . . . 196

List of Tables

4.1 Key parameters of the laser measurement setups 61

6.1 Power consumption on a non-irradiated ATLASPix3 118
6.2 Irradiation doses for the ATLASPix3 samples from HIT 161

1 Introduction

High-energy physics (HEP) needs large accelerators to reach the energies necessary to probe the properties of elementary particles [EB08, MAB+18, CEP18]. Precision measurements with these machines require large detectors and large statistics. For this reason, experiments strive for higher event rates and energies to obtain more data from particle collisions and by this testing for less probable interactions, too [ABF+20]. As large projects, planning and construction of these experiments typically takes years to decades, it can be expected that new technologies are developed before the start of operation with which performance can be improved. This development is fostered for the experiments guided by scientific curiosity which is limited by available technology. To obtain more insights for the effort, these accelerators and detectors are refined and upgraded as the technologies get ready and the operation schedule allows for it [CDE+10, ATL17].

One example are silicon sensors for particle tracking [Har09]: The common implementation of them is using a sensor die and another die for readout which are connected via one-to-one connections from the segments to the readout circuits. Each die of this hybrid process is manufactured in a highly optimised process to maximise detector performance. This means that large sensitivity and high radiation tolerance can be achieved. The drawback of these sensor modules is that both the silicon chips and the assembly – a fine-pitch bump bonding process in the case of pixel sensors – are expensive. Consequently, this detector type is only used where it is necessary, resulting in small detectors or utilisation of more affordable technologies at positions with less harsh conditions.

Basing a silicon sensor technology on a common manufacturing process, Prof. Dr. Ivan Perić invented a new sensor concept in 2007 [Per07]: This high-voltage complementary metal-oxide-semiconductor (HV-CMOS) sensor technology is based on a commercial HV-CMOS process integrating both sensing elements and electronics on the same die. Progressing the integration of more functionality into the sensor, the integrated sensor can become a fully monolithic sensor device. On a commercial process, such a monolithic device offers the opportunity of cost and material budget reduction compared to hybrid sensors at the same time. This is because not only half of the silicon wafers, but also the assembly step of precision bump bonding is omitted. Despite this change, good time and spatial resolution can be expected from charge collection by drift and the omission of the bump bonds limiting the pixel size. The cost reduction opens up the possibility to build large silicon-based tracking detectors meeting the demands for more powerful tracking detectors while improving other aspects as the material budget, too. With this technology, smaller experiments with limited budget become viable [RAD+21, Ehr21].

The main target applications, towards which monolithic HV-CMOS sensors have been developed, are accelerator experiments as the Large Hadron Collider (LHC) for example in the tracking sub-detectors of A Toroidal LHC ApparatuS (ATLAS) or the tracking detector of the Mu3e experiment. However, the technology can also be used for other applications as beam monitoring or radiation control.

Since the invention of the technology, the sensor application-specific integrated circuits (ASICs) have grown in size and complexity, creating the necessity to concept the development cycle as a collaborative effort. This development cycle consists of the design of the sensor ASIC according to specifications from the application and the manufacturing process used, the production of the ASIC at a foundry, the development of a test system for it, the execution of measurements, their analysis and evaluation of the results for feedback on performance. If the design meets all requirements, the development cycle is finished. If it does not, or the design was not intended to meet all requirements – as for example a reduced size in test sensors or systems – the next iteration of the design is prepared with the input from the characterisation.

The design on a commercial HV-CMOS process make the production an external task to be handled by the foundry charged with it. Apart from this step, the author contributed at all steps of the development cycle during the PhD thesis:

- For the **design stage**, a simulation framework for conception and dimensioning of readout architectures has been developed and applied to a design which was then implemented by Prof. Perić and other designers inside and outside the KIT ASIC and Detector Laboratory (KIT-ADL) group.
- To **prepare the tests** of the produced ASIC, the test system has been developed based on the GEneric Configuration and COntrol (GECCO) setup developed for [Ehr21], improving adaption speed and extending the functionality. This comprises hardware to connect the sensor die, firmware and software development.
- This setup has then been used for **measurements** in the **laboratory**. This comprises measurements with electrical test signals, lasers and radioactive sources. **Beam tests** have been performed with the setup at Deutsches Elektronen-Synchrotron (DESY) [DDE+19] and at Heidelberger Ionen Therapiezentrum (HIT) [OW08].
- In more mature designs, also the construction of modules belong to the validation campaign, which has been done in the form of a **beam telescope** for which hardware, firmware and software have been developed.
- To make use of the **data** recorded in the measurements, it has to be **analysed**. For this, scripts and own software have been developed and common software packages have been used and extended for elaborating and extracting the parameters of interest and visualising the results.

For the application as particle sensor, the important parameters to test are sensitivity or detection efficiency, as well as time and spatial resolution. For high event rates and long run-times in these conditions, also rate capabilities and radiation tolerance are important. Furthermore, towards a final sensor ASIC, the suitability for **system building** becomes important as a successful HEP experiment can only work with many sensors working as one system.

This thesis will present the work done on the design of sensor ASICs, the preparation of ASIC tests and the testing of one of the latest prototype sensors filling the whole mask size of the manufacturing process – ATLASPix3. The content is structured to narrow down through the thesis ending at the measurements specifically done for ATLASPix3.

At first in chapter 2, the context of the monolithic HV-CMOS sensor development is explained in an analysis-synthesis process: Starting from the field of HEP and irradiation therapy digging down to the application of these sensors before building up the theory behind the sensors themselves in chapter 3.

Then, the work on the GECCO system is presented in chapter 4. Also, the transient current technique (TCT) setup, built to extend the range of tests that can be done in the laboratory, is presented.

The contribution to the design effort of ATLASPix3 is presented in chapter 5 with the simulation framework developed for the analysis of readout architectures. After the description of the framework itself and how it compares to measurable parameters, its concept is shown on two example systems with readout of pixel groups and signal sorting. Afterwards, the simulation campaign for the readout of ATLASPix3 is presented.

The last part before the conclusions, in chapter 6, are the measurements on ATLASPix3 building up in complexity and addressing the performance measures for particle sensors.

At the end, a conclusion of the work done with the key findings is given in chapter 7.

2 Particle Measurement – Applications for Tracking Detectors

This chapter bridges the gap from general applications – with the examples HEP and medicine with radiation treatment – to the silicon sensors that can be built to satisfy the needs of these. For this, the context of the tracking detector application on the example of the ATLAS experiment at LHC is broken down to the requirements of a silicon sensor to be used there. Also, the environment for a beam monitor in irradiation therapy of tumors is explained. But first, technology options for tracking detectors and their usage for momentum measurement are described. Silicon sensors, that can be employed for these applications, will be explained in chapter 3.

2.1 Detector Technologies for Tracking

Simple detector types as cloud chambers or bubble chambers have been replaced in current experiments with detector types that are directly converting signals to electronic signals simplifying the analysis process and largely increasing the possible measurement speeds [KW16].

One class of detectors used for tracking are gaseous detectors. They detect particles from ionisation of gas contained. This type of detector has the advantage that large volumes can be covered at rather low cost and with little material introduced. The charge generated in such a detector is separated by an electric field and collected at electrodes that generate the electrical signal. Single particle detection is achieved by gas amplification at the anode by the use of very large fields to create an avalanche of secondary particles. The wire chamber is one form of such a detector with improved spatial resolution. Spatial information is gained from the wire the charge is collected at. Hence, the spacing of these wires limits the resolution of the chamber. The resolution is only good perpendicular to the direction of the wires, to obtain information about the second dimension, several ionisation chambers with perpendicular orientation can be used.

Another type are spark chambers with small distance between the plates where a high, homogeneous field is applied between parallel plates. The advantage of this arrangement is that charge does not have to drift to the volume of the high field to start an avalanche but the avalanche can start anywhere in the volume. This grants high time resolution. In resistive plate chambers (RPC), the collected charges are detected capacitively by strips on the back-side of the resistive material of the plates.

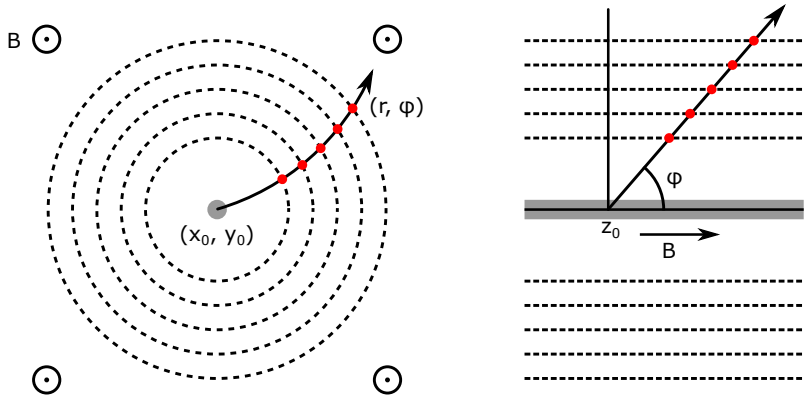


Figure 2.1: The schematic drawing shows the arrangement of detector layers (as dashed lines) around a beam pipe with the collision point (grey area). In the magnetic field along the beam direction, trajectories of charged particles are bent. From the signals (red dots), the trajectory can be reconstructed and the charge and momentum reconstructed from this. After [KW16], from [Ehr21]

More recently than the previous detector types [KW16, Har09], semiconductor based detectors are being used for building particle detectors. The generation energy for a charge carrier pair is with 3.6 eV an order of magnitude smaller in silicon compared to typical values for gas detectors at 30 eV. The charge is collected in the depletion zones of diodes that are reverse-biased. Being largely available [Shi89], silicon based sensors are the most common variant of semiconductor detectors. Segmentation is possible in different ways: in long strip-shaped segment and in areas with short edges in both directions. In both cases, the sensing diodes are connected to readout electronics individually. For the strip sensors, this can happen via wire-bonds to an external readout chip placed at one or both ends of the strip sensor. To establish a one-to-one connection of the segments of a pixel sensor to a readout chip, bump bonds in a flip-chip process have to be used.

For particle identification with a tracking detector, all of these detector types have to be implemented in several layers in a magnetic field to bend the tracks of the particles. A neutral particle will show no bend in the track in radial direction, charged particles will show bent tracks according to the charge to momentum relation. The arrangement for a barrel tracker is shown in figure 2.1. Together with the energy measurement from the calorimeters, particle identification is possible: The energy E consists of the contributions of mass m and momentum p related via the speed of light c via

$$E = \sqrt{(mc^2)^2 + p^2c^2} \tag{2.1}$$

which can be rewritten as

$$m = \sqrt{\frac{E^2}{c^4} - \frac{p^2}{c^2}} \ . \tag{2.2}$$

The Lorenz force $\vec{F} = q(\vec{v} \times \vec{B})$ can be rewritten with $\vec{F} = \dot{\vec{p}} = \gamma m \vec{v}$ to give the differential equation

$$\dot{\vec{v}} = \frac{q}{\gamma m} \left(\vec{v} \times \vec{B} \right) \ . \tag{2.3}$$

There, γ is the Lorentz factor which is $\gamma = (1 - \beta^2)^{1/2}$ with $\beta = v/c$. This equation can be solved by a helical path around the magnetic field direction. The integration over time can be used to calculate the radius of the helix to give

$$R = \frac{p_T}{|q|B} \qquad \Leftrightarrow \qquad p_T = |q|B\,R \tag{2.4}$$

with the transversal momentum p_T and magnetic flux density B. Transversal in this case denotes perpendicular to the magnetic field. The only left parameter is the charge which is an integer times elementary charge and the sign is extracted from the rotation direction in the field.

2.2 The Large Hadron Collider and the High Luminosity Upgrade

Although larger colliders are being planned, the Large Hadron Collider (LHC) is the largest particle accelerator built to date with a length of nearly 27 km [EB08]. The Future Circular Collider (FCC) [MAB+18] and Circular Electron-Positron Collider (CEPC) [CEP18] with 100 km storage rings are still in planning phase.

LHC is designed to collide two particle beams at four interaction points at which the experiments are placed. Using proton beams, the collisions happen at a center of mass energy up to 14 TeV. Using lead ions, 1148 TeV are possible. These energies are achieved using pre-accelerators. After the initial acceleration in a linear accelerator (LINAC 2), three circular accelerators are used: Proton Synchrotron Booster (PS BOOSTER), Proton Synchrotron (PS) and Super Proton Synchrotron (SPS). The first two accelerators in the chain are replaced by LINAC 3 and the Low Energy Ion Ring (LEIR) for heavy-ion operation.

Being located around Geneva at the border between France and Switzerland, the tunnel, that has been built for the Large Electron-Positron Collider (LEP), is between 50 m and 175 m under ground and accessible via several access tunnels. The two large general-purpose detectors ATLAS [AAA+08] and Compact Muon Solenoid (CMS) [CMS08] are located in caverns build for LHC. The two smaller and more specialised experiments – A Large Ion Collider Experiment (ALICE) [ALI08] and Large Hadron Collider beauty (LHCb) [LHC08] – are located in the caverns built for the experiments at LEP. The structure of LHC with the mentioned experiments is shown in figure 2.2.

Inside a ring accelerator, it can be more handy to have the particles grouped to short groups instead of a continuous stream. These groups are typically referred to as bunches.

For a particle accelerator, there are two key numbers: The energy of the particles in the collisions and the number of collisions accessible to experiments. The latter is quantified in luminosity $\mathcal{L}$. It describes the number of particles per area and time interacting and evaluates to

$$\mathcal{L} = \frac{N_1 N_2 n_b f}{A} \tag{2.5}$$

with the number of particles in the respective bunches N_1 and N_2, the number of bunches in the ring n_b, the revolution frequency f and the cross sectional area in the collision spot A. This assumes that the particles in the beam pipes are formed to bunches that arrive at the interaction points at a frequency of 40 MHz in the case of LHC [EB08]. Integrating the luminosity over time, the integrated luminosity $\mathcal{L}_{\text{int}}$ is obtained, which is commonly measured in inverse femto barn (fb^{-1} with $1\,\text{fb} = 10^{-28}\,\text{m}^2$).

The design value for the instantaneous luminosity of LHC is 10^{34} $\text{cm}^{-2}\text{s}^{-1}$ which was reached in 2016 and doubled before the beginning of the second long shut down in 2018 [HLL21]. These planned stops are to repair and upgrade accelerator and experiments. After the third run of data taking until end of 2024, LHC will undergo large upgrades for both accelerator and experiments to reach up to 7.5 times the initial design luminosity. These upgrades are prepared in the High-Luminosity Large Hadron Collider (HL-LHC) project.

The roadmap towards HL-LHC in figure 2.3 shows the time line towards operation of HL-LHC.

Figure 2.2: The location of the LHC around the CERN site nearby Geneva is shown as an overlay of an aerial view of the region. The large experiments ATLAS and CMS are located at opposite points on the ring while the smaller experiments ALICE and LHCb are located closer to the main site in caverns built for LEP. Credit: CERN, from [All13]

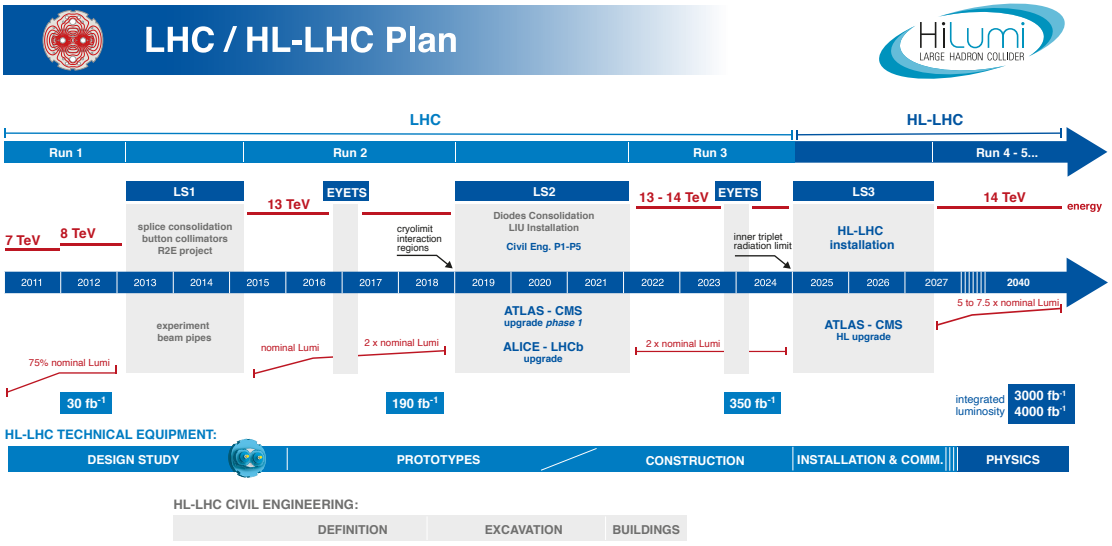


Figure 2.3: Since the initial startup in 2011, the LHC has been gone into several planned long shut downs to perform upgrades and repairs to maintain and improve the performance as visible from the achieved or envisaged luminosity. [HLL21]

The bunches in the two rings are focused for passing the interaction points at the experiments. This led to an average of 25 proton-proton collisions per bunch-crossing during the first run of LHC. With the timestamping recreating the bunch-crossing frequency, the collision products from these can not be disentangled using the timestamp of the signals. Only from reconstruction of the events they can be separated. This overlaying of events is referred to as pile-up. In the second run, pile-up was increased to an average of 32 collisions per bunch-crossing with peaks at 50 for short periods of time [CMS18].

For HL-LHC, a pile-up of 200 is expected. These numbers are not only a large increase requiring changes in the accelerator complex, but also the detectors need to be upgraded to cope with this change. The integrated luminosity is expected to reach 3000 fb^{-1} or in a best-case scenario even 4000 fb^{-1} until 2040. [ABF+20]

On the accelerator side, the increase of instantaneous luminosity to $5 \cdot 10^{34}$ $\text{cm}^{-2}\text{s}^{-1}$ or even $7.5 \cdot 10^{34}$ $\text{cm}^{-2}\text{s}^{-1}$ is achieved my increasing the number of protons per bunch to almost twice the current value at then $2.2 \cdot 10^{11}$ particles per bunch. Despite the slightly decreased number of bunches in the ring, the total number of particles in the ring is almost doubled from $3.2 \cdot 10^{14}$ to $6.0 \cdot 10^{14}$. The emittance, describing the beam size in position-momentum phase space, is also reduced by a third from 3.75 µm to 2.50 µm.

Next, the ATLAS detector as used for run 2 is described before the improvements on the tracking detector for HL-LHC are described in section 2.2.3.

The goal of a general-purpose detector for collision products is to identify all particles produced in the collisions. This is done measuring energy and trajectories of all collision products, allowing for reconstruction of charges, masses and speeds for particle identification. To measure these different parameters, the detectors are divided into sub-detectors carefully arranged around the interaction point to minimise the changes introduced for the sub-detectors on the outside of the sub-detector stack. This arrangement depends on the focus of the experiment and design choices. The symmetry of the detector is also given by the type of accelerator: A symmetric accelerator like LHC[1] colliding the same particles at the same energies results in a symmetrical detector system with the symmetry axis along the beam direction. Fixed-target colliders however will have different requirements upstream and downstream of the collision point resulting in asymmetric detectors. The beam nature leads in most cases to detector arrangements with a rotational symmetry around the beam direction.

2.2.1 Structure of the ATLAS Detector

At the center with smallest distance from the interaction point, the tracking sub-detector is located in the A Toroidal LHC ApparatuS (ATLAS) detector. Together with a magnetic field bending the trajectories of charged particles and the surrounding calorimeters, energies and bending radii determine the particle type and enable back-tracing of the particles through vertices for reconstructing events with pile-up. The calorimeters divide into electromagnetic calorimeters and hadronic calorimeters. Hadrons are not stopped in the electromagnetic calorimeter and will generate a signal in the hadronic calorimeter enabling differentiation between hadrons and lighter particles that are stopped in the electromagnetic calorimeter. To distinguish muons from hadrons, the outermost sub-detectors are muon chambers. Since a muon is not stopped in the hadronic calorimeter either, a signal from the muon chambers enables separation of muons from hadrons that are stopped in the hadronic calorimeter.

[1] LHC can be operated in symmetric proton-proton mode or with lead ions. But asymmetric modes as proton-lead-ion are possible, too.

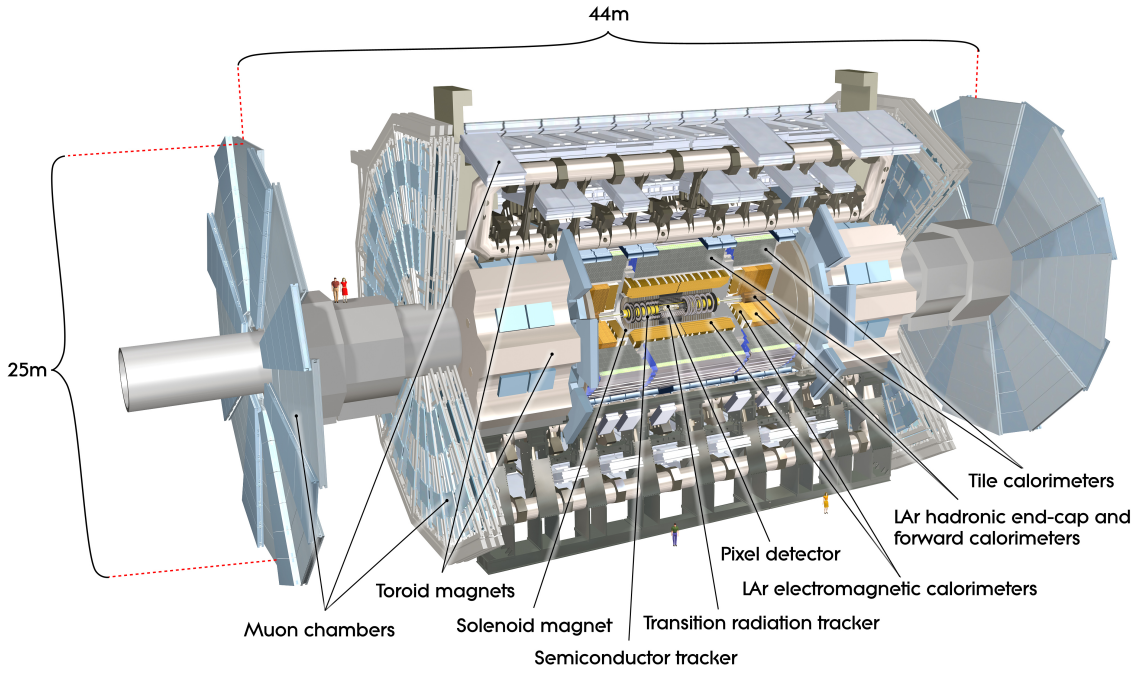


Figure 2.4: This cross section visualisation shows the sub-detectors of the ATLAS detector. Four persons are included in the figure for size reference. [Peq15]

The whole ATLAS detector has a cylindrical shape with a diameter of 25 m and a length of 44 m. The structure of the ATLAS detector is shown in figure 2.4. For size comparison, several people are added in the visualisation. The concepts comprises two structures: The barrel enclosing the interaction point with the symmetry axis being the beam direction and the disk shaped end-caps.

The magnets of ATLAS consist of a thin superconducting solenoid surrounding the inner detector and three large superconducting toroids with eight-fold symmetry around the calorimeters. One toroid is located around the barrel and the other two are located at the end-caps. The magnetic solenoid field for the inner detector is 2 T [AAA+08].

The calorimeter system of ATLAS consists of the electromagnetic calorimeter (EMCal) and the hadronic calorimeter (HCal) arranged concentrically around the inner detector and solenoid. These are built from alternating layers of absorbing and sensitive material along radial direction in the barrel and along the beam direction in the end-caps. Consequently, a particle with more energy will generate signals in more sensing layers than a particle with less energy. Both calorimeter systems use liquid argon (LAr) as sensing material. The absorbing material for the EMcal is lead, for the HCal it is stainless steel. Since both the solenoid magnet around the inner detector and the LAr require low temperatures and to reduce the material in the detector, the solenoid magnet and the EMCal share the vacuum vessel so that two vacuum walls can be omitted.

Apart from muons – and weakly interacting particles as neutrinos, no particles make it past the HCal. Consequently, any signal past the HCal can be considered to originate from muons. The muons are measured with the muon spectrometer. With the magnetic field, the tracks of the muons are bent to measure momentum and charge. Monitored drift tubes (MDT) and cathode strip chambers (CSC) are used as sensing elements. In addition, resistive plate chambers (RPC) are used in the barrel for triggering. In the end-caps, this is done with thin gas-chambers (TGC). Detailed descriptions of the detector types can be

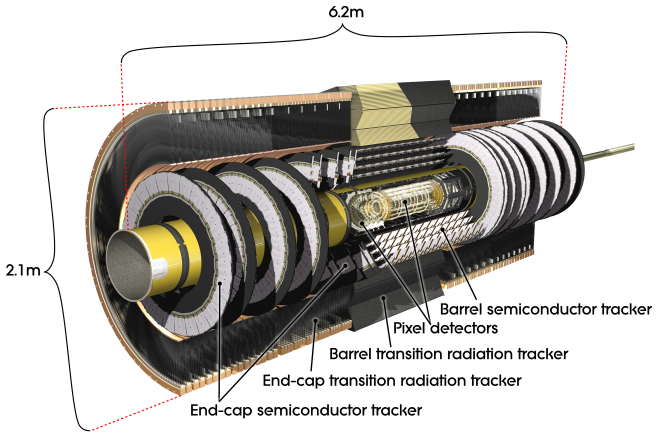


Figure 2.5: The inner detector consists of four pixel layers as barrel and end-caps close to the interaction point and the semiconductor tracker consisting of strip detector layers.

found for example in [KW16].

The inner detector contains several layers of silicon pixel and strip detectors and will be described in more detail in the next section.

A crucial part of the detector system is the event selection or triggering and the corresponding data acquisition system. With the 1.7 billion proton-proton collisions per second, the ATLAS detector generates an estimate of 60 PB per second [ATL21]. The trigger system, consisting of a hardware trigger (level-1 trigger) selects events from calorimeter and muon system information using custom electronics. The trigger decision has to be made within 2.5 µs from the event to extract the data for the event from the data stream of the detector. The level-1 trigger reduces the number of events to further evaluate from 40 million to up to 100 000 events per second. The second stage of trigger is a software trigger implemented on a large CPU farm. The software selects about 1 000 events per second for which the data from all sub-detectors is combined and sent to a storage system for off-line analysis.

2.2.2 The Inner Detector

The inner detector consists of a barrel section and end-caps like the other sub-detectors and has a diameter of 2.1 m at a length of 6.2 m. Each section consists of several layers of detectors. As for the particle density, sensors closest to the interaction point are required to have the highest spatial and time resolution. The structure of the inner detector is shown in figure 2.5. On the outer parts of the tracking detectors, the rate is reduced from distance to the interaction point so that lower granularity detectors can be used which reduces cost and complexity. Furthermore, the detectors close to the interaction point have to withstand more radiation damage over the lifetime of the detector.

For the innermost layers, the data taking puts more constraints on material and power budget as more material means more scattering diminishing the resolution of the whole system and on power as cooling structures also introduce material in the system.

The innermost layers are three pixel layers that were extended by a fourth layer between the new (thinner) beam pipe and the, until then, first layer during the first long shutdown in 2014 [CDE+10, ATL12, Per14]. The radii from the interaction point to the layers are 34 mm, 50.5 mm, 88.5 mm and 122.5 mm. The end-caps consist of four disks per side at 400 mm, 495 mm, 580m and 650 mm from the interaction point.

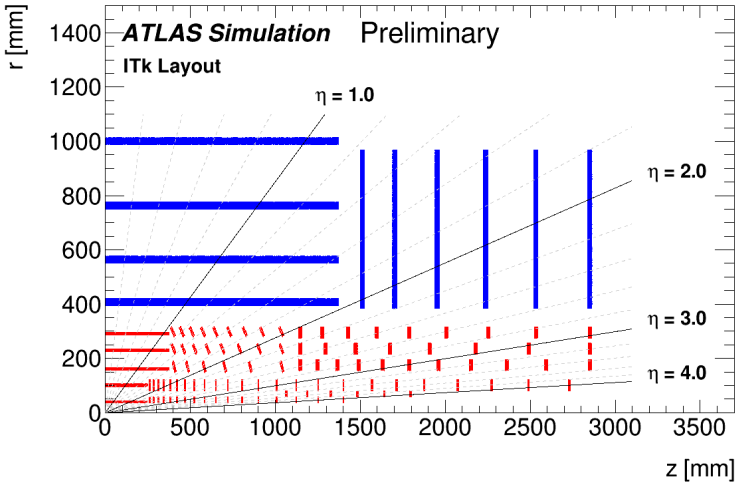


Figure 2.6: The structure of one quarter of the inner tracker upgrade as projected is shown. The pixel sub-system is drawn in red, the strip subsystem in blue. The detectors are arranged that for any track at least nine measurements are recorded. [ATL19]

Around the pixel tracker, the semiconductor tracker is placed. It consists of strip detectors with a pitch of 80 µm and a length of 128 mm. These silicon strip detectors reduce the complexity of the system as the number of channels is reduced by a factor of 1600 compared to squared pixels at the same pitch of 80 µm.

The outermost part of the inner detector is the transition radiation tracker. It consists of straws with a diameter of 4 mm that are filled with xenon and contain a sensing wire in the center. Their length is up to 1500 mm. The working principle is that the permittivity changes between two media cause transition radiation on high-energetic particles passing the interface. The sensing wire is biased against the straw to drag generated charge out of the gas. The space between the straws is filled with polymer fibers in the barrel region and polymer sheets in the end-cap region to create the transition radiation. Lighter particles create more transition radiation than heavier particles at the same energy. Consequently, the signals received in the straws are larger for light particles as electrons.

2.2.3 The Inner Tracker Upgrade

For HL-LHC, the complete inner detector of the ATLAS experiment will be replaced in long shutdown three from 2025 to 2027. This is because the current inner tracker will reach its end of life and the current design is not capable of coping with the data rates expected after the upgrade of the accelerator. The replacement is the all-silicon inner tracker (ITk).

The ITk structure consists of five pixel layers in the barrel region and of five layers of inclined or vertical rings in the end-cap region. The strip sub-system consists of four barrel layers and six disks in the end-caps. The arrangement of the detectors on the end-caps is optimised to have at least nine precision measurements per track. The projected structure of the inner tracker is shown in figure 2.6.

The inclined pixel modules in third to fifth layer are to reduce the incident angle and reduce the cluster size on these sensors and hence reduce the number of signals generated per particle passing these modules.

The baseline technology foreseen for the pixel layers are hybrid pixel detectors. Monolithic HV-CMOS detectors have been proposed as alternative for the outermost pixel layer. At a

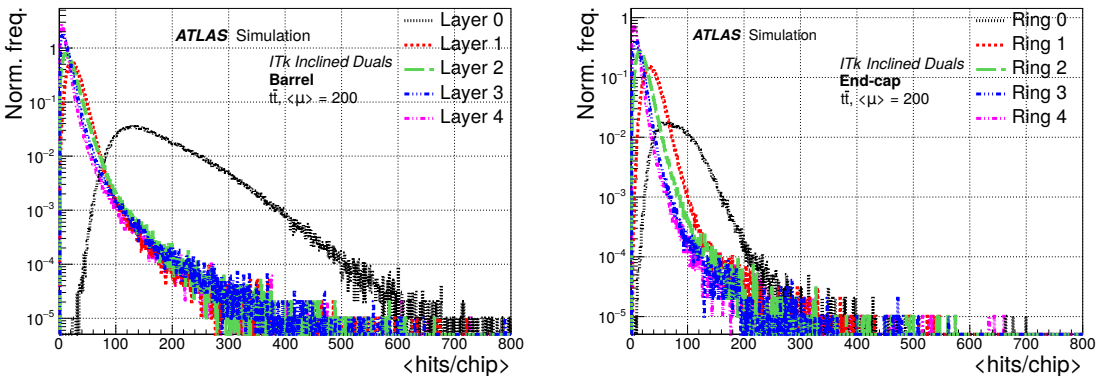


Figure 2.7: For a pixel size of 50×50 µm^2, the signal rate distributions for the chips on the inclined modules have been simulated at the expected average pile-up of 200.

radius of 291 mm from the beam axis, this layer consists of over 4000 detector chips in the flat barrel region alone. There, the requirements on radiation tolerance are moderate, so that they can be endured by monolithic HV-CMOS detectors.

The important requirements for developing sensor ASICs for this application are the following [ATL17, ATL19]:

- **Power consumption** is limited by the capacity of the cooling system which is 0.7 W/cm^2 of sensor area. This includes 0.5 W/cm^2 for the readout chip and 0.1 W/cm^2 for the sensor.
- The **pixel size** is required to be 50×50 µm^2 or 25×100 µm^2.
- A high **detection efficiency** of 99 % needs to be maintained until the end of the lifetime of the detector with the noise rate per pixel not exceeding 40 Hz.
- The detected signals have to be assigned to a single bunch-crossing. The bunch crossing frequency of 40 MHz therefore sets the **time resolution** requirement of $25\ \text{ns}/\sqrt{12} = 7.2$ ns.
- Furthermore, the **response time** is required to be moderate, which is specified to $\mathcal{O}(5\,\mu\text{s})$.
- For the **trigger scheme**, several options are listed in the technical design report: One option is to receive a trigger with an average rate of 4 MHz at a delay of 10 µs (referred to as level-0-only). The other option is the usage of the level-1 trigger which is specified to 800 kHz at a maximum delay of 35 µs. A third option is an average trigger rate of 1 MHz with a latency of 25 µs. In any case, the data has to be stored on the chip until trigger arrival and is to be transmitted off the chip via a single gigabit link.
- The **radiation tolerance** required for the detector chips is defined for the fifth pixel layer as a lifetime dose of 1.6 MGy to 3.5 MGy and a fluence of $2.8 \cdot 10^{15}$ n_{eq}/cm^2 to $3.8 \cdot 10^{15}$ n_{eq}/cm^2 depending on the position. The lowest value is expected for the barrel, the largest for the end-cap. But as these numbers are common for the three outer pixel layers, the values for the fifth layer can be expected to be below them.

Another value important for the design of a monolithic detector or readout ASIC, but not a direct requirement is the expected number of particles passing the detector area. For the expected average pile-up of 200, the signal rates for the chips on the inclined modules are shown in figure 2.7: For layer 4 – which is the fifth layer – the outcome is 12.2 signals per bunch crossing.

2.3 Treatment Monitoring in Ion Irradiation Therapy

The sensor technology can be used also for a totally different application: cancer therapy. In their last report on cancer in Germany, the Robert-Koch-Institute published data showing that in Germany, almost 500 000 new cases of cancer are found per year [Kru20]. Due to demographic change, this number is expected to increase in the future.

For treatment of cancer, several treatment principles exist: Resection, where the tumor is surgically removed, chemotherapy, where substances are used to inhibit the metabolism of the tumor, and irradiation, which uses ionising radiation to damage and destroy the tumor's DNA.

All principles have advantages and disadvantages. Resection can only be applied, if the tumor is accessible and suited for removal. This means that the affected tissue is localised and segregated from the healthy tissue.

Chemotherapy has the disadvantage that the substances do not only affect the tumor, but the whole patient, so that side effects can be more severe than the impairment from the tumor itself even up to being lethal.

Radiation destroys the tissue where the energy is deposited. This is suited for localised tumors accessible for the radiation. The challenge is to ensure that only the affected tissue is irradiated as the healthy tissue will be destroyed as well if irradiated. Therefore, precise control and monitoring is crucial for successful treatment.

For this monitoring task in irradiation treatment, HV-CMOS detectors can be used to improve the quality of monitoring data over – for example – wire-chamber systems. The higher spatial and time resolution can lead to improvements in patient therapy.

In the following, the working principle of the irradiation therapy with the differences between the particles used and the monitoring applications for which HV-CMOS detectors can be used are described: Direct monitoring of the ingoing beam and tracking of secondary particles coming from the interaction of the beam to reconstruct the interaction point.

2.3.1 Irradiation Therapy

For irradiation of cancerous tissue, in principle any particle interacting with it can be used. This can be photons, electrons, protons or even heavier ions and neutrons.

Depending on the particle type used, the profile of the energy deposition in the material differs: For photons, the energy deposition decreases exponentially with the penetration depth. For heavier particles, the penetration depth can be estimated as integration of the Bethe formula (section 3.2.2). As the cross-section increases for smaller particle energies, the energy deposition results in a Bragg-curve with the largest energy deposition directly before the particle is stopped completely. The comparison of the energy deposition profiles for photons (X-rays), protons and carbon ions is shown in figure 2.8.

The penetration depth of the heavier particles is given by the initial energy, allowing for shifting the Bragg-peak to deposit most energy inside the volume of the tumor. Typically, the energy distribution in the particle beam is sharp so that it has to be widened to irradiated the whole tumor. The energy deposited in the tissue before the tumor is smaller for heavier particles. Heavier particles are more suited for targets deeper inside the body, behind or within sensitive organs [Mei03]. The downside of irradiation with heavier particles is that large accelerator complexes are necessary. But over the last decade, more than 20 clinical facilities have come to operation in Europe [GDG+20].

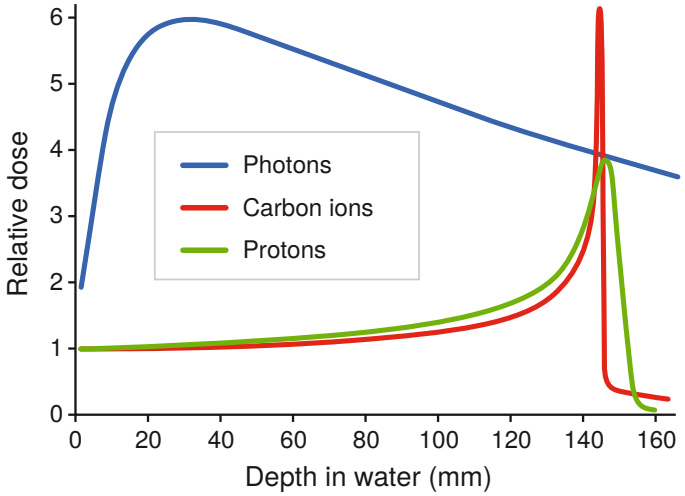


Figure 2.8: Photons show an exponential decrease of the energy deposited with penetration depth. Protons and carbon ions deposit most energy right before the complete stopping, forming the Bragg-curve. [GDG+20]

2.3.2 Direct Beam Monitoring

To avoid damaging healthy tissue, the tumor has to be mapped prior to the irradiation and the movement of the body part to be irradiated have to be prevented or tracked during the irradiation. On the other hand, it has to be made sure that the beam is directed to the correct position, with the desired energy and shape.

The energy of the particles is controlled by the accelerator, adjustment of the shape and position are controlled locally for the irradiation stations. Safe operation of these require additional instrumentation.

At HIT [OW08], this beam monitoring is currently performed by the Beam Application and Monitoring System (BAMS) which employs multi-wire proportional chambers for position monitoring and ionisation chambers for dose measurement [Pet20]. The wire chambers provide only one-dimensional information, so that at least two are necessary to obtain profile information in two dimensions. Furthermore, the resolution is limited by the distance of the wires in the chamber and the inhomogeneity introduced by the wires is not favourable for the treatment, either. Measurements with magnetic fields nearby – as an irradiation with simultaneous monitoring with magnetic resonance tomography scans would require – are an additional complication as the position measurement is affected due to the influence of the magnetic field on the movement of the electrons in the wire chambers.

Solid state detectors, as HV-CMOS detectors enable higher resolution with finer and two-dimensional segmentation as well as reduced affection from magnetic fields [ABB+19a]. The drawbacks are that a detector system to be developed stays below the material budget of less than 2 mm of water equivalent. Furthermore, the radiation hardness for five to ten years of operation has to be ensured without direct cooling [Deb17].

Just as the current system, a new system needs to incorporate redundancies to enable fail-safe operation. But with an HV-CMOS detector being able to replace two wire chambers and an ionisation chamber (via leakage current measurement as presented in section 6.6.3 and particle counting), a set of two HV-CMOS detectors can be sufficient.

The advantage of monolithic HV-CMOS sensors over hybrid pixel sensors is in this case the reduced material budget that reduces scattering on the measurement device. The impact of scattering on the monitor is severe for this application because of the large distance (in

the order of half a metre) between the beam monitor and the actual target in the patient's tissue. The details of this difference will be discussed in section 3.5.

2.3.3 Secondary Particle Tracking

Showers of secondary particles are generated when irradiating material with heavy ions. The origins of these particles hint towards the interaction point in the irradiated material. With these particles, the irradiation process can be monitored to validate the intended with the actual interaction position.

To measure these secondary particles and obtain precise information, it is necessary to place the detectors around the irradiated volume. By placing several layers of detectors, tracks of the secondary particles can be reconstructed to trace back to the interaction point. Efficient reconstruction requires high time resolution to disentangle the signals belonging to different particles.

Important for the interpretation of this track data is that the angle distribution of the secondary particles changes with the energy. As a result, an overlay of all tracks recorded will not show a maximum at the position of the Bragg peak but fade out before it as with the large energy transfers at the Bragg peak the angle distribution gets narrow and the dominant part in the position heat map will be the beam path before the peak [GHJ+13].

3 Silicon Detector Technology

In this chapter, the monolithic HV-CMOS sensor technology is composed from the underlying fundamentals of solid state materials, interaction of high-energetic particles to detect over technologies in use for this application and possible advances with the concept of monolithic HV-CMOS sensors.

3.1 Fundamentals of Semiconductors

To prepare the description of silicon detectors, the underlying theoretical description is described as quick pass-through. Starting from the description of the lattice and the concept of electronic states in it, the basic elements for the sensors with readout electronics are built up: From doping over the combination to pn junctions, the structure of the metal-oxide-semiconductor field-effect transistor (MOSFET) is described. At the end, also the production of silicon based micro-electronics and silicon sensors is sketched.

3.1.1 Theory of Semiconductors

In an ideal, infinite lattice, the electronic states can be determined from the time-independent Schrödinger equation:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r})\right)\Psi(\vec{r}) = E\,\Psi(\vec{r}) \tag{3.1}$$

with the periodic potential $V(\vec{r})$ reproducing the periodicity of the crystal lattice. Consequently, it satisfies the condition

$$V(\vec{r}) = V(\vec{r}+\vec{R}) \tag{3.2}$$

for $\vec{R}$ being integer multiples of the lattice base vectors. Bloch waves of the form

$$\Psi(\vec{r}) = e^{i\vec{k}\vec{r}}u_{\vec{k}}(\vec{r}) \tag{3.3}$$

solve this equation for $u_{\vec{k}}(\vec{r}) = u_{\vec{k}}(\vec{r}+\vec{R})$ reproducing the periodicity of the lattice. The parameter $\vec{k}$ is called reciprocal lattice momentum.

With finite crystals, the electronic states have discrete split-up energies. However, for the numbers of states in the order of 10^{23}, the states can be treated as quasi-continuum. These ranges of states are referred to as bands. In terms of conductivity, two bands are important:

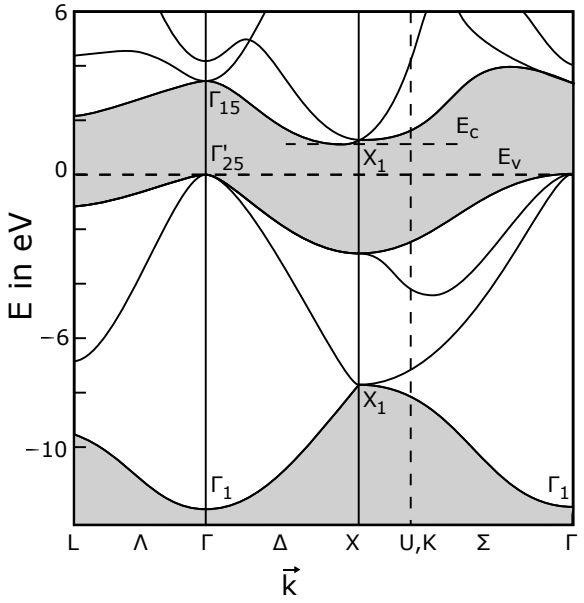


Figure 3.1: The band structure of silicon is shown over the reciprocal lattice momentum $\vec{k}$. On the X axis, special symmetry points are noted. The band gap is shown in grey around the energy of 0 eV. [Ebe13], after [CC74]

The lowest not fully occupied band in ground state – called conduction band – and the band below it – called valence band. At a temperature of $T = 0\,\mathrm{K}$, the Fermi level marks the energy below which all states are occupied and above which no states are occupied.

To conduct a current, it is necessary for electrons to occupy different states than in equilibrium. This is only possible, if there are unoccupied states available. A completely filled valence band can not conduct current because no states are available to create the movement and an empty conduction band can not conduct current as no charge carriers are available. Exciting charge carriers from the valence band to the conduction band, charge carriers are introduced in the conduction band and empty states are created in the valence band so that a current can be conducted there, too.

The arrangement of these bands to each other define the class a material belongs to: If the valence band and the conduction band overlap or the Fermi level is inside the conduction band, it is always possible for electrons to change to another state and current flow is always possible. This is the case for metals.

If the conduction band and valence band do not overlap, the energy difference between the highest energy state in valence band E_V and the lowest energy state in conduction band E_C

$$E_G = E_C - E_V \tag{3.4}$$

is referred to as band gap E_G. Depending on the size of the band gap, materials are divided into insulators and semiconductors. The value of this border varies in literature, but is typically in the range of $3 - 4\,\mathrm{eV}$.

Silicon has a diamond lattice structure. This is a face-centered cubic lattice with a unit cell containing two atoms at $(0|0|0)$ and $a_0/4(1|1|1)$ where a_0 is the lattice constant. The band structure of it is shown in figure 3.1. From the structure, two points can be read: Firstly, silicon is a semiconductor with a band gap of 1.12 eV. Consequently, excitation is necessary for silicon to conduct a current. Going to low temperatures, the conductivity will decrease and at $T = 0\,\mathrm{K}$, silicon will behave as an insulator. Secondly, that the E_V and E_C are not at the same point in $\vec{k}$-space. This makes silicon an indirect semiconductor.

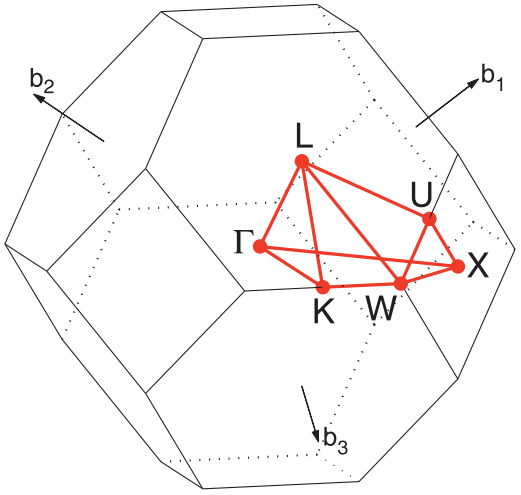


Figure 3.2: The Brillouin zone gets its shape from the lattice symmetry and hence is of the same shape for all face centered cubic lattice materials including silicon. The symmetry points marked in the band structure in figure 3.1 are shown in red. [SC10]

This means that either a larger amount of energy is necessary or momentum has to be transferred in addition to the band gap energy by creating or absorbing a phonon.

The labels on the abscissa of the band structure of silicon mark symmetry points in the Brillouin zone which are sketched in figure 3.2. From the band structure of silicon as shown in figure 3.1 it becomes obvious, that the material properties of silicon will depend on the orientation of the crystal lattice. These orientation is denoted by the help of Miller indices (see [Sau09]).

The conductivity of semiconductors as silicon depends on temperature. This dependency is described by the probability distribution for occupation of electronic states as described by the Fermi-Dirac distribution

$$f(E) = \frac{1}{\exp\left(\frac{E-E_F}{k_B T}\right) + 1} \,. \tag{3.5}$$

This distribution approaches a step function for the temperature T approaching 0 K. For higher temperatures, the step function gets smeared out as shown in figure 3.3. To actually obtain the number of charge carriers in a band, this function has to be multiplied with the state density and integrated over the corresponding energy interval.

Most semiconductor circuits are operated around room temperature. Specifications of operating temperatures of commercially available devices put ranges around this, for example from −40 °C to 125 °C [Nex17]. For semiconductor detectors, this range is typically −20 °C to 50 °C and plans aim for cooling temperatures of −30 °C at LHC [TDP+16]. At 300 K, silicon has an intrinsic charge carrier density of $\sim 5 \cdot 10^9 \mathrm{cm}^{-3}$ [Sau09] which is low compared to metals where this number is in the order of 10^{23}.

While in the conduction band, few charge carriers transport the current, in the valence band, many charge carriers transport the current with few unoccupied states. To simplify the description of this, the unoccupied states are treated as quasi-particles with opposite-sign mass and charge but equivalent absolute values as the electrons in that state. In this picture, exciting an electron from the valence band to the conduction band creates an electron-hole pair.

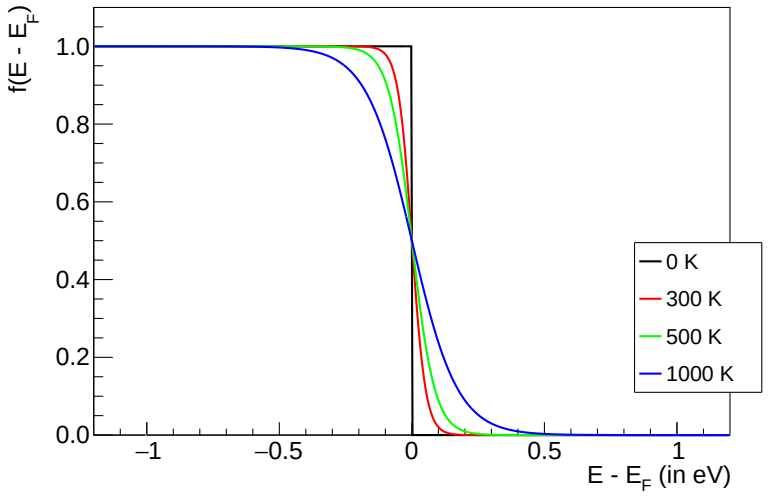


Figure 3.3: The Fermi distribution shows a step at the Fermi level for $T = 0\,\mathrm{K}$ which is smeared out for higher temperatures.

3.1.2 Doped Semiconductors

To improve the poor conductivity of semiconductors at room temperature, they are doped with other elements. These doping concentrations are orders of magnitude larger than the intrinsic charge carrier densities at room temperature in silicon. The doping concentrations are in the order of $10^{16}\mathrm{cm}^{-3}$ with a range of several orders of magnitude around this number.

For silicon with its four valence electrons in the tetrahedral bond structure, typical doping atoms are from the main groups III or V of the periodic table of elements. Elements like boron from group III have one valence electron less than silicon and hence introduce a vacancy in a covalent bond to a neighbouring atom. They are referred to as acceptors. Group V elements, like phosphor, introduce an extra valence electron that is weakly bound to the atom. These atoms are referred to as donors.

The missing electrons from group III elements introduce additional states close to the valence band which are not occupied at $T = 0\,\mathrm{K}$. At higher temperatures, electrons from valence band can be excited to those states leaving an unoccupied state in valence band behind. The equivalent is the excitation of a hole from the acceptor state into the valence band.

The additional electron from group V elements are occupied states close to conduction band. From there, the electrons can be thermally excited to the conduction band to increase the charge carrier density there.

These new states change the position of the Fermi level in the material: At $T = 0\,\mathrm{K}$, the occupation state of these additional states is opposite to the band next to them. Hence, the Fermi level will be located in the middle between the dopant states and the band.

A part of the charge carriers introduced by the doping atoms will recombine with the charge carriers of the other type so that the law of mass action

$$np = N_C N_V e^{-E_G/k_B T} \tag{3.6}$$

holds. There, n and p are the charge carrier densities in valence and conduction band, N_C and N_V the dopant concentrations in conduction and valence band and E_G the band gap.

As for the type of the majority of charge carriers, the semiconductor doped with donors and majority charge carrier electron is referred to as n-doped (or n-type) and the semiconductor doped with acceptors is called p-doped (or p-type) because of the positive charge of the holes being the majority charge carriers.

3.1.3 PN Junction

The interface between a p-doped semiconductor and an n-doped semiconductor is referred to as pn junction. The resulting device is a diode. It is a basic element in many important semiconductor elements as for example the transistor.

The pn junction is easiest to understand when following the hypothetical joining of the p-doped and n-doped parts: Both parts have different majority charge carriers and Fermi levels at different positions in the band gap. Bringing these materials in contact, the electrons from the n-doped area move to the p-doped area and the holes from the p-doped area move to the n-doped area because of the concentration gradients of the charge carriers. There, the charge carriers recombine with the charge carriers of the other type removing a charge carrier of each type per recombination. However, the occupied acceptor states and non-occupied donor states result in a space charge which counteracts the charge carrier gradient. They are fixed as from their origin at the dopant atoms. As no charge carriers are left in this volume, it is also referred to as depletion zone.

The two contributions can be described with Fick's law of diffusion

$$\vec{j}_{\text{diff}} = -qD\nabla n \tag{3.7}$$

and the Poisson equation

$$\frac{\partial^2 \Phi}{\partial x^2} = -\frac{1}{\varepsilon_0 \varepsilon_{\text{r}}} \rho(x) \ . \tag{3.8}$$

There, $\vec{j}_{\text{diff}}$ denotes the diffusion current, q the charge of a single charge carrier, D the diffusion constant, n the charge carrier density, ε_0 the vacuum permittivity and the relative permittivity ε_{r} which would be $\varepsilon_{\text{r}} = \varepsilon_{\text{Si}} = 11.68$ for silicon.

$$\rho(x) = q\left(p(x) - n(x) + N_D(x) - N_A(x)\right) \tag{3.9}$$

is the charge density consisting of the contributions of the charge carrier densities for holes and electrons (p and n) and of the dopant atom densities for donors and acceptors ($N_D(x)$ and $N_A(x)$).

The x direction is chosen perpendicular to the interface between the differently doped volumes with the origin on the interface, the p-doped area for $x < 0$ and the n-doped area for $x > 0$. Visualisations for this and the following steps are given on the left side of figure 3.4. For simplicity, it is assumed that the charge density $\rho(x)$ is constant close to the interface of the doping regions and zero elsewhere:

$$\rho(x) = \begin{cases} 0, & x \leq -x_p \\ -qN_A, & -x_p < x \leq 0 \\ +qN_D, & 0 < x \leq x_n \\ 0, & x_n < x \end{cases} \tag{3.10}$$

There, x_p and x_n are the extents of the space charge region into the p-doped and n-doped side, respectively. To satisfy charge conservation, the amounts of charge in the two regions have to be the same. This means that for different dopant densities N_A and N_D, the extensions are different. For the extent into the p-doped region x_p and the extent into the

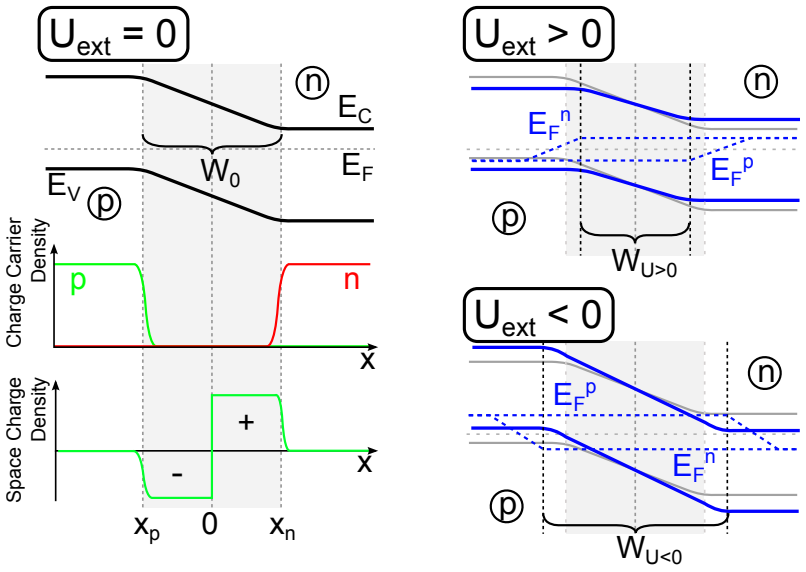


Figure 3.4: On the left side, the band diagram of the pn junction without external voltage U_{ext} is shown with the charge carrier densities and space charge. The depletion zone is shown shaded in grey. For external voltages, the width change of the depletion zone is indicated with the split-up quasi Fermi levels for electrons ${E_F}^n$ and holes ${E_F}^p$. [Sch14]

n-doped region x_n this is shown by the equation

$$0 \stackrel{!}{=} \int_{-\infty}^{\infty} \rho(x)dx \qquad \Rightarrow \qquad N_A x_p = N_D x_n \ . \tag{3.11}$$

Integration of equation (3.8) results in an equation for the electrical field:

$$E(x) = \begin{cases} \frac{-qN_A}{\varepsilon_0\varepsilon_\text{r}}(x + x_p), & -x_p < x \leq 0 \\ \frac{qN_D}{\varepsilon_0\varepsilon_\text{r}}(x - x_n), & 0 < x \leq x_n \\ 0, & x \leq -x_p \quad \text{or} \quad x_n < x \end{cases} \tag{3.12}$$

Integrating the field in equation (3.12) with the boundary conditions of $\Phi(x < -x_p) = 0$ and continuity at $V(0)$ results in

$$\Psi(x) = \begin{cases} 0, & x < -x_p \\ -\frac{qN_D}{\varepsilon_0\varepsilon_\text{r}}\left(\frac{x^2}{2} - x_n x\right) + \frac{eN_A}{2\varepsilon_0\varepsilon_\text{Si}}{x_p}^2, & 0 < x \leq x_n \\ U_\text{bi}, & x > x_n \end{cases} \qquad \frac{qN_A}{2\varepsilon_0\varepsilon_\text{r}}(x + x_p)^2, \quad -x_p \leq x \leq 0 \tag{3.13}$$

with the built-in voltage $U_\text{bi} = \lim_{x \to {x_n}^-} \Psi(x)$

$$U_\text{bi} = V(x_n) = \frac{q}{2\varepsilon_0\varepsilon_\text{r}}\left(N_D {x_n}^2 + N_A {x_p}^2\right) \ . \tag{3.14}$$

With the neutrality condition from equation (3.11) and the built-in voltage U_bi, the width w of the depletion zone evaluates to

$$w = (x_p + x_n) = \sqrt{\frac{2\varepsilon_0\varepsilon_\text{r}}{q}\frac{N_D + N_A}{N_D N_A}U_\text{bi}} \ . \tag{3.15}$$

The built-in voltage U_{bi} is given from the position of the Fermi levels in the two base materials which is located in the respective middle between the band edge and the dopant states.

For an external voltage U_{ext} applied to the pn junction, the width of the depletion zone changes. A forward bias with the positive potential connected to the p-doped side reduces the width, while a reverse bias increases the width. Assuming small voltages , i.e. $U_{\text{ext}} < U_{\text{bi}}$, the external voltage adds to the potential difference from the material and equation (3.15) is modified to

$$w = (x_p + x_n) = \sqrt{\frac{2\varepsilon_0\varepsilon_{\text{r}}}{q}\frac{N_D + N_A}{N_D N_A}(U_{\text{bi}} - U_{\text{ext}})} \; . \tag{3.16}$$

With external voltage, the Fermi levels of holes and electrons split up in the depletion region as there are no carriers to recombine with and compensate for the potential difference. The length scale of the change of the potentials is defined by the diffusion and recombination parameters of the excess minority carriers. The changes are visualised on the right of figure 3.4.

Reverse bias and breakdown

For the application as sensing element in particle physics, the diode is used reversely biased. For this, the simplification to neglect generation and recombination in the depletion zone can not be kept.

For generation of electron-hole pairs in the depletion zone, the electric field from the ionised dopant atoms will separate the charge carriers and prevent recombination of them. Instead they will reach the neutral volume where they contribute to the leakage current.

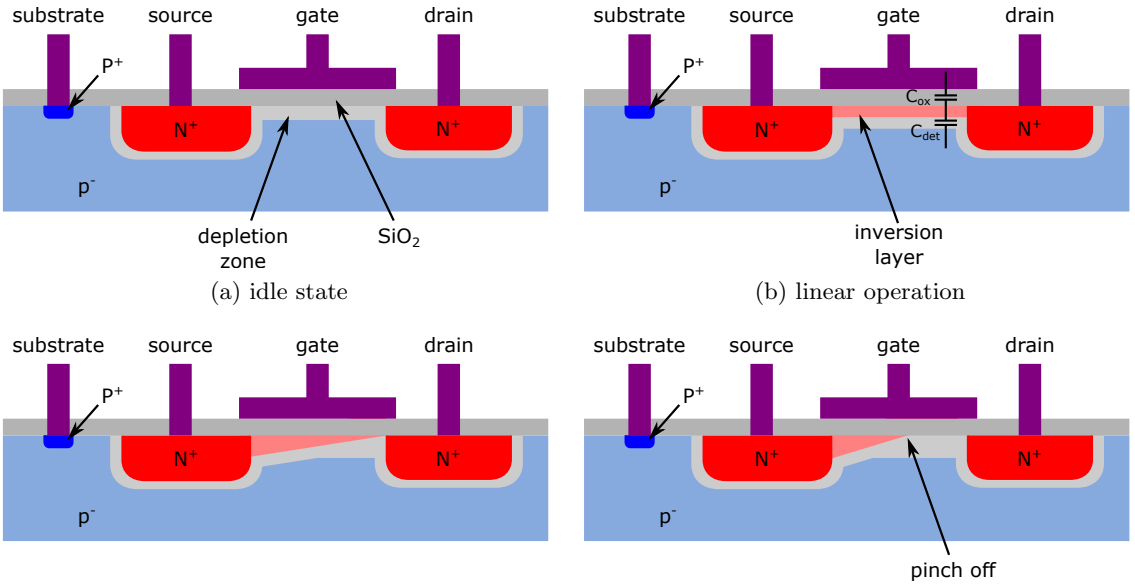
At this point three mechanisms can destroy the stationary state:

- thermal runaway
- tunneling current
- avalanche breakdown

Thermal runaway can happen because of the power dissipation at large bias currents which heats up the semiconductor and by that increasing the generation current resulting in a self-enforcing loop.

Tunneling current can occur for strong band bending where an electron from valence band can tunnel into a state in the conduction band. This charge carrier is then moved away from the junction and the depletion zone. As for the larger energy states occupied at higher temperatures, the tunneling becomes more probable and the breakdown voltage decreases for larger temperatures.

For an avalanche breakdown, a charge carrier in the conduction band is accelerated by the electric field in the depletion zone and gaining sufficient energy to generate new electron-hole pairs from impact ionisation. If the gain of this process is larger than the absorption, this will lead to an increasing number of charge carriers generated in the depletion zone forming an avalanche. With higher temperature, inelastic scattering increases for charge carriers moving in the material, consequently the energy losses increase with temperature. For this reason, the breakdown voltage for an avalanche breakdown increases with temperature compensating the increased energy losses.



(a) idle state

(b) linear operation

(c) transition to saturation

(d) saturation region

Figure 3.5: For an n-channel MOSFET, the different operation regimes are shown: without voltages connected (a), with a substrate-gate voltage U_{gs} creating an inversion layer below the gate (b), at the limit to channel saturation (c) and in saturation (d). [Ehr21], after [Thu18]

With the breakdown voltage U_{br} and the band gap energy E_G, the tunneling effect breakdown is dominant in the range of $U_{\mathrm{br}} \leq 4E_G/e$ while the avalanche breakdown is dominant for the range $U_{\mathrm{br}} \geq 6E_G/e$ [Sau09].

3.1.4 Metal-Oxide-Semiconductor Field-Effect Transistor

The transistor is an important element in electronic circuits. In HV-CMOS technologies, these are mostly realised as metal-oxide-semiconductor field-effect transistor (MOSFET).

The basic working principle is to control conductivity with another voltage. For this, the MOSFET has four connections: source, drain, gate and substrate. The control voltage is applied between substrate and gate to change the conductivity between source and drain. This is achieved with an arrangement of differently doped areas: For a p-doped substrate, source and drain are implemented as n-doped areas inside it. The area between them is covered with a thin layer of SiO_2 as insulator topped by a conductor made out of metal or poly-silicon forming the gate. From the conductive channel to be formed between source and drain carrying minority charge carriers for the substrate, this arrangement is called an n-channel MOSFET (or NMOS transistor). Reversing the dopant types (n-doped substrate and p-doped source and drain), the channel will conduct via holes resulting in a p-channel MOSFET (also referred to as PMOS transistor). The structure of an NMOS traansistor without applied voltages is shown in figure 3.5a. Packaged MOSFETs typically connect the source and substrate contacts to one resulting in a three-terminal device. If this connection of source and substrate is not present, the contacts of source and drain are equivalent.

As described in the previous section, the interface between the n-doped inplants for source and drain and the p-doped substrate form diodes. They are oriented in opposite directions so that by default no current flow is possible between source and drain.

Applying a voltage U_{gs} between the gate and substrate, a conductive channel can be created below the gate accumulating electrons for a voltage larger than the threshold voltage

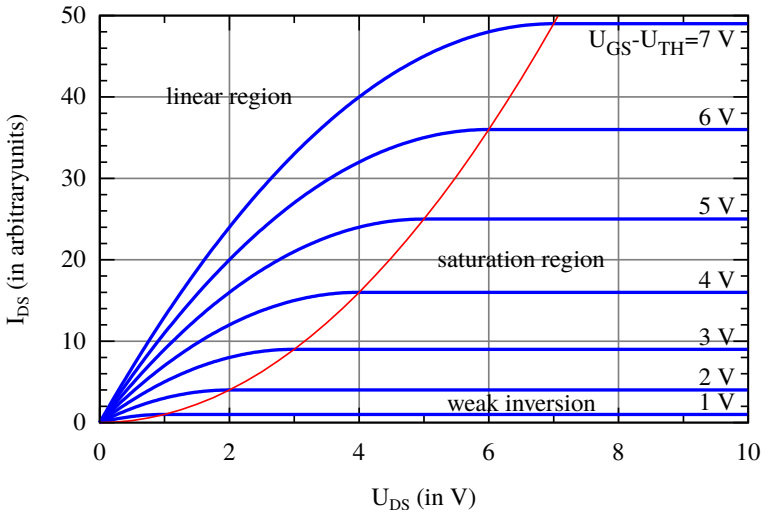


Figure 3.6: For an n-channel MOSFET, the characteristics of I_{DS} over U_{DS} with $U_{\mathrm{gs}} - U_{\mathrm{th}}$ as parameter show the two main regions separated by a red line: the linear region with linear relation of U_{DS} and I_{DS}, and the saturation region with almost no change of I_{DS} for changes of U_{DS}. The line denotes the pinch-off of the channel. After [Wik21]

$U_{\mathrm{th}} < U_{\mathrm{gs}}$. The threshold voltage U_{th} is defined as the voltage at which the number of free electrons reaches the number of holes in intrinsic silicon. As holes are the majority charge carriers in p-doped silicon, this electron excess makes it an inversion layer. Via the inversion layer, source and drain get electrically connected.

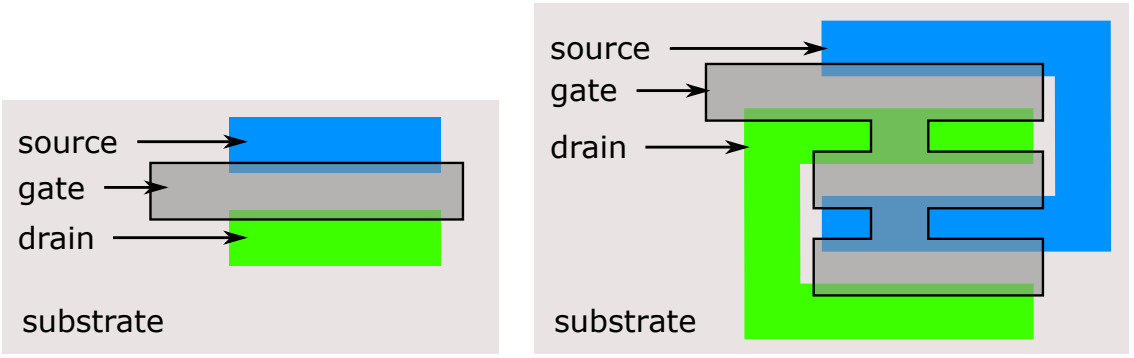
If the voltage between source an drain U_{DS} satisfies $U_{\mathrm{DS}} < U_{\mathrm{gs}} - U_{\mathrm{th}}$, the channel provides an ohmic connection between source and drain. Hence, $I_{\mathrm{DS}} \propto U_{\mathrm{DS}}$.

The voltage between source and drain changes the potential inside the channel. As a result, the electrons in the channel are dragged towards the source. At $U_{\mathrm{DS}} = U_{\mathrm{gs}} - U_{\mathrm{th}}$, the channel is pinched off at the drain implant (figure 3.5c). Further increasing U_{DS}, the pinch-off point moves towards the source implant. This shortens the channel and compensates for the larger voltage applied. As a result, the current will increase only slowly. For this reason, this operation region is called saturation region (figure 3.5d).

Collecting these operation states into a diagram of U_{DS} over the current between drain and source I_{DS} with U_{gs} as parameter, figure 3.6 is received. The pinch-off point of the channel is indicated by the red line, marking the beginning of the saturation region.

Summing up the behaviour of the MOSFET, it behaves as a voltage-controlled resistor in the linear region and as a voltage-controlled current source in the saturation region.

The most simple shape of a transistor is linear with a cross-section as shown in figure 3.5. A top view of such a MOSFET is shown in figure 3.7a. Its geometry defines the resistance of the transistor. A wider transistor conducts more current at the same current density. This is the horizontal direction in the figure. On the other hand, a longer transistor (vertical direction in the sketch) results in a longer channel and consequently a larger resistance. Transistors suited for large currents consequently can become very long and less efficient as the end of the channel will see a lower voltage from the resistance in the source and drain implants and connection material. To overcome this, such transistors are broken up and folded into comb structures (figure 3.7b). This way, space need is reduced as the source and drain contacts in the middle are used for both directions. Each triplet of a source



(a) simple geometry (b) structure with several fingers

Figure 3.7: The default structure of a MOSFET is a linear one with three parallel rectangles (a). For larger currents, the structure can be broken up and folded to obtain shorter structures (b). This example is split into three parts, called fingers. (a) is after [Sch16]

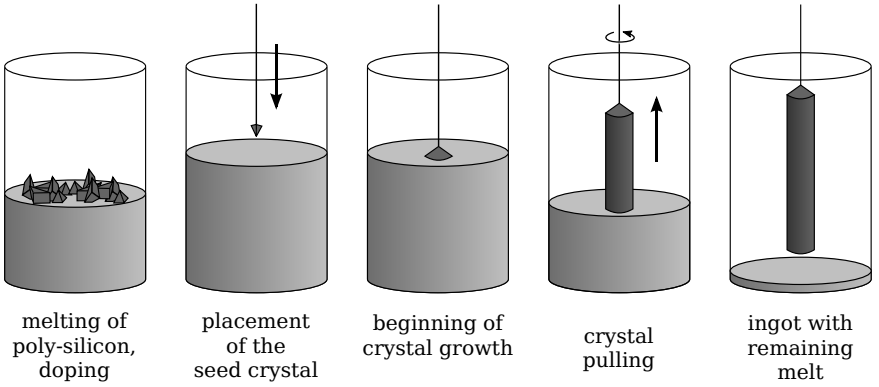


Figure 3.8: The poly-crystal silicon is molten and doped. Then the seed crystal is brought in touch with the molten material and slowly retracted to enable crystal growth at the interface between the crystal and the melt. A slow rotation ensures homogeneous growth. After [Wik08]

implant, drain implant and gate is referred to as a finger: The transistor in figure 3.7a has one finger, the more complex transistor in figure 3.7b has three fingers.

For radiation hard design of MOSFETs, also a different geometry exists which is described in section 3.3.2.

3.1.5 Production of Semiconductor Circuits

To make use of silicon as a semiconductor, it is necessary for the material to be a mono-crystal as the interfaces between the crystalline parts in poly-crystalline silicon will alter the properties. Solidification of silicon produced from silicon-dioxide tends to solidify in a poly-crystalline structure. To obtain the mono-crystal material necessary for wafer production, conversion and further purification is necessary.

One option to achieve a mono-crystal of silicon is the Czochralski process. For it, poly-cristal silicon is molten and a small seed crystal is brought into contact with the liquid material. Then, it is slowly retracted upwards while slowly rotating around the vertical axis. The crystal grows on the contact patch between the crystal and the molten silicon. The rotation around the vertical axis helps forming the so-called boule to a round shape. The process is visualised in figure 3.8.

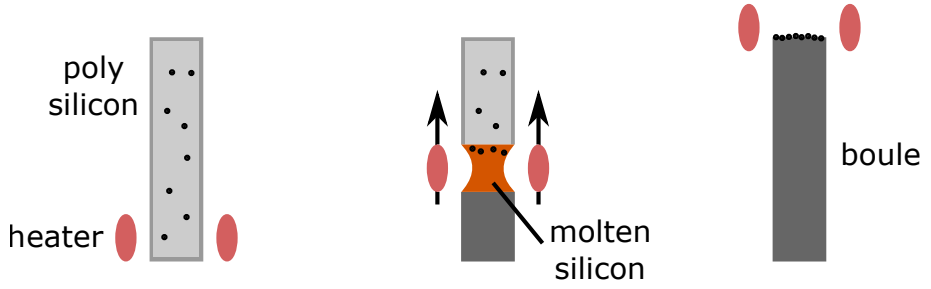


Figure 3.9: A poly-crystal silicon boule is molten in a thin layer. In this layer, impurities move to the top. By moving the heater along the boule, the material at the bottom of the molten area solidifies in an ordered state. The impurities stay in the molten part, hence collecting at the end of the finished boule. [Ehr21]

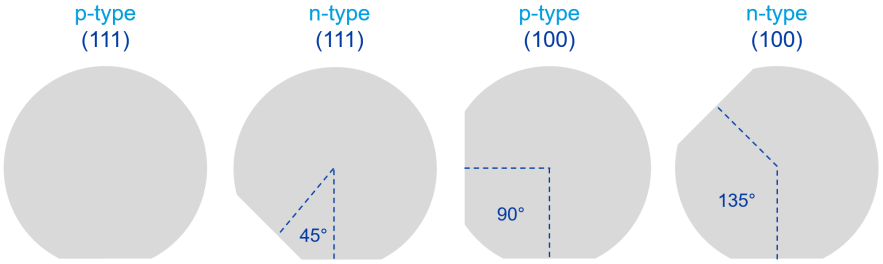


Figure 3.10: Wafers are marked with cuts on the edges for the type they are: a p-doped wafer in {111} orientation has only one large cut, while n-doped wafers and p-doped {100} type wafers have an additional smaller cut. Sometimes, n-doped {100} wafers have the second cut at 180° instead of 135°. [Sil18]

Another option is zone melting. For this, a poly-crystalline boule is the starting point. The conversion to a mono-crystal happens by melting a shallow section of the boule. The heating element is then moved along the boule melting more of the poly-crystal material. On the side pointing away from the heater, the silicon cools down and forms a mono-crystal. Impurities contained in the poly-silicon will stay in the molten part and by this be separated from the silicon as the material will be accumulated at the end of the boule. The process is visualised in figure 3.9.

After transformation to a mono-crystal, the boules are cut into disks which are then polished resulting in wafers that can be used for production of micro-electronic devices.

As mentioned in section 3.1.1, the orientation of the lattice in the wafer affects the properties of the material and consequently of the circuits realised on it. Therefore, the orientation of the seed crystal is important for predictable results. Commonly used orientations for silicon wafers – given as Miller indices – are {100} and {111}. For production wafers, the doping type and the crystal orientation are encoded in the shape of the wafer: From the initially round shape, one large cut and a small cut flatten the edge of the wafer. The positions of the cuts are shown in figure 3.10. Typical substrate resistances for p-type wafers in {111} orientation used for micro-electronics production are around $10 - 20\ \Omega$cm. By decreasing the dopant concentration, also higher substrate resistivities can be achieved.

Production of electronic devices based on these wafers is a process consisting of many steps of adding or selectively removing passivation layers, creating implants or forming connections on top of the wafer. Alignment between the single steps is crucial as for the small structures down to single-digit nanometre processes as used for current processes [TSM21]. The lithography steps require masks which have to be precisely aligned. The

size of these masks depends on the application and panelisation is often applied to reduce the mask size and by that ease constraints on the mask material. A detailed description of the production process of the electronic circuits on the wafer can be found for example in [Thu18].

3.2 Interaction of Particles with Silicon Detectors

Detection of a particle by a detector requires interaction of it with the detector. For HEP experiments, these particles can be of different type: X-rays, electrons or heavier particles are possible. These have different interaction processes that will be discussed in the following. The focus of the described interaction processes is laid on the ones important for silicon detectors as interaction with silicon is mainly electromagnetic. For charged particles, the fluctuation of the deposited charge is discussed as well as scattering which can limit the achievable resolution in tracking.

3.2.1 Interaction with Photons

Interaction of photons with a silicon detector has three contributions relevant for detection:

- absorption,
- inelastic scattering with an outer-shell electron and
- pair-production.

Rayleigh scattering is not considered here as no energy is transferred.

If a photon of sufficient energy is absorbed by an inner-shell electron it is lifted to an unbound state leaving a vacancy in the inner shell behind. This process is the *photoelectric effect*. The vacancy will be filled by an electron from an outer shell emitting a low-energy photon. The excited electron will interact with the matter around it, generating more charge carriers resulting in the conversion of the additional energy via processes described in the next sections.

Inelastic scattering with an outer-shell electron (*Compton scattering*) will result in an excited electron and a photon with smaller energy. The wavelength of the photon changes accordingly to the energy transferred $\Delta E_\gamma = E_\gamma - E_\gamma{}'$. This energy transfer depends on the scattering angle θ and the initial energy. The energy of the deflected photon is

$$E_\gamma{}' = \frac{E_\gamma}{1 + \frac{E_\gamma}{m_e c^2}\left(1 - \cos(\theta)\right)} \tag{3.17}$$

with the electron mass m_e and the speed of light c.

The third process (*pair-production*) is only possible for photons with an energy higher than twice the rest mass of the electron $E_\gamma \geq 2m_e c^2 = 1.022$ MeV. For momentum conservation, a nearby mass is required in this process.

For photon energies $E_\gamma \lesssim 100$ keV, the photoelectric effect is the dominant process. From this point up to $E_\gamma \lesssim 20$ MeV, Compton scattering is the dominant process. Only above 20 MeV, pair-production becomes the most important process, where not only a nucleus, but also an electron can serve as massive object to satisfy momentum conservation.

The running of the cross sections for interaction with carbon is shown in figure 3.11. For silicon (atomic number 14), the behaviour is similar, but the scales shift, so that the photoelectric effect is the dominant contribution up to higher energies and pair-production becomes dominant at slightly lower energies. Because of the minimum energy to create the electron-positron pair, this limit can only approach $E_\gamma \approx 1.022$ MeV.

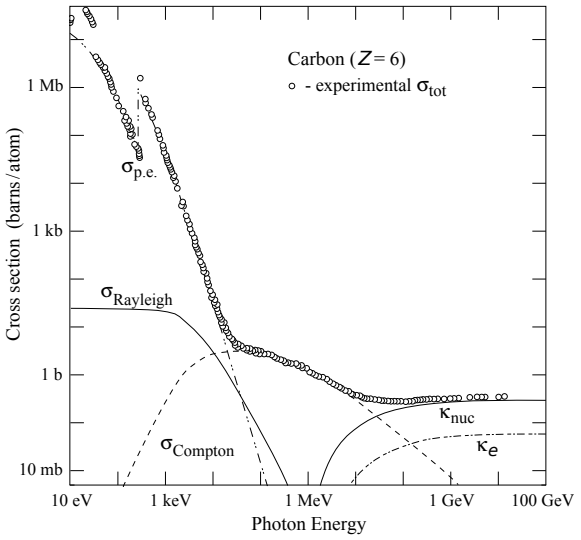


Figure 3.11: Measured on carbon, the cross sections of the different interaction processes add up to a measurable value. The photoelectric effect ($\sigma_{\text{p.e.}}$) dominates for small energies where Rayleigh and Compton scattering become more important at a photon energy around 100 keV. At photon energies around 20 MeV, pair-production becomes dominant with contributions from events close-by to a nucleus (κ_{nuc}) and an electron (κ_e). [BAB+12]

3.2.2 Interaction with Heavy Particles

As the interaction with silicon will be mainly electromagnetic, the particles looked at here have to be charged. Heavy in this case is considered a mass larger than the one of an electron. Energy is transferred to the material in collisions with little energy transfer per interaction. In thin layers of material this results in few collisions and therefore large fluctuations of the transferred energy.

The amount of energy transferred per path length in a material depends on the energy of the incoming particle and the cross section. In the range where ionisation is the dominant process, this average energy loss per length $\langle dE/dx \rangle$ is described by the Bethe equation [BAB+12]:

$$-\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = \frac{e^4 n z^2}{4\pi {\varepsilon_0}^2 m_e c^2 \beta^2} \left(\log\left(\frac{2 m_e c^2 \beta^2 \gamma^2}{I\left(1 + 2\frac{\gamma m_e}{M} + \left(\frac{m_e}{M}\right)^2\right)} \right) - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right) \tag{3.18}$$

with the speed of light c, the particle's velocity v with $\beta = v/c$ and $\gamma = (1-\beta^2)^{-1/2}$, the electron mass m_e, electron density of the material n, the vacuum permittivity ε_0, the mean excitation energy I of the material and the mass of the incident particle M. $\delta(\beta\gamma)$ is a correction for polarisation effects of the material at high incident particle energies truncating the field. For low energies, defined as $2\gamma m_e/M \ll 1$, the bracket in the denominator of the fraction in the logarithm becomes 1 and the form

$$-\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = \frac{e^4 n z^2}{4\pi {\varepsilon_0}^2 m_e c^2 \beta^2} \left(\log\left(\frac{2 m_e c^2 \beta^2 \gamma^2}{I} \right) - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right) \tag{3.19}$$

is obtained. The mean excitation energy for silicon has been found to be 173 eV [PDG21].

For muons in copper, the running of the stopping power is shown in figure 3.12. The ionisation dominated energy range is indicated by the grey bars and the label "Bethe". Towards the top end of the shown spectrum, radiative losses (bremsstrahlung) become

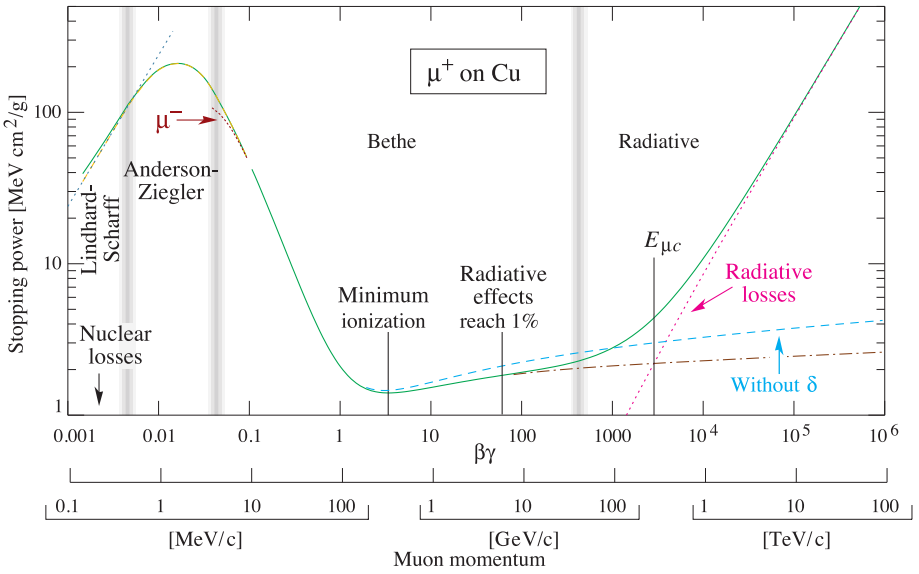


Figure 3.12: The stopping power of copper for anti-muons is shown over nine orders of magnitude in momentum. The middle range for 10 MeV $\lesssim E \lesssim$ 100 GeV where ionisation dominates the energy loss is described by the Bethe equation. [BAB+12]

dominating. For silicon, the absolute values change, but the general behaviour stays the same.

The Bethe formula shows a global minimum for the energy loss around $\beta\gamma \approx 3.0 - 3.5$ depending on the material. Particles of this energy form a worst-case scenario for detection and are referred to as minimum ionising particle (MIP).

Moving through matter, the energy of a particle is gradually absorbed by the traversed matter. As visible from the Bethe equation and corrections for small energies (see figure 3.12), the cross section increases for small energies of the incident particle. This means that for a completely stopped particle, towards the end of the path, most of the energy is deposited. This results in a Bragg curve [BK04]. It gradually increases before a steep drop-off to zero at the depth the particle gets absorbed. The path length in matter traveled before absorption depends on the energy of the particle and the stopping power of the traversed material.

3.2.3 Interaction with Light Particles

For light, charged particles as electrons, other processes contribute significantly to the interaction with matter so that the Bethe equation can not describe the interaction for the full range.

The contributing processes to the energy loss of electrons or positrons in matter are shown in figure 3.13: At energies below $\sim$ 10 MeV, ionisation is the dominant process in lead, but above this energy, bremsstrahlung becomes the dominant process and consequently, the Bethe equation can not be used to describe the interaction any more.

3.2.4 Charge Generation Fluctuation

In silicon, the energy necessary to create an electron-hole pair is 3.6 eV. The excess energy over the band gap energy of 1.12 eV goes into generation of phonons. The energy loss per length in silicon is about 390 eV/μm for a MIP. This evaluates to a mean value of 108

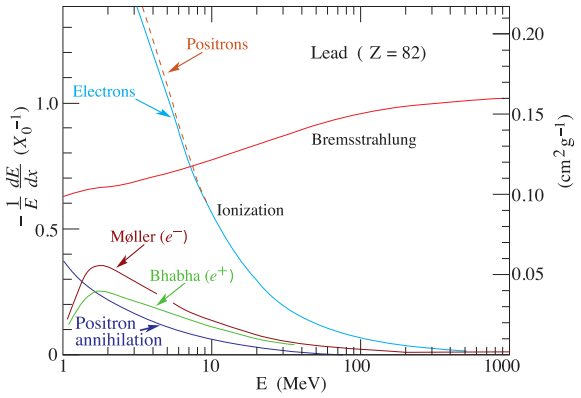


Figure 3.13: For energies below ~ 10 MeV, ionisation is the dominant interaction process for light, charged particles in lead, but above this energy, bremsstrahlung becomes more important. [BAB+12]

electron-hole pairs per micrometre on average. The most probable value, however, is 76. [Har09]

The combination of the Poisson contribution of the number of interactions of a particle per length of travel in a material and the "straggling function" describing the energy transfer in an interaction result in an asymmetric distribution first derived by Landau [Lan44] and was measured for example by Goldwasser et al in 1952 [GMH52].

In an experiment, this distribution is convoluted with uncertainty descriptions for the measurement devices, as for example amplifiers used to measure the signal introduce noise.

The tail to large energy depositions is due to δ-electrons which originate from rare events in which large amounts of energy are transferred to an electron.

3.2.5 Particle Scattering

For tracking applications, it is necessary to measure the position of a particle at several points on its trajectory to probe the momentum. The detector planes however involve interaction with matter and consequently the probability of altering the trajectory by scattering. Furthermore, the energy transferred to the tracking detectors is missing in the calorimeters. Hence, the amount of material in a tracking system is minimised.

The dominant process for scattering of the charged particles to be measured by the tracking detector in HEP is Coulomb scattering on the electric field of a nucleus. The differential cross section of this elastic process is described by Rutherford scattering:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \left(\frac{zZe^2}{4\pi\varepsilon_0\, 4E_0}\right)^2 \frac{1}{\sin^4\left(\frac{\theta}{2}\right)} \tag{3.20}$$

This equation describes the interaction of an incoming particle with charge ze and energy E_0 to a nucleus with charge Ze – where e is the elementary charge – for a scattering angle θ.

Multiple scattering will result in a distribution approaching a Gaussian distribution in central limit theorem, centered around zero. The width θ_0 of this distribution depends on the particle energy, the scattering material and the amount of it. An approximation for the standard deviation of this distribution has been developed in [Hig75, LD91] to the Highland equation:

$$\theta_0 = \frac{13.6\ \mathrm{MeV/c}}{p\,\beta} z \sqrt{\frac{x}{X_0}} \left(1 + 0.038\ \log\left(\frac{x}{X_0}\right)\right) \tag{3.21}$$

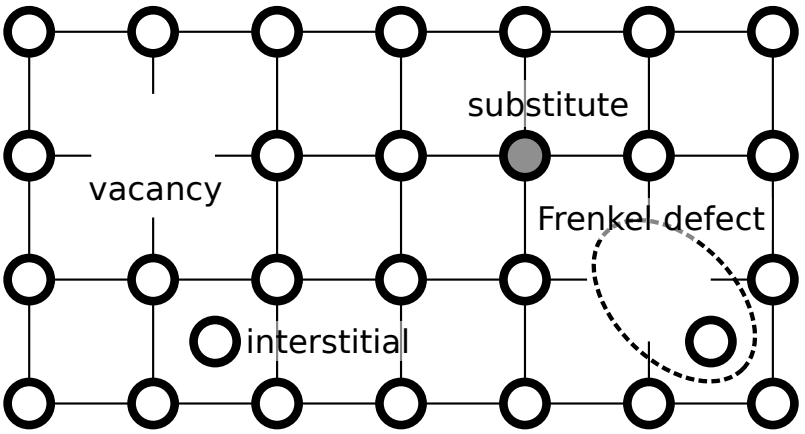


Figure 3.14: There are several basic kinds of lattice defects: Positions can be unoccupied forming vacancies in the lattice, or as interstitial in between lattice positions. A special displacement defect is the Frenkel defect where an interstitial atom is located next to a vacancy. Furthermore, atoms can be substituted by other atoms. After [KW16]

The parameters of this expression are the momentum of the scattering particle p with its charge z and velocity $\beta = v/c$ with the speed of light c and the parameters of the material it traverses, the thickness x and the radiation length X_0.

3.3 Radiation Damage in Semiconductors

With longer operation of semiconductor sensors in environments as the LHC, large numbers of high energetic charged particles traverse the sensor. The energy deposited by them damages the structure of the semiconductors.

The different parts of the sensor receive the damage from different processes. Electronics, implemented at the surface of the die, suffers from ionisation damage. The sensor diodes in the volume of the die, however, use this very effect for signal generation but decrease in performance for lattice defects introduced by the radiation. For this reason, the discussion of the damage mechanisms is split between the bulk material with the sensor diodes and the surface effects with the electronics.

3.3.1 Bulk Damage

Damage of the bulk material of a silicon sensor is mainly non-ionising energy deposition – referred to as non-ionising energy loss (NIEL). This is an effect of scattering of incident particles with the lattice structure. One effect is phonon emission which does not alter the lattice. Others are atom displacement or other lattice altering processes.

Ionising interaction of particles with silicon do not result in permanent damage as ionisation is a reversible process in a conductive material. Consequently, ionisation can only permanently alter the material properties at isolation layers and interfaces.

The created defect constellations depend on the particle type and its energy. To quantify the damage to be expected due to NIEL, the fluence is converted to an equivalent to neutron irradiation with 1 MeV energy and given as fluence per area. The typically used quantity is the neutron equivalent per square centimetre $n_{\mathrm{eq}}/\mathrm{cm}^2$. Typical values for this quantity are 10^{15} $n_{\mathrm{eq}}/\mathrm{cm}^2$ for the outermost pixel layer in the ATLAS inner tracker (ATLAS ITk) and an order of magnitude more for the innermost layer [ATL17].

The main types of single-atom lattice defects (point defects) are shown in figure 3.14: unoccupied lattice positions (vacancies), atoms in inter-lattice positions (interstitials), their combination (a vacancy next to an interstitial, called Frenkel defect) and substitutions of a lattice atom with an atom of a different element.

The underlying processes are Coulomb scattering with the nucleus (for electrons, charged pions and protons) and elastic and inelastic scattering with the nucleus for neutrons. The minimum energy necessary to displace an atom in silicon is 25 eV [VLFLN80]. The maximum energy that can be transferred to a silicon atom can exceed this minimum by several orders of magnitude for nuclei. Those silicon atoms can start a cascade of more displacements leading to cluster defects, i.e. a small region with many defects.

A significant fraction of the lattice defects recombines again. However, those defects can also form stable defect clusters. Frenkel defects are especially likely to heal. This process is called annealing. Including relocation of atoms in the lattice, thermal excitation will contribute to the process leading to accelerated annealing for slight heating (to 60 °C or 80 °C) while cooling will prevent it.

The defects can cause electronic states in the band gap that are able to accept or emit a charge carrier. Depending on their position in the band gap, the effect is different:

Generation of acceptor and donator centres States close to the valence or conduction band are created. They change the effective doping. Hence, the depletion zone can change and even type-inversion of n-doped silicon to effectively p-doped silicon is possible.

Generation of generation/recombination centres States in the middle of the band gap enlarge the probability of generation of electron-hole pairs leading to more generation current or recombination reducing the charge carrier concentration. This shot noise is proportional to the leakage current and can lead to thermal runaway.

Generation of trapping centres Defects deeper in the band gap than normal dopant states can trap electrons or holes and, by this, reduce lifetime and mean free path of the charge carrier and by that delay charge so that is does not contrinbute to signal shaping.

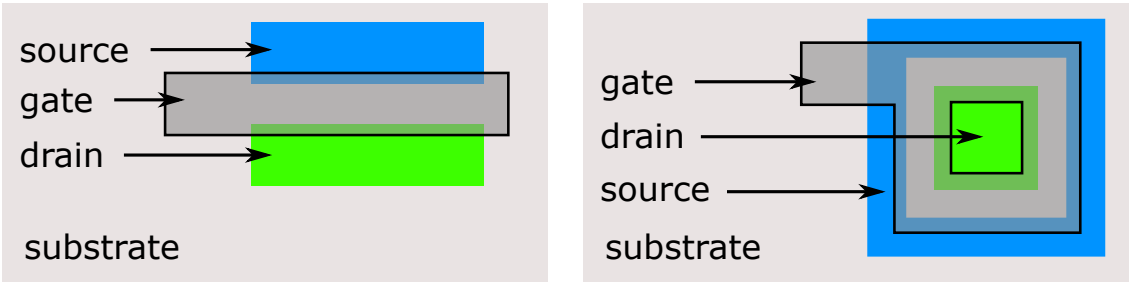
3.3.2 Surface Damage

Surface defects occur in the passivation layers at the surface of silicon devices. Especially, this happens at the interface of silicon to silicon-dioxide. The processes involved are the same as in the bulk material: ionisation and displacement.

Displacement damage in the oxide, does not contribute to the electrical properties as the structure of the oxide is amorphous.

In the oxide, the mobility of the charge carriers is lower than in silicon modifying the effects of ionisation: The charge carriers exist longer if they can not recombine. They are prevented from recombining more likely in presence of electric fields separating them. This is the case for example in the gate oxide of transistors. The electrons have a higher mobility in silicon-dioxide and can eventually leave the gate oxide, but holes move slowly and get trapped permanently in deep traps at the interface between silicon and silicon-dioxide. This leads to a positive oxide charge modifying the effective gate voltage.

Advances of silicon processes to smaller feature sizes have a beneficial effect on radiation hardness: These technologies have very thin gate-oxide layers in the order of a few nanometres [TBB+00], enabling tunneling of the holes out of the gate oxide. In this case,



(a) linear transistor (b) annular transistor

Figure 3.15: In standard linear design, current can bypass the gate after irradiation. In an annular transistor, the drain implant is surrounded by gate and source so that this leakage path is blocked. After [Sch16]

they do not contribute to the gate oxide charge and the influence on the effective gate charge is reduced.

The positive oxide charge results in larger threshold voltages for PMOS transistors to achieve conduction. NMOS transistors, however, have their threshold voltage lowered. At some point, the transistor becomes always conductive [Ehr15].

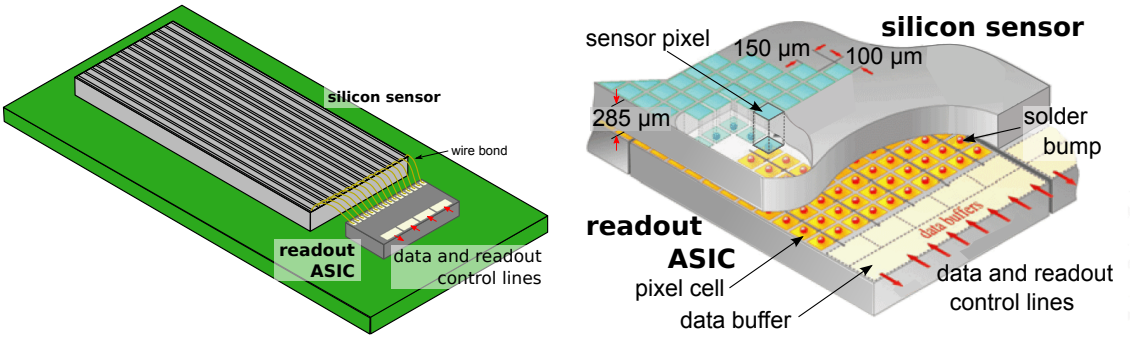
Another contribution to the radiation damage is the lattice mismatch between silicon and silicon-dioxide which leads to dangling bonds at the interface creating trap states at the interface. During production, they are saturated with hydrogen atoms. However, the hydrogen atoms are not strongly bound to the lattice, so that they can be removed by irradiation reactivating the interface traps. For the linear transistor geometry, it is possible for current to flow around the gate increasing the leakage current of the transistor. This leakage path is suppressed by an annular structure of the transistor where the drain implant is surrounded by the source implant. The geometries are shown in figure 3.15. The annular transistor layout is predominantly used for NMOS transistors since this structure occupies more space for the same conductivity and PMOS transistors can sustain larger amounts of radiation in linear layout.

3.4 Silicon-based Particle Detector Technologies

Silicon detectors have been realised in different ways over the years. In all cases, the sensing element is a pn junction with or without reverse biasing. Traversing particles generate electron-hole pairs that are separated in the depletion zone by the electric field and the charge is collected by the segmented electrodes to be amplified and discriminated. The property affecting the detector system conception most is the segmentation: Silicon sensors exist segmented in one dimension, so-called strip sensors, and segmented in two dimensions, the pixel sensors.

Despite the low charge carrier density of semiconductors compared to metals, these densities are still orders of magnitude larger than the amount of charge typically generated in interactions of particles with the detector material. However, in the depletion region around a pn junction, the number of charge carriers is further reduced, making the generated charges significant and by this detectable.

The signals from the diodes are routed to readout electronics. These amplify the signals and discriminate the outcome. Depending on the application, timestamping and amplitude measurement are done in addition to addressing. The last step is the multiplexing of the signal data to output data links. For current detectors, the readout electronics have been



(a) strip sensor (b) hybrid pixel sensor

Figure 3.16: The common approach for both strip and pixel sensors is to use separate dies for sensor and readout. Pixel sensors are connected via bump bonds in a flip-chip process, while strip sensors can be connected via wire bonds. (b) after [CMS21].

implemented as dedicated readout ASIC that is connected to the segments of the silicon sensor.

Strip sensors can be connected to the readout ASIC with wire bonds at one or both sides of the strips. The second dimension of the segmentation in pixel sensors prevents this approach. Pixel sensors are connected to the readout ASICs via bump bonds in a flip-chip process. Consequently, the readout ASIC is placed next to the strip sensor and front to front for the pixel sensor as sketched in figure 3.16.

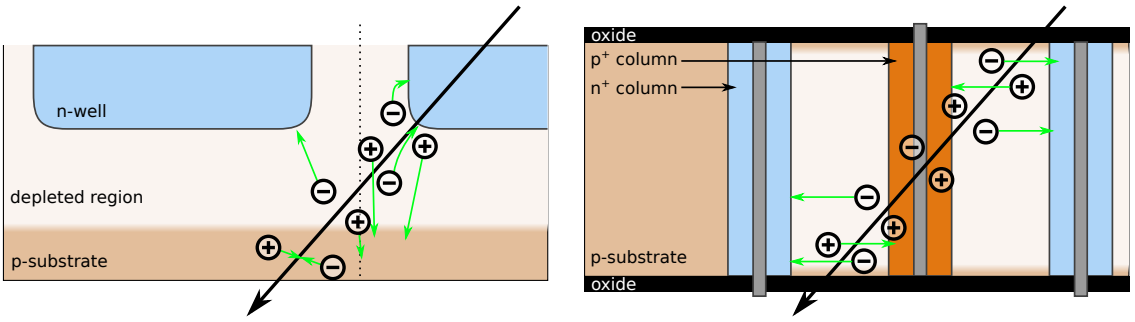
In the following, technologies to build pixel sensors are discussed with more detail. Their application would also allow for building strip detectors but it is mostly not reasonable because of process and application constraints. Furthermore, the noise contributions to be expected from silicon sensors are discussed.

3.4.1 Sensor Technologies

Most silicon sensors use a depletion voltage to enlarge the depleted volume increasing the fraction of active material of the sensor. The sensing diodes formed by the (typically) low-doped substrate and the high-doped implants result in an asymmetric extension of the depletion zone almost completely extending into the low-doped bulk material as visible from equations (3.11) and (3.15) for the approximation $N_A \ll N_D$. This doping concentration difference is even increased for some sensor types by the usage of higher-resistivity substrate. The higher resistivity is the result of a lower dopant concentration leading to an increased size of the depleted volume.

The increased depletion volume enlarges the path length inside depleted volume of a charged particle traversing the sensor. Since only from the depleted volume, the charge is collected efficiently, this results in larger signals for the same particle type. Signals of X-rays do not get larger as they are either absorbed or not. However, the larger depletion volume increases the chance of the deposited charge being detected by a segment. The processes leading to charge generation to be collected from the depleted volume are described in section 3.2.

Because of the larger mobility of the electrons compared to holes in silicon, the preferred implementation for sensors with fast charge collection are p-substrate sensors with n-doped implants for charge collection in the segments. The holes with smaller mobility are collected by the common electrode for all segments where they add to the sensor current where time resolution is not critical.



(a) planar sensor (b) 3D sensor

Figure 3.17: Charge generated by traversing particles is only collected from the depleted volume. In planar sensors, this is the area around the n-well implants. Deeper inside the bulk, the depletion zone ends and charge recombines. In 3D sensors, the whole thickness is depleted leading to large signals, but the pillars for the collection electrodes are insensitive.

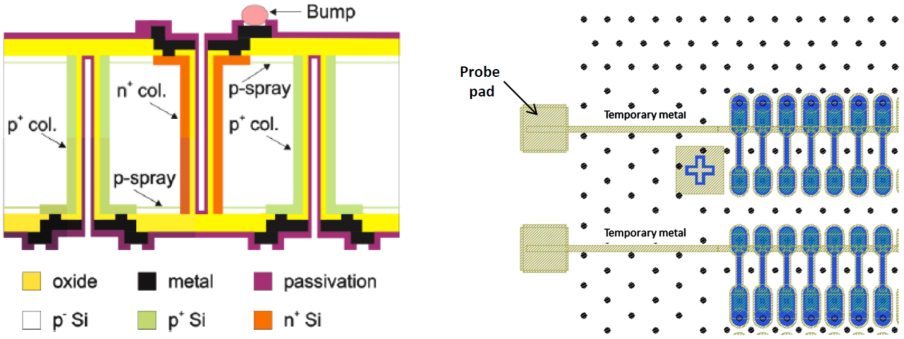


Figure 3.18: The electrodes of a silicon 3D pixel sensor are etched into the wafer material making them perpendicular to the sensor surface. This reduces the path generated charge has to drift before reaching the collection electrodes and utilises more of the sensor material for charge generation. For minimising the distance between the electrodes, they are arranged in a hexagonal grid. [Lan15]

For pixel sensors, two approaches exist to build the sensor diodes: planar sensors and 3D sensors. The former implement the diode in vertical direction between the n-implant at the top forming the collection electrode and the p-doped bulk material. Particles passing the sensor perpendicular to its surface generate charge along this direction and for charge collection, the charge is drifting (or moving by diffusion for undepleted sensors) vertically to the collection electrode. This process including charge sharing between neighbouring pixels is shown in figure 3.17a.

For a 3D sensor, the collection electrodes are created as vertical pillars through the wafer (see figure 3.18). In this implementation, the charge is collected by drift sideways. This has the advantage that on one hand the full sensor material can be depleted at lower depletion voltages and the drift length of the generated charge is shorter than for planar sensors (see figure 3.17b). This makes 3D sensors attractive for precise timing applications and high-radiation environments. To obtain the advantages mentioned, the electrode spacing has to be smaller than the depletion thichness in planar sensors. With pixel sizes of 50×250 µm^2, this is a reasonable choice compared to hybrid pixel sensors [Lan15].

The drawbacks of this technology are that the area of the electrodes can not be sensitive

and that the process is more complex than the planar sensor. Therefore, this technology is applied only where the performance gains are necessary – as for the innermost layers of a tracking detector at LHC [ATL17].

Another technology are monolithic sensors. These integrate the sensor diodes and the readout electronics on a single die. This renders the bump-bonding and flip-chipping obsolete, simplifying the assembly process and open the opportunity to reduce the material budget of the sensor layers in a tracker. They are implemented in commercial complementary metal-oxide-semiconductor (CMOS) processes. The depletion zone can be enlarged by the use of high-resistivity substrate (HR-CMOS) or a high reverse-bias voltage (HV-CMOS). This sensor type is described in more detail in section 3.5.

3.4.2 Noise Sources in Silicon Sensors

While the implementations differ between the technologies presented, the signal chain is comparable: The signal is generated in the sensor diode, amplified and shaped, and digitised. Consequently, the effects disturbing the signal are of the same kind, too. The contributions to this noise are listed in the following:

Thermal Noise is caused by thermal movement of free charge carriers (Brown's movement). From that point, thermal noise occurs both with and without a current flowing. The fluctuations expected from this contribution can be described by

$$\mathrm{d}\langle i^2\rangle = \frac{4k_bT\mathrm{d}f}{R} \tag{3.22}$$

for a resistance R, the Boltzmann constant k_B and temperature T. The term is independent from the current as stated before. Also, being temperature dependent, this contribution can be reduced by cooling.

Shot Noise is a contribution from statistical fluctuations of charge carriers when they pass a potential wall. Such a process is electron-hole pair generation in a reverse biased sensor diode leading to leakage current even before the breakdown. The reverse process, recombination, satisfies this constraint, too. The fluctuations for this process, that can be described by the Poisson distribution as for the single carrier contributions, are

$$\mathrm{d}\langle i^2\rangle = 2eI_0\mathrm{d}f \tag{3.23}$$

With the average current I_0 through the barrier.

Flicker Noise or $1/f$-noise is a noise contribution with a frequency dependency of $1/f^\alpha$ with $\alpha = 0.5 - 3$. There is no general theory for it as it appeary in many different types of systems. In MOSFETs, it appears in the conduction of the inversion channel: Charge carriers can get trapped in the substrate and the gate oxide changing the conductivity and by that cause noise.

Pick-up Noise is a noise contribution not necessarily caused by the measured system. It is activity coupled into the system from outside. System in this case may denote the whole ASIC, but also parts of a circuit. One example of pick-up noise is cross-talk. There, the system can be one pixel with amplifier and the external source is another pixel detecting a signal where the comparator output couples to the input of the other pixel.

To quantify a system and its reaction to noise, the answer to the expected signal is compared to the answer to noise. Those two answers are set into relation to obtain the signal-to-noise ratio (SNR). The essence of this value is that not the absolute size of the noise is important, but its relation to the signals expected to distinguish signal from noise.

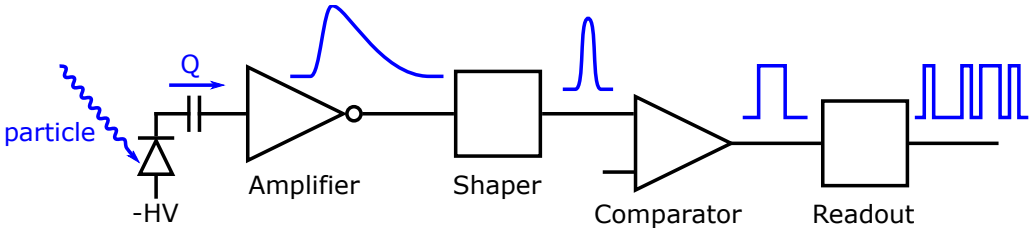


Figure 3.19: The monolithic HV-CMOS sensor includes not only the pixel, but also amplifiers, shapers, comparators and the readout structure on a single die. After [Ehr21]

3.5 Monolithic High-Voltage CMOS Sensors

Monolithic HV-CMOS sensors are active pixel sensors built in commercial CMOS processes. The sensing element is a diode formed by a deep n-well with the p-doped substrate. This means that the electrons of the carrier pairs generated by traversing particles are collected in the pixels. In-pixel electronics is placed in the deep n-well as floating logic making the pixels active. The depletion zone is enlarged with a high depletion voltage to collect the charge by drift rather than by diffusion. For this reason, these detectors are also referred to as high-voltage monolithic active pixel sensor (HV-MAPS).

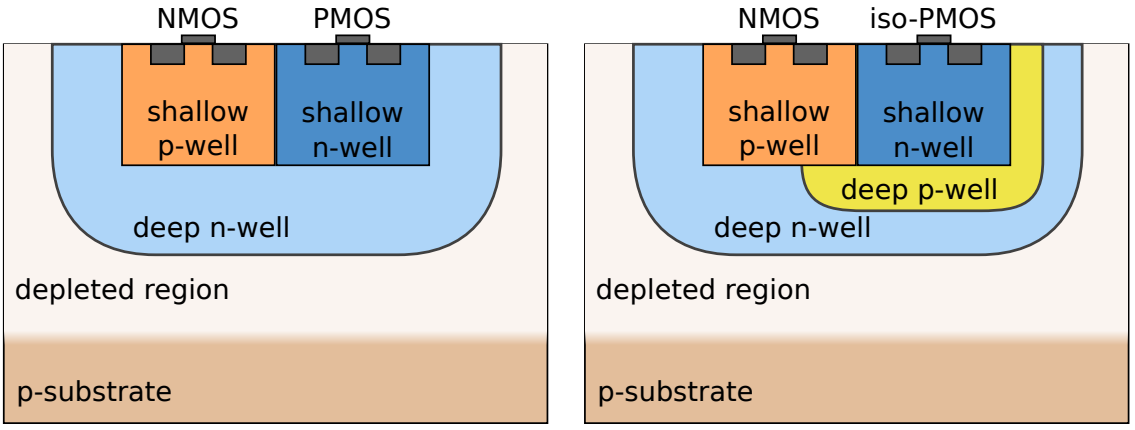
The development of HV-CMOS sensors started with active pixel sensors that were meant to capacitively couple the signal to a readout ASIC to remove the need for bump-bonds [PBC+17]. But the target was to fully integrate the whole signal processing chain into one single die [Per07]. The simplified signal chain necessary for a monolithic sensor is shown in figure 3.19. The integration of amplifier, comparator and readout into the sensor die make the monolithic HV-CMOS sensor an integrated sensor.

The implementation of the parts of the signal chain may vary from design to design, but the concept stays the same: The charge collected by drift in the pixel is amplified by a charge-sensitive amplifier (CSA). Its output signal is shaped and fed into a comparator for digitisation. This binary – but not synchronous – signal is then fed into the readout structure of the ASIC. It registers the signal and assigns time and space coordinates to it. For the application in HEP, the readout scheme is a zero-suppressed individual signal readout with address information for each signal as well as the signal arrival time and optionally amplitude information that may be encoded in different forms. From this point, the readout structure acts like an arbiter and multiplexer to transmit the information on the signals serially via a data link at speeds over a gigabit per second [PSB+19].

The die area of a monolithic sensor is split into two parts: The sensitive pixel matrix, and the insensitive periphery. The latter contains pads for wire bonding, configuration registers, biasing structures and readout circuitry.

Without additions to the standard CMOS process, it is not possible to have both PMOS and NMOS transistors inside the pixel collection well. This leads to insensitive areas inside the pixel that can be avoided when in addition to the deep n-well, and shallow n-well and p-well, also a deep p-well can be used. The cross section of these structures are given in figure 3.20.

The deep wells are created using diffusion. This means that these wells do not have acute corners or edges. Instead the transition to the surrounding material is smooth and rounded. This is indicated in the figure with the rounded shape of these wells. This is important for the depletion as corners in the pixel diodes would lead to field spikes which would cause early breakdown and by that limit the possible depletion voltage and hence reduce the signal size of charged particles.



(a) triple well process (b) quadruple well process

Figure 3.20: In the standard triple well process, PMOS transistors are not always possible as they are not completely isolated leading to insensitive areas in the matrix. In the quadruple well process with a deep p-well, the PMOS transistors can be isolated from the deep n-well enabling a completely sensitive matrix. After [Web21]

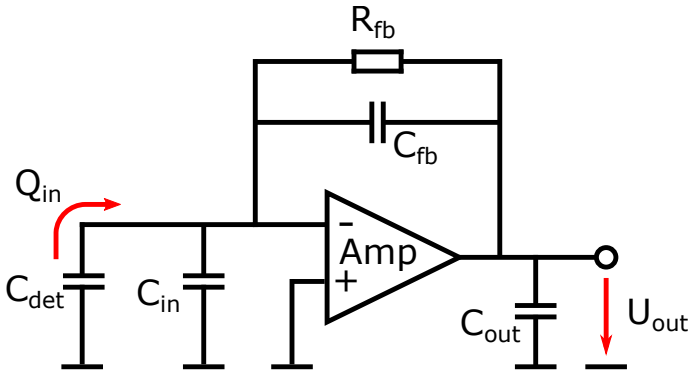


Figure 3.21: Simplified schematics of the charge sensitive amplifier with input capacity of the amplifier shown as individual component. [Ehr21]

The depletion layer thickness in HV-CMOS sensors with 200 Ωcm resistance is in the order of 30 µm to 50 µm. With the design of the sensor for the small depletion depths and no necessity for large structural rigidity, monolithic HV-CMOS sensors can be thinned down to 50 µm [ABD+17]. This reduces the material budget significantly compared to hybrid sensors requiring two dies that have to be rigid enough for bump bonding. At 50 µm, the handling of the sensors is the limiting factor, not necessarily the sensor itself. In addition, with the achievable depletion thicknesses on higher resistivity substrate can reach full depletion for which the sensors have to be prepared as surface defects on the backside can cause leakage paths when fully depleting the sensor.

Being collected by drift, the charge generated by a traversing particle results in a short current pulse at the charge collection electrode. As the information of interest is the amount of charge, it is converted to a voltage by a CSA. The concept behind the conversion is the integration of the charge. Therefore, the circuit of the CSA is similar to the integrator circuit for an operational amplifier.

The simplified schematics of the CSA are shown in figure 3.21 with the input capacity C_{in} of the amplifier as extra component. From this, the output voltage U_{out} of the circuit at

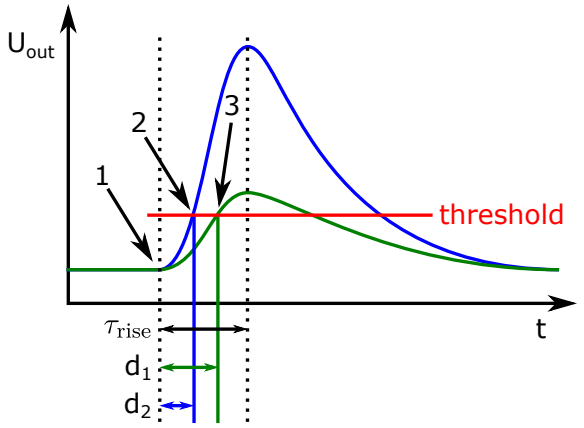


Figure 3.22: For a fixed threshold, signals of different size show a different detection time with the charge sensitive amplifier's rise time independent of the charge. This delay difference is referred to as time-walk. [Ehr21]

an input charge Q_{in} is

$$U_{\text{out}} = -\frac{Q_{\text{in}}}{C_{\text{fb}} + (C_{\text{det}} + C_{\text{in}} + C_{\text{fb}})/A} \tag{3.24}$$

with the amplification A of the amplifier, the feedback capacity C_{fb}, the detector capacity C_{det} and the output capacity C_{out}. The resistor R_{fb} discharges C_{fb} defining the resetting time scale of the circuit. In actual designs, R_{fb} is a tunable resistance for continuous reset of the circuit. Designs with an active reset already have been used in this position, too.

For detectors that are supposed to measure the time of arrival of a signal, also the rise time τ_{rise} of the CSA is important. An estimation for it is

$$\tau_{\text{rise}} = \frac{C_{\text{out}}C_{\text{fb}} + C_{\text{out}}C_{\text{det}} + C_{\text{fb}}C_{\text{det}}}{g_m C_{\text{fb}}} \tag{3.25}$$

for the assumption that the rise time is much smaller than the fall time with the transconductance g_m of the amplifier. [Per04]

The rise time is independent of the amount of charge, the charge does only influence the output voltage. This leads to different detection delays for signals of different size (see figure 3.22): Smaller signals are detected later than larger signals. This effect called, time-walk, is important for signals close to the threshold, for large signals, the delay saturates.

Using the amplitude information, time-walk can be corrected for as the characteristics of the amplifier can be measured and applied to data. As the feedback is constant, a larger signal also becomes longer, too. Therefore, instead of the amplitude, a time-over-threshold measurement can be used to estimate the amplitude. This measurement only requires a second timestamp for the end of the signal while an actual amplitude measurement requires additional circuitry.

Another contribution to time uncertainty of the signal detection is jitter – a time uncertainty for digitisation due to noise. On one hand, jitter in the timing signal itself leads to time uncertainty, but also noise as discussed in section 3.4.2 contributes to it: The signal arriving at the comparator is the overlay of the actual signal and noise contributions. These can alter the threshold crossing point. The effect is stronger for smaller signals as the slope of the rising edge is smaller.

4 Measurement Infrastructure

As part of this work, I have developed test systems for the HV-CMOS sensors. They partially build on previous work and are improved and extended in functionality and user-friendliness.

The GEneric Configuration and COntrol (GECCO) system has been developed by me starting from the GECCO setup developed by F. Ehrler in his PhD thesis [Ehr21]. It is a test system intended to be quickly adaptable to newASICs to test. The hardware has been adopted from the GECCO setup, as well as the base structures of firmware and software. However, the structure in the firmware modules has been changed for better maintainability and the software has undergone large restructuring in the work for this thesis. All parts have been written to be reconfigurable for different devices-under-test (DUTs) generalising the code to remove the need for testing it for every DUT. Interfaces to laboratory devices and management structures for large amounts of data are also included. The system is described in section 4.1 split into hardware (section 4.1.1), firmware (section 4.1.2) and software (section 4.1.3). For high data rates from the DUT, the system has been extended with an Ethernet data transmission channel from firmware to a computer with a dedicated software described in section 4.1.4.

In addition to the electrical tests possible with the GECCO system, a scanning TCT setup has been designed and built as part of this thesis. It enables partially replacing measurements with particle beams or radiation sources. It integrates with the GECCO system software, but it is designed as an independent component that can be integrated in other measurement software as well. The TCT system is described in section 4.2 with the description of hardware (section 4.2.1) and software (section 4.2.2).

4.1 Generic Configuration and Control (GECCO) System

The GEneric Configuration and COntrol (GECCO) system consists of hardware, firmware and a collection of software projects. It is designed as ecosystem for rapid test system development for sensors. All parts aim for collecting reusable structures to reduce the effort of adapting the system to new devices. In addition, the hardware aims for cost reduction of the system per tested sample by building it modular and reusing of components removing them from the device carrier. As a third point, the simplification of operation – especially during debugging with temporary signal probes – is tackled by this system providing simple access to all signals for debugging.

It has been developed as the existing test system projects do not meet all requirements for the application at KIT-ADL. With Caribou, Basil and YARR, three examples are given in appendix A for comparison.

Since the start of the project, the system has been used for nine devices in the KIT-ADL working group. Some of these chips are rather simple test chips, others full-featured prototypes for building tracking detectors. The devices tested with the GECCO system are (with publications or theses if applicable):

- ASTROPIX (see [Str21])
- ADC chip in 28 nm technology (see [Pra20])
- HitIntegratingPix
- HitPix (see [Web21])
- MPROC (see [Zha21])
- H35Demo (see [Bad20, Ehr21])
- MuPix8 / ATLASPix1 (see [Ehr21])
- MuPix10 (see [Web21])
- ATLASPix3 (content of this thesis and [PAA+21])

The versatility of the GECCO system becomes apparent as it is not only used for single devices, but also for groups of devices: The H35Demo has been tested as a sensor triplet, MuPix8 and ATLASPix1 have been used as a five-layer beam telescope and ATLASPix3 is used as single device, in a four-layer beam telescope and as a quad module [RGS+21, Rac21].

By the writing of this thesis, about 60 units of the GECCO board have been produced and the majority of the systems have been shipped to other universities with only a fifth of them left at KIT. In the silicon tracker project for future circular lepton colliders – future circular electron-positron collider (FCEPC), the system is used by collaborating institutes at:

- University of Bristol (GB)
- University of Edinburgh (GB)
- Lancaster University (GB)
- University of Liverpool (GB)
- Rutherford Appleton Laboratory (GB)
- INFN Milano (IT)
- a number of Chinese universities in the FCEPC collaboration, e.g. IHEP Beijing

For the Mighty Tracker project of the LHCb experiment [Par20, LHC19], the GECCO system is also going to be used for the test campaign of the prototype HV-CMOS detectors ASICs by a group at the University of Heidelberg.

To increase outreach, a logo has been created for the project (see figure 4.1). For other universities, it helps distinguishing the GECCO system from other systems and the name gives an easy term to refer to.

The hardware of the GECCO system is described in section 4.1.1, before the firmware (section 4.1.2) and software projects (sections 4.1.3 and 4.1.4) are presented. A more detailed description of the parts can be found in [Sch21a].

Figure 4.1: The logo of the Generic Configuration and COntrol (GECCO) readout system has been created to enhance the recognisability of the system. Merely, this is a benefit for communication between other institutes utilising it to distinguish this system from others.

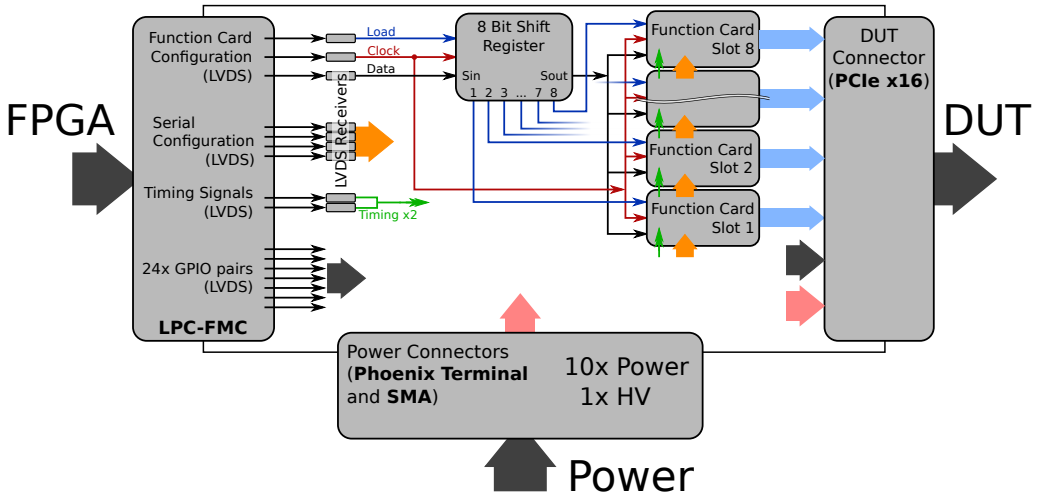


Figure 4.2: The GECCO board receives power via Phoenix terminal connectors and an SMA connector for the depletion voltage for the device-under-test (DUT). The connection to the FPGA is established via a low pin-count FPGA mezzanine connector (LPC-FMC) and the DUT is connected via a PCIe x16 connector. In addition to prepared data lines for configuration of the FunctionCards and four configuration lines for the DUT, 24 differential GPIO pairs are available. The FunctionCards share the data, clock and load lines where the load signal is de-multiplexed via a shift register in the data line before the FunctionCards. The PCIe connector for the DUT accommodates all data pins from the FPGA, power inputs and the output signals from the FunctionCard slots.

4.1.1 Hardware

The central element of the GECCO system is the GECCO board designed by F. Ehrler and described in his PhD thesis [Ehr21]. The GECCO board sits in between an FPGA board and the DUT. It collects data and power lines to provide them via a single connector to the DUT carrier board. Reconfigurability is achieved by the implementation of eight function card slots that can be equipped with cards of different functionality and sustaining the line mapping as differential pairs for the lines from the field-programmable gate-array (FPGA). The eight outputs of the function cards are routed to pins on the DUT connector. A sketch of the structure of the GECCO board is shown in figure 4.2.

The function cards can be configured via a shift register interface consisting of a clock, a data line and a load signal. The implementation of this scheme is shown in the centre of the figure and takes into account that not all function card slots have to be equipped: The data and clock lines are connected to all slots in parallel, the load line is demultiplexed for the eight slots using a shift register in the data line before the slot inputs. Hence, the load signal connected to the shift register is activating the signal for the desired slot.

The connector to the FPGA board is a low pin-count FPGA mezzanine card (LPC-FMC) connector. The pin mapping is the default one for this connector [VIT08] with 34 differential pairs, of which 24 are directly routed to the DUT connector. The other differential pairs are

reserved for function card configuration, and special signals for configuration and test signal generation. The special function pins present on the FMC connector (as a multi-gigabit transceiver) are routed to the DUT connector as well.

The DUT connector is mechanically a peripheral component interconnect express (PCI express, PCIe) connector for 16 lanes. Its 164 pins are not used according to standard but instead power inputs, function card outputs and FMC signals are connected in a way to enable also a reasonable mix of function card slots, power and differential lines when using the smaller form factor PCIe connectors (with 64 pins or 96 pins).

The function cards group commonly used features on reusable boards enabling adaption of the system to different DUT. On the connectors for the function cards, six single-ended signals generated on the GECCO board from differential signals from the FPGA are present. In addition, a 3.3 V and the DUT supply voltage VDD18 are present for powering the circuits on the function card and for the output levels of the signals generated. Currently, there are five function card types:

ConfigCard
: The ConfigCard routes four of the six special signals to four of the eight output lines of the card slot. They are intended for serial configuration of DUTs using a two-phase clock shift register scheme. The direct connection from the FPGA for these lines enables fast data transmission.

InjectionCard
: The InjectionCard contains two circuits for generating charge pulses of adjustable size for test signal generation on DUTs. The pulse size is set via digital-to-analogue converters (DACs) on the card while the other two special signals are used as timing signals for pulse generation.

VoltageCard
: The VoltageCard contains eight DACs with an amplifier each for generating digitally controllable analogue voltages. The applications are analogue references as comparator threshold or baseline settings.

RegisterCard
: The RegisterCard houses an 8 bit shift register driving the eight outputs of the function card. Hence, it provides digital output signals that can be used for slow changing signals as reset inputs for the DUT.

PowerCard
: The PowerCard is an option between the power inputs of the GECCO board and the VoltageCard. It provides between two and four power connections for medium currents due to the output over the signal lines of the function card. At least two output lines are used per power line, but two power inputs each can be grouped on the card with jumpers to double the current limit of the inputs.

Several elements of the concept enable operation of a DUT without a common ground potential between FPGA and DUT: For the configuration – which is typically driven by single-ended signals – the ConfigCard can be used. Slow changing single-ended signals can be driven with RegisterCards. The configuration of the function cards uses differential signals with line receivers on the GECCO board, too. If no other single-ended signals are to be transmitted from the FPGA to the DUT carrier or vice-versa, the ground potentials can be separated by removing a jumper on the GECCO board.

For debugging, the DUT connector footprint is present twice on the GECCO board, one for connecting the DUT and the other one for probing the signals. The GECCO board and a selection of function cards is shown in figure 4.3.

(a) GECCO Board (b) function cards close-up

Figure 4.3: The GECCO board is the central hardware element of the GECCO measurement system connecting the FPGA board (LPC-FMC connector), power supplies (Phoenix terminal connectors) and the device under test (PCIe x16). To adapt the setup to the needs of different devices, eight one-lane PCIe connectors provide slots for function cards.

Schematics and more detailed information can be found in [Ehr21] and [Sch21a].

4.1.2 Firmware

The firmware is written in Verilog and built as a framework for the DUT-specific user code. Functionality that is used for every DUT is already included. It is based on the firmware used in previous firmware projects developed by the author and F. Ehrler (see also [Ehr21]). The target device is the Digilent NexysVideo[1] board. A structure diagram of the firmware indicating the pre-included modules and the places where adaption work for a specific DUT may be necessary is shown in figure 4.4: The dark modules and solid connection lines are prepared, only the light-grey modules and the dashed connections are to be user-generated as they are DUT-specific.

The most prominent example for reusable code is the communication with the software running on the computer via the USB data connection. For this, the state machines to interface with the USB driver IC (the FTDI FT2232h present on the Digilent NexysVideo board) is provided with modules to multiplex and demultiplex the data using an addressing scheme. The transmission system is described in more detail in [Sch21a]. Only the content of additional entries has to be added by the user of the framework. Furthermore, modules for signal generation timing and time reference (trigger) input signals are pre-included. An implementation of the digital readout is included to serve as template for user implementations. The code of these modules were restructured to improve their versatility and reusability in new projects.

Ethernet Readout of Digital Readout Data

Since for some applications the USB 2.0 connection data throughput to the computer is too small, the firmware implements modules to send data via the user datagram protocol (UDP) connection [Pos80] over Gigabit Ethernet. USB 2.0 is capable of transmitting data at up to 48 MB/s. However, the data transmission protocol used for communication with the software limits the throughput to about 4 MB/s in optimal conditions. It requires initiation

[1] `https://reference.digilentinc.com/programmable-logic/nexys-video/reference-manual`

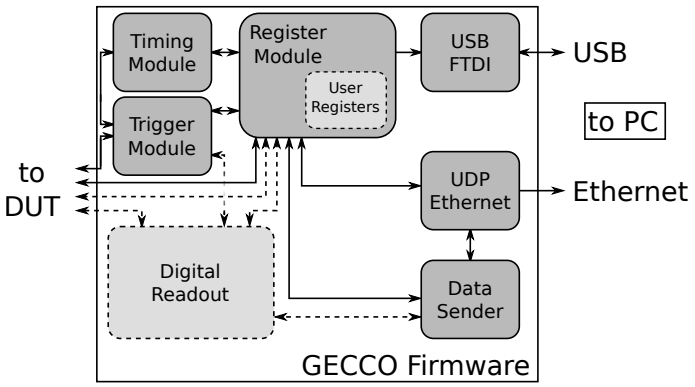


Figure 4.4: The GECCO firmware consists of a number of reusable elements as the pattern generator in the timing module, the trigger manager or the UDP readout, the register file and the connected USB communication modules configuring all other parts of the firmware and the additional DUT-specific modules that are to be written by the user. The content of the register file has also to be adapted by the user for the DUT to match the requirements of the other user written modules. The framework modules are drawn with solid lines, the user-created content with dashed lines.

of all operations from software, requiring a bidirectional communication for single direction data transmission. The UDP connection is not limited by such a protocol enabling data transmission reaching about 30 times the achievable speed of the USB connection.

The UDP package sending is implemented without a soft-core processor and is based on work of A. Forencich [For19]. The setup of the module with network configuration is included in the register module and configured via the USB connection. The network configuration includes the media access control (MAC) address, internet protocol (IP) address, net mask, gateway address and the port for the UDP protocol to work on. Two operation modes are implemented: A mode for establishing a connection as well as for testing it and one for actual data transmission off the FPGA.

As UDP is a connection-less transmission protocol, the address (i.e. combination of IP address and port) to send the data to is the origin address of the last UDP package received. The connection establishing mode will answer to a received package to test the connection. The data sending mode will not make use of any data received apart from updating the destination address for its own packages and not answer on received data. Instead, the data sending mode reads data from a FIFO and builds data packages from it. They contain a package identification number to identify data loss and the package size is optimised for writing the data efficiently to hard disk.

Digital Readout

The digital readout module receives digital data from the DUT and preprocesses them. But as it is a very DUT-specific module, the included implementation is intended to be used as template for the user to adapt it to the DUT. To help with this, the module is built with substructures:

1. a data receiver module for deserialising the incoming data,
2. an alignment module for finding the data words in the stream,
3. a decoder module for resolving the data link encoding,
4. a timestamp generator module,

5. optionally, a trigger-signal generator module and
6. a data interpreter module for sorting, decorating and storing the data in FIFO buffers.

The serial data is received in the deserialising module which hands over blocks of data to the alignment module that implements structures for automatic alignment to data words. For an 8b/10b Aurora style encoding [Xil14] (as used for example on MuPix8), this is the identification of two specific data words (for example `0b1100000101` and `0b0011111010`) and an implementation for this is already included in the firmware. For the 64b/66b version of the Aurora encoding, the identification is done via the exclusive-or relation of the two first bits in the data word: For a number of consecutive data words, their relation is tracked and for any other than the correct alignment, the exclusive-or relation will be eventually broken. An implementation for this is included in the firmware, too. The output of the alignment module consists of the data words as they are finished and are routed to the decoding module.

Further content of the digital readout module are timestamp and trigger signal generation modules. The timestamp generation module mimics the timestamp generation of the DUT but with more bits to decorate the signal data from the DUT for off-line analysis. For DUTs with triggered readout, the trigger signal generation module provides a structure to obtain trigger signals. Both modules receive the same synchronisation reset signal as the DUT to synchronise them.

The decoding module output and the timestamps generated by the timestamp generator module are handed to the interpreter module. It contains a state machine that traces the on-chip readout state machine using the received data. This enables interpretation of the data, its decoration (for example with the extended timestamps) and storing of the data in a FIFO buffer with zero suppression. To actually use this data, each signal has to fit in one entry of the FIFO (with an input width of 64 bit) or an identification pattern has to be included.

The global trigger module of the firmware can provide more information as an identification index and timestamp for an input signal in the firmware that can be added to this data as well. An example for such a signal can be a scintillator that is hooked up with the firmware. This signal input can be used to start the DUT trigger module or just be stored as time reference in the data written to the readout FIFO.

4.1.3 Software

The GECCO software project is written in C++ with usage of the Qt framework[2] for the user interface. It is built to be quickly adapted to new DUTs. To enable low-level testing of DUTs, the software is supposed to both allow for run-time modification of all operation parameters of the DUT while at the same time provide an easy-to-use framework for higher-level routines for complex measurements.

To serve both applications, a well-structured data management is necessary. This also means that the user interface is only one entity accessing the configuration data. Hence, the software is structured in a back-end part holding the configuration and the higher-level routines, and a front-end part with the user interface that can display and alter the configuration structures in the back-end as well as control the higher-level functions.

A third requirement on the software is that it has to be quick to adapt to new DUTs even for people who are not programming experts. This creates the need for detailed predefined structures for all parts with good abstraction towards the higher-level functionality. Also, automatisation of tasks is to be fostered in the design of the framework.

[2] `https://www.qt.io/`

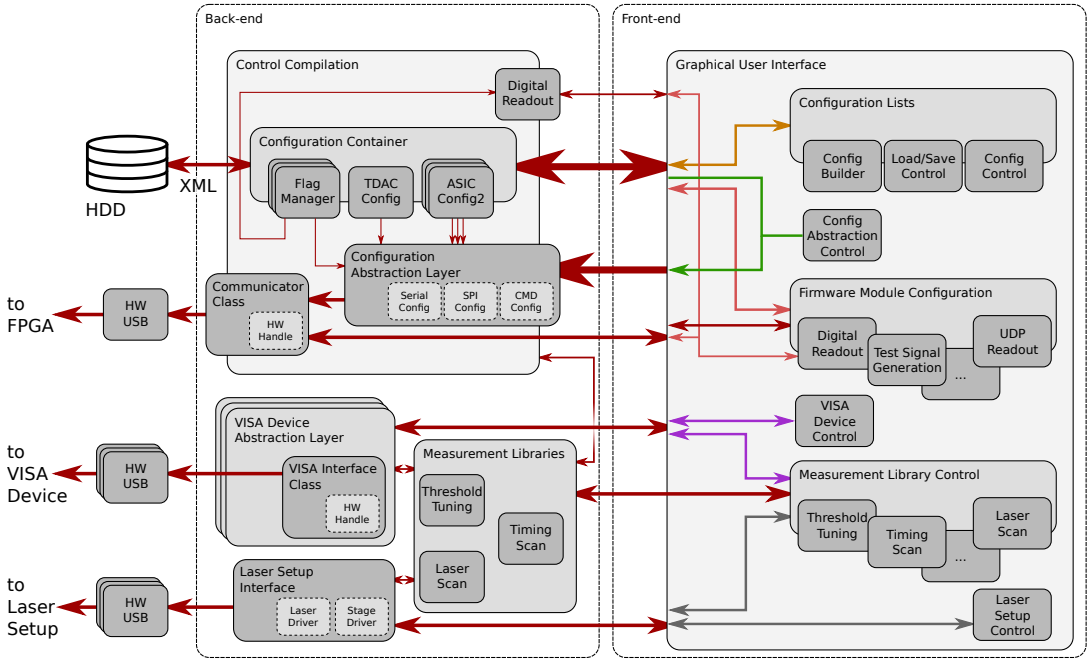


Figure 4.5: The GECCO software is divided into front-end and back-end. The back-end stores the data which the front-end is displaying and providing means to modify. Configuration of the chip and communication with other measurement devices is hidden behind abstraction layers to ease the use of different equipment and to enable measurement algorithms independent of the configuration method used. The configuration can be saved to and loaded from hard drive to provide a simple way of managing different chip configurations.

The resulting structure of the software is sketched in figure 4.5: It is divided into back-end and front-end.

The back-end contains the configuration abstraction layer (see section 4.1.3.2) for writing the configuration to the DUT via the different interfaces. Also the configuration data collected in dedicated containers (see section 4.1.3.1) can be stored to hard-disk and loaded from there. Data transmission happens via a class wrapping the USB hardware interface and managing the transmission protocol to the firmware. The digital readout data decoding-class also belongs to the low-level functionality structures contained in the back-end. It has to be adapted to the digital readout module in the firmware to decode its output data.

The higher-level structures in the back-end contain measurement libraries (as described by reference of an example in section 4.1.3.4) as used for pixel tuning and libraries to interface with other lab infrastructure. This can be the laser measurement setup from section 4.2 or instruments implementing the virtual instrument software architecture (VISA) interface specified by the IVI foundation[3] (see section 4.1.3.5). For complex measurements, also data management libraries allowing for hierarchical structures and data compression are included.

The front-end contains controls for modifying the configuration of the DUT and selecting the interface to use for the transmission of it, controls for configuration of firmware modules and controls for using the measurement libraries (see section 4.1.3.6).

The classes storing the configuration data in the back-end all contain methods for import and export of data from or to files. The format used in the files is extensible markup

[3] `https://www.ivifoundation.org/specifications/`

language (XML). This enables direct modification of the files and also cross-usage of configuration files between versions of the software for which the content may have been changed. For ease of use, one file is used to collect all configuration trees from the objects containing the configuration data.

The container for per-pixel configuration data which is stored in a second file. This split is intentional as for the same configuration file can be used for measurements on different pixel sets without the risk of other parameter modifications between them.

The user interface presents the loading and saving section with the list of file names used, sorted by the time of last usage. This reduces the time to select the configuration for the measured sample avoiding the use of hard-coded default values in the software that may change between samples and hence cause problems with repeatability of measurements.

4.1.3.1 Configuration-Data Container Structures

The GECCO software makes use of several configuration data container structures. Three categories of settings exist covered by a data container class each:

- Configuration for the DUT is stored in the ASIC_Config2 class.
- Per-pixel configuration for the DUT is stored in the TDAC_Config class
- Configuration for firmware modules is stored in the FlagManager class

All three classes are optimised for their application: The content of the containers is generated at run-time which means that adaption to a new DUT is only a matter of reconfiguration rather than of changing the program. Also, import from XML files and export to them is implemented for all three classes with the same interface.

In contrary to the other two, the DUT configuration does not have addressing as the whole configuration has to be written every time. Also, the amount of data in the DUT is larger and it is more complex than in the byte based configuration of firmware modules. And the per-pixel configuration contains data for all pixels while the FlagManager contains data for one object.

The benefit of the run-time generation is that the code generating the data for transmission does not change and hence does not require retesting on changes on the configuration fields. In addition, adaption of the data content consists only of simple calls of the methods to add the fields.

The interface to XML files includes not only direct file access with optional ZIP compression, but also tree structures of the XML library used (tinyxml2 written by L. Thomason[4]). The latter option enables combination of several trees in one file. For this reason, the configuration data containers all contain identification strings that are used for the XML import and export.

In the following, the three configuration data containers and the class, managing their combination into an XML file and the separation of them from XML file input, are described.

Chip Configuration Container

The ASIC_Config2 class represents the data for a shift register for the configuration of the DUT. For example, this configuration can be settings of biasing structures or settings for digital circuits on the DUT.

The data is stored as an integer value per field in the configuration. Each of these field values is accompanied by a name for identification and the information of the bit order.

[4] https://github.com/leethomason/tinyxml2

Access to the fields is possible via the index in the list or via the name. For the latter option, the names have to be unique. Uniqueness is not enforced by the class for larger flexibility.

The field names are intended to make code modifying the configuration data more readable favouring development speed over execution speed. For perfomance-critical parts, faster access via indexing is possible, too.

Only upon request, the bit vector to write the shift register is generated from the data contained in the data fields. To ease setting a system up, reversing the bit orders is included in the implementation.

Other features of the class are aliases for single bits, bit inversion and optional fields:

- An alias is an alternative name for accessing one bit of another register. This is useful for bits that have several functions depending on the context, because then the bits can be accessed with informative names for all use-cases. The aliases have a higher priority than the field names themselves and add to the list of names. Hence, the alias names combined with the field names have to form a set of unique entries.
- Each bit output can be inverted for the bit vector. An example use-case of inverted bit usage is to achieve a non-zero default configuration from a register written with all the same bit values.
- The optional fields are a structure to combine two configurations in one object if one is a subset of the other. The bit vector can be generated with these optional bits or without them resulting in two configuration streams. If a shift register is placed in the data path for the configuration and can be bypassed, the data for the shift register can be marked as optional. When bypassing the extra shift register, the optional parameters are left out for the bit vector and are included if using the extra shift register. Hence, the two configurations do not require separate data containers that would have to be synchronised.

All data is added to the containers at runtime enabling complete reconfiguration at any time. With the XML import, fields are created replacing the content inside the object. The import and export functions are complete, that means that when exporting the configuration to a file and loading it, the same content is created.

The large benefits of the usage of this class are the time reduction of the development and testing of the code: Firstly, the transition from the integer value to the bit order is written generally so that this code does not require testing for each field. Secondly, the fields are created with one function call giving all necessary information and thirdly, import and export are already present, too. The created XML format also enables simple exchange of settings between sites testing the same DUT even if the test systems are not the same.

Tuning Parameter Container Class

Some DUTs implement memory for every pixel for example to equalise thresholds. The TDAC_Config class is built to store this per-pixel information for the whole matrix in one object. There is only one integer value per pixel with a limited interval of legal values taking care of the number of bits available on the DUT. There are two options for addressing the entries for the pixels: Either, they can be addressed by their position in the matrix or by a custom addressing with a value per row and a custom column address per pixel in each row.

The size of the matrix is defined at runtime and the values can be imported from and exported to XML files that can be ZIP compressed. The repetitive pattern of the matrix grants large space savings in this case.

Firmware Module Configuration Container

The FlagManager class is the equivalent to the ASIC_Config2 class for firmware modules: It holds integer data fields for the configuration of firmware modules that are given a name for accessing them. The bit order can be specified similarly, too. But bit inversion and optional parameters are not available. The same is for aliases that do not exist for the FlagManager. Also, the output is not a bit vector, but an integer value that is supposed to be casted to a byte as this is the standard size of the registers accessible in firmware. For this register access structure, the FlagManager class also contains a member storing the register address to which its data are supposed to be written to.

The data fields however are not restricted to the 8 bit of the output, but to the 32 bit of the underlying integer data type. For such "non-writeable registers", the plausibility check of the data fields can be turned off. Normally, it would prevent using bits twice.

In the default firmware, there are already several modules to be configured making use of the FlagManager class. While the existence of several objects in the software is not a problem, the identification string is crucial when combining the configuration output in one file to be able to distinguish between the data trees for the different objects when loading the data.

Import and Export Data Distributor

For collecting the tree representations of the configuration objects and writing them to a file, the ASIC_Config2_Manager was created. Initially, it was written for the ASIC_Config2 objects of the DUT configuration which are internally converted to tree representations. However, the tree representations of the other FlagManager objects were to be added to the combined file, too. Consequently, the ASIC_Config2_Manager class can take tree structures of the internally used tinyxml2 library as well as the ASIC_Config2 objects it was intended for.

The remaining difference between ASIC_Config2 objects and FlagManager objects is that for the latter ones the tree representation generation has to be called manually for adding them to the managing object.

The ASIC_Config2_Manager object acts as a collector for configuration trees and for writing them to an XML file or loading them from such a file. After loading, the tree structures for contained elements can be requested by the element tag and the identifying name.

4.1.3.2 Configuration Abstraction Layer

At the transition from the low-level access to the parameters of the DUT to the higher-level functionality, the necessity of an abstraction and data collection entity arises. The Configuration class implementing this abstraction layer collects all configuration for the whole test system as references and the routines for writing the configuration to its respective destinations via the available interfaces.

The configuration data consist of all shift register configurations for the DUT (ASIC_Config2) and for the firmware modules (FlagManager). The collection as references enables simple backup and restore functionality for higher-level functions, where setup parameters are to be altered or scanned, by creating a copy of the configuration and replacing the reference in the Configuration class. The configuration procedure becomes therefore independent of the location of the data and also nested procedures can operate without interfering with the changes to the configuration made by the calling function: Any routine receives a single Configuration object to work on.

The other part of the Configuration class is the collection of routines for transferring data to DUT or firmware registers. Partially, these routines are integrated in the class itself, but more complex parts are outsourced to dedicated classes:

- The NexysIO class contains the implementation of the data transfer protocol for the USB connection and includes the interface to the hardware library. This library is taken over from previous software projects.
- The communication via the serial peripheral interface (SPI) is collected in the SPINexys class which uses the NexysIO class for the transition to the FPGA enabling SPI data transmission controlled by software and by a state machine in firmware.
- The data transmission for the command decoder implemented on the ATLASPix3 integrated sensor (see section 6.1) is done with a state machine in firmware. The data encoding and the transmission to the FPGA is collected in the CMDConfig class. With the Configuration class, it also manages the data transfer rate to not lose data at the FIFO buffer of the firmware module.

4.1.3.3 Digital Readout

The software counterpart of the digital readout module in the firmware is the FastReadout class. It receives the data recorded by the firmware module for decoding. In addition, the configuration of the basic functionality of the firmware module is stored in this class.

The data from the digital readout module is read from the FPGA by the user and provided to the decoding methods using the user interface functions. The data is converted to a Dataset object containing all information provided as decoded values. These Dataset objects can be converted to text for display and output in files.

The decoding procedure requires input from the firmware module design as well as from the DUT design: While the state machine in the digital readout module traces the on-chip state machine, data is collected and combined to blocks of 64 bit to be written to the data FIFO for transmission to the computer. These blocks contain several header bits for identification to determine the content. With this information, the data content from the DUT can be separated into the different fields and the values can be decoded. Typically, this includes address corrections and Gray decoding of timestamps.

Depending on the complexity of the data to be decoded, the data for one signal can consist of several parts requiring buffering of incomplete data waiting for the remaining parts to be decoded.

4.1.3.4 Measurement Libraries

For more complex measurements, the code needs to be structured to keep it reusable and adaptable to new DUTs. This is illustrated at the example of the threshold tuning procedure for the pixels: This tuning procedure consists of several measurements and parameter scans that have to be repeated to obtain the information needed. Details on the measurement algorithm are given in section 6.4. Here, only the program and data structures are sketched.

To determine the detection threshold of a pixel, the size of the test signal or the comparator threshold has to be scanned. This scan has to be performed for every pixel and for several settings of the per-pixel threshold adjustment.

To handle the multi-dimensional problem, the atomic elements are written as individual methods that are called by composition methods. This splits the task into manageable parts: One composition method scans one parameter and calls other composition methods until

the atomic methods are reached. The data from the subordinate methods are decorated with the information from the own parameter scan, before being analysed and the result together with the decorated data is passed back to the superordinate method. For this decoration, appropriate data containers have to be used where advantage of the associative containers of the Standard Template Library can be taken.

In the example of threshold tuning, composition methods are scans over pixels or groups of pixels, over the pixel threshold adjustment settings or over the test signal size. The atomic measurement would be the calculation of the detection efficiency for the given parameters. Analysis of the data can be a fit to the data which can again be structured in the same way. The decision algorithm for the following steps takes the data measured and returns a result for another method for adjusting parameters and writing them to the DUT.

This disentanglement of the steps increase the development speed and reduce the risk of errors in optimisation of the procedure by separating preparation and execution steps and breaking down the complexity of the steps. Using the strategy design pattern, adaption of the code to new algorithms, procedures or applications is possible, too. The latter point can be observed again on the threshold tuning procedure: The methods and data containers are reused for other measurements as timing measurements or calibration of the signal size measurement.

For storing complex data with associations between blocks of data, also a library for data management is provided with the Measurement class: It is intended for aggregated data and to be stored in – optionally ZIP compressed – files. The data is structured in measurements, that consist of events. Each event contains a number of data items. These can be waveform or histogram data which are tailored for data from oscilloscopes or general data stored as textual data. Encoding for the last type is up to the user.

The data is exported in an XML tree structure. This grouping reduces the number of files potentially generated in measurement campaigns keeping the data managable. Furthermore, the measurements can be supplied with additional information to ease documentation. With compression hard disk space is saved and the usage of a common library to base the data structure increases development speed for the analysis.

4.1.3.5 Measurement Device Integration

For integration of data from laboratory devices, a virtual instrument software architecture (VISA) interface is integrated into the software. To ease adaption to new hardware, the interface makes use of the strategy design pattern, where a base class defines a wrapper interface for the VISA driver and depending on the device to open, the appropriate implementation is used.

With this implementation, the differences in commands and data structure between different devices can be hidden from the user. Consequently, the measurement setup does not rely on one specific oscilloscope or device.

The structure builds upon the oscilloscope readout software developed for [Sch16].

Compared to the separate software, the direct integration of the VISA device interface in the software simplifies the synchronisation of DUT control and data taking with the VISA devices. Consequently, complexity of the measurement procedures is reduced without inter-software communication and their development is sped-up.

4.1.3.6 User Interface

The GECCO software is designed not as a static collection of code for testing devices, but as a starting point for modification, adaption or extension to the needs of the DUT. This

explicitly includes development during the testing of the DUT. For this reason, the user interface of the GECCO software is designed to grant the user direct access to as much of the functionality of the system as possible. As a consequence, the user interface is explicitly designed to pick usage speed for the experienced operator and during development over ease-of-use.

The main adaptions for this concept are described in the following: They include generation of control elements for DUT configuration and integration of elements for newly added functionality.

Interface Building

For fast testing, all parameters of the DUT configuration are to be available for modification in the user interface. In previous control software projects, a time consuming part of the preparation has been the setting-up of the elements in the user interface for parameter modification in the configuration of the DUT. To reduce this, the user interface in the GECCO software builds the elements for modification of the DUT configuration automatically from the configuration data itself: Data fields with one bit are created as a check-box, fields with several bits are created with a horizontal slider and a spin-box. For the connection to the back-end, the elements are registered in data structures for generic access to them and event handlers for user interaction are generated. Apart from the time saved during development, the automatic generation of the user interface also reduces the risk of errors in this modification structure.

An important part of the testing procedure is documentation. In case of the control software, this involves the parameter settings used for a measurement. Therefore, elements for import and export functionality of the complete configuration are prominently placed in the main view. The file names used for this are listed in chronological order in the drop-down lists for the file names with the most recently used one at the top.

For development of new procedures and debugging, direct read and write access to the values of the registers in the firmware is integrated to the user interface, too. This way, new functionality of the firmware can be tested before controls for it have been added to the user interface. After initial tests, the debug access is replaced by dedicated elements in the user interface that modify a configuration container structure (e.g. FlagManager objects, see section 4.1.3.1) in the back-end and send its resulting configuration to the firmware. These data containers are to be integrated into the import/export files and in the transfer functions for data from back-end to front-end.

Interface Structure

With all content for DUT configuration, firmware module configuration and test procedure control elements, the user interface for ATLASPix3 is shown in figure 4.6: In the ASIC Settings tab starting on the left, the elements for the USB connection to the FPGA are placed. Below this, configuration data import and export elements follow, before the elements for direct firmware register access. The rest of this column is occupied by elements for sending the configuration for the DUT. The empty space is filled only for some configuration interfaces.

In the middle, the tab widget contains the automatically generated elements for the DUT configuration with the check-boxes, sliders and spin-boxes mentioned above. Each of the shift registers of ATLASPix3 is placed in a dedicated tab and for RAM writing and test signal control, a summary and control tab Col/Row/TDAC was added.

The right half of the user interface is occupied mostly by elements for firmware module configuration as for the timing module (with the group Injections as for its application), the UDP

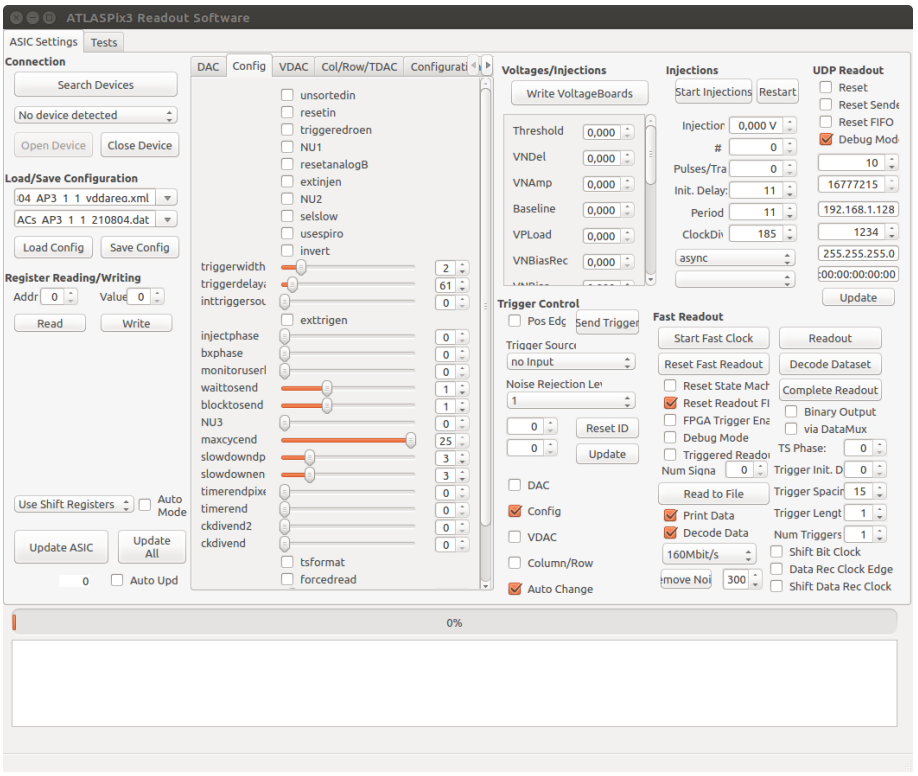


Figure 4.6: The user interface of the GECCO software is made to grant direct access to as much of the functionality of the system as possible choosing functionality over ease-of-use. On the left, the most basic elements for the USB connection, data import/export and direct firmware register access are placed. Next to it, the tab area contains the DUT configuration and the rest of the user interface is filled with elements for control of firmware modules as the timing module (labelled Injections), the UDP data readout module or the digital readout module (Fast Readout). The content of the DUT configuration tab area is generated from the configuration provided at run-time.

data readout module (UDP Readout), the two optional VoltageCards (Voltages/Injections), trigger input control for the FPGA (Trigger Control) and the digital readout (Fast Readout). The controls for the VoltageCards is generated from the configuration, too. The five check-boxes at the bottom (DAC, Config, VDAC, Column/Row and Auto Change) control which shift registers of ATLASPix3 are to be written when clicking the Update ASIC button on the left. The last check-box (Auto Change) enables automatic modification of the list to transmit the currently selected register – in case of figure 4.6 the Config register. The exclamation mark on the Update ASIC button indicates modification of the data since the last transmission of the configuration to the DUT.

At the bottom of the window, a progress-bar and a log window are placed. Text can be added to the log window with a function adding the same text to the log window and the terminal output of the software. It is a public method of the class containing the user interface. The progress-bar is intended for longer operations and available to the user via the Configuration class also able to provide information on whether it is used or not. This feature is for writing nested functions where only the function called first is supposed to use the progress-bar to show the execution progress.

Controls for the more complex tests as threshold tuning or timing scans are placed in the Tests tab of the software (as shown in figure 4.7). The threshold tuning procedure to equalise the detection thresholds of the pixels is controlled in the Trimming group and

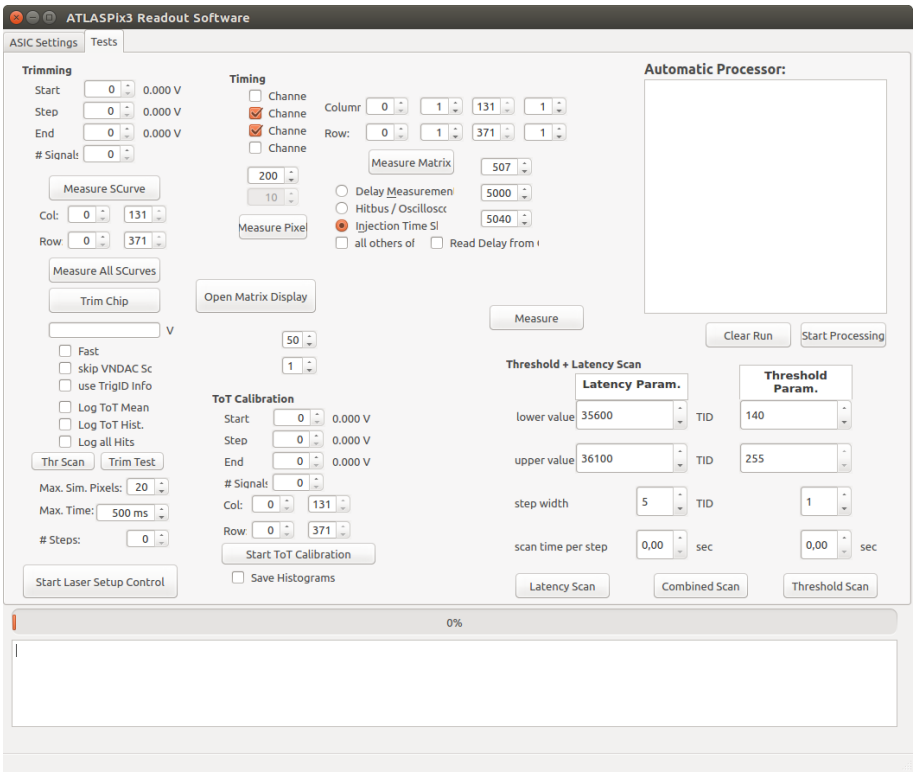


Figure 4.7: The user interface of the GECCO software has a second tab for control elements for higher-level measurement procedures. This includes threshold tuning (Trimming), matrix timing measurements (Timing) or calibration of the amplitude information via ToT (ToT Calibration). Also, a command list execution system is included on the top right (Automatic Processor).

will be used for the measurements presented in section 6.4. The control elements for the amplitude measurement calibration via ToT to be used in section 6.5.3 are grouped below the ToT Calibration header. For matrix timing, the measurement control elements are included in the Timing group.

Three other elements are to be mentioned here: Firstly, the Automatic Processor giving the possibility to describe procedures similar to clicking elements of the user interface that can be executed one step after the other automatising small measurement campaigns. It is to be used for one-time measurements where the effort of creating the user interface exceeds the benefit. The execution of entered commands is logged into a file with timestamps of execution start and end. In the command list window, finished commands are marked as *done* and in addition as *failed* if an error occurred during their execution. On an error, the execution of the list is stopped. On re-execution, finished steps are skipped to enable continuing a started procedure after fixing the error. To repeat the complete run, all completition and failure markers can be cleared bringing back the content to its initial state. Being intended for one-time use, the content of the command list is not saved for re-execution, only the executed commands are contained in the log file.

Secondly, a visualisation display of matrix data is included which has been adapted from the UDP data readout software in section 4.1.4. It is opened with the Open Matrix Display button and used to display different kinds of data over the position of the matrix. This can be the position of modified pixel settings or hit maps from digital readout.

Thirdly, the Start Laser Setup Control button opens the user interface for the laser setup described in section 4.2 and VISA device control.

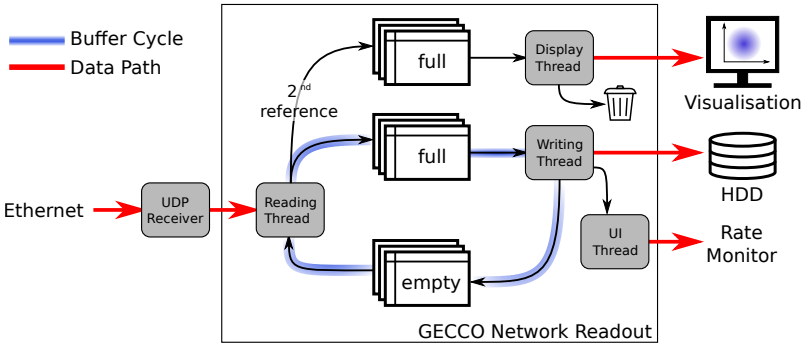


Figure 4.8: The data received from the network card is buffered in the software to write the data to hard disk and optionally to generate hitmaps from it. The data is held in memory that is passed between the threads via three queues.

4.1.4 Readout via Network

For high-speed data readout with the GECCO system, a UDP data sender module has been added to the firmware. The software counterpart is the UDP readout software. Both are part of the author's work for this thesis. Its main purpose is to write the received data unaltered to hard disk. It follows a multi-threaded approach buffering data in RAM to compensate for operating system or hardware delays. The implementation as an independent software from the GECCO software is chosen to separate the configuration and data taking effort and enable the usage of separate machines for the tasks increasing the robustness of the system. This way, an error in one part will not directly affect the other.

The high data rates, the firmware module can generate on the network, have to be read from the input buffer of the network card quickly as its buffer will only be able to store about 80 UDP packages.

To cope with this, the software runs with four threads: The first thread reads the data from the network card to a buffer structure in RAM. The second thread runs analysis methods on the data in the buffer for generating hit maps if an appropriate data decoding is provided. The third thread takes the data from this buffer to write it to hard disk. The fourth thread runs the graphical user interface and calculates metrics to display the data rate.

Using a three-queue system, the memory allocated is passed between the reading, writing and displaying threads as sketched in figure 4.8: The data buffer in RAM consists of a number of memory blocks for the UDP packages and three queues. At the beginning, the memory blocks are allocated and a reference for each block is filled in the queue for empty blocks. The reading thread takes a reference from this queue, fills the data from the UDP package into it and adds the reference to the other two queues for filled blocks. The monitor thread reads the references from the full queue and analyses the data. The writing thread reads the references from the other full queue, writes the data to hard disk, clears the block content and adds the reference again to the queue for empty blocks. This way, no memory is allocated or freed during execution. The number of buffers to use can be defined in the user interface and is adapted on starting a new recording. Since the monitor thread does not require access to hard disk, it is faster processing the data than the writing thread and it is possible for the two threads to work on the same data.

This recording thread structure is started from the user interface to take the data. On start-up of the software, it is not running. The user interface of the software presents the user with the address-port combinations for the FPGA and the computer to use. It can read and write UDP data packages after opening the port on the network.

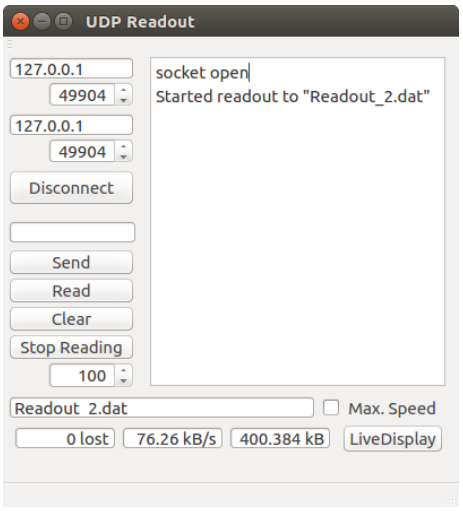


Figure 4.9: The user interface of the UDP readout software consists of connectivity widgets and widgets for testing the connection as well as for data taking: At the bottom of the window, three field show the data taking performance with number of lost packages so far, data rate and total amount of data taken in the current recording.

The user interface is shown in figure 4.9. The main part of the user interface is taken up by the log window. On the right are the elements for the connection at the top, then the reading and writing elements for single packages, clearing the log window and the button to start and stop recording data to a file with the spin-box to set the amount of buffers to use for the operation. Below the log window, the file name for the file output can be entered. At the bottom, monitor data for the currently running recording is shown next to the button to open the visualisation of the hit map. This data calculation can be turned off with the Max. Speed check-box. The data rate is averaged over 5 s and calculated for completely filled packages giving an upper limit for the amount of data. The total amount of data makes use of the same assumption. However, the data packages are stored in a 1024 byte block pattern, i.e. smaller packages are padded with zeroes to the same size giving the amount of disk space used.

All settings in the user interface are automatically stored in a configuration file that is loaded on start-up. This way, the configuration does not have to be entered on every start of the software reducing the time between start-up and data recording. If the output filename ends with a number, the value is incremented on termination of the recording to directly restart the acquisition with the next file name.

For visualisation of the hit map, the recorded data have to be decoded before insertion into a histogram. Similar to the VISA implementation in the software (see section 4.1.3.5), the decoding is implemented using the strategy design pattern. The general functionality is implemented in a base class on which a class for each readout mode of each DUT is based. Using runtime polymorphism, the decoder for the user-selected DUT is instantiated and used on the data. These decoder classes only need to decode a subset of all data provided by the data stream: Since only the number of signals or their amplitude measures are used, only the pixel address and amplitude measures have to be decoded. The other data fields can be skipped reducing execution time. This data are then added to the data container for the visualisation. The starting point for the development of the visualisation is test code for the visualisation of the UDP data written by A. Nürnberg.

For the visualisation, the matrix size and aspect ratio of the pixels is adapted to the DUT selected by the user. To be able to rescale the visualisation, the data are stored as

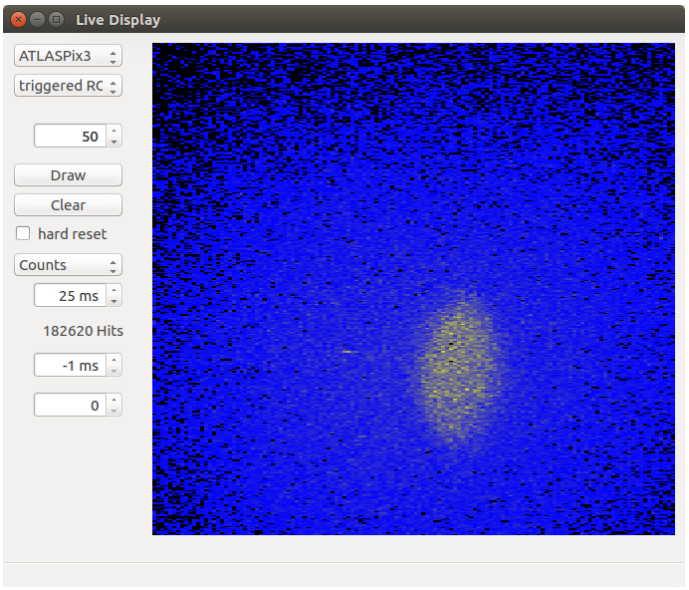


Figure 4.10: Optionally, the recorded data can be decoded and visualised in an extra window. There, the decoding method can be selected from the implemented options and the display style can be adjusted. The data shown are a carbon ion beam through a copper collimator measured at HIT.

per-pixel value (counts or summed-up amplitude information) and colour scale. From this information, the actual visualisation is built.

The visualisation is integrated into a second window of the readout software and is shown in figure 4.10: In the top left corner, the DUT and the readout mode for it are selected. Below, the display parameters can be adjusted: Starting with the maximum value for the colour scale, over redrawing and data clearing options, metric to use, update time, integration time to layer selection.

The data visualisation takes the rest of the available space in the window and for its purpose of visualisation, this window also implements rescaling of the hit map display to the window size. However, the aspect ratio of the pixels is maintained so that the smaller value of width and height determines the drawing size.

The update time is the interval at which the image shown is updated, the integration time is the duration, after which the content of the data container is cleared to only show the most recent data. For reference, the number of signals in the visualisation is given, too.

4.2 Laser Measurement System

To replace a part of beam test measurements and measurements with radiative sources, a scanning TCT measurement system (see [Kra15]) has been built as part of this thesis. For this, a pulsed laser is positioned above the sensor to generate signals in the depleted volume. Such a system consists of a pulsed laser of appropriate wavelength to achieve the penetration depths for probing the depletion layer of the sensor and translational stages for moving the focusing optics over the DUT. With an attenuator, the amount of charge generated can be varied enabling adaption to generate signals comparable to other signal sources.

In contrary to the test signal injection circuits implemented on most detector ASICs developed at KIT-ADL, the laser-generated charge represents a real signal that can be generated at different positions in the active volume in the sensor diode.

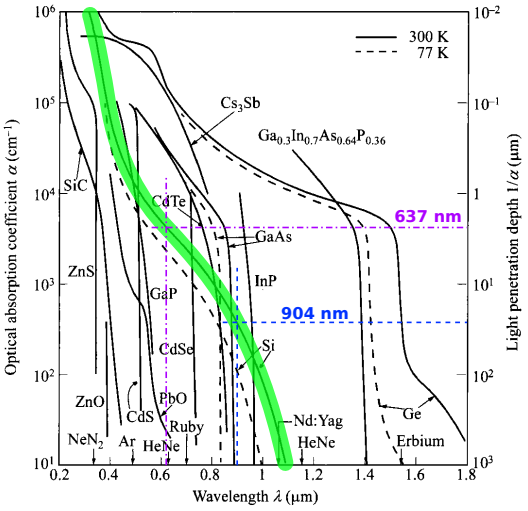


Figure 4.11: Light of different wavelengths is absorbed differently in different semiconductors. As for the application to measure silicon based particle detectors, the absorption line for silicon at room temperature is highlighted in green. The laser wavelengths used for the TCT setup are marked in blue and purple. [SK007]

For the HV-CMOS sensors developed at KIT-ADL, two wavelengths have been picked according to penetration depths from literature as in figure 4.11: The data line for silicon at room temperature (300 K) has been highlighted in green. The penetration depth for the 637 nm laser of about 2 μm and for the 904 nm laser of about 20 μm have been added as well.

The reasoning for these wavelengths is that in some sensors, the charge is to be generated very close to the surface and in others, the charge collection region is located at the bottom of the deep n-well which is at about 10 μm below the surface of the sensor.

For the application to the HV-CMOS sensors developed at KIT-ADL, the spatial resolution of the stages need a positioning resolution below 10 μm and time differences in the order of 1 ns needs to be resolvable.

Short pulse lengths enable measurements of the timing which the used laser systems[5] provide pulses as short as 60 ps (for the 637 nm laser) or 40 ps (for the 904 nm laser).

The translational stages are stepper motor stages[6] and provide a repeated positioning accuracy of 1 μm. Three axes are necessary to scan the sensitive area of the DUT and to compensate for different mounting heights. The key parameters of the system are collected in table 4.1.

One example for the application of the Z height adjustment is the back-side illumination of the sensor: Since the top side of HV-CMOS sensors is covered by aluminium traces, the laser is absorbed before even penetrating the silicon. For thinned sensors, the depleted region can also be probed from the back-side of the sensor where no such structures are present. To perform the measurement, the DUT carrier is mounted upside down. For standard-thickness sensors at about 700 μm the penetration depth of the laser is not large enough to reach the depleted region.

[5]The laser systems are PiLas modules from the company Advanced Laser Diode Systems.
[6]The stages are Mercury class stages from Physik Instrumente GmbH & Co. KG

Table 4.1: For precision measurements in time and space, the setup needs to suffice minimum requirements to be used for measurements on the sensors developed at KIT-ADL.

	red laser	infrared laser
wavelength	637 nm	904 nm
focal diameter (FWHM)	10 µm	
penetration depth in silicon	2 µm	20 µm
repeated position resolution	1 µm	
scanable volume (X × Y × Z)	(100 × 25 × 25) mm^3	
maximum pulse rate	40 MHz	1 MHz
minimum pulse length	60 ps	40 ps

Combining the high positioning accuracy of the stages with the micro-focus of the laser, the sensitivity of the DUT can be studied on sub-pixel level. Hooking up the laser driver with the DUT control system, also time resolution and delay measurements are possible.

In the following, the hardware built is described in more detail, before the software written for it is presented.

4.2.1 Hardware

For measurements at sub-pixel resolution ($\sim$ 1 µm), the pulsed laser is equipped with a micro-focus (FWHM of $\sim$ 10 µm) at the end of an optical fibre. This micro-focus is mounted on translational stages for positioning above the sensitive sensor elements. The combination of laser and translational stages is contained in an aluminium box shielding the system from external light sources and the outside from leakage radiation.

The translational stages are adapted to the HV-CMOS sensors the setup is designed to test: Since the reticle size limits the size of the sensors, a scannable area of 25 × 25 mm^2 is sufficient for the whole sensor. One direction is equipped with 100 mm of moving distance for additional space during installation and removal of the sample. This results in a reachable area of 25 × 100 mm^2 and 25 mm of distance range covered perpendicular to the measurement plane.

Since one of the probing lasers is invisible and both probing lasers are not to be operated without protective measures, the support structure for the micro-focus also holds a cross-line laser for positioning of the probing laser micro-focus above the sample. A close-up photograph of the micro-focus above a sample with activated cross-line laser is shown in figure 4.12.

The box is large enough to house the DUT setup, too. A split setup of DUT inside the laser housing and parts of the readout and control system outside the housing has two problems: Firstly, this can not rule out the necessity of modifications of the setup altering the conditions for the DUT. Secondly, each feed-through in the housing contains the risk of leakage radiation. In case of the GECCO system, the DUT fits inside the laser housing including the GECCO board and the NexysVideo FPGA board.

Connections to power and control computer are enabled through flanges for laboratory measurement lines (also referred to as “banana cable”), BNC and SMA cables, one USB connection which is also used for control of the laser setup itself. In addition to this, an opening with a light trap is available. Through it, also the connection of oscilloscope probes is possible. These flanges on the housing are shown in figure 4.13.

Inside the container box, the DUT carrier is clamped with PVC blocks on aluminium profiles for flexible positioning. The translational stages are mounted on the same profiles

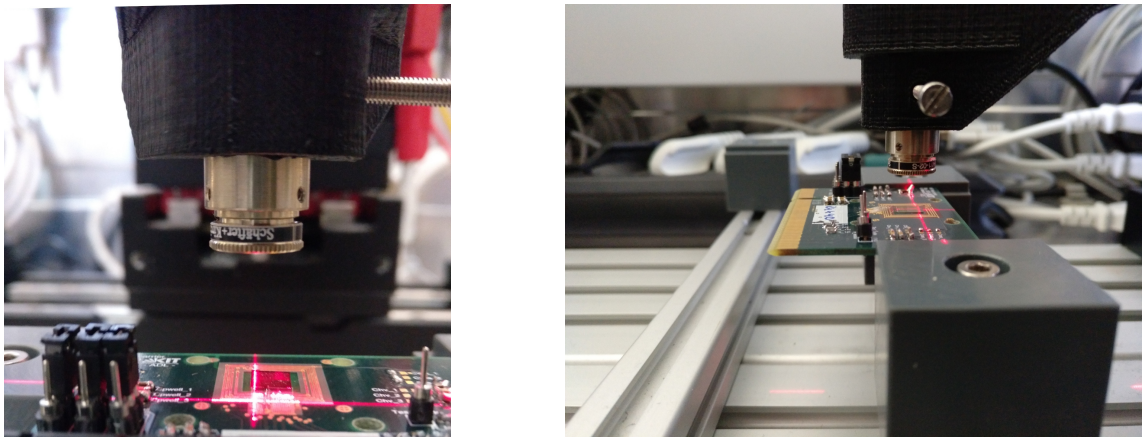


Figure 4.12: The laser beam is transmitted via an optical fibre to a micro-focus that is mounted on the translational stages together with a guiding laser to ease positioning above the sensor. On the photographs, the micro-focus is positioned above a small sensor with the help of the red cross-line laser.

Figure 4.13: The ten flanges for laboratory measurement lines, the three SMA flanges and the two BNC flanges (right to left) enable the operation of the DUT setup inside the box mostly without cables fed through the light trap at the top. The USB flange is used for both the DUT system and the laser setup itself via an USB hub inside the box.

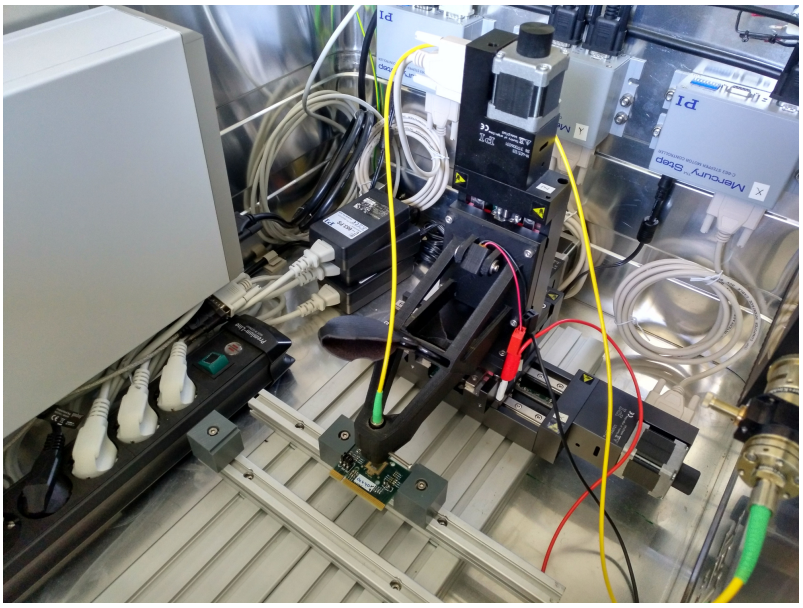


Figure 4.14: The DUT carrier is clamped to the same aluminium profiles as the translational stages giving the rigidity to measure with high precision. The drivers for the stages are mounted on the side wall of the box in the background of the photograph. The laser driver is mounted on the back wall on the left and the laser with the attenuator attached is mounted on the front wall on the right. The yellow fibre connects the laser to the microfocus on the mounting structure with the cross-line positioning laser.

giving the system the necessary rigidity to perform measurements down to the micrometre level. The drivers for the laser and the stages are mounted on the back and side walls inside the box to ease transportation of the setup. The solid state laser itself is mounted on the front wall and connected to the optical fibre leading to the micro-focus via an attenuator enabling the adjustment of the size of the signal over several orders of magnitude. The arrangement inside the box for a small DUT carrier without the rest of the readout and control system is shown in figure 4.14. Despite the laser setup being designed to replace a part of test beam measurements, the components are mounted in the laser setup in a way, that they can be quickly unmounted for use in beam tests or other measurements. Hence, the cables have not be shortened to sufficient length.

The lid of the box is equipped with an interlock system with two switches turning off the laser as soon as one switch opens. The overlap of the closed lid with the walls of the box prevent leakage radiation from the laser while the interlock turns off the laser. The two switches are placed in a way that one person alone can not activate the laser with the lid open.

The driver of the laser implements several signal sources: An internal trigger generating signals at a selectable frequency or two trigger inputs at TTL level and with an adjustable voltage. For the latter, the chopper signal as received on the InjectionCard of the GECCO system (see section 4.1.1) can be used. The debug pins on the InjectionCard enable attaching a cable directly to the card for hooking the laser setup up with the GECCO setup and the GECCO setup can drive the measurement the same way as for other measurements with test signal injections.

For the configuration of the laser setup, the serial interface (RS-232) is used via USB-to-RS-232 converters. This is because the ground potential of the data inputs of the drivers is connected to their housings and hence to the whole box. USB however carries significant

noise on the ground potential. For this reason, the ground potential connection of the flange to the housing is removed and converters are used to not have a common ground potential between the USB side and the serial interface side. With an USB hub at the inside of the USB flange, it is possible to connect the converter for the laser driver, the converter for the translational stages and the connection to the readout and control setup for the DUT with one USB connection.

4.2.2 Software

During this thesis, a control class has been developed for the control of the laser setup and on top of this class, a graphical user interface has been created to increase the integration speed in measurement systems. Both parts are written in C++ and make use of the Qt framework.

The LaserSetup class uses only the non-user interface classes of the Qt framework to reduce the necessary elements for its application. It provides functionality to open the connections to the laser and stage drivers automatically testing the available ports and their return values to specific commands to find the devices. With the open connection, the laser parameters can be configured and the translational stages can be operated.

The LaserControl class uses the Qt framework to create a user interface to grant the user interactive access to operation of the hardware. Additionally, it implements measurement procedures to scan areas (2D) or volumes (3D) using an oscilloscope or an answer from another function to be called. For access to the oscilloscope, the VISA libraries described in section 4.1.3.5 are used. The class is made to be integrated into readout and control software as the GECCO software. The scan procedures accept a function pointer that can be adapted by the user to match the DUT so that only the procedure to measure one point needs to be implemented using the scanning procedure of the LaserControl class.

The measurement with the oscilloscope will record several waveforms for each position and extract the average signal height from them filling this value in a summary file. All data taken is stored using the Measurement class structure mentioned in section 4.1.3.4.

The user interface of the LaserControl class is shown in figure 4.15: On the top left, the connections are managed. On the top right, the laser is configured and activated. A broken interlock will be detected and the user interface will be updated. Since the interlock is implemented directly on the driver, this is only a monitor feature that will react to the break with a delay which for this reason not a security issue. The next part below is the control part for the translational stages with the current position, motor status and reference drive options for the three points available (start, middle and end of the range). The next group is for the measurement scans giving starting and ending position as well as the step size for selectable axes. The scan intervals can also be imported from the current position with ranges via the Import Pos. button. The connection to the oscilloscope necessary for some measurements, is controlled via the Oscilloscope control group below. At the bottom, the log window and a progress-bar complete the window.

From the LaserControl class, a standalone software with user interface for control for the translational stages was derived. It only contains the control elements for the translational stages and is intended for the alternative uses of the translational stages mentioned in the previous section.

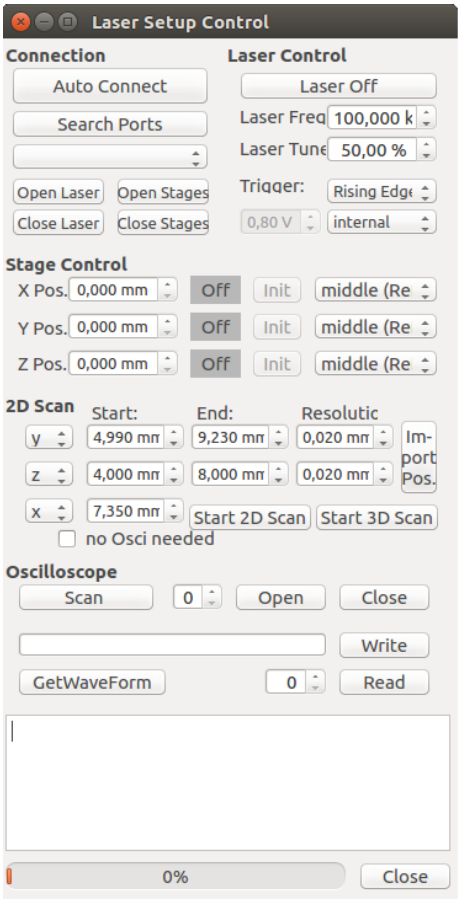


Figure 4.15: The LaserControl GUI is a collection of UI elements to interface with the LaserSetup class in the background of the software. At the top, the connection management and the laser control are located. Below, manual positioning with the stages can be done as well as referencing of them. The next part is for setting up 2D and 3D scans. At the bottom, a VISA device can be controlled that can be used as signal source for the scans.

5 Readout Simulation

The idea of a simulation for the readout architecture comes from a sensor with grouped readout and data compression developed at KIT-ADL: Its projection encoding of the pixels in a group of pixels led in some cases to ghost signals, that resulted from compression loss in the decoding but were generated by the data encoding. For more special cases, these ghost signals were not even recoverable. For this, a small software project was created by F. Ehrler to estimate the severeness of this problem. It is described in [Ehr21]. The simulation has the advantage over the ASIC that every part of the system is known to reconstruct the creation of the problem.

From the resulting toy Monte-Carlo simulation it became clear, that such a simulation can do more than estimating the probability of a special event happening. Assuming accurate input data, also memory size and data loss estimates are possible, even before the design of the sensor has started. The result of this insight is the ReadOut Modelling Environment (ROME) simulation framework. It was started by the author and F. Ehrler, but after the initial conception, the development was fostered by the author alone. Additional testing of the framework has been done by C. Röck in his bachelor's thesis under the author's supervision [Rö18]. With the applications, the concept has been extended and generalised to support more use-cases. As Monte-Carlo simulation, it complements static rate calculations for the dimensioning of the readout structures of detectors. The possible coverage of corner cases as well as implementation details affecting timing enrich the collection of insights that can be gained from this simulation. As versatile extension for development efforts, ROME has been recognised in reviews of sensor developments.

For the application of ROME in the conception of prototypes of the ATLAS inner tracker upgrade (ATLAS ITk) for HL-LHC, a dataset of 100 000 events has been generated. It covers the fifth layer of the pixel tracker for which HV-CMOS detectors are an option for future upgrades. The data has been extracted from the official physics simulation campaign for the ATLAS ITk upgrade [ATL17]. This simulation campaign was gradually increasing the amount of details and design decisions into the system to obtain improved simulation results for the planned detector system. This state is described in simulation steps what are denoted by numbers. First datasets were based on the simulation step 2.2. Later, event data containing the expected collision pile-up has been generated from step 3 simulations. Data extraction and conversion have been performed by E. Zaffaroni (Université de Genève).

The description of the configuration in an XML document including the detector description has been chosen over a hardware description language (HDL) implementation for a good

Figure 5.1: For recogisability, a logo has been created for the simulation framework. The Colosseum was chosen as it is a famous landmark in Rome (Italy) and resembles a barrel particle tracker.

reason: ROME aims at higher-level concepts allowing for an abstraction from the precise description. On one hand, the design description in XML allows for less strict design constraints that can also lead to a faster implementation for the simulation. On the other hand, the different description enforces translation of the concept, giving the opportunity to discover pitfalls of it. For some cases, also the execution of the simulation can be faster with the chosen implementation than with the HDL approach due to a less detailed low-level simulation and separation of timing considerations from structure conception.

During the development of ROME, the system has been used to investigate grouped readout and the influence of partial signal sorting on the sensor. Afterwards, the simulation framework has been extended to support more complex detectors with fractional timing relations to help with the design of the readout architecture of ATLASPix3. For the development of the sensor for the Mighty Tracker upgrade of LHCb, ROME simulations are planned to verify that the design is capable of handling the data rates at all places in the detector and to identify bottlenecks. Furthermore, the beam monitor project for the ion beam at HIT plans to use the simulation for evaluation of different detector arrangement and readout concepts.

For recognisability, a logo has been created for the project. It is shown in figure 5.1. The Colosseum in the logo bridges the name and the application: It both is a landmark in Rome (Italy) and resembles a barrel tracking detector.

In the following, the structure of the simulation framework, the user interfaces and the comparability to measurements (as beam tests) are described, before the development applications grouped readout and partial signal sorting are covered. The grouped readout example serves as showcase for the necessity of a tool for validating readout architectures and that small changes can have a large impact on the design as a whole. Then, the simulation campaign for the development of the readout architecture of ATLASPix3 is treated followed by a conclusion of it. The simulation campaign consists of optimisations regarding the state machines, buffer size dimensioning and data word considerations. For the partial sorting, the input data has been taken from the ATLAS ITk simulation at step 2.2. For the ATLASPix3 simulations, the simulations at step 3 have been used. The grouped readout design was fed with artificially enlarged charge distributions not from the ATLAS ITk simulation campaign. A more detailed description of the framework can be found in [Sch21b] and on the code repository with the attached documentation page at [Sch21c].

5.1 Structure of the Simulation

The simulation execution in ROME is structured into several parts: First, the signals to process through the readout architecture are generated or loaded. Then, the readout architecture is instantiated. After the instantiation, the actual data processing is done.

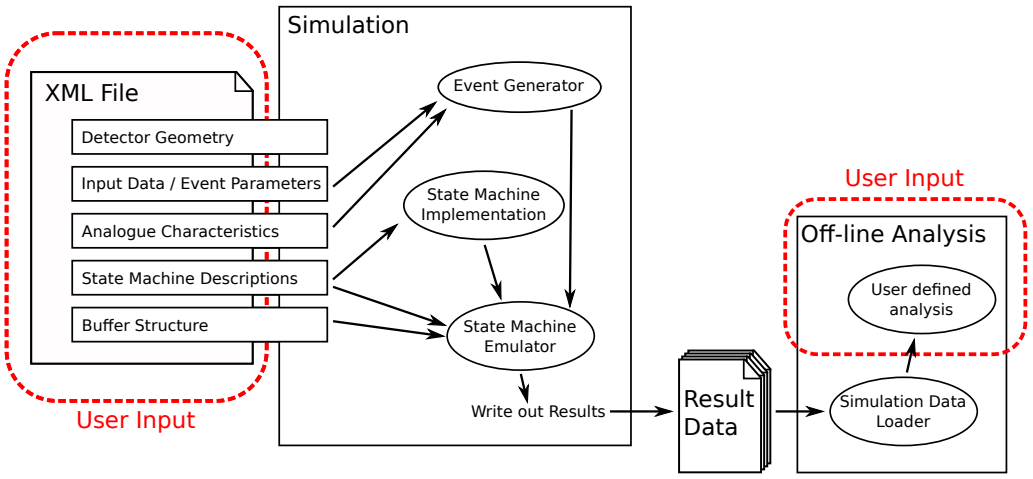


Figure 5.2: Data flow scheme of ROME: The user provides the description of the simulation to the framework which processes this data and generates output files with information on timing and path of the signals through the detector structure. This data can be analysed off-line by a user defined script.

A data-flow diagram of the simulation is shown in figure 5.2 with parts that require user input marked in red.

By design, user input is necessary only for the definition of the simulation and the analysis of the output data: With the configuration file, the geometry of the detector including analogue characteristics, the input data with generation information for the events and the readout structure of memory and state machines moving the information are defined. Typically, applying ROME to a new detector readout design does not include extending the code of the framework. The provided options for the implementation of the parts aim for versatility, so that extensions of the framework are necessary in the least number of cases reasonable. For processing the output data of the simulation, scripts are necessary that match the readout structure of the detector described in the configuration file. A generic analysis would always suffer from being incomplete and slow due to aspects covered that are not of interest. Therefore, user input is required to analyse the data with the help of generic analysis code, that can be applied for most analyses.

In the following, the conception of the simulation framework is sketched starting with the representation of the detector itself (section 5.1.1), over the treatment and house-keeping of data generated (section 5.1.2) and signal pattern generation (section 5.1.3). Furthermore, the structure of the simulation progress is described (section 5.1.4) before sketching the concepts of the user-input (section 5.1.5).

5.1.1 Detector Representation

The detector representation in ROME consists of several parts: Firstly, the array of pixels that contain information on their respective position, analogue characteristics and sensitivity. Secondly, the tree structures of the memory (buffers) for readout and thirdly, the state machines driving the transfer of the data through the memory structure.

These three parts represent the whole design of the detector: The first part covers parameters as matrix size and sensitive fraction and amplifier parameters. The second part describes the connections between memories down to how this memory treats signal arbitration or whether data is overwritten. The actual readout procedure is encoded in the readout state machines to transfer the signal data from the pixels to the output of the detector.

Matrix and Analogue Characteristics

The matrix consists of a number of pixels that are defined with their position and size, threshold and analogue characteristics.

The position given in the representation is the sensitive area of the pixel. If this area does not match the area physically occupied by the pixel, this is represented by gaps between the sensitive areas of neighbouring pixels.

From the amount of charge generated for the signal in a pixel, the digital signal representations are created. The detection threshold giving the minimum amount of charge to detect a signal can be defined individually for each pixel. The length and delay of this signal are defined by the dead time and time-walk characteristics, respectively. These characteristics are defined globally to represent the behaviour of the pixel amplifier-comparator combination (an example is given in section 3.22). Also, the detection efficiency of the pixels can be tuned individually as fraction of signals that are detected of all signals larger than the threshold. To mimic the differences in dead time between pixels on real sensors, the dead time for the signals can be scaled for each pixel individually from the value of the general characteristics.

Time-walk and dead time characteristics are provided as support points that are interpolated with cubic splines. Time-walk denotes the delay from generation of the charge in the pixel until the detection of it using the comparator. This means, that both the rise time of the amplifier and of the comparator are included in this delay. Dead time denotes the time the amplifier output stays above the threshold of the comparator preventing the detection of another signal. It is equal to the value obtained for a ToT measurement.

Detector Memory Structure

Each detector, buffer and pixel has its own address consisting of an address-part name and a value. The complete address of a structure builds up following the path through the memory tree to the root node.

Buffers represent the nodes of the memory tree structure and can have an arbitrary number of subordinate buffers and pixels. Each buffer has a depth representing the number of memory spaces inside it. Pixels are the leaves of the memory tree structures and are insertion points of signals. The root node of a memory tree is the detector and represents the output of the system.

The behaviour of the buffers can be configured individually for each buffer in three aspects: The readout from pixels, the readout from subordinate buffers (i.e. incoming data) and the management of the own memory spaces for outgoing data.

- The pixel readout can be the `or`-function of the pixels connected to the buffer or more complex logic between the pixels contained. For the latter case, there is a differentiation between contribution to the acceptance decision and contribution to the resulting pixel address.
- The readout from subordinate memory buffers can be configured for memory space usage and abort conditions (e.g. behaviour on a completely filled buffer).
- The data management method for readout to superordinate structures can be selected from a FIFO behaviour or a prioriy chain behaviour. The latter searches for the first occupied space to read, keeping the positions of the other data the same.

Readout State Machines

Each detector can have multiple state machines for readout of signals. The state machines can run at different frequencies and starting phases. For the state machines, also a pooled memory for integer values is implemented. The memory entries have a name and can be accessed by every state machine. Consequently, independent counters in the state machines are required to have unique names. Identical names can be used to transfer data between the state machines. The implementation supports both Mealy and Moore machines defining the procedures on the state transitions or on the states of the state machine description. Even a combination of both types is possible.

To transfer signal data from one buffer to another, the state machines can send transfer signals to pixels or buffers, specified with an address-part name. This signal is put into the detector and passed on through the buffer tree until the buffers with the selected address name are reached. On these buffers, the read operation is executed.

Simulating an ASIC on a CPU, the actions to execute are treated as simultaneous. This way, also parallel transmission of several datasets can be depicted as two read commands in the same state of the state machine.

Structure Abstraction

The simulation purposely does not include data bus limits or transmission timing. Data is treated as a collection of information without size measures and is treated as atomic object.

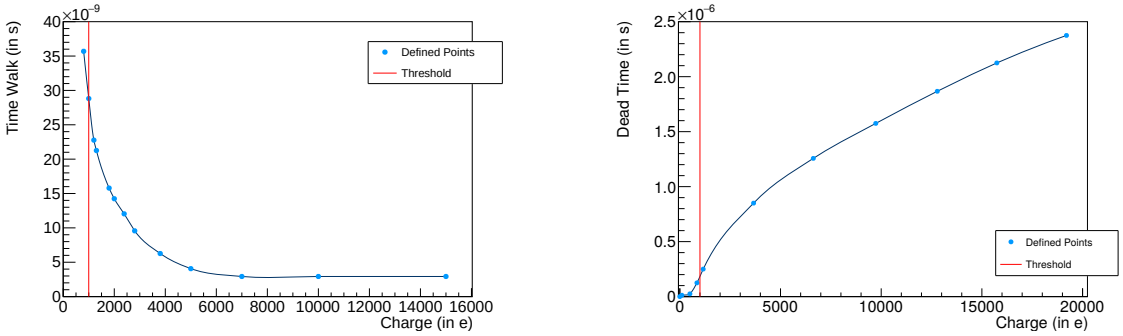
Moving the limitation of data bus width to the analysis, one simulation can be used to estimate the outcome for several widths: The information loss included in the limitation of any bus width can be exactly reproduced in off-line analysis of the output data. For example, the timestamp transmission in the proposed design of a detector is 6 bit wide. If a simulation with this limit results in the statement that the bus width is too small, it is likely that the simulation has to be rerun with larger bus width to find the necessary width. For the analysis-time restriction, this estimation can be done directly on the existing dataset. The same applies to addressing of the data and consequences for the output data format. Additional clock cycles occupied by the data due to extra bits can be mimicked with lower frequencies of the corresponding state machine or with dummy data added to the data stream as datasets are treated as atomic objects by the simulation.

For the goal to estimate the potential of different architectures, ROME is designed to not model signal transmission timing, simplifying the creation process of the architecture in the simulation. An extensive simulation including the delays requires knowledge on the complete implementation and layout which contradicts this paradigm.

5.1.2 Data Management

The simulation is designed to prepare data for off-line analysis. This requires logging every part of it. For this, the signals generated in the pixels log their own path through the memory structure. Also, loss of signals is logged. Together with the list of signals generated in the pixels, the performance of the detector architecture can be determined.

The Hit object representing a signal from a pixel contains not only the address of the pixel it is generated in (as collection of the name-value pairs for the trace through the buffer structure) and signal size, but also the resulting dead-time of the pixel and the timestamps for the arrival in each buffer. For buffers with a depth larger than one, also the index of the entry is saved. If a hit is lost – for example by overwriting it – also the time of this action is added as a timestamp.



(a) Time-Walk Characteristics (b) Dead Time Characteristics

Figure 5.3: For hit generation from charge clouds, the front-end characteristics for time-walk and dead time are extracted from cubic splines through the supplied points. The values from the splines are only applied if the signal surpasses the threshold of the pixel. The initial characteristics were measured on a CCPDv1 detector and then adapted to the simulated values to be expected for the pixel of ATLASPix3.

The Hit objects that are transferred to the detector object as root node of the memory tree are stored in one file. The Hit objects that are lost on the way are stored in a second file.

This leads to a collection of data traces through the detector architecture. This data can be mined to analyse the structure for bottlenecks or errors.

To indentify mismatched signals, the Hit object also contains an event identification number making it unique in combination with the pixel address. As a consequence, mismatching of signals can not only be recreated, but also reconstructed.

5.1.3 Event Generation

As first step of the simulation, the signals to process through the architecture have to be generated. The most simple source of this data can be a file defining the signals for the current detector. Alternatively, charge distributions can be used to generate the signal information.

For loading of predefined signals, it is necessary that the signals match the addressing scheme of the detector defined in the simulation.

For charge distributions, the geometry and analogue characteristics of the pixels are used. These analogue characteristics consist of a time-walk function and a dead time function (as explained in section 5.1.1). Both are defined via data points that are connected using cubic splines. Extrapolation of these functions will be used if the covered interval of the characteristics is smaller than the one provided by the charges used in the input data. Hence, the splines should be defined in a way, that the extrapolation is stable. The threshold defines a cut-off below which no signal will be generated. Example characteristics as used for the ATLASPix3 simulations in section 5.5 are shown in figure 5.3.

The charge distributions are mapped to the active volumes of the pixels and if the threshold is exceeded for a pixel, a signal is generated. From the charge, the time-walk and dead time are derived with the characteristics mentioned above and all three parameters are stored in the signal description.

For reuse and analysis, the signals generated can be written to a file. This ensures comparable results when testing variations of the same architecture against each other. In

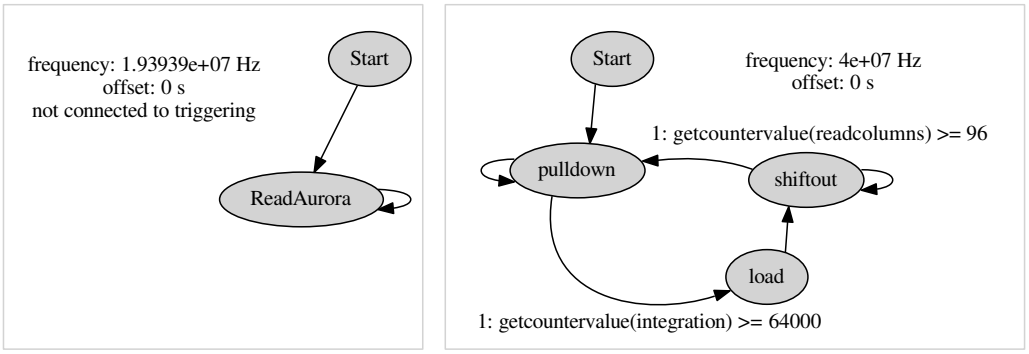


Figure 5.4: This is a visualisation created by the simulation framework from the state machine description in the configuration file. It consists of two independent state machines driving different parts of the readout. Each state machine is running at its own frequency with a well-defined starting phase. The event scheduler of the simulation will execute the actions defined for the state machines at the given times while inserting the hits into the pixel diodes.

addition, loading of the data from hard disk is faster than generation of the signal list from the charge distributions.

As source for the charge distributions, three options are available:

- Charge generation around lines in 3D space passing through the detector. Without tuning of the parameters of the charge distributions, this method can result in very different signal patterns than expected from particles.
- Simple charge distributions in ROOT[1] files. The content of these events in the file are reclustered to identify particles inside each event. These are separated and for usage, the separate particle signatures are combined to give the selected particle rate.
- Charge distributions in ROOT files that already contain pile-up, i.e. the correct number of simultaneous events as expected for the application of the sensor.

For all options, the similarity to real events has to be assured. However, such data is typically the result of physics simulations of particle interactions so that the task is to be solved in the generation software.

5.1.4 Simulation Engine

The core element of the software framework running the simulation is the Simulator class. It manages the detectors and signal insertion keeping track of the simulation time.

Time is treated as continuous variable in ROME to enable different frequencies for the state machines and event generation. To execute the events in correct order, the Simulator implements an event scheduling system sorting the events of the different simulation components: Each state machine hands the time of the next event to execute to the event scheduler for the detector. The event schedulers for the detectors and the insertion times for the signals and generation of readout trigger signals are sorted by a global event scheduler. With this structure, time sampling granularity is adapted to the needs of the simulation. At the event times, the Simulator calls the respective methods for the events registered. After executing the event, the time for the next event of the structure is calculated and returned to the event scheduler.

[1]ROOT data analysis framework developed at CERN [BR97, BRC+19], `https://root.cern.ch`

The benefit of this structure can be seen at the example state machines shown in figure 5.4: One state machine is running at a slightly lower frequency for the data encoding it mimics adding 2 bit on 64 bit of payload data. With a fixed base timestamp, the simulation would have to run at a much higher precision with most steps just containing waiting commands to make up for the frequency difference. Further optimisations as the hint for the simulation that a state machine is not affected by readout trigger generation can help reducing the execution time of the simulation. The visualisation itself is there to help the user verify that the implemented state machines match the intended ones and provides material for documentation of the system.

As a consequence of the implementation of time as a floating point number, the simulation time needs to be smaller than the time spanned by the precision of the variables containing the time. With the usage of the G++ compiler[2] and the floating point variables used at 64 bit width (double), the usable range of approximately 15 decimal digits results in a simulation time interval of about 1.5 weeks at a precision of 1 ns.

Instead of the simulated time, the limiting factor for the simulations is computational power: All operations of the readout have to be executed on every buffer down to the level at which the operation is targeted, making the process CPU-intensive. And synchronisation between threads working on the same detector requires additional effort. With the design as a testing environment, often parameters of the design are modified. The repetitions of the simulations for the different parameters are parallelised with each parameter set running on a single execution thread.

The runtime for a simulations depends on the complexity of the detector and its size. Furthermore, the rate at which the signals are inserted affects the runtime. For the design used for the development of the ATLASPix3 readout structure in section 5.5, the simulation of 1.2 ms took about $(1:16\pm0:14)$ h or (3.80 ± 0.68) s per simulated microsecond. These differences are due to different triggered fractions of the data altering the processing effort.

5.1.5 User-Input to the Simulation

As shown at the beginning of the section in figure 5.2, user input is necessary at two places in the simulation and evaluation procedure: Firstly, at the beginning, defining the structure to simulate and the data to transport through the described structure. Secondly, for the analysis of the data where only supportive structures can be provided.

Configuration File

For the definition of the simulation to execute, a single XML file is used. It contains the detector memory structure with the pixels as the leaves of the tree structure including the geometry of the detector. The structures inside the detector can be multiplied with a special tag that defines the number of repetitions and the shift at which its content is to be added. As second part of the detector, the description of the state machines is included. The state machine description includes the states and as subordinate tags, the detector operations to execute in the state and the state transitions. The state transitions contain conditions to go to the next state and optionally also detector operations to be executed. The different state transitions are checked in the order of occurrence in the configuration file reducing the complexity of the conditions necessary. Further contents of the file are the analogue characteristics for the pixels and the inputs for the signal data. If pre-generated signal data are used, the analogue characteristics is not required.

Important settings are also the output files and the output format: The processed signals are stored in a file for the lost signals and one for the signals successfully passed through

[2]Part of the GNU Compiler Collection (GCC); `https://gcc.gnu.org/`

the whole detector. In addition, the input data is collected. With the state machine visualisation, the data can be generated as single files or be compressed into a ZIP archive. The textual storage format of the signal data with each line decodable individually enables large reduction of data size with compression.

An important feature – applicable to all parts of the configuration file – is the tag to scan parameters. Every parameter of any tag can be modified generating a list of parameter sets and by that sub-simulations. Linking several scanned parameters at different places is also possible to enable modification of the output file names for parameter scans.

The simulation can be called with several configuration files at once. Bash pipelines can be combined with normal parameters in the function call. The former kind will be processed first if both are given.

Data Analysis

The simulation provides three lists of signal data: The signals created inside the detector structure, the successfully read out signals and the signals lost on the way. The latter two lists contain timestamp data for the arrival in each buffer, as well as additional information on the buffer position used in buffers with a depth larger than one or the reason and time for the signal loss. Also dummy hits are included in the output data as they were generated in the simulation. This comprehensive logging enables drawing conclusions on the performance of the simulated structure.

For the analysis in the ROOT data analysis framework, a class is provided that enables loading and filtering of the data generated in the simulation. If used with C++11 onwards, even direct access for data inside ZIP archives is possible. Also several plots can be directly generated from the data, as a histogram or a corresponding integration plot for the delay between two stages in the readout procedure.

However, due to the flexibility granted in the configuration of the architecture, it is not possible to provide a general extensive analysis. Especially the search for bottlenecks can require detailed knowledge of the architecture simulated. Therefore, the approach with a class is provided to handle the data in the output files of the simulation and prepare the user for writing the analysis scripts for the simulated architecture on his own.

5.2 Comparability of the Simulations to Beam Tests

At first, the comparison of this architecture simulation to beam test data seems straight forward: In both cases, there is a list of signals from the DUT and a list of reference data for where to expect signals from the DUT.

The data for the DUT represents the same data. However, the reference data represents a different kind of data: While the reference signal for the beam test is an approximation of the position where the particle traversed the DUT with a spatial and temporal uncertainty, for the simulation the exact pattern is known. To compensate for the uncertainties on the interaction of the particle with the DUT in the beam test, time and location offsets are accepted. Assuming square pixels with an uncertainty of one pixel length, for a track trough the middle of one pixel, nine pixels can satisfy the condition of a matched track. Also, the time uncertainty added can disguise time-walk effects as uncertainties from the reference system and the DUT can not be disentangled.

In the simulation, these uncertainties do not exist. Consequently, signal delays due to time-walk and missing pixel signals can be identified. In this situation, lost signals due to pile-up of signals in a pixel contribute to the reduction in efficiency and are multiplied

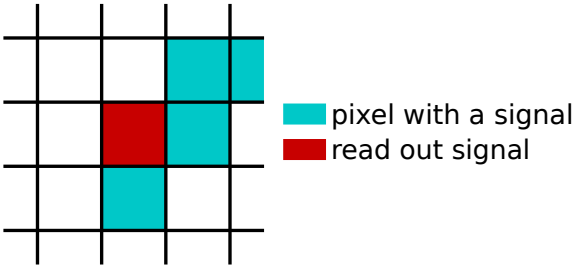


Figure 5.5: In contrary to beam test measurements where only reconstructed particle tracks going through the detector under test are known, in the simulation all detectable signals are known for the analysis. To harmonise these measures from beam tests and simulations, particle efficiency can be used to mimic the information from the experiment.

by the fact that a cluster from a single particle can result in several missing pixel signals, whereas a missing signal in the beam test analysis can only count as a single missing signal. This results in much lower efficiencies calculated from the simulation compared to beam tests.

The uncertainties in beam test measurements can not be recovered as the information is not available. On the other hand, the information loss necessary to match the simulation measure with the beam test is possible. For this, the pixel detection efficiency defined as fraction of pixel signals read out of the emplaced signals is replaced with the particle efficiency: The particle efficiency is defined as the number of clusters of which at least one signal is read out over the total number of clusters that were emplaced in the detector. This means that if one pixel signal even for a large cluster is read out, the particle is accounted for as found. This concept solves the difference due to the spatial uncertainty. The concept is visualised in figure 5.5.

The time uncertainty can be matched with the beam test by allowing for the signal to be in the time bin of emplacement or one of the following time bins depending on the timing constraints set for the beam test reference.

The actual delay possible for the signals is known for the simulation as it is the value from the time-walk characteristics. Hence, putting a longer acceptance window than the time-walk can only result in false-positive matches. But in contrary to beam tests, these mismatched signals can be found by comparing the Hit object identification numbers.

5.3 Grouped Readout Studies

With increasing granularity of pixel sensors, the amount of data to transmit increases. On one hand, addressing smaller pixels requires more bits for identification. On the other hand, smaller pixels lead to higher probabilities of signal clusters. This adds further to the amount of data to transmit for a particle signal detected on such a sensor device.

Typical signals generated in HV-CMOS sensors are still small compared to the achievable pixel dimensions of 25×25 µm^2. However, grouped readout of pixels for signals with increased charge distribution diameters to the same scale as the pixel edges is a test case for the simulation. In addition, it can visualise the interconnectedness of the whole system on comparably simple structures.

Clusters appear when a charge distribution is larger than one sensor segment or spans the area around a border between segments depositing sufficient charge in several segments to be detected by the respective comparators. As a consequence, the probability of clusters

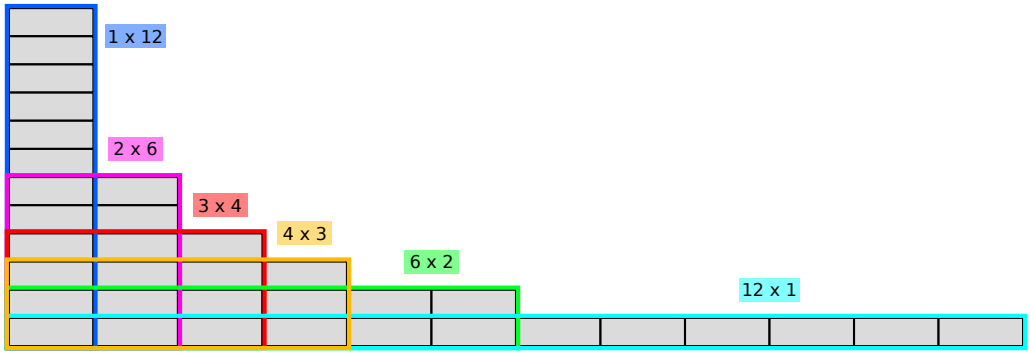


Figure 5.6: To form a rectangular group from 12 pixels at an aspect ratio of 3:1, six options exist. The option with two columns and 2×6 pixels results in a square group minimising the length of the group border.

increases with the length of the segment borders. Grouping segments on a pixel detector and reading them out in these groups, clusters inside them do not affect the readout, i.e. the readout is the same whether one element in the group detects a signal or all. Hence, only the outlines of the groups contribute to cluster generation for the readout. The benefit of such a grouped readout system is the reduced bandwidth from sending parts of the data only once per group. This means, that for several segments in a group with a signal, less data is transmitted compared to individual readout of these elements.

Such a system has been implemented on the ATLASPix_M2 and on the LFATLASPix integrated sensors as the parallel pixel-to-buffer (PPtB) readout [SBC+19, Pra20]: A group of pixels shares output lines for the comparator going to the PPtB buffer. This buffer contains four memory spaces. Each of them can store the timestamp of the event and a snapshot of the comparator lines. On a rising edge of one of the comparators, the timestamp is stored and – after a short delay to compensate for time-walk – the pattern of the comparator output lines is stored. This is done in the first empty memory block of the buffer. The readout of this data happens with a priority-chain, too. This means, that the temporal order of the signals stored in the four memory blocks is not necessarily maintained for readout. In this case, the timestamp, column address and with the group ID a part of the row index is transmitted for all pixels in the group combined. For single pixel events, more data is transmitted since the whole pattern for the group is transmitted, but starting from two pixels in the group, less data has to be transmitted. This does not only save output bandwidth, on-chip data processing is faster, too.

The performance of a readout structure depends on the geometry of these groups: Assuming pixels at a size of 120×40 µm^2 and groups of 12 pixels, there are six rectangular options for grouping as shown in figure 5.6. With the 2×6 pixel option resulting in a square group at 240×240 µm^2, only the three options around it (1×12, 2×6 and 3×4 pixels) will be considered in the following. From the geometric point of view, the square group option offers the largest saving in readout bandwidth as the least number of groups has to be read out what has been derived in [Rö18] under my supervision. The length of the border for the 2×6 pixel arrangement is 20 % shorter than for the 1×12 pixel shape and 8 % shorter than for the 3×4 pixel shape. Taking this numbers, the square option with 2×6 pixels can be expected to perform better than the other options.

Pairing these groups with four memory spaces per group, a column based readout and charge distributions enlarged in diameter by a factor of about 20, simulations were performed with ROME. For all signals successfully processed through the structure, the delay between emplacement in the detector and the readout was calculated, histogrammed and integrated for figure 5.7: For each group shape the data points mark the number of signals read out

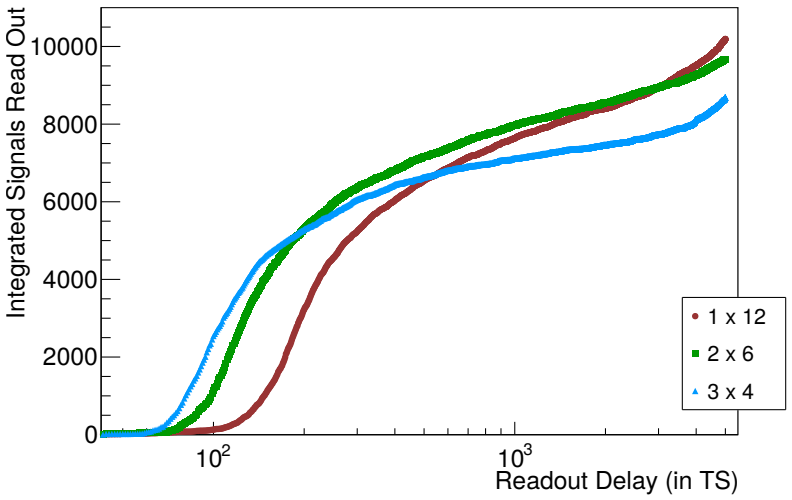


Figure 5.7: The readout delay distribution differs between the group shapes. Several crossover points show that the readout performance depends on more than the group shape. In addition, also the amount of memory and readout procedure affect the result.

with a delay as marked on the X-axis or faster, resulting in monotonically increasing curves. The higher the value at a given delay, the more signals can be processed by the structure. These data rows show several crossover points indicating to a more complex structure behind them than the geometric shape of the groups.

To explain this behaviour, it is necessary to have a closer look at the simulated structures: The first difference is that the different widths of the groups result in an altered number of readout columns. Also, the number of groups in such a readout column changes by a factor of three as the geometric height of the groups changes from 480 µm to 160 µm. The effect of this change is connected to the readout scheme used: The memory in each readout column is arranged in a priority chain where the first filled entry is read. At the end of each readout column, one memory space is filled with this entry. After filling this end-of-column (EoC) memory with data, the valid data is read one at a time from these buffers clearing them again.

After all EoC memory buffers are cleared, the next set of signals is transferred from the memory of the groups to the EoC memory. Consequently, reading two entries from the same readout column is slower than from two different columns. With the round clusters used for the simulation, wider groups read more data at the beginning, reading the data from the edge of the charge distribution (a visualisation is shown in figure 5.8). The longer the cluster is in the direction of the columns, the more groups have to be read out. longer groups have an advantage for this, since less groups from the same readout column have to be read. If the detector is processing many signals at the same time and all readout columns contain data to read, also the architecture with more readout columns will read more data as the loading of new data to the EoC memory will take up a smaller fraction of time.

The result of this is the expectation that a wider group reads a larger number of signals at a short delay, but the additional loading cycles will result in a smaller efficiency for much data contained because of the smaller number of EoC memory buffers filled at once. With this, the behaviour from the figure can be explained.

In addition, at the very beginning of the increase of the curves, the readout procedure can

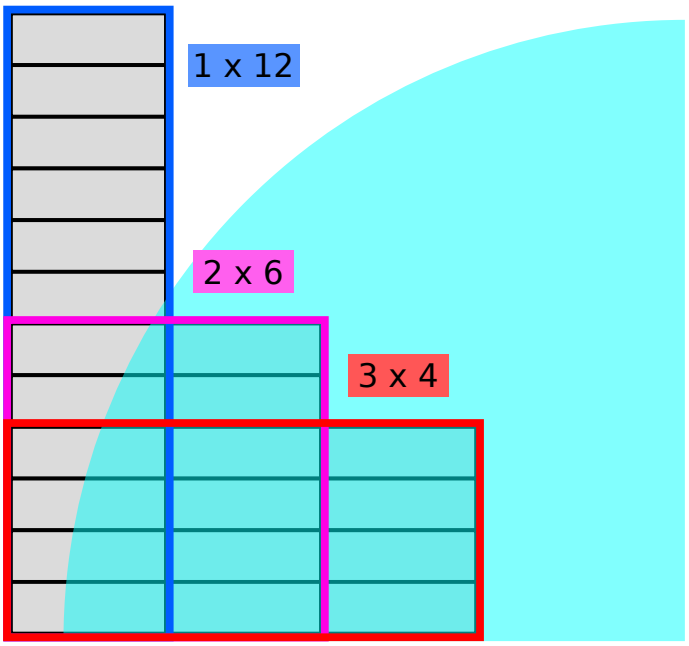


Figure 5.8: For charge clusters with a size larger than the long edge of the pixels, the shape of the cluster has an influence on the readout speed: The first groups read from a cluster contain more pixels if the group is wider, while in the later read cycles, more signals are collected in longer groups.

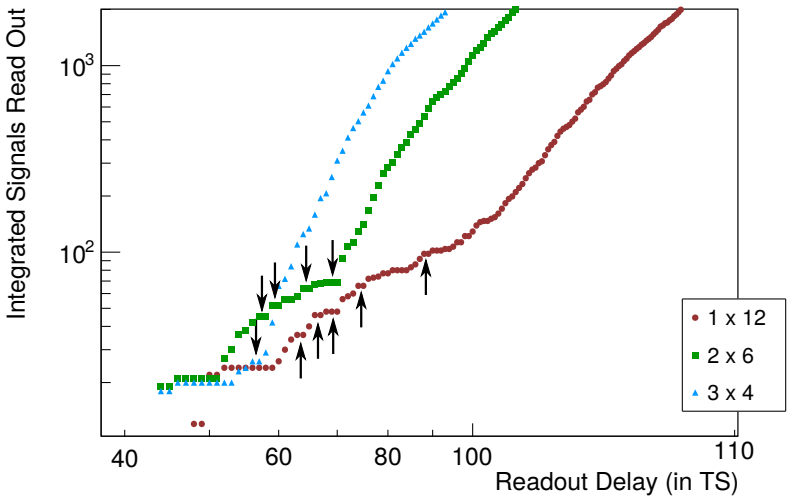


Figure 5.9: The beginning of the readout delay curves directly shows the influence of the readout procedure: During the loading of data in the end-of-column (EoC) buffers, no data is read out leading to a reduced slope for the delay curve. This wave structure becomes less visible for larger delays as the distribution of the loading cycles averages out between the delays. Several delays at which no data has been transferred are marked with vertical arrows.

be seen directly as change of the slope (see figure 5.9): While new datasets are loaded into the EoC memory, no signals are sent off the detector and as a consequence, the delay curve will have a smaller slope for this delay. For short delays, this effect is visible, because the number of cycles spent is small. For large readout delays however, many other signals are read out before the signal with long delay. Consequently, there are many combinations of signal writing and loading cycles leading to this delay averaging out for longer delays.

On the sensor, structures are kept as simple as possible to reduce the amount of possible error sources and to keep the insensitive fraction of the area as small as possible. For this reason, the different group shapes would likely result in a changed number of readout columns as it was implemented for the simulations. This example shows, that this adaption has side effects on the detector performance stressing the necessity of a detailed analysis of the architecture as a whole. In this case, even the effect of interest created by the different group shapes can not be studied because of the other adaptions of the system to accommodate the group shape. In combination with the input data, the modifications cover up the effects that were to be studied.

5.4 On-Chip Signal Sorting

The sensors developed at KIT-ADL use priority-chains for selecting signals for readout. This results in a spatial sorting and only if the signals are sparse enough they are in temporal order. For analysis, data needs to be sorted temporally to ensure that all of it is taken into account. On-line analyses as performed at LHC depend on this for efficient data reduction and fast processing.

Data sorting on FPGAs requires much memory space. Especially, the exact amount of memory needed for the task depends on the particles passing through the detector. As a consequence, sorting of signal data takes up much space on the readout electronics and consumes power. To overcome this problem, readout system design concepts envisage data sorting directly on the detector readout ASICs. In case of monolithic HV-CMOS detectors, this is the sensor itself. Despite of access to more information on the data present directly on the ASIC, the task of sorting the data temporally does not become trivial. It involves keeping track of the timestamps available on the chip. This takes either space on the ASIC or time in the readout procedure. If the response latency of the detector does not have to be as short as possible, the latter approach is typically favoured to avoid insensitive area in the detector system.

Combining the sorted readout with a trigger signal to read out only a fraction of the timestamps, this can be implemented using a FIFO: The triggered timestamp values are written to the FIFO and the readout electronics take one entry at a time and transfer all signals with this timestamp off the chip. The mentioned drawbacks of this are that for each triggered timestamp the value has to be fetched individually from the FIFO before the data can be transferred. On the other hand, the limited memory issue from the FPGA can occur, too: If too many timestamps are triggered shortly after each other, the FIFO can run full and timestamp values are lost.

For trigger and signal rates sufficiently high that several signals are ready for readout (i.e. the trigger for the second timestamp arrived before the data from the first one were processed), also the efficiency of the readout can be affected. This can be visualised with a simple example: For a readout structure similar to the ones from the previous section with column-based data collection to EoC buffers that are read out, two pixels in the same column detect a signal for the first event at time t_1. The next event at timestamp $t_2 = t_1 + 1$ generates also two signals in two pixels of another column. Without time sorting, the first signal of each event would be loaded to the EoC buffers simultaneously and read out. The

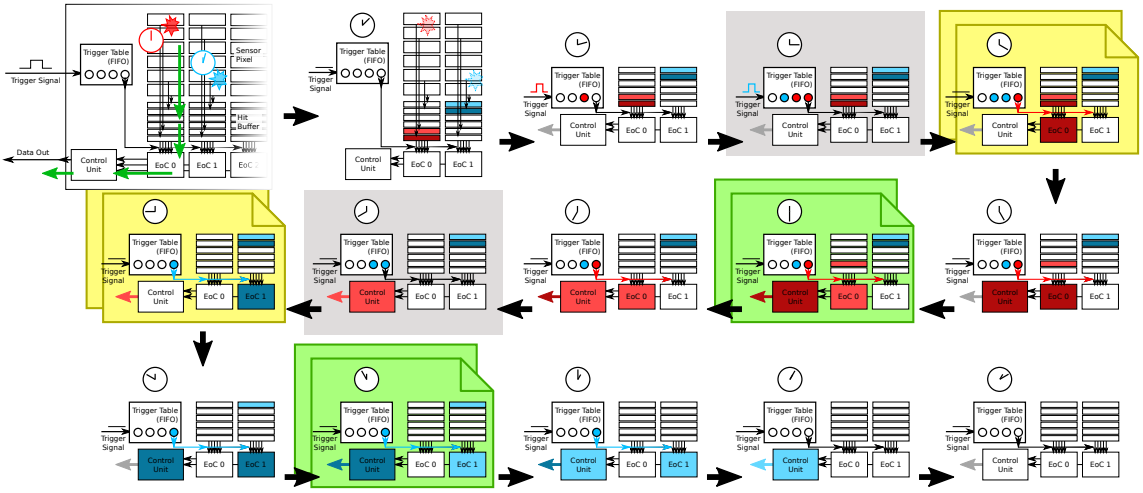


Figure 5.10: For two events close after each other, a two-pixel cluster is generated each. For event sorted readout with triggering, the full procedure is necessary. For unsorted readout, the event loading steps at 3:00 and 8:00 (shaded in grey) are omitted and the steps 4:00 and 9:00, as well as 6:00 and 11:00 (marked with a copy background) are combined taking a step less for each pair.

same would happen for the other two signals. In total, this includes six moving operations. With sorting for events however, first the event timestamp has to be selected before only the first event at t_1 is loaded with four moving operations. The same applies to the second event at t_2: four operations are necessary to transfer the signal data to the output of the ASIC. All in all, the sorted readout of these two events requires ten operations for the same data. This is a 40 % increase of readout effort compared to the unsorted procedure. This difference can be even larger depending on the readout procedure of the detector because of intermediate steps between the moving operations. The processing scheme is visualised in figure 5.10 for the sorted readout. The extra steps compared to unsorted readout are marked.

The readout efficiency reduction can cause the whole architecture to fall below the requirements for the expected particle rates in the experiment. To overcome this problem, a split approach between sorting on the sensor and the FPGA has been evaluated with ROME: If the sensor assures that data is presorted to larger time intervals, the FPGA only has to sort the content of these intervals in a divide and conquer approach, as the best comparison based sorting algorithms scale at $\mathcal{O}(n \log n)$ meaning that dividing sorting into two sub-tasks of half the size reduces the effort by about 30 % for any side.

For low trigger frequencies, event data will be transmitted individually. Only for high trigger rates, this feature is necessary to increase the readout efficiency of the sensor. Consequently, the sorting effort is best distributed if the lowest bits of the timestamp are left unsorted for the readout. For the simulation, this is done by ignoring the values of the lowest bits of the timestamp for the selection in the readout. This explicitly does not affect the actual selection of the signals with the trigger, only the readout is altered.

For comparison, complete sorting, skipping the least-significant bit (LSb; “mask 1”), the two LSbs (“mask 3”) and the three LSbs (“mask 7”) are simulated. The readout delay integral curves for these options are shown in figure 5.11: After the trigger delay of 800 timestamps (TS)[3], the number of read out signals rises faster with smaller sorting effort as expected. The reduction of the slopes results from the data put into the detector: If an event with many signals is being processed while another triggered event becomes ready for readout,

[3] A timestamp is 25 ns long which equals the event spacing

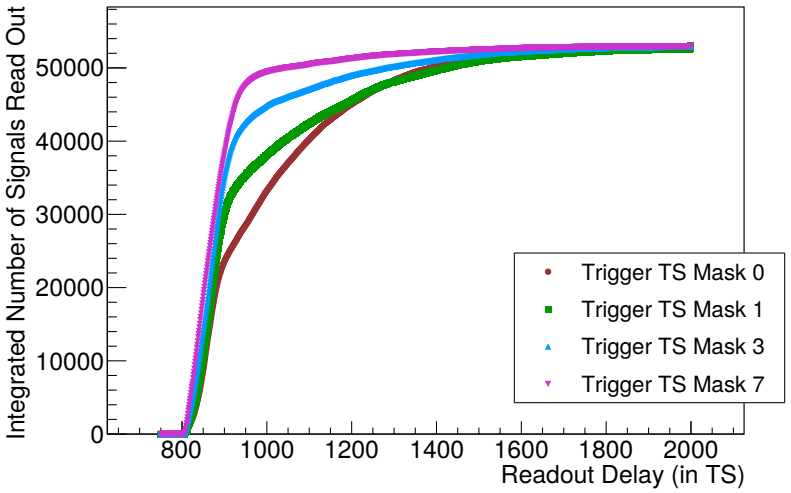


Figure 5.11: The bits ignored for the sorting of the data are shown as decimal representation in the legend: complete sorting has no masked bits (0), skipping the least-significant bit for sorting equals a decimal one, two bits are a three and three bits left out from sorting results in a seven. The smaller the sorting effort, the more efficient is the data transfer off the sensor – which is visible as larger value for a given delay.

these signals are delayed by the readout of the signal data of the previous event. At a readout delay of about 1300 TS (or 500 TS after the trigger delay), the completely sorted readout surpasses the readout of sorting with one masked bit. This is because the long delays are dominated by large events: If two events are read out combined it is possible that a signal from the first event at a position at the end of the priority chain is read after signals for the later event which are closer to the front of the priority chain. The combined readout consequently results in a longer delay for one signal compared to the separate readout of the events.

The long delays are a sign that the readout is not capable of reliably sending the data off the sensor. Assuming a 10 bit timestamp, the value of it becomes non-unique after $(1023 + 800)$ TS $= 1823$ TS. This means that the data can not be mapped to one event in all cases. The larger readout delay compared to the timestamp interval results from the fact that the trigger delay is the same for all signals and consequently is just an offset not interfering with uniqueness of the timestamp. The choice of the trigger rate slightly above the capabilities of the readout architecture is on purpose as the differences become larger towards the threshold rate.

In this case, the change between the variants of the detector was isolated allowing for comparison of the behaviour. The expectation of the reduced overhead from the smaller number of individually selected timestamps matches the simulation. The readout efficiency is reduced less for the smaller sorting efforts and consequently, higher data rates are possible. The cross-over point between complete sorting and sorting apart from the LSb shows again, that details, corner-cases or side-effects are crucial to be studied for building sensors at cutting-edge level for applications as in the ATLAS ITk upgrade, the Mighty Tracker upgrade of LHCb or for the Mu3e experiment.

5.5 Development of the ATLASPix3 Readout Structure

The ATLASPix integrated sensors are a development line with the goal to create a sensor device that can be used in a particle tracker. One possible application at the beginning of

this development was the outermost pixel layer of the inner tracker of the ATLAS upgrade for HL-LHC. The design of the sensors is built on the specification for this environment. While the goal of ATLASPix1 was to show that monolithic active pixel sensor (MAPS) devices work at the largest scale possible for the process – the reticle at about 2×2 cm^2, ATLASPix2 was testing means to improve time resolution. ATLASPix1 was 19.5 mm high, but the width of 10.5 mm was divided into three independent matrices. ATLASPix2 was a small ASIC at 4.2×3.7 mm^2. The third iteration – ATLASPix3 – is supposed to fill the whole reticle without subdivisions. Furthermore, this large detector is supposed to be suited for building modules. This means that bond connections are only possible on one side avoiding insensitive area between the samples placed next to each other. The data for the signals to write out are required to assign signals to the correct bunch-crossing at 40 MHz and to reduce the data rate, the modules receive a trigger signal at a fixed delay after the event to decide whether the data is to be read out or discarded [ATL17].

From the physics simulation data for the fifth layer in the inner tracker of the ATLAS experiment at HL-LHC, it becomes clear, that the achievable pixel sizes for HV-CMOS sensors are large compared to the generated charge clouds and signal clusters are rare. Measurements on previous sensors as the ATLASPix1 or MuPix8 show the same spatially small signals resulting in over 90 % of the signals being single-pixel events [PPA+19].

Consequently, the design of ATLASPix3 does not feature the grouped readout as described in section 5.3. Instead, an individual-pixel column-drain readout approach is chosen: Each pixel is connected to one buffer. The buffers in each column are read to the end of the respective column and from there, the data is collected from the different columns. The readout order of the buffers in the column is controlled by a priority chain. For triggered readout, the end of the column is modified: Instead of transferring data directly to an EoC buffer for readout, the data is transferred to column-assigned memory where the data is stored until arrival of the delayed trigger signal. From this column-assigned memory, the data is moved to the EoC buffer for readout after the trigger arrived. This is to keep the pixel active during the time from signal detection to trigger arrival. Furthermore, the data read out from the sensors are supposed to be sorted by the events they belong to.

The signal data are transferred from the memory connected to the pixel to this column-assigned memory after the signal finishes. The trigger signal is binary information received per time bin at a fixed delay after the event. After the trigger delay, the data are either marked as triggered or are deleted depending on whether the trigger signal was received or not. The trigger signal also creates an entry in a FIFO from which one entry at a time is read to select data that belongs to it for readout. This FIFO will be denoted as trigger table.

The output data are chosen to be encoded with a 64 bit to 66 bit Aurora encoding according to the requirements for the ATLAS ITk upgrade [ATL17]. This does not match the timing of timestamping. This speed difference has to be incorporated in the simulation as well. The transition is implemented using a FIFO between the output from the buffer tree to the encoder. This FIFO will be referred to as Aurora FIFO as for the encoding used.

As the previous ATLASPix integrated sensors, ATLASPix3 is supposed to feature an asynchronous matrix. This means that no timing signal is transmitted to the matrix, but all timing signals are constrained to the periphery. This requires an individual connection from every pixel to the corresponding buffer that assigns the timestamp to the signal. This is beneficial for reducing the noise level in the matrix, but – with the specification of the design process in line width and number of metal layers – puts constraints on the number of pixels possible.

From these constraints and experimental considerations, a pixel size of 130×50 µm^2 was envisaged for the design studies. The width of the pixel is adopted from ATLASPix1, the

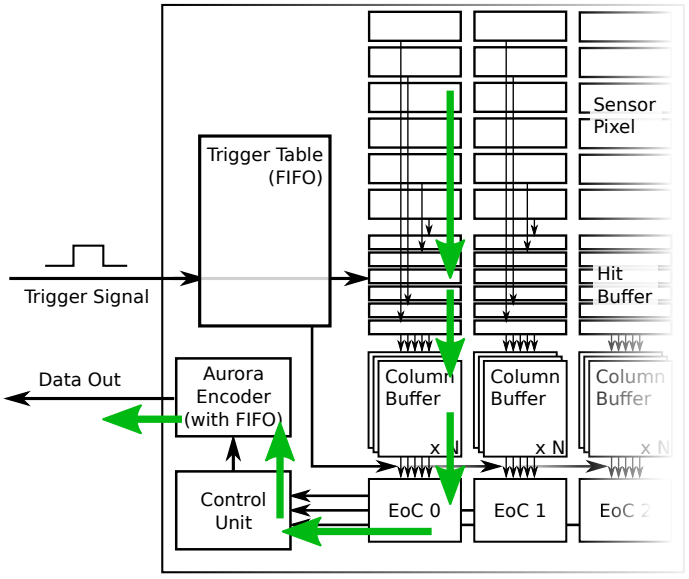


Figure 5.12: The design structure of ATLASPix3 for the ROME simulations consists of the pixels with associated hit buffers, the column memory for awaiting the trigger signal, EoC buffers for sorting and FIFOs for the trigger signal IDs and the data for the encoder.

height a result from the design process constraints. To fill the reticle leaving space for the periphery, the matrix size for the simulations is chosen to be 154×372 pixels. The other aspects as the number of memory spaces in the column memory, the state machines driving the readout process or the size of the trigger table are to be optimised in the simulations. The delay and rate of the trigger has been tailored to the plans and numbers used by the groups working on HV-CMOS sensors for the ATLAS ITk upgrade in 2018. This is a two-level trigger scheme. The first trigger level (level 0) has a delay of up to 12.5 µs at a rate of 4 MHz. The second trigger level (level 1) has a delay of up to 25 µs at a rate of about 1 MHz. The HV-CMOS detector has been envisaged to use the level-1 trigger at 25 µs of delay. The delay of 35 µs specified in [ATL17] at a different section is not considered for this study and would alter the outcome.

A sketch of the resulting structure for the simulation is shown in figure 5.12. The spatial separation of the pixels and hit buffers is not part of ROME but a hint to the planned implementation. The control unit is the equivalent to the EoC buffers for the readout of the columns. The design on the ASIC for the control unit includes the state machines driving the readout, too. This placement of the state machine circuitry is not part of ROME, either, but the name is kept for consistency. The placement and size of the trigger table is not representative, either.

In the following, the considerations for the state machine, sizes of the memories and the data format used for the readout are described, starting with the state machine adjustments in section 5.5.1. Then, the memory per column, the trigger timestamp FIFO and the depth of the Aurora FIFO in front of the encoder is discussed in section 5.5.2. As third point, the data format for the output is discussed in section 5.5.3. It is not directly a part of the simulation, however, the data format has implications on the simulation which have to be modelled. All three parts rely on the optimisations of the other two to work. Therefore, the sections highlight the different aspects and complement each other to give a complete picture.

5.5.1 State Machine Optimisation

The readout of ATLASPix3 is controlled by one or several state machines. They drive the signals controlling data transmission between the different buffers and finally to the data encoder and sending the data off the ASIC.

The two extra bits of the data encoding prevent synchronous operation of all parts of the readout at frequencies lower than half the bit transmission frequency. For driving large networks as for transferring data between buffers, the necessary frequencies for the bit transmission are too fast for the state machine. As a consequence, the state machine is split between the matrix readout part running at a base frequency of 40 MHz and the encoder part which is running at a fraction of 64/66 of half the frequency at 19.4 MHz. The part driving the encoder will be referred to as encoder state machine, the other part as matrix readout state machine.

On previous sensors, the data for the decoder was provided directly and consequently, data for sending had to be provided for every work cycle of the encoder. This data was used for sending debug data or synchronisation words, too. But with the Aurora FIFO on ATLASPix3, it is possible to not send data in some states of the readout state machine of the matrix. On an empty FIFO, extra data still has to be provided. The difference is that the amount of debug data adapts to the signal data to send: With more signal data, the number of debug data words is reduced potentially improving the efficiency of the readout.

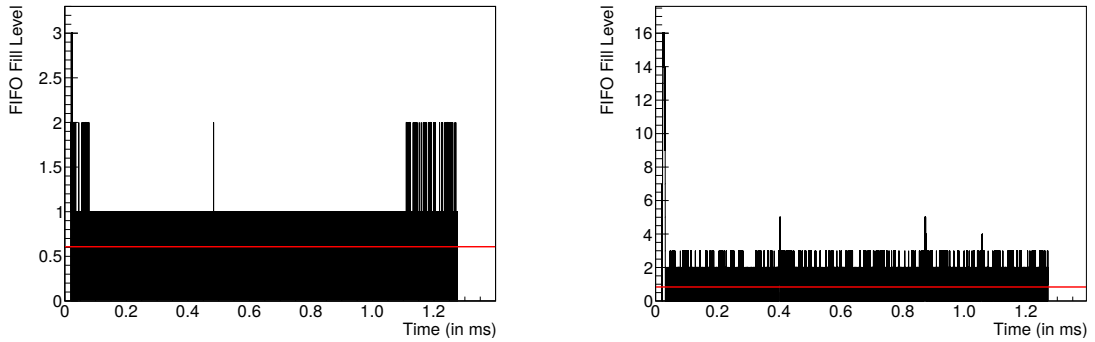
The state machines on the predecessors of ATLASPix3 were running at lower speeds providing 32 bits of data per clock cycle to the encoder which was sending them out during four time bins of the timestamp. With the input width of 64 bit for the encoder on ATLASPix3, this data is not filling up the data words. Simulations confirm what can be read from the flow chart of a state machine that is not sending data in every state: If the matrix readout state machine is running at the 66/64 of the encoder state machine speed, the Aurora FIFO will be empty for a large fraction of the time and debug data is sent.

Since some of the states, which are not generating data, are necessary for stable operation of the readout, the matrix readout state machine can be run at twice the speed to compensate for the unused bandwidth. If data is provided by the matrix readout state machine on every second state and it is running at 80 MHz while the encoder state machine is running at 19.4 MHz, it is mathematically possible to write more data into the Aurora FIFO than is read out leading to data loss at the FIFO. For this reason, the matrix readout state machine is suspended if the Aurora FIFO is full. In the simulation, this is mimicked by an extra state in which no data is written to the Aurora FIFO.

In practice, the state machine will generate data when loading data from the EoC buffers to the Aurora FIFO, not while loading data into the EoC buffers. This reduces the effective number of datasets transmitted per time. For a suitable combination of Aurora FIFO depth and number of columns, this single empty cycle is sufficient to compensate for the two extra bits transmitted by the encoder so that the FIFO will not loose data. In addition, not all columns of the matrix will contain data for a single event in the intended operation environment.

The fill state of the Aurora FIFO over the simulation time is shown in figure 5.13 for 40 MHz and 80 MHz for the matrix readout state machine. An average fill state below the number of entries processed per cycle of the encoder means that the encoder is not used to its full capacity for data. The data being processed for the two plots are the same which means that in the case of 40 MHz, not all data is processed and signals are lost earlier in the signal chain.

Actually, the matrix readout state machine itself consists of two parts: One part that is transferring the signal data from the buffers connected to the pixels to the column-assigned



(a) 40 MHz (b) 80 MHz

Figure 5.13: The fill state over time for an Aurora FIFO with a depth of 16 entries is shown for a frequency of 40 MHz and 80 MHz for the matrix readout state machine. The horizontal red line indicates the average fill state of the FIFO. At an average fill below the number of entries read per cycle of the encoder, reading is faster than filling.

memory for trigger discrimination (pixel readout). The other part is selecting the signals that belong to one event and transfers them to the Aurora FIFO (column readout). This is for two reasons: Firstly, data transfer of the signals to the column buffer needs to be faster than the trigger delay or they will not be deleted as deletion happens at a fixed time after signal detection with the absence of the trigger signal for this time. Hence, an entry in the column memory is blocked. Secondly, the data throughput from the column buffers to the Aurora FIFO and from the pixel associated memory to the column memory would be reduced if they shared a state machine.

The part for column readout also needs to be efficient when processing empty events, i.e. events that are triggered but no signals were acquired for. This is because this process is taking time from processing of events with a large number of signals. The target is consequently a short loop from loading a timestamp from the trigger table until reaching this state again.

The three state machines driving the readout are visualised in figure 5.14. The three-step procedure of the pixel readout state machine does not match the three plus n step procedure of the column readout state machine. Consequently, their separation enables simpler state machines resulting in better readout efficiency. The WaitFIFOSpace state in the column readout state machine is the representation used for ROME to mimic the suspension of the state machine on a full Aurora FIFO. In the LoadColumn state, the data from the column memory is transferred to the EoC buffers and with the ReadColumn state, the data from the EoC buffers is written to the Aurora FIFO. The ReadFIFO1 state loads a new entry from the trigger table. The second state is for the comparators in the column memory to select the signal data belonging to this event timestamp.

In the pixel readout state machine, the LdPix ("load pixel") state prepares the available signal data for transfer to the column memory. In the RdPix ("read pixel") state, the data is transferred. The PullDown state is for separation of the states to prevent data loss. Despite simulating the readout architecture detached from the actual implementation, some aspects still have to be taken into account to receive a design that can be implemented and the PullDown state is such an example.

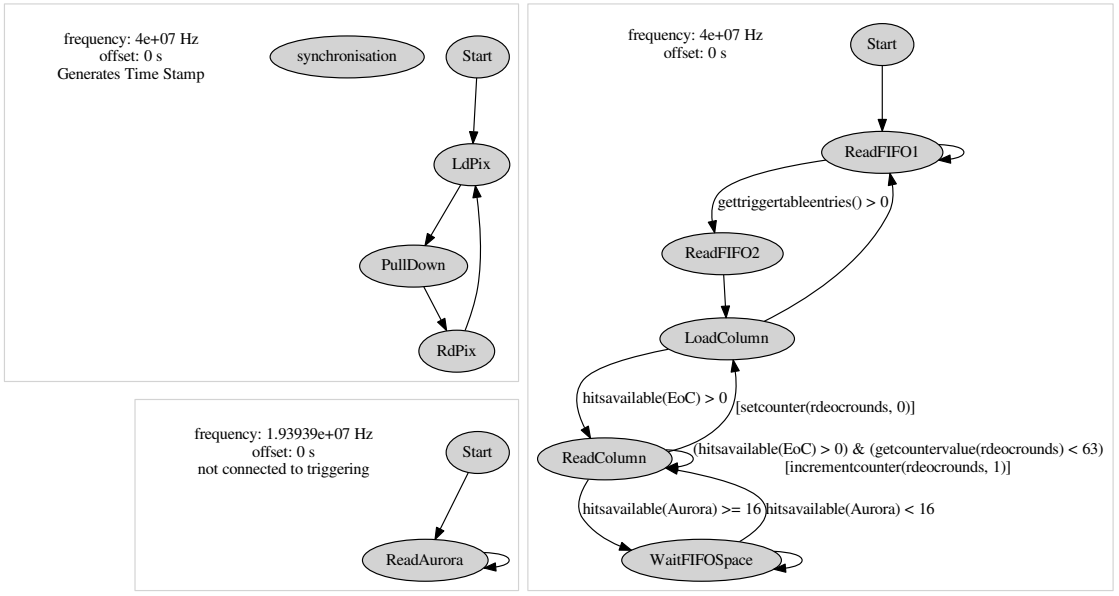


Figure 5.14: The three state machines for the readout of ATLASPix3 are separated by the boxes: At the top left, the pixel readout state machine, at the right the column readout state machine and on the bottom left the slower encoder state machine.

5.5.2 Buffer Dimensioning

The design contains three types of memory entities of which the size is a design parameter that can be chosen: The trigger table, the column memory and the Aurora FIFO. The second one exists once per column while the other two exist only once in the design.

They will be covered in the reverse order of their position in the readout structure: The size of the Aurora FIFO will be addressed first, followed by size considerations for the column memory, and lastly, the trigger table size implications will be discussed.

Aurora FIFO

The necessary size of the Aurora FIFO depends on the input data rate fluctuations that are to be compensated for. In the simulated design, these rate fluctuation sources are:

- the different frequencies of the column readout state machine and the encoder state machine,
- the design of the column readout state machine to not write data to the Aurora FIFO in every state and
- the trigger rate and the number of signals per triggered event, as well as the arrangement in the matrix.

The different state machine speeds result in one missing read cycle in 33 cycles of the column readout state machine. Consequently, the maximum number of cycles of the column readout state machine between states that do not write data to the FIFO gives the number of additional buffer spaces necessary after division by 33: For the 154 columns of the matrix, the ReadColumn state can be repeated 154 times if the number of repetitions is not limited lower – assuming the worst case that all columns contain data for the event. In figure 5.14, this limit is 63. For this value, the number of additional spaces is 2. For the whole matrix, this number is 5. This value is to be rescaled by the width ratio of the inputs and outputs

of the Aurora FIFO. Assuming for example 32 bit data words, two of them can be encoded at once in the 64 bit payload of the encoding meaning that the number of entries in the FIFO needs to be doubled. The assumption for this is that the ReadColumn state does not generate data for the Aurora FIFO.

The phase between the state machines makes it necessary to provide space for another two sets of entries. With the entry for the current dataset, this is a total of five datasets. In the example of 32 bit data words, this is then a FIFO with 10 entries of which at most 9 can be used.

One point on the list reduces the amount of data being written to the Aurora FIFO and does not require additional spaces: If no events are triggered, the column readout state machine stays in the ReadFIFO1 state not writing data. For many events with small numbers of signals per event, the fraction of cycles in the ReadFIFO states not writing data becomes also larger compared to less events with the same total number of signals.

The arrangement of the signals in the matrix can have an impact on the readout speed as well: This becomes obvious for the example of a delta-electron in the detector plane: If the track is in the direction perpendicular to the column, at most two pixels per column detect a signal from it, but pixels in many columns do so. Consequently, the LoadColumn state occurs only two times in the event. For an orientation of the delta-electron trajectory parallel to the column, many pixels in very few columns detect a signal. This means that the LoadColumn state occurs every three or four cycles.

If data is not transmitted every time in the LoadColumn state, this reduces the amount of data written. However, with data being sent in this state, the amount of data increases with arrangement of events parallel to the column direction. Assuming a delta-electron as object with one of the largest extents in the direction of the column, the amount of memory necessary can be calculated: If the delta-electron track spans the whole height of the matrix at the border between two columns, this is 744 pixels to detect a signal. Consequently, this requires 23 spaces in the Aurora FIFO just for the different state machine speeds.

However, this worst case scenario is highly unlikely in both angle and position: To cover the length of the whole matrix inside the depletion layer at a thickness of about 50 µm, the angle deviation from the detector plane has to be smaller than 0.15°. Furthermore, the typical width of a charge track in the order of a few micrometres reduces the phase space for such a track.

Consequently, for normal operation a FIFO depth of 16 entries in combination with the suspending of the column readout state machine on a full Aurora FIFO enable lossless operation at more reasonable circuit complexity. This size allows for the readout of about 500 data words before the column readout state machine has to be suspended. Loading the next triggered event timestamp decreases the FIFO fill state again resulting in unsuspended operation of the state machines in most cases.

For the case of 32 bit data words with two encoded at once and a trigger rate of 4 MHz the fill state over time is shown in figure 5.15. The data output at two data words per encoder cycle was almost completely exploited with an average fill state of the Aurora FIFO at 1.99 entries. But despite this, the column readout state machine was not suspended as the maximum capacity of the FIFO at 16 entries was not used. The 60 % security factor has to account for events not depicted in the 50 000 simulated events.

Column Memory

The description of the distribution of the content of the column buffers over time is complicated as it consists of non-constant input and output rates and a subdivision in 154

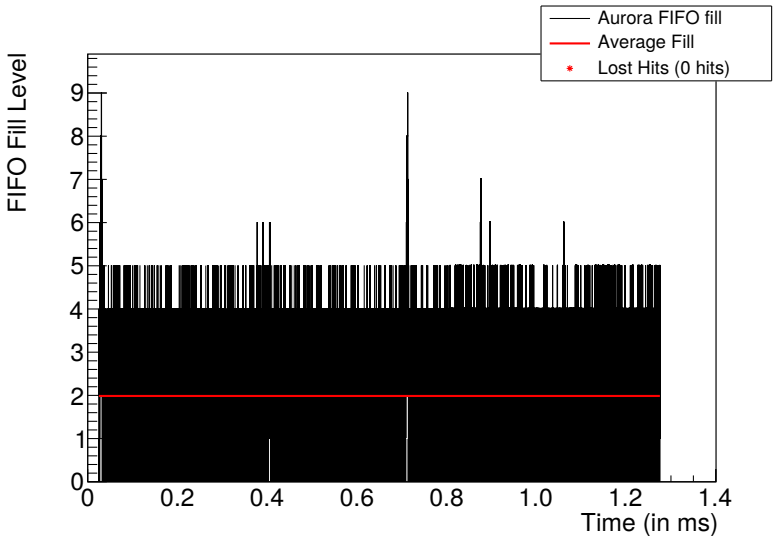


Figure 5.15: The fill state of the Aurora FIFO does not reach the maximum number of entries (16) provided in the simulation at any time. This is despite a triggered fraction of 10 % of all events. The average fill state is at 1.99 entries meaning that the output capacity was almost completely exploited throughout the simulation.

parts that can be empty individually. Consequently, a closed-form description is replaced by Monte-Carlo simulations. The fill state is not logged directly by the simulation, but from the traces of the signals through the structure, both in successfully read out and lost or discarded signal lists, the fill state can be reconstructed. This fill state is histogrammed for each column individually and is shown in figure 5.16 for one parameter setting. In this case, it is a trigger delay of 25 µs and a trigger rate of 1 MHz or a fraction of 2.5 % of the events. The number of buffers was chosen large (120 buffers per column) to be able to obtain information on the necessary size without running simulations for many different settings.

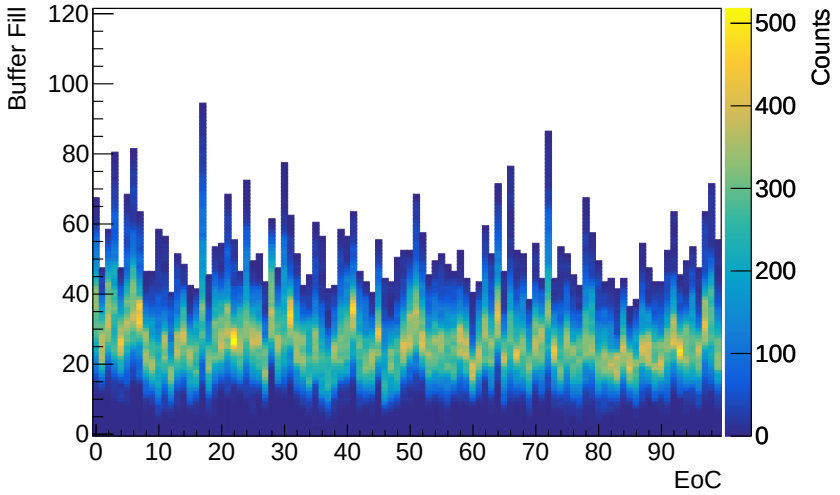
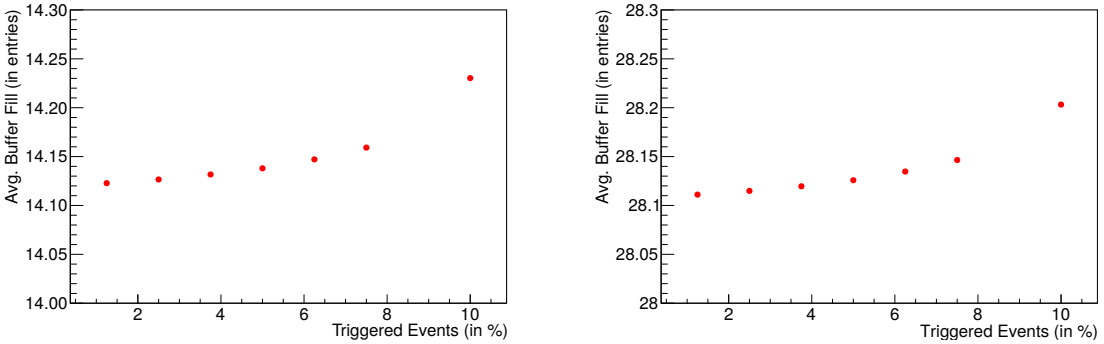


Figure 5.16: The distributions for the single columns are all correlated as more triggered data in one column delays the readout of all other columns as well. The average of the means per column is 26.8 entries.



(a) Trigger delay: 12.5 µs

(b) Trigger delay: 25.0 µs

Figure 5.17: With twice the delay for the trigger signal, the average fill state of the column buffers doubles. However, the change of the triggered fraction by a factor of four from 2.5 % to 10.0 % changes the average fill state by less than 0.1 entries.

The average fill state for columns 0 to 99 of the matrix is (26.8 ± 5.6) entries with all column averages weighted equally. The columns 100 to 154 have a slightly lower fill state due to a discontinuity in the input data.

The content inside the column buffers consists of two groups of signals: The data for which the trigger signal did not yet arrive and the data that has not been read out. The former group is larger than the latter for two reasons: Firstly, only the triggered fraction of the data is kept after the trigger where triggered fractions around 2.5 % or 1 MHz are envisaged [ATL17]. Secondly, the trigger signal delay is typically larger than the delay from the trigger signal arrival until the data is read out. This is necessary to not loose the uniqueness of the timestamp and being able to assign the signals reliably.

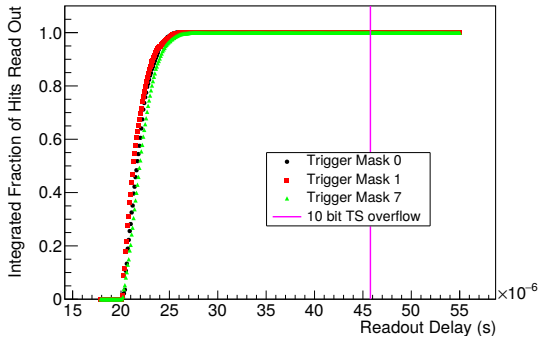
As a consequence, the major contribution is the data that has not been triggered or discarded. This amount is directly proportional to the trigger delay as it is integrating the signals from the pixels. A changed fraction of triggered events will influence this fill state, too, as more data is left for reading out if the fraction is increased. As long as the readout architecture can cope with the data rate, this change will not be comparable to the data before the trigger.

This relation becomes apparent in figure 5.17: While doubling the delay of the trigger signal results in twice the amount of data in the column buffers, issuing four times as many trigger signals changes the fill state by less than 0.1 entries. But this statement only holds as long as the readout architecture can still process all data.

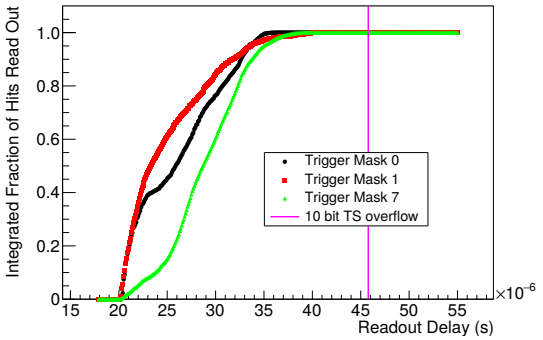
While the average values can be used to determine the influence of the contributions, the final design needs to take the upper tail of the fill state distribution into account. From the distributions in figure 5.16, it can be seen that for a total of 60 buffers per column, 22 columns would have lost data for a trigger signal delay of 25 µs. Even at 80 buffers per column, 5 columns would have lost data. The loss in the shown example is equivalent to 0.3 % of all data contained. A larger number of entries in the column memory enlarges the size occupied on the die and consequently the target to maximise the sensitive area on the die favours smaller numbers of buffers in the column memory. Consequently, the number of buffers is to be chosen as large as possible within the limits of the spatial constraints.

Trigger Table

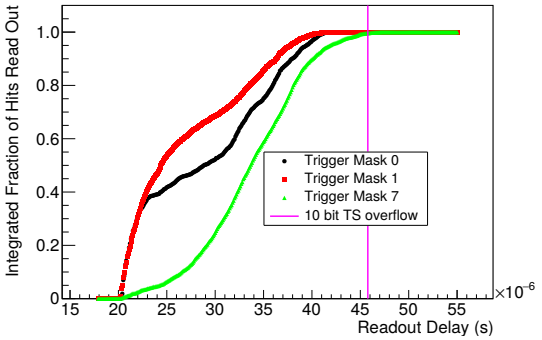
The trigger table is an important element of the readout with sorting of the signals into events. Its length also plays an important role for the performance of the whole architecture.



(a) Trigger Table Depth 4



(b) Trigger Table Depth 16



(c) Trigger Table Depth 24

Figure 5.18: The influence of the trigger table on the readout delay is only visible if those extra entries are actually used. This happens for larger trigger rates. For the figures, a trigger delay of 20 µs has been used. The integrated delay curves for different trigger table depths show increased readout delays for the longer trigger tables. In combination with more data being processed with partial sorting (masking the lowest three bits), the timestamp interval is exceeded for the trigger table with 24 entries.

If the trigger table has too little entries, trigger signals are lost. This has two implications: As the readout of an event is controlled by the content of the trigger table, the data will not be read out reducing the efficiency of the readout. On the other hand, the trigger signal is not only generating an entry in the trigger table, but it is also controlling the deletion of untriggered signals. This feature results in not deleting the triggered signals that belong to the lost trigger signal. Since the signals will not be read, they block buffers in the column memory reducing its effective size and affecting later events, too.

A longer trigger table has other drawbacks: The more entries are stored, the longer is the delay until the last event is processed for readout. Eventually, this delay gets so long that the timestamp loses its uniqueness and data can not be mapped to an event any more. In the simulation, this event is very unlikely as the timestamp is sufficiently large. But on an ASIC, the number of bits for the timestamp is limited and as it is not clear which signals have been read out too late, all data has to be discarded at a certain point.

This development is shown in figure 5.18: The difference between the plots is only the trigger table depth. For the short trigger table with four entries, the data that is not skipped from lost signals is read out quickly, whereas it takes longer for the trigger table with 16 entries. For the trigger table with 24 entries, even the timestamp interval for the 10 bit timestamp at 40 MHz is exceeded when using partial sorting masking the lowest

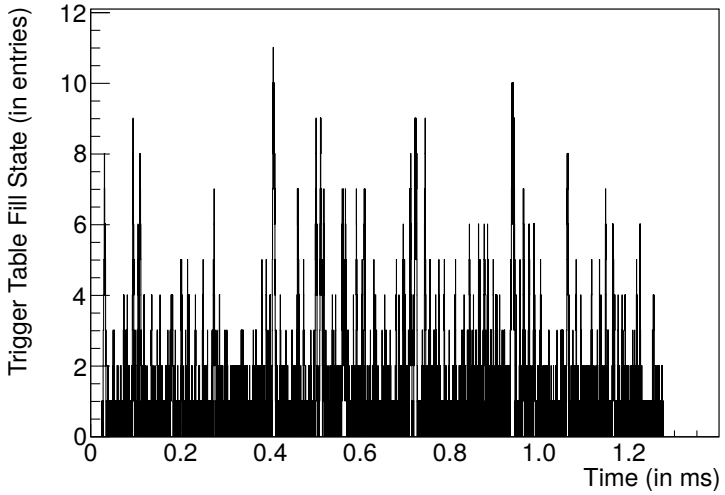


Figure 5.19: The fill state of the trigger table does not reach the maximum of 16 entries for three times the design rate of trigger signals on the final state of the simulated design.

three bits. These results disregard the fraction of lost trigger signals that decreases from 7.2 % lost for the shortest trigger table at four entries to 5.3 % for the longest trigger table with 24 entries. The individual rescaling of the curves to 1 hides differences in the total number of signals read out. Operating close to the maximum capacity of the data link, these small changes in the number of triggers being processed can lead to large changes of the delay.

The system, to which the plots in figure 5.18 belong, is a non-final version of the design with smaller data throughput. However, it emphasises the influence of the trigger table. Starting from this data, a trigger table depth of 16 entries is chosen for the designs simulated afterwards.

If operated at average trigger rates below the link capacity, the length of the trigger table is determined by the size of the trigger rate fluctuations that can be compensated. For a trigger rate of 3 MHz or a fraction of 7.5 % which is three times the design target, the trigger table of the final design is in the simulation never filled to more than 11 of 16 entries. This is shown in figure 5.19 for the 50 000 simulated events.

The difference to figure 5.18 emphasises the strong connections between all parts of the readout architecture: The possible depth of the trigger table depends on the time spent on processing one triggered event. This includes the state machine moving the signal data on its own and also the transmission structure, as it will be shown in section 5.5.3.

5.5.3 Data Word Format Considerations

To grant larger flexibility and reusability of simulation output, ROME does not impose restrictions on data widths apart from the size of the data containers used. The internal data type is a 32 bit integer per field. Instead, the amount of bits to be used per field is to be calculated off-line for the analysis.

In case of ATLASPix3, the matrix requires 8 bit for the column address and 9 bit for the row address. The leading-edge timestamp (which equals the event timestamp) is 10 bit to accommodate 25 µs of trigger signal delay and the trailing-edge timestamp for ToT measurement is planned to be 7 bit. This is a total of 34 bit of data per signal. In the

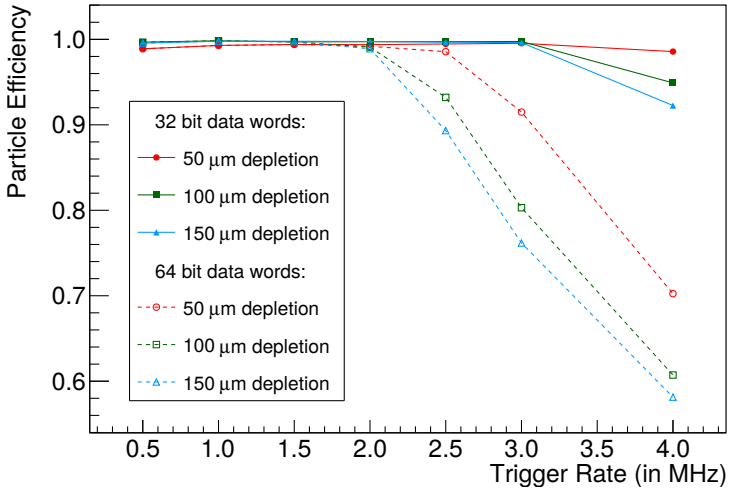


Figure 5.20: The different depletion depths represent different amounts of charge collected by the pixels per event. With this, the cluster size increases as more pixels can detect tails of charge distributions. At trigger rates above 2 MHz, the readout with 64 bit data words starts to loose efficiency. For the 32 bit data words, the same starts to happen between 3 MHz and 4 MHz. The 50 µm data line at 32 bit data words just starts to decrease at 4 MHz.

64 bit payload of the Aurora encoder, only one complete data word can be fitted leaving 30 bit unused without further measures.

However, the event timestamp is the same for all signals in one event enabling to skip the leading-edge timestamp for the signals and instead send it in a dedicated data word. This way, the data word size shrinks below 32 bit enabling transmission of two data words in one encoding cycle of the Aurora encoder and reducing the number of unused bits.

Another benefit from the smaller data word format is that the column readout state machine can be run faster without operating in an unbalanced state for the Aurora FIFO. This fact has already been used in the state machine section.

As long as the data link is occupied less than 50 % of the time with the 32 bit data words, there will not be a difference between 64 bit data words and 32 bit data words. For larger occupancies however, more data throughput is possible with 32 bit data words. In simulations depicting the extra data words needed for the transmission of the timestamp, the difference becomes visible at trigger signal rates above 2 MHz. The particle efficiency is shown in figure 5.20. The readout with 32 bit data words shows the beginning of the detection efficiency decrease at higher trigger frequencies with the decrease just starting at 4 MHz for the 50 µm depletion depth data line.

This detection efficiency decrease originates solely from the readout, the injected signals are the same for all different trigger rates, only more events are selected for readout. In the readout, the drop in particle efficiency originates from lost trigger signals which show the same characteristics (see figure 5.21).

Even before loosing trigger IDs in the trigger table, the readout with the 64 bit data words is slower than the 32 bit readout. This can be seen from the integrated number of signals read out (see figure 5.22). The curves for 32 bit data words are drawn with solid markers and rise faster to the total number of signals read out. For the 3 MHz trigger rate, the significant loss of trigger signals becomes also visible for the 64 bit data words as a smaller total number of signals read out than for the 32 bit data words where the same events

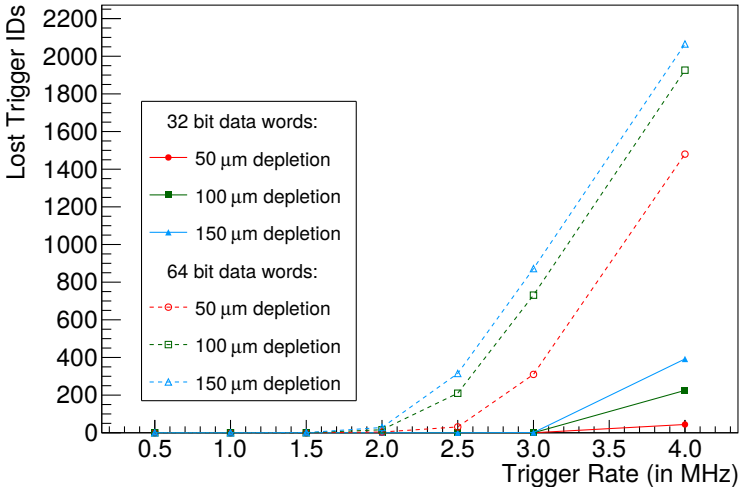


Figure 5.21: The drop in particle efficiency coincides with an increase of the number of lost trigger signals for the readout. A lost trigger ID means that no signals from the event of the lost ID can be found in the read out data.

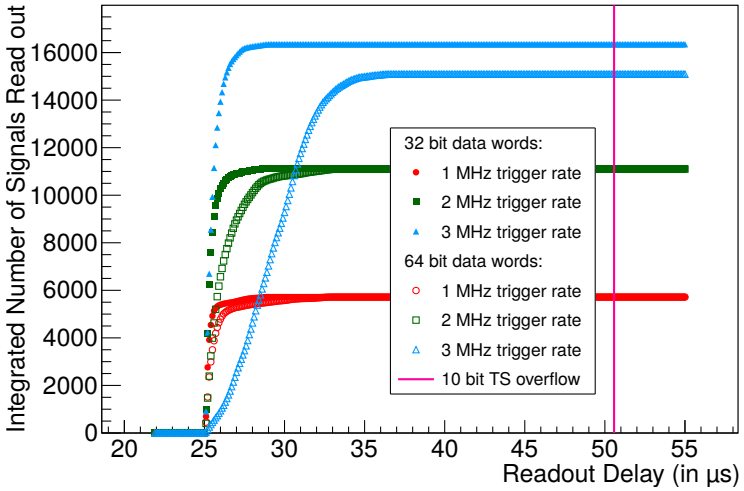


Figure 5.22: The readout of signals becomes slower with 64 bit wide data words compared to 32 bit data words. Most significantly for readout overload, but also for lower trigger and data rates. Despite the overload, the design does not exceed the timestamp interval keeping the data that are written out usable.

were triggered. In both cases, the readout delay is well short enough to ensure that the timestamp remains unique as indicated by the vertical line at 50.6 µs. The total number of signals read out matches the expectation of the ratios of the trigger rates between the data lines.

To obtain a larger security factor in the readout performance, the 32 bit data words are to be favoured enabling reliable performance above two times the envisaged trigger rate. This equals the statement that large clusters can be treated more efficiently leading to larger rate fluctuation compensation capability.

5.6 Insights from the Simulations

The simulations have shown the complexity and interconnectivity of all parts of the readout architecture of a monolithic sensor. Even small adaptions to implement a changed feature can alter the results as it has been shown in the simulations for the readout of pixel groups. Consequently, both the structure and the input data have to be modelled precisely to give sound results.

Isolating aspects in the simulation, these adaptions can be quantified for given conditions. While confirming the assumption that combining close-by trigger signals for readout speeds up the processing of their data, the analysis can point out side effects or corner cases as the possibility of slower readout for single signals in the combined and generally more efficient grouped readout of close-by events seen in the analysis for the simulations for partial sorting.

Splitting the effort into system description and analysis of the simulation results, the design conception and evaluation process is structured and the analysis can shed light on aspects of the design that were not the initial target of the simulation. The more general approach of modelling with aspects as not fixed data widths (e.g. for timestamps) and the differences compared to beam test measurements foster in-depth discussion of the results. An example for this is the comparison of the detector efficiency that can be measured at beam tests: To mimic the uncertainties from the beam test, the particle efficiency has been defined.

The ROME simulations have led to a modified readout state machine structure for the ATLASPix3 integrated sensor. Instead of one, three state machines generate the signals for the readout of the signals. These have been improved to allow for higher data throughput for example with the higher frequency of the column readout state machine compared to previous designs.

Without existing ASICs implementing a structure like the sorted readout with triggering, a size estimate for the memory storing the data during the trigger delay has been provided. The interplay of this memory size with the trigger table and the Aurora FIFO has been studied, too, and size estimates for them are included in the results of the simulations as well.

From calculations based on simulation results and adaption of the simulated system to them, also the advantages of 32 bit data words have been elaborated for the ATLASPix3 integrated sensor. The result is an increased viable trigger rate which is more than 50 % higher compared to 64 bit data words. This improvement is not as straight forward as it seems, as the benefit only materialises for sufficiently large data rates or fluctuations to them.

The final design of ATLASPix3 features a different pixel size compared to the simulations: The pixels are 150 µm wide instead of 130 µm because of the layout of the single elements of the column memory: The layout has been shrunk by the chip designers to a width of 75 µm to be arranged in double columns with respect to the pixel columns. This size was not known prior to the simulations. As a consequence, the number of columns of the matrix is 132 instead of 154. However, this 15 % increase of area per pixel – and consequently per column – does not invalidate the results from the simulations. This is because the larger pixels result in a slightly decreased number of pixel signals per cluster which was already close to one pixel per cluster. In contrary to the pixel group simulations where the group width changes were larger (200 % compared to 15 %) in combination with much larger clusters (about a factor 20 larger than for ATLASPix3), the effect of the changed pixel size on the ATLASPix3 design is small: The readout has spare capacities to compensate for reduced readout efficiency from more the signals distributing over less columns and requiring more cycles for readout as long as the column memory is still sufficient. If

necessary, ATLASPix3 can be operated at smaller trigger delays for the tests to compensate for the increased area per column as it was shown in the discussion of the column buffer size recovering the space needed.

6 The ATLASPix3 Integrated Sensor

From insights of measurements on previous generations of HV-CMOS sensors and simulations of analogue circuits, of digital structures and of the architecture, the ATLASPix3 integrated sensor has been developed in under the leadership of Prof. Ivan Perić. Architecture design choices have been made according to extensive ROME [Sch21b, Sch21c] simulations by the author as presented in chapter 5. The final design is presented first in section 6.1.

The target application of the design is the fifth layer of the inner tracker upgrade of the ATLAS detector for HL-LHC. According to this target, a set of requirements is defined [ATL17]. This includes limits on power consumption, detection efficiency, time resolution and radiation hardness (numbers are given in section 2.2.3). Furthermore, due to limited space for cables, the design concept for the upgrade defines constraints on the powering scheme and form factor to build modules and from them the whole tracker layer.

To test the produced ASIC, the GECCO system [Sch21a] has been prepared for ATLASPix3 by the author and is described in section 6.2. The measurements performed with it test all aspects mentioned above.

Power consumption and powering schemes are tested first (section 6.3). ATLASPix3 features three interfaces for configuration (direct access, serial peripheral interface (SPI) and a command decoder) and a division of the configuration data into several shift registers opposed to previous sensors. The new configuration scheme enables much faster configuration which is important for large systems. Considerations for this step are included in appendix B.

The detection efficiency is tested in the laboratory using the test signal generation circuit of ATLASPix3 including tuning of the pixels to obtain a homogeneous behaviour of all pixels. The measurements for this are presented in section 6.4.

Measurements complementing these with particles are described in section 6.6. Particle signals are used to calibrate the test signals used before and to test the amplifier and comparator characteristics.

Time resolution measurements are presented in section 6.5. These include measurements of the signal timing of the pixels, but also the amplitude measurement with calibration which can be used to improve timing correcting for detection delay differences – the time-walk.

Radiation hardness is tested by comparison of irradiated and unirradiated samples (section 6.7).

Fitness of ATLASPix3 for module building is investigated on a telescope built from four ATLASPix3 layers. This telescope is described in section 6.8 before measurements with it are presented in section 6.9. These measurements are to be understood a proof of concept for the telescope and the fitness of ATLASPix3 for building modules as the time at beam tests has not been sufficient for commissioning and measurement campaigns.

With its large active area, ATLASPix3 is also a candidate for other applications. For an ion beam monitor as used at HIT (section 2.3, ATLASPix3 has been used as technology representative to characterise the signals generated by the beam in HV-CMOS detectors. These measurements are included in the section on the measurements with particles (section 6.6.3) and the inhomogeneous irradiation profile to be expected for a beam monitor used there have been studied, too (section 6.7.2). Large signals, as expected from the ion beams at HIT, also enable timing improvements making use of the threshold tuning structures where the finite rise time is used to shift the timing of pixels by setting different thresholds (section 6.5.2).

Beyond the content of this thesis, ATLASPix3 has been used to build a quad module prototype based on the structures built for the telescope in the working group of Prof. Attilio Andreazza at INFN Milano (IT). The readout system for it is the GECCO system of which the adaption has been done by B. Raciti in her master's thesis [Rac21] under the author's co-supervision. More information on the quad module is included in appendix C. Furthermore, the telescope itself has been used for tracking of secondary particles at HIT as part of the master's thesis of C. Klauda [Kla22] under the author's supervision.

6.1 ATLASPix3 Design

The ATLASPix3 integrated sensor is a full reticle size detector with a size of 20.2×21 mm^2. The design has been started in 2018 in a 180 nm HV-CMOS process and has been submitted for production in April 2019.

The requirement specification for the application has been taken from the ATLAS ITk upgrade in the fifth pixel layer[ATL17]. Parts of the design have been created during the PhD thesis of M. Prathapan [Pra20]. The insights from the ROME simulations, presented in section 5.5, have been used for the design and dimensioning of the readout. However, the ASIC implementation details are not part of this thesis.

The layout of ATLASPix3 is shown in figure 6.1. The sensor is divided in two parts: The sensitive matrix and the insensitive periphery.

The periphery is located at the bottom of the area with a height of 2 mm spanning the whole width of the sensor. It contains configuration buffers, regulators, biasing structures and the readout electronics as well as the pads. The pads are located only at the bottom edge of the die to allow for three-side butting for module building.

The matrix consists of 132×372 pixels, each with a size of 150×50 µm^2, spanning the other 19 mm of the height of the design. Every pixel contains an amplifier and comparator sending digital asynchronous hit information to the periphery. The charge generated by a traversing particle is collected via drift from the depleted region below the large collection electrodes. This charge is amplified and compared to a threshold voltage to generate the binary hit information. For threshold tuning, each pixel contains four bit of memory. Three bits shift the detection threshold while the fourth bit disables the amplifier to turn the pixel off.

In the following sections, the design is described starting from the logic level description and then following the signal chain: The logic arrangement of the detector is sketched first. Then, the elements for detecting and storing signals are described. This is followed

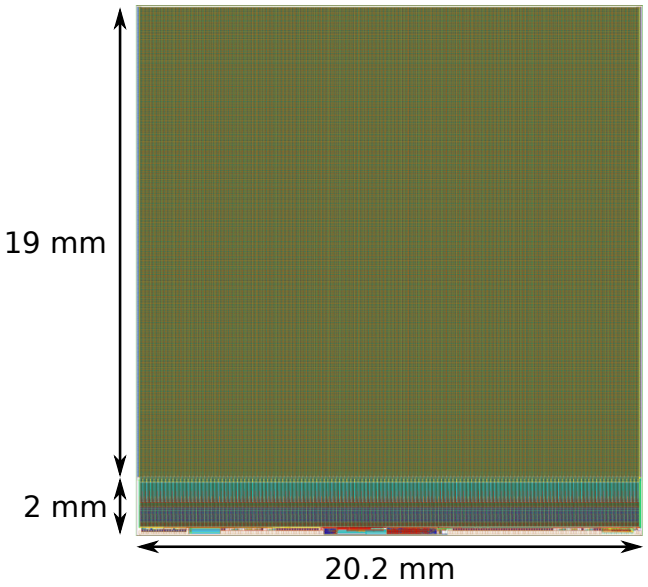


Figure 6.1: The layout of ATLASPix3 is divided into two parts: The matrix at the top and the periphery at the bottom. The matrix is the sensitive area with 132×372 pixels while the periphery contains readout electronics, configuration, powering and pads. The pads are only on the bottom edge of the sensor making it 3-side buttable. [Per19]

by the description of the elements necessary to operate the whole system starting at the configuration and timing signal inputs. At the end, the arrangement of these elements in the layout is presented before modifications for the ATLASPix3.1 sensor are listed. More detailed information on the design can be found in [Per19]. Schematics and layouts are extracted from the design sources.

6.1.1 Matrix and Column Structure

The matrix of ATLASPix3 consists of 132 independent column structures. The column based structure is a direct result from the decision for the column-drain readout that enables a repeating structure combining sensing elements with readout structures located at the bottom of each column. In case of ATLASPix3, the column structures contain not only elements for readout, but also for configuration in addition to the pixels. The readout structures are hit buffers for receiving the hit signals from the pixels, content-addressable memory (CAM) for selecting triggered events in the generated data and EoC buffers connecting the column structures readout-wise. The configuration elements integrated in the column structures are used for pixel-wise configuration, global settings are not included here.

To reduce the noise in the pixel matrix, ATLASPix3 is separated in two parts: the sensitive matrix containing the pixels and the insensitive periphery where the digital processing is confined. The pixel matrix is asynchronous, reducing noise from digital readout activity and power dissipation from shorter transmission lengths in the periphery. The arrangement of the readout elements is driven by the two readout modes (hit-driven and triggered) and is sketched in figure 6.2 with the possible readout paths: Both paths start at the 372 hit buffers receiving the signals from the pixels are the readout structure closest to the pixels and are connected to one pixel each. The next structure is the EoC buffer for hit-driven readout. For triggered readout, the data is not shifted to these EoC buffers but to one of the 80 the CAM elements, the content addressable buffers (CABs), below it to await the trigger decision for readout or deletion – hence the label trigger buffer. The hits for

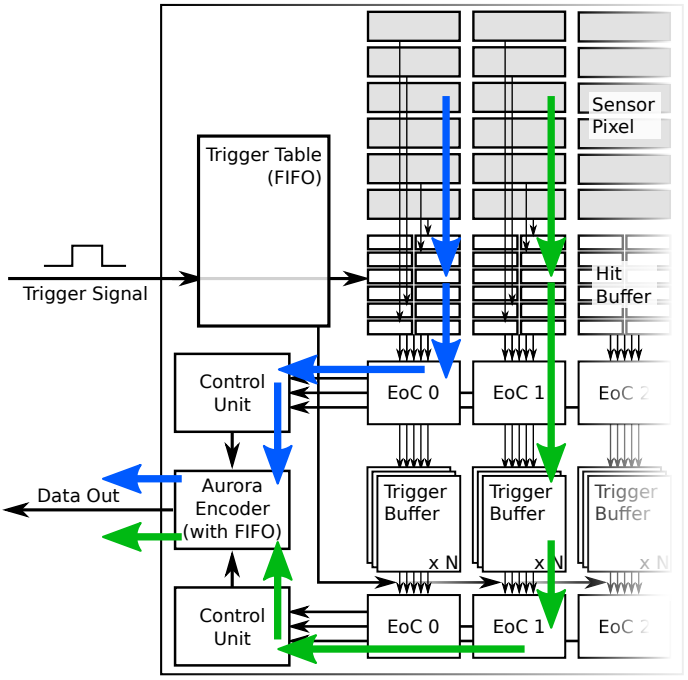


Figure 6.2: The periphery of ATLASPix3 consists of one hit buffer per pixel, end-of-column (EoC) registers and $N = 80$ trigger buffers per column. The pixels digitise the signal and transmit this information to hit buffers in the periphery via a one-to-one connection. From this point, two readout options exist: Either hit-driven readout or triggered readout can be used: For hit-driven readout, the hits are loaded into the first set of EoC buffers and are then written out encoded and serialised. For triggered readout, the hits are first transferred to trigger buffers where they stay until the arrival of the delayed trigger signal. Without a trigger signal, they are deleted. Otherwise, the readout occurs in a similar way to the hit-driven readout, but the trigger table is used to select only one event at a time resulting in time sorted hit data written out via the second set of EoC buffers and the triggered RCU.

the triggered events are then moved to the second set of EoC buffers sorted by events. Both hit buffers and CAM elements are arranged in double columns inside the column structure as for their smaller width. In both readout modes, the data from the EoC buffers is transmitted to a serialiser and encoder module.

Between the pixels and the hit buffers, the configuration elements are placed. Since they are inactive without configuration activity, they are not emitting noise from switching enhancing the shielding of the pixel matrix from noise generated by digital activity. The configuration elements include the configuration bits for column control, row control and tuning DAC control in three independent shift registers going through all columns. The row control register is implemented as mapping from the column to three rows: The first column structure contains the row register bits for the rows zero to two, the second column structure for the rows three to five and so on.

The one-to-one connection of the pixels to the hit buffers makes use of three metal layers. Each one connects 124 pixels to the hit buffers. In order to equalise the timing between the pixels with a physical distance of 19 mm, the transmission lines were designed to be all of similar length to obtain similar capacities resulting in similar timing. A visualisation of the mapping from pixels to hit buffers is shown in figure 6.3.

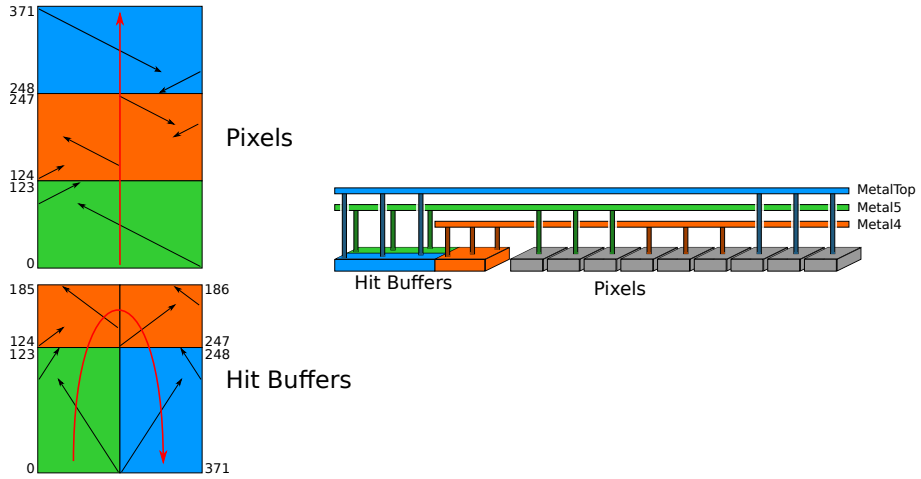


Figure 6.3: The pixel-to-hit-buffer connection is done using three metal layers (Metal4, Metal5 and MetalTop). The double column arrangement of the hit buffers results in a U shaped mapping from the pixels to the hit buffers. The pixels are split in three routing groups of 124 pixels each to be routed via one metal layer as indicated by the colour code. The black arrows indicate the order of the mapping as Y point and the horizontal position of the line as X point. The jumps inside the routing groups are due to the horizontal position of the hit buffer input being crossed. The numbers next to the blocks indicate the pixel row at that position.

6.1.2 Pixel

The active pixels of ATLASPix3 contain a charge-sensitive amplifier and an individually tunable comparator. In addition, each pixel features a test signal injection circuit for testing purposes. The simplified schematics of the amplifier is shown in figure 6.4. The grey arrows indicate tunable parameters from the configuration of the chip.

The pixels are biased with high-voltage U_{HV} applied from the substrate to the n-wells and the analogue supply voltage $VDDA$ with adjustable biasing strength. This results in a depletion voltage of $U_{\mathrm{HV}}+$ VDDA. VDDA is specified to be 1.8 V while U_{HV} is specified by the process to be up to 120 V. The VDDA supply is important as it protects the low-voltage electronics inside the pixel from the depletion voltage.

Capacitive coupling ensures that the in-pixel electronics are not supplied with the depletion voltage at the amplifier input. The charge-sensitive amplifier is described in section 3.5. The supply current of the amplifier can be adjusted with the parameters VPLoad and VN. Various amplifier components are driven by currents and voltages globally configurable to tune the behaviour of the pixel electronics: The length of the signals can be adjusted with the parameter VNFB a current controlling the feedback loop of the amplifier. The capacitive coupling at the output of the amplifier enables tuning the size of the signal for the comparator with the baseline resistance parameter BLRes. The baseline voltage BL is the reference for the threshold set for the comparator.

For pixels in row 0 of the matrix, the analogue output of the amplifier is accessible on a pad of the chip for one pixel at a time.

To test a pixel, charge can be injected through the Inject input. This is a circuit charging a capacitor to an adjustable voltage to inject a defined charge into the pixels. The signal EnInj (and its complement EnInjB) can be turned on for individual pixels or groups of pixels (see section 6.1.4 for details). For all other pixels, this input is connected to ground potential to prevent such signal generation.

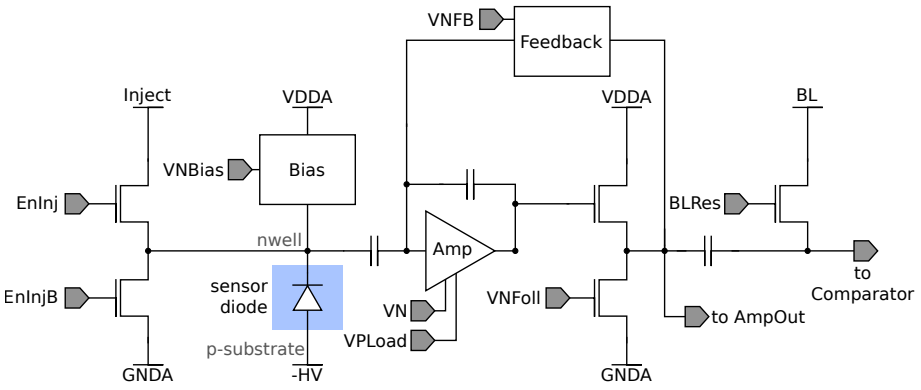


Figure 6.4: The first part of the in-pixel electronics consists of the charge sensitive amplifier with source follower getting a capacitively coupled signal from the sensor diode between the deep nwell and the p-substrate. A test signal injection circuit sketched on the left and capacitive coupling of the signal for the comparator on the right side complete the diagram. By capacitive coupling between the sensing diode and the amplifier, the low-voltage amplifier is protected from the large voltage depleting the sensor volume.

The comparator consists of two stages, the first one contains an additional differential current DAC draining a part of the current of the branches for adjusting the threshold. The simplified schematics is shown in figure 6.5. The grey arrows are again inputs connected to the circuit, and – apart from VCasc which is fixed internally to about 1.08 V – adjustable. Gate is an external supply voltage which is higher than VDDA and has a value of $U_{\mathrm{Gate}} = 2.1\,\mathrm{V}$. AmpOutAC is the output from the amplifier circuit in figure 6.4, Th the global threshold voltage. VNComp adjusts the bias current of the comparator.

To control the tunable comparator and to disable the amplifier, the pixel contains four bits of RAM implemented as two NOT gates connected in a loop. Three bits adjust the tuning DAC for the comparator and one turns off the amplifier.

The differential threshold tuning DAC connected to idacp and idacn consists of three transistors with one, two and four fingers connecting the outputs OutS1 and OutS1B to analogue ground potential. These are driven by the VNDAC parameter of the chip. current

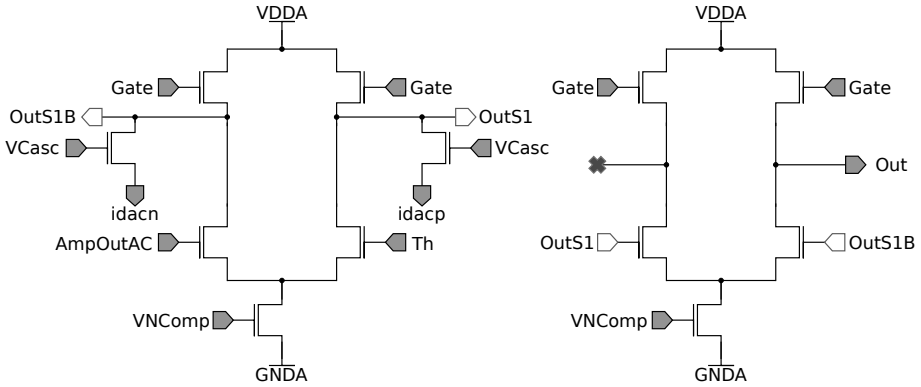


Figure 6.5: The in-pixel comparator of ATLASPix3 has two stages with the threshold tuning implemented in between the two stages: The threshold tuning is done with a differential current DAC that is draining a part of the current from the branches of the first stage shifting the transition point for the second stage providing the single-ended output. The differential design of the tuning feature enables both rising and lowering of the threshold.

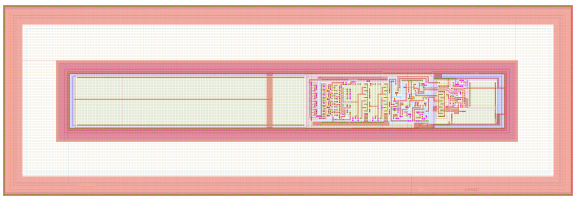


Figure 6.6: The pixel is 150×50 µm^2 large and contains an n-well with amplifier, comparator and memory for tuning (right to left) inside the deep n-well in the middle of the pixel.

flow through them is activated on the idacp side for a set RAM bit and on the idacn side for a cleared RAM bit. Hence, the parameter VNDAC adjusts the amount of current drained by the tuning DAC and therefore adjusts the step width of the comparator threshold change between the different settings.

This structure has the property that the lower four settings for the three RAM bits lower the comparator threshold while the upper four settings raise it keeping the current consumption independent of the setting of the tuning digital-to-analogue converter (DAC) (TDAC).

As mentioned before, these structures are implemented in the n-well of the pixel which has a size of 150 × 50 µm^2. The layout of the pixel is shown in figure 6.6.

6.1.3 Readout Buffers

Generally speaking, the readout of the pixel matrix multiplexes all 49104 pixels to one readout channel. This multiplexing requires arbitration at the nodes of this readout tree structure. For this, memory cells are necessary to temporarily store the hit signal information before it can be read out.

For the two readout modes, ATLASPix3 implements three full-custom buffers for pixel readout that are described in this section: The hit buffer receiving the hit signal from the pixel, the EoC buffer multiplexing between the columns of the chip and the CAM for storing hit information until the trigger decision for triggered readout.

Hit Buffer

The hit buffer is receiving the comparator output from the pixel and consists of two parts: The first part is the memory to store the hit information and the second part is the readout logic for transferring the data to the EoC buffer for hit-driven readout or to the CAM for triggered readout.

Because of the one-to-one connection between pixel and hit buffer, the pixel address (which is the row in the column) is fixed requiring no memory. This is implemented as a hard-coded input to this block. The column-drain readout also removes the necessity to store the column address in the hit buffer since all hit buffers connected to an EoC block share the same column address.

Hence, the data to be stored in the hit buffer are the leading edge and trailing edge timestamps. These two independent timestamps are provided Gray-coded to the block and stored on the rising and falling edge of the comparator signal from the pixel. The leading edge timestamp has 10 bits and the trailing edge timestamp is stored with 7 bits. The reasoning is to be able to cover the trigger delay for the ATLAS ITk upgrade system at 25 ns resolution. At design time, this delay was foreseen to be up to 25 µs. The smaller trailing edge timestamp is chosen as it provides sufficient resolution for the charge measurement while reducing the bandwidth necessary for readout. The technical design report for the

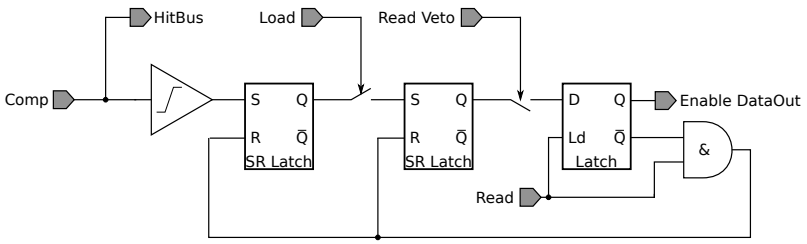


Figure 6.7: The hit buffer consists of two parts: The memory to store the timestamps of a signal and the logic to read out the information. This simplified schematic shows only the readout logic: The hit buffer receives the comparator output signal from the pixel and a rising edge detector creates a pulse on a signal start that is stored in a set-reset latch. The Load signal selects all currently filled hit buffers for readout, so that signals detected after this signal are not directly considered for readout. The Read signal then leads to transmission of the data from the hit buffer selected by the priority chain that creates a veto for all subsequent buffers with data. In addition, this reading operation from the buffer clears the latches freeing the buffer for the next signal.

upgrade of the ATLAS inner tracker [ATL17] provides several different numbers on the maximum trigger delays. To cover the largest number of up to 35 µs, the design would have to be changed adding an extra bit to the leading edge timestamp. The values used for the design have been a trigger rate of 1 MHz at a delay of 25 µs which is a level-1 trigger for a two-level triggering scheme.

In addition to the timestamps, status flags are implemented in the hit buffer to determine its state. These flags and their connecting logic are shown in figure 6.7: The comparator output from the pixel is transformed into a pulse by an edge detector which sets the first set-reset (SR) latch. Its inverted output $\overline{Q}$ is used to enable updating the leading edge timestamp. Hence, updating of the timestamp memory stops on receiving a hit. The trailing edge timestamp is only updated while the comparator output is high. As a consequence, it stops updating on the falling edge. The next latch is set on a load signal issued by the readout control unit (see sections 6.1.6 and 6.1.7). The Read Veto signal is generated by a priority chain over all hit buffers in a column ensuring that only one hit buffer at the time sets the last data latch and enabling the data output drivers. The Read signal – again issued by the readout control unit – activates the data output drivers and clears the status flags freeing the buffer for the next hit signal.

The scan logic for activating only one hit buffer data output at a time was developed in previous ASICs [Web16, WAB+19] to ensure sufficient speed for the operation on sensors with fast readout: In principle, the priority chain has to go through all buffers to find the first one in the line with data deactivating all other outputs independent from their state. For this, the output of one element is the input of the next one. As the logic for each element introduces a delay and the chain contains 372 elements for ATLASPix3, the signal chain would limit the readout speed of the hit buffers. Therefore, the scan logic has been extended by a second input for a fast signal and grouping of the elements as shown in figure 6.8: The output of the last element in a group is not only connected to the input of the first element in the next group, but to all elements of the next group. Therefore, the output of all groups following the group with the first filled element is delayed only by the signal delay of one scan chain element. This speed-up is also visible in the simulated delay plot for a sysetem with groups of ten elements in the bottom right of the figure. The signal input for the scan chain is the second SR latch as shown in figure 6.7. The hit buffers on ATLASPix3 are arranged in 12 groups of 31 buffers per column. The same concept will be used for the CAM and the EoC buffers.

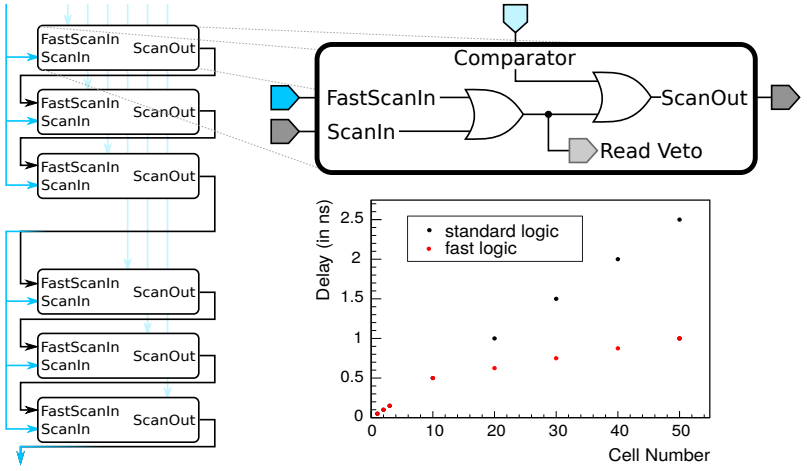


Figure 6.8: The priorisation of the hit buffers for readout is done with a priority chain of which a simplified schematic is shown. However, a simple chain is not fast enough for the 372 pixels in one column of ATLASPix3. Therefore, the hit buffers are sorted into groups with n buffers that receive a common fast signal. With this signal, $n-1$ buffers can be skipped in groups after one with a full buffer. The timing improvement is shown for an estimation created for MuPix8 with groups of 10 buffers [W+18, WAB+19].

The layout of the buffer has a size of $75 \times 4.2\mu m^2$. This way, an arrangement in double columns is possible reducing the height of the insensitive periphery. This means that for every column of pixels, two columns of buffers with the same total width are placed.

End-of-Column Buffer

The end-of-column (EoC) buffer is located at the end of the hit buffers for hit-driven readout and at the end of the content-addressable memory for triggered readout. It contains the hard-coded column address that is added to data read out from it and propagates the information from the preceding memory: row address and the two timestamps.

In addition, the EoC buffer receives the end signal of the priority chain over the preceding buffers. This information is used by the readout control units to determine whether there are hits available for readout. The block also distributes the readout signals and the timestamps from the control units to the preceding memory.

The output of the EoC buffer is built in a similar way as for the hit-buffers: With a priority chain, the buffer to send the data down-stream is determined and then read out.

Content-Addressable Memory (CAM)

The trigger buffer structure is implemented as content-addressable memory (CAM). Its purpose is to select the triggered hits and sort the triggered data event-vise. Each column contains memory for 80 datasets which are filled using a priority chain similar to the one for the readout of the hit buffers (see section 6.1.3).

Selecting the triggered signals is done by deleting not triggered signals at the time the trigger signal is configured to arrive. Sorting by event is done by selecting only signal information that belongs to one event for readout at once. After selection, the data is read out from the memory again using a priority chain as for the hit buffers. After all signals of an event are read out, the next event is selected repeating the procedure.

This functionality is achieved with two digital 10-bit comparators and logic. The simplified schematic of one CAM cell is shown in figure 6.9: The input data is the one from the

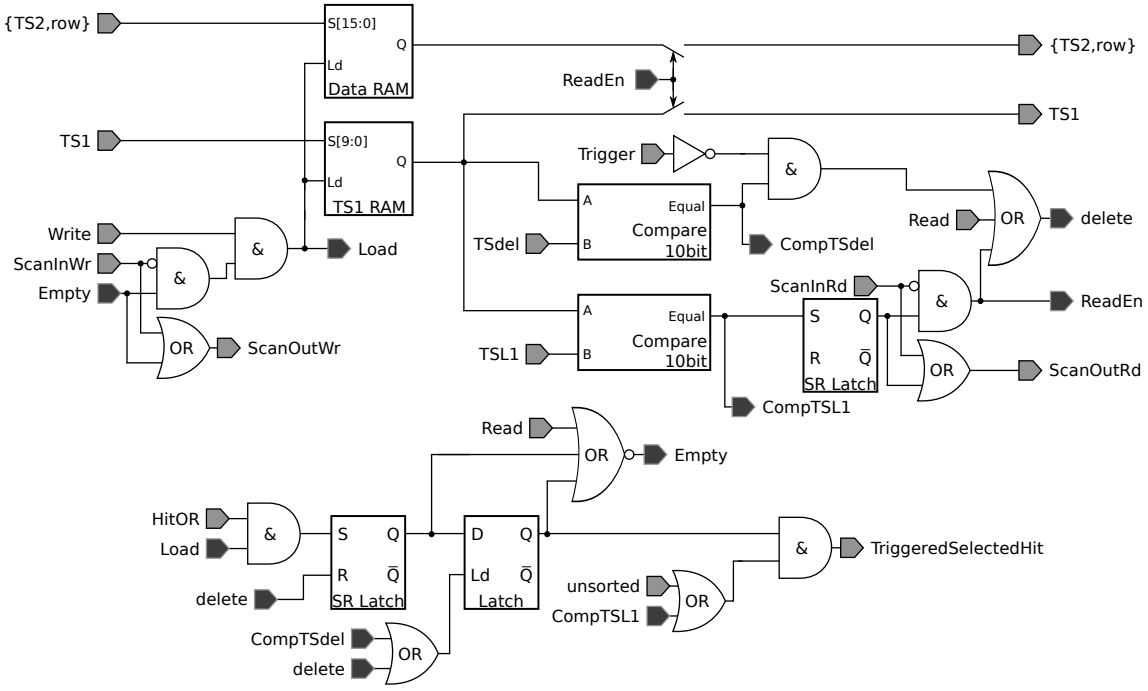


Figure 6.9: The content addressable memory cell of ATLASPix3 stores the two timestamps (TS1 and TS2) and the row address of a signal. The first of the two digital comparators deletes hits without trigger by comparing the TS1 with the delayed timestamp TSdel. The second digital comparator selects hits with a TS1 value matching the current L1 timestamp TSL1 for readout. Input and output of the buffer is controlled by a scan logic similar to the one for the trigger buffers.

hit buffer: row address and the two timestamps. It is split into two memories: trailing edge timestamp TS2 and row address are stored in the data RAM, and the leading edge timestamp TS1 is stored in the TS1 RAM. The latter is the addressed content giving the name of the buffer. The digital comparators compare the TS1 to the delayed timestamp TSdel for issuing deletion of the information: The TSdel is the same timestamp as the timestamp for the leading edge of the signal but delayed by the set trigger delay. This way, the trigger delay for a hit signal has passed when TS1 and TSdel match. The combination of the trigger input and this comparator output generates a deletion signal for the memory. If the trigger signal does arrive, the hit is marked as triggered with the latch at the bottom of the figure.

Sorting is done with the second digital comparator: It compares the TS1 with the level-1 timestamp TSL1. The TSL1 is created from the delayed timestamp and its value is stored in a FIFO and one value after the other is read by the readout control unit (see section 6.1.7) and provided to the CAM as TSL1. A match of TS1 and TSL1 leads to a preparation of the memory content for readout. The output drivers are activated on selection from the priority chain for readout of the CAM content.

Just as for non-triggered events, the reading of the data leads to deletion of the content in the buffer after reading is finished. As a recovery feature, the event sorting can be turned off with the unsorted signal. It overrides the output of the second digital comparator enabling readout for all stored datasets.

The unsorted readout has two purposes: Firstly, due to the limited size of the FIFO buffer storing the TSL1 values, such values can be lost. In this case, the datasets are stored for readout and occupy the memory, but they are not read out until the same timestamp is triggered again. On one hand this is bad, because data is mapped to a wrong timestamp

and on the other hand, these datasets block memory buffers reducing the number of buffers available and by this the maximum signal rate that can be processed with the system. Secondly, the sorting into events requires sending the framing data of the events reducing the available bandwidth for the actual signal data. Grouping all hits into the same event, the time needed to read out all datasets can be reduced.

Compared to the priority logic of the hit buffers, the group size for the CAM is ten instead of 31 and present once for filling and once for reading the data.

The layout of a single cell has a size of 75×15 µm^2 and the 80 cells per column are arranged in double columns.

6.1.4 Configuration

ATLASPix3 is configured using shift registers with a two-phase clock. Compared to previous designs, the single shift register for all configuration has been split into six parts. This enables faster configuration as less bits have to be transmitted for repetitive changes in a part of the configuration. This is for example the case for writing of the in-pixel RAM.

In the following, first the interfaces to the configuration shift register controls are described, then the six shift registers including the process of writing the in-pixel RAM. At the end, the structure of the single bits in the shift registers is described.

Configuration Interfaces

The shift registers share the two clock signals and the data input. Only the load signals are independent.

Write access to the shift registers is possible via three interfaces: direct access to the signals on pads of the die, via an SPI interface and via a one-line differential command based signal, the command decoder (CMD). The latter two also implement ways to read back configuration data: via the Master-in-Slave-out (MISO) line for SPI and via the data output used for hit signal data for the CMD.

The SPI interface grants access to a register driving the signals for the shift registers for writing and 64 bit of data for reading. In addition to the shift register interface signals, also a signal to generate charge injection signals in the pixels is available. For reduced complexity systems, also the digital readout data can be sent via the MISO line. However, this puts a low limit on the data rate possible.

The CMD interface is a command based interface. The commands can change signals similar to the SPI register but also send blocks of data bits that will be written to the shift registers by an on-chip state machine. Furthermore, trigger signals for triggered readout can be sent. The data is encoded carrying 10 bit of data in 16 bit transmitted. The concept also allows for reconstructing a reference clock for the phase-lock loop (PLL) from the data stream (see section 6.1.5). This clock recovery requires the line to be constantly active even without actual data transmitted. Therefore, special data words for clock reconstruction and for synchronisation are to be sent if no data is available. The state machine has no buffer for the data to process. This means that while the state machine is busy processing one command, the synchronisation words should be sent to not corrupt the data being processed.

The state machine for the CMD is driven with a clock at a quarter of the speed of the data output clock. This means, that for a readout speed of 1.28 Gbit/s the readout clock is 640 MHz because double data rate is used and the CMD is then driven with a 160 MHz clock.

Shift Registers on ATLASPix3

The six shift registers of ATLASPix3 are:

- The Config Register contains digital settings for the readout and also a reset flag deactivating the whole detector.
- The DAC register contains the settings for the biasing DACs on the detector. Also, an un-lock pattern is included here which will turn off the biasing block if set wrong.
- The Voltage DAC (VDAC) register contains the settings for analogue voltages on the detector.
- The Column register contains flags for debug input and output configuration.
- The Row register contains flags, just as the column register, flags for debug inputs and RAM writing.
- The TDAC register contains the data to be written to the in-pixel RAM.

The reset and un-lock flags require that on first configuration after power-up, the config register is written first, followed by the DAC register. Then the other registers can be written in any order.

All registers are located in the periphery at the bottom of the detector. The mapping from the row register from the horizontal orientation to the vertical row direction is done with a diagonal column-to-row mapping from one column to three consecutive rows starting at rows zero to two in column 0.

The debug input mentioned in column and row registers is the configuration of the charge injection circuit. This circuit is activated in the pixels in a projection encoded way: A circuit is activated in a pixel if both the bit for the column in the column register and the bit for the row in the row register are activated.

For the first row, the output of the pixel amplifiers can be connected to an output pad of the detector. This is activated with a configuration bit per column. Also, the or of all comparator outputs in a column can be connected to another output pad of the detector with another bit. This output is called hitbus.

RAM Writing

To write the four bits of in-pixel RAM, the row register, the TDAC register and the four write RAM signals are necessary. Each of the write RAM signals is for one bit of the in-pixel RAM. The actual write enable signal for the RAM bits in the pixels is the and-function of the write RAM signal and the RAM writing flag for the row in the row register. Hence, the row register defines in which row the RAM is written. The TDAC register has one bit per column which contain one bit of data for writing per column.

To write the RAM of a row, the row register has to be written first to select the row. Then, the data for the first bit of each pixel in the row is loaded into the TDAC register and written to the RAM with a pulse on the corresponding write RAM signal. Next, the data for the second RAM bit is loaded into the TDAC register and the second write RAM signal is fed with a pulse. This procedure is repeated for the third and fourth bit, too.

To write the whole matrix, the procedure described above is repeated for every one of the 372 rows.

This means that the TDAC register has to be written 1488 times. In combination with the separation of the write RAM signals from the shift registers, this greatly reduces the amount of data to be written for configuring the RAM of the matrix. Compared with the configuration scheme used previous versions with only one shift register, this is a reduction of the number of bits written for RAM configuration by 91.2 %.

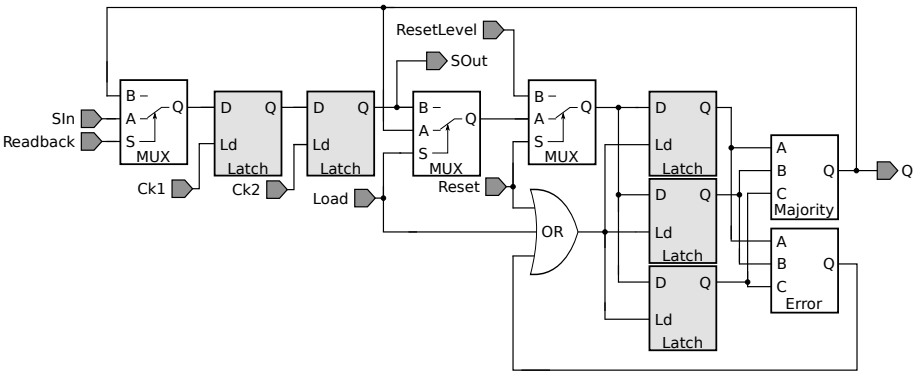


Figure 6.10: The shift registers on ATLASPix3 are configured with a two-phase clock and feature triple-redundancy to prevent errors due to SEU. The outputs of the three latches are monitored for errors and written back via majority logic to recover the flipped bit.

The Structure of the Shift Register Bits

The single bits of the global configuration registers consist of several latches. Their combination is shown in figure 6.10: For configuration, the Readback signal is low, and the serial input Sin is used which is connected to the serial output Sout of the previous bit. The first two latches are driven by the two clocks Ck1 and Ck2. A high level at the Ld input of the latch makes it take the level at the D input. With the two latches in a row, the two-phase clocking is implemented. This ensures that the shift register will not suffer from timing problems for longer chains as in this case, the configuration speed can be reduced.

With the Load signal, the information stored in the second latch is written to the three latches in on the right in parallel. These form a triple redundancy to cope with single-event upset (SEU): If one of the three latches is flipped by an SEU, this will be detected by the error detection circuit and the value generated by the majority logic block will be written back into all three latches to recover the SEU.

The register bit also has a reset input, which will restore a hard coded configuration connected to ResetLevel.

6.1.5 Clocking Scheme

ATLASPix3 can be fed a readout clock in three ways: directly with the clock for readout, with a reference clock for a phase-lock loop (PLL) at a quarter of the readout speed and with the CMD interface of which the data stream is used for clock data recovery (CDR).

ATLASPix3 has to be configured accordingly for each option. The simplified structure of the clocking inputs is shown in figure 6.11: The clock signals are shown in red, the configuration pads in grey and the configuration flags in black.

The reference clock for the PLL and the CMD signal are received with a phase detector to be used in the PLL shown as its components: phase detector, charge pump and voltage controlled oscillator. The phase detector to use is defined by the EnCDR signal, while the phase detectors are enables by the EnPLL signal. The output speed of the PLL can be divided by two with the SelSlow and takefast signals. With SelExt, the direct readout clock is selected over the PLL source.

The resulting clock is provided to the two readout control units. One of them has to be deactivated via the untriggeredRoEn pad and configuration bit.

For probing the working state of the PLL, the readout clock divided by four by the triggered readout control unit is sent out on a differential pair.

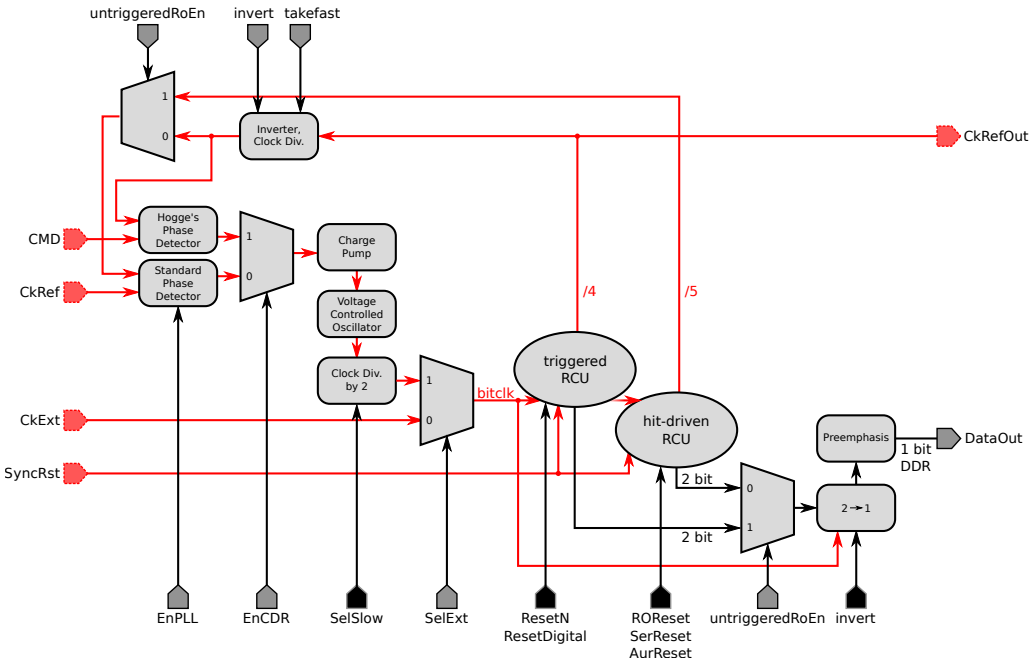


Figure 6.11: ATLASPix3 has several options for providing the clock for the digital readout: It can either be provided directly, as reference clock for a phase-lock loop or as data stream of the command decoder for clock data recovery. The selection is done via several configuration flags. Some of these are pads of the chip, indicated as grey arrows with black border. Others are bits in the configuration register indicated by the black arrows with light border. The clocking inputs and outputs are indicated by red arrows with dashed border. After [Per19]

6.1.6 Hit-Driven Readout

The readout control unit (RCU) for hit-driven readout is very similar to the one used on the MuPix8[1]. Its state machine is only extended to send one block of data more for ATLASPix3.

The readout control unit drives the readout from the hit buffers to the first set of EoC buffers and from there, the data is transferred to an 8b/10b Aurora encoder and a serialiser to send the data off the chip. The whole system is controlled by a state machine visualised in figure 6.12. To avoid timing problems, the signals for operations are issued always with one cycle (i.e. state) in between. Matching this structure, the data is sent to the encoder in larger blocks of 32 bit with four status bits, one per byte. The states in which data is sent to the encoder and serialiser are highlighted in green with the data sent in the states.

The data sent in the Load Column 2 state identifies the start of a readout cycle which can be identified by the state machine decoding the data. For all other states, a spacer word, called *comma word* is sent. This data word can have two encodings that can be identified by the receiver to align the data stream.

The working principle of the state machine is to load the currently available hit datasets from the hit buffers to the EoC buffers in the Load Column states. Then the new hit datasets in the hit buffers are prepared for readout in the Load Pixel states. Finally, one dataset is transferred to the serialiser per loop through the Read Column states. If no more hits are available in the EoC buffers, the loop is restarted at the Pulll Down states at which no operation is executed and synchronisation words are sent. If no hit datasets are present

[1]MuPix is the HV-MAPS detector foreseen for the Mu3e experiment. MuPix8 is a prototype detector for this experiment.

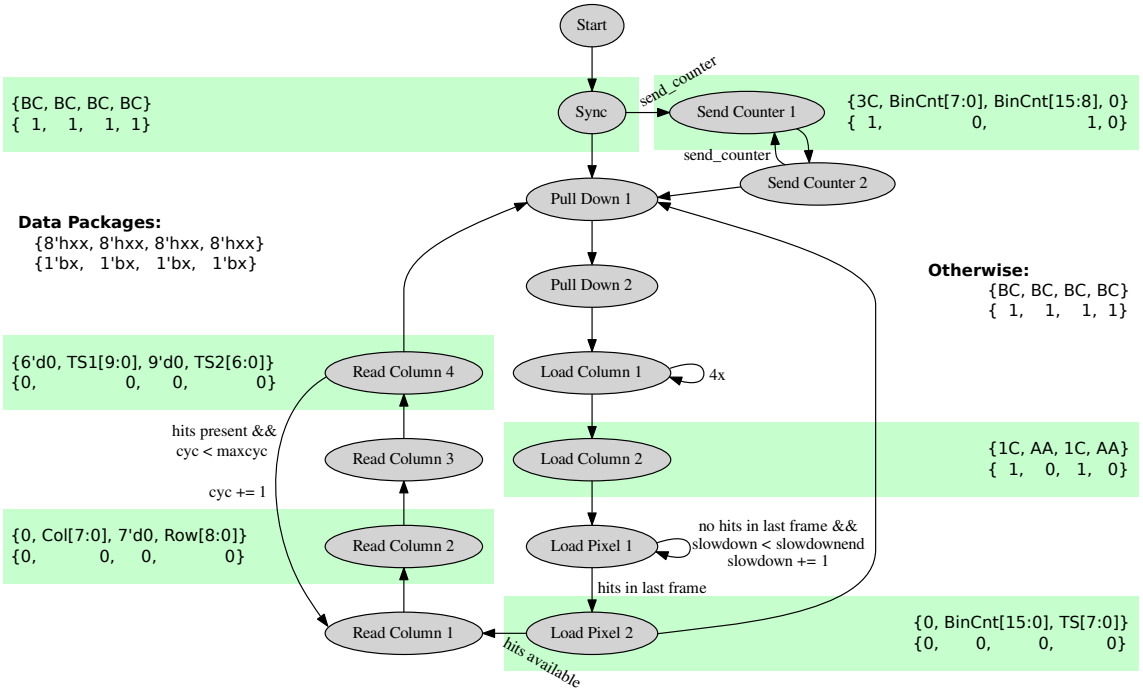


Figure 6.12: The readout of the detected signals on ATLASPix3 in hit-driven readout mode is performed by an on-chip state machine with the depicted structure: The states appear twice each to prevent data corruption due to hold violations. The first of each set of two activates a signal, whereas the second one deactivates it again so that there is no overlap of the signals activated by the states. Per write cycle, four sets of nine bits are written to the serial output. The states with payload data are marked with the green background and the data contained on the side.

after the Load Column operation in the state Load Pixel 2, the state machine directly jumps back to the Pull Down states.

6.1.7 Triggered Readout

The RCU for triggered readout is more complex than the one for the hit-driven readout and not just a simple adaption: On one hand, the state machine is split into two independent parts and the data serialiser and encoder is preceded by a FIFO to optimise bandwidth utilisation of the data link. This FIFO is also referred to as Aurora FIFO as for the name of the encoding. The encoding uses a 64b/66b scheme, reducing the overhead compared to hit-driven readout. On the other hand, grouping of hit datasets into events changes the data structure.

The first part of the state machine moves the hit datasets from the hit buffers to the CAM. The second part is responsible of grouping the datasets into events and sending them out. A visualisation of the state machines with highlighted states generating data for the readout is shown in figure 6.13.

To sort the datasets, the timestamp delayed by the trigger delay setting is stored in a FIFO, called trigger table. From this FIFO, the state machine loads one entry to access all datasets that have a matching leading edge timestamp in the CAM. These datasets are then read out to the Aurora FIFO. After moving all datasets for one event from the CAM to the Aurora FIFO, the next timestamp is loaded from the trigger table and the process is repeated.

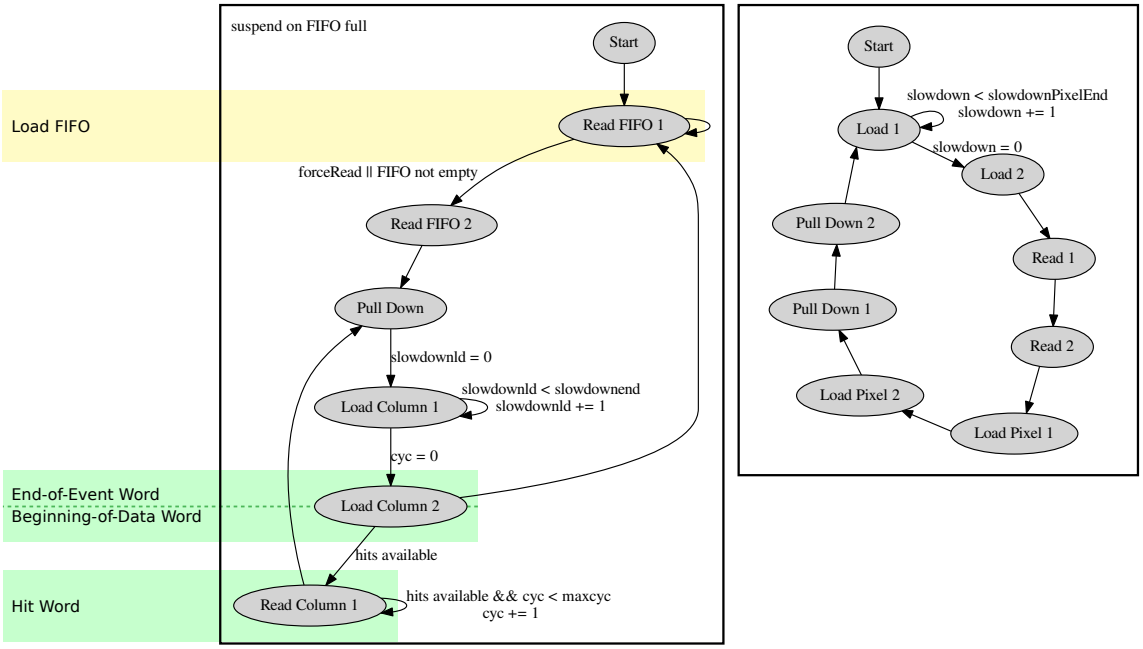


Figure 6.13: The readout state machine for triggered readout of ATLASPix3 consists of several parts operating different structures: The state machine in the right box is driving the transmission of the hits from the hit buffers to the trigger buffers. The state machine in the left box is transmitting the data from the trigger buffers to the Aurora FIFO sorting the hits into events. The data generating states are marked with green boxes including the data transmitted. From the state Load FIFO 1, the next entry from the trigger table is loaded. Also, the left state machine is suspended on a full Aurora FIFO to prevent data loss.

The implementation with the Aurora FIFO enables the state machine to only send data when available as visible in the figure. On an empty Aurora FIFO, idle words can be sent. Also, the RCU can be configured to send out other debug data as configuration read-back or user defined data words.

In order to fit the information of a dataset into 32 bit, the grouping into events has to be used: The pixel address requires 17 bit as the two timestamps, exceeding the 32 bit by two bits. Therefore, the timestamp matching the event timestamp is only sent once per event in the end-of-event word. In the beginning-of-event word sent on every loading of datasets into the EoC buffers, debug information is contained. For encoding, two of the 32 bit data words are combined to form a 64 bit data word. If an event has an odd number of data words a spacer word is added to fill the last encoder package.

In the design of this structure, ROME has been used as described in section 5.5. The results justify for example the increased complexity of the data words by pointing out the necessity of the transition to meet the requirements of the inner tracker upgrade of ATLAS.

6.1.8 Bias Block

The bias block contains digital-to-analogue converters (DACs) for currents used on the detector. Each DAC has a resolution of 6 bit which drive transistors of different width that conduct current from supply voltage VDD to the output of the DAC or to its inverted output. The transistor width is doubled for each bit starting from the least significant bit.

Since either the transistor for the output or the transistor for the inverted output is driven, the current consumption of the DAC is independent from its setting. A changed current

consumption can shift voltage levels on the ASIC possibly reducing detector performance. Consequently, a current consumption independent from the setting leads to less side effects of the settings of the bias DACs.

In addition to the bits for the DACs, the bias block also contains an "unlock code". If this code is not set to the defined pattern, the whole block is switched off. This ensures that the chip is put in a safe state on failed configuration.

6.1.9 Regulators

To reduce the number of necessary supply voltages, ATLASPix3 implements regulators. Using current inputs, the number of necessary supply voltages can be further reduced in detector systems applying serial powering.

ATLASPix3 features four regulators, two for the supply voltage and two low-dropout regulators for lower voltages used on-chip. These regulator circuits drive large transistors distributed over the width of the die to distribute the heat dissipated. They are designed in a way that their use is optional. This means, that the output of all regulators is put on output pads, close to the power inputs for the detector itself. To use the regulators, the output pads of the regulators are connected to the respective power input pads for the rest of the detector.

For serial powering, two regulators are designed to receive a current from which the input side voltage is to be used as gate voltage $\mathsf{V}_{\mathsf{gate}}$ and the output side voltage as supply voltage for the analogue and digital parts of the detector VDDA and VDDD. From these output side voltages, the two low-dropout regulators generate the lower amplifier supply voltage VSSA and the signal transmission voltage $\mathsf{V}_{\mathsf{minus}}$.

The choice to drive the latter two regulators from the output of the first one is to reduce the number on input voltages or currents. With this design choice, the low-dropout regulator [TI 06] completes the structure: The abscence of switching noise and no necessity of inductivities make them suitable for the detector application: The insensitive area is kept small and the decreased noise compared to other regulator implementations enables the detection of smaller signals improving detector performance.

6.1.10 Sensor Layout

The layout of ATLASPix3 is prepared for usage in building of larger detector modules. This means that pads are only present on one edge and the sensor is mountable side-by-side with the next sensors. This feature is also referred to as being 3-side buttable. Also, the matrix – the sensitive part of the detector – reaches these three edges minimising the dead area between adjacent detectors. This design is displayed in figure 6.1: The different colour of the periphery is only visible at the bottom where also the pads are located.

The largest part of the insensitive periphery is occupied by the readout structures: hit buffers, CAM and EoC buffers. Their arrangement is visible in figure 6.14: Below the matrix at the top of the enlarged part, the column and row configuration shift registers are located and below them, the hit buffers, EoC buffers for hit-driven readout, CAM and EoC buffers for triggered readout follow. Of the 2 mm height of the periphery, 1.38 mm are occupied by the hit buffers and CAM. Below the EoC buffers for the triggered readout, the RCUs are located. Inside the triggered RCU, also the location of the trigger table is hinted.

The space usage of this lowest band of the periphery is shown with more detail in figure 6.15: This band contains the connection pads distributed over the full width of the layout. In the gaps between the pad groups and above the pads, the registers and DACs for the DAC register, the VDAC register as well as the config register and the RCUs are placed. The remaining space is filled up with the regulators and the power transistors of the output stages of the regulators.

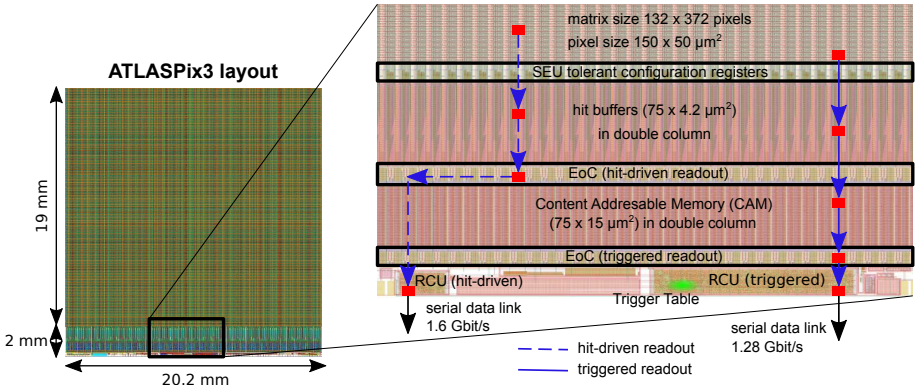


Figure 6.14: The readout buffers and control units are located in the periphery of ATLASPix3. The hit buffers have a one-to-one connection to the pixels of a column and are arranged in double columns. From there, the hit data is transmitted to the EoC buffers for hit-driven readout or to the trigger buffers which are implemented as CAB for triggered readout. In case of hit-driven readout, the hits are then transmitted to the RCU and from there written out serialised over the data link at up to 1.6 Gbit/s. For triggered readout, the hits in the CABs are selected for readout with information from the trigger table FIFO before being loaded to the EoC buffers for triggered readout and sent out via a different serialiser at up to 1.28 Gbit/s. After [Pra20]

6.1.11 Design Modifications for ATLASPix3.1

For a second production run in December 2020, several small changes in the metal layers have been made to the design. The limitation to changes in the metal layers reduces the number of changed masks which reduces the price for the production run. These changes were triggered by measurements that are part of this thesis.

The biasing structure of the pixels has been modified to remove a potential cause of early breakdown. With this, the matrix and pixel substrate biasing structures have been separated and are to be supplied with different voltages to shape the electric field.

Changes inside the pixel reduced the input capacity of the amplifier from 250 fF to 120 fF according to simulations. This results in larger signals reducing the influence of time-walk on timing.

Capacities have been added to the control signals of the supply voltage regulators to avoid oscillations. Measurements on the regulators of the initial version of ATLASPix3 have shown that the regulators oscillate without externally added capacities.

With the restructured metal layers above the pixels, the shielding of the test signal injection line was weakened resulting in signal generation in all pixels of the columns the injection goes to. Since it is not the direct transmission as with the circuit intended for it, but as cross-talk, the signals in other pixels are significantly smaller than the signal generated in the configured pixel.

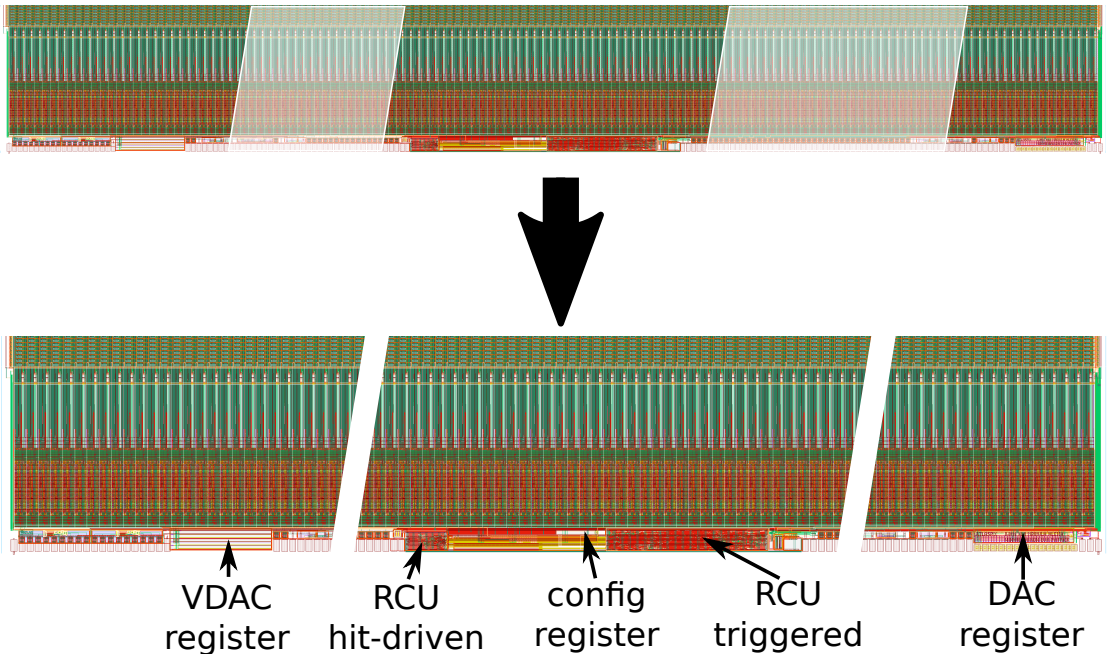


Figure 6.15: Apart from the readout elements shown in figure 6.14, the configuration registers biasing structures and regulators are located in the periphery. These elements fill the space between the pads and the readout structures enabling a little insensitive area as possible. The regulators for V_{DD} and V_{Gate} are located at the left end above the pads. The first gap in the pads is filled with the voltage DACs. The other large gap in the pads is filled by the two control units for hit-driven (left) and triggered (right) readout. In between large transistors driving the current with the control signals from the regulators and the global configuration register are located. At the right end, probing pads for bias voltages and the DAC register with bias generation are located.

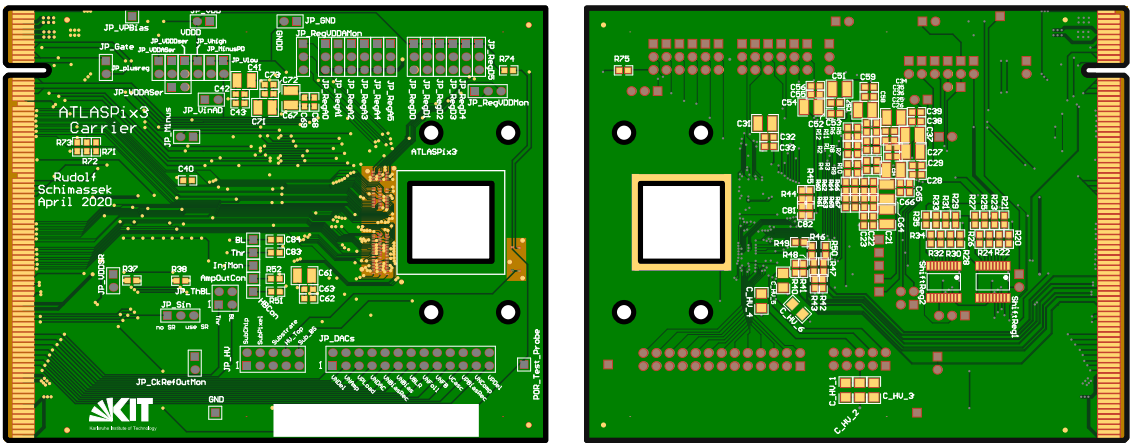
6.2 The ATLASPix3 Measurement Setup

The ATLASPix3 integrated sensor is the target around which the GECCO system has been developed. As a consequence, it is the first ASIC measured with the system described in section 4.1.

The GECCO board is equipped with a ConfigCard and optionally with two VoltageCards for overwriting bias voltages generated on-chip. The printed circuit board (PCB) carrier for ATLASPix3 contains mostly passive elements as resistors for termination of differential signal lines or capacitors for voltage decoupling. It provides access to all features implemented on the ASIC with probe points. Optional features as the regulators are configurable via jumpers to be placed on pin headers.

The only active elements on the PCB – apart from the ASIC – are two shift registers that are used to reduce the number of signal lines used for serial configuration. They are placed in front of the serial data input of ATLASPix3 with the data out of the second shift register connected to the input of ATLASPix3. The load signals and write enable signals for RAM writing are driven from the shift registers. Bypassing the shift registers is possible, too, supplying the signals directly from the FPGA. In this case, the data signal is connected directly to the input pad of the ASIC.

The layout of the carrier PCB for ATLASPix3 is shown in figure 6.16: Below the ASIC (on the right in top view), a cut-out is placed in the PCB to enable measurements with particles trespassing the sensor area. The bond pad area is left without solder stop to



(a) top view (b) bottom view

Figure 6.16: The carrier circuit board provides the bond pads for the ATLASPix3 integrated sensor, configuration jumpers and decoupling of the supply voltages. In addition, probe pads for several debug signals are available. For normal operation, all necessary signals and supply power are provided via the PCIe connector at the left end of the board (top view).

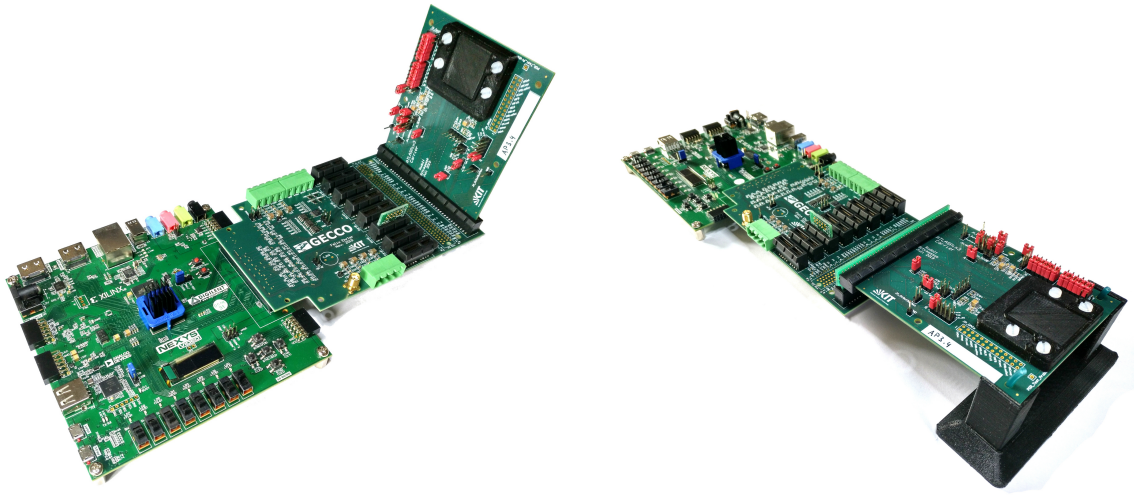


Figure 6.17: The ATLASPix3 carrier board is plugged directly into the PCIe connector on the GECCO board. Power and configuration lines are all provided via the GECCO board. The powering features are configured using jumpers on the carrier itself and the chip is protected by a 3D-printed cap with lids covering openings to the matrix. With an adapter, the carrier board can be brought to a horizontal position as visible in the right photograph.

enable repairing of bonds bonding not only on the pads but also on the lines leading to them. Above the sensor, pin headers for configuration of the VDDA and VDDD regulators are placed. The powering configuration (e.g. direct powering or via regulators) is defined with the pin headers in the top left area of the board. The bias voltages present at pads are connected to the large pin header below the ASIC footprint. Placing jumpers between upper and lower pin row, the pin is connected to the output line of the function card slots for the VoltageCards. On the left of this pin header, another one is placed for connecting the depletion voltage to the different options available on ATLASPix3. The shift registers for configuration are placed on the bottom side of the PCB close to the PCIe connector.

The configuration to use or bypass the shift registers is done via jumpers on the pin headers JP_Sin and JP_VDDSR. Next to the shift registers, but on the top side, probe points for debug signals are placed.

The sensor itself is protected by a 3D printed cover allowing for opening a window above the matrix while keeping the bonds protected. The cover is secured via the holes around the sensor footprint.

Power and data connections are all provided via the PCIe connector, only debugging signals are to be probed on the carrier PCB. The setup without power cables is shown in figure 6.17. Between the GECCO board and the carrier board, an adapter can be added to bring the carrier in a horizontal orientation.

6.3 Commissioning Measurements

Any test campaign of a new sensor starts with the most easily accessible property: power consumption. For some of these tests, it is also necessary to be able to configure the sensor. The tests for the different interfaces available on ATLASPix3 are described in appendix B. In this section, only the power consumption of the different parts on the non-irradiated sensor are discussed.

Compared to previous sensors, the power consumption on ATLASPix3 is more important as it is not a test detector for the technology, but a prototype intended for building detector systems with. Building detector systems from hundreds of such sensors, the power consumption adds up and from high integration density in a particle tracker, cooling and power losses in cables become significant. With ATLASPix3 being designed on the specifications for the ATLAS ITk upgrade, target values to achieve are defined which are to be tested.

As from the design of the HV-CMOS detectors – the category ATLASPix3 belongs to, the electronics supply voltage is used to shield the electronics from the depletion voltage, this discussion for the power consumption for them is presented first.

Providing several individual voltages to each sensor in a tracker is also a problem for which regulators are implemented on ATLASPix3. Their properties are presented next and their impact on the total power consumption is discussed. Finally, the properties of the depletion voltage are looked upon.

6.3.1 Power Consumption

For application in particle tracking, power consumption of the single detector has to be reduced as much as possible. This is due to the large number of detectors necessary on the layers of the tracker in a minimum volume. Consequently, the power density is high and space for cooling structures is limited both by the available space between the modules and the aim for minimising the amount of material inside the tracking volume. For the ATLAS inner tracker, this is a limit of 500 mW/cm^2 for the readout chip and 100 mW/cm^2 for the sensor [ATL17].

ATLASPix3 spans an active area of 384 mm^2 leading to a power consumption limit of 2.3 W combining the power dissipation limits for sensor and readout ASIC. Due to irradiation, the power consumption can increase from radiation damage itself or from different settings necessary to compensate for decreased detector performance. Therefore, power consumption needs to be lower for the non-irradiated sensor.

The total power consumption of ATLASPix3 consists of the power from the low-voltage supplies and the depletion voltage. The latter does not contribute significantly to the total

Table 6.1: For default operation, the contributions to power consumption of ATLASPix3 are listed. The area assumed for the last column is the area of the matrix at 20.2×19 mm^2 as the modules are to be placed overlapping with the periphery to avoid insensitive areas.

Input	Voltage	Current	Power	Power per Area
Amplifier Supply	1.15 V	113 mA	129 mW	34 mW/cm^2
Analogue Supply	1.80 V	184 mA	331 mW	86 mW/cm^2
Digital Supply	1.80 V	45 mA	81 mW	21 mW/cm^2
Data Link Supply	1.60 V	10 mA	16 mW	4.2 mW/cm^2
Total			580 mW	145 mW/cm^2

power consumption as the leakage current is in the order of 50 nA for the non-irradiated sensor. The values of a parameter set not optimised for a specific aspect as time resolution or minimum power consumption are listed in table 6.1. The target readout speed for these settings is 320 Mbit/s with triggered readout, which is a quarter of the design speed.

The power $\mathsf{V}_{\mathsf{minus}}$ is not included in the list as its current consumption is less than 1 mA and its contribution can be neglected.

The amplifiers in the pixels consume about 2.3 µA each for default operation settings. This value can increase by 50 % for improved time resolution bringing the total power per area to 165 mW/cm^2 which is still well below the limit proposed for the ATLAS ITk upgrade.

The power for the comparators is included in the contribution for the analogue supply voltage. The current for the comparators can be isolated by adjusting the bias DAC for the comparators and evaluates to 94 mA contributing a power consumption of 174 mW or 45 mW/cm^2.

The data link supply separates out the power for the data output links of the chip. This is the actual data link for the signal data and the reference clock output. This power contribution can be tripled for for sharper transitions to enable higher link speeds. However, the reference output clock is not necessary for operation and can be left unconnected. Then no current to drive a differential line is consumed there and the power consumed by this is cut in half. Consequently, the power is increased by 2.1 mW/cm^2 for operation at full readout speed of 1.28 Gbit/s without connecting this debug feature compared to the values in table 6.1.

The digital supply provides power to the readout control units. The value provided in the table is for a readout speed of 320 Mbit/s. The power increase for the full readout speed at 1.28 Gbit/s is about 4 mW/cm^2 compared to the values in the table.

The change in power consumption introduced by the usage of the on-chip regulators will be discussed at the end of the next section.

6.3.2 On-Chip Regulators

The regulators on ATLASPix3 are a required feature for integration in large detector systems: The design requires at least four independent low-voltage power-inputs of which some have to be fine-tuned for optimal operation for the individual samples. This is not a viable solution for systems as the fifth pixel layer of the ATLAS ITk upgrade the design of ATLASPix3 is based on: There, 7800 samples of ATLASPix3 would be necessary to populate the layer with 2600 modules. Instead, the sensors are supposed to generate the necessary voltages and currents from as few power inputs as possible. Also, concepts to

power the sensors serially are pursued to reduce the number of necessary inputs for two reasons: The necessary number of cables necessary does not fit into the detector structure and high currents over the cables lead to significant losses in the cables. Consequently, using larger voltages to provide several sensors in series promise reduced transmission losses and fewer cables. [ATL17]

ATLASPix3 implements regulators to be powered with only two input currents for low-voltage power and one depletion voltage input.

To characterise the regulators on ATLASPix3, the implementation as optional features is made use of: For the measurements, the regulator output is connected to a load resistor. This enables tests of the regulator independent from the specific load formed by the circuitry on ATLASPix3. To do so, the power inputs are connected to the designated inputs on the GECCO system. The outputs are connected to the load from the jumpers for powering the detector with the regulator outputs.

ATLASPix3 does not embed capacities to stabilise the control signals in the regulator for VDD which causes oscillations if not provided with external ones. These capacities were added in the design of ATLASPix3.1. The measurements for this regulator have been performed on a sample of ATLASPix3.1. The regulators plusreg for VSSA and minusreg for V_{minus} have not been changed for ATLASPix3.1 and were measured on ATLASPix3.

As voltage changes by 10 mV can compromise the performance of the detector, the precision of the measurement has to match this scale. From the currents drawn by the circuitry on ATLASPix3 as shown before, the ranges are derived: For the regulators generating VDD and V_{Gate}, the current drawn can be up to 500 mA, the regulator generating VSSA needs to provide at least 150 mA for operating the detector. These currents at voltages of less than 2 V result in load resistances in the order of 10 Ω. Consequently, for the required precision, the resistances of the cables, connectors and lines on the PCB have to be accounted for as well.

The circuits for the regulators for VDDA and VDDD – the analogue and digital supply voltages – are identical and therefore only the measurements on the regulator for the analogue regulator that also generates the gate voltage V_{Gate} are presented. To test the regulator, the regulator input is provided with a defined current from a source-measure unit. The output is connected to the load resistor. The generated voltages are measured as close as possible to the regulator and from combination of measurements with different loads and supplementary measurements, the resistances of the connections are determined.

Accounting for the resistances of the wires at the current input and the output of the regulator, the characteristics shown in figure 6.18: If the current provided to the regulator is smaller than the load would require for the target voltage of the regulator, the resulting voltage is smaller than the target. For currents larger than the required amount for the load, the voltages on both the input and the output side of the regulator stay constant. The resulting voltages are larger than necessary for operation of ATLASPix3 with 1.97 V for VDDA on the output side and 2.32 V for V_{Gate} on the input side. The value for VDDA can be used as it is, V_{Gate} is not drawing a significant current (< 1 mA) which can be reduced to the desired value at 2.1 V with a voltage divider.

The regulator plusreg for VSSA is powered from VDDA and consequently optionally from the output of the first regulator. To isolate the behaviour of the regulator, direct powering for VDDA is used. This regulator is configured via the VDAC register to provide an adjustable voltage level. Consequently, ATLASPix3 needs to be configured during the measurement in contrary to the previous measurement on the regulator for VDDA. Also, the input is a voltage and not a current. The same adjustable load resistor is used instead of the ASIC circuitry.

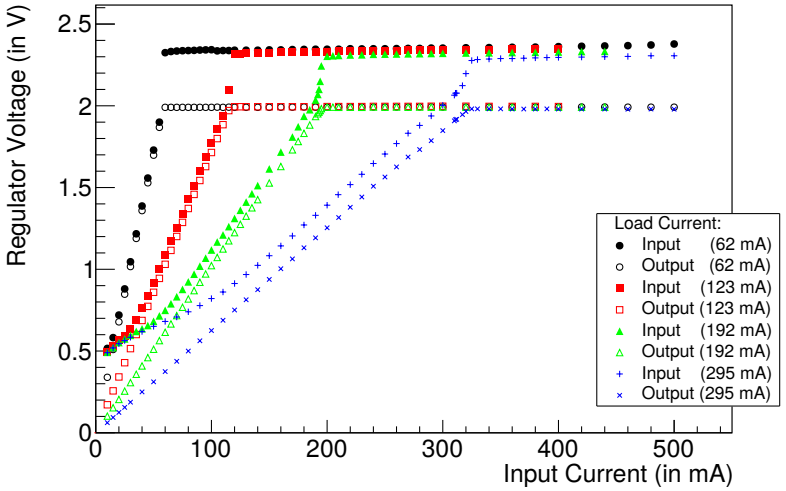


Figure 6.18: The main regulator on ATLASPix3.1 shows a stable voltage on both input and output sides supplying a current larger than the required current for the load on the output side. For the measurement, a resistive load has been used.

For different loads, the current drawn from VDDA changes as the LDO regulator does not convert the current. As for the resistive load, the current drawn via the regulator changes with the setting of the DAC adjusting the regulator. This is shown in figure 6.19. The currents rise at different slopes over the DAC setting according to the loads used. The voltages for these measurements are visualised next to it. Towards the end of the range, a saturation is visible both for the current and for the output voltage. But before this value, the behaviour is linear and for the voltage independent of the load current.

The regulator minusreg for V_{minus} is a current sink in contrary to the regulator plusreg. Furthermore, the currents to be sinked are about two orders of magnitude smaller: The typical currents for V_{minus} are $\sim 1\,\text{mA}$ compared to over 100 mA for plusreg. Therefore, the influence of voltage drops is not as severe as for the other regulators.

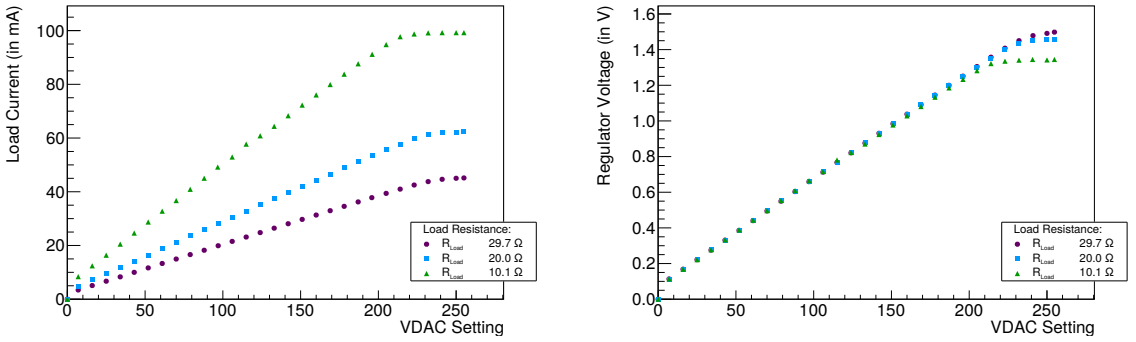
For an input current of 1 mA, the voltage output of a subset of the range of the VDAC register parameter is shown in figure 6.20. The resulting voltage is proportional to the setting of the DAC controlling the regulator.

The regulators used on ATLASPix3 have been designed to avoid generation of noise for example by using the LDO regulator. This noise can be measured for example on the amplifier output of one pixel in the row closest to the periphery or from digitisation effects as discussed in section 6.4. Both measures show no large change between operation with direct powering of VDDA and VSSA and powering of both from the regulators providing ATLASPix3 with a current for the regulator generating VDDA. V_{Gate} has been provided directly for this measurement as for the too high value provided by the regulator.

For the amplifier output, the noise measured is decreased from $\sigma_{\text{direct}} = (7.93 \pm 0.05)$ mV to $\sigma_{\text{regulator}} = (6.79 \pm 0.05)$ mV for switching from direct powering to using the regulators for VDDA and VSSA. Compared to the maximum amplitude of the amplifier of about 350 mV, this change is 0.3 %. Even when compared to small signals as from ^{55}Fe decays (see section 6.6.2) or MIPs, the change is at a scale of 1 %.

For the estimation from signal digitisation, a similar result can be observed: The noise estimate decreases by 6 % switching from direct powering to the regulators.

Therefore, it can be concluded that the regulators do not significantly impair the detector performance. Similar changes as in the transition from direct powering to the use of the



(a) Load current (b) Output voltage

Figure 6.19: The output current of the regulator changes with the load connected to it and the output voltage scales linearly with the setting of the VDAC parameter adjusting it. The output voltage is independent of the load current before saturation. The linear increase of the current with the setting of the DAC is due to the resistive load used which draws more current with the increasing voltage for the higher DAC settings.

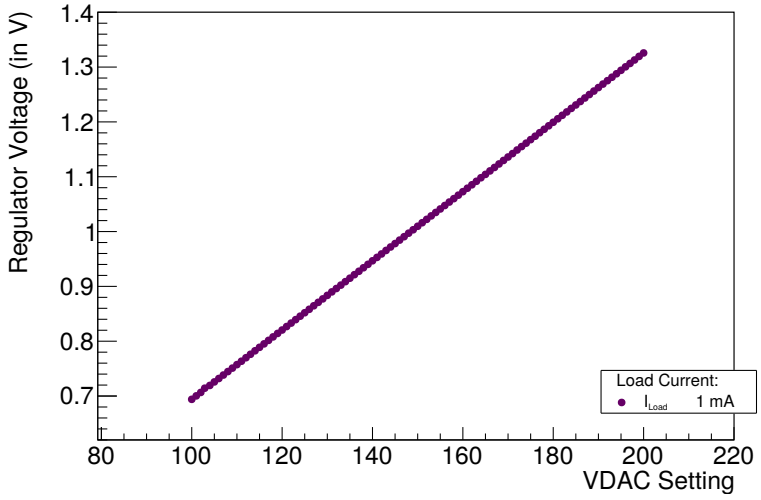


Figure 6.20: The voltage for a current of 1 mA sinked by the minusreg regulator is linear over the setting of the DAC controlling the regulator.

regulators can be obtained by changes of the settings and supply voltages. Hence, to quantify the change, extensive parameter scans would be necessary to be conducted to ensure exclude such changes from the differences.

The regulators have an inpact on power consumption of ATLASPix3. Since the lower voltages as the amplifier supply voltage VSSA (1.15 V) are derived from the higher voltages, the additional power is consumed in the regulators: For example, the regulators for VSSA and V_{minus} are supplied with the supply voltage of 1.8 V meaning that the difference to their output is consumed in the regulator. Similarily, the power consumed from the full supply voltage at 1.8 V has to be accounted for when using the regulator for V_{Gate} and VDD itself. With the input voltage of the VDD regulator at 2.35 V as highest voltage present on the module, the power consumption of ATLASPix3 increases to 215 mW/cm^2 for the same ASIC settings as used for the values in table 6.1. This is an increase of 48 % over the configuration without the regulators. But the power consumption requirement for ATLAS

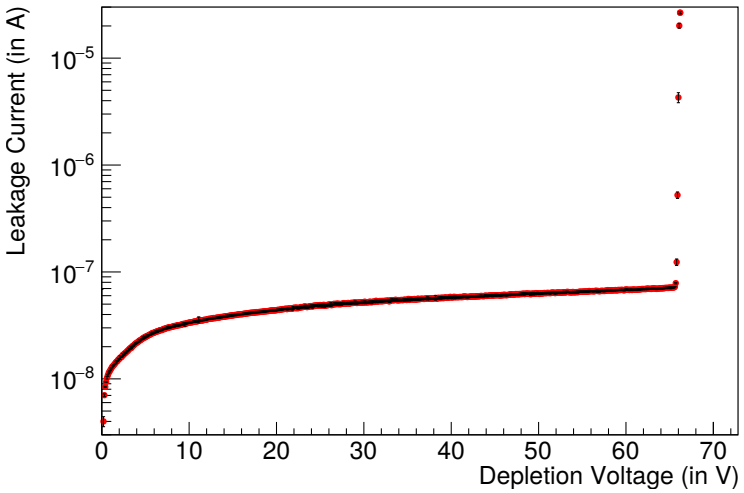


Figure 6.21: The leakage current for the reverse bias voltage for depleting the sensor diodes shows a breakdown voltage of about 65 V. The leakage current per area for operation at 20 V is 1.2 pA per Pixel and 160 pA/mm^2.

ITk upgrade is still met.

6.3.3 Depletion Voltage

ATLASPix3 has been designed for depletion voltages up to 120 V. This voltage depletes the volume below the pixel matrix to improve particle detection. From the concept of the sensor with the supply voltage shielding the in-pixel electronics from the depletion voltage, the chip has to be powered during any time when a depletion voltage is connected.

Careful design of the guard ring structures around the individual pixels and around the matrix as a whole prevents currents different than the leakage through the sensor bulk.

A measurement of the current for the depletion voltage is shown in figure 6.21 for a measurement at room temperature. The leakage current at 20 V is about 60 nA for the whole matrix. This equals a leakage current of 1.2 pA per pixel or 160 pA/mm^2. The exponential increase of the leakage current – called breakdown – starts at the breakdown voltage at 65 V. This is about half the design target. For the revised version ATLASPix3.1, this voltage did not change despite removal of possible causes for an early breakdown in the metal layers above the pixel (see figure I.9).

A modified version of the pixel from ATLASPix3 has been used for the HitPix integrated sensor [Web21]. There, a breakdown voltage of more than 100 V has been measured. Consequently, a geometry change from ATLASPix3 to HitPix included removal of the cause of the early breakdown.

In the pixel, the n-well is surrounded by the deep n-well which is generated by diffusion. This leads to round corners of the deep n-well preventing high fields at the corners. In the periphery however, structures without diffusion generated wells exist that can generate such field spikes: There, the n-well has kinks close to a p-well bias with the depletion voltage. While the outward facing corners of the n-well are rounded off by the deep n-well, the p-well is not surrounded by a diffusion-generated structure. At outward facing corners of this p-well, the field will be increased and can lead to the early breakdown measured.

An example of this structure is shown in figure 6.22: The n-well is biased to VDD at 2.8 V while the p-well is at the negative potential of the depletion voltage. The inward

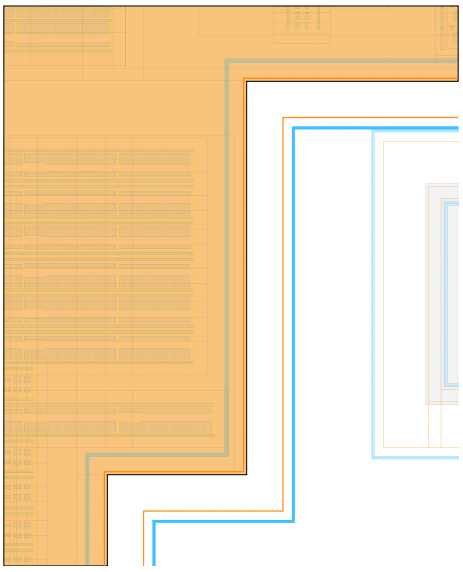


Figure 6.22: The deep n-well (marked by the orange area) of the HV-CMOS process is created via diffusion. Consequently, the corners are not sharp but round. At the periphery, some p-wells (blue outline) are not enclosed in a diffusion-generated well. The kinks creating corners pointing towards the gap create points with high field. Those points can lead to the early breakdown observed on ATLASPix3. The shown structure is one of several examples at the bottom of the periphery – in this case located next to the triggered RCU.

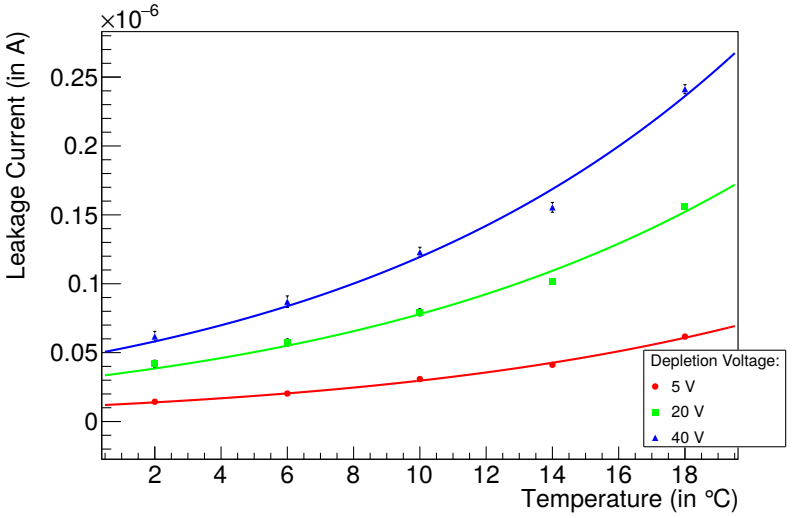


Figure 6.23: The leakage current dependency for different depletion voltages on ATLASPix3 matches with the expectation for charge generation as main source.

facing corner of the p-well does not create a field spike because of the diffusion of the deep n-well. But the two other corners in the figure lead to areas of high field that can cause the breakdown as measured at a depletion voltage of 65 V.

The temperature depencence of the leakage current follows the expected behaviour from charge generation as described in [Chi13]:

$$I(T) \sim T^2 \cdot \log\left(-\frac{E_{\text{eff}}}{2\,k_B T}\right) \tag{6.1}$$

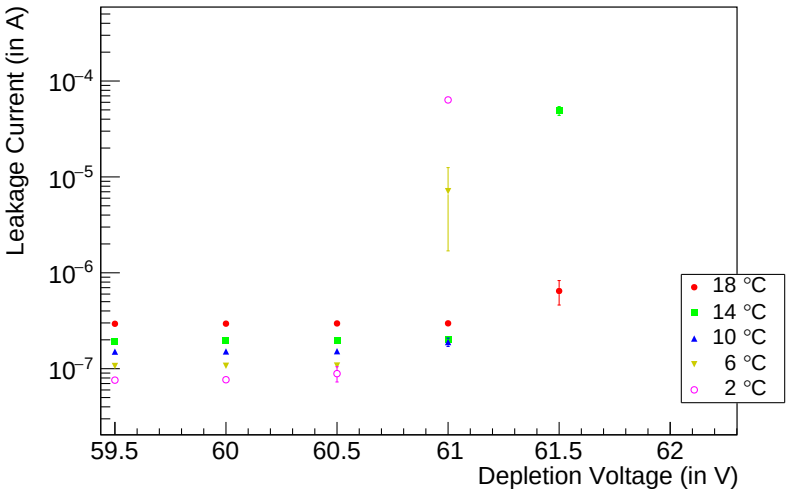


Figure 6.24: The breakdown voltage shifts to smaller voltages with decreasing temperature. This change becomes visible comparing the extremal temperatures where the 2 °C data line is already in breakdown whereas it is still 0.5 V before the data line at 18 °C starts to rise.

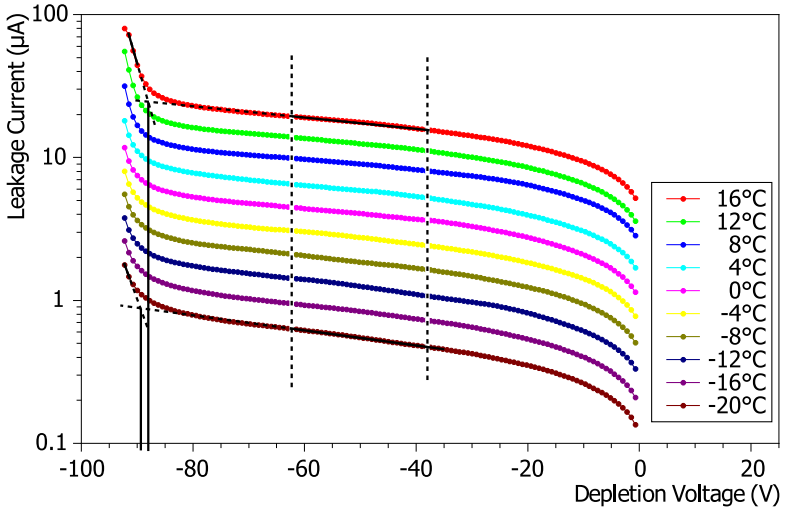


Figure 6.25: The Temperature dependence of the leakage current of the HitPix detector shows a reversed breakdown characteristics compared to ATLASPix3: The breakdown voltage decreases with increasing temperature. The breakdown voltage is indicated for the extremal temperatures in the plot to highlight the change. From [Web21]

where $b = E_{\text{eff}}/2\,k_B$ is the constant fitted to the measured data and E_{eff} denotes an effective energy. The measured dependency is visualised in figure 6.23. The found value for $E_{\text{eff}} = (1.12 \pm 0.04)$ eV is compatible with the value for mid-gap generation of 1.21 eV [Chi13].

A feature observed in the temperature dependence of the depletion voltage is a shift of the breakdown voltage. It moves to smaller values for lower temperatures as visible in figure 6.24. This change can be explained by more inelastic scattering at higher temperatures which increase the amount of energy to be transferred to a charge carrier to maintain the avalanche. This is a hint that the breakdown is caused by an avalanche breakdown as opposed to a breakdown caused by band-to-band tunneling. For tunneling, the breakdown voltage would

move to lower values for higher temperatures because of the larger energies of the charge carriers.

Typically, band-to-band tunneling is expected at lower external voltages than the avalanche effect. The dominance of the avalanche effect in the breakdown matches the explanation from before that the shape of the wells in the periphery creates an area with enlarged electric field where the avalanche limit is reached earlier.

On the HitPix (see [Web21]), the pixel structure used is an adaption of the pixel of ATLASPix3. But there, the depletion voltage is decreasing with higher temperatures as expected for the band-to-band tunneling breakdown. This behaviour is shown in figure 6.25.

6.4 Threshold Tuning

With the in-pixel tuning circuits for threshold adjustment, differences in detection efficiency between the pixels can be compensated. As for the large number of pixels, the measurement process is structured and automatised.

In the following, the basic measurement to determine the detection threshold – the S-curve – is explained and the influence of the TDACs and control parameters is described. Following this, the tuning procedure of the matrix is presented with an example result. As last part, the noise estimate that is obtained by the S-curve measurement is discussed.

6.4.1 Measurement Methods and Circuit Characterisation

To determine the detection threshold of a pixel, several options are available: For a fixed signal size, the threshold setting is varied or for a fixed threshold setting, the signal size is changed. With the comparator being an analogue circuit, the result can be affected by non-linearities of the comparator. ATLASPix3 implements a circuit to inject an adjustable amount of charge into the pixels. Therefore, the latter method is used here.

The data is taken as the number of detected signals per signal size where the total number of injected signals is known. In an ideal system, this would result in a step function of the detection efficiency curve with no detected signals for signals below the threshold and all signals detected above it.

In the real system however, the injected signal is altered by noise from the leakage current and the electronics. A signal slightly smaller than the threshold hence can be lifted above the threshold by noise leading to its detection. On the other hand, a signal slightly larger than the threshold can be diminished by noise leading to no detection by the comparator. This leads to a smearing out of the step function of the ideal system. Such a measurement scanning the amount of charge controlled by a voltage leads to an efficiency curve as shown in figure 6.26. The width of the transition from zero efficiency to full efficiency is a measure of the noise in the system as the detection result can only be altered if the maximum signal level is in proximity of the threshold level where proximity is the amplitude of the noise at the comparator input. With the complexity of the circuits involved in the detection, the noise can be assumed Gaussian. Consequently, the efficiency curve can be fitted with a Gaussian error function as the integral over a Gaussian distribution. Such a scan is referred to as “S-curve” due to its shape.

The injection voltage used for the plot can be calibrated and converted to an electron equivalent using radiation of known energy, enabling to convert the noise in the system to a charge equivalent.

Making use of S-curve scans, the effect of the in-pixel TDACs can be evaluated. For this, the detection threshold for a larger number of pixels is used and for each setting of the

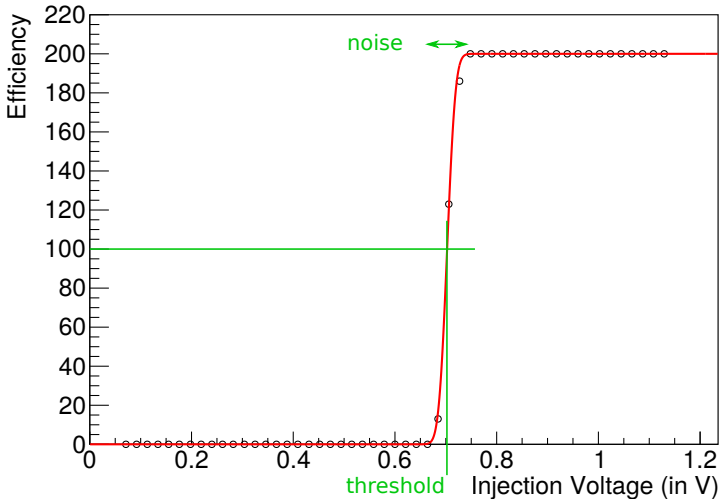


Figure 6.26: To determine the detection threshold, test signals of different size are injected into the pixel and the number of detected signals is counted. The step function is smeared out by noise and – assuming a Gaussian distribution of the noise – can be fitted with a Gaussian error function. The symmetry point of the fit function gives the detection threshold and the transition width is linked to the noise present in the system.

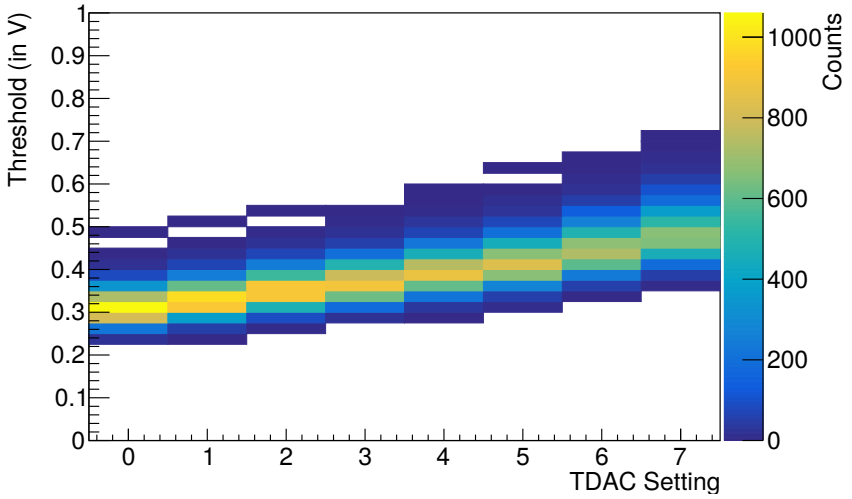
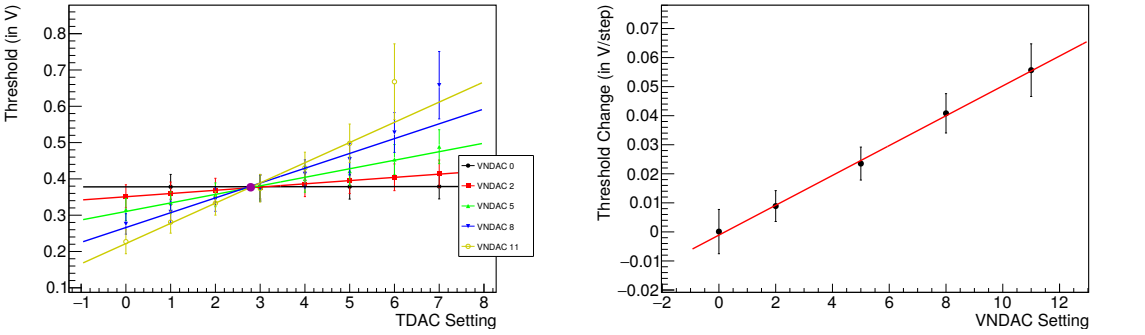


Figure 6.27: The detection threshold changes linearly with the TDAC setting over the full range. The lower four settings (0-3) decrease the threshold while the upper four settings (4-7) increase it with respect to the globally set threshold. The differences between the thresholds can be minimised if the threshold shift from TDAC = 0 to TDAC = 7 equals the width of the threshold distribution.

3 bit TDAC, the S-curve measurement is repeated. The expectation is a monotonous shift of the detection threshold to higher values with larger values of the TDAC setting. The result of such a scan is shown in figure 6.27. For this figure, 3300 pixels on one sample have been measured. The monotonous behaviour enables not only the intended use for detection threshold compensation, but also the implementation of optimisations in the measurement procedure to reduce the measurement time.

The other parameter involved in the adjustment of the threshold is the current that is



(a) Mean Thresholds for different VNDAC settings

(b) Slope of the threshold gradient over the VNDAC setting

Figure 6.28: The step size of the TDACs is determined by the bias current controlled by the VNDAC setting. A lower setting results in a lower current and hence in a smaller step width. By measuring the detection threshold for different VNDAC settings and TDAC values on several pixels, the mean detection threshold can be extracted for each VNDAC-TDAC pair (see (a)). These data rows are fitted and the resulting slopes are presented in (b). The step dependency on VNDAC is linear on the measured range.

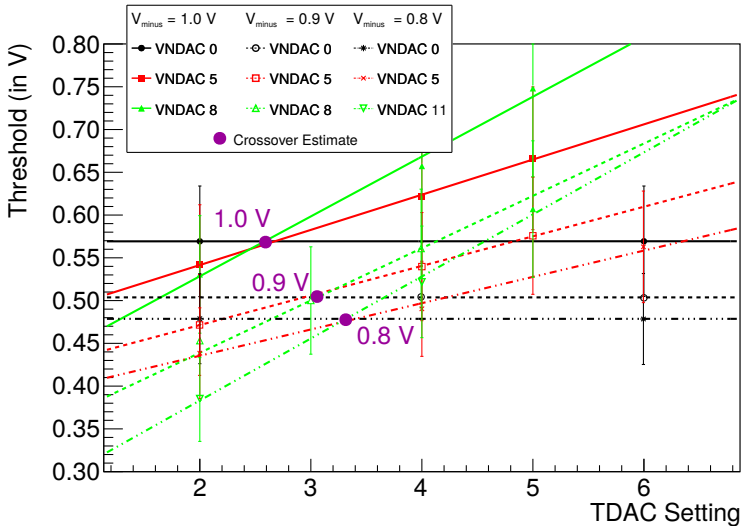


Figure 6.29: The cross-over point for the TDAC settings is altered with the value of V_{minus}. Also the absolute detection threshold level is changed with this voltage.

sinked by the TDAC circuits from the comparator. This current controls the change of the detection threshold per step of the TDAC value. To test this influence, the scan over the TDAC settings is repeated for different current settings of the TDACs.

For visualisation, the detection threshold distribution per TDAC setting is fitted with a Gaussian distribution and the mean and standard deviation replace the colour-coded histogram. For five different settings, the detection threshold dispersion is shown in figure 6.28: For a value of the DAC controlling the current for the TDACs of zero, no current is sinked from the comparator branches and the detection threshold remains unchanged for the different TDAC settings. For the settings larger than zero, the slope of the dispersion increases with the setting as expected. The connection of the slope is linear to the setting of the responsible DAC as visible in subfigure (b). The differential design is

visible as cross-over point of the different threshold dispersion curves in the middle of the range.

The cross-over point at a TDAC setting equivalent of 2.8 is due to the setting of $\mathsf{V}_{\mathsf{minus}}$ differing from the design value used for the dimensioning. The amplification factor in the order of ~ 10 of the comparator results in a change of the working point with a change of $\mathsf{V}_{\mathsf{minus}}$. This running of the cross-over point is shown in figure 6.29: With higher setting of $\mathsf{V}_{\mathsf{minus}}$, the cross-over point is shifted to lower settings of the TDAC. Furthermore, the detection threshold increases with the setting of $\mathsf{V}_{\mathsf{minus}}$. The cross-over point for $\mathsf{V}_{\mathsf{minus}} = 0.8\,\mathrm{V}$ is with a TDAC setting of 3.31 close to the design target of 3.5.

6.4.2 Matrix Tuning

Optimising the detection thresholds of the matrix requires several steps preparing the actual optimisation: Firstly, the tuning target needs to be set. Secondly, the global threshold needs to be set appropriately to the tuning target. Thirdly, the step size determined by the VNDAC setting has to be adapted to the actual spread of the detection thresholds.

The target for the detection threshold can be chosen manually from external requirements, or it can be set automatically by the software: To determine the tuning target, the software measures a fraction of the matrix at the highest TDAC setting and selects the target as $\mu - 3\,\sigma$ of the Gaussian fit to the distribution of the detection threshold distribution found in the measurement (μ denotes the mean of the Gaussian distribution and σ its standard deviation). The reason is that from the highest TDAC setting, the threshold can only be lowered not increased and by picking the upper end of the range, the risk of setting a pixel below the noise level is reduced.

If the target detection threshold is chosen manually, the global threshold setting is to be adjusted to result in detection thresholds measured for middle settings of the TDACs on average at the target value.

The step size setting can be adjusted after measuring the threshold distribution at different TDAC settings: The target for this is to set the VNDAC parameter just as large that the difference in the mean values of the distributions of the detection thresholds at maximum and minimum setting equals the distribution width of each of the distributions. This aims for usage of the whole range of the TDAC settings maximising the available granularity.

With the S-curve being measured many times for the tuning, it is optimised to adjust the step width according to the requirements: High granularity is only required at the transition, the intervals at zero detection efficiency and full efficiency do not require many points measured. For this reason, the step size of the test signal is tripled compared to the set value. If a transition is detected, the skipped points are measured, too.

The tuning of the matrix itself is then done with a binary search on the TDAC setting comparing the found detection threshold with the target value: Starting from the maximum value at 7, the most-significant bit is cleared for measuring the first S-curve. Depending on the resulting detection threshold being too low the bit is set again or left cleared for values higher than the target. The same procedure of clearing a bit, measuring an S-curve and setting the bit again for detection thresholds lower than the target is repeated for the other two bits. If a pixel has a setting of seven after the third S-curve, the fourth S-curve with this setting is measured, too. For all other outcomes of the TDAC value after the three S-curves, the fourth S-curve for a TDAC value of 7 is not necessary.

Tests have shown that the signal size of the charge injection is affected when measuring too many pixels at once. For this reason, the number of pixels measured at once is chosen as 22 as it is a sixth of the number of columns on ATLASPix3.

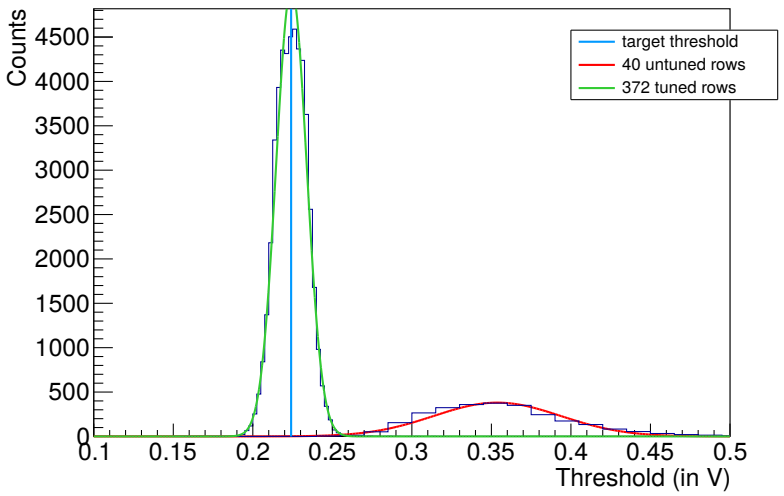


Figure 6.30: From the untuned threshold distribution measured on a subset of the pixels (shown in red), the tuning target is derived as $\mu - 3\,\sigma$ of the gaussian distribution. This measurement is performed at the highest threshold setting for these pixels. The pixels TDAC settings are then scanned to find the value for which the measured detection threshold matches the target value best. The result is typically a width reduction of a factor of four with a mean of the distribution at the lower end of the initial distribution. [SAA+21]

While one group of pixels is measured, all other pixels remain at a TDAC setting of 7 to avoid interference with the measured pixels. After finishing a group, the TDAC values for these pixels are set back to 7 before measuring the next group.

After measuring the S-curves for a group at one step, the S-curves are fitted and after the last step the setting best matching the target value is chosen for each pixel. This is not necessarily the last measurement: If the second to last measurement is just below the target (e.g. with a TDAC setting of 5), the last measurement (with a TDAC setting of 6 in the example) will result in a detection threshold above the target. However, the distance of this last value to the target can be larger than the one before.

For one ATLASPix3 sample, the resulting threshold distribution and the estimate from the untrimmed detection threshold distribution are shown in figure 6.30: The integral of the untuned distribution is larger as the whole matrix has been measured for it while the untuned distribution only contains data for 40 of 372 rows. The tuning target derived from the untuned distribution is shown, too. The value of the target as well as the mean of the resulting distribution is equivalent of a charge of $1830\,\mathrm{e}^-$ with a standard deviation of $70\,\mathrm{e}^-$ for the latter case. The calibration for this conversion will be described in section 6.6.2.

The distribution of the TDAC settings used for the result is shown in figure 6.31: The distribution shows a Gaussian shape with a width spanning about the whole range. With a mean of 2.7, it indicates that the global threshold setting could have been slightly lower shifting the optimum TDAC settings up. Also – as the TDAC setting of 7 is almost unused, the step size could be decreased. The value of 8 represents the pixels that have been deactivated because no adequate setting was found.

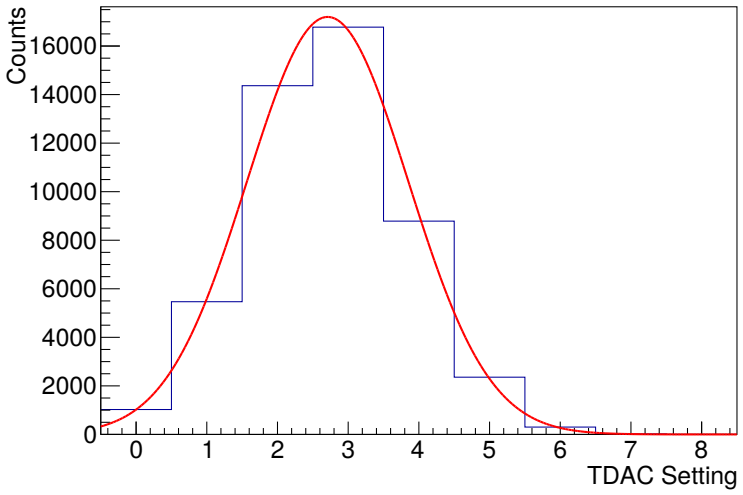


Figure 6.31: If the global threshold matches the target threshold for the tuning, the distribution of the used TDAC values for all pixels resembles a Gaussian distribution around 3.5. For lower target thresholds this distribution is shifted to lower TDAC settings and to higher values for higher target thresholds. In this case, the target threshold has been too low for the global threshold setting. The width of the distribution is determined by the scaling of the TDACs and is adjusted with the VNDAC setting of the chip. Higher VNDAC settings make the distribution narrower, lower settings widen the distribution. [SAA+21]

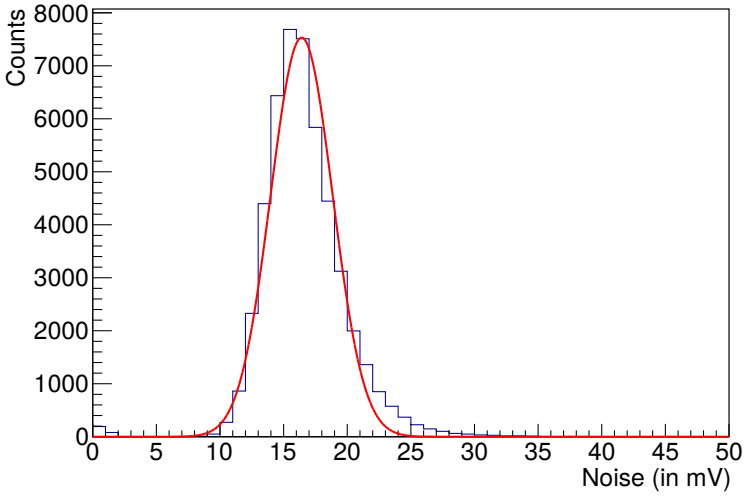


Figure 6.32: Since the noise measured with the S-curves depends on signal fluctuations around the detection threshold and not on the detection of noise signals without stimulus, the distribution of the noise does not change significantly with the tuning process. For the same reason, this voltage does not represent the actual amplitude of the noise a the comparator input but is a charge equivalent converted to a voltage by the injection circuit. [SAA+21]

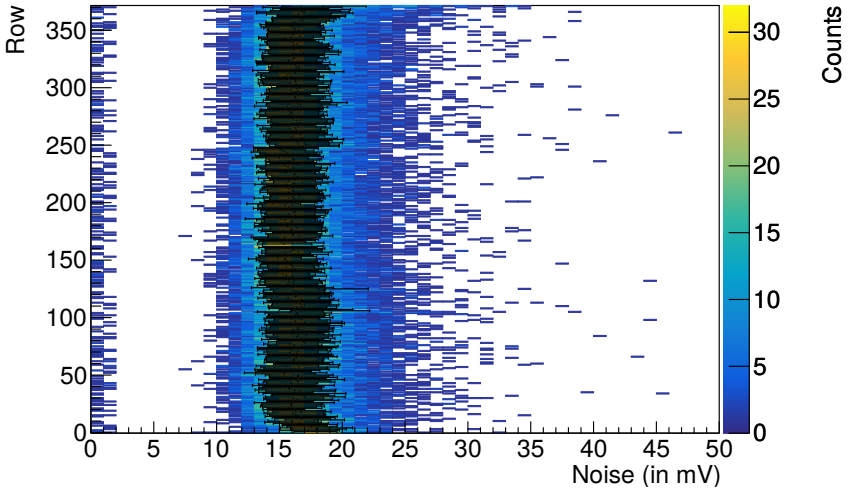


Figure 6.33: The uppermost and lowermost rows of the matrix show increased noise levels compared to the rest of the matrix. This contributes to the non-Gaussian fraction of the distribution in figure 6.32.

6.4.3 Noise Estimation

From the same measurement, the noise in the pixels can be estimated from the transition widths of the S-curves. The distribution of the noise in the pixels matching the results from the last section is shown in figure 6.32. The voltage shown on the X axis is a voltage equivalent for the charge injected, not the actual amplitude of the signal at the comparator. Using the conversion from section 6.6.2, the mean of the Gaussian fit equals $124\,\mathrm{e}^-$ with a standard deviation of $19\,\mathrm{e}^-$.

The deviation from the Gaussian shape includes contributions from a correlation of the noise to the row in the matrix. This correlation is shown in figure 6.33: The histogram with the noise value correlation to the row is overlayed with data points of Gaussian fit parameters showing the mean and standard deviation. At the top of the matrix above row 350 and at the bottom in the rows below 30, the noise is increased. Furthermore, the layout of the pixel to hit buffer connection is visible: At the row 124, a small drop of the noise distribution is visible, and at row 248 the distributions jump back up again.

6.5 Time Resolution Measurements

Time resolution on a monolithic HV-CMOS sensor is limited by several contributions:

- One is the signal distribution on the ASIC,
- another the finite signal rise time resulting in time-walk.

This section is addressing the signal distribution first, before discussing the calibration of the amplitude measurement to compensate for time-walk.

Making use of the external timing signal for the test signal injection, the timing of the pixels can be probed. Compared to other measurement methods as TCT and source measurements with time reference, the injection offers simple operation. The TCT measurements takes time involving mechanical movements of the laser and the shielding of the matrix by the metal layers leaving only small areas where the pixels can actually be tested with a laser. The source measurement requires the rate to be small enough to separate the signals from

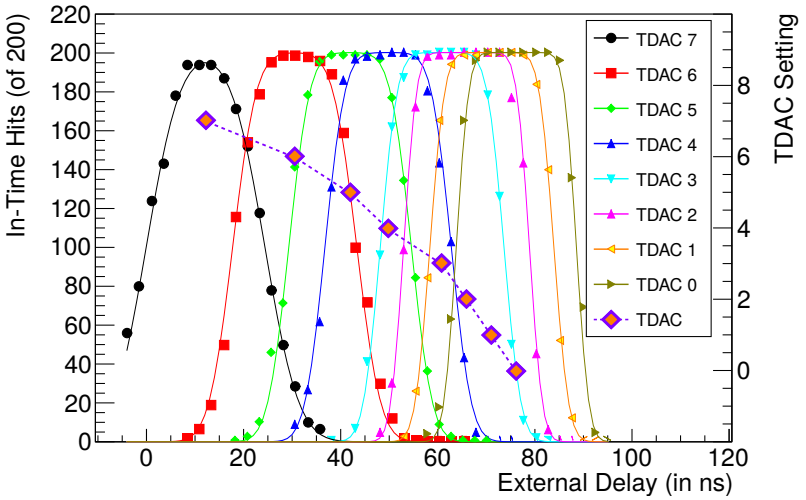


Figure 6.34: With an external time reference, test signals are generated at increasing delays from the timestamp epoch of the chip. From the detected hits, the fraction with one timestamp value is extracted. This leads to a pulse-shaped function. Time jitter leads to a smearing out of the edges of the pulse. The finite rise time of the amplifier leads to a timing shift with the change of the TDAC value. This shift can be measured and leads to a characteristics as shown with the orange squares with purple border. [SAA+21]

the time reference to be able to match them to the signals from the sensor which increases the measurement time.

The injection has the advantages of a controlled environment. It comes at the disadvantage of a bias on the measured value due to the signal distribution over the sensor. But since this is a static system, the bias can be corrected for in a dedicated measurement with a source or a TCT setup.

With synchronisation to the reference clock of the DUT, the signals can be generated precisely to the needs of the measurement. This enables measuring the timing with much higher precision than the timestamp generated by the DUT making use of noise and jitter in the system: Shifting the injection signals with respect to the timestamp epoch of the DUT at a high precision, the transition from one timestamp value to another can be measured. Due to noise and jitter, the detection time will be altered between the signals resulting in a smooth transition from one timestamp value to the next similar to an S-curve (see section 6.4.1). Shifting the signal further, also the transition to the next timestamp value can be probed doubling the statistics for the transition time for the timestamp. By assuming the jitter to be Gaussian, the fraction of the signals with a specific timestamp over the external delay of the injection signal can be fitted by the product of two Gaussian error functions with opposite sign, the same jitter value and a fixed offset between the two symmetry points being the timestamp length of the DUT. Such pulse shapes are shown in figure 6.34: For a fixed timestamp value from the DUT, the fraction of the signals read out is shown for different external delays.

The used time reference at 400 MHz enables time shifts at a granularity of 2.5 ns which is ten times as fine as than the timestamp duration ATLASPix3 is designed for.

The high time granularity does not only enable the measurement of the delay differences between different pixels, but it also enables timing compensation between them for large signals.

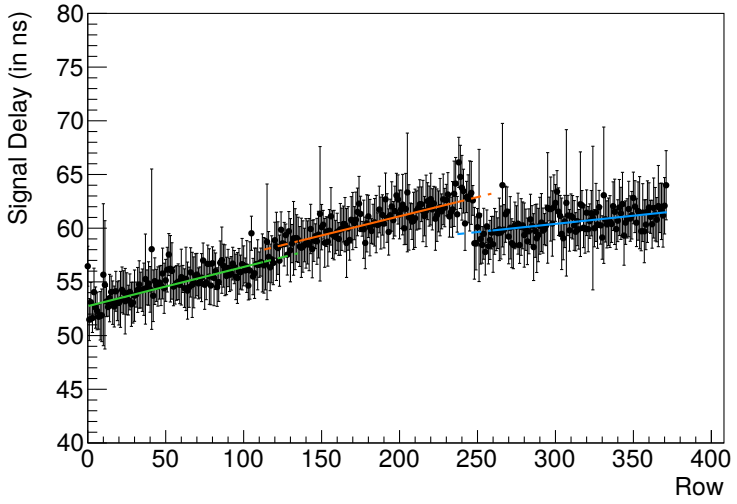


Figure 6.35: With an external time reference at 400 MHz, the signal detection delay of the pixels in different rows have been measured. The plot shows the average delay of 10 columns. At changes in the signal routing, steps in the delay dispersion are visible. The big steps at the rows 124 and 248 match the layer change for the comparator output signal routing. [SAA+21]

The other addressed contribution limiting time resolution is **time-walk** originating from the finite rise time of the amplifier in the pixel. However, the amplitude information from the ToT measurement can be used to correct it. But in order to use this information, the pixels need calibration as the length of the signal for a specific charge depends on the feedback current which is small and resource consuming to control precisely. Hence, the design is made to be calibrated and corrected off-line to obtain comparable results between the pixels.

In the following sections, first the signal delay over the sensor is discussed, before a compensation with the threshold TDACs is presented. The calibration of the ToT information is included as last point.

6.5.1 Row Dependence of Signal Timing

Using the measurement procedure described above, the timing of the pixels in the matrix can be probed with test signal injections. For the measurement, the symmetry of ATLASPix3 is exploited: The statistics for the measurement is increased by measuring several columns that are copies of each other placed next to each other in the layout. This way, the effects of the row dependence can be disentangled from the individual fluctuation of the single pixels. The influence of the test signal routing to the pixel from being located in different columns will contribute in all rows similarly. Hence, it will alter the shape of the delay distribution for each row, but not the shifts between the rows.

For ten adjacent columns, the delays have been measured on all pixels resulting in mean and standard deviation values for each row. These values are shown in figure 6.35, the individual delays are shown in appendix E. The delay dispersion shows three sections separating at the rows 124 and 248. These are the rows at which the routing from the pixels to the hit buffers changes to a different layer. The colours for the fit lines match the ones used in the description of the layout in section 6.1.1 with figure 6.3. The solid parts of the lines denote the interval used for the fits and the dashed part are extrapolations for comparing the delay changes. The total delay spread is (10.1 ± 0.7) ns. The largest delay

is reached at the end of the second group (row 247), the smallest at the beginning of the matrix (row 0).

The found characteristics still contains the bias from the injection circuit used in the measurement. However, the structure of the circuitry to conduct the injection signal into the pixel does not change over the matrix and with increasing distance to the periphery, the signal lines get longer. Hence, the injection bias will add to the characteristics seen in the measurement making the value an upper limit for the detector performance. In addition, the delay difference of 10 ns over the matrix leads to the conclusion that the bias introduced by the injection circuitry has to be smaller than this value.

6.5.2 Matrix Timing Optimisation

For large signals, the exact value of the detection threshold is not the dominant contribution to detection efficiency. In this case, large denotes signals several times the charge necessary to exceed typical detection thresholds. On ATLASPix3, signals above 4000 e^- can suffice this condition. For signals this large, the timing does not change significantly with a changed signal size as the rise time is limited by the amplifier bias current.

The rise time is constrained with a lower limit by the bias current of the amplifier. This is a consequence of the limited power consumption for the whole detector which originates from the application environment in the ATLAS ITk upgrade for HL-LHC. The finite rise time results in a timing shift of the signal detection if the comparator threshold is changed. This is visible in figure 6.34: To keep the detection time the same for the different TDAC settings, the external delay has to be increased the more the threshold is lowered. The data line with tilted squares indicates the characteristics between the TDAC setting and the signal delay.

The characteristics of the detection threshold to TDAC setting connection enables compensation for timing differences between the pixels. Compared to the threshold tuning, the characteristics can not only be modified by the step size of the TDACs but also by the amplifier current changing the rise time. In contrary to the TDAC step size setting, this setting changes power consumption and can influence other aspects as well.

To tune the matrix for equal delays, global threshold and step size (with the VNDAC setting) are adjusted so that all pixels can be used at all settings of the TDAC and pixels do not reach the baseline of the amplifier output with the threshold. Then, the tuning target is computed from the mean of the distributions for the extremal TDAC settings: For the lowest TDAC setting at 0, the signals are detected first resulting in long external delays in the delay measurement described at the beginning of this section. This is because the timestamp value from the DUT is kept constant. For the highest TDAC setting, the signal detection happens later resulting in shorter external delays. The delay distributions for these two settings give the scale on which the delay can be tuned. If there is a small overlap between the two distributions, the characteristics are set up in a way that the whole range of the TDACs can be used to get the optimum result. The distributions from which the target is computed are shown in figure 6.36.

Similarly to the threshold tuning, the pixels are then measured scanning the TDAC value to find the value matching the target best. Compared to the distribution for the highest TDAC setting, the delays will be shifted to longer external delays. The result for the tuning with the data in figure 6.36 is shown in figure 6.37. The distribution width is decreased from 9.06 ns for the untuned distribution to 2.34 ns in the tuned distribution.

The TDAC setting distribution, shown in figure 6.38, presents a Gaussian distribution centered around 4.3 with a standard deviation of 1.04. The deviation from the middle

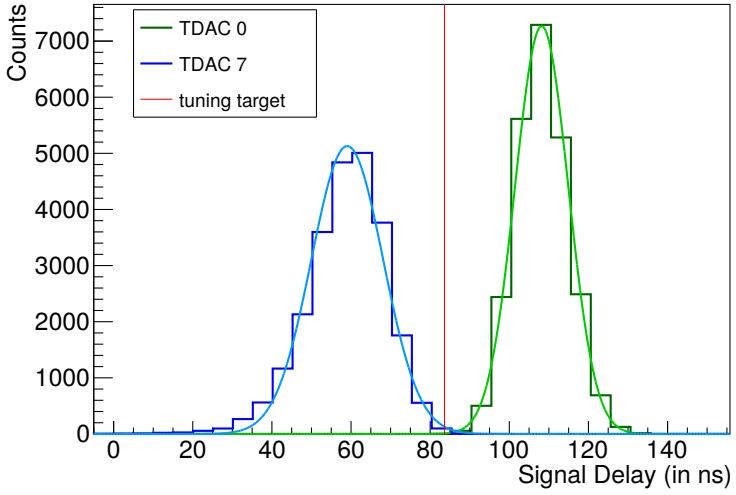


Figure 6.36: From the highest (TDAC = 7) and lowest (TDAC = 0) threshold settings, also the earliest (TDAC = 0) and latest (TDAC = 7) detection times can be extracted. Since the external delay to shift the signal into the timing window is shown on the X axis, longer delays on the chip require a smaller external delay and are drawn more to the left in this plot. From these two distributions, the delay target value is evaluated as average of the two means for the extremal settings.

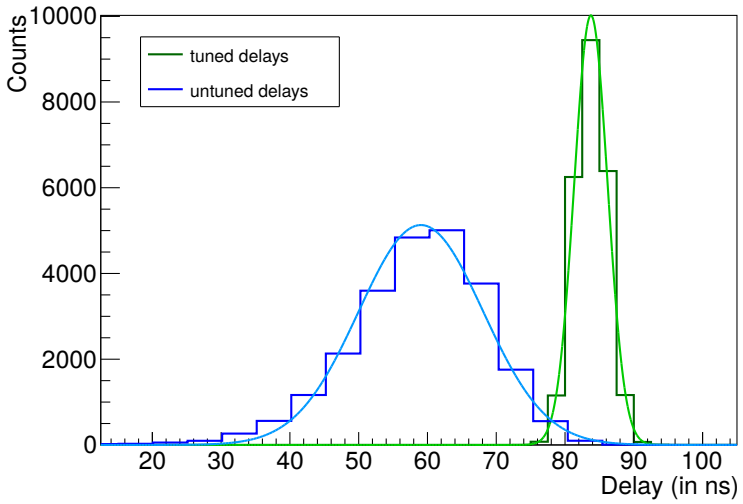


Figure 6.37: Compared to the settings prior to tuning (at TDAC = 7), the on-chip delay got smaller (requiring a longer external delay) and the width of the distribution was reduced by a factor of 3.7.

value of 3.5 can be explained from the non-linearity visible in figure 6.34 in the TDAC setting to delay data row.

With the insight from the row dependence of the signal delay in the previous section that the injection bias is small (i.e. in the order of less than 10 ns) and the achievable delay shifts of about 50 ns, it can be concluded that the described procedure works independently from the injection circuit replacing it for example by a TCT laser for signal generation. The signal delay differences can be compensated for large signals.

Comparing the resulting settings for detection threshold tuning and timing tuning it

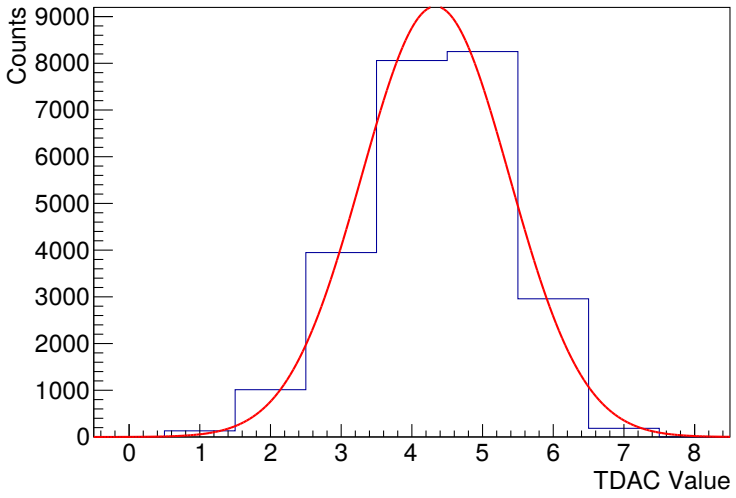


Figure 6.38: The TDAC distribution for optimum timing with the set target is of Gaussian shape. From the non-linear characteristics (as visible in figure 6.34), the distribution is shifted towards larger values.

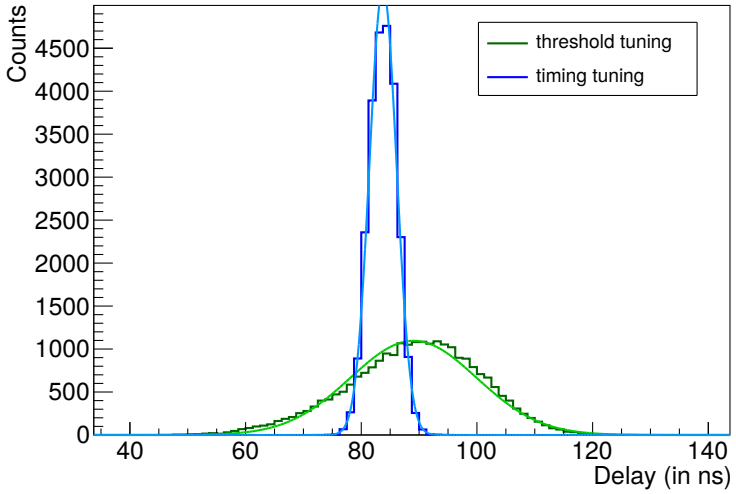


Figure 6.39: The result of threshold optimisation and timing optimisation do not match with the delay distribution being significantly wider for the threshold optimisation. The timing distribution for threshold tuning is even wider than for a single TDAC setting on all pixels. Hence, threshold tuning and timing tuning mutually exclude each other on ATLASPix3.

becomes clear that the results are not the same. Timing depending on the bias currents for the amplifier and comparator and threshold tuning depends – among other parameters – on the relative properties of the branches of the comparator. This can be seen in figure 6.39: The timing tuning results in a much narrower delay distribution than the threshold tuning. The delay distribution of the threshold tuning is even wider than the distribution for all pixels at the same setting when comparing the width to the distributions in figure 6.36. The consequence is that timing and threshold can not be optimised at once with the TDACs on ATLASPix3.

6.5.3 Time-over-Threshold Calibration

For on-chip measurement of the signal amplitude, ATLASPix3 implements a time-over-threshold (ToT) measurement. It is measured with the main timestamp on the leading edge of the signal and a secondary timestamp that stores the value at the trailing edge of the signal. Here, start and end denote the time at which the comparator output goes high and low. The signal shape generated by the amplifier results in a complex connection between the charge generating the signal and the length of this signal. The leading edge of the amplifier signal is affected by the bias current, the trailing edge signal is affected by the feedback current of the amplifier. From the design of the detector concept, the feedback current needs to be in the order of 60 pA [PAA+21]. This small current is hard to control precisely by circuit design while keeping the circuit efficient and small. As a consequence, the feedback current varies between pixels resulting in ToT differences of a factor of two between pixels for the same signal at equivalent detection thresholds.

Instead of tuning the values on-chip, calibration and off-line correction are favoured. For a given pixel with its signal characteristics, the ToT values measured will change with the setting of the threshold. Consequently, threshold tuning has to precede the ToT calibration and the calibration will only be valid for the threshold setting it was made on.

For calibration, the ToT value is measured for test signals of different size. For each signal size, a sufficient number of signals is measured to obtain a significant distribution from which the mean and standard deviation are extracted with a Gaussian fit. The extracted values are collected and fitted with an approximation of the characteristics motivated from the amplifier characteristics. The used equation is

$$ToT(u) = a \log\left(\frac{u - u_0}{u_0}\right) + b\,u + c \tag{6.2}$$

with the injection voltage u and parameters a, b, c and u_0 at which the function diverges and approximates the detection threshold. For a sound calibration, the points at which the signals are not detected have to be accounted, too. This is important as the fit function will be only defined for values larger than u_0. An example of a characteristics is shown in figure 6.40.

For measurements without calibration of ToT, it is important to be aware, that the feedback current as the parameter with variation does not enter linearly to ToT. Instead, it enters inside a logarithm. Consequently, the ToT distribution of many pixels for one signal size will not form a Gaussian distribution over the ToT value, but a log-normal distribution. This distribution is the result of a normal distributed parameter entering in the value inside a logarithm. For test signal injections of a charge equivalent of about 8200 e^-, the distribution of ToT values of the pixels of the whole matrix is shown in figure 6.41 fitted with a log-normal distribution.

For application, the calibration function fitted to the data points has to be inverted accepting the ToT value as variable and resulting in a charge equivalent voltage if measured with test signal injections. This voltage can be converted into a charge using a calibration for the test signal injections as described in section 6.6.2. The used fit function leads to a Lambert W function in the inverted function evaluating to

$$u(ToT) = u_0 + \frac{a}{b}\,\mathrm{W}\left(\frac{a}{b} u_0 \exp\left(\frac{ToT - c - b\,u_0}{a}\right)\right) \tag{6.3}$$

for the injection equivalent voltage u with the parameters from the fit before. The analysis scripts for ToT data use the implementation of the Lambert W function from [Veb12].

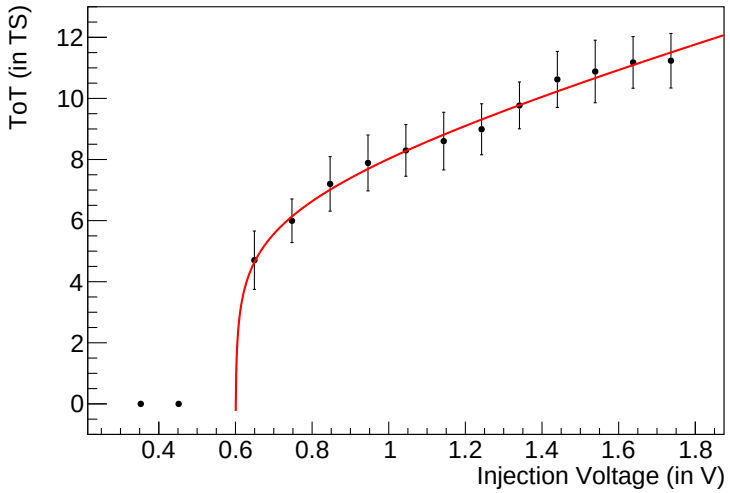


Figure 6.40: The time-over-threshold shows a logarithmic dependency on the signal size for small signals. For signals smaller than the detection threshold, the measurement is not possible. Above the detection threshold, the signal follows a function with logarithmic and linear contributions.

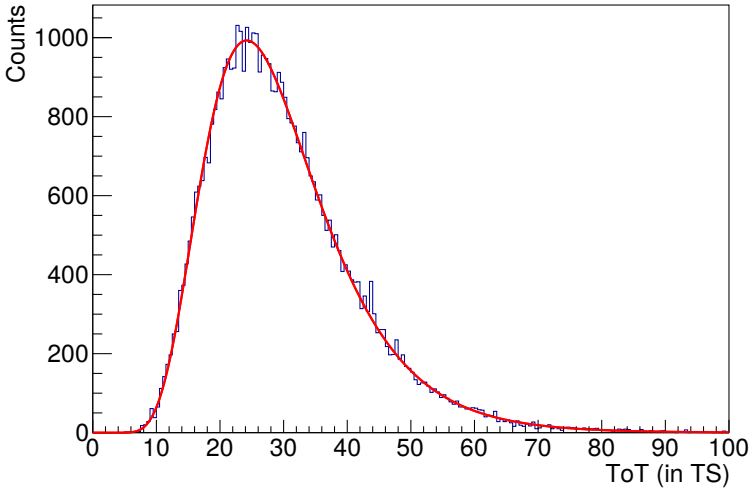


Figure 6.41: The ToT distribution for a fixed signal size of all pixels of the matrix is plotted. Because of the logarithmic ToT to signal size relation, the resulting distribution is not a Gaussian distribution but a log-normal distribution.

This calibration is applied to the data in section 6.9 (see figure 6.90). It can be used to compensate for time-walk as it is measurable with the procedure from the beginning of this section (section 6.5) or from external time references as they can be available in beam tests in section 6.9 in figure 6.94.

6.6 Measurements with Particle Sources

To complement the tests with charge injections, the sensor is tested with radiative sources. On one hand, ^{55}Fe is used as its mono-energetic X-ray photons generate a small defined signal used as benchmark for silicon-based radiation detectors. On the other hand, ^{90}Sr as a β-source is used as the electrons emitted generate signals in silicon that are similar to the signals of MIPs [FP+15]. Using the amplifier output to record the signals generated, a

calibration of the charge injection circuit can be performed or the signal-to-noise ratio can be calculated. With digital readout, the functionality of all pixels can be verified.

The ^{90}Sr measurements allow for probing the amplifier dynamic range, depletion volume tests and to give an estimate of the SNR that can be expected in a general application environment.

Measurements with ^{55}Fe enable putting absolute numbers on the signal scales by allowing for calibrating test signals against it.

Following the characterisation in laboratory, ATLASPix3 has been used as technology representative to test HV-CMOS detectors in proton and carbon ion beams. These measurements are predominantly testing the pixel diodes and amplifiers to prepare a foundation on which dedicated sensors can be built for this environment. But directly measuring a beam also offers the opportunity to probe the readout structure of ATLASPix3 enabling comparison to the simulations in ROME in chapter 5.

6.6.1 Beta Particles from Strontium-90

The electrons from the decay of ^{90}Sr will be used first to test the output of a pixel amplifier. Then, the whole matrix will be looked at by measuring the signal size via ToT measurement. The electrons are not stopped immediately as photons, instead the average energy deposition can be assumed constant per length. Therefore, the depletion layer thickness change over the depletion voltage can be probed, too.

The amount of charge generated by a traversing electron can be expected to follow a Landau distribution (see section 3.2.4). Measuring the output signal from the amplifier of a pixel, uncertainties will be introduced to the measurement that will be modelled with a Gaussian distribution resulting in a convolution with the Landau distribution.

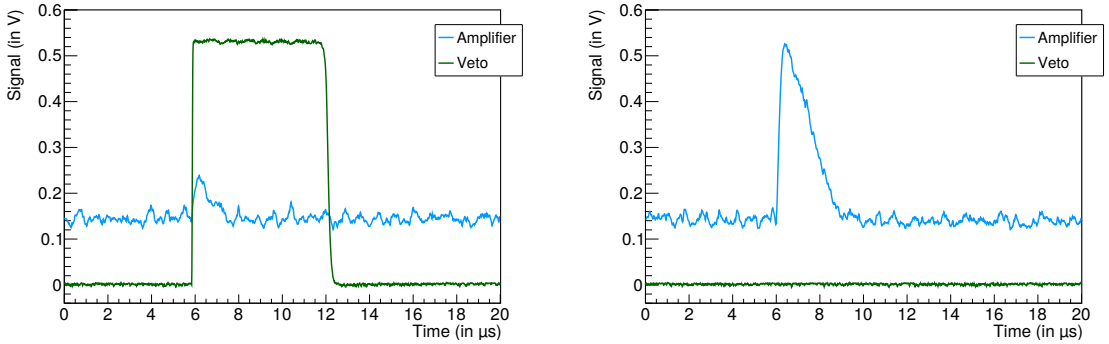
Amplifier Output

One point to take into account is charge sharing between neighbouring pixels. Because then not the full charge of the cluster is collected in the measured pixel altering the resulting signal size distribution. Therefore, the neighbouring pixels will be used with the hitbus output as veto for the acquired signals.

With the pixels with accessible amplifier output located at the bottom end of the matrix, five pixels surrounding the probed pixel are monitored. For this, the thresholds of all pixels apart from the five pixels of interest are set to a high value so that the threshold can not be crossed. Especially the pixel, of which the amplifier output is measured, is set to this high threshold. The five neighbour pixels are set to a threshold as low as possible. These thresholds are at or below $1400\,\mathrm{e}^-$ as the calibration done with the ^{55}Fe source will show later. During the measurement, both the amplifier output and the hitbus are recorded and signals with a pulse on the hitbus are discarded as visible in figure 6.42.

The accepted signals are then analysed for their height and length. The correlation of these parameters is shown in figure 6.43.

It is important to understand the shape of the histograms of the single parameters: The saturation of the signal height means that the histogram content above 500 mV will be squeezed resulting in higher counts in the respective bins. Consequently, this part of the signal height histogram can not be used for a Landau-Gaussian fit. The signal shape for the saturated amplifier and its implications are described in appendix F. Compared to other HV-CMOS detectors, the amplification is larger leading to the observed saturation effect. On HVStripV1 for example, this saturation was not visible [Sch16]. Also on the



(a) discarded capture (b) accepted capture

Figure 6.42: The hitbus is active for the pixels surrounding the pixel measured at the amplifier output. A signal at the hitbus hints to charge sharing which means that the signal is not to be used for determination of the signal height from the ^{90}Sr decay electrons. (a) shows a signal to be discarded, (b) shows an accepted signal. For better visibility, the amplifier output has been shifted vertically and the hitbus has been scaled to 50 % of its actual amplitude.

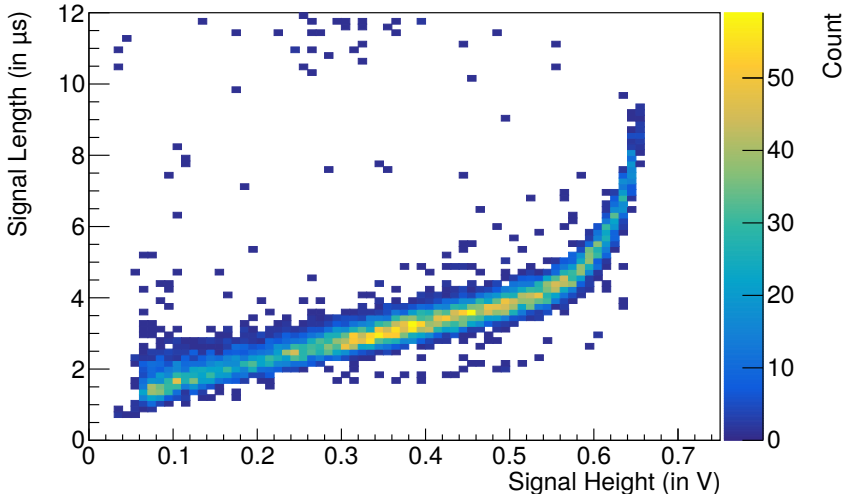


Figure 6.43: The signals recorded show a linear correlation between the length of the signal and the height for signals smaller than 500 mV. Above this point, the signal height saturates while the length continues to increase. In the colour axis, the peaks in the distribution become visible: one noise peak at a signal height of 100 mV and the signal peak around 350 mV for the used depletion voltage of 50 V.

HitPix integrated sensor, the amplification is smaller for the different target application with larger signals to be expected [Web21].

On the other hand, the noise peak at a signal height of 100 mV becomes almost indistinguishable in the signal length parameter as for the overlap of the peaks. Consequently, it has to be kept in mind that the rising edge of the signal length distribution is altered.

Looking at the signal height distribution, the noise contribution at the smallest heights recorded becomes clearly visible in figure 6.44. This background is fitted with an exponential decay on the interval indicated by the solid line. Extrapolating this distribution, the background contribution is removed. Errors for the signal heights above 500 mV with this

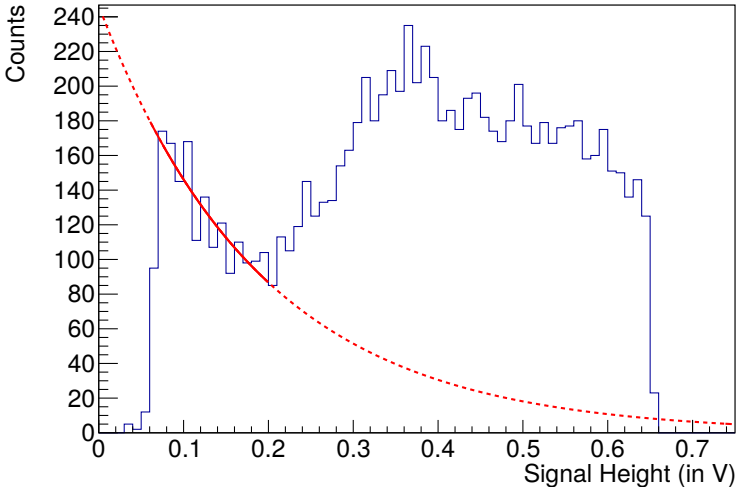


Figure 6.44: The signal height distribution at a depletion voltage of 50 V shows the noise peak at the smallest heights recorded and the signal distribution with the squeezed tail at heights above 500 mV assuming a Landau-Gaussian height distribution. The background is fitted with an exponential decay on the interval with solid line.

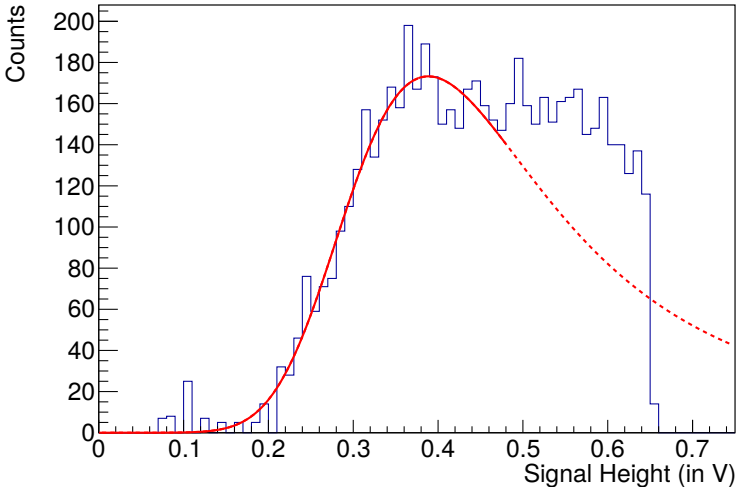


Figure 6.45: The remaining signal distribution is fitted for signals below 500 mV with a Landau-Gaussian distribution to find the MPV. The decay of the distribution expected for a Landau-Gaussian distribution is compromised by the saturation of the amplifier's signal height.

correction are ignored as this part will not be used for the next steps.

With the background removed, the lower end of the signal height distribution becomes accessible (see figure 6.45). It is fitted with a Landau-Gaussian distribution for signals below 500 mV to obtain an MPV of (378 ± 10) mV at a depletion voltage of 50 V.

Comparing this value to the noise measured on the base line of the amplifier output of the same data, an estimation for the SNR can be given. Histogramming the baseline from many signals, the histogram in figure 6.46 is obtained. The noise evaluates as standard deviation of the Gaussian fit to (9.19 ± 0.02) mV. Combining these two values, the SNR results in 41.1 ± 1.2.

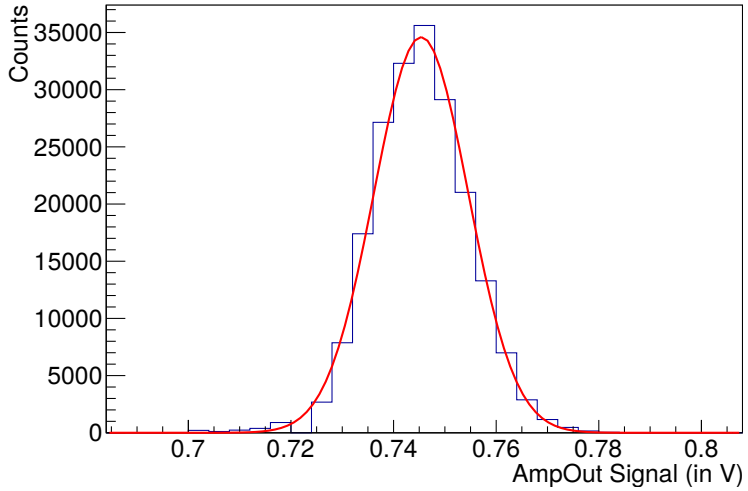


Figure 6.46: Histogramming the baseline the shown histogram is obtained. The noise is defined as the width of this distribution.

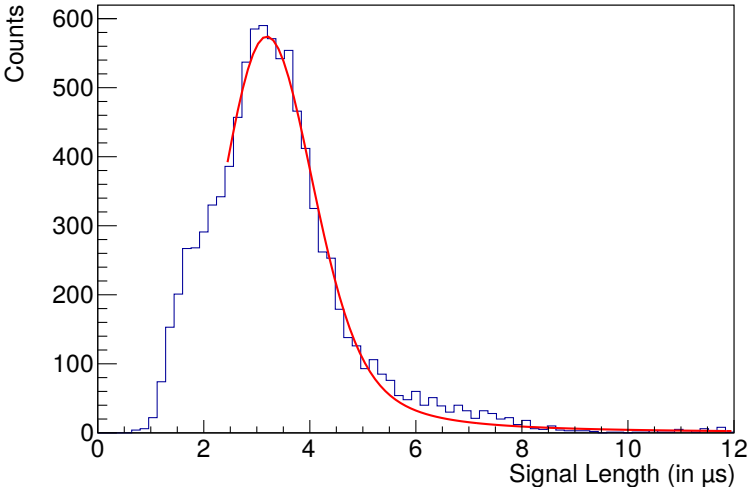


Figure 6.47: The rising edge of the distribution is altered from the expected shape by the noise signals. The rest of the distribution follows coarsely a Landau-Gaussian shape.

Using the calibration from the ^{55}Fe X-ray signals from the next section, this signal amplitude equals to a charge of $(5350 \pm 160)\,\mathrm{e}^-$. Assuming an average energy loss of a minimum ionising particle in silicon of 390 eV/µm [Har09], this results in 108 electron-hole pairs per micrometre and consequently to a depletion depth of (49.5 ± 1.4) µm which matches the expectation for the 200 Ωcm substrate material.

The signal length distribution for the same data is shown in figure 6.47. The overlap of the background signals with the signals of interest prevents separation as in the case of the signal height. However, the correlation can be used to find the point at which the signal distribution becomes the major contribution and a lower limit on the fit interval can be derived from it. For the depletion voltage of 50 V, this is chosen at 2.4 µs. The MPV of the length distribution results in (2.96 ± 2) µs.

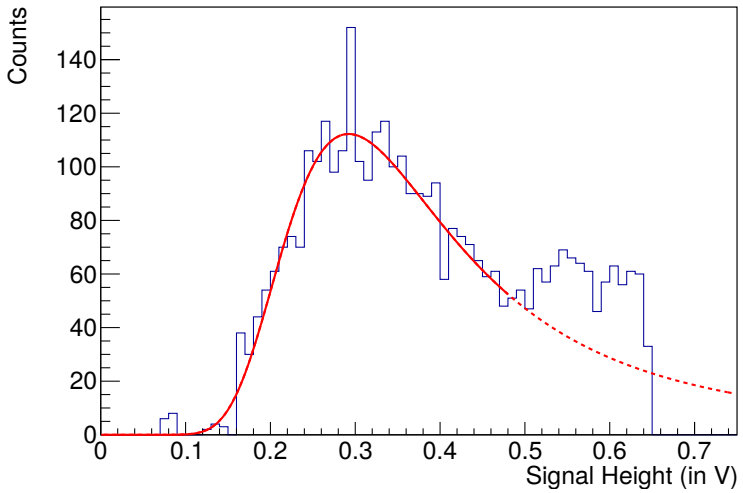


Figure 6.48: The signals are smaller for a depletion voltage of 20 V compared to the signals at 50 V. Consequently, a larger fraction of the Landau-Gaussian distribution is well below the saturation resulting in a smaller uncertainty on the MPV.

Influence of Depletion Voltage on Signal Size

As it can be expected from the description of the pn-junction (section 3.1.3), the amount of collected charge is reduced for smaller depletion voltages. For a depletion voltage of 20 V, the signal height distribution with removed background is shown in figure 6.48: The MPV of the distribution evaluates to (292 ± 3) mV. The smaller uncertainty can be explained from the smaller signals enabling the fit of a larger fraction of the distribution below the saturation limit of the amplifier. The charge equivalent is $(4130 \pm 50)\,e^-$ with a depletion depth of (38.3 ± 0.5) µm. Evaluating the noise in the same way as for the measurement at a depletion of 50 V, the noise evaluates to (11.3 ± 0.02) mV and results in a SNR of 25.8 ± 0.3.

Matrix Measurement

With the matrix tuned, the detection thresholds of the pixels are similar resulting in ToT values measured at comparable fractions of the signals. However, a calibration of ToT is not applied. Consequently, the count rates of the pixels are expected to follow the gradient from the position of the source above the matrix but the measured values for ToT will form a log-normal distribution.

The signal distribution over the matrix is shown in figure 6.49. The spot from the source being placed over column 65 below the vertical center is clearly visible.

Also according to expectation, the distribution of the ToT values follows a log-normal distribution in figure 6.50. The clock dividers have been chosen to cover the expected range of values with the pixel-to-pixel variations. The MPV of this distribution recorded for a depletion voltage of 5 V evaluates to (1.83 ± 0.01) µs. Repeating the measurement with increasing depletion voltage, the running of the MPV shown in figure 6.51, is recorded. From theory, a square root connection between the depletion voltage and the thickness of the depleted volume is expected which translates to the same connection between depletion voltage and charge. The same connection should hold for ToT looking at the correlation of the signal height and length in figure 6.43. Increasing the depletion voltage from 5 V to 50 V, the MPV of the ToT increased by 48 % and shows the expected sub-linear behaviour.

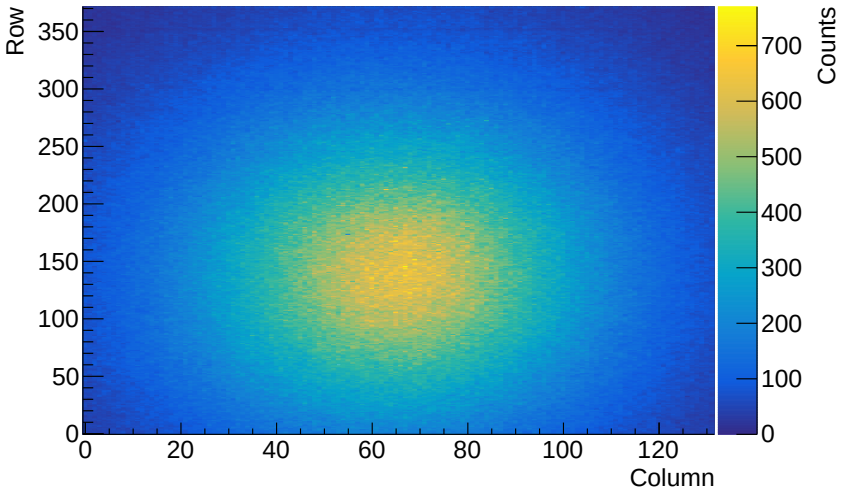


Figure 6.49: The histogram contains data from 10 million signals and clearly shows the position of the ^{90}Sr source.

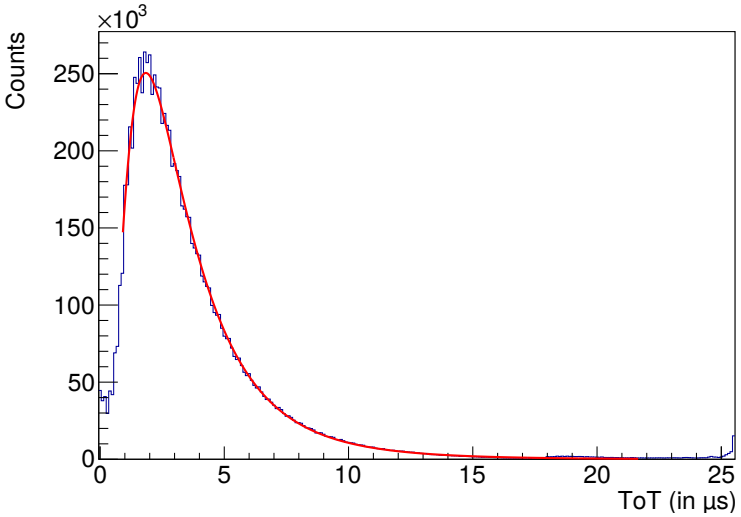


Figure 6.50: The ToT distribution follows a log-normal distribution. As for the amplifier output measurement before, the low end of the distribution is spared out from the fit as there, data are compromised by noise.

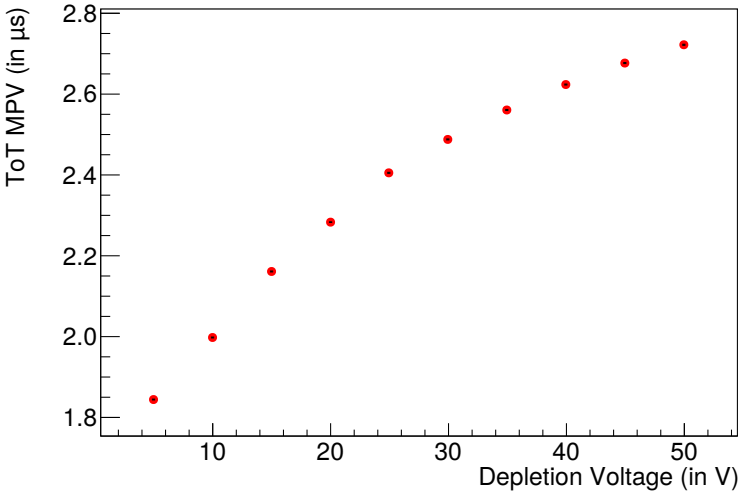


Figure 6.51: The MPV for the ToT increases with the depletion voltage sub-linearly as it is expected from the depletion thickness in a reverse-biased pn-junction.

6.6.2 X-rays from Iron-55

The ^{55}Fe source emits characteristic X-ray radiation at an energy of about 5.9 keV from the two K_α transitions (with an energy difference of 1 eV). Hence, the two transitions are treated as mono-energetic. In silicon, this energy results in 1638 electron-hole pairs from the generation energy of 3.6 eV per pair [Har09].

The amplifier output of one pixel is used to record the signals generated by the X-rays from the ^{55}Fe source. The signals are recorded with an oscilloscope to extract the signal height. To check for consistency, the correlation of the signal height with the signal length is formed allowing to separate signals from random noise spikes. This is possible because the amplifier has a characteristic signal shape for signals received from the pixel that gets longer and taller with increasing charge. In contrary, pulses generated by noise are short and these pulses do not change in length with signal height. Figure 6.52 shows the correlation for the dataset measured with pixel (65|0) in the middle of the lowest row of the matrix: The signals of interest form the distribution around the accumulation point at a signal height of 115 mV and a duration of 2.1 µs. The tail to smaller signals can be generated by charge sharing with neighbouring pixels and other signals. The cluster detection method used in the previous section is not possible here due to the smaller signals.

Taking the signal height distribution alone from figure 6.53, two parts can be identified: The signal peak on the right and a noise peak cut off on the left at small signals. The signal peak is fitted with a Gaussian distribution as X-rays transfer a defined amount of energy being absorbed which is then widened again by the measurement system and thermal fluctuations. The peak evaluates to a signal height of (115.7 ± 0.3) mV.

Calibration of Test Signal Injections

With the knowledge on the amount of charge deposited by the X-rays from the ^{55}Fe source, the test signal injection circuit can be calibrated. Since this circuit is a capacitor that is discharged into the pixel, the calibration is independent of the settings used for the amplifier in contrast to the threshold tuning or the ToT calibration. To calibrate the test signal injections, such signals of different sizes are generated in the pixel that has been used to measure the size of the signal from the ^{55}Fe X-rays with the same settings for the sensor.

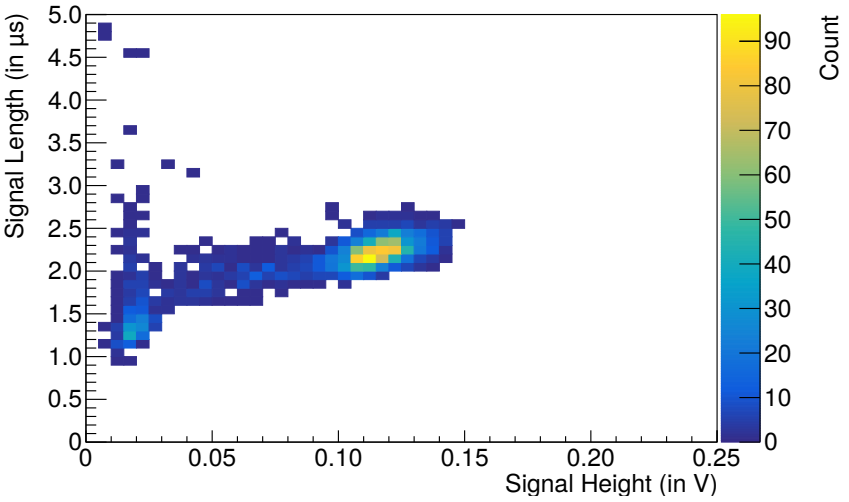


Figure 6.52: Correlating the height and length of the signals recorded at the amplifier output, consistency of the data can be checked. At small signal sizes, the length measurement shows large uncertainties as large spread in length. Charge sharing will lead to smaller signals which are shorter, too. Hence, the presence of these effects results in a tail of the distribution to the bottom left.

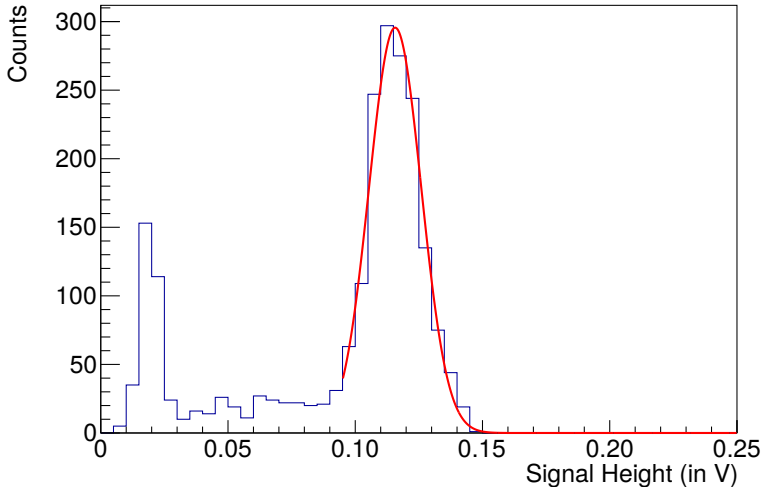


Figure 6.53: Measured at the output of the amplifier of one pixel, the signal height distribution for decay signals from ^{55}Fe shows an average signal height of (115.7 ± 0.3) mV. The tail of the distribution consists of noise and signals with charge sharing.

The charge injections are generated in equidistant steps in the parameter controlling the voltage to which the capacity is charged. The signal height and length distributions are shown as correlation in figure 6.54. The decreasing distance between the peaks in both signal height and length indicates towards a saturation of the test signal generation circuit.

From this data, the connection between the parameter controlling the circuit and the signal height generated is extracted and visualised in figure 6.55. The saturation for settings equivalent to a voltage of 850 mV, the slope of the connection starts to decrease. However, for smaller signals the connection is linear.

Fitting a straight line to the points for voltage equivalents smaller than 1 V, the test

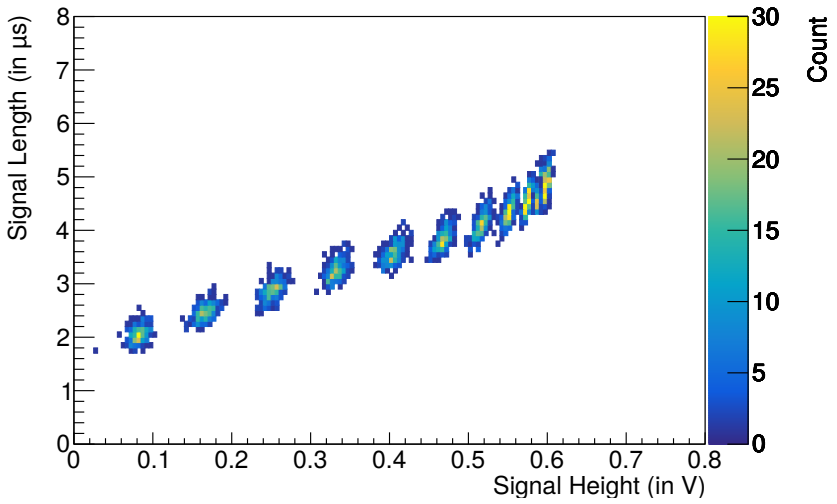


Figure 6.54: The test signals show the same linear behaviour with as the electrons in the previous section. However, the spacing between the peaks gets smaller to large signals indicating a saturation. And since both height and length stop to increase, it indicates towards a saturation of the test signal generation circuit.

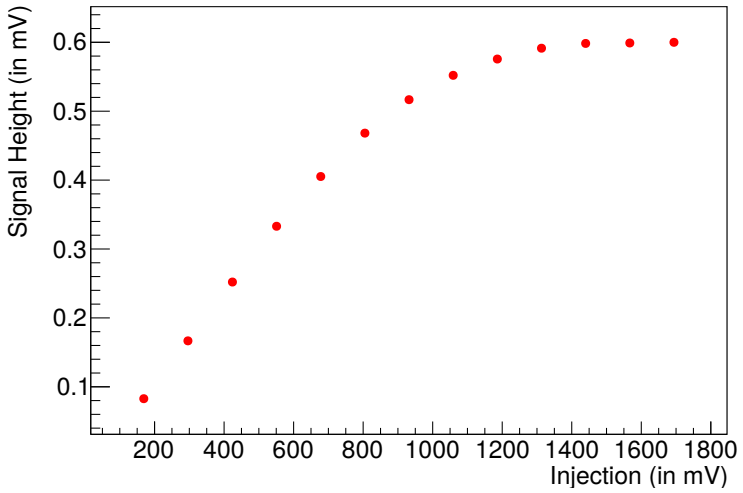


Figure 6.55: The connection from the parameter controlling the test signal generation circuit is linear for voltage equivalents smaller than 850 mV. Above this point, saturation of the circuit starts.

injection size equivalent to the signal from the ^{55}Fe x-rays is (211 ± 21) mV (or a setting of the 8 bit parameter of 30 ± 3).

6.6.3 Ion Beam Measurements

ATLASPix3 has been used for measurements with proton and carbon ion beams at HIT. For the possible application of HV-CMOS detectors as direct beam monitor for such beams, the goal of these measurements was to find the properties of the signals generated by the beam and implications for dedicated designs as it has been realised with the HitPix integrated sensor [Web21].

For high-intensity beams like at HIT, single signals are not important, but the position, shape and intensity of the beam. The measurements are performed on ATLASPix3 as it

was the most recent design and is easy to set up because of the full-reticle size.

Directly measuring the beam, the amount of charge generated inside the matrix is comparable to or larger than the leakage current of the sensor. Consequently, these measurements of the leakage current at different beam conditions have been conducted and are presented here.

The next step is to estimate the expected charge cluster size generated by the beam's particles. This is important as a pixel size smaller than the cluster size to expect results in a larger number of individual signals that has to be dealt with in addition to higher power consumption of the larger number of pixels necessary to cover the monitor surface without a benefit from it. Such an estimation is performed with the lowest intensity beams in order to avoid the limit of the readout of individual signals as implemented on ATLASPix3.

While for the measurement of the cluster size the system has been tweaked to obtain the data of interest, the beam can also be used to probe the limits of the readout structure of ATLASPix3 to compare the behaviour to the ROME simulations in chapter 5. This measurement is described last in this section.

Beam-Induced Sensor Current

The depletion voltage is used to increase the depleted volume in the pixel diodes and to separate generated charge. This adds to the sensor leakage current. The electrons generated in the depleted area are collected and amplified in the charge-sensitive amplifier. The holes are drained to the negative electrode. Consequently, traversing particles increase the sensor current temporarily. For single particles, this additional change is negligible: Typical signals generated in HV-CMOS sensors have less than 10 000 electrons (see for example figure 6.90 in section 6.9) which equals to $1.6 \cdot 10^{-15}$ C. This is seven orders of magnitude smaller than the leakage current on ATLASPix3 (see section 6.3.3). The leakage current refers to the leakage current without signal sources apart from background radiation.

A particle beam can contain a sufficient number of particles to make the generated change comparable to the dark current. Especially for irradiation applications as at HIT, high particle currents are desired. From the specifications, particle rates of up to $8 \cdot 10^7$ particles per second (carbon ions) or $3.2 \cdot 10^9$ particles per second (protons) are in use at HIT which are sufficiently large to expect measurable sensor current changes due to the particle beam in an HV-CMOS sensor.

During a beam test at HIT, the sensor current of ATLASPix3 has been monitored at different beam parameters. Matching this data with the irradiation plans, the dependency of the sensor current on the beam parameters can be reconstructed. The particle beam consists of spills with a length in the order of 5 s with gaps in the same order. The precise lengths of the spills and the gaps depend on the beam settings and the particle type. The leakage current has been sampled at 18.7 Hz to resolve the start and end of the spills and changes during the spills.

The sensor current measured on ATLASPix3 changes abruptly at the beginning and the end of spills as shown in figure 6.56. The difference between the dark sensor current before the step and the current during the spill equals the current induced by the particles traversing the detector. The spills in the figure had different parameters resulting in different sensor currents induced. Taking the average values for the intervals between the steps, the beam induced sensor current can be quantified. The resulting numbers are shown in figure 6.57b.

However, this result includes a correction factor which is visualised in figure 6.57a: the beam size is energy dependent. For small energies, the beam diameter becomes larger. This effect is larger for the proton beam and at some point, the beam becomes larger than ATLASPix3

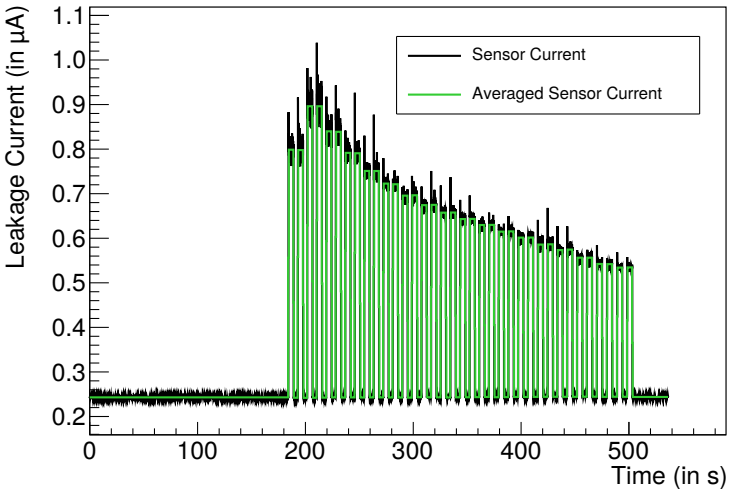


Figure 6.56: The leakage current measured on ATLASPix3 during a beam test with carbon ions contains jumps at the beginning and the end of spills. From the plateaus in between, the leakage current induced by the beam is calculated.

and consequently, only a part of the particle current actually passes the detector. The fraction of the beam passing the detector is estimated from a two-dimensional Gaussian distribution with a FWHM matching the characterisation chart for the accelerator with the center of the beam in the middle of the sensitive area. The measured current is then rescaled with this fraction to obtain an estimate for the whole beam. For the smallest energies, the fraction passing ATLASPix3 can go down to about 30 % according to the model. In addition, the focus setting for the beam has to be changed between the settings for the carbon ion beam. This is marked in the figure with different marker shapes and colours. To make use of the best-maintained beam settings – which are the ones used for patient treatment – the different focus settings have been used.

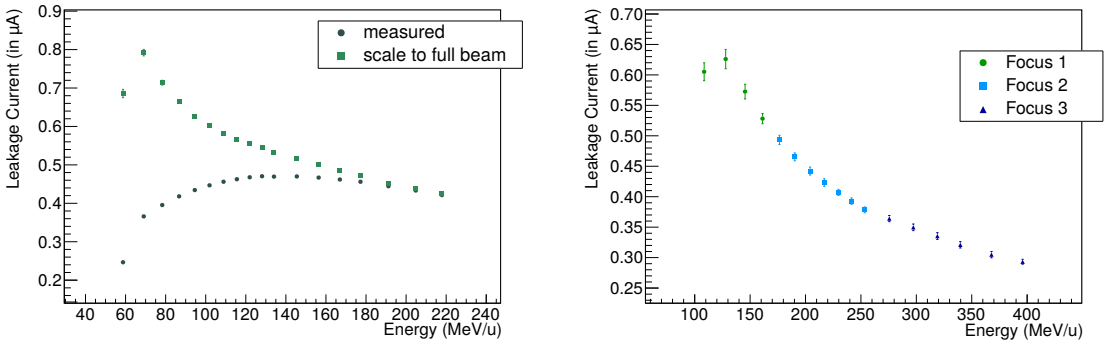
For a beam monitor application, this correction will not be necessary as not only one sensor but a matrix spanning approximately 25×25 cm^2 will monitor the beam. With this arrangement, the whole beam passes through sensors for all energy settings.

A linear dependency of the sensor current change is expected for the different beam intensities. For carbon ion beams, a measurement is shown in figure 6.58. The data rows for the three energies are linear and scale to each other as seen in figure 6.57b.

These measurements are match the findings presented in [Ehr21] and refine them with the correlation to the sensor size and finer granularity of the energy measurement. They support the hypothesis that the leakage current can be used for verification of the beam parameters and if the leakage current is measured on the sensors of the matrix separately, a coarse position measurement is possible, too.

Beam Signal Parameters

For the application of monolithic HV-CMOS detectors as beam monitor for a particle beam, the properties of the signals generated by the beam on the sensor need to be known. Since other beam facilities use lighter particles – for example electrons at DESY [DDE+19] – especially the signals generated by the carbon ions used at HIT are not known for HV-CMOS detectors. Therefore, data has been taken with existing sensors at the commissioning line at HIT. One of the detectors used is ATLASPix3.



(a) proton beam (b) carbon beam

Figure 6.57: For the proton beam, the leakage current has to be rescaled in order to compare the different energies as the beam diameter changes with the energy setting of the beam and consequently a changing fraction of the beam passes through ATLASPix3. This affects the amount of leakage current generated. The diameter of the carbon beam is smaller than the one of the proton beam, resulting in small changes of the fraction of the beam passing through ATLASPix3. After applying the corrections for the beam size, both beams show a smaller leakage current with higher beam energy.

The beam parameters scanned at HIT are the energy of the particles, the particle type, the focus of the beam and the beam intensity. As for different foci and beam intensities only the number of particles per area will change and ATLASPix3 is built for tracking of collision products not direct monitoring of a particle beam, only the lowest beam intensity and widest beam diameter are used to study the signals generated by the particles at different energies. In addition, a copper collimator has been placed in front of the detector to further reduce the particle current reaching the detector. It consists of two symmetrical parts that allow for adjustment of the aperture. However, this shielding with the two 1 cm thick copper plates is not sufficient to stop the particles at higher energies so that low rates are only achievable for the low particle energies. A schematic drawing of the setup is shown in figure 6.59.

Summing up these considerations leaves the remaining parameters to test as particle type and particle energy.

From the signals read out for the different beam settings, the signal clusters are extracted to estimate the size of the charge clouds generated by the traversing particles. Such a cluster size histogram is shown in figure 6.60. The frequency of the sizes follows an exponential decay. In the following, the decay length of this fit is taken as a measure for the size of the clusters.

Performing this fit for all measured energies, the dispersion of the cluster size can be extracted. It is shown for the proton beam in figure 6.61. The energies at which the particle beam can pass through one and two of the layers of the collimator are marked as vertical lines. For energies above these limits the cluster size drops which is connected to the low energetic particles generated from the collimator. With a maximum cluster size of about 1.24 ± 0.25 for the largest energies, the effects of the shielding can be considered small.

The same measurement has been performed for carbon ions, for which the cluster size dispersion is shown in figure 6.62. The low-energetic carbon ions just passing through the copper shielding create large clusters, exceeding the scale of the plot. In addition to the particles passing the aperture and the material, collision products from the beam with

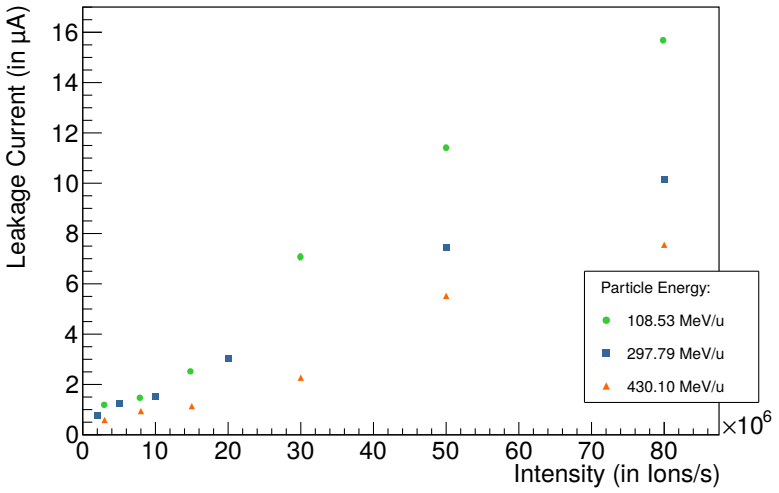


Figure 6.58: The beam-induced leakage current increases linearly with the particle rate. For the different energy settings, the data rows scale according to the measurement in figure 6.57b with lower leakage currents for higher energies.

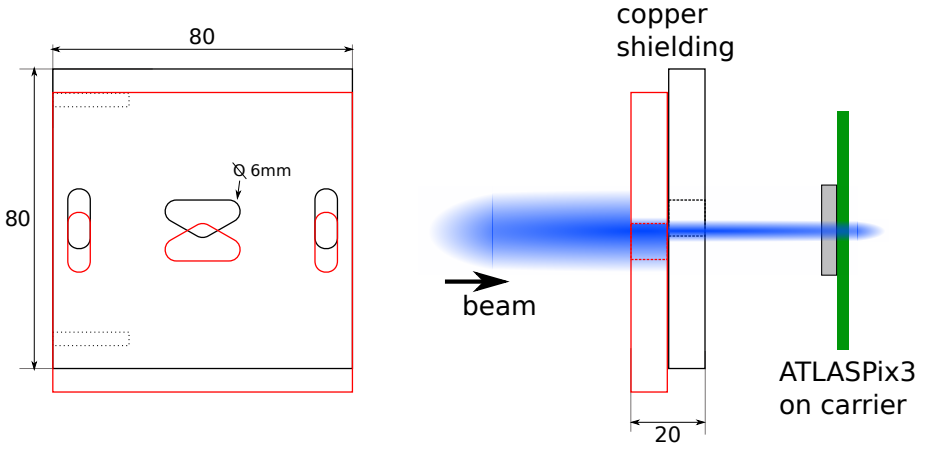


Figure 6.59: The two times 1 cm thick copper plates are movable against each other giving a changeable aperture with lens shape. The DUTs are placed behind it.

the shielding hit the sensor and overlay the data of interest. With higher energies, the amount of collision products is reduced again. The average cluster size is smaller than for the proton beam with a smallest cluster size of 1.04 ± 0.03 at the highest energy of 217.9 MeV/u.

The largely increased cluster size after reaching the penetration depth of the first layer of the shielding hints to the creation of secondary particles creating a shower of particles traversing ATLASPix3. The overlay of all clusters for this measurement is shown in figure 6.63.

High-Rate Readout Tests

The high-intensity beam at HIT results in a particle rate one order of magnitude larger than for the target application in the ATLAS ITk upgrade [ATL17]. Therefore, the signal rate will exceed the rates the readout is designed for and the limitations of the readout encountered in the simulations can be identified in the measurement.

If the readout speed is reduced, this excess is further increased. For the beam test measurements presented in the following, a quarter of the design readout speed at 320 Mbit/s has been used.

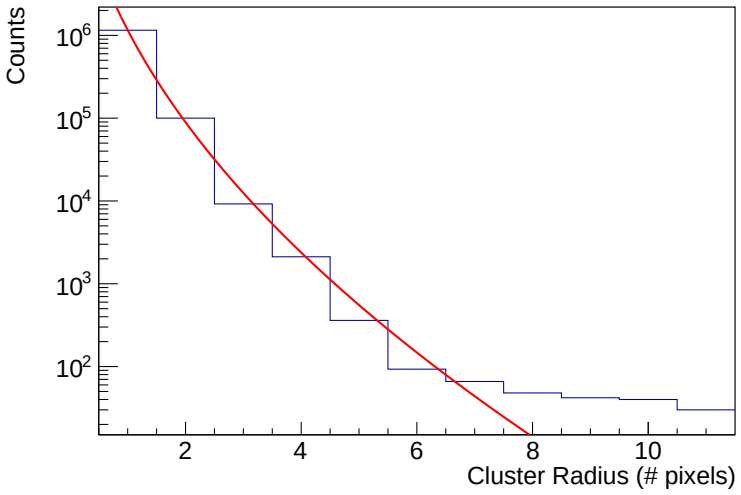


Figure 6.60: The cluster size is shown as number of pixels in the cluster. This is an approximative measure since the aspect ratio of the pixels on ATLASPix3 is 3:1. The effects giving the tail are not depicted in the fit function as they are partly not beam-related.

To compensate for the difference, special settings have been used:

- Since the amount of memory necessary in the trigger buffers is proportional to the product of signal rate and trigger delay, the trigger delay has been chosen as short as ten timestamps. This is possible as the trigger is activated at a fixed rate and no preprocessing is necessary for trigger generation.
- To further increase the memory available for a pixel, only every third pixel is activated in a staggered pattern between the columns.
- The timestamp length has been extended by a factor of 16 to enable readout of whole clusters in one event triggering longer time intervals at once.

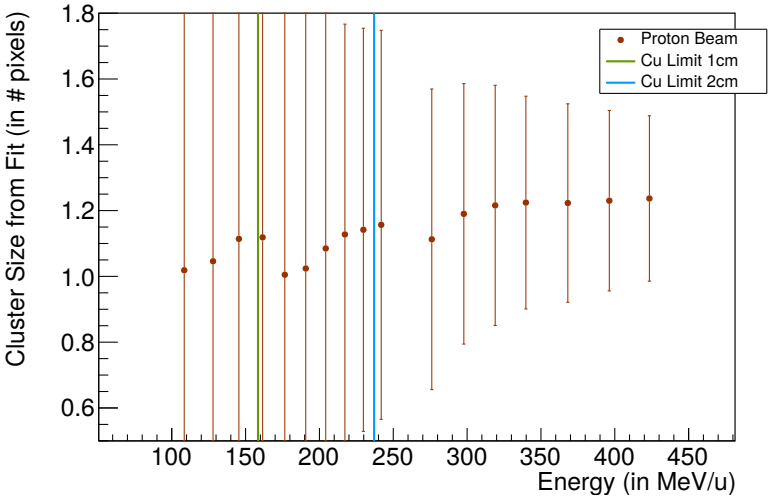


Figure 6.61: The low statistics for the small particle energies result in large uncertainties. At energies above the penetration depth of the two parts of the copper collimator, the cluster size drops. But with larger energy, the cluster size increases again.

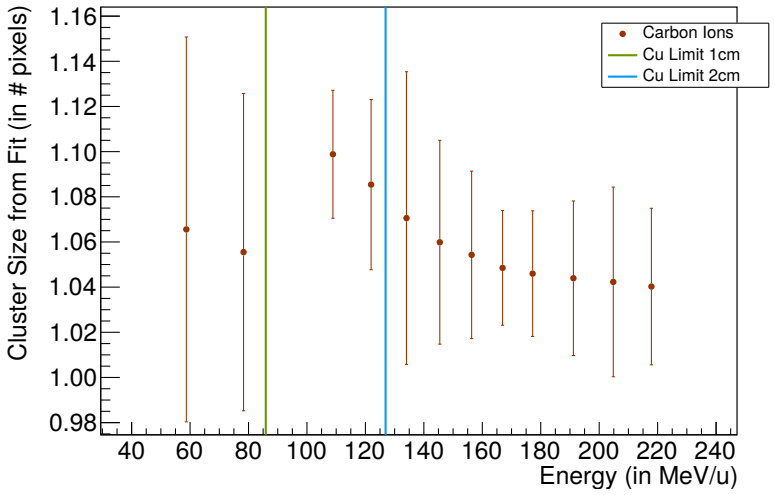


Figure 6.62: The cluster size is decreasing with increasing energy of the carbon ions passing through ATLASPix3. At the energy when the first half of the shielding becomes transparent, the cluster size increases above the scale shown but decreases again.

- The triggered fraction has been reduced to 5 %. Combined with the timestamp downscaling, this allows for the readout of about 300 signals during the event on average.
- The trigger delay is chosen in a way that more signals can be transferred from the hit buffers to the CAM cells than there are CAM cells to ensure maximum efficiency.

With these settings, the beam profile of a carbon ion beam at lowest intensity of $2 \cdot 10^6$ particles per second can be recorded as it is shown in figure 6.64. The bins have been extended to cover three pixels in vertical direction as of these triplets always only one pixel is active.

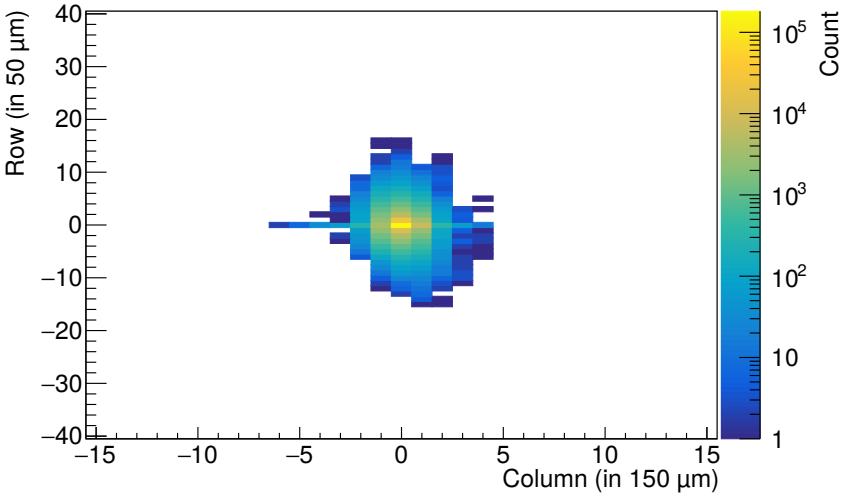


Figure 6.63: For a carbon beam at an energy that just penetrates through the first layer of the copper shielding, the largest clusters reach a diameter of almost 1 mm. For this plot, all clusters identified on the measurement have been shifted to position (0|0).

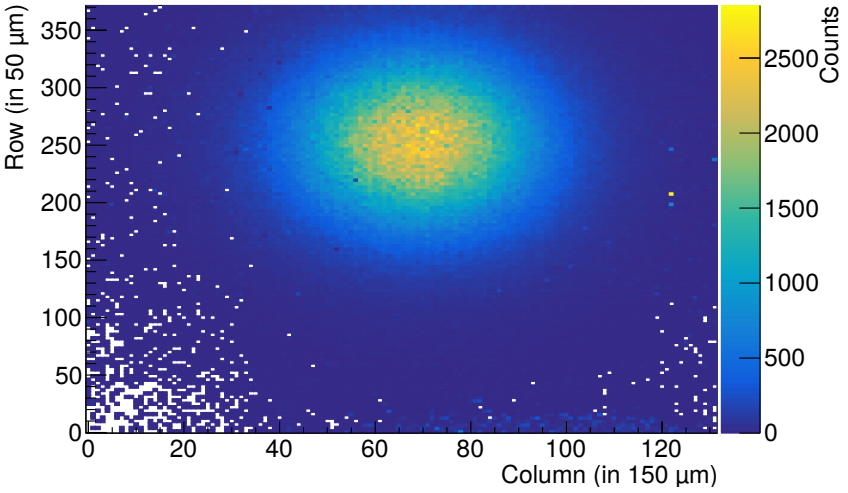


Figure 6.64: With adapted settings, the full beam profile can be captured with ATLASPix3. In this case, two thirds of the pixels have been deactivated to enhance the CAM cell to pixel ratio which is necessary because of the reduced readout speed used for this beam test.

However, with slightly less extreme settings for the triggered fraction of the trigger delay, the readout structure reaches its limits: The trigger table gets filled up from slower processing of events than new event IDs are added leading to loss of trigger signals. The consequence is that the signals block the CAM cells preventing their usage. This leads to a reduced readout efficiency which can go down to zero from this.

A completely blocked column however reduces the amount of signals that can be part of an event eventually leading to a data rate that can be managed by the readout again. With the constant trigger frequency and the limited timestamp size, the same timestamps will be triggered again eventually freeing the CAM cells again. This increases the amount of data per event again and the process starts again.

The signal rate is not distributed equally over the sensitive area. Hence, the overload of the readout is locally constrained and will alter the shape of the beam profile measured. This is shown in figure 6.65: The vertical position of the area with high signal count is aligned with the true beam profile (as in figure 6.64). This is because in the middle of the profile, the highest rate generates most signals taking the space from the signals occurring later in the outer parts of the profile. The height of the distribution is limited by the amount of CAM cells available in a column. The horizontally constant shape is a result from the reduced data rate from the middle columns with blocked CAM cells: At lower particle rates of the columns to the sides of the beam profile, the same procedure applies, just that the time until all CAM cells are blocked is longer. After all columns taking the major part of the beam are completely filled, the readout data rate is reduced below the readout limit and when the timestamps of the lost trigger signals are triggered again, the trigger buffers are cleared and the process starts again.

This efficiency loss down to no readout at all and eventual recovery can be seen from the visualisation of the same data as before in figure 6.66: The columns in the middle receiving the highest rate are completely blocked first and then the columns to the sides get inefficient until all columns covered by the beam are fully inefficient. The recovery starts at about 800 ms later with triggering the same signals again. The signals read out in this case are the old signals which becomes obvious at the end of the spill: Despite no more particles arriving on the sensor – which can be seen for the outermost columns where the rate is low

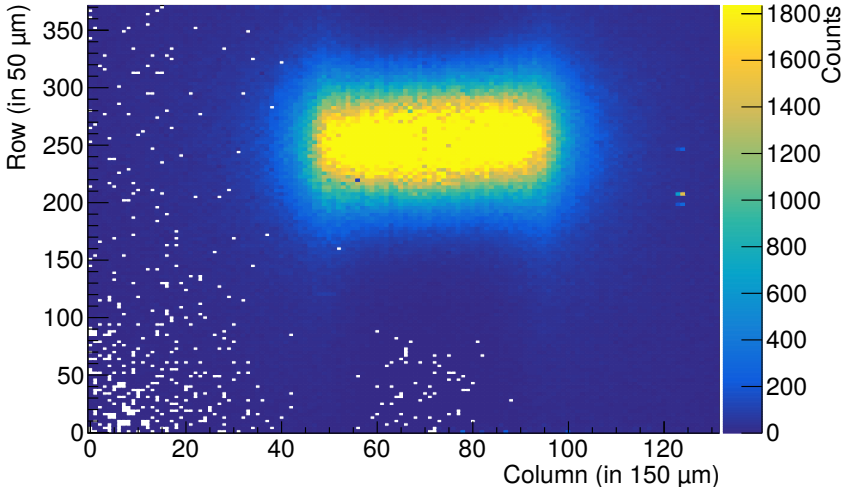


Figure 6.65: The beam profile of the carbon ion beam at HIT is cut off at top and bottom because of the lack of trigger buffers: The higher rate in the middle of the beam profile creates more hits faster than in the outer regions. As a result of the assignment of the trigger buffers to a column, the profile is only cut off in row direction.

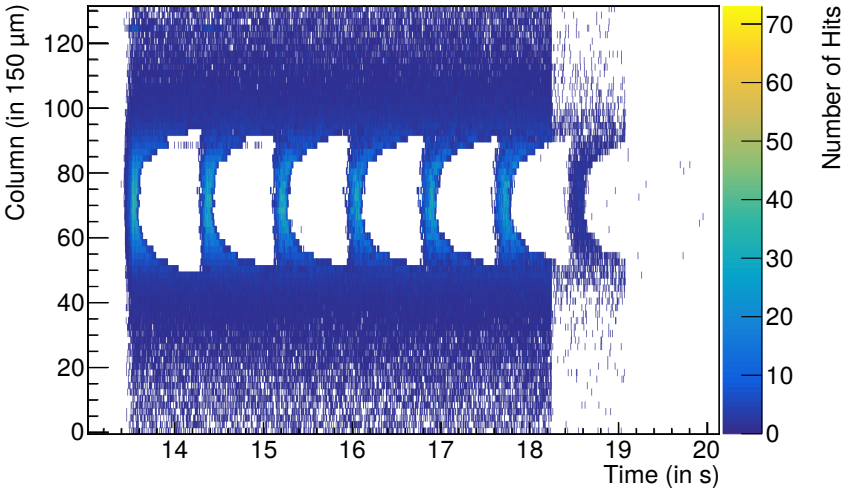


Figure 6.66: The beam generates more triggered signals than the readout can write out at the set trigger rate resulting in loss of trigger IDs. The corresponding hits stay in the trigger buffers and block the CAM cells until the same ID is triggered again later when they are read out. This blocking and late readout scheme repeats until the particle beam stops. The hits blocking the trigger buffers are then read out after the beam already stopped.

enough to not cause problems – signals are read for the columns on the center of the beam. The length of the readout also gives an indication on how long it took until all buffers were blocked.

This behaviour matches the expectations from the ROME simulations in section 5.5.

6.7 Measurements on Irradiated Sensors

To investigate the radiation hardness of ATLASPix3 and of the HV-CMOS technology in general, several samples of ATLASPix3 have been irradiated. Once with protons to the fluence expected for the design target in the ATLAS ITk upgrade and once with a combination of protons and carbon ions targeting the application of HV-CMOS detectors as beam monitor at HIT.

The proton irradiation is performed homogeneously to show fitness of the design for particle tracking in collider experiments. There, the distribution of the radiation is similar over the whole area of the die [ATL17].

For the new application field as beam monitor, ATLASPix3 has been irradiated as technology example with the beam used at HIT. ATLASPix3 is not built for this application and does not meet all requirements, but it shares radiation hard design with the dedicated sensors that are being developed at KIT-ADL for this application. The beam monitor will not receive a flat distribution of radiation, but one with a peak in the middle of the detector array and less towards the edges. The dose will not only change significantly between the single sensors of the array, but also within some. Therefore, the inhomogeneous irradiation is necessary and enables new insights on the performance of HV-CMOS detectors.

In the following, first the measurements from the homogeneous proton irradiation are presented in section 6.7.1, before the ones on the inhomogeneous irradiation at HIT in section 6.7.2.

6.7.1 Homogeneous Proton Irradiation

In the cyclotron of the ZAG[2] at Campus North of KIT, ATLASPix3 samples have been irradiated with 25 MeV protons to a fluence of $1 \cdot 10^{15}$ n_{eq}/cm^2. The irradiation has been performed to result in a homogeneous distribution over the whole area of the sensor. This fluence matches the requirement for the application in the ATLAS ITk upgrade ATLASPix3 has been designed for.

One of the irradiated samples turned out to be broken with a way too large current on the analogue supply voltage. This kind of defect has been observed on other non-irradiated samples, too. Hence, the defect was not necessarily created by the radiation. The other sample is operational and measurement results for it are presented in the following.

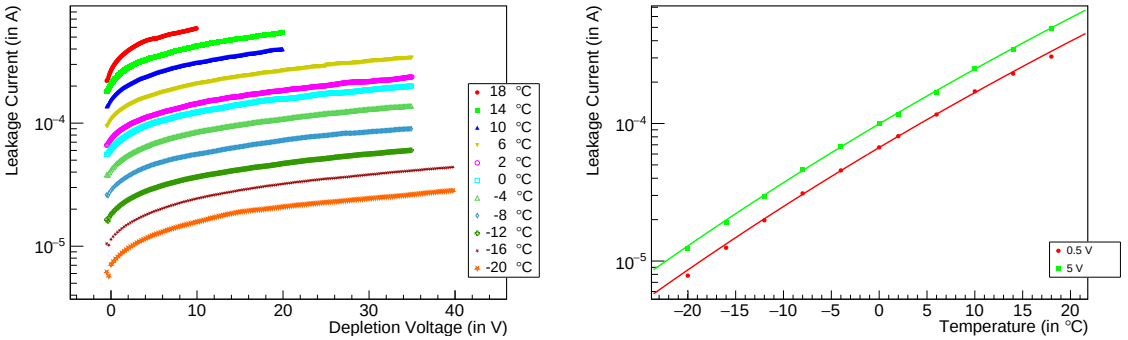
These measurements start with the influence of the irradiation on the leakage current on the depletion voltage and go on over measurements to determine the noise in the pixels both with dependency on temperature. Furthermore, the threshold tuning circuits and the digital readout features are tested before the sensitivity to ^{55}Fe gamma-rays is shown.

Leakage Current

According to the expectations, the leakage current of the irradiated sensor is increased compared to the non-irradiated sample. Around room temperature (18 °C), the leakage current surpasses 600 µA before reaching a depletion voltage of 10 V (see figure 6.67a).

Reducing the temperature, the leakage current is reduced with a characteristics matching charge generation at an effective energy of $E_{\text{eff}} = 1.12$ eV. The characteristics at 0.5 V and 5 V are shown in figure 6.67b. The over the temperature range of 38 K, the leakage current is reduced by almost two orders of magnitude for the same depletion voltage. On average, this equals a reduction of 29 % per 4 K of temperature decrease.

[2] Zyklotron AG



(a) IV curves at different temperatures

(b) Temperature dispersion of the leakage current

Figure 6.67: On the left figure, IV curves for the depletion voltage of the sensor diodes are shown for different temperatures. No breakdown is visible at the measured voltages. On the right figure, the temperature dispersion of the leakage current for fixed depletion voltage is shown: On average, the leakage current is reduced by 29 % for a temperature decrease by 4 K.

For the large currents at room temperature, the scan has been finished at lower depletion voltages. Consequently, the breakdown voltage of the non-irradiated sensor has not been reached. In this reduced interval, no breakdown has been encountered. Consequently, the breakdown is still higher than the end of the scanned interval at 40 V.

Noise

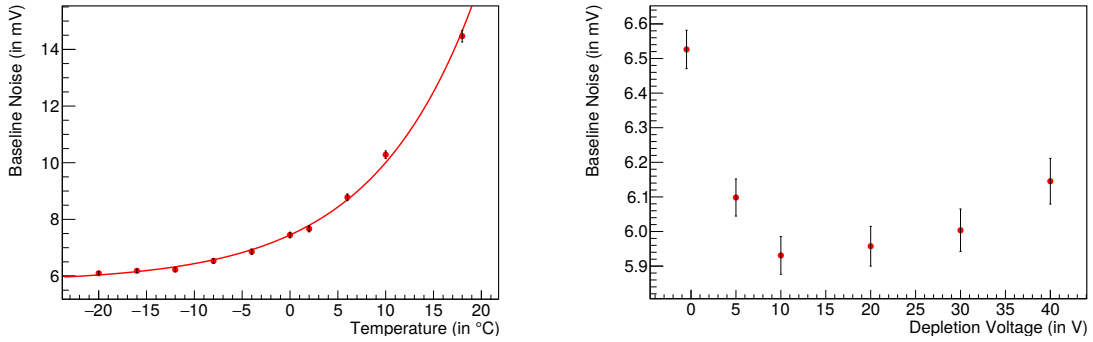
Measuring the amplifier output of one pixel for the different temperatures used for the IV curves, a decrease of the noise on the signal can be observed, too. To quantify it, the output signal of a pixel is recorded over time and histogrammed. The width of this distribution fitted with a Gaussian distribution gives a measure for the noise reaching the amplifier of the pixel. For a constant depletion voltage of 5 V, the standard deviation parameter of the Gaussian fit is shown in figure 6.68a. The increase with higher temperature does not reproduce the behaviour of the leakage current itself. Instead, the behaviour can be approximated on the measured interval by an exponential function of the form

$$\delta U(T) = a\,T\,\exp\left(b\,T\right)) + u_0 \tag{6.4}$$

with temperature in Kelvin, scaling factors a and b and an offset u_0.

For the lowest temperatures measured, a saturation is visible where the resulting noise is not significantly reduced between the measurement at −12 °C and −20 °C. Since the leakage current decreased further between these temperatures (as shown above), this indicates that there, the measured noise is not dominated by leakage current any more.

For the lowest temperature measured at −20 °C, also the depletion voltage has been scanned to probe the shielding of the in-pixel electronics from noise. The measured values are shown in figure 6.68b: For increasing but low depletion voltages, the measured noise drops before it starts to rise again for the points at 20 V and above. This increase is smaller than the decrease from 0 V to 10 V. The minimum value for the noise measured on the in-pixel amplifier is located around 10 V. The increased leakage current from the higher depletion voltage is a contribution that can increase the noise. However, this increase has to be set into context from the increased depletion volume leading to larger signals generated by charged particles.



(a) temperature dependence

(b) depletion voltage dependence

Figure 6.68: The increased leakage current from irradiation is dominating the noise generation on the sensor as visible from the measurement at the amplifier output of a pixel for decreasing temperatures: With temperature, the leakage current drops and so does the noise as shown in the left figure. The dependence on depletion voltage in the right figure shows a minimum around $-10\,\mathrm{V}$.

Content-Addressable Memory (CAM)

For measurements with triggered readout on the irradiated sample, a reduced readout efficiency on charge signal injections has been observed. A fraction of $\sim 50\,\%$ is missing even for large signals where the comparator signal hints to a detection efficiency close to $100\,\%$.

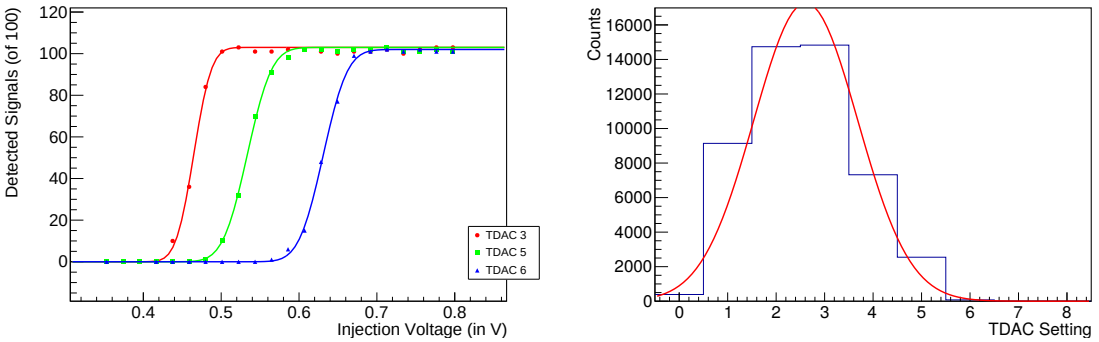
The remaining signals can be read out using the unsortedin readout option which overrides the selection of signals for an event but reads all signals present in the CAM cells. Combining the data read with the normal event-sorted readout and with the unsorted readout, the number of read signals matches the number measured with the comparator output. This assumes that the number of missing signals per column is less than the 80 which is the number of CAM cells that are available per column.

Since the signals are deleted using the first timestamp comparator and all generated signals are stored, both the timestamp and the timestamp comparator can be considered working. Accounting errors, where the actual signals are not detected but noise signals are received instead, are ruled out by choosing large signals with higher thresholds and triggering only the timestamps in which the test signals are generated.

For event-sorted readout, two components are responsible: The second timestamp comparator and the trigger table. As the two timestamp comparators in the CAM buffer are the same structure and directly adjacent, the probability of one working while the other is not is low when considering that this behaviour was visible on all tested pixels. The probability is further reduced, as it would always be the second comparator not working.

This leaves the trigger table as source of the error. It is part of the digital design of the RCU for triggered readout and present only once on the sensor. The single instance matches the homogeneity of this error on the whole matrix.

Since only one irradiated sample could be tested, it can not be ruled out that this is not a problem of the sample that existed before irradiation. But as the signals are read out sometimes and sometimes not, this hints to a problem that is introduced by irradiation. The other irradiated samples of ATLASPix3 of which the measurement results are discussed in section 6.7.2 are inhomogeneously irradiated where the dose on the periphery is close to zero. Hence, cross-checking is not possible with them.



(a) SCurves on pixel (82|119)

(b) TDAC distribution after tuning of the matrix

Figure 6.69: The SCurves measured on the ATLASPix3 sample irradiated with protons show a larger noise than the ones for the non-irradiated sensor (see figure 6.26) despite being measured at −10 °C. However, the detection threshold shift from the TDAC is still present. The functionality is not put out of commission by the irradiation. The tuning procedure results in a Gaussian shaped distribution indicating comparable behaviour of the pixel TDACs.

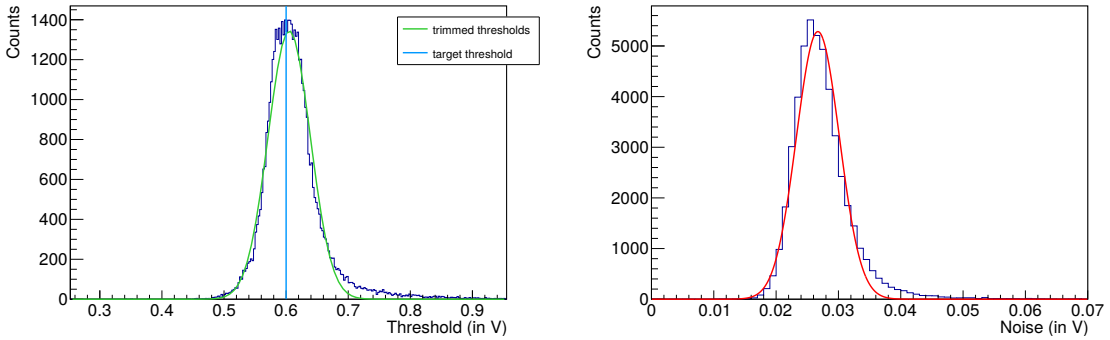
Tuning

Radiation damage on a sensor is expected to lead to changed properties and increased performance differences between the pixels. Consequently, the threshold compensation circuit becomes more important on irradiated sensors. To be able to use this structure, it first has to be verified that it is still working on the irradated sample.

This verification is done by measuring S-curves for different settings of the TDAC controlling the local threshold for the pixel. Three S-curves for one pixel are shown in figure 6.69a. The shift between the different settings matches the behaviour on a non-irradiated sample. However, the transition width of the S-curves is increased by 50 % when comparing the data taken at −10 °C for the irradiated sample with the value for an non-irradiated sample at 18 °C.

The successful tests of the TDAC circuits on the irradiated sample allow for threshold tuning. For this, the same routine as in section 6.4 is used. The target threshold is chosen higher with 3200 e^- to separate the test of the tuning circuitry from measurements to test the achievable thresholds. Furthermore, the step size for the TDACs has been increased to account for outliers in the distribution and generally a wider untuned threshold distribution. The setting of 14 results in a step size almost twice as big as used for the non-irradiated tests in section 6.4. Consequently, it is expected that the resulting distribution of the used TDAC values is narrower than the available range. The distribution is shown in figure 6.69b.

The distributions for the detection threshold and noise are shown in figure 6.70. As expectable from the larger step size for the TDAC, the resulting threshold distribution is twice as wide for the irradiated sensor. The distribution is centered around the target threshold equivalent to 3200 e^-. One contribution to the tail for higher thresholds is from pixels that can not be set to a lower threshold because of noise signals compromising the performance at these settings. The noise distribution shows the same shift to higher values that has been addressed on the S-curves measured prior to the tuning. The tail of the distribution to larger noise values comes from the uppermost rows of the matrix where the values rise towards the end of the die. The corresponding figures are included in appendix D.



(a) Threshold Distribution (b) Noise Distribution

Figure 6.70: The threshold and noise distributions resemble the ones from tuning before irradiation (see figures 6.30 and 6.32). But the target threshold was chosen higher with an injection signal of 0.6 V and the noise in the pixels is larger.

For the measurement, the triggered readout with the unsortedin option has been used as explained in the previous section. This feature enables the tests and leads to the insight that the tuning circuits on ATLASPix3 are working after proton irradiation irradiation.

Iron-55 Signal Detection

The thresholds set in the previous section are too high to be able to detect signals from ^{55}Fe decays. This requires a detection threshold below 1600 e^-. On the irradiated sample, this threshold is not achievable for all pixels and the threshold has to be adjusted to control the noise signal rate to distinguish the signals from the ^{55}Fe source from noise. This measurement is only possible at low temperatures because the noise signal rate would exceed the signal rate from the source preventing the identification of these signals. Therefore, the measurement has been conducted at $-20\,°C$ to bring down leakage current and noise signal rate.

The TDAC settings have been adjusted in the columns 59 to 90 to be able to detect the ^{55}Fe signals. The remaining pixels in the rows above row 207 (inclusive) are set to a higher threshold which results in some signals being detected, but at a very low rate. The other pixels in the rows below 207 have been set to the maximum TDAC value to reduce the amount of noise signals during the measurement. Also, several pixels in the bottom left corner have a lower setting of the TDAC value.

The source is placed at the bottom of the matrix horizontally around column 65. This leads to the expectation of a rate gradient from this point. The hitmap for a measurement of 15 hours is shown in figure 6.71: A significant fraction of the pixels in the prepared area in the column range from 59 to 90 show a much higher number of signals than expected from the surrounding pixels. Not taking these into account, the expected gradient is visible.

From this measurement, it can be concluded that the X-rays from the ^{55}Fe decays can still be detected. Thresholds below the 1639 e^- can be achieved after irradiation. It requires low temperatures to reduce the noise signal rate sufficiently for identifying the signals from the source.

6.7.2 Inhomogeneous Irradiation with Protons and Carbon Ions

At the beam dump at HIT, five ATLASPix3 samples have been irradiated with protons and carbon ions with energies around 110 MeV/u. This irradiation was performed while

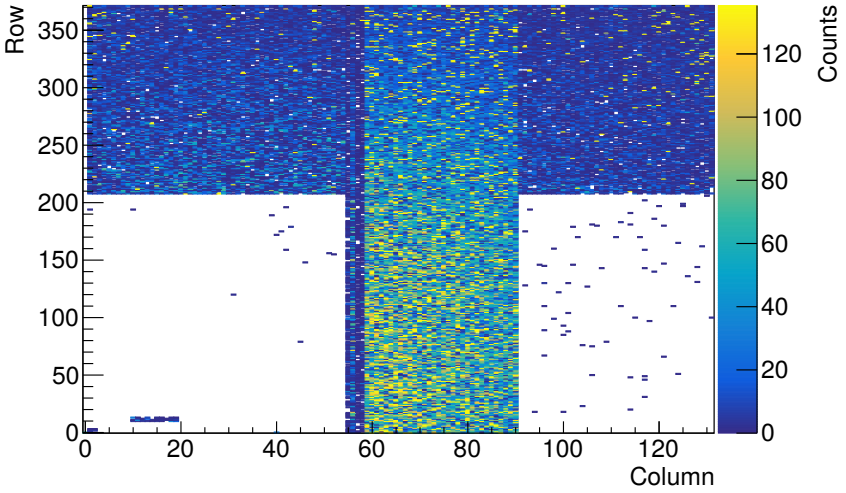


Figure 6.71: The iron-55 source was placed centered at the bottom end of the matrix. The columns 59 to 90 have been adjusted to detect the signals as well as the rest of the pixels in the rows 207 and above with a higher threshold. The remaining pixels are set to a higher threshold to to avoid high hit rates from noise on those pixels. These noise hits also exist in the adjusted columns, but there, a compromise between sensitivity and noise hit rate was made. In between these noisy pixels, a circular gradient is visible indicating to the position of the source.

the beam was not used for patient treatment, meaning that the irradiation process took several months. The samples were mounted in a stack that was traversed by the particles. To simplify the irradiation setup, the samples were not irradiated homogeneously, but the beam was focused on one spot of the samples and remained there for the whole irradiation. This setup has two advantages: Firstly, no mechanics are necessary to move the samples during the irradiation so that a static mounting structure could be used. Secondly, the irradiation time is greatly reduced as not the whole matrix is irradiated.

Furthermore, this inhomogeneous irradiation scheme is more similar to the conditions that will be experienced by beam monitoring detectors: Most of the time, the beam will be passing through the middle of the beam monitor leading to much higher radiation doses in the center compared to the sensors on the edges of the detector.

One after the other, the samples were removed from the beam dump to obtain samples at different fluences. As the irradiation has been performed as a parasitic application in

Table 6.2: The irradiation doses of the ATLASPix3 samples irradiated at HIT are split into the doses for proton and carbon ion irradiation and given as percentages of the annual doses expected in operation. Due to the irradiation setup, these numbers include an uncertainty of at least 10 %.

Sample	Annual Proton Dose	Annual Carbon Dose
AP3.H1	37 %	19 %
AP3.H2	85 %	45 %
AP3.H3	92 %	49 %
AP3.H4	153 %	85 %
AP3.H5	155 %	85 %

the accelerator at HIT, the exact doses on the samples are not known. Estimates from average properties of the particles bunches used for the irradiation are taken to quantify the doses on the samples. These estimates contain an uncertainty of at least 10 %. For expected annual fluences of particles of $6 \cdot 10^{12}$ protons/cm^2 and $6 \cdot 10^{11}$ carbon ions/cm^2, the relative annual doses on the samples are given in table 6.2.

All but the fourth sample (AP3.H4) work. The broken sample has been examined to find that the defect seems to be located outside the irradiated area indicating that the sample has been broken before irradiation. Without a probing possibility to test the functionality of the samples before irradiation, this risk had to be taken.

The last sample (AP3.H5) has been taken from the beam dump shortly after the fourth sample and for the additional dose on the fifth sample, the numbers contain an even larger uncertainty.

With the irradiated spot being smaller than the pixel matrix of ATLASPix3, the placement on the Kapton foil in the beam dump mounting structure had not to be very precise. A shift of the irradiated area influences only the selection of pixels that are investigated.

With all working samples showing a consistent behaviour, only the first and the last sample will be considered in this discussion reducing the length. In the following, the irradiation profile on the samples will be discussed first. It is slightly different on each chip, but the shape stays the same for all of them. Then, the leakage current behaviour and dispersion with temperature are discussed on the example of AP3.H1 with the plots for AP3.H5 in the appendix. This is followed by the influence of the depletion voltage on the operation of the digital readout of the sensors. As the pixels with accessible amplifier output are not among the irradiated pixels, the noise estimation and temperature influence on it are studied afterwards using detection threshold measurements using S-curves. The test of the charge injections are then complemented with source and TCT measurements on AP3.H1. At the end, a measurement to recover the insensitive pixels on AP3.H5 is presented including an estimation for necessary changes on the threshold tuning circuits to cope with the inhomogeneous irradiation expected at HIT.

Irradiation Pattern

The irradiation of the samples at the beam dump of HIT results in a spot-shaped irradiation pattern. It is smaller than the chip area and pointed always at the same position in the matrix. This procedure has two advantages for the irradiation of the sensors: Firstly, no mechanics for movement have to be included in the irradiation setup making the system simpler. Secondly, the beam is concentrated to one spot reaching significant doses in shorter time. This is especially important as this irradiation can only happen when the beam is not needed for treatment of patients and beam time is consequently limited.

As a result, the effects of the radiation on the pixel and the in-pixel electronics can be separated from the radiation effects on the readout electronics in the periphery. Furthermore, irradiated and non-irradiated pixels can be compared on the same die, giving a result closer to tests before and after irradiation.

The proton and carbon ion beams used for the irradiation are not tuned as they are on the patient treatment sites. The shapes of the two beams differ slightly in shape and position. The proton beam is more elongated than the carbon beam.

These shapes are visualised in figure 6.72 as a noise signal map of one of the irradiated samples overlayed with markers to identify the beam shapes. These shapes have been recorded with films placed in front of the ATLASPix3 samples.

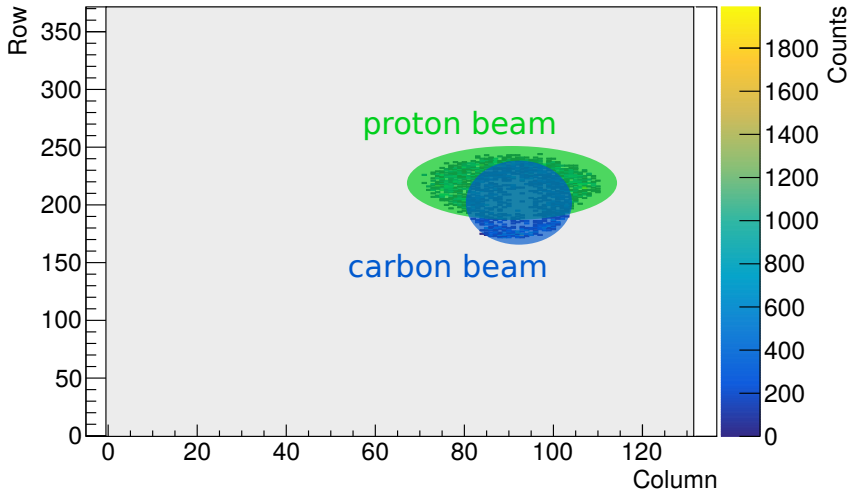


Figure 6.72: The beam spots of the proton and carbon ions used for irradiation at HIT overlap but are not of the same shape: The proton beam spot is more elongated and hitting the sample at higher rows than the carbon beam spot. The entries in the histogram are noise signals distinguishing the irradiated area from the non-irradiated area. The matrix area is marked with a light grey background to keep the aspect ratio for the plot.

The samples have been mounted in a frame using Kapton tape. The placement has been done by hand resulting in a shifted position of the irradiated area between the samples. However, as only a part in the middle of the matrix is irradiated, this does not influence the measurement. The downside of this choice is that the amplifier output available only for the pixels in row 0 is not available for any of the irradiated pixels.

The area covered by the irradiation is about 5 % taking the region detecting noise signals from figure 6.72 as reference. This means that for a full area irradiation of the chip, the leakage current will be higher by a factor in the order of 20.

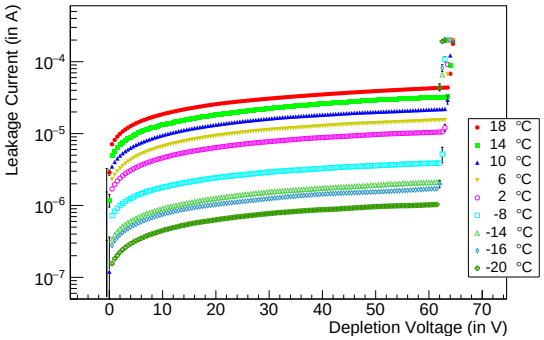
Leakage Current

When measuring the leakage current of the samples irradiated at the HIT beam dump, it is important to remember that only a fraction of the matrix has been irradiated, so that the leakage current has to be scaled up to obtain the numbers for the whole chip being irradiated to the dose of the irradiated spot.

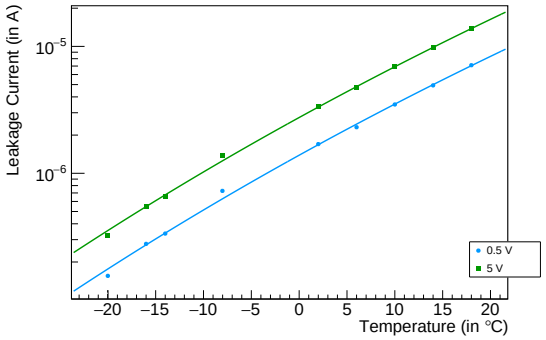
With the leakage current of the non-irradiated ATLASPix3 at around 60 nA, the expected changes in leakage current are measurable despite only a fraction of the matrix being irradiated. For the proton irradiation in section 6.7.1, the leakage current was increased by three to four orders of magnitude – depending on the temperature measured at. For a similar increase of the leakage current density in the irradiated area, this results in an increase by one to three orders of magnitude.

The leakage current of the sample AP3.H1 is shown in figure 6.73. The leakage current at room temperature is larger by three orders of magnitude compared to the non-irradiated value. With decreasing temperature, the leakage current drops according to the expectation for the generation current hypothesis as used in section 6.3.3.

The breakdown voltage around 63 V matches the range from the non-irradiated samples. With temperature, a shift of the breakdown voltage has been observed which evaluates in about 0.5 V per 10 K. This behaviour is shown in figure 6.74. The breakdown point has been

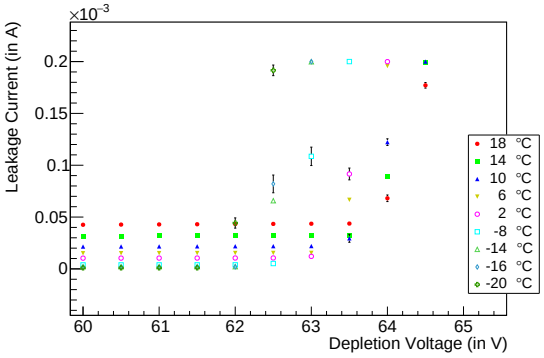


(a) Sensor diode leakage current

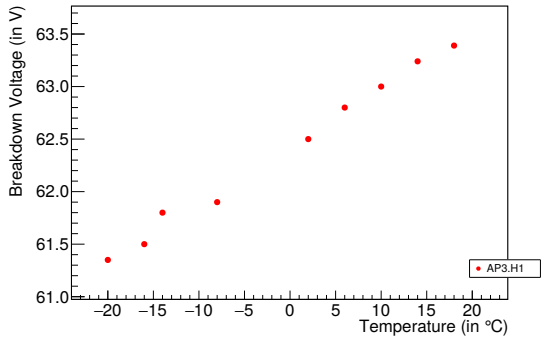


(b) Temperature dispersion of the leakage current

Figure 6.73: The sensor diode leakage current for the whole matrix consists of the contributions from the irradiated spot with a high leakage current and the unirradiated part of the pixel matrix of which the contribution to the total leakage current should be small. Compared to the leakage current from the homogeneous irradiation with protons (see figure 6.67), the leakage current is more than an order of magnitude smaller as expected from the size of the irradiated spot. An early breakdown is not visible compared to the non-irradiated sensor and the temperature dispersion (right figure) follows the generation current model.



(a) breakdown of the sensor



(b) breakdown points

Figure 6.74: With decreasing temperature, the sensor diode leakage current is lower for depletion voltages below the breakdown. But the breakdown point moves towards smaller depletion voltages with decreasing temperature. This point is evaluated as intersection of line fits to the points after the breakdown and the ones before. The decrease over the 38 K investigated is 2 V.

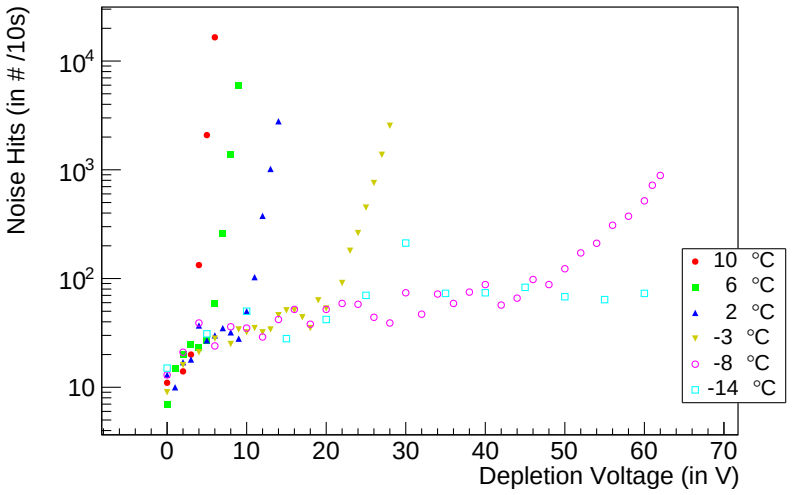


Figure 6.75: The increased leakage current from the irradiated pixels results in leakage-induced hits. The frequency of these noise hits is determined by the amount of leakage current. For lower temperatures, the leakage current is reduced (see figure 6.73) and higher depletion voltages can be applied without significant increase of noise hit rate.

evaluated as intersection from straight line fits to the points above the breakdown voltage and the last points before it. This results in a linear shift to lower breakdown voltages at lower temperatures. The same behaviour has been measured for a non-irradiated sample, too (see figure 6.24).

For the sample AP3.H5 with the highest irradiation dose, the leakage current behaviour is similar to the one of AP3.H1 but with higher currents. Since the behaviour did not change between the working samples, only the plots for AP3.H5 are included in the appendix, omitting the ones for the samples in between.

From the unchanged temperature dispersion and absolute value, it can be concluded that the irradiation did not introduce a breakdown mechanism limiting the depletion voltage to smaller values than the avalanche breakdown present on the non-irradiated sensor.

The additional plots are included in appendix G.

Noise Signal Threshold

For irradiated detectors, it is important to use the depletion voltage to collect the generated charge in the pixels. The increased leakage current after irradiation, however, prevents the usage of large depletion voltages at some point. Then, the leakage current creates a sufficient amount of noise to exceed the signal detection threshold of the pixel. Partially, this can be compensated by higher thresholds for pixels generating such noise signals. But at some point, the amount of noise generated in the pixel becomes so large that signals to detect are not larger than noise any more and the pixel becomes insensitive.

Decreasing the temperature, at which the irradiated sensor is operated, reduces the leakage current and consequently the amount of noise. For this reason, it can be expected that larger depletion voltages become achievable without losing pixels due to noise.

To test this, the depletion voltage is scanned at different temperatures and the recorded number of noise signals per time is evaluated. The resulting plot is shown in figure 6.75: As the leakage current drops with decreasing temperature, the noise signal rate drops, too.

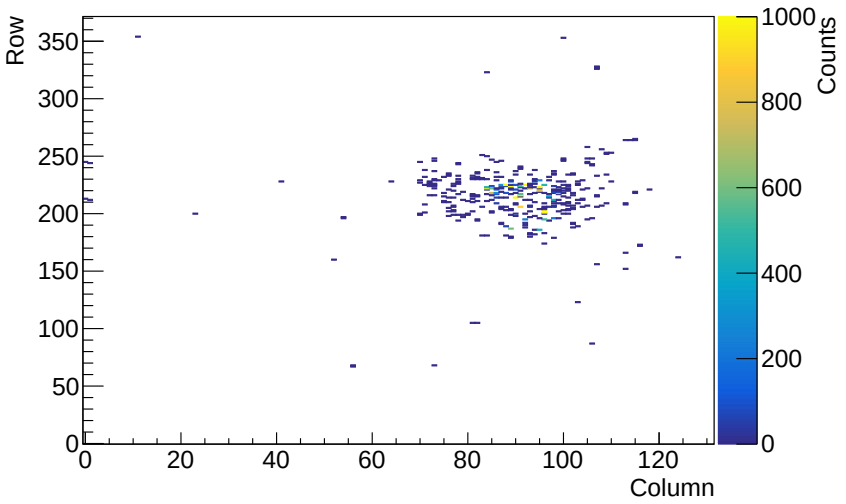


Figure 6.76: Almost all noise signals recorded during the depletion voltage scans are located in the irradiated area. It has been measured on AP3.H1, the same sample for which the noise map to show the irradiation spot in figure 6.72 is taken.

Consequently, the maximum possible depletion voltage increases for lower temperatures. At room temperature, the possible depletion voltages are close to no depletion voltage at all. Therefore, only temperatures at 10 °C and below are included. For 10 °C, the noise signal rate increases rapidly at 4 V. At −8 °C, the noise signal rate stays stable until above 40 V before an increase starts. It is much slower than for the data line at 10 °C. For −14 °C, no similar increase can be identified until the depletion voltage where the problem arises independent from irradiation.

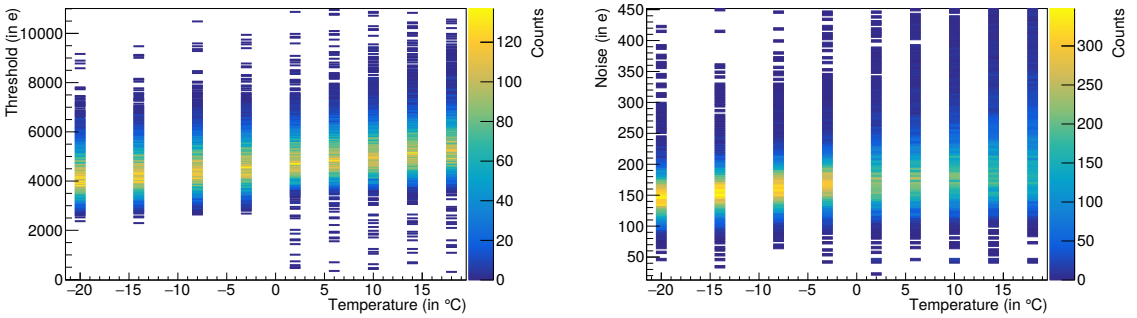
The noise signals detected for the respectively highest depletion voltages at each temperature are combined in one hit map to verify that apart from few exceptions, the signals originate from the irradiated area. The signals outside the irradiated area can be justified from cosmic particles passing through the sensor matrix during the measurement. The hit map is shown in figure 6.76.

Threshold Dispersion and Noise Estimation

As directly measuring the noise inside the irradiated pixels is not possible on the inhomogeneously irradiated samples, the noise in the pixels has to be estimated using the S-curve measurement from threshold tuning. Furthermore, the fraction caused by the leakage current can be estimated from comparing the noise estimation at different temperatures. For this, a rectangular region in the center of the irradiated region is tested at temperatures ranging from room temperature down to −20 °C. As transistor properties change with temperature, not only the noise estimation is expected to change, but a shift of the threshold is expectable, too.

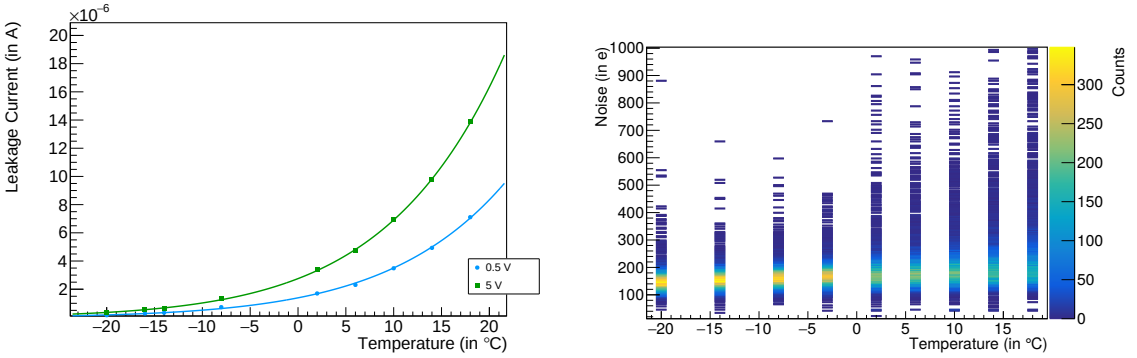
The distributions of the thresholds and the noise estimates for the different temperatures are shown in figure 6.77. The injection voltages have been converted to an electron equivalent for the plots. The average threshold value drops over the 38 K temperature difference by 1200 e^-. The MPV for the noise gets smaller over this range, too. The change is 40 e^- for the MPV, but the major difference is the larger width with more pronounced tail for higher temperatures.

To compare the noise estimate to the leakage current, the corresponding dispersion plots are printed side by side in figure 6.78. The vertical axes of both plots have been adapted



(a) Threshold dispersion (b) Noise dispersion

Figure 6.77: The threshold distribution in the irradiated area of the ion irradiated sample shows a decrease of the detection threshold with decreasing temperature. This behaviour does not differ from the behaviour of the non-irradiated sample measured as reference (see appendix H). The decrease of the noise and its spread between pixels in the right figure with decreasing temperature in addition allows for a lower setting of the threshold resulting in lower possible settings.

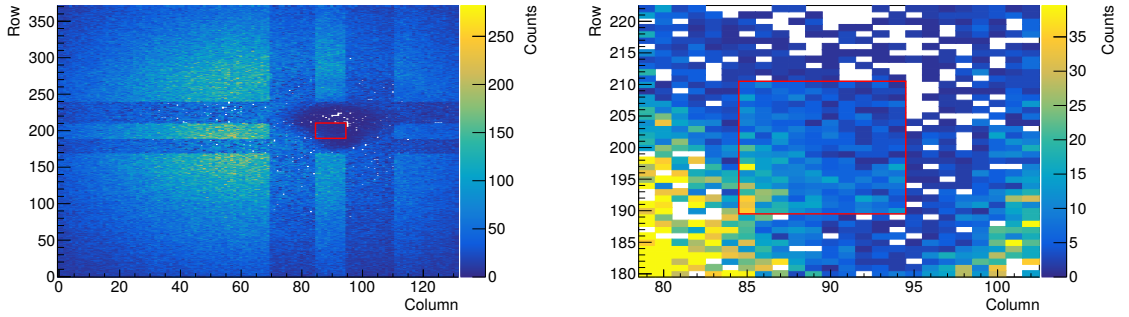


(a) Leakage current (b) Noise estimate from S-curves

Figure 6.78: The leakage current dispersion with temperature for two different depletion voltages from figure 6.73b – drawn without logarithmic vertical axis – shows the same behaviour as the tail of the noise estimate dispersion from the S-curve measurement in figure 6.77b.

to ease the comparison. The leakage current dispersion matches the length of the noise distribution tail from the S-curve noise estimation. The MPV of the distribution does not show the same characteristics, indicating that – for the majority of the pixels – charge generation is not the main cause of leakage current and not the dominant source of noise.

On an non-irradiated sample, the detection threshold changes with temperature, too. This is shown in appendix H. The temperature range covered by this measurement is smaller. But the change on this temperature interval is larger for the irradiated sample, too. One aspect of this change are the adapted settings used for the irradiated sample, that are necessary to obtain a working state for the irradiated pixels. This includes a larger feedback current setting (VNFB) as well as a stronger coupling of the comparator input to the baseline voltage (BLRes). However, the properties of the transistors in amplifier and comparator will change from the radiation as well influencing the performance. Such changes in transistor properties have been measured for example in [Ehr15].



(a) full matrix hit map

(b) zoomed-in hit map on the tuned region

Figure 6.79: A rectangle of 10×21 pixels in the beam spot has been tuned and the thresholds for the irradiated area around the tuned pixels has been increased. This has been done as whole columns and rows. The resulting hitmap for exposure to a ^{90}Sr source with a rectangle indicating the border of the tuned area is shown in the left figure. On the zoom-in figure on the right, it is visible that all tuned pixels detected a similar amount of signals whereas the untuned pixels have a much more wide-spread number of counts.

Response to Particle Signals

The charge injections for testing are injected into the diode of the pixel. However, this process differs from the charge generation by particles passing through the sensor. There, the charge is generated inside the depleted volume around the pn junction of the pixel. In this volume, the charge has to be separated and collected. The charge injection is a unipolar process where only one carrier type is involved.

Therefore, the test signal injections have to be complemented with measurements with particles. For the inhomogeneously irradiated ATLASPix3 samples, this complementary test has been done with a ^{90}Sr source. The Landau distribution of the amount of charge generated by the electrons traversing the pixels' depletion zones offer a large range of signals to also detect a small number of signals for pixels with high thresholds.

As the irradiation has been done with a stationary particle beam, the irradiation dose in the middle of the irradiated area is the highest. The changes from the radiation are therefore expected to be the largest there. For the measurement on AP3.H1, the outer part of the irradiated area has been set to the highest TDAC setting of 7 and the middle part has been tuned to obtain comparable results between them. The tuning target is comparable to the thresholds of the non-irradiated part of the matrix. These pixels with the TDAC value of 7 will be referred to as less sensitive area. To achieve the reduction of the noise signal rate close to zero on most pixels, also the step size of the TDACs had to be increased reducing the granularity for tuning in the region of interest. Then, the ^{90}Sr source has been placed over the matrix.

The resulting hit map is shown in figure 6.79. With the pixels at the border of the irradiated area, also the rest of these columns and rows have been set to the high thresholds at the TDAC value of 7 adding to the less sensitive area. This results in lower numbers of detected signals in these pixels framing the tuned area. This tuned area is in addition marked with a red rectangle in the plots.

From the columns zero to 69, the position of the source above the pixel matrix can be reconstructed: The source has been placed centrally above the sensor.

In the less sensitive region, the central part of the irradiated area can be identified as darker

area with less signals than the rest of the less sensitive area. Interestingly, the tuned part of the matrix in the irradiated spot does not show a significantly higher number than the surrounding pixels in the less sensitive region. From the enlarged view in figure 6.79b, it is visible, that all pixels have detected signals, in contrary to the surrounding pixels set to the high threshold.

Furthermore, the gradient in the count rate between the pixels on the bottom left of the tuned region compared to the top right end is not visible using the test charge injections. This hints to a difference in charge collection efficiency from the pixel volume that is made visible from the dose gradient over the tuned area.

Together with the different settings necessary to operate the pixels and detect the signals from the charge injections, this leads to the conclusion that both the in-pixel electronics and the charge collection – i.e. the bulk material – have been compromised by the irradiation.

The different effects of the irradiation become prominent at different fluences which becomes visible from the irradiation profile: The outermost part of the irradated area shows a reduced signal rate when exposed to a radiative source as ^{90}Sr. Charge injections work normally. Going further inside of the irradiation profile, the noise signal rate is largely increased preventing sound statements on the detection efficiency of particles without usage of external time references. At the center of the irradiation profile with the highest fluence, neither particles nor charge injections are detected any more.

TCT Efficiency Measurement

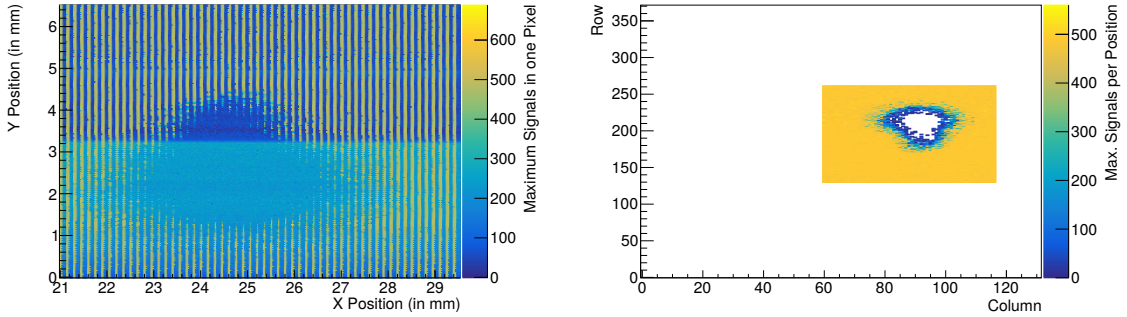
Since the particles traversing the detector at HIT are protons and carbon ions, their signal is likely larger than the one from the β particles from the ^{90}Sr source. Therefore, the irradiated area is also tested with the TCT setup described in section 4.2. The laser is capable of generating these larger signals.

As the irradiated samples are standard-thickness dies, the laser can be used only from the front-side, because the signal will not reach the depletion zone then illuminating the sensor from the backside. This is especially the case as this measurement had to be conducted at room temperature at which only minimal depletion voltage can be used.

From the front-side, the sensor diodes are shielded by the metal layers. Therefore, it is not possible on the whole area of the pixels to detect a signal.

With light pulses that saturate the amplifier output in the unirradiated part, the matrix surface is scanned around the irradiated spot resulting in the plots in figure 6.80. From the geometry inside the TCT setup, the vertical axis in TCT coordinates is mirrored with respect to the other plots shown here. For each position of the laser collimator, the number of signals for the pixel with most signals is shown. This assumes that the number of pulses generated is larger than the number of noise signals detected by a pixel at the same time. The 500 pulses sent at each position are only reached in narrow vertical lines along the pixel columns with areas in between where the signals are not detected. This is not a radiation effect as described above and justified by the lack of difference to the top right corner of the TCT coordinates plot where the irradiation dose has been low compared to the center. The background between the sensitive lines are filled with numbers from pixels detecting noise signals. At these positions, the assumption that the pixel with the largest number of signals is the pixel located below the laser, does not hold.

In the irradiated spot, the sensitive lines get narrower before vanishing completely in the central part leaving all bins filled with the signal counts from pixels with noise signals. The horizontal step in the background counts are likely from changed conditions during the measurements.



(a) Sensitivity in TCT coordinated

(b) Sensitivity in ATLASPix3 coordinates

Figure 6.80: The irradiated area on one ATLASPix3 sample has been scanned with a pulsed laser. The left figure shows the maximum number of signals detected by any pixel on the chip for each position of the laser. The vertical stripes with less signals are due to the metal layers on top of the sensor diode which shield the sensor from the laser photons. In the irradiated area, also the stripes with high detection efficiency vanish exposing the irradiation pattern. The pattern ist mirrored due to the axis directions of the kinematic system of the measurement setup. On the right figure, the same data is shown but with the pixel address as spatial information: For any pixel, the maximum number of signals for any position of the laser are put on display exposing again the shape of the irradiated area.

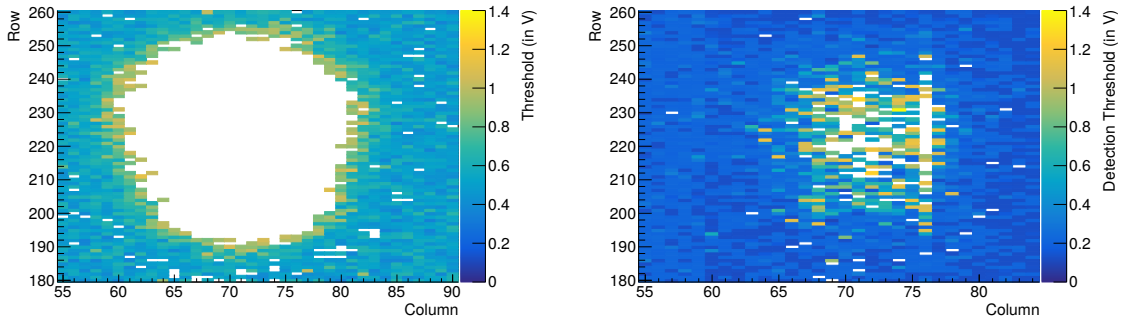
Reversing the relation and going through the pixel addresses of ATLASPix3 and checking for the maximum number of signals detected for any position of the laser collimator, figure 6.80b is received. From it, it can be seen that towards the non-irradiated part in the corners of the scanned area, all pixels have seen the full 500 signals generated for some position. But towards the center of the irradiated spot, a transition with less received signals exists leaving the pixels in the middle completely without signals for any position.

With illumination from the front side, it can not be ruled out, that radiation damage close to the surface of the detector absorbs the laser light, leaving no photons to generate charge in the depletion region and hiding that the pixels would actually be sensitive. However, since this pattern matches the findings of the ^{90}Sr source measurement, it is unlikely that the laser light is completely absorbed before reaching the sensor diode. In this case, the charge collection or signal amplification has to be compromised.

Recovery of Detection Efficiency

On AP3.H5 using settings as for AP3.H1, even charge injections are not detected any more by the pixels in the middle of the irradiation spot. A measurement of the detection threshold over the irradiated region reveals a spot without a sensitive pixels for any signal size achievable with the charge injection feature implemented on ATLASPix3. The threshold map of this measurement is shown in figure 6.81a.

The large step size, necessary for the pixels at the outer part of the irradiated area, prevents precise adjustment of the thresholds for the pixels in the middle of the irradiated area with the highest dose, further diminishing the sensitivity of this area. These contradicting needs for fine adjustment of the threshold in the middle of the irradiation profile and large shifts at the periphery of the irradiation profile arise from the inhomogeneous irradiation for which ATLASPix3 is not built. In its application, the radiation distribution is expected to be homogeneous where higher thresholds can be achieved with a higher global threshold.



(a) unoptimised efficiency

(b) combined result from the scan

Figure 6.81: On the sample irradiated to the highest dose, the detection efficiency with charge injection shows an insensitive area with default chip settings in the left figure. Optimising the settings for the chip especially with increasing the amplifier supply voltage can recover most pixels: The right figure shows the detection thresholds for the same area of the same chip. However, this plot is a combination of several measurements with slightly different threshold settings to find working settings for the maximum number of pixels.

On this application however, the low thresholds are required for the pixels in the middle of the irradiation profile. This spread can only be achieved with larger step size for the TDACs.

Comparing the measurements with injections and electrons for the AP3.H1 and AP3.H5 it becomes clear, that both the charge collection is impaired and the properties of the amplifier and comparator are changed: On AP3.H1, the injections are still working comparable to the non-irradiated case while the signals from the ^{90}Sr source show a reduced detection efficiency. This hints towards a change in charge collection. The lack of detection efficiency for the charge injections on AP3.H5 points towards changes in the properties of the in-pixel electronics as the charge collection presents itself to be less important for the charge injections.

To compensate for the changed properties of the in-pixel electronics, the VSSA voltage has been increased from 1.25 V to 1.60 V. This larger supply voltage changes the voltage levels in the electronics. These shifts can partially compensate for the radiation induced changes. With the larger VSSA voltage, the insensitive area in the middle of the irradiation profile got smaller meaning that the pixels at the edges of this region gained back sensitivity to charge injection signals.

The large steps between the TDAC settings lead to some pixels being below the noise threshold after this VSSA increase for a setting at which no signals could be detected before. However, the next higher TDAC setting may already be too high again.

To obtain an estimation of the fraction of the pixels that can be set to working conditions, the global threshold is gradually lowered and pixels that fall below the threshold level at which noise signals are repelled get their TDAC setting incremented. After the TDAC setting distribution shifted by more than one TDAC setting on average, the best detection threshold found for any setting is taken for combining them in one plot. This plot gives an indication on the achievable performance if the TDAC steps were as fine as the global threshold steps. The scan can be terminated after this shift, as not much gain can be expected after a shift by one TDAC setting. The independent shifts from global threshold and TDAC will reproduce the characteristics already measured.

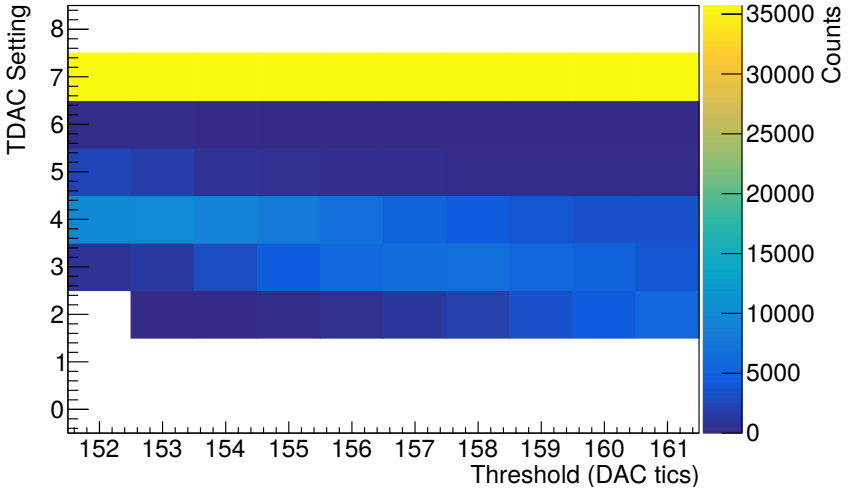


Figure 6.82: For the different global threshold settings, the distribution of the TDAC settings is shown: With decreasing threshold, the pixels start to show increased noise and eventually the threshold falls below the baseline. In this case the TDAC value is increased. In the covered interval of the global threshold, the TDAC setting of the pixels shifts by two steps indicating a sufficient size of the scanned interval. The large number of pixels with the maximum setting of 7 are from the rest of the matrix that is not looked at here.

The resulting detection threshold map is shown in figure 6.81b. The distribution of the TDAC values during the scan of the global threshold is shown in figure 6.82. Since the whole matrix is included in this plot, the highest setting, which was used for all pixels outside the area shown in figure 6.81, these bins contain the largest number of entries. The measurement scan is ended at the shift of about two TDAC steps to make sure it is at least one step for each pixel in this area. With the wider distribution at the highest global threshold setting, some pixels may have not changed the setting when stopping at an average setting of three.

The pixels in the irradiated region started at a TDAC setting of 2 if possible for the global threshold setting of 161. The steps of this 8 bit DAC equal a step of 7 mV. This is the actual voltage supplied to the comparator and can not directly be compared to the voltage numbers used for the injection signals. With the decreasing global threshold towards 152, gradually more pixels need to get the TDAC value incremented until the distribution of the settings for the irradiated pixels peaks at a setting of 4 for the threshold setting of 152.

With this procedure, 94 % of the pixels in the irradiated area could be set to working conditions. For any individual setting of the global threshold did not exceed 82 % as shown in figure 6.83. This means that granularity is missing in the TDAC setting to compensate for the differences between the pixels.

For future sensors that are to face this kind of inhomogeneous irradiation, from this measurement, an estimation of the number of necessary TDAC bits can be derived: The shift of the global threshold by 10 DAC steps is covered by a shift of the average TDAC value by 2 steps. This means that to obtain a similar resolution with the TDAC circuits as with the global threshold with this step size, two additional bits are necessary.

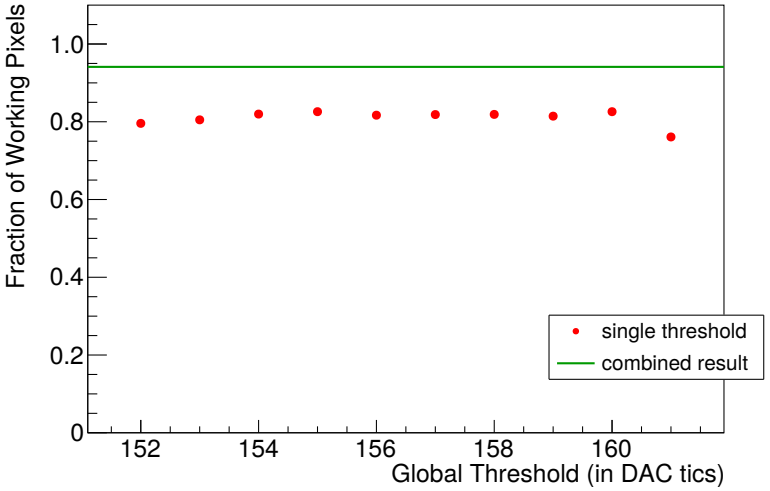


Figure 6.83: For any setting with the corresponding TDAC settings, not more than 82 % of the pixels are working. Combining the optimum conditions for the pixels from the different settings of the global threshold, for 94 % of the pixels settings exist at which they become sensitive to charge injections again.

6.8 The ATLASPix3 Beam Telescope

The four-layer beam telescope has been designed reusing as many components as possible from the single detector setup. Hence, the telescope uses the carrier PCBs from the single sensor tests. They are connected to the GECCO board via an additional PCB that is interfacing between the four PCIe connectors for the four layer PCBs and the PCIe connector to the GECCO board. A photograph of the system is shown in figure 6.84.

The telescope PCB has been produced in several iterations. The first one has been used for a beam test at DESY in 2019 and for secondary particle tracking tests at HIT in 2021. The latest iteration of the telescope PCB is – just as the single sensor PCBs – going to be distributed in the FCEPC project.

In the following, the adaptions in hardware, firmware and software from the single detector setup to the telescope system are sketched. They are also the blueprint for the quad-module firmware and software developed by B. Raciti for her master thesis [Rac21] under co-supervision of the author. For the quad-module, the hit-driven readout used for the telescope is replaced with triggered readout.

Hardware

It distributes power and data lines and routes the outputs of the layers to different inputs of the FPGA.

For configuration, the serial configuration option without external shift registers is not possible to reduce the necessary number of signals and the load line is vetoed with `OR`-gates to keep the configuration procedure after layer selection identical to the single detector setup. The same is done for the chip-select lines of the SPI interface.

The data outputs for digital readout and the CMD are distributed via fanout buffers. On systems as the quad-module, this is not required, but the single chip carriers have a termination for each of the lines and do not suffice the short line extension limitation for supplying several receivers from one differential line. As for the high power consumption of

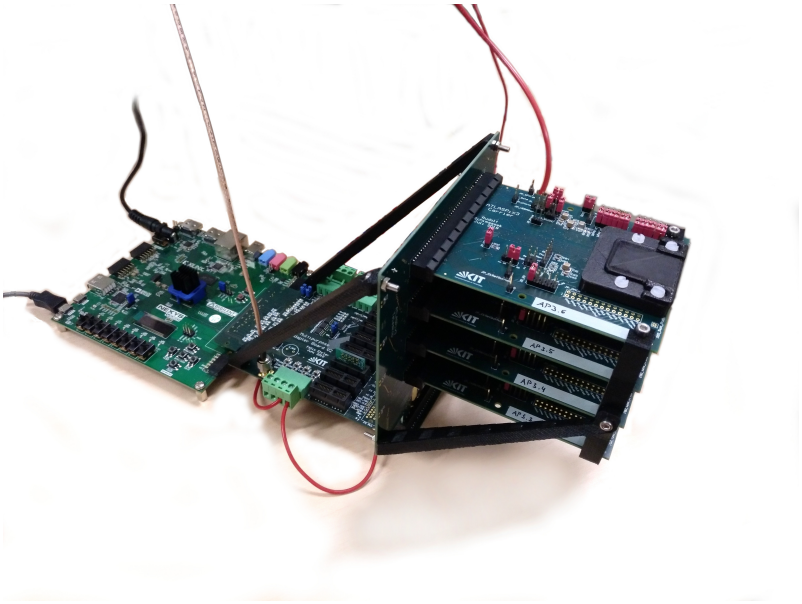
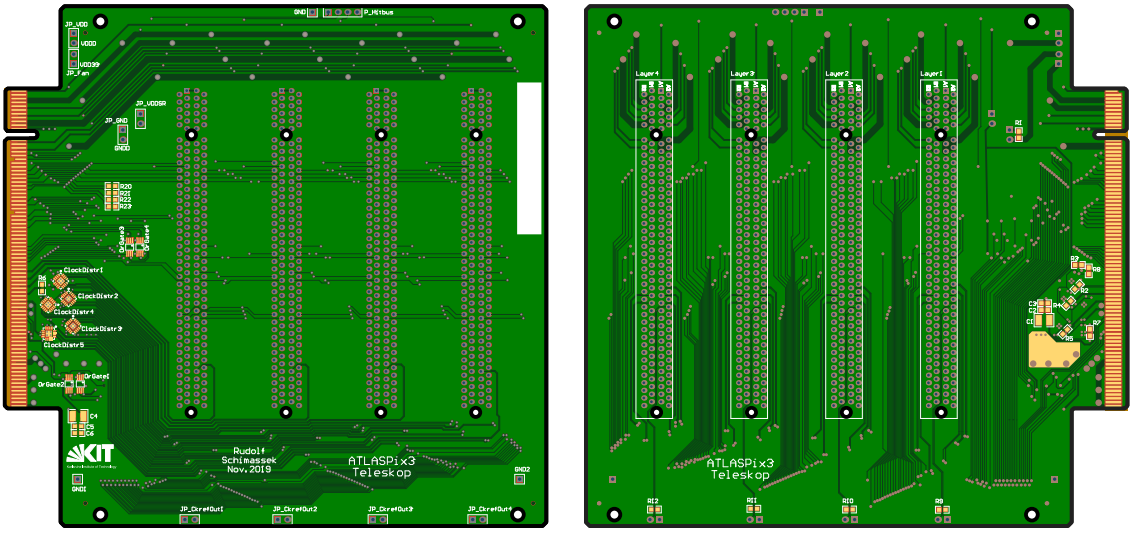


Figure 6.84: The ATLASPix3 beam telescope uses four single chip carrier boards connected to the telescope board to integrate in the GECCO readout system. The system is held in place by stabilisers between GECCO board and telescope board, as well as between the layers and the telescope board. This system enables simple adaption to different beam test environments as the setup has only one mounting point at the GECCO board and is self-supporting apart from that.



(a) top view (b) bottom view

Figure 6.85: The telescope base board is designed as a four-fold extension of the pins used on the PCIe connector for the single chip carrier. Therefore, four PCIe connectors for the four layers are placed on the board with a pitch of 2.5 cm. The LVDS signals from the FPGA are provided to the chips via fanout buffers in close to the input PCIe connector. For heat dissipation, the area next to these QFN packaged parts is filled with a copper plane connected to the base of the fanout buffers.

the fanout buffers, they are placed with a copper plane beneath to dissipate the heat. The layout of the telescope PCB is shown in figure 6.85.

Mechanically, the telescope uses the telescope PCB and stabilisers between the parts. The system with four layers, stabilisers and NexysVideo FPGA board is shown at the beginning of the section in figure 6.84. The placement of the PCIe connectors for the layers at a pitch of 2.5 cm defines the layer spacing of the setup. The other end of the carrier PCBs are held in place by spacing brackets that are fixed to the bottom of the telescope PCB. The telescope PCB itself is stabilised with diagonal stays to the far end of the GECCO board. With the screws fixing the FPGA board to the GECCO board, the whole system can be screwed to stages with the mounting holes on the GECCO board. This creates a compact, self-supporting system easy to set up.

The drawback of this simplicity is that adjustments to the position of the layers are not possible. Furthermore, the first iteration of the telescope PCB – as shown in the figures – does not allow for usage of the regulators as all voltages are connected between the layers. The powering scheme has been modified on second iteration to enable both direct powering and usage of the regulators individually per layer. On the third iteration, the fanout buffers were removed reducing the component count and requiring remapping of the signals on the connection to the FPGA.

From the pixel geometry of ATLASPix3 with (150×50) µm^2 and assuming uniform sensitivity across the pixel, the position resolution of the single layer would be expected to be the edge length divided by $\sqrt{12}$ which results in 43.3 µm and 14.4 µm for the long and short edges.

Firmware

Most parts of the firmware have undergone no changes in the transition from the single detector setup to the telescope. Only the pin mapping has been adapted for the signals of these parts.

For configuration via shift registers and SPI, a register with veto signals for the load signal and the chip-select bits has been added. The output from the FPGA stayed the same, on the telescope PCB, the signal is combined with the veto signals to obtain the signals for the individual chips.

The major change is the quadruplication of the modules for the digital readout of the sensors including multiplexing modules and additional registers for configuration of the additional modules.

The readout structure shares one clock generation module for all four inputs but includes a phase shifter per layer to compensate for different data line delays. The alignment, transmission layer decoding, data decoding and storing to a FIFO is the same as for the single detector firmware, just four times in parallel. The four FIFOs are multiplexed into one FIFO, from which the data are transmitted off the FPGA. Each data input of the multiplexer can be deactivated to select the layers of interest. Data is read from the active data inputs in an interleaved pattern. This means that after each data word, the channel is switched to equalise the amount od data read from the channels in high occupancy situations. This way, no layer can block the readout of the other layers. A layer identification is inserted in each dataset on bits that were unused in the single detector firmware. From these, the software will be able to demultiplex the data from the layers.

The module structure including the clock relations is shown in figure 6.86. A consequence from the operation with one clock generation and synchronisation signal for all four layers is that synchronisation is trivial: The timestamps of the layers are running synchronous by design.

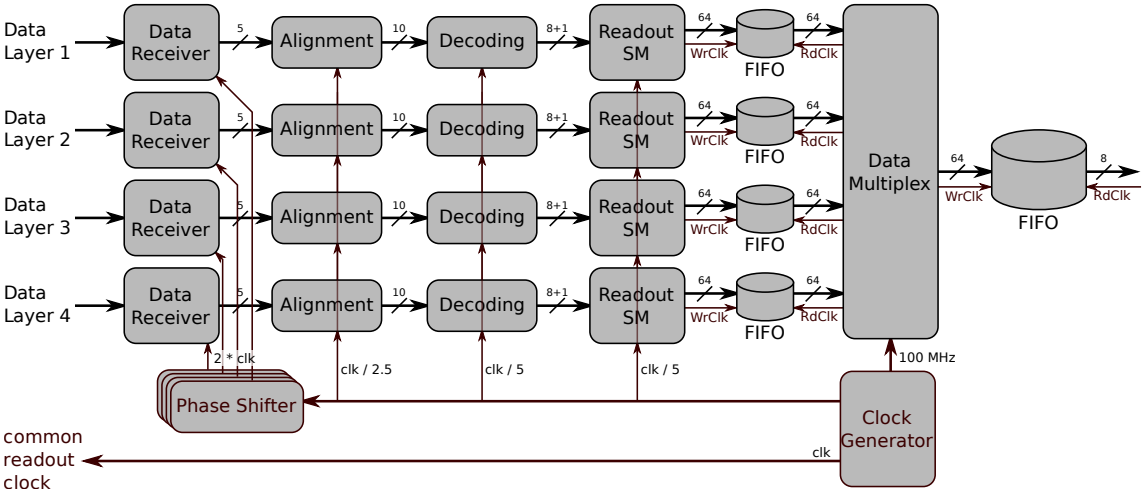


Figure 6.86: In order to obtain an efficient readout system for possibly unbalanced data rates, the digital readout of the telescope layers is extended with a multiplexer module. The sketch shows the situation for hit-driven readout. Each layer keeps a small FIFO for buffering. It is filled with the data after aligning to the data blocks, decoding and interpreting the data in the respective modules. The data widths in bits is given at the arrows. Data source identification is included in the 64 bit of data from the readout state machine. To save hardware complexity, the layers share the readout clock which means that phase adjustments have to be done on the receiver with a phase shifter per data stream. The relative clock speeds for the modules compared to the readout clock is given next to the clock distribution arrows

Software

The software changes are bigger compared to the changes in firmware. There, the configuration for each layer has to be managed individually. For the digital readout, the controls become more complex as each layer can be configured individually, too, and the interleaved data has to be demultiplexed and decoded.

The configuration is managed by quadruplication of the configuration containers and assigning one of each kind to one layer. The layer selection is added as new control and it manages the configuration of the veto signals introduced in firmware at the same time. The configuration storage files are kept separate for the layers. As they are mechanically single detector setups, this way, the calibration and tuning can be done on single sensor setups in parallel before placing the sensor PCBs in the telescope.

For the digital readout, the controls for the single detector readout path are quadrupled and completed by controls for the data multiplexer module added for the telescope.

To decode the interleaved data from the four layers, one decoder object per layer is created and only fed with the data for the respective layer. The signal data container is extended with the layer information so that all signal data can be processed the same way as before.

6.9 Telescope Measurements

The ATLASPix3 telescope has been used at a beam test at DESY [DDE+19, DSY21a] and at beam tests at HIT [OW08]. For the beam test at DESY, the goals were to build and commission a telescope from ATLASPix3 and to gain insights on the performance of it. At HIT, the telescope has been used for tracking of secondary particles generated by the

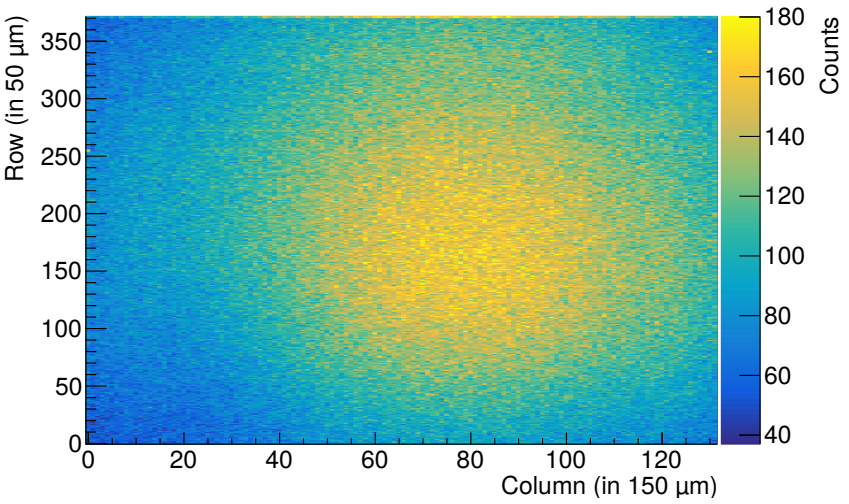


Figure 6.87: The profile of the electron beam is visible and smaller than the active area of ATLASPix3 enabling a complete profile with a single sensor.

carbon ion beam when penetrating a PMMA[3] cylinder in the context of the master's thesis of C. Klauda under the author's supervision [Kla22]. Here, only the results from the beam test at DESY are presented.

While all four layers are used for tracking at HIT, one layer is used as DUT at the beam test at DESY. Since tracking with three layers does not allow for two half tracks intersecting on the DUT, straight line fits are used for the tracks and the leverage is increased by using the first and the last two layers for tracking and the second layer as DUT. This choice is possible after the measurement as long as the settings of the layers are equivalent.

For the analysis, the Corryvreckan framework [DDH+21, WKH+20] is used. Corryvreckan is a modular 4D track reconstruction and analysis software intended for beam test measurements. It has been extended by the author with a module for loading and integrating the data taken with the ATLASPix3 telescope.

The beam time at DESY has been too short for both commissioning and large parameter scans. Therefore, a proof of concept is presented here. It includes general aspects extracted from single layer data as well as data combined from all telescope layers. As from the measurement the detector performance can be expected to be reduced, the analysis is not optimised for highest efficiency, either.

Single Layer Data

Measuring with a particle beam, the first point to check for is that the beam is passing the detectors and generates signals as expected. This does not require a full telescope, hence, data from a single layer can be used.

A hit map from the electron beam is shown in figure 6.87. The beam profile is smaller than the active area of ATLASPix3. But the tails of the profile extend over the edges of ATLASPix3. Consequently, the whole matrix can be probed without movement of the telescope.

In the data, signals forming clusters, i.e. groups of neighbouring pixels detecting signals at the same time, are found with frequencies as shown in figure 6.88: Even clusters with two pixels are more than an order of magnitude less probable than single pixel events.

[3] poly methyl methacrylate

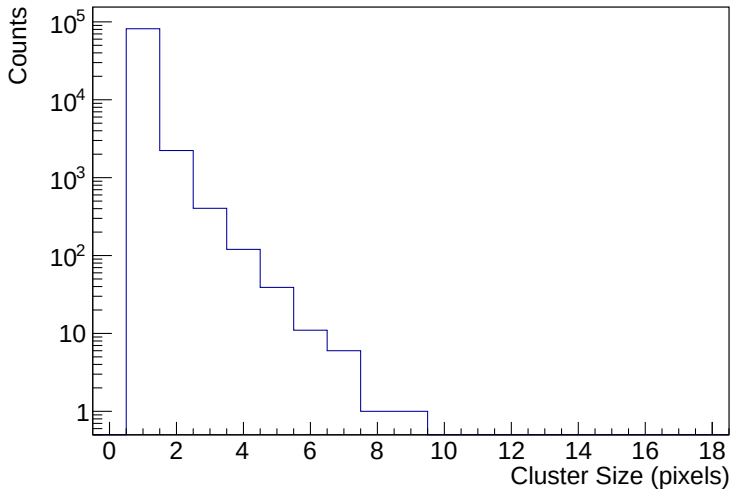


Figure 6.88: Clusters of pixel signals are detected on ATLASPix3 with decreasing frequency to larger clusters.

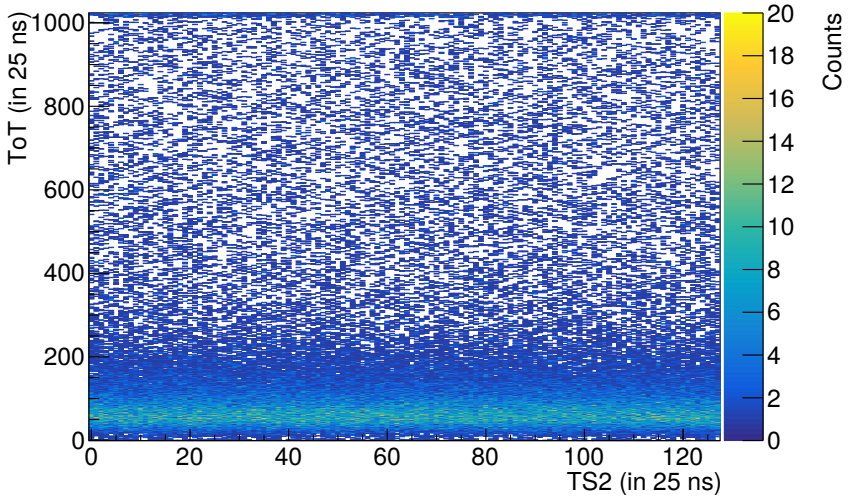
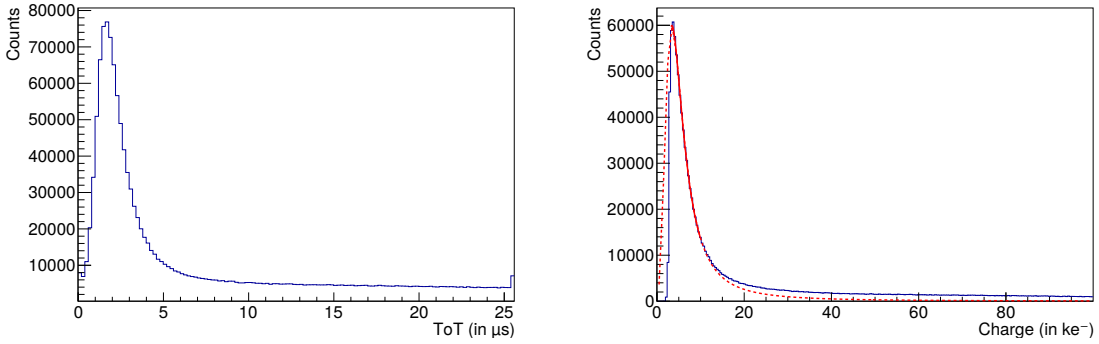


Figure 6.89: The ToT distribution along the vertical axis is independent from the value of the trailing edge timestamp and shows no artifacts. This hints towards a consistent configuration and faultless operation of the timestamps. An example of an error is given in appendix J.

This is consistent with measurements done on other monolithic HV-CMOS detectors as ATLASpix1 or MuPix8 [PPA+19]. The aspect ratio of 3 : 1 is not corrected for in this plot, it shows the combination of the contributions from the clusters extending in row direction (with a pitch of 50 µm) and the clusters extending in column direction (with a pitch of 150 µm).

In addition to the spatial distribution of the signals, also the arrival time and amplitude distribution of the signals is forming the ToT measurement of ATLASPix3. With the electrons passing the depletion layer of ATLASPix3, the charge distribution is expected to be a Landau distribution. With the electronics (as amplifiers with feedback and comparators), this distribution is convoluted with a Gaussian distribution, leading to a Landau-Gaussian distribution.

Since the timestamp of the telescope and the beam run independently, the distributions



(a) uncalibrated signal lengths

(b) cluster charge from calibration

Figure 6.90: The summed-up ToT values of the clusters form a distribution similar to a Landau-Gaussian distribution. Applying the calibration to convert the ToT value to charge, both the non-linearity and the pixel-to-pixel differences can be removed. The resulting distribution is cut off at the lower end by the detection threshold. The solid line represents the interval the Landau-Gaussian fit has been performed on, while the dashed line is the extension of this fit.

of both timestamps for the ToT measurement are supposed to be flat. To ensure correct operation of the timestamps, the ToT distribution can be correlated to one of the timestamps. This correlation is shown in figure 6.89: No artifacts are visible and only the tail of the distribution reaches the maximum value of the available range for the ToT measurement.

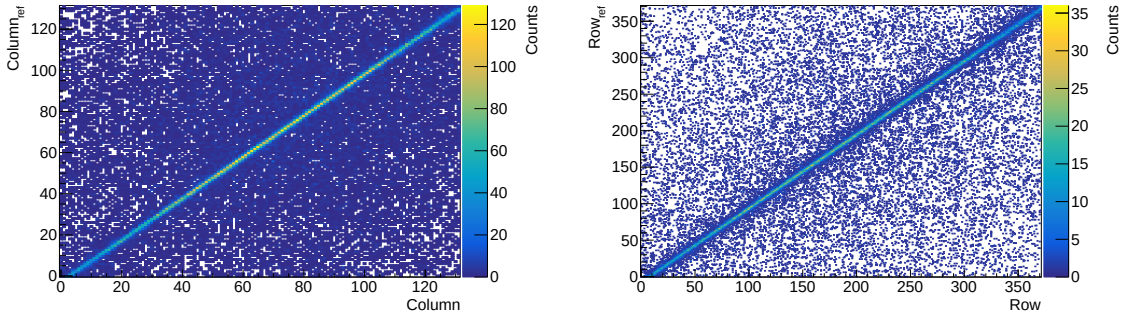
As it has been described in section 6.5.3, the signal lengths used for amplitude measurement will differ between the pixels on one sample. Therefore, it can be expected that the Gaussian fraction is enlarged without calibration. Furthermore, the ToT measurement is not linear for converting charge to time. The ToT distribution is not a real Landau-Gaussian distribution as a result.

With a calibration as performed in section 6.5.3, these effects can be removed. The remaining effects altering the distribution are then the detection threshold cutting off the lower end of the distribution and imperfections of the calibration. For one measurement, the calibration has been performed and is applied to the data in figure 6.90: The distribution is cut off for small signals. The tail for long signals is less pronounced for the calibrated signals, as it can be expected from pixels with small feedback currents. The tail in the calibrated distribution hints towards a contribution from particles with lower energy generating more charge than the primary beam energy electrons.

Multi-Layer Data

With all layers connected to the same FPGA and receiving the same timing signals via the fanout buffers, synchronisation of the data streams from the layers is trivial and it is integrated directly in firmware.

The simplest test for data consistency between the layers is the correlation between them. To do so, all signals with a time difference smaller than a set limit are evaluated against each other. For the ATLASPix3 telescope, all layers are oriented in the same direction leading to the expectation of a correlation line as angle bisector which is altered by the shifts and rotations of the sensors against each other. For the correlation between the first and the last layer in the telescope, the plots are shown in figure 6.91. The large lever between these layers amplifies shifts due non-vertical incident of the beam. Despite the large correlation window of 20 µs – which is 800 timestamps long, the correlation lines are



(a) Column correlation (b) Row correlation

Figure 6.91: The correlation plots between the first and last layer at a 20 µs time window shows clear lines. Shifts from the angle bisectors indicate towards shifts and tilts of the layers against each other.

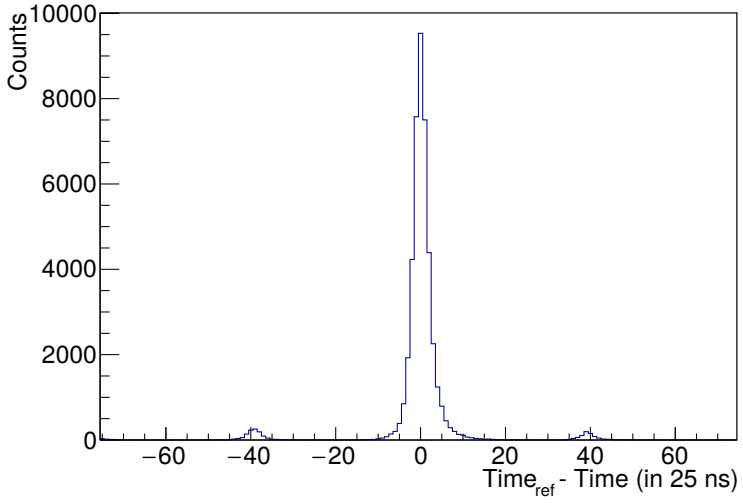


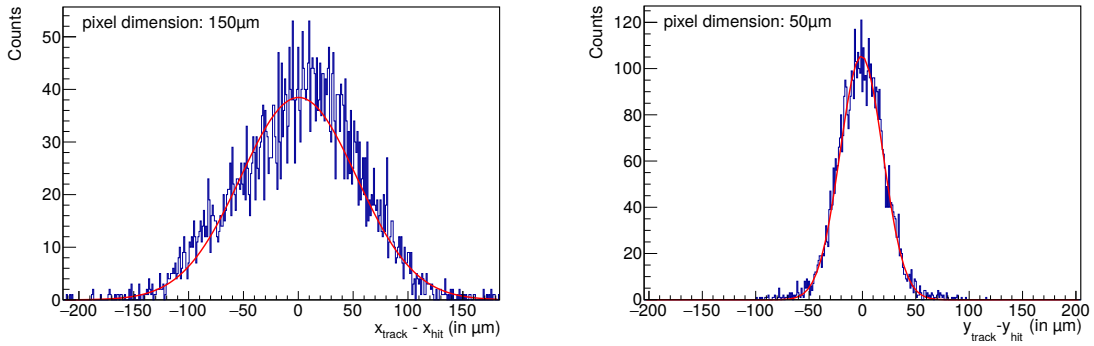
Figure 6.92: The time correlation plot shows a narrow accumulation around zero. The entries around −40 and 40 timestamps indicate towards beam structures, namely the DESY II revolution frequency of 1 MHz [DSY21b].

clearly visible indicating that the data is dominated by signals from particles traversing the telescope.

The time difference between the signals on the telescope layers also form a narrow distribution as visible in figure 6.92. The accumulations at −40 and 40 timestamps difference indicate to systematics from the beam. The difference matches the revolution frequency of the DESY II ring [DSY21b].

Starting from these correlation lines, the alignment of the telescope layers can be performed: Firstly, the correlation line shifts are taken to find a starting value for the shifts of the layers against each other. Then, the tracks through the detector are fitted and the residuals are looked at: The target is to have the deviation of the clusters from the tracks centered around zero. To achieve this, the layers are shifted and rotated and the track finding and fitting is repeated.

With the finished alignment, the deviation of the signals on the DUT layer with respect to the track defined from the other three layers can be investigated. From 10 000 tracks, the residual distributions for the DUT on the second layer, shown in figure 6.93, are obtained.



(a) residuals in column direction

(b) residuals in row direction

Figure 6.93: The residual distribution for column direction is about three times as wide as in row direction. In both directions three standard deviations approximate the pixel length.

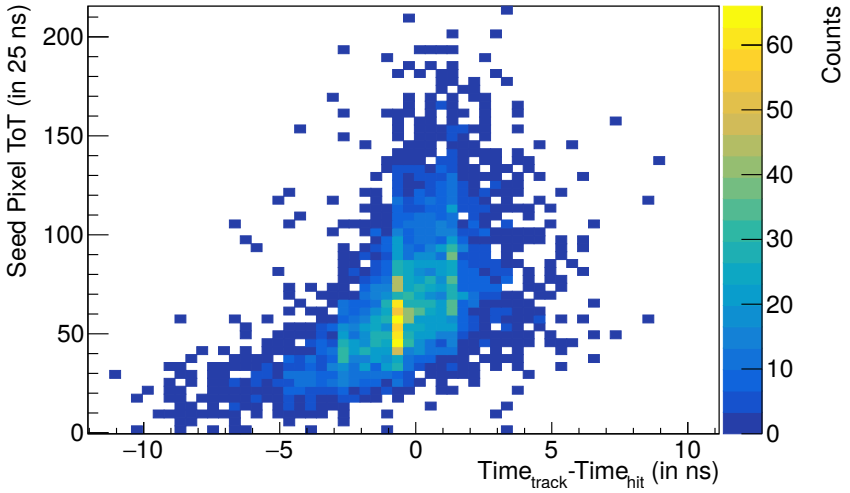


Figure 6.94: For small signals on the DUT, the signal is delayed with respect to charge generation. From the three clusters forming the track, the probability of one being large enough to not suffer from time-walk is larger than on the DUT. Consequently, for large ToT values on the DUT, no tail towards larger differences is visible.

In both directions, about three to four standard deviations of the Gaussian fit equal the length of the pixel edge which is 150 µm in column direction and 50 µm in row direction. The values of (37.3 ± 0.3) µm for the 150 µm edge and (16.1 ± 0.4) µm for the 50 µm edge are in the order of the values expected for the uniform distribution over the pixel edge lengths of 43.3 µm and 14.4 µm, respectively (see section 6.8).

One contribution to this uncertainty is time-walk that has not been corrected for. This becomes visible correlating the ToT measured for the DUT with the time difference to the track which is shown in figure 6.94. Together with the ToT calibration from section 6.5.3, the timing shift for small signals can be corrected for.

7 Conclusions

Over the last years, monolithic active high-voltage complementary metal-oxide-semiconductor (HV-CMOS) sensors have gotten mature when comparing the HVStripV1 sensor from 2014 [Sch16] or the H35Demo sensor from 2015 [Ehr21] with the most recent large scale designs ATLASPix3 [PBB+19] or MuPix10 [APSW20, Web21] from 2019: The size of the sensor grew to fill the whole mask size of the manufacturing process with a single pixel matrix They feature zero-suppressed readout at high data transmission rates at over one gigabit per second. But most importantly, ATLASPix3 and MuPix10 are not just prototypes for the sensor technology any more, but for detector system building [SAA+21, APSW20].

Current tracking detectors use hybrid pixel sensors, as in the upcoming upgrades of the detectors for High-Luminosity Large Hadron Collider (HL-LHC) [ATL17, Bac19]. Monolithic HV-CMOS sensors have been under consideration for a part of the inner tracker upgrade for A Toroidal LHC ApparatuS (ATLAS) [ATL17], but were not chosen. But with the advances of the HV-CMOS technology, future experiments or upgrades will consider this option for its advantages. As first experiment, the Mu3e experiment at PSI in Switzerland will use a tracker entirely made from monolithic HV-CMOS sensors [RAD+21]. More experiments are considering the technology for the next tracking detectors to be built, as Large Hadron Collider beauty (LHCb) with the Mighty Tracker [Par20] or for a detector at Circular Electron-Positron Collider (CEPC) [CZA+19].

The reason for this are on one hand the cost reduction compared to hybrid detectors as only one die instead of two is necessary and the assembly is simplified without a bump bonding process [Ehr21]. On the other hand, the material budget of a monolithic HV-CMOS sensor module can be lower than the one of a hybrid module: The HV-CMOS sensors for the Mu3e experiment are going to be thinned down to 50 µm [ABD+17].

The application, ATLASPix3 is designed for, is the ATLAS inner tracker (ATLAS ITk) upgrade for High-Luminosity Large Hadron Collider (HL-LHC). It requires triggered readout with data buffering on the sensor until arrival of the delayed trigger, and means being restricted on the data output bandwidth [ATL17]. For the optimisation of the architecture of the readout of ATLASPix3, the ReadOut Modelling Environment (ROME) simulation framework (chapter 5) has been developed as part of this thesis. It fills the gap between static rate calculations and in-depth hardware description simulations. Built as a framework for generic synchronously read out sensors, it can be used for optimisation of other designs, too, extending the simulation toolbox for the design of monolithic sensors

and readout application-specific integrated circuits (ASICs). ROME is planned to be used for the prototypes for the Mighty Tracker of LHCb.

The result of the ROME simulations for ATLASPix3 led to changes in the readout state machine with reduced data word size and size estimations for the buffers that together increased the maximum data throughput above the numbers required for the ITk expectations [ATL17] including a large safety margin of three.

The increased complexity of the ASICs raises the requirements for the readout setups, too. For this reason, building on the GEneric Configuration and COntrol (GECCO) setup hardware [Ehr21], software and firmware have been developed to thoroughly apply the reusability approach from hardware to software and firmware. The goal to obtain a framework that can be quickly adapted to new ASICs has been achieved by separating validation of code and configuration: A major part of this work is debugging and code validation. Hence, the speed improvement is achieved by providing configurable code that separates the configuration for the ASIC from the actual implementation of the functionality. This way, the code can be validated once and, for a new ASIC, only the configuration has to be validated. With the provided changes, the average setup time for a new ASIC has been cut down from weeks to days for the experienced user.

As a result, it has been used for several other ASICs bringing down the development time in various projects. The rebuilt system has been used for eight of the nine projects the GECCO hardware has been used for.

In addition to the usage at KIT ASIC and Detector Laboratory (KIT-ADL), the GECCO system has been exported to other universities in Germany, Great Britain, Italy and China (see section 4.1). At INFN Milan (IT), the GECCO system is used for the tests of the ATLASPix3 quad-module that has been developed there[1]. To provide all groups with the number of boards needed, another production has been done in 2021 resulting in a total of 65 GECCO boards in existence today. This large number showcases the advantageous properties of the system.

On top of the improvements, the GECCO system has been extended by Ethernet readout for measurements requiring high data throughput as for example beam tests. The transient current technique (TCT) setup developed as part of this thesis also integrates with the GECCO system. But the software for it can be run independently, too. It enables probing sensors on a sub-pixel scale which is otherwise only available on beam tests with heavy particles because scattering prevents sub-pixel resolution for light particles. It has been used for six different sensors: ATLASPix3 as part of this thesis (section 6.7.2), ATLASpix1_M2 [Ehr21], HitPix[2], HitIntegratingPix[3], H35-SPADV1 [Bla19] and the LFoundry LFATLASpix (also referred to as ALPHA) [SBC+18].

After production, the ATLASPix3 integrated sensor was measured with the GECCO readout system to probe the requirements from the design target – the ATLAS ITk upgrade.

- **Power consumption** is at 145 mW/cm^2 below the limit of 700 mW/cm^2 for direct powering and using the regulators at 215 mW/cm^2, too. Operation using the regulators on ATLASPix3.1 works at least as good as with direct powering (section 6.3).
- **Uniformity of the matrix** has been tested with signal delay measurements resulting in (10.1 ± 0.7) ns as delay spread over the rows of the matrix (section 6.5.1) and

[1] The adaption of the GECCO system to the quad-module has been done as part of [Rac21] under the co-supervision of the author.

[2] HitPix is an integrated counting pixel sensor for direct beam monitoring. It is presented in [Web21], the TCT measurement is to be published soon.

[3] HitIntegratingPix is an integrating pixel sensor for direct beam monitoring.

threshold tuning (section 6.4). The differential tuning digital-to-analogue converter (TDAC) of the pixels has been characterised and the mechanism enables threshold difference compensation to a standard deviation in the order of $70\,e^-$.

- **Sensitivity** to particles has been tested with source measurements and on beam tests leading to an most probable value (MPV) of $(5350 \pm 160)\,e^-$ for ^{90}Sr β-decay electrons at a signal-to-noise ratio (SNR) of 41.1 ± 1.2.
- The short **response time** has not been measured in a dedicated measurement. However, the information can be extracted from the other measurements: The digitisation is visible in the example measurements for the ^{90}Sr test with the amplifier output and the hitbus and the total delay can be estimated from the ROME simulations which indicate short delays, too, if not operated in overload conditions.
- **Radiation hardness** has been tested with a proton-irradiated sample at a fluence of $1 \cdot 10^{15}$ n_{eq}/cm^2 which is still operational (section 6.7.1).
- The **triggered readout** with sorting of signals by event works which has been used for the beam tests at Heidelberger Ionen Therapiezentrum (HIT) and especially in the high-rate readout tests (section 6.6.3).

In addition to those points, ATLASPix3 has been used to characterise the environment for a direct beam monitor for ion beams as used at HIT. This includes beam tests to measure the signals generated by the beams in HV-CMOS sensors (section 6.6.3) and inhomogeneous irradiation (section 6.7.2). The amount of radiation passing the sensor differs by orders of magnitude within the length scale of one sensor (2 cm) which is a large difference to the application in a tracking detector for collision products in the outer layers as it has been proposed for ATLAS ITk [ATL17]. The large differences found in those sensors between the pixels in the centre of the beam profile and at the tails of it require a larger range of adjustment for the TDACs at least at the step size used for homogeneously irradiated sensors. Consequently, the number of bits of the TDAC for dedicated beam monitor sensors has to be increased. The measurements (section 6.7.2) point to an increase from the current three bits to at least five bits. These results are going to be used for the development of the beam monitor sensors developed at KIT-ADL, as the HitPix prototype sensors.

The breakdown significantly below the expected value of 120 V has been investigated and in the geometry of the well structure in the periphery, a feature has been identified that can cause the breakdown behaviour. The HitPix integrated sensor, which was designed after ATLASPix3, the expected depletion voltages have been achieved (section 6.3.3 and [Web21]).

For improved timing on large signals, also the utilisation of the threshold tuning digital-to-analogue converters (DACs) has been investigated on ATLASPix3 leading to a standard deviation of the timing distribution of only 2.34 ns. This is an improvement by a factor of 3.9 compared to an untuned distribution. However, this timing optimisation does not reproduce the result of the tuning of the thresholds. Consequently, for large scale sensors to be used for timing with a resolution in the order of 1 ns, this effect has to be addressed in the design as well.

With a four layer telescope (section 6.8) and a quad-module prototype ([RGS+21] and appendix C), it has been shown that ATLASPix3 is suited for module building. The telescope has been used in a beam test at Deutsches Elektronen-Synchrotron (DESY) (section 6.9) and for tracking of secondary particles at beam tests at HIT as part of a master thesis [Kla22] under the author's supervision. The high-intensity beam at HIT has also been used to verify results from the ROME simulations (section 6.6.3).

Together with the latest prototype for the Mu3e experiment – MuPix10 – ATLASPix3 is likely the last development step before monolithic HV-CMOS sensors that are installed in

particle physics experiments and by that foster knowledge gain in particle physics. The work presented in this thesis contributed to the design of the sensor architecture of ATLASPix3 and tests leading to insights for more specialised sensors as for beam monitors or sensors with high time resolution, but proving the general fitness of the design for detetor system building. Furthermore, the GECCO system has been improved and extended decreasing the setup time for new sensor ASICs and providing structures for measurements at beam tests or with the TCT setup built. All in all, this work has contributed to the whole development loop of monolithic HV-CMOS sensors: With ROME, to the design of the sensor architecture, with the improvements and extension on the GECCO system, to the tests and their preparation and with the measurements on ATLASPix3, feedback for the next generation of sensors has been provided. Among these next-generation sensors are the MightyPix prototypes for the LHCb Mighty Tracker upgrade and the HitPix2 and HitPix3 for beam monitoring.

Appendix

A Collection of alternatives to GECCO

Here, several examples of existing test system projects are listed and compared to GECCO. It is meant to give a reasoning for the decision to not build on top of these systems.

A.1 Caribou Readout System

The Caribou readout system [LBC+19] is a collection of open source hardware, firmware and software for laboratory and high-rate beam tests that is maintained by developers at Brookhaven National Lab, CERN and DESY. The design aims for reduction of time until first data taking by minimising device integration effort.

It uses a system-on-chip (SoC) board as the Xilinx ZC706 to run the data acquisition and control software on the embedded processor while the interfacing to the DUT is performed with the FPGA. The hardware connection to the DUT is established via the Control and Readout (CaR) board which provides configurable features as 32 adjustable voltage references, eight adjustable current references, four injection pulsers and several analogue-to-digital converters (ADCs). The carrier board for the DUT is connected to the CaR board via a full-pin-count FPGA mezzanine-card (FMC) connector.

The software run on the embedded processor is a custom Yocto-based Linux distribution[1].

Compared to GECCO, the system components come at much higher cost – the SoC board costs about 2600 € (as of 11.11.2021). Furthermore, the CaR board comes with all components equipped increasing the cost of the board itself and the FMC connector for the DUT carrier increases the cost of the single carrier as well as making reflow-soldering for its assembly necessary.

An advantage of this system is the more efficient connection between software and firmware where the USB connection on the GECCO system restricts the data transfer rates. However, the usage of a custom operating system and a command-line interface software to control the system are significant drawbacks for managing many systems and for low-level debugging at the beginning of the testing cycle.

[1] `https://www.yoctoproject.org/`

A.2 Basil

Basil [SiL21] is a modular data acquisition system and system testing framework written in Python. For simulation, it makes use of the cocotb project [Ctb21] for the hardware. It is developed by the SiLab at Bonn University.

The project is cross-platform and supports various combinations of hardware.

At KIT-ADL, the integration of the hardware simulation is not necessary for chip testing and it is designed around Kintex 7 based hardware which comes at higher cost than the Artix 7 based NexysVideo board used for GECCO. As Caribou, basil does not have a GUI which turned out to greatly improve ASIC debugging efficiency.

A.3 YARR

The Yet Another Rapid Readout (YARR) system [Hei17] uses off-the shelf hardware in form of PCIe FPGA cards to ease the system development for developers not familiar with hardware development. With the installation of the FPGA card on the PCIe connector in the computer, a fast and efficient connection between firmware and software becomes possible. The firmware is intended to just be used as reconfigurable I/O interface.

The shift of complexity from firmware to software is a commonality with GECCO. However, the integration of the FPGA card into the computer housing has drawbacks in the applications, GECCO is intended for: Firstly, the test system has to be transported to different laboratories. There, being able to control the system via a laptop makes it easier to move the system. Secondly, the integration limits the access to the hardware which makes probing signals on the transition to the FPGA card difficult.

On the other hand, YARR is targeted towards detector systems which – as for the other readout systems – means that it is dimensioned larger than necessary for early stage tests which GECCO is built for. Consequently, the price for the supported FPGA boards is higher than for the NexysVideo used for GECCO.

B Configuration of ATLASPix3

All three configuration interfaces of ATLASPix3 are operational. For the application of ATLASPix3 in larger systems the achievable configuration speeds from the sensor ASIC itself and from the readout system are discussed here. Since both the SPI interface and the CMD internally access the shift register interface, the shift register interface is taken as reference for the comparison.

The SPI interface uses an extra shift register to demultiplex the signals for the shift registers increasing the amount of data to write. This means that as long as the SPI shift register can not be written 24 times as fast as the configuration shift registers of the sensor ASIC itself, the configuration speed will be decreased. The reason for this is that for changing one line of the configuration shift register interface at least 24 bits have to be written to the SPI shift register with extra cycles for the chip select line.

For the CMD, there are two options: The first option is to change the data lines for the configuration shift registers in a similar way as for the SPI interface. The second option is to use the on-chip state machine that receives blocks of 10 bit that are then written to the shift registers.

The first option transmits more data than the SPI interface as the number of bits to be sent for such a command is 48 bit for including the identification number for the receiving chip, the command type and the payload data as three 16 bit blocks.

The second option transmits the same 48 bit for transmitting 10 bit of configuration data. After a command for writing 10 bit of data, 50 blocks of 16 bit must not be such a writing command, or the data to be written to the configuration shift registers will be corrupted. This results in 848 bit transmitted for 10 bit in the configuration shift registers. This is more than for the SPI interface, but the CMD runs at a higher transmission frequency of 160 MHz from the design resulting in faster configuration.

In any case, both interfaces can not be faster than direct access to the configuration shift registers. However, their benefit is addressing through the chip select line for the SPI interface or through a chip identification number for the CMD. This reduces the number of lines necessary for the configuration of the sensor ASIC and therefore is a viable option for larger systems.

In contrary to the direct shift register access and the SPI interface, the CMD requires the readout clock to be provided to the sensor ASIC to work. In a system where a reference clock is provided, this also requires the PLL to start up before configuration can be done. To ensure this, a suitable configuration to start with has to be implemented in the design for the PLL to start up correctly. This becomes even more critical if clock-data recovery (CDR) is to be used where the PLL needs to start up from the command line signal with a dedicated phase detector.

The advantage from this is the reduction of the inputs for the sensors to only one differential input that can be shared between 16 ASICs whereas SPI requires 18 single-ended lines plus the reference clock as differential input for this. The direct shift register access requires up to 163 single-ended signals plus the reference clock to configure 16 sensors. Depending on the implementation, the number of signals can be reduced to 43 signals by splitting the clock signals instead of the ten load signals for the individual chips.

C ATLASPix3 Quad-Module

The quad-module developed at INFN Milan (IT) is a prototype of a demonstrator for a sensor module that could be used for ATLAS ITk. It comprises four ATLASPix3 integrated sensors arranged in a matrix of two by two dies. the pads are facing two opposing sides and the dies are glued to the bottom side of the flex circuit. A photograph of the assembled quad-module is shown in figure C.1.

Power is provided via a flex cable from the bottom side of the figure. The data lines are established via another flex cable (coming from the top of the figure) connected to the clamping connector in the middle of the module.

The powering scheme is an intermediate step, where the supply voltages VDD and $\mathsf{V_{gate}}$ are supplied directly and the voltages VSSA and $\mathsf{V_{minus}}$ are generated by the regulators of the ATLASPix3 integrated sensors. This is because of the oscillation encountered with the regulator for VDD on ATLASPix3.

Configuration of the sensors is possible using the SPI interface or the CMD. For SPI, the data and clock lines are shared between the sensors, only the data outputs of the sensors and the chip-select lines are separate. All differential inputs to the sensors are shared between them. The configuration via CMD uses addressing to select the sensor to write the configuration to.

The readout system used is the GECCO system. The two flex cables are connected to the PCIe connector using an adapter PCB. The firmware uses the triggered readout with the multiplexing structure from the telescope firmware. The software has been adapted accordingly from the single-sensor setup with elements from the software for the telescope.

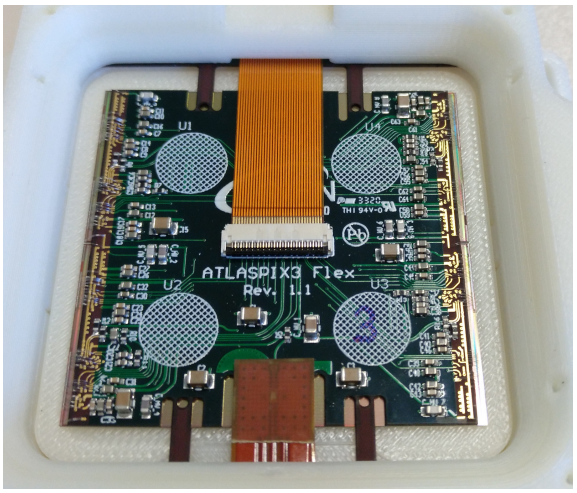


Figure C.1: The quad-module has four ATLASPix3 integrated sensors glued to its bottom side and bonded to the edges on the left and the right. Power is provided by an orange flex cable from the bottom and data connections are established via the other orange flex cable connecting to the clamping connector in the middle of the module.

The design of the flex circuit for the quad-module has been done by F. Sabatini in the group of Prof. A. Andreazza. The readout system has been adapted by B. Raciti for her master's thesis [Rac21] under the author's co-supervision.

D Threshold Tuning of a proton-irradiated ATLASPix3

In this section, additional plots for the threshold tuning on the proton-irradiated sensor in section 6.7.1 are included. This tuning procedure has been performed at $-10\,°C$.

The noise measured for the top end of the matrix is increased as visible in figure D.2

The detection threshold map (see figure D.3) does not show a similar increase to the top rows. Scattered over the whole matrix, pixels with higher detection threshold exist with an increase towards the lower right corner. But this is not reproduced as general trend of the majority of the pixels.

The noise map (see figure D.4) shows the same pattern of flashy pixels as the detection threshold map. In addition, an increase of the noise values to the top end and to both sides can be identified.

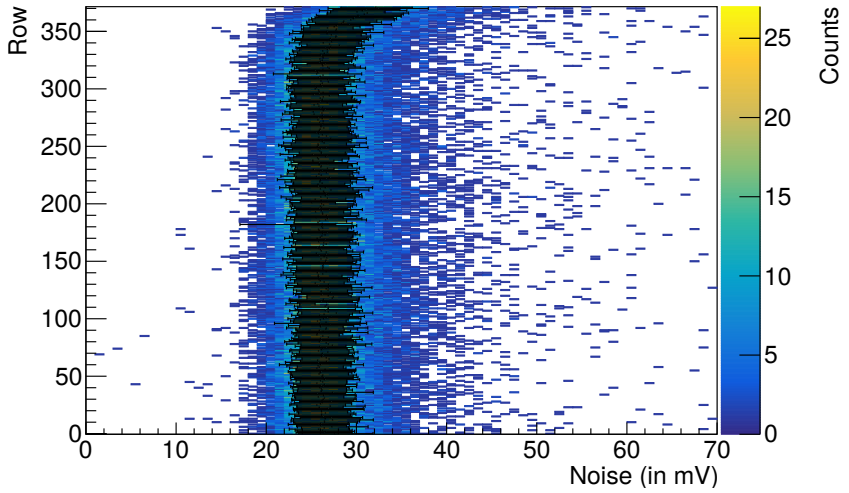


Figure D.2: The noise measured with the S-curves shows increased values at the top end of the matrix. Otherwise the distribution stays the same over the row index.

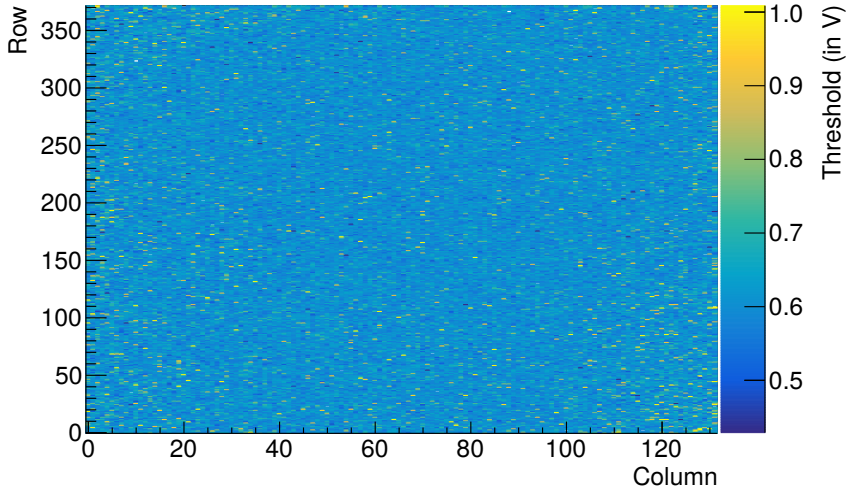


Figure D.3: The threshold map does not show significant gradients for large fractions of pixels. But pixels with higher detection threshold are scattered over the whole matrix.

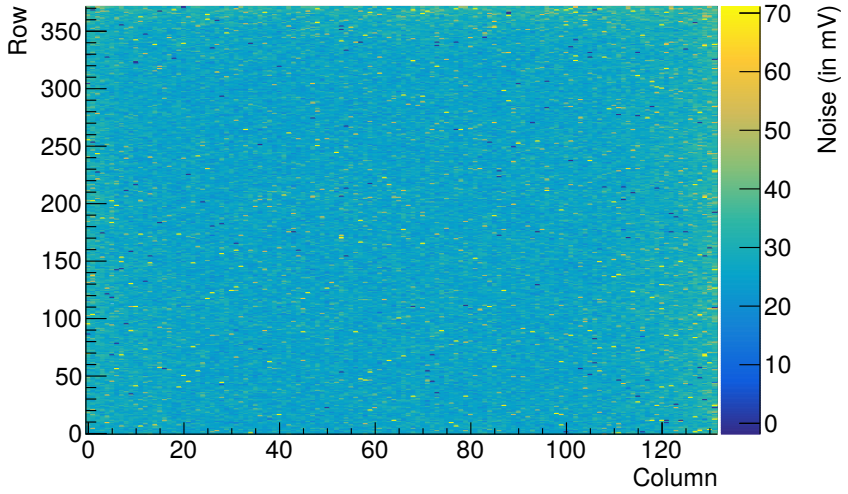


Figure D.4: The noise distribution over the matrix is flat in the middle apart from the same pixels that showed an increased detection threshold. But towards the top end and both sides, an increase can be identified.

E Matrix Timing on ATLASPix3

The values averaged for figure 6.35 in section 6.5.1 are shown in figure E.5. The step at row 248 can be clearly identified, but the smaller step at row 124 is less obvious.

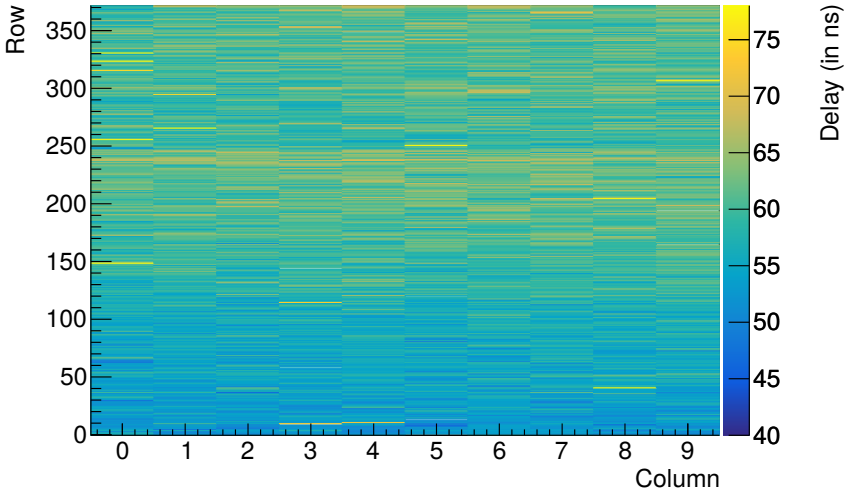


Figure E.5: The measured columns on the matrix of ATLASPix3 show a similar behaviour as expected from the design as identical elements. Also the steps between the routing layers are visible at rows 124 and 248.

F Amplifier Saturation on ATLASPix3

From the large amplification of the in-pixel amplifier of ATLASPix3, it is possible that charges saturating the amplifier's output are generated by traversing particles. In this case, the signal can not get any higher with increased charge generated. However, an increased charge still increases the input signal which results in an extended reset time of the amplifier. This means that the falling edge of the signal will include a transition from

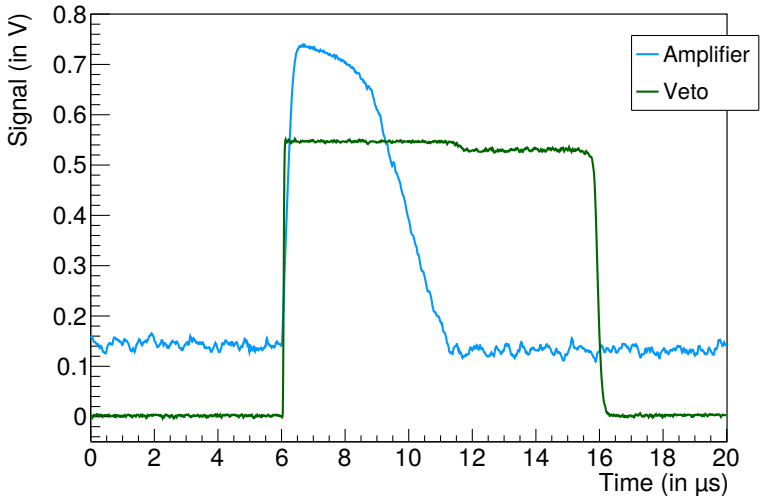


Figure F.6: The charge generated by a ^{90}Sr electron exceeded the amount saturating the amplifier output. This results in a kink in the slope of the falling edge at the point when the amplifier output leaves saturation. The veto signal is the logical OR of the comparator outputs of the surrounding pixels. As in figure 6.42, the hitbus signal is scaled down by 50 % and the amplifier output level is shifted for better visibility.

a small slope to a large slope at the point saturation ends. This is shown in figure F.6 for a signal from a ^{90}Sr electron.

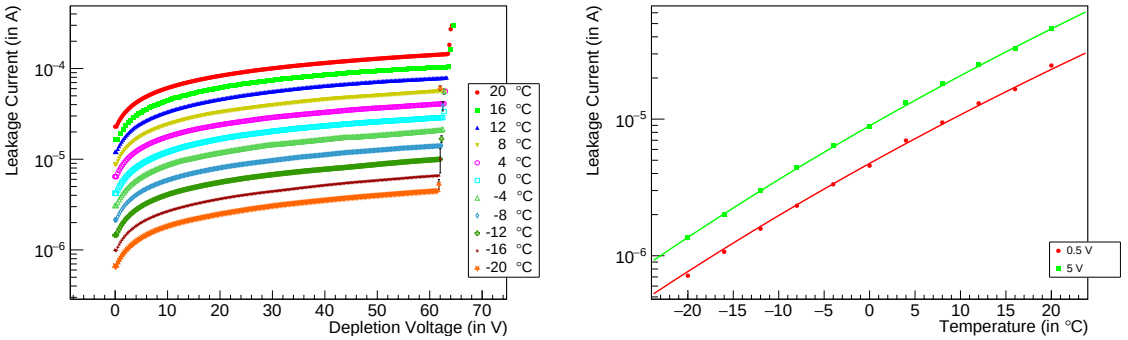
As visible from the correlation plot of signal height and length in figure 6.43 and justified from the description above, the signal length continues to grow with charge. Consequently, the charge measurement via ToT still generates usable data.

The signal detection occurs on the rising edge of the comparator output. This means that as long as the amplifier output stays above the comparator reference level, the pixel is insensitive for new signals. The saturation happens on the amplifier output not on the sensor diode, hence, the insensitive time enlarges with the amount of charge generated, but not more significantly more than for signals below the saturation limit.

G Further Leakage Current Measurements on ATLASPix3

The leakage current on AP3.H5 with about 150 % of the annual proton dose expected for HIT is almost an order of magnitude larger than the one on AP3.H1 and a factor of about 50 above a non-irradiated sensor. The temperature characteristics of the sensor IV-curve is shown in figure G.7: The leakage current is decreased matching the proportionality for the generation current (see section 6.3.3.

Furthermore, the temperature dependence of the breakdown voltage becomes visible in the plot. The breakdown happens at the same range as for the non-irradiated sensor and changes to lower voltages for lower temperatures. This indicates to an avalanche breakdown. Hence, the behaviour did not change compared to the non-irradiated sensor.



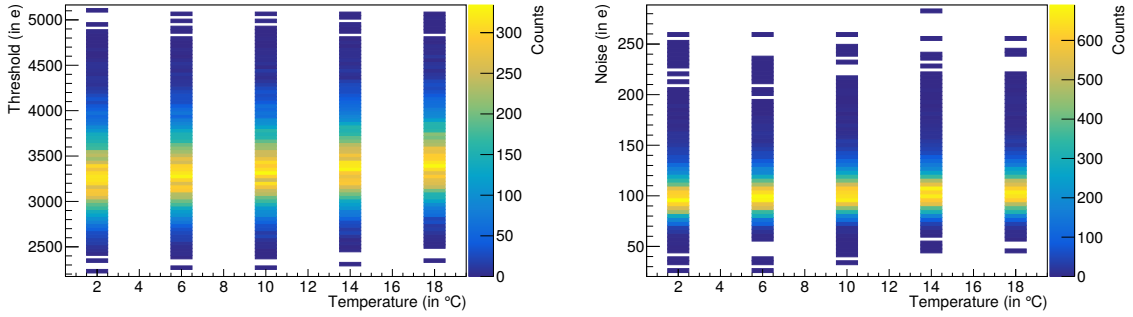
(a) Sensor diode leakage current

(b) Temperature dispersion of the leakage current

Figure G.7: The leakage current on AP3.H5 is almost an order of magnitude larger than on AP3.H1. The leakage current still drops with decreasing temperature according to the assumption of the current coming from charge generation as visible in (b). Furthermore, the shift of the breakdown voltage to smaller values at lower temperatures can be seen as well in (a).

H Threshold Dispersion on a non-irradiated ATLASPix3

For a non-irradiated sample of ATLASPix3, the threshold distribution shows a dependency on temperature just as the irradiated samples: For lower temperatures, the threshold gets lower. This change translates in a charge equivalent of about 150 e^- over the 16 K measured. The reduction of the noise average is about 10 e^- for the same temperature range.



(a) Threshold dispersion

(b) Noise dispersion

Figure H.8: The threshold measured with S-curves on charge injections decreases by 100 e^- over the 16 K temperature decrease shown. At the same time, the noise estimation from the same S-curve measurement decreases by 5 e^-.

I ATLASPix3.1 Leakage Current

The leakage current characteristics of ATLASPix3.1 matches the one of ATLASPix3. The IV curve of ATLASPix3.1 at room temperature is shown in figure I.9. The breakdown voltage is at about the same voltage. The larger absolute values of the points compared to the initial version can be explained by ambient light: Below the cover of the sensor, gaps exist through which light can enter. For the low leakage current of ATLASPix3, this can more than double the leakage current.

These points support the hypothesis that the changes made on ATLASPix3.1 did not solve the early breakdown problem of ATLASPix3.

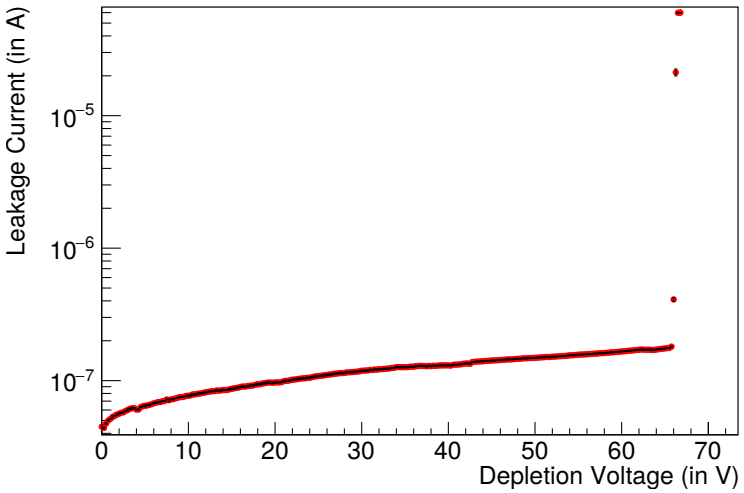


Figure I.9: The leakage current over the depletion voltage applied shows the same behaviour as for ATLASPix3 with the breakdown at 63 V.

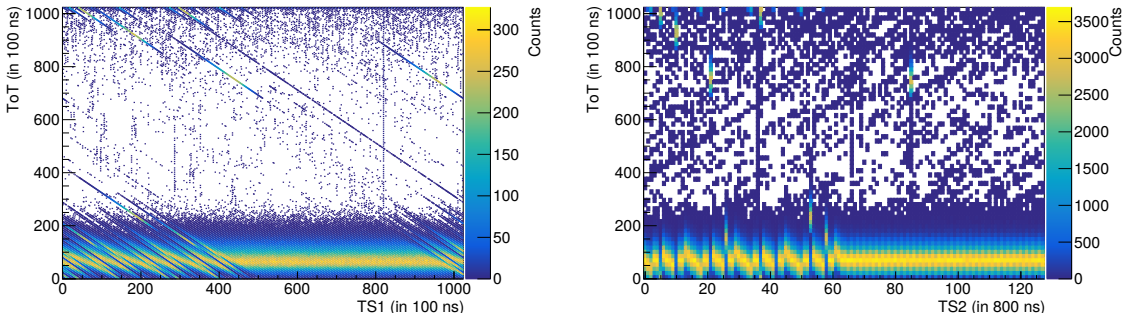
J Timestamp Error Effects

The ToT is measured on ATLASPix3 as two individual timestamps that can run at different speeds. These clock dividers are part of the sensor configuration. The value for ToT is the time difference between the two timestamps saved at the beginning and the end of an amplifier pulse.

Since the amplitude distribution of signals should not change with the signal start time, any correlation of the ToT with one of the timestamps should result in a distribution independent of the abscissa.

If one of the timestamps includes an error, this will be visible in the correlation of both timestamps with the ToT value. This is shown in figure J.10. If the error originates from the timestamp currently used as abscissa, the distribution will show a jumps of the data between the different values for the abscissa. On the other timestamp however, the error is incorporated via the difference needed for the ToT measurement. Consequently, the jumps will be at a 45 ° angle. In the example in the figure, the least significant bit of the trailing edge timestamp had an issue. Via the Gray code used for transmission, this led to the most significant bit containing the error.

If the clock dividers are set wrong for a dataset, this error will become visible as not the whole area of the correlation plots is occupied as the run length of the timestamp counter will not have the same length.



(a) leading edge timestamp

(b) trailing edge timestamp

Figure J.10: The effect of a wrong decoding problem materialises differently in the correlation of the two timestamps: If the error is in the timestamp on the X axis, the error pattern will be vertical. If the other timestamp has the problem, the pattern will be diagonal. In (a), the disruptions are diagonal hinting to a problem in the trailing edge timestamp. There, the error pattern is vertical and hints to a problem occuring for a cleared most significant bit.

Abbreviation Index

ADC analogue-to-digital converter

ALICE A Large Ion Collider Experiment

ASIC application-specific integrated circuit

ATLAS A Toroidal LHC ApparatuS

CAB content addressable buffer

CAM content-addressable memory

CDR clock-data recovery

CEPC Circular Electron-Positron Collider

CERN European Council for Nuclear Research – Conseil Européen pour la Recherche Nucléaire

CMOS complementary metal-oxide-semiconductor

CMS Compact Muon Solenoid

CSA charge-sensitive amplifier

DAC digital-to-analogue converter

DESY Deutsches Elektronen-Synchrotron

DUT device-under-test

EoC end-of-column

FCC Future Circular Collider

FCEPC future circular electron-positron collider

FIFO first-in-first-out buffer

FMC FPGA mezzanine-card

FPGA field-programmable gate-array

FWHM full width at half-maximum

GECCO GEneric Configuration and COntrol

HDL hardware description language

HEP high-energy physics

HIT Heidelberger Ionen Therapiezentrum

HL-LHC High-Luminosity Large Hadron Collider

HV-CMOS high-voltage complementary metal-oxide-semiconductor

HV-MAPS high-voltage monolithic active pixel sensor

ATLAS ITk ATLAS inner tracker

KIT-ADL KIT ASIC and Detector Laboratory

LEP Large Electron-Positron Collider

LHC Large Hadron Collider

LHCb Large Hadron Collider beauty

MAPS monolithic active pixel sensor

MIP minimum ionising particle

MOSFET metal-oxide-semiconductor field-effect transistor

NIEL non-ionising energy loss

PCB printed circuit board

PLL phase-lock loop

PPtB parallel pixel-to-buffer

ROME ReadOut Modelling Environment

SNR signal-to-noise ratio

SoC system-on-chip

SPI serial peripheral interface

TCT transient current technique

TDAC tuning DAC

ToT time-over-threshold

UDP user datagram protocol

VISA virtual instrument software architecture

Publications

Publications by or with contribution from the author of this thesis:

[1] R. Schimassek, A. Andreazza, H. Augustin, M. Barbero, M. Benoit, F. Ehrler, G. Iacobucci, A. Meneses, P. Pangaud, M. Prathapan, A. Schöning, E. Vilella, A. Weber, M. Weber, W. Wong, H. Zhang, and I. Perić, "Test results of ATLASPIX3 — A reticle size HVCMOS pixel sensor designed for construction of multi chip modules," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 986, p. 164812, 2021. [Online]. Available: https://doi.org/10.1016/j.nima.2020.164812

[2] R. Schimassek, R. Blanco, R. Casanova, F. Ehrler, I. Perić, E. Vilella, and H. Zhang, "Monolithic sensors in LFoundry technology: Concepts and measurements," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018. [Online]. Available: https://doi.org/10.1016/j.nima.2018.09.033

[3] R. Schimassek, F. Ehrler, and I. Perić, "HVCMOS Pixel Detectors - Methods for Enhancement of Time Resolution," in *Proceedings, 2016 IEEE Nuclear Science Symposium and Medical Imaging Conference: NSS/MIC 2016: Strasbourg, France*, 2016, p. 8069903. [Online]. Available: https://doi.org/10.1109/NSSMIC.2016.8069903

[4] I. Perić, A. Andreazza, H. Augustin, M. Barbero, M. Benoit, R. Casanova, F. Ehrler, G. Iacobucci, R. Leys, A. M. Gonzalez, P. Pangaud, M. Prathapan, R. Schimassek, A. Schöning, E. Vilella, A. Weber, M. Weber, W. Wong, and H. Zhang, "High-voltage cmos active pixel sensor," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 8, pp. 2488–2502, 2021. [Online]. Available: https://doi.org/10.1109/JSSC.2021.3061760

[5] K. Arndt, H. Augustin, P. Baesso, N. Berger, F. Berg, C. Betancourt, D. Bortoletto, A. Bravar, K. Briggl, D. vom Bruch, A. Buonaura, F. Cadoux, C. C. Barajas, H. Chen, K. Clark, P. Cooke, S. Corrodi, A. Damyanova, Y. Demets, S. Dittmeier, P. Eckert, F. Ehrler, D. Fahrni, S. Gagneur, L. Gerritzen, J. Goldstein, D. Gottschalk, C. Grab, R. Gredig, A. Groves, J. Hammerich, U. Hartenstein, U. Hartmann, H. Hayward, A. Herkert, G. Hesketh, S. Hetzel, M. Hildebrandt, Z. Hodge, A. Hofer, Q. Huang, S. Hughes, L. Huth, D. Immig, T. Jones, M. Jones, H.-C. Kästli, M. Köppel, P.-R. Kettle, M. Kiehn, S. Kilani, H. Klingenmeyer, A. Knecht, A. Knight, B. Kotlinski, A. Kozlinskiy, R. Leys, G. Lockwood, A. Loreti, D. La Marra, M. Müller, B. Meier, F. M. Aeschbacher, A. Meneses, K. Metodiev, A. Mtchedlishvili, S. Muley, Y. Munwes, L. Noehte, P. Owen, A. Papa, I. Paraskevas, I. Perić, A.-K. Perrevoort, R. Plackett, M. Pohl, S. Ritt, P. Robmann, N. Rompotis, T. Rudzki, G. Rutar, A. Schöning, R. Schimassek, H.-C. Schultz-Coulon, N. Serra, W. Shen, I. Shipsey, S. Shrestha, O. Steinkamp, A. Stoykov, U. Straumann, S. Streuli, K. Stumpf, N. Tata, J. Velthuis, L. Vigani, E. Vilella, J. Vossebeld, R. Wallny, A. Wasili, F. Wauters, A. Weber, D. Wiedner, B. Windelband, and T. Zhong, "Technical design of the phase I Mu3e

experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1014, p. 165679, 2021. [Online]. Available: https://doi.org/10.1016/j.nima.2021.165679

[6] I. Perić, M. Prathapan, H. Augustin, M. Benoit, R. Casanova, D. Dannheim, F. Ehrler, F. G. Messaoud, M. Kiehn, A. Nürnberg, R. Schimassek, M. V. Barreto, E. Vilella, A. Weber, W. Wong, and H. Zhang, "A high-voltage pixel sensor for the ATLAS upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 924, pp. 99–103, 2019, 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors. [Online]. Available: https://doi.org/10.1016/j.nima.2018.06.060

[7] H. Augustin, N. Berger, S. Dittmeier, F. Ehrler, C. Grzesik, J. Hammerich, A. Herkert, L. Huth, J. Kröger, F. M. Aeschbacher, I. Perić, M. Prathapan, R. Schimassek, A. Schöning, I. Sorokin, A. Weber, D. Wiedner, H. Zhang, and M. Zimmermann, "MuPix8 — Large area monolithic HVCMOS pixel detector for the Mu3e experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018. [Online]. Available: https://doi.org/10.1016/j.nima.2018.09.095

[8] F. Ehrler, M. Benoit, D. Dannheim, M. Kiehn, A. Nürnberg, I. Perić, M. Prathapan, R. Schimassek, T. Vanat, M. Vicente, A. Weber, and H. Zhang, "Characterization results of a hvcmos sensor for atlas," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018. [Online]. Available: https://doi.org/10.1016/j.nima.2018.08.069

[9] M. Kiehn, F. A. D. Bello, M. Benoit, R. Casanova, H. Chen, K. Chen, S. D.M.S., F. Ehrler, D. Ferrere, D. Frizell, S. G. Sevilla, G. Iacobucci, F. Lanni, H. Liu, C. Merlassino, J. Metcalfe, A. Miucci, I. Peric, M. Prathapan, R. Schimassek, M. V. Barreto, T. Weston, E. Vilella, M. Weber, A. Weber, W. Wong, W. Wu, E. Zaffaroni, H. Zhang, and M. Zhang, "Performance of cmos pixel sensor prototypes in ams h35 and ah18 technology for the atlas itk upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018. [Online]. Available: https://doi.org/10.1016/j.nima.2018.07.061

[10] M. Prathapan, M. Benoit, R. Casanova, D. Dannheim, F. Ehrler, M. Kiehn, A. Nürnberg, P. Pangaud, R. Schimassek, E. Vilella, A. Weber, W. Wong, H. Zhang, and I. Perić, "Towards the large area HVCMOS demonstrator for ATLAS ITk," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018. [Online]. Available: https://doi.org/10.1016/j.nima.2018.11.022

[11] H. Zhang, F. Ehrler, R. Schimassek, and I. Peric, "Measurement results on capacitively coupled particle detector with PHOTON readout chip," *Journal of Instrumentation*, vol. 15, no. 09, pp. P09 041–P09 041, sep 2020. [Online]. Available: https://doi.org/10.1088/1748-0221/15/09/p09041

[12] H. Augustin, N. Berger, C. Blattgerste, S. Dittmeier, F. Ehrler, C. Grzesik, J. Hammerich, A. Herkert, L. Huth, D. Immig, A. Kozlinskiy, M. Köppel, J. Kröger, F. Meier, A. Meneses, M. Müller, L. Noehte, I. Perić, M. Prathapan, T. Rudzki, R. Schimassek, A. Schöning, I. Sorokin, F. Stieler, A. Tyukin, T. Wagner, F. Wauters, A. Weber, D. Wiedner, H. Zhang, and M. Zimmermann, "Performance of the large scale HVCMOS pixel sensor MuPix8," *Journal of Instrumentation*, vol. 14, no. 10, pp. C10 011–

C10 011, oct 2019. [Online]. Available: https://doi.org/10.1088/1748-0221/14/10/c10011

[13] I. Perić, R. Blanco, R. Casanova, F. Ehrler, F. G. Messaoud, C. Krämer, R. Leys, M. Prathapan, R. Schimassek, A. Schöning, E. Vilella, A. Weber, and H. Zhang, "Status of HVCMOS developments for ATLAS," *Journal of Instrumentation*, vol. 12, no. 02, pp. C02 030–C02 030, feb 2017. [Online]. Available: https://doi.org/10.1088/1748-0221/12/02/c02030

[14] R. Blanco, H. Zhang, C. Krämer, F. Ehrler, R. Schimassek, R. Casanova, E. Vilella, F. G. Messaoud, R. Leys, M. Prathapan, A. Weber, and I. Perić, "HVCMOS monolithic sensors for the high luminosity upgrade of ATLAS experiment," *Journal of Instrumentation*, vol. 12, no. 04, pp. C04 001–C04 001, apr 2017. [Online]. Available: https://doi.org/10.1088/1748-0221/12/04/c04001

[15] F. Ehrler, I. Perić, and R. Schimassek, "HVCMOS pixel detectors first measurements on the reticle size prototype for the ATLAS pixel layers," in *Proceedings, 2016 IEEE Nuclear Science Symposium and Medical Imaging Conference: NSS/MIC 2016: Strasbourg, France*, 2016, p. 8069901. [Online]. Available: https://doi.org/10.1109/NSSMIC.2016.8069901

[16] A. Schöning, J. Anders, H. Augustin, M. Benoit, N. Berger, S. Dittmeier, F. Ehrler, A. Fehr, T. Golling, S. G. Sevilla, J. Hammerich, A. Herkert, L. Huth, G. Iacobucci, D. Immig, M. Kiehn, J. Kröger, F. Meier, A. Meneses, A. Miucci, L. O. S. Noehte, I. Peric, M. Prathapan, T. Rudzki, R. Schimassek, D. Sultan, L. Vigani, A. Weber, M. Weber, W. Wong, E. Zaffaroni, and H. Zhangh, "MuPix and ATLASPix – Architectures and Results," 2020, accessed: 07.11.2021. [Online]. Available: https://arxiv.org/abs/2002.07253

[17] M. Prathapan, R. Schimassek, M. Benoit, R. Casanova, F. Ehrler, A. Meneses, P. Pangaud, D. M. S. Sultan, E. Vilella, A. Weber, W. Wong, H. Zhang, and I. Perić, "Atlaspix3: A high voltage CMOS sensor chip designed for ATLAS Inner Tracker," in *Proceedings of Topical Workshop on Electronics for Particle Physics, TWEPP 2019, Santiago de Compostela, Spain, 2 - 6 September 2019*, ser. Proceedings of Science, vol. 370, 2019, p. Code 160323, 54.02.03; LK 01. [Online]. Available: https://doi.org/10.5445/IR/1000121410

[18] M. Prathapan, P. Barrillon, M. Benoit, R. Casanova, F. Ehrler, P. Pangaud, S. Pusti, R. Schimassek, E. Vilella, A. L. Weber, W. Wong, H. Zhang, and I. Perić, "Design of a HVCMOS pixel sensor ASIC with on-chip readout electronics for ATLAS ITk Upgrade," in *Proceedings of Topical Workshop on Electronics for Particle Physics — PoS(TWEPP2018)*, vol. 343, 2019, p. 074. [Online]. Available: https://doi.org/10.22323/1.343.0074

[19] B. Raciti, Y. Gao, R. Schimassek, J. Velthuis, H. Fox, A. Andreazza, H. Zhu, J. Martin, T. Jones, Y. Li, Y. Han, Z. Feng, H. Zhang, and I. Perić, "Characterisation of hv-maps atlaspix3 and its applications for future lepton colliders," *The 12th International Conference on Position Sensitive Detectors*, 2021, to be published in Journal of Instrumentation, accessed: 07.11.2021. [Online]. Available: https://indico.cern.ch/event/797047/contributions/4455962/

Reports

In addition to publications, reports providing documentation on the general-purpose projects have been written and put online for reference. These are listed for completeness, too.

[20] R. Schimassek, "Description of the GECCO Test System," Karlsruhe Institute of Technology (KIT), Karlsruhe, Tech. Rep., 2021, accessed: 07.11.2021. [Online]. Available: https://git.scc.kit.edu/adl_documents/gecco_documentation

[21] R. Schimassek, "Description of the ROME Simulation Framework," Karlsruhe Institute of Technology (KIT), Karlsruhe, Tech. Rep., 2021, accessed: 07.11.2021. [Online]. Available: https://git.scc.kit.edu/adl_documents/rome_description

Bibliography

[AAA+08] G. Aad, M. Ackers, F. A. Alberti *et al.*, "ATLAS pixel detector electronics and sensors," *Journal of Instrumentation*, vol. 3, no. 07, pp. P07 007–P07 007, jul 2008. [Online]. Available: https://doi.org/10.1088/1748-0221/3/07/p07007

[AAB+21] K. Arndt, H. Augustin, P. Baesso, N. Berger, F. Berg, C. Betancourt, D. Bortoletto, A. Bravar, K. Briggl, D. vom Bruch, A. Buonaura, F. Cadoux, C. C. Barajas, H. Chen, K. Clark, P. Cooke, S. Corrodi, A. Damyanova, Y. Demets, S. Dittmeier, P. Eckert, F. Ehrler, D. Fahrni, S. Gagneur, L. Gerritzen, J. Goldstein, D. Gottschalk, C. Grab, R. Gredig, A. Groves, J. Hammerich, U. Hartenstein, U. Hartmann, H. Hayward, A. Herkert, G. Hesketh, S. Hetzel, M. Hildebrandt, Z. Hodge, A. Hofer, Q. Huang, S. Hughes, L. Huth, D. Immig, T. Jones, M. Jones, H.-C. Kästli, M. Köppel, P.-R. Kettle, M. Kiehn, S. Kilani, H. Klingenmeyer, A. Knecht, A. Knight, B. Kotlinski, A. Kozlinskiy, R. Leys, G. Lockwood, A. Loreti, D. La Marra, M. Müller, B. Meier, F. M. Aeschbacher, A. Meneses, K. Metodiev, A. Mtchedlishvili, S. Muley, Y. Munwes, L. Noehte, P. Owen, A. Papa, I. Paraskevas, I. Perić, A.-K. Perrevoort, R. Plackett, M. Pohl, S. Ritt, P. Robmann, N. Rompotis, T. Rudzki, G. Rutar, A. Schöning, R. Schimassek, H.-C. Schultz-Coulon, N. Serra, W. Shen, I. Shipsey, S. Shrestha, O. Steinkamp, A. Stoykov, U. Straumann, S. Streuli, K. Stumpf, N. Tata, J. Velthuis, L. Vigani, E. Vilella, J. Vossebeld, R. Wallny, A. Wasili, F. Wauters, A. Weber, D. Wiedner, B. Windelband, and T. Zhong, "Technical design of the phase I Mu3e experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1014, p. 165679, 2021. [Online]. Available: https://doi.org/10.1016/j.nima.2021.165679

[ABB+19a] H. Augustin, N. Berger, C. Blattgerste, S. Dittmeier, F. Ehrler, C. Grzesik, J. Hammerich, A. Herkert, L. Huth, D. Immig, A. Kozlinskiy, M. Köppel, J. Kröger, F. Meier, A. Meneses, M. Müller, L. Noehte, I. Perić, M. Prathapan, T. Rudzki, R. Schimassek, A. Schöning, I. Sorokin, F. Stieler, A. Tyukin, T. Wagner, F. Wauters, A. Weber, D. Wiedner, H. Zhang, and M. Zimmermann, "Performance of the large scale HV-CMOS pixel sensor MuPix8," *Journal of Instrumentation*, vol. 14, no. 10, pp. C10 011–C10 011, oct 2019. [Online]. Available: https://doi.org/10.1088/1748-0221/14/10/c10011

[ABB+19b] H. Augustin, N. Berger, C. Blattgerste, S. Dittmeier, F. Ehrler, C. Grzesik, J. Hammerich, A. Herkert, L. Huth, D. Immig, A. Kozlinskiy, M. Köppel, J. Kröger, F. Meier, A. Meneses, M. Müller, L. Noehte, I. Perić, M. Prathapan, T. Rudzki, R. Schimassek, A. Schöning, I. Sorokin, F. Stieler, A. Tyukin, T. Wagner, F. Wauters, A. Weber, D. Wiedner, H. Zhang, and M. Zimmermann, "Performance of the large scale HV-CMOS pixel sensor MuPix8," *Journal of Instrumentation*, vol. 14, no. 10, pp. C10 011–C10 011, oct 2019. [Online]. Available: https://doi.org/10.1088/1748-0221/14/10/c10011

[ABD+17] H. Augustin, N. Berger, S. Dittmeier, C. Grzesik, J. Hammerich, Q. Huang, L. Huth, M. Kiehn, A. Kozlinskiy, F. Meier Aeschbacher, I. Perić, A.-K. Perrevoort, A. Schöning, S. Shrestha, D. vom Bruch, F. Wauters, and D. Wiedner, "The MuPix system-on-chip for the Mu3e experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 845, pp. 194–198, 2017, proceedings of the Vienna Conference on Instrumentation 2016. [Online]. Available: https://doi.org/10.1016/j.nima.2016.06.095

[ABD+18] H. Augustin, N. Berger, S. Dittmeier, F. Ehrler, C. Grzesik, J. Hammerich, A. Herkert, L. Huth, J. Kröger, F. M. Aeschbacher, I. Perić, M. Prathapan, R. Schimassek, A. Schöning, I. Sorokin, A. Weber, D. Wiedner, H. Zhang, and M. Zimmermann, "MuPix8 — Large area monolithic HVCMOS pixel detector for the Mu3e experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018. [Online]. Available: https://doi.org/10.1016/j.nima.2018.09.095

[ABF+20] I. B. Alonso, O. Brüning, P. Fessia, M. Lamont, L. Rossi, L. Tavian, and M. Zerlauth, *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN, 12 2020, vol. 10. [Online]. Available: https://doi.org/10.23731/CYRM-2020-0010

[ALI08] ALICE Collaboration, "The ALICE experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 002–S08 002, aug 2008. [Online]. Available: https://doi.org/10.1088/1748-0221/3/08/s08002

[All13] R. Allen, "The higgs bridge," *Physica Scripta*, vol. 89, 11 2013. [Online]. Available: https://doi.org/10.1088/0031-8949/89/01/018001

[APSW20] H. Augustin, I. Perić, A. Schöning, and A. Weber, "The MuPix sensor for the Mu3e experiment," *Nucl. Instrum. Meth. A*, vol. 979, p. 164441, 2020. [Online]. Available: https://doi.org/10.1016/j.nima.2020.164441

[ATL12] ATLAS Collaboration, "ATLAS Insertable B-Layer Technical Design Report Addendum," CERN, Tech. Rep., May 2012, addendum to CERN-LHCC-2010-013, ATLAS-TDR-019, accessed: 07.11.2021. [Online]. Available: https://cds.cern.ch/record/1451888

[ATL17] ATLAS Collaboration, "Technical Design Report for the ATLAS Inner Tracker Pixel Detector," CERN, Geneva, Tech. Rep., Sep 2017, accessed: 07.11.21. [Online]. Available: https://cds.cern.ch/record/2285585

[ATL19] ATLAS Collaboration, "Expected Tracking Performance of the ATLAS Inner Tracker at the HL-LHC," CERN, Geneva, Tech. Rep., Mar 2019, accessed: 07.11.21. [Online]. Available: https://cds.cern.ch/record/2669540

[ATL21] ATLAS Collaboration, "Trigger and Data Acquisition System," 2021, accessed: 07.11.21. [Online]. Available: https://atlas.cern/discover/detector/trigger-daq

[BAB+12] J. Beringer, J. Arguin, R. Barnett *et al.*, "Review of Particle Physics (RPP)," *Phys.Rev.*, vol. D86, p. 010001, 2012. [Online]. Available: https://doi.org/10.1103/PhysRevD.86.010001

[Bac19] M. Backhaus on behalf of the CMS Collaboration, "The upgrade of the CMS inner tracker for HL-LHC," *Journal of Instrumentation*, vol. 14, no. 10, pp. C10 017–C10 017, oct 2019. [Online]. Available: https://doi.org/10.1088/1748-0221/14/10/c10017

[Bad20] J. N. Bader, "Construction and Characterization of a Detector Module made of Monolithic HV-CMOS Senors," Master's thesis, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2020.

[BK04] W. Bragg and R. Kleeman, "On the ionization curves of radium," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 8, no. 48, pp. 726–738, 1904. [Online]. Available: https://doi.org/10.1080/14786440409463246

[Bla19] R. Blanco, "Customized integrated circuits for scientific and medial applications," Ph.D. dissertation, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2019. [Online]. Available: https://dx.doi.org/10.5445/IR/1000091922

[BR97] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," *Nucl.Instrum.Meth.*, vol. A389, pp. 81–86, 1997. [Online]. Available: https://doi.org/10.1016/S0168-9002(97)00048-X

[BRC+19] R. Brun, F. Rademakers, P. Canal, A. Naumann, O. Couet, L. Moneta, V. Vassilev, S. Linev, D. Piparo, G. Ganis, and et al., "root-project/root: v6.18/02," *Zenodo*, Aug 2019. [Online]. Available: https://doi.org/10.5281/zenodo.3895860

[BZK+17] R. Blanco, H. Zhang, C. Krämer, F. Ehrler, R. Schimassek, R. Casanova, E. Vilella, F. G. Messaoud, R. Leys, M. Prathapan, A. Weber, and I. Perić, "HVCMOS monolithic sensors for the high luminosity upgrade of ATLAS experiment," *Journal of Instrumentation*, vol. 12, no. 04, pp. C04 001–C04 001, apr 2017. [Online]. Available: https://doi.org/10.1088/1748-0221/12/04/c04001

[CC74] J. R. Chelikowsky and M. L. Cohen, "Electronic structure of silicon," *Phys. Rev. B*, vol. 10, pp. 5095–5107, Dec 1974. [Online]. Available: https://doi.org/10.1103/PhysRevB.10.5095

[CDE+10] M. Capeans, G. Darbo, K. Einsweiller, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet, "ATLAS Insertable B-Layer Technical Design Report," CERN, Tech. Rep., Sep 2010, accessed: 07.11.21. [Online]. Available: https://cds.cern.ch/record/1291633

[CEP18] CEPC Collaboration, "CEPC Input to the ESPP 2018," Institute of High Energy Physics, CAS, Tech. Rep., 2018, accessed: 07.11.2021. [Online]. Available: http://cepc.ihep.ac.cn/CEPC_Accelerator_Addendum/5304_CEPC_European_strategy_accelerator-v9Submit_version.pdf

[Chi13] A. Chilingarov, "Temperature dependence of the current generated in si bulk," *Journal of Instrumentation*, vol. 8, no. 10, pp. P10 003–P10 003, oct 2013. [Online]. Available: https://doi.org/10.1088/1748-0221/8/10/p10003

[CMS08] CMS Collaboration, "The CMS experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, p. S08004, aug 2008. [Online]. Available: https://doi.org/10.1088/1748-0221/3/08/s08004

[CMS18] CMS Collaboration, "How CMS weeds out particles that pile up," 2018, accessed: 07.11.2021. [Online]. Available: https://cms.cern/news/how-cms-weeds-out-particles-pile

[CMS21] "Silicon pixels," 2021, accessed: 07.11.21. [Online]. Available: https://cms.cern/detector/identifying-tracks/silicon-pixels

[Ctb21] "cocotb github repository," 2021, accessed: 11.11.2021. [Online]. Available: https://github.com/cocotb/cocotb

[CZA+19] L. J. Chen, H. Zhu, X. C. Ai, M. Fu, R. Kiuchi, Y. Liu, Z. A. Liu, X. C. Lou, Y. P. Lu, Q. Ouyang, X. Shi, J. Tao, K. Wang, N. Wang, C. F. Yang, Y. Zhang, and Y. Zhou, "Characterization of the first prototype CMOS pixel sensor developed for the CEPC vertex detector," *Radiation Detection Technology and Methods*, vol. 3, p. 45, 2019. [Online]. Available: https://doi.org/10.1007/s41605-019-0124-0

[DDE+19] R. Diener, J. Dreyling-Eschweiler, H. Ehrlichmann, I. Gregor, U. Kötz, U. Krämer, N. Meyners, N. Potylitsina-Kube, A. Schütz, P. Schütze, and M. Stanitzki, "The DESY II test beam facility," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 922, pp. 265–286, 2019. [Online]. Available: https://doi.org/10.1016/j.nima.2018.11.133

[DDH+21] D. Dannheim, K. Dort, L. Huth, D. Hynds, I. Kremastiotis, J. Kröger, M. Munker, F. Pitters, P. Schütze, S. Spannagel, T. Vanat, and M. Williams, "Corryvreckan: a modular 4D track reconstruction and analysis software for test beam data," *Journal of Instrumentation*, vol. 16, no. 03, p. P03008, mar 2021. [Online]. Available: https://doi.org/10.1088/1748-0221/16/03/p03008

[Deb17] J. Debus, "BAMS Detectors Specification," HIT, Requirement Specification, 2017.

[DSY21a] "Test Beams at DESY," 2021, accessed: 07.11.21. [Online]. Available: https://particle-physics.desy.de/e252106/

[DSY21b] "DESY II Description & Status," 2021, accessed: 07.11.21. [Online]. Available: https://particle-physics.desy.de/e252106/e252106/e252334

[EB08] L. Evans and P. Bryant, "LHC machine," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 001–S08 001, aug 2008. [Online]. Available: https://doi.org/10.1088/1748-0221/3/08/s08001

[EBD+18] F. Ehrler, M. Benoit, D. Dannheim, M. Kiehn, A. Nürnberg, I. Perić, M. Prathapan, R. Schimassek, T. Vanat, M. Vicente, A. Weber, and H. Zhang, "Characterization results of a hvcmos sensor for atlas," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018. [Online]. Available: https://doi.org/10.1016/j.nima.2018.08.069

[Ebe13] R. Eber, "Investigations of new Sensor Designs and Development of an effective Radiation Damage Model for the Simulation of highly irradiated Silicon Particle Detectors," Ph.D. dissertation, Karlsruhe Institute of Technology (KIT), 2013, accessed: 07.11.2021. [Online]. Available: https://publish.etp.kit.edu/record/20930

[Ehr15] F. Ehrler, "Development of active CMOS sensors for particle physics experiments," Master's thesis, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2015.

[Ehr21] F. Ehrler, "Characterization of monolithic hv-cmos pixel sensors for particle physics experiments," Ph.D. dissertation, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2021. [Online]. Available: http://dx.doi.org/10.5445/IR/1000133748

[EPS16] F. Ehrler, I. Perić, and R. Schimassek, "HVCMOS pixel detectors first measurements on the reticle size prototype for the ATLAS pixel layers," in *Proceedings, 2016 IEEE Nuclear Science Symposium and Medical Imaging Conference:*

NSS/MIC 2016: Strasbourg, France, 2016, p. 8069901. [Online]. Available: https://doi.org/10.1109/NSSMIC.2016.8069901

[For19] A. Forencich, "Verilog ethernet components," Source Code Repository, 2019, accessed 07.11.2021. [Online]. Available: https://github.com/alexforencich/verilog-ethernet

[FP+15] S. Feigl, I. Perić *et al.*, "Radiation-hard Active Pixel Sensors for HL-LHC Detector Upgrades based on HV/HR-CMOS Technology," in *Proceedings of Technology and Instrumentation in Particle Physics 2014 — PoS(TIPP2014)*, vol. 213, 2015, p. 280. [Online]. Available: https://doi.org/10.22323/1.213.0280

[GDG+20] C. Grau, M. Durante, D. Georg, J. A. Langendijk, and D. C. Weber, "Particle therapy in europe," *Molecular Oncology*, vol. 14, no. 7, pp. 1492–1499, 2020. [Online]. Available: https://doi.org/10.1002/1878-0261.12677

[GHJ+13] K. Gwosch, B. Hartmann, J. Jakubek, C. Granja, P. Soukup, O. Jäkel, and M. Martišíková, "Non-invasive monitoring of therapeutic carbon ion beams in a homogeneous phantom by tracking of secondary ions," *Physics in Medicine and Biology*, vol. 58, no. 11, pp. 3755–3773, may 2013. [Online]. Available: https://doi.org/10.1088/0031-9155/58/11/3755

[GMH52] E. L. Goldwasser, F. E. Mills, and A. O. Hanson, "Ionization loss and straggling of fast electrons," *Phys. Rev.*, vol. 88, pp. 1137–1141, Dec 1952. [Online]. Available: https://doi.org/10.1103/PhysRev.88.1137

[Har09] F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*, ser. Springer tracts in modern physics ; 231 : Elementary particle physics. Berlin: Springer, 2009, accessed: 07.11.2021. [Online]. Available: http://link.springer.com/book/10.1007/b106762

[Hei17] T. Heim, "YARR - A PCIe based readout concept for current and future ATLAS Pixel modules," *J. Phys.: Conf. Ser.*, vol. 898, p. 032053. 8 p, 2017, accessed: 11.11.2021. [Online]. Available: https://cds.cern.ch/record/2297459

[Hig75] V. Highland, "Some practical remarks on multiple scattering," *Nuclear Instruments and Methods*, vol. 129, no. 2, pp. 497–499, 1975. [Online]. Available: https://doi.org/10.1016/0029-554X(75)90743-0

[HLL21] "The HL-LHC Project," 2021, accessed: 07.11.2021. [Online]. Available: https://hilumilhc.web.cern.ch/content/hl-lhc-project

[KBB+18] M. Kiehn, F. A. D. Bello, M. Benoit, R. Casanova, H. Chen, K. Chen, S. D.M.S., F. Ehrler, D. Ferrere, D. Frizell, S. G. Sevilla, G. Iacobucci, F. Lanni, H. Liu, C. Merlassino, J. Metcalfe, A. Miucci, I. Peric, M. Prathapan, R. Schimassek, M. V. Barreto, T. Weston, E. Vilella, M. Weber, A. Weber, W. Wong, W. Wu, E. Zaffaroni, H. Zhang, and M. Zhang, "Performance of cmos pixel sensor prototypes in ams h35 and ah18 technology for the atlas itk upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018. [Online]. Available: https://doi.org/10.1016/j.nima.2018.07.061

[Kla22] C. Klauda, "Suitability Studies with an HVCMOS Sensor for Continuous Use and Detection of Secondary Particle Tracks with a Beam Telescope during Ion Irradiation at a Medical Irradiation Facility," Master's thesis, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2022, to be submitted by January 2022.

[Kra15] G. Kramberger, "Advanced TCT Setups," *Proceedings of Science*, vol. Vertex2014, p. 032, 2015. [Online]. Available: https://doi.org/10.22323/1.227.0032

[Kru20] B.-M. Kruth, "Krebs in Deutschland für 2015/2016," *Robert-Koch-Insitut*, 2020, accessed: 07.11.21. [Online]. Available: https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebs_in_Deutschland/kid_2019/krebs_in_deutschland_2019.pdf?__blob=publicationFile

[KW16] H. Kolanoski and N. Wermes, *Teilchendetektoren*, 1st ed. Springer, 2016. [Online]. Available: https://doi.org/10.1007/978-3-662-45350-6

[Lan44] L. Landau, "On the energy loss of fast particles by ionization," *J. Phys. (USSR)*, vol. 8, pp. 201–205, 1944.

[Lan15] J. Lange, "Recent Progress on 3D Silicon Detectors," *Proceedings of Science*, 2015, accessed: 07.11.2021. [Online]. Available: https://arxiv.org/abs/1511.02080

[LBC+19] H. Liu, M. Benoit, H. Chen, K. Chen, F. Di Bello, G. Iacobucci, F. Lanni, M. Pinto, W. Wu, and L. Xu, "Development and application of a modular test system for the HV-CMOS pixel sensor R&D of the ATLAS HL-LHC upgrade," *Radiation Detection Technology and Methods*, vol. 3, p. 47, 2019. [Online]. Available: https://doi.org/10.1007/s41605-019-0126-y

[LD91] G. Lynch and O. Dahl, "Approximations to multiple coulomb scattering," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 58, no. 1, pp. 6–10, 1991. [Online]. Available: https://doi.org/10.1016/0168-583X(91)95671-Y

[LHC08] LHCb Collaboration, "The LHCb detector at the LHC," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 005–S08 005, aug 2008. [Online]. Available: https://doi.org/10.1088/1748-0221/3/08/s08005

[LHC19] LHCb Collaboration, "Mighty Tracker: Design Studies for the Downstream Silicon Tracker in Upgrade Ib and II," CERN, internal document LHCb-INT-2019-007, 2019, LHCb internal document.

[MAB+18] M. Mangano, P. Azzi, M. Benedikt, A. Blondel, D. A. Britzger, A. Dainese, M. Dam, J. de Blas, D. Enterria, O. Fischer, C. Grojean, J. Gutleber, C. Gwenlan, C. Helsens, P. Janot, M. Klein, U. Klein, M. P. Mccullough, S. Monteil, J. Poole, M. Ramsey-Musolf, C. Schwanenberger, M. Selvaggi, F. Zimmermann, and T. You, "FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1. Future Circular Collider," CERN, Geneva, Tech. Rep., Dec 2018, accessed: 07.11.2021. [Online]. Available: https://cds.cern.ch/record/2651294

[Mei03] J. Meikle, "How particles can be therapeutic," PhysicsWorld.com, 2003, accessed: 07.11.21. [Online]. Available: https://physicsworld.com/a/how-particles-can-be-therapeutic/

[Nex17] "Product data sheet 74lcv8t595: Dual supply 8-bit serial-in/serial-out or parallel-out shift register; 3-state," 2017, accessed: 07.11.21. [Online]. Available: https://assets.nexperia.com/documents/data-sheet/74HC_HCT595.pdf

[OW08] D. Ondreka and U. Weinrich, "The Heidelberg Ion Therapy (HIT) Accelerator Coming into Operation," *EPAC08 Proceedings*, 2008, accessed: 07.11.2021. [Online]. Available: https://accelconf.web.cern.ch/e08/papers/tuocg01.pdf

[PAA+21] I. Perić, A. Andreazza, H. Augustin, M. Barbero, M. Benoit, R. Casanova, F. Ehrler, G. Iacobucci, R. Leys, A. M. Gonzalez, P. Pangaud, M. Prathapan, R. Schimassek, A. Schöning, E. Vilella, A. Weber, M. Weber, W. Wong, and H. Zhang, "High-voltage cmos active pixel sensor," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 8, pp. 2488–2502, 2021. [Online]. Available: https://doi.org/10.1109/JSSC.2021.3061760

[Par20] C. Parkes, "Towards a mighty tracker," 2020, CERN, accessed: 07.11.2021. [Online]. Available: https://indico.cern.ch/event/897697/contributions/3786517/

[PBB+19] M. Prathapan, P. Barrillon, M. Benoit, R. Casanova, F. Ehrler, P. Pangaud, S. Pusti, R. Schimassek, E. Vilella, A. L. Weber, W. Wong, H. Zhang, and I. Perić, "Design of a HVCMOS pixel sensor ASIC with on-chip readout electronics for ATLAS ITk Upgrade," in *Proceedings of Topical Workshop on Electronics for Particle Physics — PoS(TWEPP2018)*, vol. 343, 2019, p. 074. [Online]. Available: https://doi.org/10.22323/1.343.0074

[PBC+17] I. Perić, R. Blanco, R. Casanova, F. Ehrler, F. G. Messaoud, C. Krämer, R. Leys, M. Prathapan, R. Schimassek, A. Schöning, E. Vilella, A. Weber, and H. Zhang, "Status of HVCMOS developments for ATLAS," *Journal of Instrumentation*, vol. 12, no. 02, pp. C02 030–C02 030, feb 2017. [Online]. Available: https://doi.org/10.1088/1748-0221/12/02/c02030

[PBC+18] M. Prathapan, M. Benoit, R. Casanova, D. Dannheim, F. Ehrler, M. Kiehn, A. Nürnberg, P. Pangaud, R. Schimassek, E. Vilella, A. Weber, W. Wong, H. Zhang, and I. Perić, "Towards the large area HVCMOS demonstrator for ATLAS ITk," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018. [Online]. Available: https://doi.org/10.1016/j.nima.2018.11.022

[PDG21] "X-ray Mass Attenuation Coefficients," 2021, accessed: 07.11.21. [Online]. Available: https://physics.nist.gov/PhysRefData/XrayMassCoef/tab1.html

[Peq15] J. Pequenao, "Computer generated image of the whole ATLAS detector," 2015, accessed: 07.11.2021. [Online]. Available: http://cds.cern.ch/record/1095924

[Per04] I. Perić, "Design and realisation of integrated circuits for the readout of pixel sensors in high-energy physics and biomedical imaging," Ph.D. dissertation, Universität Bonn, Aug 2004.

[Per07] I. Perić, "A novel monolithic pixel detector implemented in high-voltage CMOS technology," *IEEE Nuclear Science Symposium Conference Record*, vol. 2, pp. 1033–1039, 2007. [Online]. Available: https://doi.org/10.1109/NSSMIC.2007.4437188

[Per14] H. Pernegger, "IBL Installation into the ATLAS Experiment," 2014, accessed: 07.11.21. [Online]. Available: https://cds.cern.ch/record/2138466

[Per19] I. Perić, *ATLASPIX3 user manual*, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2019, accessed: 07.11.2021. [Online]. Available: https://adl.ipe.kit.edu/downloads/ATLASPIX3_um_v1.pdf

[Pet20] A. Peters, "Medical applications – instrumentation and diagnostics," 2020, accessed: 07.11.2021. [Online]. Available: https://arxiv.org/abs/2005.08729

[Pos80] J. Postel, *User Datagram Protocol*, ISI, August 1980, RFC 768, accessed: 07.11.2021. [Online]. Available: https://tools.ietf.org/pdf/rfc768.pdf

[PPA+19] I. Perić, M. Prathapan, H. Augustin, M. Benoit, R. Casanova, D. Dannheim, F. Ehrler, F. G. Messaoud, M. Kiehn, A. Nürnberg, R. Schimassek, M. V. Barreto, E. Vilella, A. Weber, W. Wong, and H. Zhang, "A high-voltage pixel sensor for the ATLAS upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 924, pp. 99–103, 2019, 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors. [Online]. Available: https://doi.org/10.1016/j.nima.2018.06.060

[Pra20] M. Prathapan, "High voltage and nanoscale cmos integrated circuits for particle physics and quantum computing," Ph.D. dissertation, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2020. [Online]. Available: http://dx.doi.org/10.5445/IR/1000105789

[PSB+19] M. Prathapan, R. Schimassek, M. Benoit, R. Casanova, F. Ehrler, A. Meneses, P. Pangaud, D. M. S. Sultan, E. Vilella, A. Weber, W. Wong, H. Zhang, and I. Perić, "Atlaspix3: A high voltage CMOS sensor chip designed for ATLAS Inner Tracker," in *Proceedings of Topical Workshop on Electronics for Particle Physics, TWEPP 2019, Santiago de Compostela, Spain, 2 - 6 September 2019*, ser. Proceedings of Science, vol. 370, 2019, p. Code 160323, 54.02.03; LK 01. [Online]. Available: https://doi.org/10.5445/IR/1000121410

[Rac21] B. Raciti, "Analysis, characterisation and performance studies of the ATLASPix3 ASIC," Master's thesis, INFN Milano (IT), 2021.

[RAD+21] T. Rudzki, H. Augustin, M. Deflorin, S. Dittmeier, F. Frauen, D. M. Immig, D. Kim, F. M. Aeschbacher, A. Meneses, M. Menzel, I. Perić, S. Preuß, A. Schöning, L. Vigani, A. Weber, and B. Weinläder, "The Mu3e experiment: Toward the construction of an HV-MAPS vertex detector," 2021, accessed: 07.11.2021. [Online]. Available: https://arxiv.org/abs/2106.03534

[RGS+21] B. Raciti, Y. Gao, R. Schimassek, J. Velthuis, H. Fox, A. Andreazza, H. Zhu, J. Martin, T. Jones, Y. Li, Y. Han, Z. Feng, H. Zhang, and I. Perić, "Characterisation of hv-maps atlaspix3 and its applications for future lepton colliders," *The 12th International Conference on Position Sensitive Detectors*, 2021, to be published in Journal of Instrumentation, accessed: 07.11.2021. [Online]. Available: https://indico.cern.ch/event/797047/contributions/4455962/

[Rö18] C. Röck, "Architecture simulation of pixelsensors," Bachelor's thesis, Karlsruhe Institute of Technology (KIT), 2018.

[SAA+20] A. Schöning, J. Anders, H. Augustin, M. Benoit, N. Berger, S. Dittmeier, F. Ehrler, A. Fehr, T. Golling, S. G. Sevilla, J. Hammerich, A. Herkert, L. Huth, G. Iacobucci, D. Immig, M. Kiehn, J. Kröger, F. Meier, A. Meneses, A. Miucci, L. O. S. Noehte, I. Peric, M. Prathapan, T. Rudzki, R. Schimassek, D. Sultan, L. Vigani, A. Weber, M. Weber, W. Wong, E. Zaffaroni, and H. Zhangh, "MuPix and ATLASPix – Architectures and Results," 2020, accessed: 07.11.2021. [Online]. Available: https://arxiv.org/abs/2002.07253

[SAA+21] R. Schimassek, A. Andreazza, H. Augustin, M. Barbero, M. Benoit, F. Ehrler, G. Iacobucci, A. Meneses, P. Pangaud, M. Prathapan, A. Schöning, E. Vilella, A. Weber, M. Weber, W. Wong, H. Zhang, and I. Perić, "Test results of ATLASPIX3 — A reticle size HVCMOS pixel sensor designed for construction of multi chip modules," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 986, p. 164812, 2021. [Online]. Available: https://doi.org/10.1016/j.nima.2020.164812

[Sau09] R. Sauer, *Halbleiterphysik*. Oldenbourg, 2009. [Online]. Available: https://doi.org/10.1524/9783486598506

[SBC+18] R. Schimassek, R. Blanco, R. Casanova, F. Ehrler, I. Perić, E. Vilella, and H. Zhang, "Monolithic sensors in LFoundry technology: Concepts and measurements," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018. [Online]. Available: https://doi.org/10.1016/j.nima.2018.09.033

[SBC+19] R. Schimassek, R. Blanco, R. Casanova, F. Ehrler, I. Perić, E. Vilella, and H. Zhang, "Monolithic sensors in LFoundry technology: Concepts and measurements," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 679–680, 2019, frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors. [Online]. Available: https://doi.org/10.1016/j.nima.2018.09.033

[SC10] W. Setyawan and S. Curtarolo, "High-throughput electronic band structure calculations: Challenges and tools," *Computational Materials Science*, vol. 49, no. 2, pp. 299–312, 2010. [Online]. Available: https://doi.org/10.1016/j.commatsci.2010.05.010

[Sch14] R. Schimassek, "Funcionality tests of digital pixel modules for the cms upgrade," Bachelor's thesis, Karlsruhe Institute of Technology (KIT), 2014, accessed: 07.11.2021. [Online]. Available: https://publish.etp.kit.edu/record/21008

[Sch16] R. Schimassek, "Evaluation of CMOS Sensors as Particle Detectors at HL-LHC," Master's thesis, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2016, accessed: 07.11.2021. [Online]. Available: https://publish.etp.kit.edu/record/21336

[Sch21a] R. Schimassek, "Description of the GECCO Test System," Karlsruhe Institute of Technology (KIT), Karlsruhe, Tech. Rep., 2021, accessed: 07.11.2021. [Online]. Available: https://git.scc.kit.edu/adl_documents/gecco_documentation

[Sch21b] R. Schimassek, "Description of the ROME Simulation Framework," Karlsruhe Institute of Technology (KIT), Karlsruhe, Tech. Rep., 2021, accessed: 07.11.2021. [Online]. Available: https://git.scc.kit.edu/adl_documents/rome_description

[Sch21c] R. Schimassek, "ROME Code Repository," 2021, accessed: 07.11.2021. [Online]. Available: https://git.scc.kit.edu/jl1038/Readout_Simulation

[SEP16] R. Schimassek, F. Ehrler, and I. Perić, "HVCMOS Pixel Detectors - Methods for Enhancement of Time Resolution," in *Proceedings, 2016 IEEE Nuclear Science Symposium and Medical Imaging Conference: NSS/MIC 2016: Strasbourg, France*, 2016, p. 8069903. [Online]. Available: https://doi.org/10.1109/NSSMIC.2016.8069903

[Shi89] F. Shimura, *Semiconductor Silicon Crystal Technology*. Academic Press, 1989.

[Sil18] Sil'tronix Silicon Technologies, "Silicon Crystal Structure," 2018, accessed: 07.11.21. [Online]. Available: https://www.sil-tronix-st.com/fr/actualites/Silicon-crystal-structure

[SiL21] "Basil github repository," 2021, accessed: 11.11.21. [Online]. Available: https://github.com/SiLab-Bonn/basil

[SK007] S. M. Sze, Kwok K. Ng, "Physics of Semiconductor Devices", Hoboken, NJ, 2007. [Online]. Available: https://doi.org/10.1002/0470068329

[Str21] N. Striebig, "Development of Integrated Sensors for Gamma Ray Astronomy," Master's thesis, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2021.

[TBB+00] G. Timp, J. Bude, F. Baumann, K. Bourdelle, T. Boone, J. Garno, A. Ghetti, M. Green, H. Gossmann, Y. Kim, R. Kleiman, A. Kornblit, F. Klemens, S. Moccio, D. Muller, J. Rosamilia, P. Silverman, T. Sorsch, W. Timp, D. Tennant, R. Tung, and B. Weir, "The relentless march of the mosfet gate oxide thickness to zero," *Microelectronics Reliability*, vol. 40, no. 4, pp. 557–562, 2000. [Online]. Available: https://doi.org/10.1016/S0026-2714(99)00257-7

[TDP+16] P. Tropea, J. Daguin, P. Petagna, H. Postema, B. Verlaat, and L. Zwalinski, "CO_2 evaporative cooling: The future for tracking detector thermal management," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 824, pp. 473–475, 2016, frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors. [Online]. Available: https://doi.org/10.1016/j.nima.2015.08.052

[Thu18] F. Thuselt, *Physik der Halbleiterbauelemente*. Springer, 2018. [Online]. Available: https://doi.org/10.1007/978-3-662-57638-0

[TI 06] TI Literature, "Understanding Low-Dropout (LDO) Regulators," Texas Instruments, Tech. Rep., 2006, accessed 07.11.2021. [Online]. Available: http://www.ti.com/lit/slup239

[TSM21] "Logic technology," TSMC, 2021, accessed: 07.11.21. [Online]. Available: https://www.tsmc.com/english/dedicatedFoundry/technology/logic

[Veb12] D. Veberič, "Lambert W function for applications in physics," *Computer Physics Communications*, vol. 183, no. 12, pp. 2622–2628, 2012. [Online]. Available: https://doi.org/10.1016/j.cpc.2012.07.008

[VIT08] "American National Standard for FPGA Mezzanine Card (FMC) Standard," American National Standards Institute, Inc. (ANSI), Tech. Rep. ANSI/VITA 57.1-2008, 2008.

[VLFLN80] V. Van Lint, T. Flanagan, R. Leadon, and J. Naber, *Mechanisms of radiation effects in electronic materials*. Wiley-Interscience, 1980, vol. 1.

[W+18] A. Weber *et al.*, "MuPix8 — Large Area Monolithic HVCMOS Pixel Detector for the Mu3e Experiment," 2018, Poster at the Conference "Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors", accessed: 07.11.2021. [Online]. Available: https://agenda.infn.it/event/17834/contributions/83583/

[WAB+19] A. Weber, H. Augustin, N. Berger, S. Dittmeier, F. Ehrler, C. Grzesik, J. Hammerich, A. Herkert, L. Huth, J. Kröger, F. M. Aeschbacher, I. Perić, M. Prathapan, R. Schimassek, A. Schöning, I. Sorokin, D. Wiedner, H. Zhang, and M. Zimmermann, "Mupix8 — large area monolithic hvcmos pixel detector for the mu3e experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 936, pp. 681–683, 2019, Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors. [Online]. Available: https://doi.org/10.1016/j.nima.2018.09.095

[Web16] A. Weber, "Design of a Pixel Sensor Chip for Particle Physics," Master's thesis, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2016.

[Web21] A. Weber, "Development of Integrated Circuits and Smart Sensors for Particle Physics Experiments and Medial Beam Imaging," Ph.D. dissertation, Heidelberg University, 2021. [Online]. Available: http://dx.doi.org/10.11588/heidok.00030650

[Wik08] "Czochralski Process DE," 2008, accessed: 07.11.21. [Online]. Available: https://de.wikibooks.org/wiki/Datei:Czochralski_Process_DE.svg

[Wik21] "MOSFET," 2021, accessed: 07.11.21. [Online]. Available: https://upload.wikimedia.org/wikipedia/commons/1/18/IvsV_mosfet.svg

[WKH+20] M. Williams, J. Kröger, L. Huth, P. Schütze, and S. Spannagel, "Corryvreckan – A Modular 4D Track Reconstruction and Analysis Software for Test Beam Data," *Zenodo*, Dec. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4384186

[Xil14] Xilinx Inc., "Aurora 8b/10b protocol specification," Xilinx Inc., Tech. Rep., 2014, accessed: 07.11.2021. [Online]. Available: https://www.xilinx.com/support/documentation/ip_documentation/aurora_8b10b_protocol_spec_sp002.pdf

[ZESP20] H. Zhang, F. Ehrler, R. Schimassek, and I. Peric, "Measurement results on capacitively coupled particle detector with PHOTON readout chip," *Journal of Instrumentation*, vol. 15, no. 09, pp. P09 041–P09 041, sep 2020. [Online]. Available: https://doi.org/10.1088/1748-0221/15/09/p09041

[Zha21] H. Zhang, "Development of integrated detectors for charged particles and photons," Ph.D. dissertation, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2021.

Acknowledgements – Danksagung

At this point I would like to thank the many people who supported me in the making of this thesis. An dieser Stelle möchte ich all jenen danken, die mich während meiner Arbeit unterstützt haben.

Als erstem möchte ich Herrn Prof. Dr. Thomas Müller für die Gelegenheit danken, die Doktorarbeit durchführen zu können und mir als Referent zur Seite zu stehen.

Prof. Dr. Ivan Perić danke ich für die Unterstützung bei der Arbeit. Ohne ihn und seine Koordination der Projekte im und um das KIT-ADL und dessen Führung wäre diese Arbeit in dieser Form nicht möglich gewesen. Außerdem danke ich ihm für die Übernahme des Korreferats.

Allen aktuellen und ehemaligen Mitstreitern vom KIT-ADL – Felix Ehrler, Alena Weber, Roberto Blanco, Richard Leys, Horacio Mateos, Mridula Prathapan und Hui Zhang – möchte ich für die gute Zusammenarbeit, die gegenseitige Unterstützung, die auch öfters zu unerwarteten Geistesblitzen führte, und das angenehme Arbeitsklima danken.

Allen Kollegen am IPE möchte ich danken für die große Hilfsbereitschaft und die zur Verfügung gestellte Infrastruktur. Dies gilt insbesondere für die Teams in den Werkstätten und im Reinraum: Uwe Bauer, Alexander Bacher, Peter Hoffmann, Benjamin Leyrer, Bernhard Osswald, Tibor Piller und Peter Schöck; Daniel Kompalla für die Unterstützung in IT- und IT-Hardware-Fragen und Marc Schneider gilt mein Dank dafür, dass er auch für die ungewöhnlichen Probleme und kurzfristig mit Tipps und Materialien ausgeholfen hat. Auch die Mitarbeiterinnen der Administration am IPE mit Christiane Buchwald-Kayser, Antje Martin, Saskia Pulch und Tamara Wezel möchte ich explizit in den Dank einschließen, dafür, dass sie mir bei während der gesamten Zeit bei den organisatorischen Herausforderungen geholfen und so die Konzentration auf meine Arbeit ermöglicht haben.

Den Mitarbeiter der Hardware-Gruppe im Institut für Experimentelle Teilchenphysik möchte ich ebenfalls für die Hilfsbereitschaft und zur Verfügung gestellte Infrastruktur danken. Besonders bei Bestrahlungen und aufwändigen Messungen war die Hilfe und Expertise von Alexander Dierlamm, Tobias Barvich, Stefan Maier, Marius Neufeld, Andreas Nürnberg, Hans-Jürgen Simonis, Julian Stanulla und vielen weiteren unersetzlich. Besonderer Dank geht auch an Pia Steck für die vielen gebondeten Sensoren.

Weiterhin möchte ich mich bei Prof. André Schöning und seiner Arbeitsgruppe für die Unterstützung und gute Zusammenarbeit insbesondere bei der Teststrahl-Messung im Dezember 2019 am DESY[2] bedanken.

Bedanken möchte ich mich bei KSETA für die Unterstützung in der Zeit meiner Arbeit und besonders für die Organisation von Workshops und Kursen, die so wertvoll für den wissenschaftlichen Austausch sind.

Bei meiner Familie Walter, Irene, Martin und Karin möchte ich mich auch für die immerwährende Unterstützung bedanken.

[2] genauer, an der Teststrahl-Einrichtung am DESY in Hamburg (Deutschland), einem Mitglied der Helmholtz-Gemeinschaft