UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros de Telecomunicaciones

On-Chip Readout Architecture for the Novel ITS3 Wafer-Scale Pixel Sensor: From Behavioral Modeling to Physical Implementation

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Manuel Viqueira Rodríguez
Máster Universitario en Sistemas Electrónicos Avanzados

Madrid, 2025

UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros de Telecomunicaciones

Doctoral Degree in Ph.D. in Electronic Systems Engineering

On-Chip Readout Architecture for the Novel ITS3 Wafer-Scale Pixel Sensor: From Behavioral Modeling to Physical Implementation

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Manuel Viqueira Rodríguez
Máster Universitario en Sistemas Electrónicos Avanzados

Under the supervision of:
Dr. María Luisa Lopez Vallejo

Madrid, 2025

Title: On-Chip Readout Architecture for the Novel ITS3 Wafer-Scale Pixel Sensor: From Behavioral Modeling to Physical Implementation
Author: Manuel Viqueira Rodríguez
Doctoral Programme: Ph.D. in Electronic Systems Engineering

Thesis Supervision:

Dr. María Luisa Lopez Vallejo, Catedrática de Universidad (University supervisor)

External Reviewers:

Thesis Defense Committee:

Thesis Defense Date:

This thesis has been fully funded by the CERN.

A mi mujer Rihab, que dejó a su familia y toda su vida en Túnez para venir a un pais completamente diferente a empezar de cero por mí. Sin ti, nada hubiera sido lo mismo.

Acknowledgement

Many thanks to the CERN doctoral program for fully funding this thesis, giving me the opportunity to pursue my thesis fully inside CERN. The effort behind this thesis does not correspond only to me, but also to all the people who supported and helped me during this path. To my CERN supervisor, Gianluca Aglieri Rinella, for all the support during these three years. To my UPM supervisor Marisa, who is always available to give me a hand with whatever is needed. To all the MOSAIX development team, from whom I learned a lot. Special dedication to Ola, Pedro and Giulio for their close help and supervision. To my three amazing office-mates, Ioannis, Jelena and Symon for making coming to work full of unexpected fun. And to the verification team, whose “hello Manuel” messages make my heartbeat run faster than a Ferrari, the debugging never-ending story...

Beyond work, many thanks to all the nice friends I have the pleasure of meeting at CERN, and to all the gang back in Spain, who always allocate time to the few days I am back around. I truly appreciate it.

Finally, a special mention to my family. To my wife, who started a new life with me here. To my mom, dad and brother for their constant support, and to all my amazing uncles, aunts and cousins I have back home. To my grandfather who died during my PhD period and my grandmother who always wished me “health and luck”.

Abstract

The LHC's Long Shutdown 3 is scheduled to begin in 2026 and conclude in 2029. During this 3-year stop, many of the LHC's detectors will be replaced with improved versions. In this context, the ALICE collaboration is developing the Inner Tracker System 3 (ITS3), a new detector that will replace the current ITS2 in the three innermost layers of the ALICE experiment.

The ITS3 consists of three cylinders, each made up of two wafer-scale monolithic pixel detectors that were created using commercial CMOS imaging technology. These pixel detectors have a size of up to 27 cm x 9 cm, which is an unprecedented size achieved by exploiting a novel fabrication technique known as stitching. These sensors are bent around the beam pipe to form a pure cylindrical sensor, which significantly reduces the material budget. However, there are many design challenges associated with its usage, such as the power distribution, the in-chip data transmission over cm-scale links, or the yield.

This thesis focused on the ITS3 readout system modeling, design, and physical implementation. It includes modeling the readout chip using physics data from the ALICE experiment; RTL and physical implementation (synthesis, place-and-route...) of the in-chip readout architecture, the elaboration of low-leakage and DFM robust cell libraries, and the study of the ITS3 CMOS technology through transistor test structures under irradiation at different temperatures and substrate voltages.

This effort is part of the Monolithic Stitched Active Pixel (MOSAIX) chip project, which is the final research and development, and functional iteration before the ITS3 installation. The MOSAIX chip was sent to tape-out in July 2025, and the first silicon test results are expected in December 2025.

Resumen

El *Long Shutdown 3* del LHC está programado para comenzar en 2026 y durará hasta 2029. Durante esta parada de 3 años, muchos de los detectores del LHC serán reemplazados por versiones mejoradas. En este contexto, la colaboración ALICE está desarrollando el Inner Tracker System 3 (ITS3), un novedoso detector que reemplazará al actual ITS2 en las tres capas más internas del experimento ALICE.

El ITS3 consta de tres cilindros, cada uno compuesto por dos detectores de píxeles monolíticos del tamaño de una oblea completa, fabricados utilizando tecnología de imagen CMOS comercial. Estos detectores de píxeles tienen un tamaño de hasta 27 cm Œ 9 cm, una dimensión sin precedentes alcanzada mediante una novedosa técnica de fabricación conocida como *stitching*. Estos sensores se doblan alrededor del tubo de haces de partículas, logrando un sensor cilíndrico puro y una reducción importante en el *material budget*. Sin embargo, su uso plantea numerosos desafíos de diseño tales como la distribución de potencia, la transmisión de datos a largas distancias (cm) dentro del chip, o el rendimiento de fabricación.

Esta tesis se centra en el modelado, diseño e implementación física del sistema de lectura (*readout*) del ITS3. Esto incluye el modelado de la arquitectura de lectura utilizando auténticos datos físicos del experimento ALICE, el RTL y la implementación física (síntesis, place-and-route...) de la arquitectura de lectura, la elaboración de bibliotecas de celulas estándar adaptadas a *low-leakage* y robustas a las variaciones de fabricacion (DFM), así como el estudio del comportamiento bajo radiación de la tecnología CMOS utilizada para el ITS3 ante distintas temperaturas y voltajes en el sustrato.

Este esfuerzo forma parte del diseño del chip Monolithic Stitched Active Pixel (MOSAIX), la última iteración de I+D y funcional antes de la instalación final del ITS3. El chip MOSAIX fue enviado a *tape-out* en julio de 2025, y se esperan los primeros resultados de tests sobre el propio chip para diciembre de 2025.

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

ALICE A Large Ion Collider Experiment

ATLAS A Toroidal LHC Apparatus

BNL Brookhaven National Laboratory

CDC Clock Domain Crossing

CERN European Organization for Nuclear Research

CMS Compact Muon Solenoid

DFM Design for Manufacturability

DEPFET Depleted P-channel Field Effect Transistor

DESY Deutsches Elektronen-Synchrotron

DPRAM Dual Port Random Access Memory

DUT Devices Under Test

ECAL Electromagnetic Calorimeter

ECC Error-Correcting Code

EDA Electronic Design Automation

ER Engineering Run

ESD Electrostatic Discharge

FAT Fast Analytic Tool

FIFO First-In First-Out

FSM Finite State Machine

HCAL Hadronic Calorimeter

HEP High Energy Physics

IB Inner Barrel

ITS Inner Tracking System

LEC Left Endcap

LHC Large Hadron Collider

LHCb LHC-beauty

LOCOS Local Oxidation of Silicon

LPE Layout Parasitic Extraction

MAPS Monolithic Active Pixel Sensors

MIP Minimum Ionizing Particle

MLR Multi-Layer Reticle

MOS Metal Oxide Semiconductor

MOSAIX Monolithic Stitched Active Pixel

MOSS Monolithic Stitched Sensor

MOST Monolithic Stitched Sensor with Timing

NASA National Aeronautics and Space Administration

OB Outer Barrel

PCB Printed Circuit Board

PE Priority Encoder

PRBS Pseudo-Random Binary Sequence

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

QFD Quantum Flavourdynamics

QGP Quark Gluon Plasma

REC Right Endcap

RF Radio Frequency

RPCs Resistive Plate Chambers

RRU Region Readout Unit

RTL Register-Transfer Level

SBB Stitched Back Bone

SDC Synopsys Design Constraint

SEE Single Event Effect

SEL Single Event Latch-Up

SET Single Event Transient

SEU Single Event Upset

SLVT Super Low V_{TH}

SM Standard Model

SPD Silicon Pixel Detectors

SPS Super Proton Synchrotron

SSD Silicon Strip Detectors

STI Shallow Trench Isolation

SVT Standard V_{TH}

TCAD Technology Computer-Aided Design

TID Total Ionizing Dose

TMR Triple Modular Redundancy

ToT Time-over-Threshold

TPC Time Projection Chamber

TRU Top Readout Unit

UPM Universidad Politécnica de Madrid

WNMU Window Management Unit

Chapter 1

Introduction

This chapter outlines the motivation behind the thesis. It begins with a brief overview of the Standard Model and the Quark-Gluon Plasma (QGP), followed by an introduction to the ALICE experiment at CERN, including its purpose and objectives. The chapter then presents the ITS3 project, highlighting its motivation and the key challenges it poses. It concludes by outlining the author's contributions to the ITS3 project and explaining how these are discussed in the subsequent chapters.

1.1 The Standard Model

The Standard Model (SM) of Physics is a theory that describes the elementary particles of nature and the interactions among them [1]. Although it remains incomplete, it stands as the most accurate theory for the behavior of quantum particles. In Figure 1.1, we can see how the SM divides the particles into fermions (subdivided into quarks and leptons) and bosons. These particles interact with each other through four fundamental forces: electromagnetism, strong force, weak force, and gravity. The theories that study these forces are quantum electrodynamics (QED) for electromagnetism, quantum Chromodynamics (QCD) for the strong force, quantum flavourdynamics (QFD) for the weak force, and Geometrodynamics for gravity. Geometrodynamics has yet to provide a fully satisfactory theory that explains gravity at the quantum level. At this moment, gravity is assumed to be too weak compared to the other three forces to play a significant role at the quantum level.

Each force finds representation through a gauge boson or force carrier: the photons govern the electromagnetic interactions, while W and Z bosons regulate the weak interaction, and gluons manage the strong force. A gravitational force carrier (graviton) was also hypothesized, but its existence has not been proven yet. Additionally, a scalar boson, known as the Higgs boson, determines the manifestation of mass.

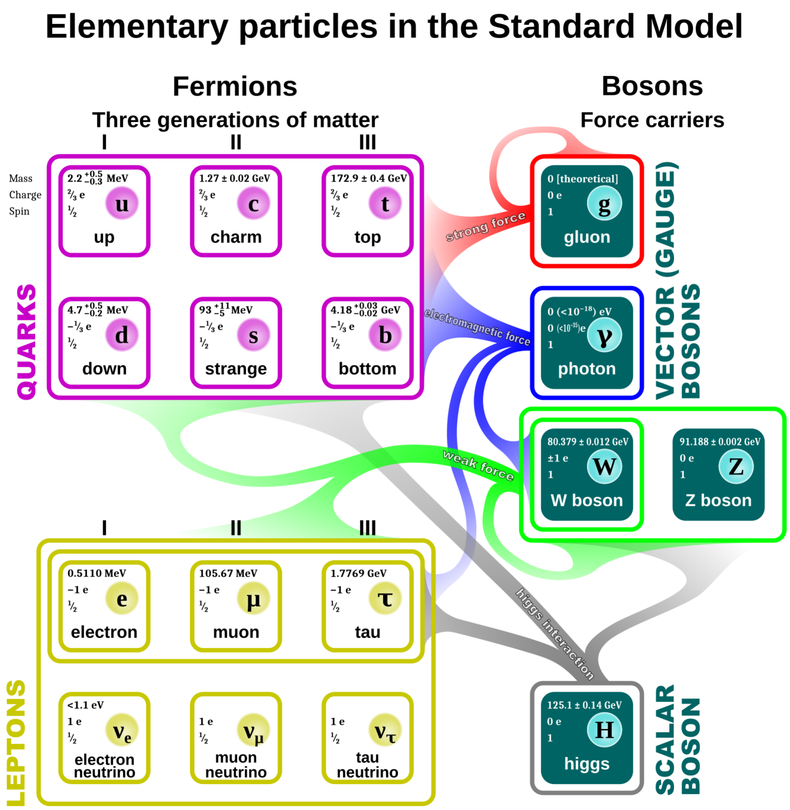


Figure 1.1: The Standard Model (SM) of Elementary Particles (Source [2]).

1.2 The Quark Gluon Plasma (QGP)

The Quark-gluon plasma (QGP) [3] is an extreme state of matter from which it is believed that the universe used to be made of $\sim 1\mu s$ after the Big Bang, as seen in Figure 1.2. In this extreme state, the quarks and the gluons are not confined within individual protons and neutrons as in normal matter. Instead, they move freely, forming a hot and dense soup-like medium. QGP is of significant interest to physicists because it provides a window into the conditions of the early universe.

1.3 The LHC

The Large Hadron Collider (LHC) [5] is the world's largest particle accelerator. It is located at CERN (the European Organization for Nuclear Research) in Geneva, Switzerland. The LHC is buried approximately 100 meters underground in a tunnel that has 27 kilometers of circumference.

The LHC is designed to accelerate two beams of particles that travel in opposite directions around the LHC circular tunnel. The LHC uses a series of superconducting magnets to steer the particles and a radio frequency (RF) cavity to accelerate them. When these beams are

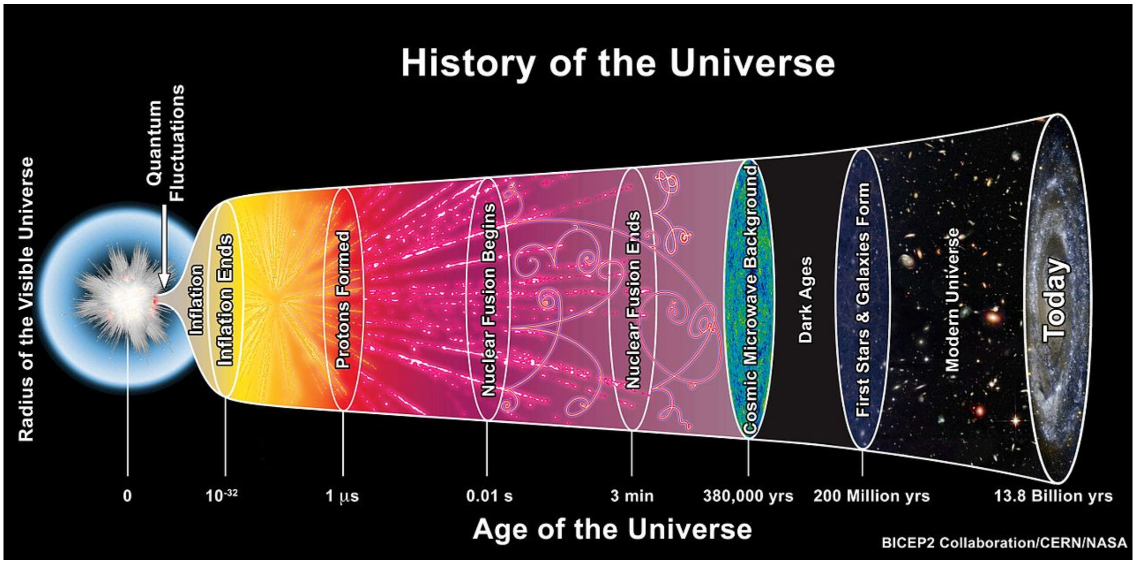


Figure 1.2: The history of the universe. The quark-gluon plasma became protons around $1\mu s$ after the Big Bang (Source [4]).

brought into collision, the particles smash into each other, generating a cascade of particles, which decay into other particles.

Around the LHC tunnel circumference, there are four major experiments in which the beams can collide. These experiments are: ATLAS, CMS, ALICE, and LHCb. The objective of these experiments is to observe these collisions and reconstruct what happens in them.

The operation of the LHC is divided into Runs, separated by Long Shutdowns in which the different experiments and acceleration infrastructure of the LHC are maintained and upgraded. Two long LHC runs have been carried out so far, Run 1 (2010 - 2013) and Run 2 (2015 - 2018). Currently, the LHC is performing the Run 3 (2022 - 2026) that will end with the Long Shutdown 3, which will turn off the LHC from 2026 to 2029.

During these Runs, the LHC primarily operates with proton-proton (p-p) collisions. However, approximately one month per year, it is dedicated to its heavy-ion physics program [6], [7]. During this period, the LHC collides heavy ions, primarily ^{208}Pb ions in lead-lead (Pb-Pb) and lead-proton (Pb-p) collisions. The goal of these collisions is to recreate the extreme temperature and pressure conditions necessary to produce QGP, as discussed in Section 1.2. Although all LHC experiments participate in heavy-ion physics, the ALICE experiment is the only one specifically optimized for the reconstruction of heavy-ion collision events.

1.4 The ITS3

During the Long Shutdown 3, the ALICE collaboration is planning to replace the three innermost layers of the current Inner Tracker System 2 (ITS2) with the Inner Tracker System 3 (ITS3) [8]–[10]. This novel detector consists of six half-cylinder layers constructed from a single-die Monolithic Active Pixel Sensor (MAPS), as illustrated in Figure 1.3.

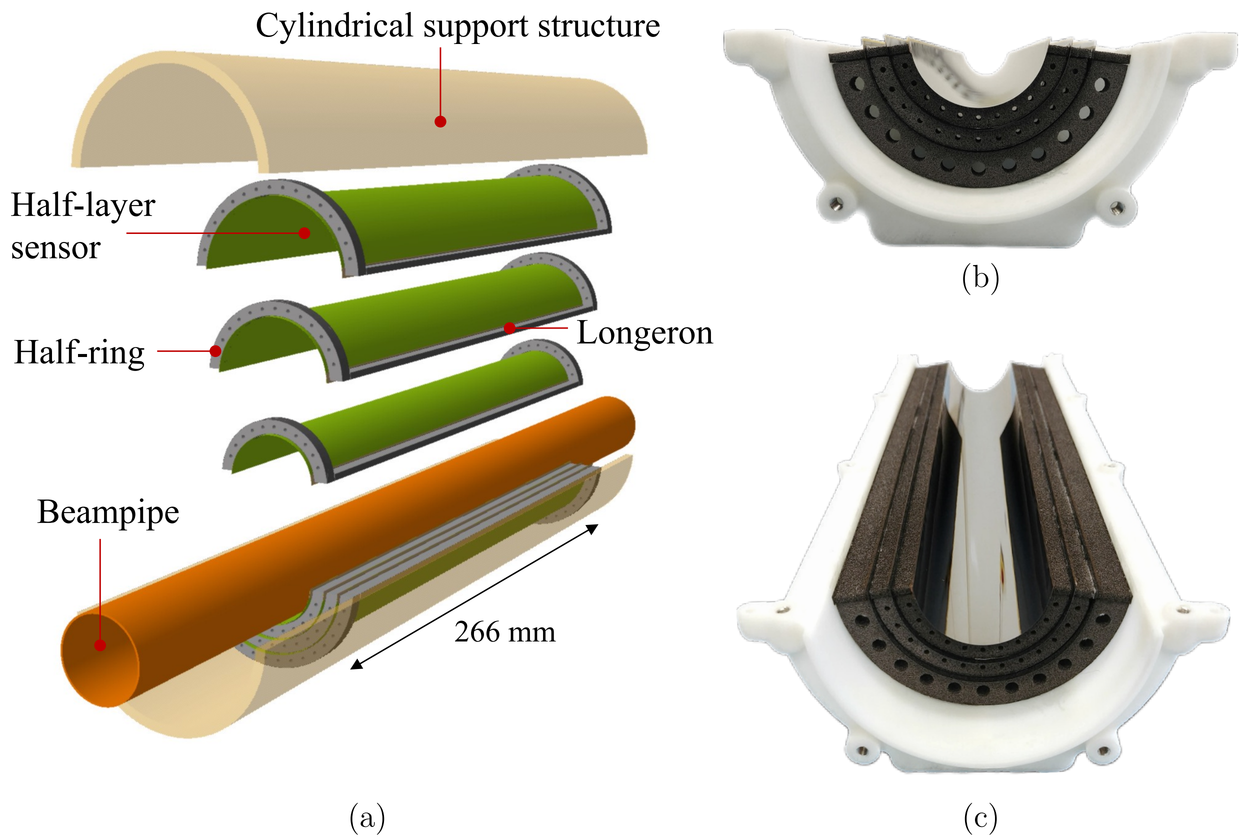


Figure 1.3: The ITS3 detector surrounding the beam pipe. Each layer is composed of two bent MAPS sensors forming a cylindrical structure around the pipe. (a) Exploded schematic view of the three ITS3 layers, showing the half-sensor modules and mechanical support. (b, c) Front and elevated views of a mechanical mock-up with three half-layers and the carbon foam spacers between them. (Source [9]).

With respect to the ITS2, the ITS3 plans to present several improvements, with the most important highlighted in Table 1.1. The ITS3 aims to remove all non-silicon material of the MAPS: PCB, glue, Kapton layers, aluminum, and water cooling systems. As a result, the ITS3 consists only of the silicon detector, an ultra-light carbon fiber foam (shown in Figure 1.3), and a cooling airflow. Removing these materials reduces the detectors radiation length X_0 from 0.35% to 0.086%, as seen in Figure 1.4. This X_0 improvement, combined with placing the first layer closer to the interaction point and the reduction in the pixel pitch, doubles the position resolution compared to ITS2 as shown in Figure 1.5.

Each ITS3 layer is bent around the beam pipe at radii of 19, 25.2, and 31.5 mm. These sensors are manufactured using commercial 65 nm CMOS imaging technology on wafer-scale substrates up to 266 mm $\times$ 93 mm. This unprecedented chip size is made possible through a fabrication technique known as *stitching*, which allows the production of chips larger than the standard reticle size that photolithography tools can achieve [11].

To use stitching, the chip design is divided into smaller sections called sub-reticles that fit within the lithographic field. These sub-reticles are precisely aligned and seamlessly tiled both

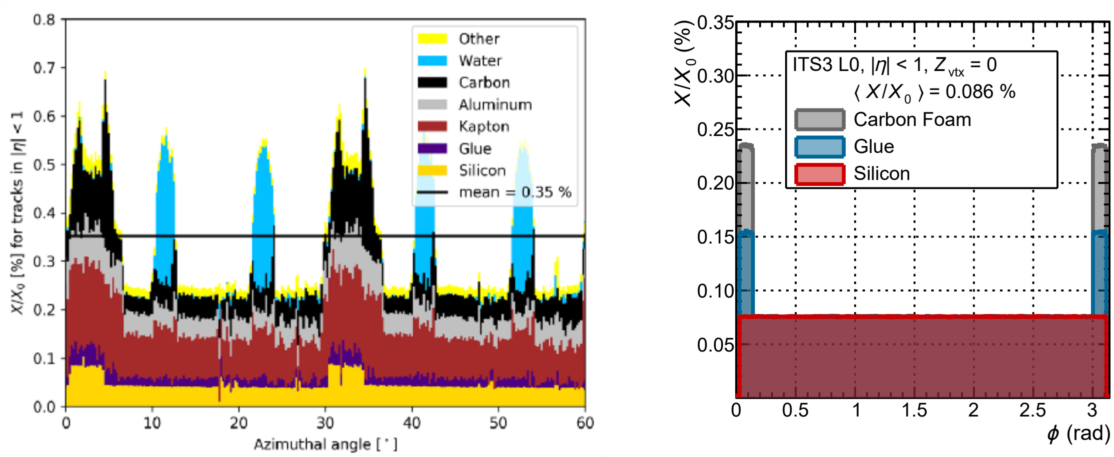
Table 1.1: Key improvements of ITS3 with respect to ITS2.

Parameter	**ITS2**	**ITS3**
Technology node	180 nm	65 nm
Material budget (X_0)	0.35%	0.086%
Distance of first layer	24 mm	19 mm
Pixel size	29.24 × 26.88 μm	22.8 × 20.8 μm
Impact-parameter resolution ($r\varphi$, $\rho_T = 1.3$ GeV/c)	20 μm	10 μm

horizontally and vertically by connecting the boundaries of adjacent sub-reticles through metal interconnect layers. The main challenge of this technique is the accuracy of the alignment, as a misalignment even of a few nanometers can produce defects that render the entire chip unusable [12]. For building the ITS3, three sub-reticle modules were designed: The Repeated Sensor Units (RSUs) that hold the sensitive part of the chip, the Left Endcap (LEC) that hosts the power and readout Input/Outputs, and the Right Endcap (REC) that hosts only power connections.

Each half-layer ITS3 chip requires a different dimension depending on the layer where it will be installed. To fabricate the chips for each layer, each half-layer ITS3 chip is divided into three, four, and five common design blocks called segments, as shown in Figure 1.6. Each segment measures 266 mm × 18.5 mm, and is as well subdivided into 12 Repeated Sensor Units (RSUs) and the two endcaps in the edges. To improve fault tolerance, each RSU is further divided into 12 tiles, which can be selectively disabled if a defect is detected in them.

As the LEC contains all I/O of the segment, all data from the tiles must be sent to it through distances up to 26.6 cm. As these long-distance communications are performed on top of the pixels, they are routed as a differential signal with a low swing, thereby reducing the induced noise in the pixels and the power consumption. This signal is re-buffered every 1.08 cm in



(a) (b)

Figure 1.4: Material budget expressed in terms of radiation length for a section of the currently installed ITS2 (a) and the expected ITS3 (b) (Source a) [8], b) [9]).

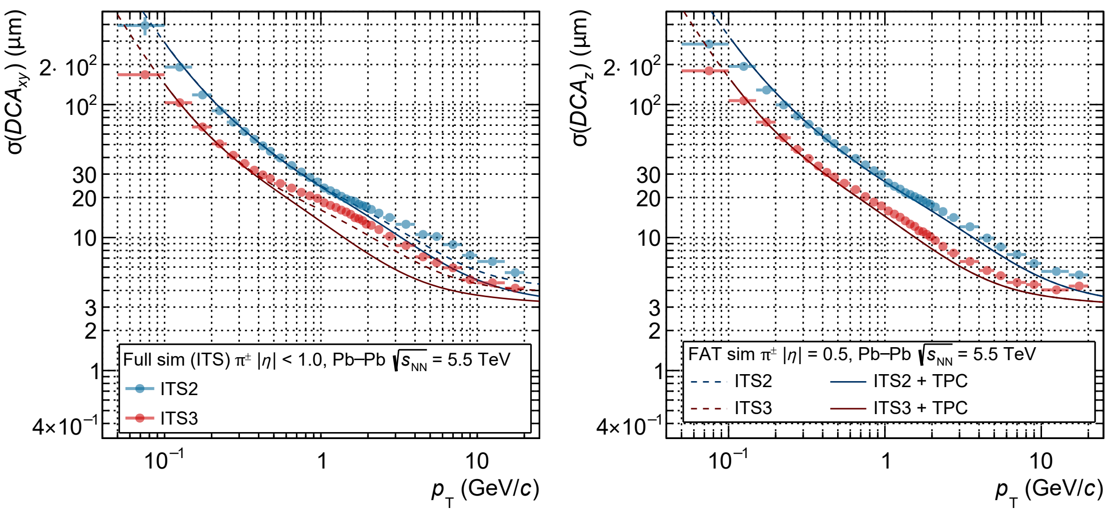


Figure 1.5: Impact-parameter resolution in the transverse ($r\varphi$, left) and longitudinal (z, right) directions for primary charged pions with pseudorapidity $|\eta| < 1$ as a function of transverse momentum p_T. Only tracks with a hit on each ITS layer are considered. Results are obtained using the Fast Analytic Tool (FAT) for ITS-only tracking and with the inclusion of the Time Projection Chamber (TPC). ITS3 improves impact-parameter resolution by a factor of 2 (Source [9]).

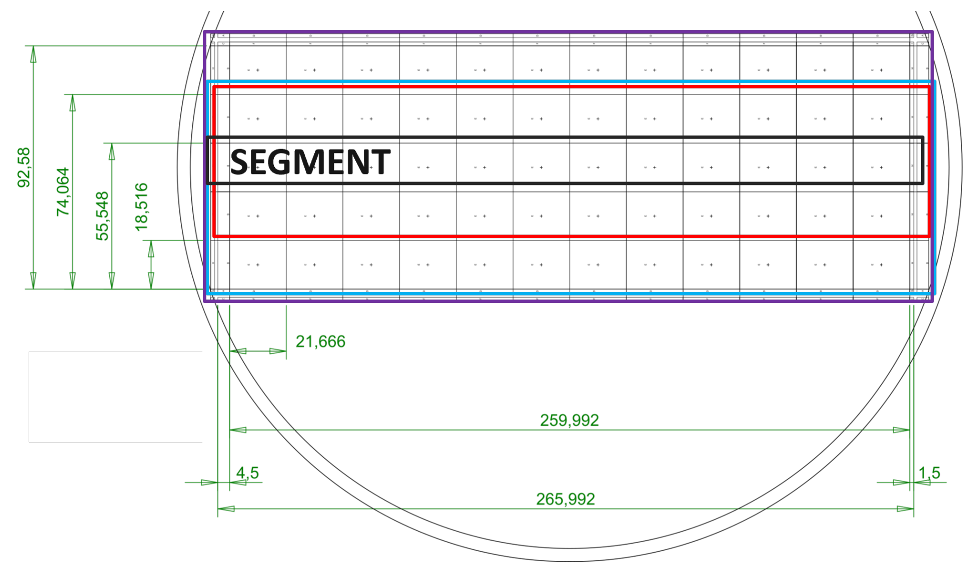


Figure 1.6: Wafer-scale ITS3 sensor dies fabricated using 65 nm technology.

special custom differential buffers known as Stitched BackBone (SBB).

1.4.1 On-chip readout architecture

The tiles contain the sensitive area of the ITS3, the pixel matrix. The addresses of the pixels that detected a particle crossing them during a certain period of time must be read, temporally stored, and finally shipped to the LEC. The system that performs this read process is called the readout architecture. A schematic of the readout architecture of a tile is shown in Figure 1.7.

As seen, each pixel contains two registers: an *integration register* and a *readout register*, which implement a frame-based readout process. This frame-based process consists of a global frame signal that defines consecutive integration windows. The duration of these integration windows is configurable, with a minimum period of 2 μs.

If the front-end circuitry detects a charge above a programmable threshold, the corresponding pixel sets its integration register, marking a hit. At the end of the frame, the data in the integration register is copied to the readout register and cleared for the next frame. The readout register is read through a zero-suppressed fully combinatorial circuit into a FIFO. To improve throughput, the pixel matrix is partitioned into regions that are read in parallel. These FIFOs are read by a top-level readout that serializes the data and transmits it to the left endcap. All addresses of the pixels hit within an integration window are shipped together in an integration window packet, which contains a time-stamp corresponding to the integration window in which they were detected.

Due to fluctuating occupancy, readout congestion may occur. The top readout system may fail to read all data from the readout register within an integration window due to insufficient bandwidth or FIFO depth. In this event, the information from the next integration window will be dropped, providing an extra time to complete the previous readout.

Each of the 144 tiles sends data via a 160 Mb/s serial link to the LEC, as shown in Figure 1.8. Each tile readout architecture is fully independent from the rest, and can potentially drop integration window packets. As the information from all the tiles is required for a correct

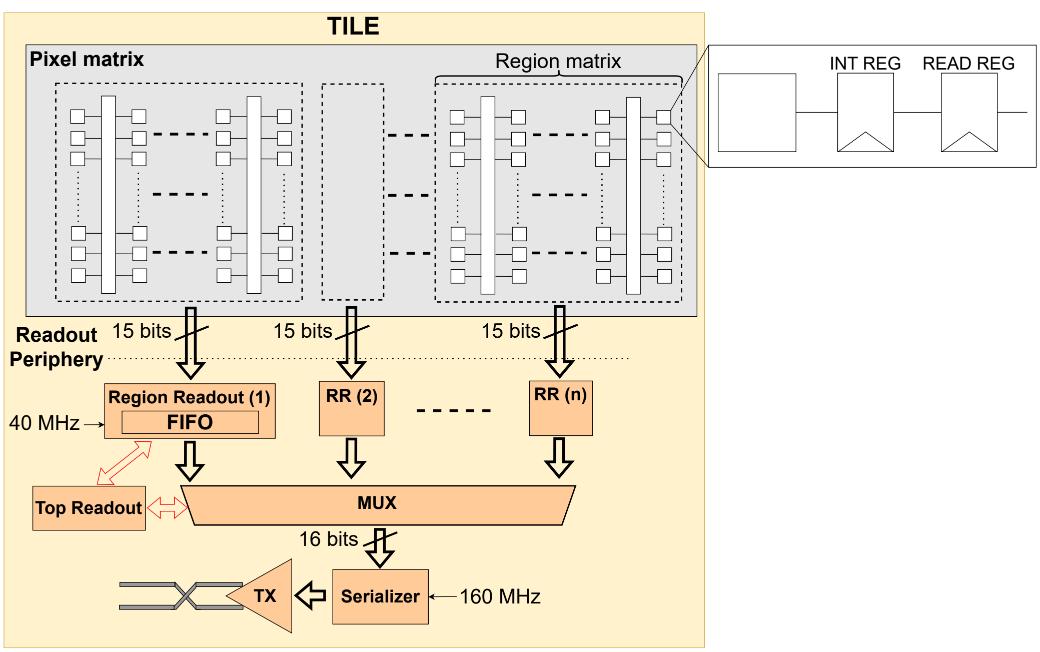


Figure 1.7: Schematic of the readout architecture of an individual tile.

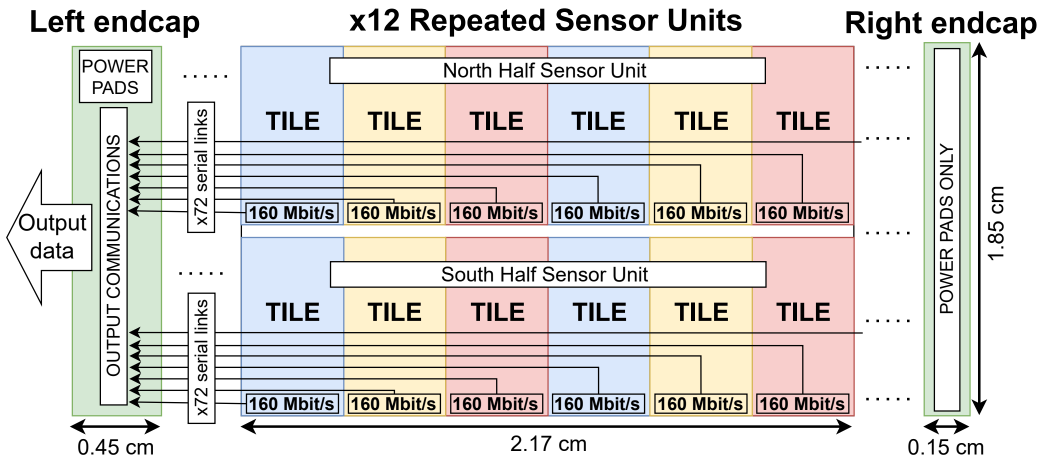


Figure 1.8: Breakdown of a sensor segment into two endcaps and 12 RSUs with 12 tiles each.

event reconstruction by the physicists, each integration window packet must be sent complete.

The readout performance of the readout architecture is therefore measured in the fraction of integration windows in which at least one tile dropped an integration window packet. This metric depends on several parameters: the number of FIFOs, the FIFO depth, the tile clock frequency, the number and bandwidth of links to the endcap... To meet ITS3s stringent challenges, the readout must minimize data loss, area, and power consumption.

1.4.2 Challenges

The design of a stitched sensor like ITS3 presents several engineering challenges. Among them, the most important are:

- **Bending:** One of the initial challenges presented in the ITS3 design was the behaviour of the sensor under bending forces. This challenge was deeply studied and analyzed in several works before the development of this thesis [13]–[15] presented in Section 2.5.4.
- **Yield:** Larger chips are more prone to manufacturing defects. To mitigate this, Design for Manufacturing (DFM) techniques must be used to enhance robustness.
- **Signal Integrity:** A large chip needs large wires. These wires must be periodically repeated to avoid potential signal integrity issues.
- **Power Distribution and Consumption:** To minimize the material budget, flexible cables across the chip must be avoided. As a result, power connections are restricted to the left and right edges of the cylindrical MAPS. This means that transistors near the edges will operate close to the supplied voltage, while central areas of the chip will suffer from voltage drops. To ensure correct functionality in these regions, the chip must be designed using low-voltage analysis corners and follow a power-aware design methodology targeting an ultra-low power design.
- **Dead Area:** MAPS detectors combine readout and sensitive area in the same piece of silicon. Due to this, the larger the readout, the less area is available for the sensitive

part of the chip. This non-sensitive part of the chip is known as the dead area. As the ITS3 is composed of three MAPS layers bent towards the beam pipe, the dead area must be minimized to guarantee the best reconstruction possible of the secondary particles generated from the Pb-Pb collisions.

1.4.3 ITS3 Project Planning

The ITS3 project involves multiple teams located across several institutions and cities, forcing the ITS3 project to coordinate through a robust set of collaborative tools and processes. Communication is handled via Mattermost, email, and Zoom, while CERNBox, Git, and Cliosoft SOS facilitate version control, and the exchange of documentation, code and analog modules.

Beyond software tools, regular meetings are essential for maintaining alignment and enabling finer-grained management. A weekly general meeting is held every Tuesday, allowing management to identify topics requiring additional resources or attention. Furthermore, when specific technical challenges become critical, dedicated topical meetings are convened. These sessions focus expertise on solving or mitigating potential issues.

Due to its complexity, it was structured into a sequence of prototype runs:

- **MLR1 (2021):** The Multi-Layer Reticle 1 (MLR1), aimed at optimizing the fabrication process and qualifying the chosen technology [16].
- **ER1 (2023):** The Engineering Run 1 produced the first stitched sensor prototypes, providing valuable data on stitching feasibility, manufacturing yield, and wafer-scale sensor performance. This run included two stitched structures: the Monolithic Stitched Sensor (MOSS) [11] and the Monolithic Stitched Sensor with Timing (MOST) [17].
- **ER2 (2025):** The Engineering Run 2 delivered the first fully functional ITS3 prototype. This first ITS3 prototype chip was named the Monolithic Stitched Active Pixel Sensor (MOSAIX). It must include all components and implement all the specifications of the final ITS3 chip.
- **ER3 (2026):** Scheduled for mid-2026, the final ITS3 tape-out will take place in Engineering Run 3 (ER3). The final chip will include only minimal changes from the MOSAIX prototype, avoiding the introduction of potential last-minute problems. The installation of the final ITS3 detectors is planned between 2026 and 2029.

1.5 Thesis Objectives

The core objective of this thesis is the design of the on-chip readout architecture for the MOSAIX ER2 chip, which will be the same architecture used in the final ITS3. This goal has been divided into the following sub-objectives:

- **Model:** To properly dimension the components of the on-chip readout architecture, the expected pixel hit data load in the ALICE cavern and its readout architecture were modeled. From this model, a suitable balance between resource utilization and performance was derived.

- **Design:** Once the required resources were evaluated, the RTL and circuitry of all readout components were designed. This included in-pixel logic, matrix readout, and transmission to the endcap. The design also included support for corner scenarios, as well as various operation modes and configurations that the user controls.
- **Debug and Verification:** Extensive verification was performed to validate all operating modes, configurations, and corner conditions.
- **Physical Implementation:** After RTL completion, the design was implemented in the physical tile area using a digital-on-top Electronic Design Automation (EDA) flow with CADENCE tools such as Genus and Innovus.
- **Characterization of technology under Radiation:** Since the chip will operate in a radiation environment, the impact of radiation on the selected technology was investigated.

1.6 Author's Contribution

The author actively participated in the MOSAIX project from 2022 to 2025, with a primary focus on the modeling, design, and implementation of the readout architecture. Due to the complexity of the project, all the work reported in this thesis is the result of a collaborative effort among the members of the MOSAIX team; the author contributed to the design, coding, debugging, and physical implementation.

The author collaborated on the development of the pixel hit physics data model with the support of physics specialists. The readout model was primarily written by the author.

The design and implementation of the pixel matrix was carried out by other members of the MOSAIX team; however, the author contributed to the design, documentation, and verification of the in-pixel logic and overall matrix readout. The RTL of the readout architecture was mainly written by the author, while some sub-modules (FIFOs, serializer, and timebase handler) were implemented by other team members.

The standard-cell library was a joint effort involving several engineers, including the author. The physical implementation was mostly performed by other team members, with the author being responsible for timing closure, power reduction techniques, and overall review and support.

The Transistor Test Structures (TTS) used for the radiation tests were designed by other team members. The irradiation and subsequent analysis were performed by the author in collaboration with the corresponding experts.

As explained, the author actively participated in several MOSAIX and CERN working groups, contributing to many collaborative activities. The author wishes to emphasize the team effort behind the results presented in this thesis and to share credit with all the colleagues involved.

1.7 Thesis Structure

This thesis is organized into several chapters. These chapters are:

- **Chapter 2: Background**
  This chapter provides an introduction to pixel detectors, covering the fundamental physics principles behind them, key characteristics, and the various types used in high-energy physics (HEP) experiments. It also discusses different pixel detector technologies and presents the current state of the art in monolithic pixel detectors.

- **Chapter 3: ITS3 Pixel Hit Sources Model**
  To properly size the data links and on-chip memory in the MOSAIX, it is crucial to understand the expected hit rate and its distribution. This chapter presents an accurate pixel hit sources model based on real physics data from the ITS3 TDR [9].

- **Chapter 4: MOSAIX Readout Model Design and Analysis**
  This chapter describes a simplified model of the readout architecture. This model was used under the hit source model from Chapter 3 to test various configurations of protocols, data links and memory sizes. From this analysis, an optimal configuration of these parameters is obtained and used for the RTL implementation presented in Chapter 6.

- **Chapter 5: The MOSAIX Pixel Matrix**
  This chapter introduces the pixel architecture of the MOSAIX chip. It begins with a brief overview of the pixel sensor and analog front-end, followed by a detailed description of the in-pixel digital logic (digital pixel). The operation of the pixel matrix and its interaction with the MOSAIX readout logic are also described.

- **Chapter 6: MOSAIX Readout Architecture Design and Implementation**
  This chapter provides an in-depth explanation of the readout architecture's design and physical implementation, highlighting the MOSAIX design challenges such as timing closure along critical paths and strategies for reducing power consumption.

- **Chapter 7: ITS3 Technology Characterization under Radiation**
  This chapter presents a thorough characterization of the 65 nm technology used for the MOSAIX chip. For this characterization, radiation tests using TTS structures were conducted at different temperatures and bulk voltages.

- **Chapter 8: Conclusion**
  A summary of the thesis, highlighting key challenges, solutions, and the main findings.

Chapter 2

Detector Principles and State of the Art

The ALICE experiment consists of multiple sub-detectors surrounding the interaction point as depicted in Figure 2.1. These sub-detectors enable an indirect analysis of the QGP produced in the Pb-Pb collisions. This analysis is performed by measuring the energy of all particles generated in the Pb-Pb collisions and reconstructing their trajectories as shown in Figure 2.2.

Not all particles generated in these collisions can be detected. For a particle to be detectable, it must satisfy two essential criteria:

- **Sufficiently long lifetime:** If a particle decays too quickly into other particles, it cannot be directly observed, as it will not leave a measurable track in the detectors. Particles with long enough lifetimes include electrons (e^-), positrons (e^+), protons (p^+), neutrons (n), photons (γ), neutrinos (ν), muons (μ^-), and antimuons (μ^+), among

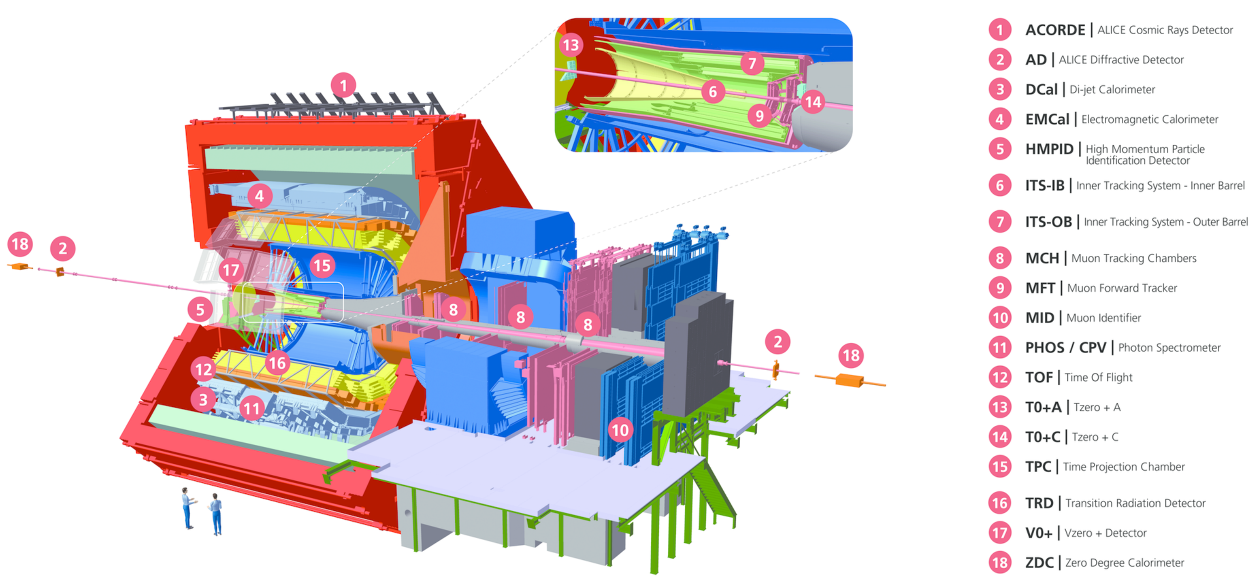


Figure 2.1: The ALICE experiment, showing its various sub-detectors (Source [18]).

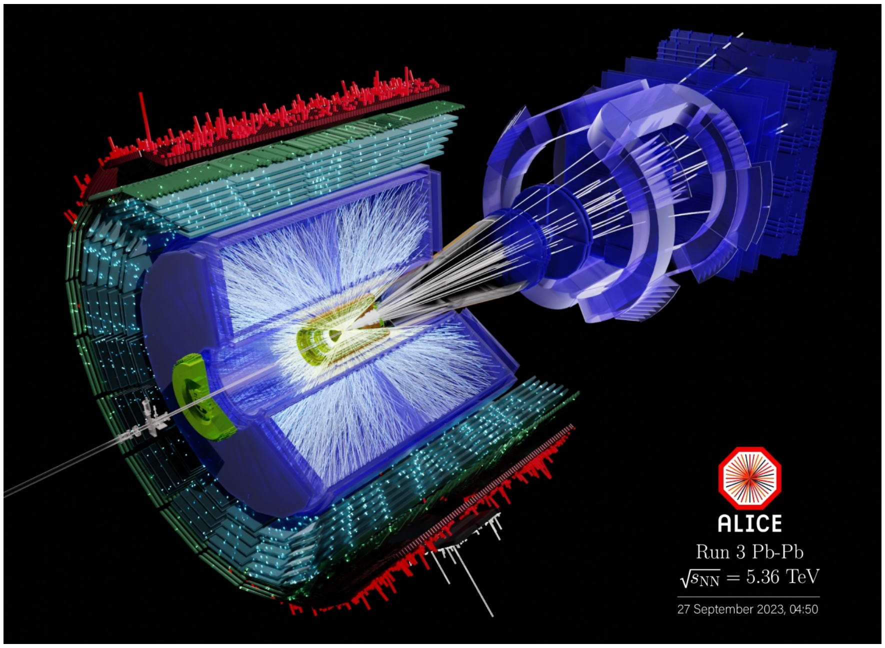


Figure 2.2: Event displays of the first Pb-Pb collision of Run 3 taken at ALICE (Source [19]).

others. In contrast, short-lived particles, such as the Z or $Higgs$ bosons, decay too quickly to be detected directly. Instead, their existence is inferred by reconstructing the particles into which they decay.

- **Interact with the detector material:** The crossing particle must interact with the detector in a way that produces an electric signal. Once such a signal is generated, various measurements can be extracted, including the particle's exact position, the time it crossed the detector, and the energy it deposited. By analyzing these data, physicists can reconstruct the entire event, identifying both directly detected particles and deducing those that cannot be detected.

This chapter provides a concise introduction to particle detectors. It covers the fundamental physical principles behind their operation, the various types of detectors and their characteristics, with a detailed focus on semiconductor detectors. Additionally, it includes a description of the current ITS2, the predecessor of the ITS3.

2.1 Interactions between Particles and Matter

The way particles interact with the detector material depends on the type of particle. Charged particles such as electrons, positrons, and protons interact electromagnetically. Protons, neutrons, and other quark-based particles can interact via the strong nuclear force. Neutrinos, however, have no charge to interact electromagnetically, nor do they interact via the strong force. Because of this, neutrinos pass through LHC detectors without leaving a trace. Dedicated neutrino detectors exist [20], [21], but their size and complexity prevent their installation at the LHC. Instead, the presence of neutrinos in LHC experiments is inferred through the observation of missing energy in collision event reconstructions.

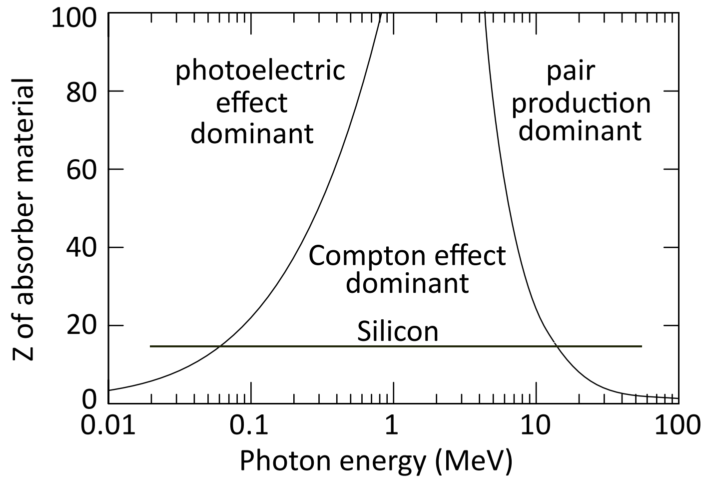


Figure 2.3: Regions in which the photoelectric effect, Compton effect, and pair production are dominant as a function of photon energy and atomic number of the absorber Z. Silicon Z is depicted as a straight horizontal line (Source [22]).

Detectors exploit the different ways particles interact with matter to extract information about them. This section explains the different interaction mechanisms based on the type of incident particle.

2.1.1 Photon Interactions

As explained in Section 1.1, the force carrier of the Quantum Electrodynamic force (QED) is the photon. Photons can interact with the charged particles of the detector, resulting either in a full absorption of the electron *(photoelectric effect, pair production)* or being deviated (scattered) through a relatively large angle (*Compton effect*).

A photon beam is attenuated exponentially through the material following the equation:

$$I = I_0 e^{-\mu x},$$

where μ is the *mass attenuation coefficient*, related to the cross section of the various attenuation processes a photon beam can suffer when crossing a certain material. The *mass attenuation coefficient* can be dominated by the photoelectric, Compton, or pair production effects, depending on the energy of the photon and atomic number of the absorption material, as seen in Figure 2.3.

2.1.1.1 Photoelectric Effect

For low photon energies (from ionization energy until ∼100 keV), the photoelectric effect dominates. In this process, a photon is completely absorbed by an atomic electron. This effect cannot occur for a free electron because, due to momentum conservation, a third interaction partner is required, typically, the entire atom, which ensures momentum conservation within the bound system.

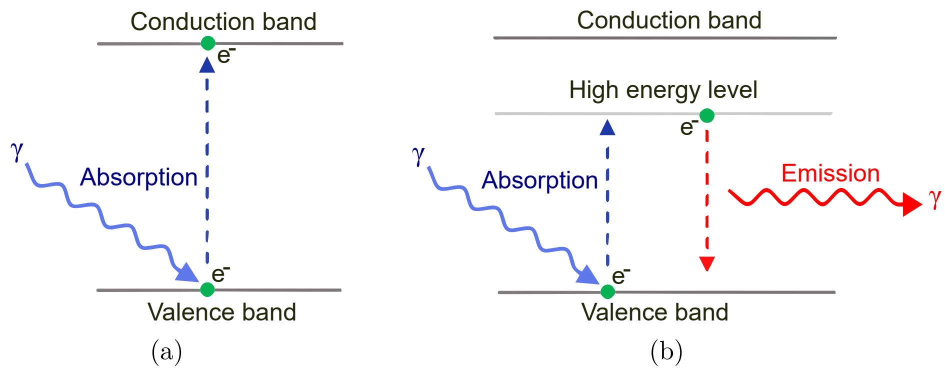


Figure 2.4: Atomic electron absorbing an incident photon: (a) Generate an electron-hole pair if the photon's energy is sufficient. (b) Gain energy and later release it by emitting a photon if the energy is insufficient to create an electron-hole pair.

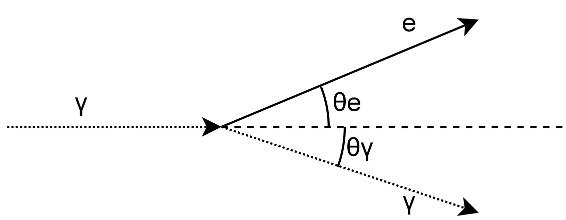


Figure 2.5: Compton effect scattering. In it, an incident photon is absorbed by an electron. As a result, the electron and the photon change their momentum, deviating their trajectory by the angles θ_e and θ_γ respectively.

As a result of this absorption, the electron gains energy. If the absorbed energy is sufficient to excite the electron from the valence band to the conduction band, the electron becomes a mobile charge carrier, generating an electron-hole pair

$$\gamma + atom \rightarrow atom^+ + e^-.$$

If the excited electron does not acquire enough energy to reach the conduction band, it will remain in a higher energy state. At this state, it can either gain additional energy and transition to the conduction band, forming an electron-hole pair, or return to the valence band by emitting a photon. These two behaviors are depicted in Figures 2.4.

2.1.1.2 Compton Effect

For medium energies (typically in the range of 100 keV $\leq E_\gamma \leq$10 MeV, depending on the atomic number, Z, of the material), the Compton effect dominates. This phenomenon occurs when a photon interacts with a quasi-free electron, absorbing part of the energy and causing a deviation of the electron and the photon. This is depicted in Figure 2.5 and can be expressed as

$$\gamma + e^- \rightarrow \gamma + e^-.$$

2.1.1.3 Pair Production

For high energies ($E_\gamma > 10$ MeV), pair production becomes the dominant interaction mechanism. Pair production is a fundamental quantum electrodynamic (QED) process in which a high-energy photon interacts with the electromagnetic (Coulomb) field of a charged particle, typically an atomic nucleus, leading to the creation of an electron-positron pair. This Coulomb field can be understood as a cloud of virtual photons surrounding the charged particle that mediate the interaction:

$$\gamma + nucleus \rightarrow e^- + e^+ + nucleus.$$

This process is subject to an energy threshold condition, as the photon must possess sufficient energy to create an electron-positron pair. The minimum required photon energy corresponds to the total rest energy of the electron and the positron as:

$$E_\gamma = 2m_e c^2.$$

If pair production takes place through interaction with the Coulomb field of an atomic electron instead of a nucleus (triplet production), the photon must have twice the energy as:

$$E_\gamma = 4m_e c^2.$$

This increase is due to the involvement of three final-state particles (two electrons and a positron) requiring additional energy to satisfy momentum conservation:

$$\gamma + e^- \rightarrow e^- + e^- + e^+.$$

As a result, pair production predominantly occurs in the vicinity of the strong Coulomb field of an atomic nucleus rather than through interactions with electrons.

2.1.2 Interaction of Charged Particles

When a charged particle travels through space, it is not just the bare particle that moves. It is accompanied by a surrounding cloud of virtual photons. These virtual photons represent the electromagnetic field of the particle.

As a charged particle, such as an electron or a proton, passes through a material, this electromagnetic field can interact with the electrons and atoms of the material. This interaction can be manifested in the excitation and ionization of electrons, multiple scattering, the bremsstrahlung effect, direct electron-positron pair production, or by Cherenkov radiation.

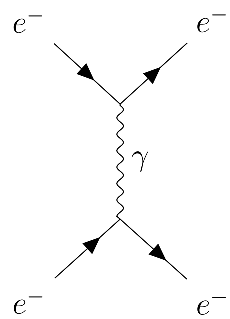


Figure 2.6: Feynman diagram of an electron-electron scattering by a virtual photon.

2.1.2.1 Excitation and Ionization of Electrons (Scattering)

One common interaction is the excitation or ionization of atoms, illustrated as:

$$\begin{aligned} e^- + atom &\rightarrow atom^* + e^- \\ &\qquad\qquad \hookrightarrow atom + \gamma. \end{aligned}$$

In this process, the charged particle loses energy, which is transferred to the electrons or nuclei of the atoms in the material as photons. This physical process is also known as scattering, a process where two charged particles (e.g., an incident and a bound electron) exchange virtual photons. In Feynman diagram [1], [23], this can be depicted as shown in Figure 2.6.

If the transferred energy is sufficient to raise a bound electron to a higher energy level within the atom, it leads to excitation. If the energy exceeds the electrons binding energy, the electron is ionized, leaving the atom and creating an electron-ion pair.

Excitation and ionization are the dominant energy loss processes of incident charged particles up to very high particle velocities, when the radiation effects start to play a more dominant role. The energy lost per unit distance of an incident charged particle due to excitation and ionization is described by the Bethe-Bloch equation:

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

where:

- $\frac{dE}{dx}$: The energy loss per traveled distance through the material in MeV/(g/cmš).
- z: Charge of the incident particle (in units of elementary e).
- Z: Atomic number of the absorber material.
- A: Atomic weight of the absorber material.
- m_e: Mass of the electron.
- r_e: Classical electron radius.

- c: Speed of light in vacuum.
- N_A: Avogadro's number.
- $\beta = \frac{v}{c}$: Velocity of the particle as a fraction of c.
- I: Mean excitation energy of the absorber material, approximated as:

$$I = 16\ Z^{0.9}\ eV \text{ for } Z > 1.$$

- $\gamma = \frac{1}{\sqrt{1-\beta^2}}$: The Lorentz factor.
- δ: The *density correction factor* [24] which describes how much the extended transverse electric field of incident relativistic particles is screened. The factor δ becomes important for high-energy particles, typically at relativistic speeds where the particle's velocity approaches the speed of light. This correction accounts for the fact that materials have a finite ability to polarize, reducing energy loss in dense media at high energies.

In crystalline materials, additional considerations to the Bethe-Bloch are needed. The periodic structure of the crystal can lead to anisotropic energy losses, meaning that the energy deposited in the material may vary depending on the angle of incidence relative to the crystal axes.

From the Bethe-Bloch equation, we can note that the rate of energy loss is inversely proportional to β^2, implying that faster particles lose less energy per unit length. Additionally, the energy loss does not depend on the mass of the incident particle, only on its charge and speed.

This leads to an important implication: Although an electron and a muon have the same electric charge, the muon is about 200 times more massive. At the same speed, they interact similarly with matter in terms of ionization energy loss. However, due to its greater mass, the muon penetrates much deeper before losing all its energy.

2.1.2.2 Multiple Scattering

When an incident charged particle traverses a material, it undergoes repeated electromagnetic interactions with the Coulomb fields of atomic nuclei and electrons. These interactions, typically involving the exchange of virtual photons, result in multiple Coulomb scatterings that generate deflections of small angles that accumulate and can significantly deviate the particle from its original trajectory.

The scattering angle distribution is described by Molières theory [25]–[27]. The root mean square of the projected scattering-angle distribution can be calculated as

$$\sqrt{\langle \Theta^2 \rangle} = \frac{13.6\ \text{MeV}}{\beta c p} z \sqrt{\frac{x}{X_0}} \left[1 + 0.038 \ln \left(\frac{x}{X_0} \right) \right],$$

where:

- p is the particle momentum (in MeV/c).
- βc is the velocity of the particle.

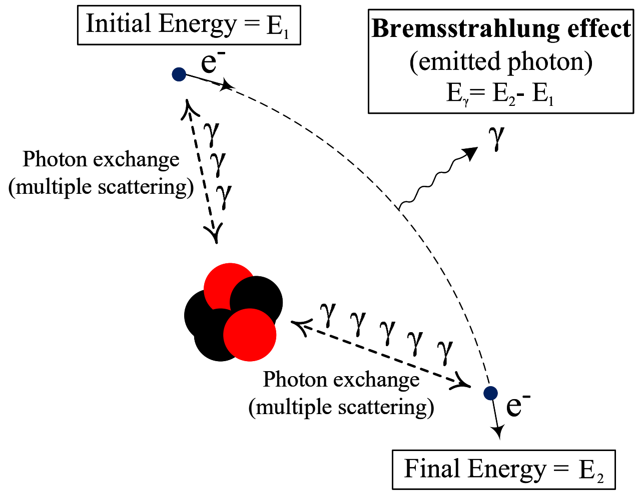


Figure 2.7: An electron undergoing multiple scattering decelerates due to the electric field of an atomic nucleus. As a result, the electron emits a Bremsstrahlung photon, reducing its energy.

- z is the charge of the particle in units of the elementary charge.
- x/X_0 is the thickness of the material traversed, in units of radiation length X_0.

The radiation length X_0 is a material-specific constant that characterizes the mean distance over which an incident electron loses $1/e$ of its energy via the Bremsstrahlung effect. It is calculated as:

$$X_0 = \frac{A}{4\alpha N_A Z^2 r_e^2 \ln\left(183/Z^{1/3}\right)} \ \mathrm{g/cm^2}.$$

2.1.2.3 Bremsstrahlung Effect

When an incident particle interacts with the nucleus of an absorber material via multiple scattering, it can be decelerated. Due to this speed change, the incident charged particle might generate a photon. This effect is known as the Bremsstrahlung effect, and it results in a reduction of the incident particle energy. The multiple scattering and Bremsstrahlung effect are shown in Figure 2.7. This effect is represented in the Feynman diagram of Figure 2.8.

To calculate the energy loss caused by the Bremsstrahlung effect, we use [28]:

$$-\frac{dE}{dx} \approx 4\alpha \cdot N_A \cdot \frac{Z^2}{A} \cdot z^2 \left(\frac{1}{4\pi\epsilon_0} \cdot \frac{e^2}{mc^2}\right)^2 \cdot E \ln\left(\frac{183}{Z^{1/3}}\right)$$

where:

- $\frac{dE}{dx}$ is the rate of energy loss per unit distance (Bremsstrahlung energy loss).
- Z, A is the atomic number and weight of the absorber material.

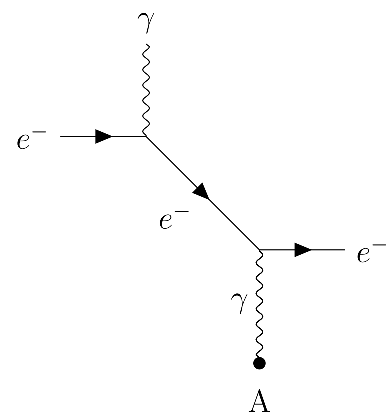


Figure 2.8: Feynman diagram of an electron scattering off an atom with a Bremsstrahlung photon emission.

- z, m, E is the charge, mass, and energy of the incident particle.
- ϵ_0 is the permittivity of the vacuum ($\epsilon_0 \approx 8.854 \times 10^{-12}\,\mathrm{F/m}$).
- α is the fine-structure constant. It is a fundamental constant in physics that quantifies the strength of the electromagnetic interaction between elementary particles with charge. It is defined as:

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137},$$

  where $\hbar$ is the reduced Planck constant.

Bremsstrahlung energy loss increases linearly with the particle's energy and is inversely proportional to the square of its mass. Consequently, electrons, as the lightest charged particles, are the most affected by this effect.

It should also be noted that Bremsstrahlung is primarily the result of interactions between incident particles and nuclei, but atomic electrons contribute by screening the nuclear charge. Taking both effects into account leads to a refined expression for the radiation length for electrons as:

$$X_0 = \frac{716.4 \cdot A}{Z\,(Z+1)\ln\left(289/\sqrt{Z}\right)}\ \mathrm{g/cm^2}.$$

2.1.2.4 Direct Electron-Positron Pair Production

As explained in Section 2.1.1, the pair production is the physics process in which two photons transform into an electron and a positron pair. At high energies, electron-positron pairs can be produced when a photon interacts with the virtual photons that surround a charged particle as

$$\gamma + X \rightarrow e^- + e^+ + X^*,$$

where X and X^* represent the ground and excited state of the incident charged particle.

This process is known as *trident production* and it can occur as well between the photons surrounding a charged particle like an electron, a muon, or an ion and the nucleus of a medium such as silicon or iron.

The probability of this process depends strongly on both the energy of the incoming charged particle and the atomic number (Z) and mass number (A) of the material. Due to their higher mass and reduced bremsstrahlung losses, muons or ions are more likely to undergo trident production at high energies compared to electrons [29].

2.1.2.5 Cherenkov Radiation

The speed of light in a vacuum is denoted by c. However, when light propagates through a dielectric medium such as water, glass, or air, its speed is reduced depending on the mediums refractive index n. The phase velocity of light in a medium is given by:

$$v = \frac{c}{n}$$

Cherenkov radiation occurs when a charged particle travels through a medium with a speed greater than the phase velocity of light in that medium. This results in the emission of photons at a characteristic angle relative to the particles trajectory. The Cherenkov angle is given by

$$\cos\theta_C = \frac{1}{\beta n}.$$

2.1.3 Hadronic Interaction

Hadrons such as protons, neutrons, and mesons can interact via the strong force. Analogous to the radiation length X_0, hadronic interactions are described by the *absorption length*, denoted by λ_a. This parameter quantifies the average distance a hadron travels in a given material before undergoing an inelastic interaction. The number of hadrons $N(x)$ that remain after traversing a thickness x of material is given by the exponential attenuation law [30] as

$$N(x) = N_0 e^{-x/\lambda_a},$$

where N_0 is the initial number of incident hadrons, and $N(x)$ is the number of hadrons that crossed x material material depth without interacting.

Hadronic interactions are predominantly inelastic, leading to the production of secondary particles such as pions or kaons. The inelastic hadronic cross-section σ_{inel} remains approximately constant for high-energy hadrons (around 45 mb per nucleon). Based on this, the absorption length can be estimated as:

$$\lambda_a = \frac{A}{N_A \rho \sigma_{inel}} \underset{\sim}{\propto} \mathcal{A}^{-\frac{2}{3}},$$

where:

- ρ is the density of the material,
- σ_{inel} is the inelastic cross-section,
- $\mathcal{A}$ is the mass number (number of nucleons).

For materials with a high atomic number (Z), the hadronic interaction length (λ_a) is typically much larger than the radiation length (X_0). This results in hadronic calorimeters needing to be significantly deeper than electromagnetic calorimeters to fully contain hadronic showers and measure the total energy deposition of incident hadrons.

2.1.4 Invisible Particles (Neutrinos)

Neutrinos are electrically neutral and do not carry color charge, meaning they do not interact via the electromagnetic or strong forces. Their only known interactions occur through the weak nuclear force. Neutrinos also have a very small mass, making their interaction cross-section with matter extraordinarily low. For example, a neutrino of moderate energy would require approximately a thousand light-years of lead to have a significant probability of interaction.

Due to this elusive nature, detecting neutrinos directly is extremely challenging. Experiments dedicated to neutrino physics often rely on extremely large detector volumes to maximize the chance of interaction. These typically consist of massive targets such as water or liquid argon, allowing for rare weak interactions to be observed. A notable example is the Deep Underground Neutrino Experiment (DUNE) [21], which is being constructed more than a kilometer underground at the Sanford Underground Research Facility in Lead, South Dakota. Its deep underground location minimizes background noise from cosmic rays, making it more sensitive to neutrino interactions.

In collider experiments such as those conducted at the LHC, neutrinos and other weakly interacting particles cannot be directly detected. However, their presence can be inferred indirectly through the phenomenon of *missing transverse momentum*, where the momentum of visible particles does not balance. This indicates that an invisible particle, such as a neutrino, has carried away energy and momentum. This concept is humorously captured in the classic "neutrino joke" shown in Figure 2.9.

2.2 Characteristics of Detectors

The performance of a detector is characterized by a set of measurable parameters that determine its ability to detect, localize, and measure particles [31]. Key characteristics include:

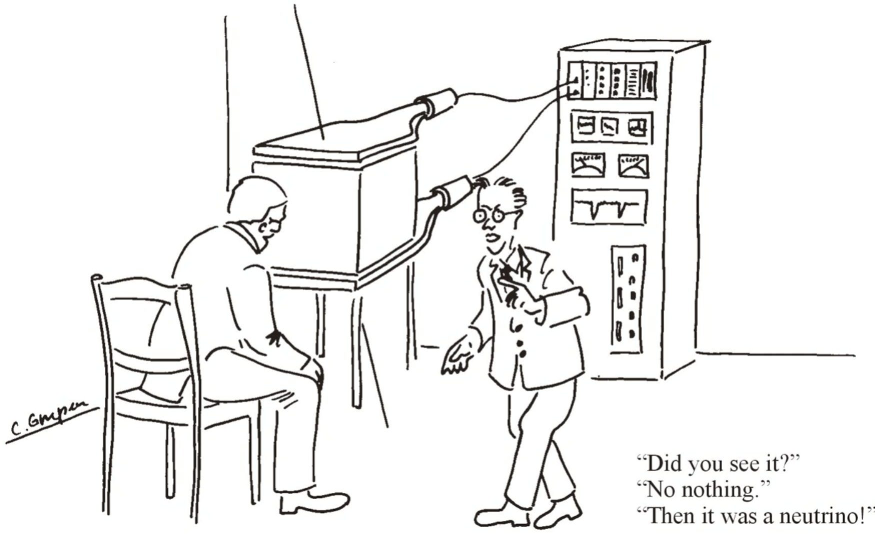


Figure 2.9: Neutrinos as invisible particles (Source [31]).

- **Resolution**: The resolution (σ) of a detector quantifies the smallest measurable difference that the detector can distinguish with confidence. For example, a spatial resolution of 10 μm in a tracking detector means that it can distinguish two particles whose trajectories are separated by at least 10 μm. High resolution is crucial for precise particle tracking and vertex reconstruction.
- **Efficiency**: The efficiency (ε) of a detector refers to the probability that a particle passing through the detector is successfully detected. Detector efficiency depends on the particle type and the detection technology. For instance, a gas detector may have an efficiency of only a few percent for detecting γ-rays, while a scintillator in a calorimeter may approach nearly 100% efficiency for detecting charged particles.
- **Dead Time (recovery time)**: The dead time (τ_D) is the interval after the detection of a particle during which the detector is unable to register new particles. This recovery time is needed for the detector to return to its sensitive state. Note that immediately following the dead time, the detector may still have reduced sensitivity until it fully recovers.
- **Dead Area**: The dead area refers to the fraction of a detector's physical area that is insensitive to particles. This can result from structural or electronic constraints, such as gaps between sensors or inactive regions needed for readout electronics. Minimizing dead areas is important for maximizing coverage and uniformity of detection.
- **Radiation Length**: The Radiation length (X_0) is a material-specific constant that characterizes the mean distance over which an incident electron loses $1/e$ of its energy via the Bremsstrahlung effect, as discussed in Section 2.1.2.2.
- **Time-Walk**: Time-walk refers to the delay between the moment a particle traverses the detector and the instant the front-end electronics are triggered, as illustrated in Figure 2.10. This effect is influenced by the amplitude of the detected signal and the

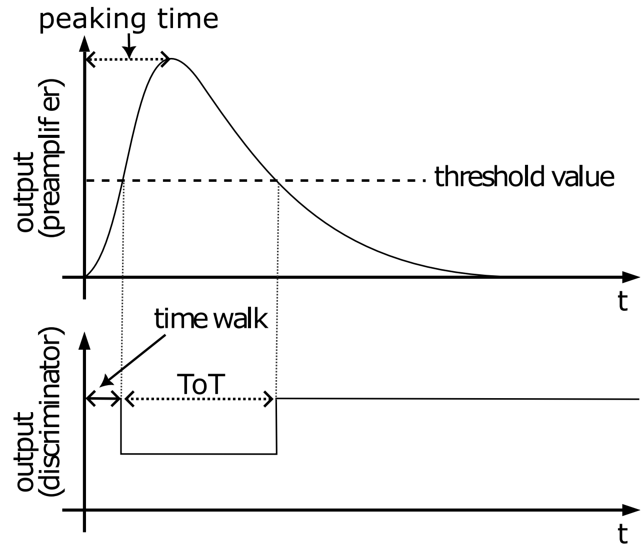


Figure 2.10: Schematic of a sensor output signal (top) and the corresponding processed signal after a discriminator (bottom). Three key timing parameters are illustrated: time-walk, time-over-threshold (ToT), and peaking time.

configured threshold.

- **Peaking Time**: The peaking time is the interval from the moment a particle interacts with the detector until the moment the signal reaches its maximum amplitude, as shown in Figure 2.10.
- **Time over Threshold (ToT)**: Time over Threshold (ToT) is the duration for which the output signal remains above a predefined threshold level. It corresponds to the time interval between the signal exceeding the threshold (assertion) and falling back below it (de-assertion), as illustrated in Figure 2.10.

2.3 Types of Detectors in High-Energy Physics

In high-energy physics experiments, such as those performed at the LHC, the particles produced in the collisions are measured by a collection of specialized detectors. To determine the momentum and electric charge of these particles, a strong magnetic field is applied. Within this field, the trajectories of charged particles are bent in opposite directions depending on the sign of their charge. In addition, the radius of curvature depends on the momentum and charge magnitude of the particle. This magnetic field is provided by a large superconducting solenoid located between the experiment sub-detectors.

Except for the LHCb, the other three major LHC experiments use a common detector layout consisting of concentric layers of sub-detectors. Each sub-detector is optimized to measure a specific property of the particles, such as their energy or trajectory. Figure 2.11 shows the layout of the CMS detector as an illustrative example of this structure.

As seen, the CMS detector consists of four main sub-detectors: the inner trackers, the

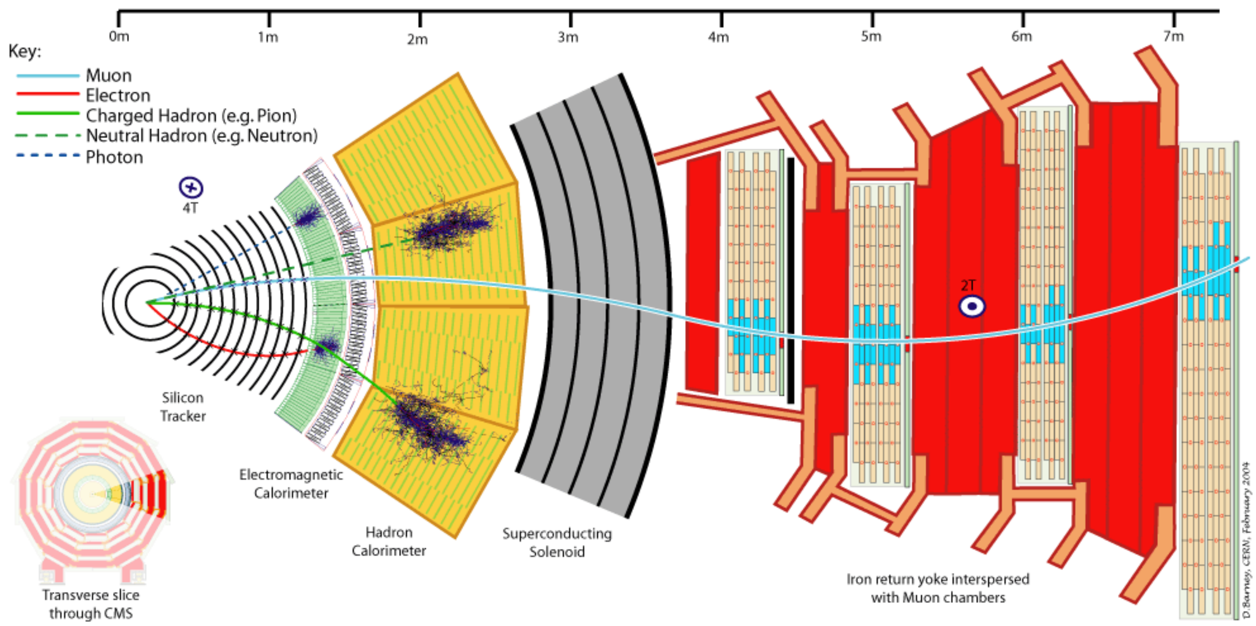


Figure 2.11: Sub-detectors that comprise the CMS detector. Particles that leave tracks in the trackers are depicted as continuous lines, while neutral particles that do not interact with the trackers are depicted as dashes (Source [32]).

electromagnetic calorimeter, the hadronic calorimeter, and the muon spectrometer.

2.3.1 Trackers

Trackers measure the trajectories of charged particles (e.g., electrons, muons, charged pions). This measurement can be performed by exploiting the electromagnetic interactions in Section 2.1.2, but avoiding deviating or decelerating the incident particle. Therefore, trackers must have a very low X_0.

As shown in Figure 2.11, trackers are located close to the interaction point to maximize spatial resolution for initial particle detection.

Historically, gaseous detectors such as cloud and bubble chambers were widely used for tracking. Modern experiments predominantly use silicon-based detectors, including Silicon Strip Detectors (SSD) and Silicon Pixel Detectors (SPD), that provide superior spatial resolution and radiation hardness. A detailed discussion of tracker semiconductor technologies is provided in Section 2.5.

2.3.2 Electromagnetic Calorimeters

Electromagnetic calorimeters (ECALs) measure the energy of electrons and photons. Like the tracking systems, they rely on electromagnetic interactions. However, in contrast to trackers, ECALs must have a large radiation length X_0 in order to fully absorb the particle energy. The energy loss of an electron with energy E can be described by

$$-\left(\frac{dE}{dx}\right)_{rad} = \frac{E}{X_0}.$$

When an electron or photon enters the absorber material (typically tungsten, copper, or lead), it initiates an electromagnetic shower. For example, after traversing one radiation length, a high-energy photon produces an e^--e^+ pair. Then, the electron and positron produce bremsstrahlung photons after another radiation length, which in turn convert into additional electron-positron pairs. The resulting low-energy charged particles ionize a material such as liquid argon, producing electrical signals collected by the readout electronics.

2.3.3 Hadronic Calorimeters

Hadronic calorimeters (HCALs) are designed to measure the energy of hadrons such as pions, protons, and neutrons by absorbing them and detecting the hadronic showers they produce. Unlike electromagnetic calorimeters, HCALs use hadronic and electromagnetic interactions to generate, absorb, and measure the energy of the incident hadrons.

HCALs are typically composed of dense absorber materials (e.g., steel or copper) interleaved with active detection media like plastic scintillators or liquid argon. These dense materials are selected for their ability to contain and fully absorb the energy of the hadrons, while the scintillators convert this energy into signals that can be read by some readout electronics.

2.3.4 Muon Spectrometers

Muons are minimally ionizing, heavy, charged particles that penetrate deeply through matter. Because of their weak interaction with both ECAL and HCAL materials, muon spectrometers are placed at the outermost layer of the detector, beyond the calorimeters, as shown in Figure 2.11.

These spectrometers consist of multiple layers of tracking detectors such as drift tubes, Resistive Plate Chambers (RPCs), and sometimes silicon detectors. These detectors are embedded within iron absorbers that filter out remaining non-muon particles. While their spatial resolution is generally lower than inner trackers, muon systems are crucial for identifying muons and measuring their momenta over long paths.

2.4 Track Reconstruction

Track reconstruction is a crucial component of data analysis in HEP experiments. Once the trajectories (tracks) of charged particles are reconstructed and their identities inferred, it becomes possible to determine their points of origin. This enables physicists to infer the presence of short-lived or undetectable particles that decay into detectable ones. Such methods have been instrumental in major discoveries, including those of the Z boson and the *Higgs* boson at CERN.

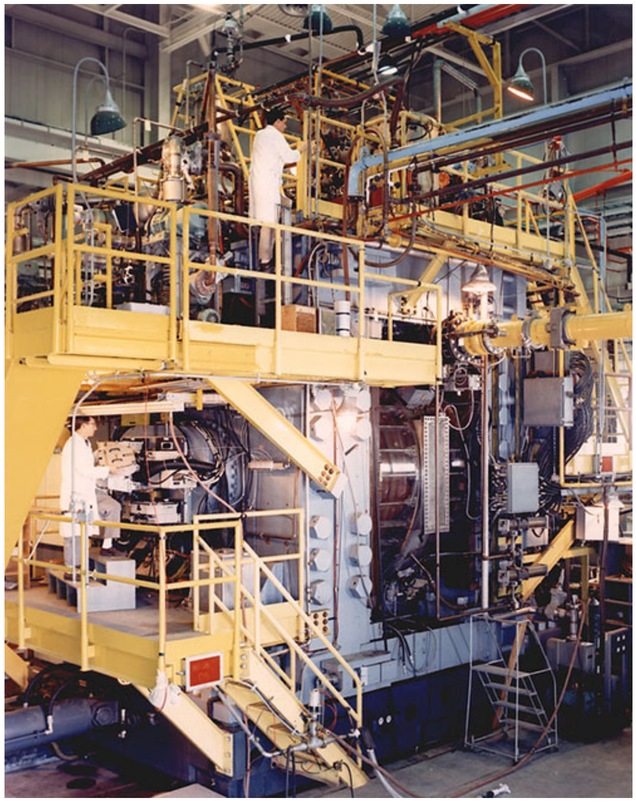


Figure 2.12: The BNL 80 liquid hydrogen bubble chamber at the Brookhaven National Laboratory (Source [33]).

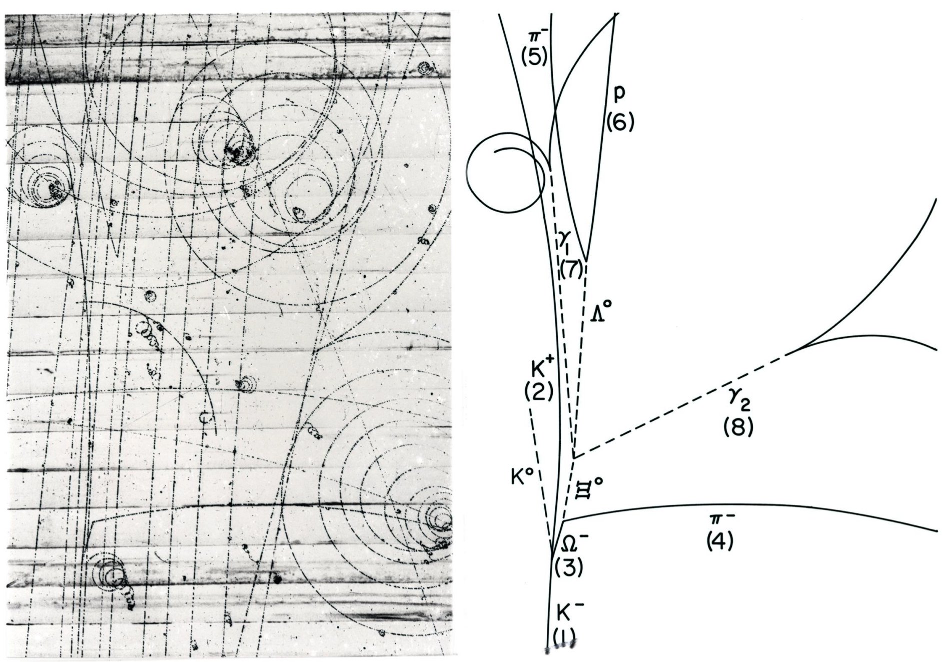


Figure 2.13: Discovery of the Ω^- baryon at the Brookhaven National Laboratory 80-inch liquid hydrogen bubble chamber in 1964. Left: the original photograph from the bubble chamber. Right: a schematic illustration of the reconstructed event (Source [34]).

As an illustrative example, we consider the discovery of the Ω^- (Omega-minus) baryon, which was first observed in 1964 at the Brookhaven National Laboratory (BNL) [34]. This observation was made using the BNL 80 liquid hydrogen bubble chamber depicted in Figure 2.12.

In Figure 2.13, the left image shows the original photograph of particle tracks captured in the bubble chamber, while the right image provides a schematic representation of the event for improved clarity. In the schematic, tracks of charged particles are shown as solid lines, while neutral undetectable particles are represented by dashed lines and are reconstructed from the kinematics of their decay products.

In this event, the collision between a K^- meson and a proton leads to the production of an Ω^- baryon, a K^0, and a K^+. The newly created Ω^- undergoes a weak decay, generating a π^- (pi-minus) meson and a Ξ^0 (Xi-zero) baryon. Following this, the Ξ^0 decays into a Λ^0 (Lambda-zero) and a π^0 (pi-zero) meson. The Λ^0 travels some distance until decaying into a proton and a π^-. The π_0 has a very short lifetime, decaying rapidly into two photons (γ), which ultimately create an electron-positron pair. This discovery was key to establishing the quark model, as it confirmed Murray Gell-Mann's 1961 predictions of the existence of a particle composed of three strange quarks (sss).

2.5 Semiconductor Trackers State of the Art

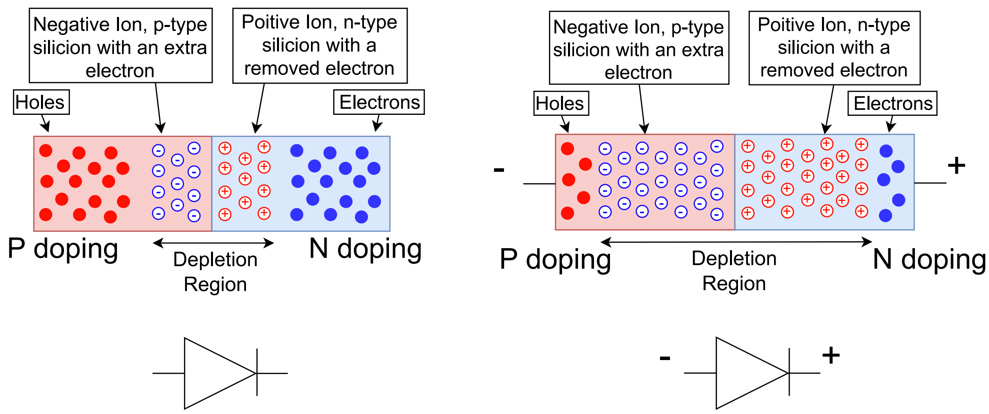
Since their invention in the 1980s, spatially sensitive semiconductor detectors have become widely used in tracking systems in particle physics experiments. Compared to the wire chambers, predominantly used before their introduction, semiconductor detectors improved the spatial resolution of particle tracking by up to two orders of magnitude. This advancement enabled the precise reconstruction of secondary vertices, which is necessary to measure the lifetimes of heavy fermions.

Several semiconductor materials can be used for tracking detectors, including silicon (Si), germanium (Ge), gallium arsenide (GaAs), and cadmium telluride (CdTe). Among these, silicon is by far the most widely used material in HEP experiments. This preference is due to several key advantages: silicon has a bandgap well-suited for operation at or near room temperature, it benefits from extensive use in the electronics industry (which reduces cost), and it can be manufactured with extremely high precision and uniformity using mature techniques.

2.5.1 Silicon Detectors Types and Basic Principles

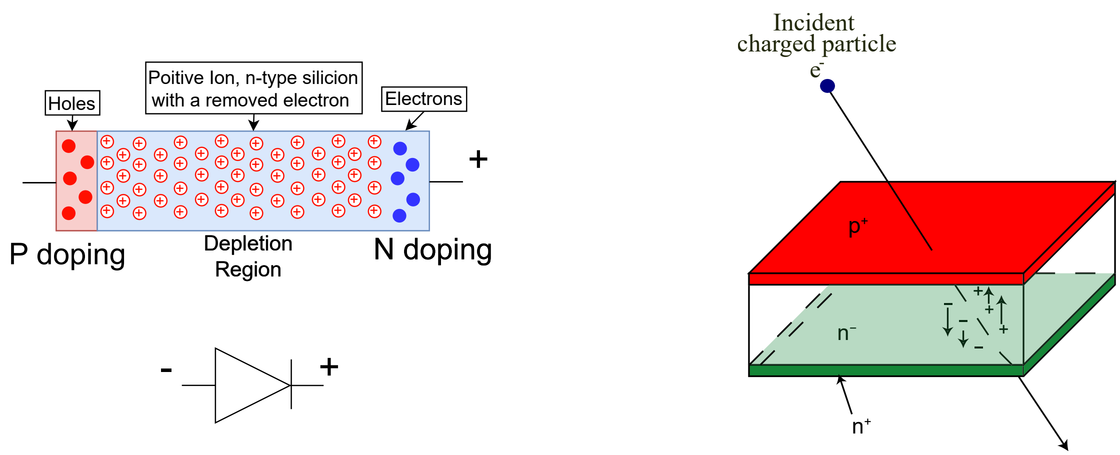
Silicon tracking detectors operate based on the principles of a P-N junction diode. In a typical diode (Figure 2.14(a)), the junction between p-type and n-type regions forms a depletion region devoid of free charge carriers, generating positive and negative ions depending on the type of implant present. Due to the diffusion of electrons and holes across the junction, this region naturally hosts an internal electric field.

In normal operation, a diode is forward biased (positive voltage on the p-side, negative on the n-side), which reduces the width of the depletion region and allows current to flow. However,



(a) (b)

Figure 2.14: P-N junction diode. a) A standard diode unconnected. b) A diode under reverse bias, showing a widened depletion region.



(a) (b)

Figure 2.15: Basic silicon detector concept. a) Depleted reverse-bias diode with a low-doped n-type material, increasing the depleted volume (Source [30]). b) Cross-sectional view of a typical planar silicon detector as a vertical diode. (Source [30]).

in a silicon detector, the diode is operated under reverse bias, applying a negative voltage to the p-side and a positive voltage to the n-side as shown in Figure 2.14(b). This increases the width of the depletion region and suppresses the flow of leakage current, making the device sensitive to ionizing radiation.

In practice, silicon detectors are designed to maximize the depleted volume, typically using low-doped n-type material as shown in Figure 2.15(a). When a charged particle passes through the depletion region, it ionizes the silicon atoms along its path, creating electron-hole pairs. The internal electric field quickly separates these charge carriers and drifts towards their respective electrodes, as illustrated in Figure 2.15(b).

This charge movement induces a current, which is detected by the readout electronics. The signal corresponds to the passage of a charged particle, and the spatial location of the interaction can be inferred from the geometry of the readout electrodes.

2.5.1.1 Hybrid Pixel Detectors

Hybrid pixel detectors [35] are a type of pixel detector that consists of two separate components, as illustrated in Figure 2.16(a). The first component is the pixel sensor, typically made of a silicon diode segmented into an array of pixel cells. The second component is the readout chip, which processes the signals generated by particle hits in the pixel sensor. Each pixel sensor is connected to its corresponding pixel readout chip via a conductive microconnection, typically a bump bond. These pixel pairs are arranged in matrices to form the detector structure, as shown in Figure 2.16(b).

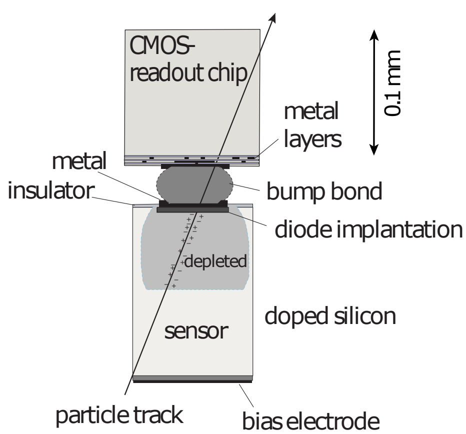
Prominent examples of hybrid pixel detectors include the ATLAS Inner Tracker (ITk) [36], [37], the LHCb VELOPix [38], and the CMS Outer Tracker [39].

2.5.1.2 Monolithic Pixel Detectors

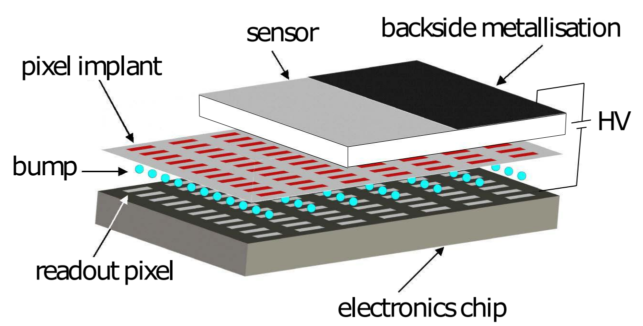
Monolithic pixel detectors [40] integrate both the sensor and the readout electronics within a single piece of silicon. This architecture has long been dominant in commercial imaging technologies. In recent years, it has garnered increasing interest in HEP due to several advantages: significantly lower cost, reduced power consumption, minimal material budget (lower radiation length), and simplified assembly processes.

Despite these benefits, most HEP experiments still rely on hybrid detectors because they offer superior signal-to-noise ratio, greater radiation tolerance, and the ability to operate under high particle flux conditions.

There are several types of monolithic pixel detectors. Among them, the most relevant for



(a)



(b)

Figure 2.16: (a) Schematic of a hybrid pixel detector cell, comprising a pixel sensor and readout electronics connected via a bump bond. (b) Matrix of hybrid pixels, each subdivided into individual cells. For illustration purposes, pixel implants are shown separated from the sensor (Source [30]).

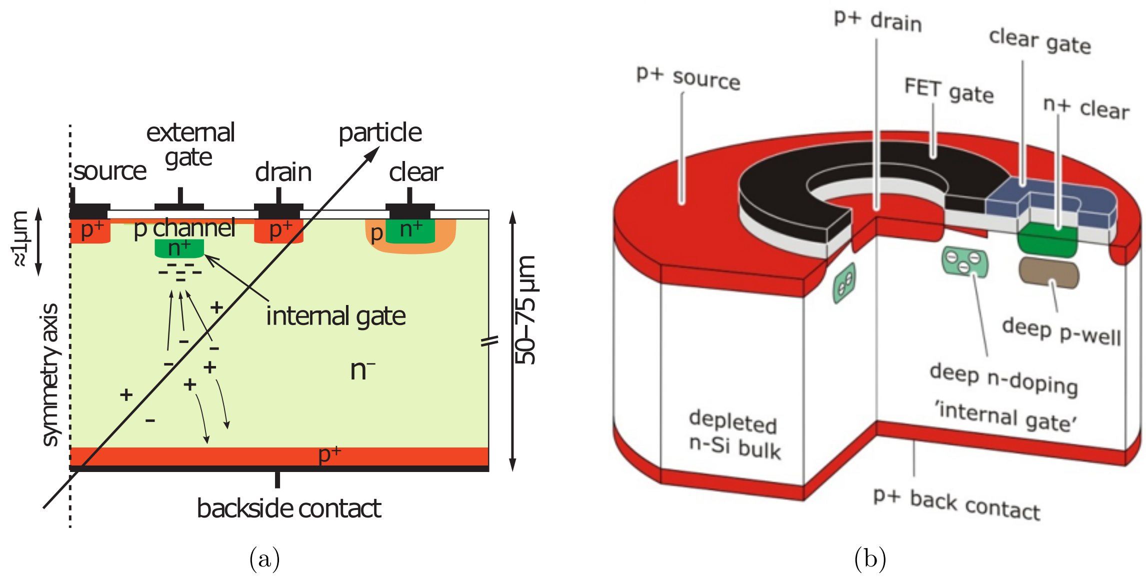


Figure 2.17: DEPFET pixel: (a) Two-dimensional cross section and charge collection. (b) three-dimensional view (Source [30]).

HEP applications are the Depleted P-channel Field Effect Transistor (DEPFET) [41] and the Monolithic Active Pixel Sensors (MAPS) [42].

A schematic cross-section of a DEPFET sensor is shown in Figure 2.17(a) and 2.17(b). As seen, it consists of a p-channel transistor with a deep n-implant below. This implant can capture the ionized electrons generated by a crossing charged particle, modulating the channel depending on how many electrons are trapped. To release these electrons, a positive voltage can be applied to the *clear* contact.

DEPFET detectors feature extremely low detector capacitance, which enables excellent signal-to-noise performance and high energy resolution. However, DEPFET technology requires specialized, non-commercial, and non-CMOS fabrication processes, which result in higher production costs and reduced flexibility compared to CMOS-based approaches. An example of DEPFET detector is the Belle-II [43] vertex detector, located at the e^+ e^- collider in KEK (Tsukuba, Japan).

MAPS detectors, by contrast, utilize the same commercial CMOS imaging technology as the ones used in digital cameras. In MAPS, the sensor consists of a p-type epitaxial layer situated beneath NMOS and PMOS transistors (see Figure 2.18). When a charged particle traverses the detector, the resulting charge carriers diffuse through the epitaxial layer and are collected by an n-well collection electrode. To isolate PMOS transistors hosted in n-well doping layers, a deep p-well layer is introduced, preventing unintended charge collection and interference.

Examples of MAPS detectors are the Heavy Flavor Tracker (HFT) in the STAR experiment [44], [45] or the ALPIDE used in the Inner Tracker System 2 (ITS2) [46], [47] in the ALICE experiment.

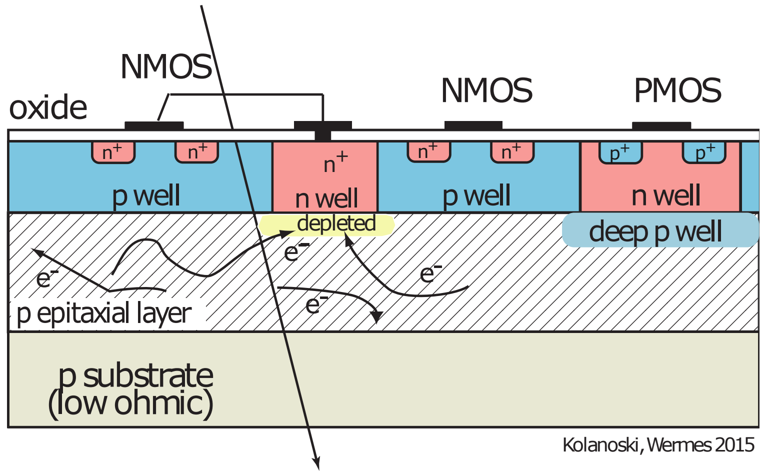


Figure 2.18: Schematic of a monolithic detector with charge collection occurring within the epitaxial layer of the silicon. NMOS and PMOS transistors are embedded in the electronics layer, while an n-well collection electrode gathers the signal. Note that the n-well hosting the PMOS transistors is electrically shielded from the epitaxial layer by a deep p-well layer (Source [30]).

2.5.2 The ITS2

The ITS2 was an upgrade of the ALICE Inner Tracking System carried out during the Long Shutdown 2 of the LHC between 2019 and 2021. It consists of 192 staves forming an active silicon area of 10 m^2, containing approximately 12.5×10^9 pixels. The layout of the detector is shown in Figure 2.19, with radii ranging from 23 mm to 400 mm. The three innermost layers constitute the Inner Barrel (IB), while the remaining layers are grouped into the Outer Barrel (OB).

This upgrade pursued three major improvements:

- Improve the impact parameter by a factor of 3 and 5 in the $r\varphi$ and z coordinates respectively. This results from the combination of several factors, such as placing the detector closer to the interaction point (from 39 mm to 23 mm), reducing the material budget of the inner layers (from 1.14 X_0 to 0.35 X_0) and the reduction of the pixel size (from 50 μm $\times$ 425 μm to 29.24 μm $\times$ 26.88 μm).
- Improve the tracking efficiency and transverse momentum resolution in the low p_t range (low momentum particles in the $r\varphi$ direction). This is achieved by increasing the number of layers from 6 to 7.
- Increase the readout rate for Pb-Pb interactions from 1 kHz to 100 kHz.

The core sensing element of the ITS2 is the ALPIDE chip [48]–[51], a MAPS sensor produced using a 180 nm CMOS imaging process with a high-resistivity (>1 kΩ·cm) epitaxial layer substrate of thickness ranging between 18-30 μm.

Figure 2.20 illustrates the ALPIDE pixel cross-section. As in other monolithic designs (see Section 2.5.1.2), a deep p-well enables the integration of complex CMOS logic by isolating

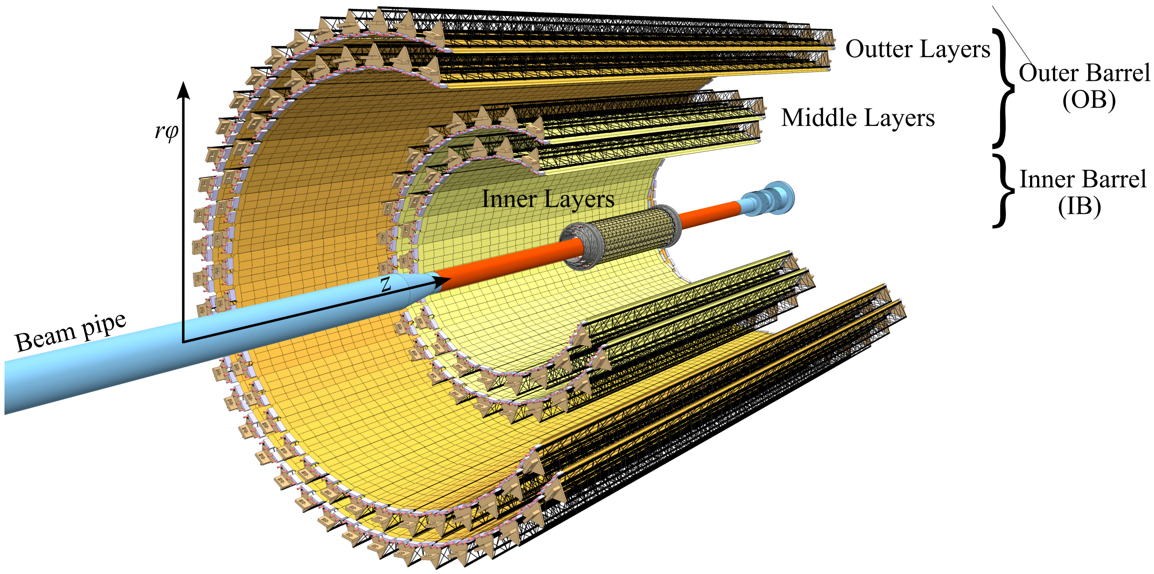


Figure 2.19: Layout of the ALICE ITS2. The concentric detector radii range from 23 mm to 400 mm (Source [47]).

PMOS transistors from the charge-collecting epitaxial layer.

Figure 2.21 shows a schematic block diagram of the ALPIDE in-pixel logic. As seen, each pixel contains a sensing diode, a front-end amplifier, a shaping stage, a discriminator, and a digital section with three hit-storage latches. A test charge can be injected via a pulse injection capacitor for calibration. The front-end and discriminator are always active. The front-end peaks at around 2 μs, and the discriminator output pulse (ToT) lasts approximately 5-10 μs. A global `STROBE` signal enables the value of the discriminator output to be written into one of the three storage latches. By propagating the `STROBE` at the same time to all pixels of the matrix, it creates a snapshot of the matrix status.

The `STROBE` signal can operate in two modes: trigger and continuous. In trigger mode, the `STROBE` is generated by the ALICE Central Trigger Processor [47], in synchronization with the physical events. In continuous mode, `STROBE` pulses are generated periodically, independently

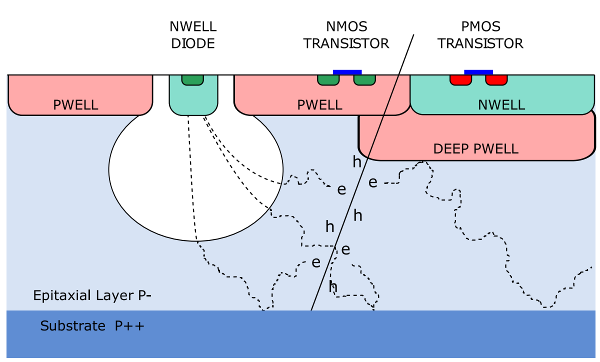


Figure 2.20: Schematic cross-section of the ALPIDE well structure, showing charge collection from an incident particle (Source [47]).

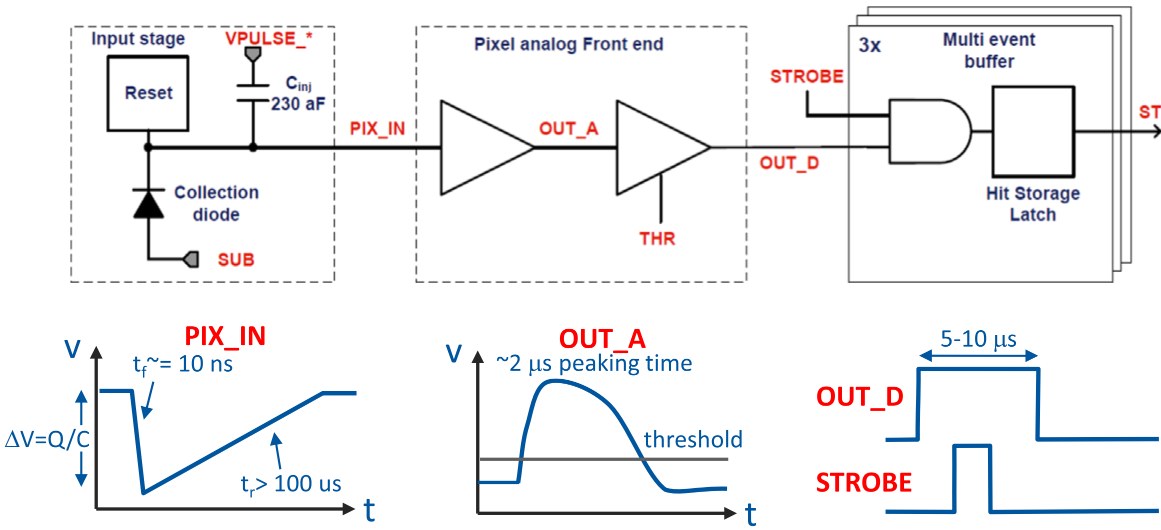
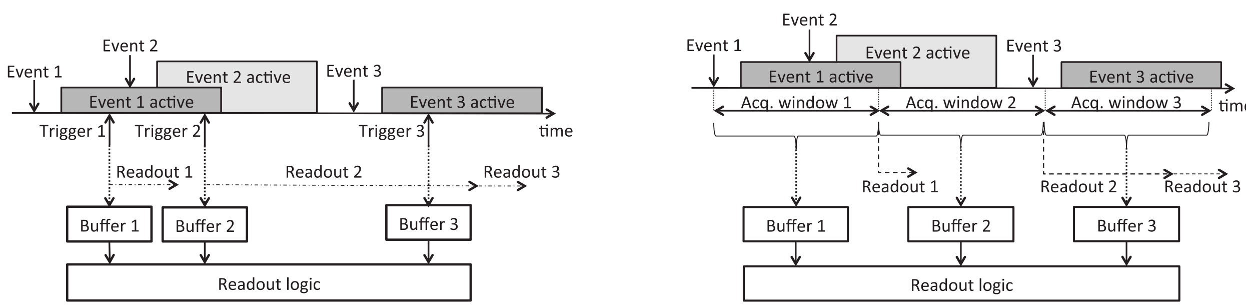


Figure 2.21: Block diagram of an ALPIDE pixel. A charged particle generates a current in the epitaxial layer which is collected and processed by the analog front end. The analog front end outputs a binary signal pulse of a duration between 5 and 10 μs. This hit can be stored in one of the three storing latches when the `STROBE` signal is high (Source [48]).



(a) Trigger Mode (b) Continuous Mode

Figure 2.22: ALPIDE `STROBE` operation modes: (a) Trigger mode with event-based `STROBE` pulses from the ALICE Central Trigger Processor; (b) Continuous mode with periodic `STROBE` pulses. [50]

of the physical events. Figure 2.22 shows both `STROBE` operation modes. As seen, hits are level-sensitive and can be stored in several storage latches if the `STROBE` is activated multiple times while the discriminator output is high.

The ALPIDE chip measures 15 mm × 30 mm and includes a matrix of 512×1024 pixels, shown in Figure 2.23. Pixels are grouped into double columns, with each double column having a zero-suppression fully combinatorial readout circuitry that encodes the pixel address, allowing the sequential read of the pixels.

As shown in Figure 2.24, double columns are grouped into four regions, each read out into a dedicated 128×24-bit Dual Port Random Access Memory (DPRAM). DPRAMs are read out via a parallel interface (4×80 Mbps) or a configurable serial link at 1200 or 400 Mbps.

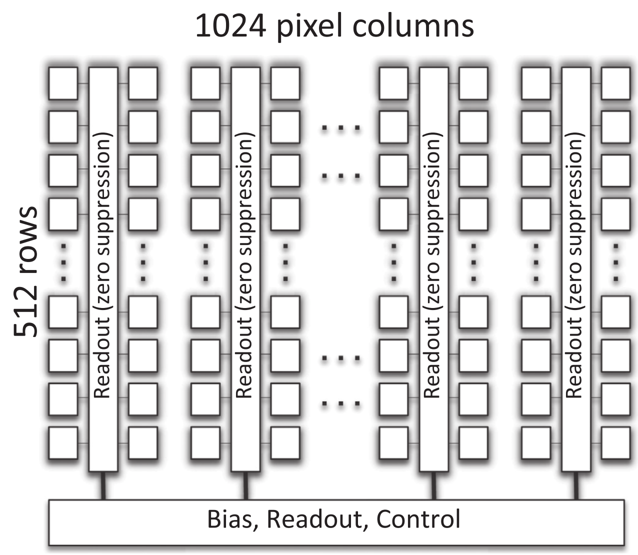


Figure 2.23: Architecture of the ALPIDE pixel matrix (Source [48]).

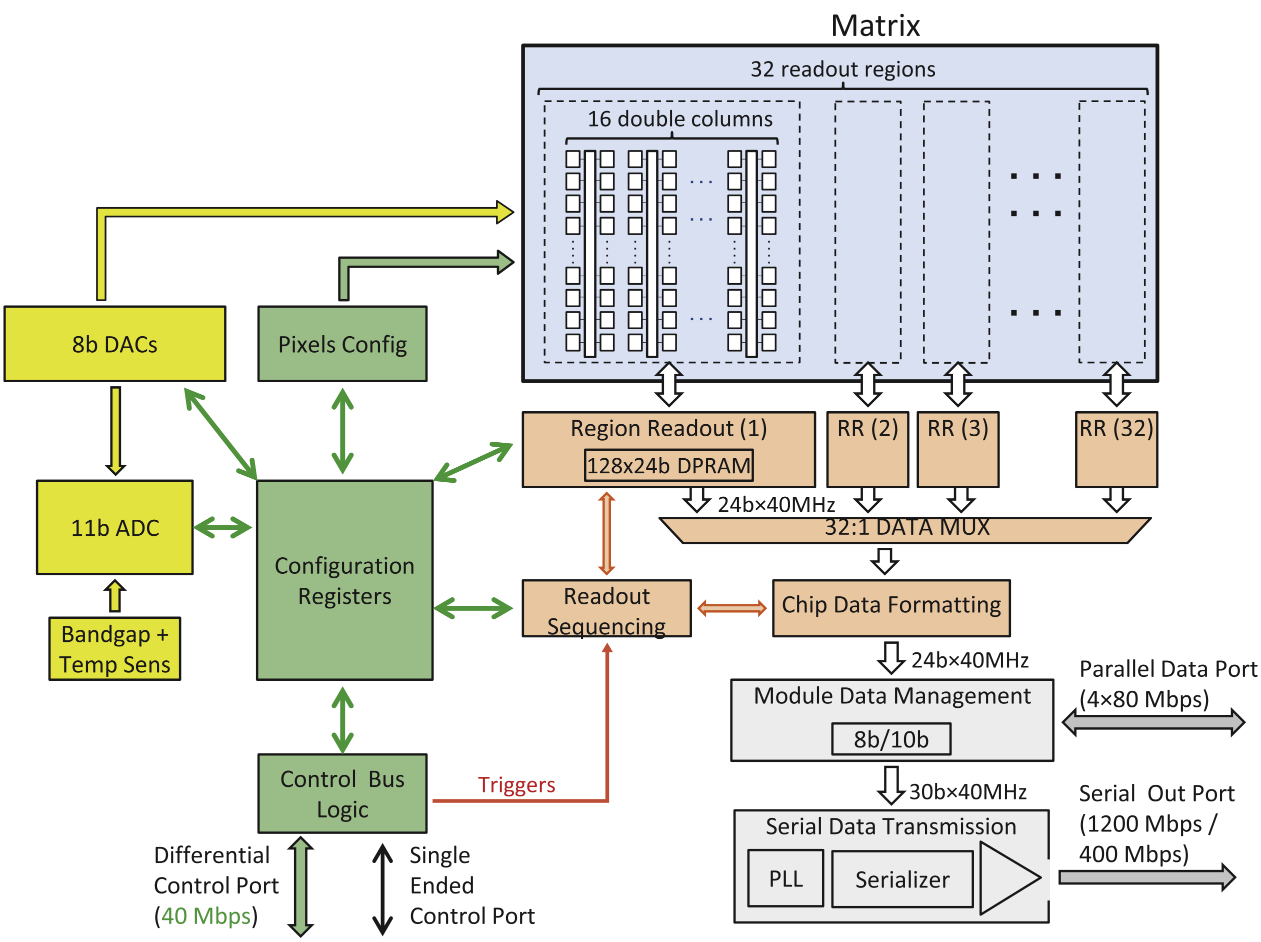


Figure 2.24: ALPIDE block diagram. The pixel matrix is read in parallel into four DPRAMs, then serialized and sent off-chip (Source [48]).

As the ITS2 is the predecessor of the ITS3, many of the ITS2 configurations, logic and operation modes were updated and ported into the ITS3. Its in-pixel logic is similar, but with some improvements and with two memory flip-flops instead of three latches. The ITS2 `STROBE` trigger mode was proven unnecessary in the real ALICE operation. Therefore, only the continuous mode was ported to the ITS3. The matrix architecture readout based on double columns and zero-suppressed fully combinatorial logic was also used as a base for the ITS3 pixel matrix readout. More information on the pixel and matrix logic of the ITS3 can be found in Section 5.

2.5.3 Stitched Chips State of the Art

Stitched chips enable die sizes that exceed the limit of the lithography reticle. By dividing a design into multiple sub-reticles and precisely aligning their exposures on the wafer, manufacturers can create a single monolithic chip spanning areas far larger than what a single photolithographic field permits. This technique is now used in several advanced commercial and scientific applications.

One of the most prominent examples is the Wafer Scale Engine 2 (WSE–2) [52] developed by Cerebras [53]. The WSE–2 is a wafer-scale AI accelerator fabricated on a 300 mm wafer using a 7 nm process. It reaches a total silicon area of 46,225 mm^2 and integrates more than 850,000 AI-optimized compute cores. Thanks to its unprecedented chip scale, the WSE–2 provides extremely high on-chip memory capacity and bandwidth, enabling neural network models to run at significantly higher performance compared to GPU-based clusters. Its architecture directly exploits the benefits of stitching, as the network-on-chip interconnect extends seamlessly across the reticle boundaries.

Beyond AI acceleration, stitched wafer-scale chips are also widely adopted in imaging applications, particularly for X-ray detection in the medical domain. Here, large-area sensors are essential for capturing wide diagnostic fields without requiring tiling or stitching at the module level. Representative examples include the wafer-scale CMOS X-ray detector proposed by Kim *et al.* [54] and the Sagara1212 sensor developed by IMASENIC [55]. These devices rely on stitching to form contiguous large-area pixel matrices suitable for high-resolution, low-noise medical imaging.

While stitching is now an established technique in medical imaging, it has not been previously employed in HEP detectors. The ITS3 project constitutes the first use of a stitched wafer-scale chip in HEP, marking a significant departure from traditional sensor tiling approaches and opening the door to future large-area, ultra-thin detectors.

2.5.4 Effects of Bending on Silicon Chips

Ultra-thin silicon chips (with a thickness below 50 μm) exhibit high flexibility and mechanical compliance, allowing them to be bent without breaking [56]. Such bendable chips have been explored in a range of applications, from medical devices [57], [58] to energy harvesting [59]. However, their use in high-energy physics (HEP) experiments has remained unexplored. The ITS3 project is therefore pioneering in the HEP field, introducing curved MAPS sensors with bending radii as small as 19 mm.

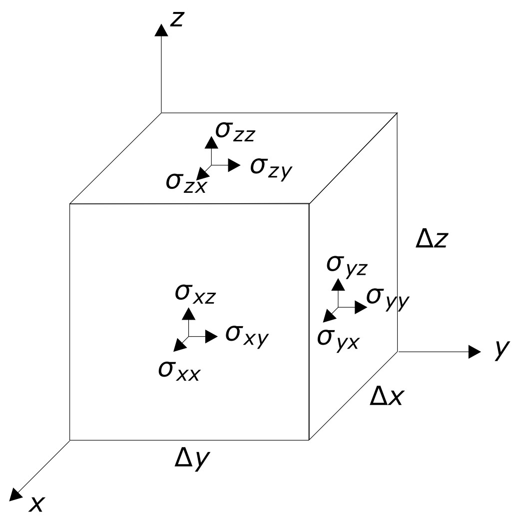


Figure 2.25: Schematic illustration of a differential volume with side lengths Δx, Δy and Δz under normal (σ_{ii}) and shear (σ_{ij}) stresses (Source [60]).

Bending a modern semiconductor chip induces mechanical stress and strain in the various layers of the device, altering its electrical properties. Before the design of the ITS3, several bent tests were performed, proving the viability of bent MAPS and providing the designer with a series of points to consider for the final design.

In this subsection, a general introduction to the electrical effects experienced by silicon devices is presented, followed by a review of the bending tests performed previously to the development of the ITS3.

2.5.4.1 Introduction to Material Stress

The stress of a material section is the force per unit area experienced by that material section. A schematic representation of the stress that a differential volume element is experiencing on each of its surfaces can be seen in Figure 2.25. To mathematically represent these stresses, the so-called stress tensor is used:

The stress in a material is defined as the force per unit area acting on a given surface. A schematic representation of the stresses acting on a differential volume element is shown in Figure 2.25. The state of stress is described mathematically by the stress tensor as:

$$\sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{bmatrix}.$$

In it, the diagonal components σ_{xx}, σ_{yy} and σ_{zz} represent the stresses orthogonal to the different axis and receive the name of normal stresses, while the off-diagonal components σ_{ij} $(i \neq j)$ represent the stresses acting along the different surfaces, denominated shear stresses. Because the stress tensor is symmetric ($\sigma_{ij} = \sigma_{ji}$), shear stresses are invariant under coordinate rotations.

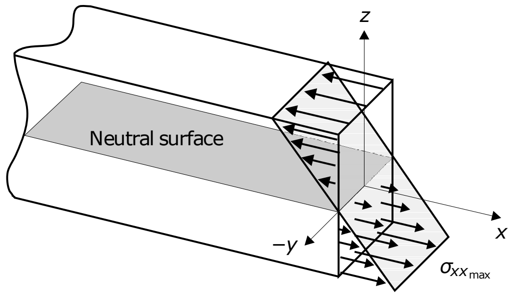


Figure 2.26: Schematic illustration of the stress distribution in a material subjected to pure bending. The component σ_{xx} takes negative or positive values depending on whether the material is under compression or tension (Source [60]).

Despite the complexity of the stress tensor, in most scenarios, some of the tensor components can be simplified. In the case of pure bending (illustrated in Figure 2.26), all components of the stress tensor vanish except σ_{xx}, which becomes negative in compression and positive in tension.

When a stress is applied to a crystal lattice, it can generate a strain deformation ($\epsilon = \Delta l/l$). Stress and strain are related through the material compliance tensor S as:

$$\epsilon_{ij} = S_{ijkl}\sigma_{kl}.$$

2.5.4.2 Strain Effects on Electrical Resistance

The electrical resistance of a uniform conductor depends on its resistivity ρ, length l and cross-sectional area A as $R = \rho l/A$. When the conductor is strained, its resistance changes according to

$$\frac{\Delta R}{R} = \frac{1}{R}\left(\frac{\partial R}{\partial l}\Delta l + \frac{\partial R}{\partial A}\Delta A + \frac{\partial R}{\partial \rho}\Delta\rho\right) = \frac{\Delta l}{\partial l} - \frac{\Delta A}{\partial A} + \frac{\Delta\rho}{\partial\rho} = (1+2\nu)\,\epsilon + \frac{\Delta\rho}{\rho},$$

where ν is the Poisson ratio, a material constant that describes the transverse contraction or expansion of the volume under longitudinal loading.

For metallic conductors, the relative change in resistivity $\Delta\rho/\rho$ is typically negligible, resulting in a resistance change equivalent to $\Delta R/R = (1+2\nu)\epsilon$.

However, in silicon, $\Delta\rho/\rho$ is strongly strain-dependent due to the piezoresistive effect. The strain modifies the interatomic spacings, thereby altering the structure of the semiconductor band and the effective mass of the charge carriers (m^*). Furthermore, because silicon is a crystalline material, the magnitude and sign of the piezoresistive response depend on the direction of the applied strain.

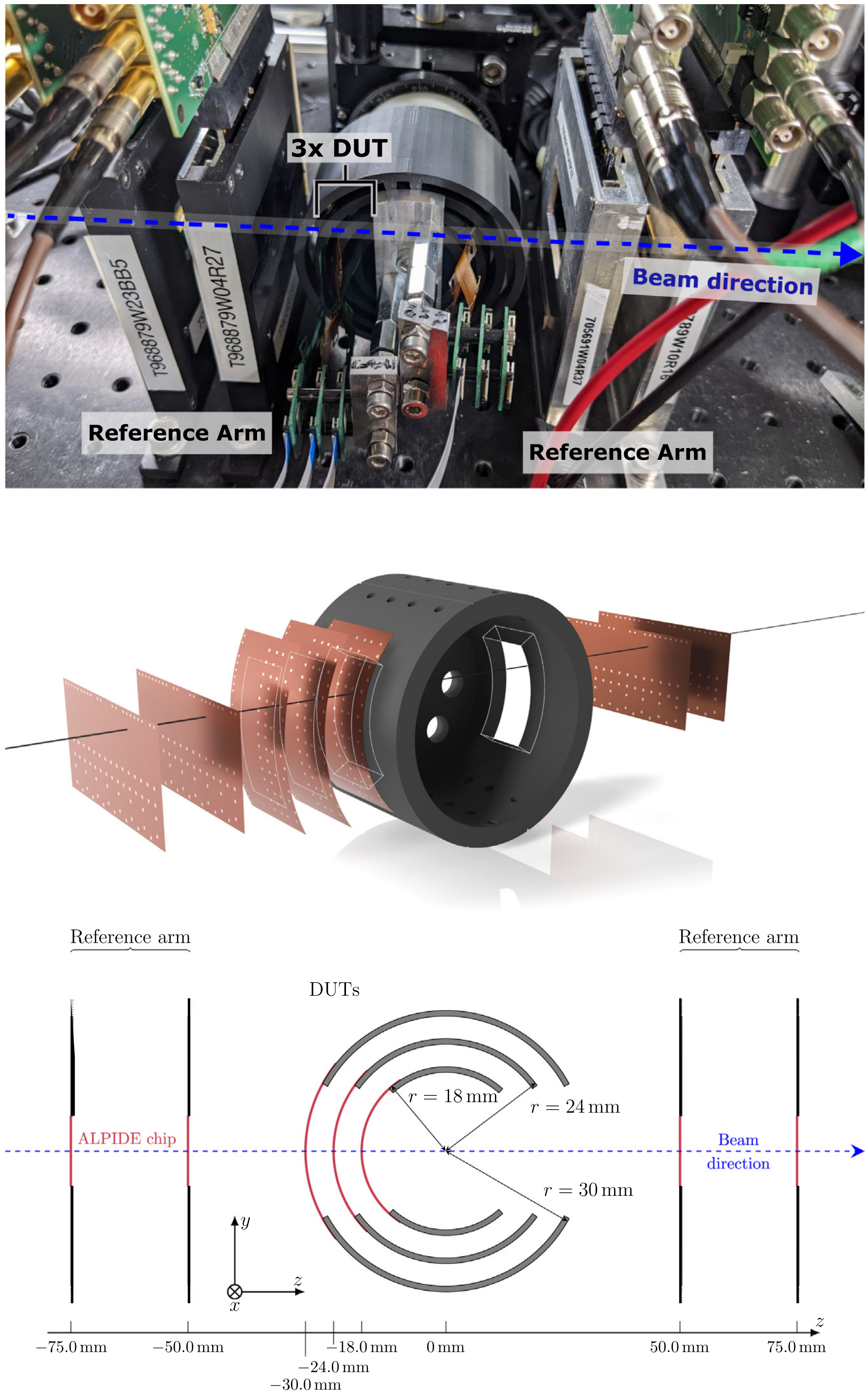


Figure 2.27: (Top) Photograph of the bent ALPIDE beam test setup. (Center) 3D rendering of the setup showing three bent ALPIDE chips and four flat ALPIDE chips used as a reference. (Bottom) Schematic of the experiment setup geometry (Source [14]).

2.5.4.3 ITS3 Preparation Bending Tests

As a proof of concept prior to bending ITS3 sensors, several beam-test campaigns were conducted using bent ALPIDE chips [13], [14] with bending radii of 18, 24, and 30 mm. These campaigns were conducted using beams from the Super Proton Synchrotron (SPS) at CERN [61] and the Deutsches Elektronen-Synchrotron (DESY) facility in Hamburg [62]. The setup of the DESY beam-test campaign is shown in Figure 2.27. As seen, the setup consists of bent ALPIDE chips acting as Devices Under Test (DUTs), together with four flat ALPIDE reference chips used to correlate and validate the measured particle tracks.

These studies produced the inefficiency curves shown in Figure 2.28. For the nominal ALPIDE threshold (100 e^-) and below, the detection efficiency exceeded 99.9%, independent of both particle incident angle and position on the chip surface. At higher thresholds, the efficiency increases with the incident angle due to the larger energy deposition at oblique incidence.

These beam tests demonstrated that bending does not cause any noticeable degradation in detection performance, confirming the feasibility of the ITS3 concept and motivating the further development of bent detectors.

However, during these tests, it was observed that the power-supply current (I_a) changed far more than expected from the small mechanical strain applied ($\sim 0.1\%$), reaching variations on the order of +10% or −5% depending on the bending direction [15].

To further investigate this, a second analysis was performed monitoring the power supply current consumed by the analog and digital parts of the chip under different strains and bent through the long and short chip edges, as plotted in Figure 2.29. As seen, the power

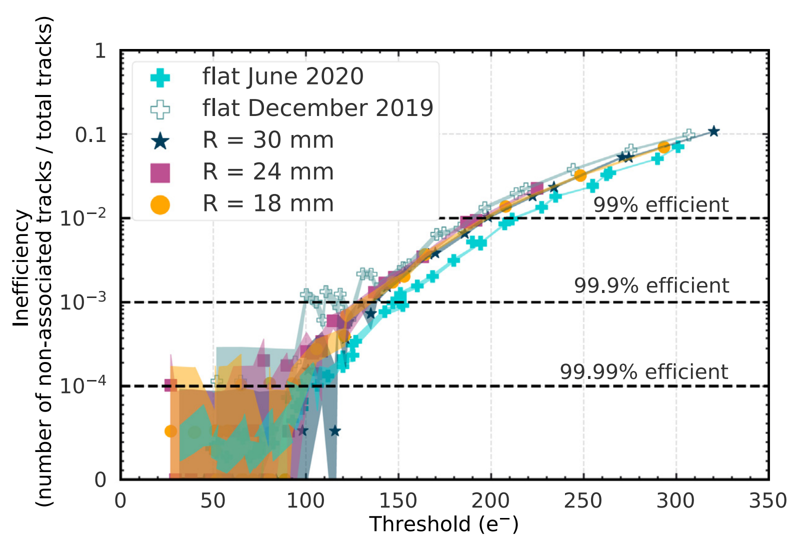


Figure 2.28: Inefficiency of bent ALPIDE sensors as a function of threshold for three bending radii, compared with two flat sensors. The nominal ALPIDE threshold is between 100 and 150 e^-. The colored bands represent statistical uncertainties. For better illustration, the y-axis is shown with a linear scale for inefficiencies below 10^{-4} and logarithmic scale when above 10^{-4} (Source [14]).

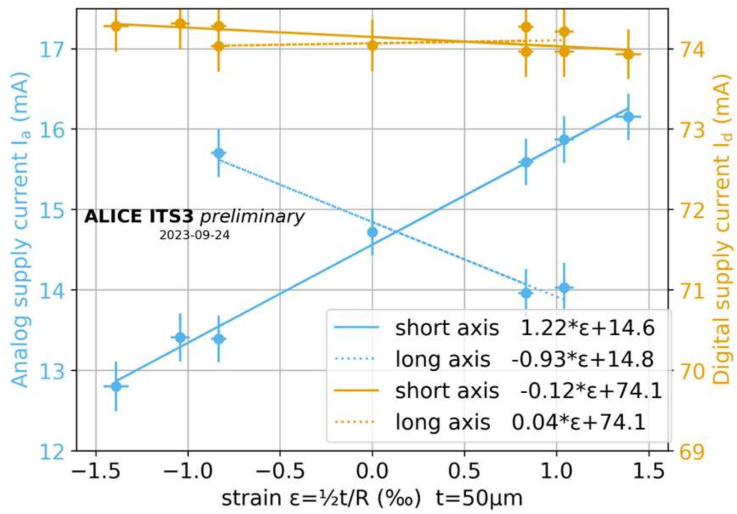


Figure 2.29: Analog and digital power-supply currents of ALPIDE chips bent along their short and long edges as a function of strain (ϵ). Positive and negative ϵ correspond to tensile and compressive strains at the chip top surface (where the circuitry is located) (Source [15]).

consumption of the digital devices experiences negligible changes, while the analog components experience a drastic increase or decrease depending on the direction in which it is bent.

Further investigation traced this behavior to the pixel biasing circuit shown in Figure 2.30. Here, the current mirror transistors highlighted in a green box (M_i and M_0) were laid out with perpendicular orientations with respect to each other. As a result, the strain applied parallel to the gate of M_i acts perpendicular to the gate of M_0, causing a mismatch in their piezoresistive response and leading to significant current variations.

This study confirmed that standard-cell digital libraries exhibit minimal strain sensitivity when aligned correctly, but also highlighted the need for careful analog design in bent detectors. In particular, transistors should be oriented consistently to ensure that strain affects them uniformly.

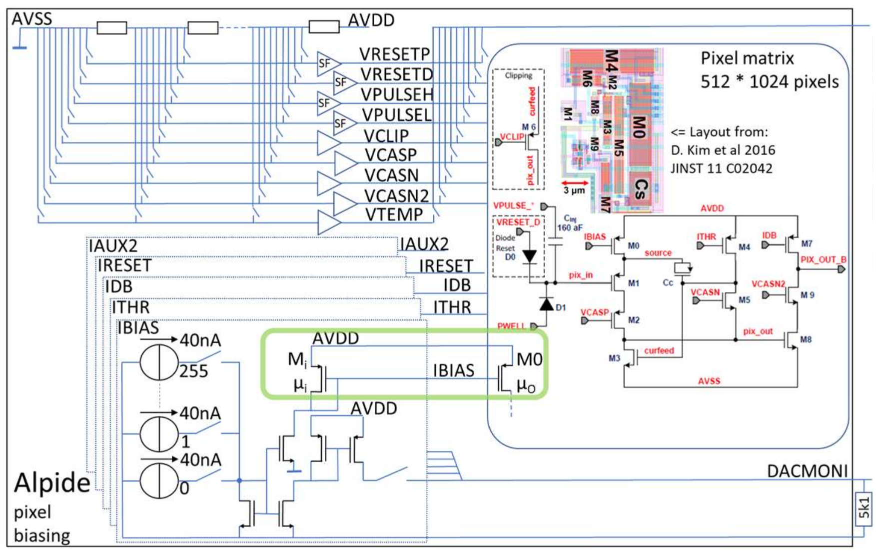


Figure 2.30: ALPIDE pixel biasing circuit. All pixels of the chip share the same biasing. Inside the green box, the current mirror composed of transistors M_i and M_0 is highlighted. These two transistors were designed in different orientations, which led to different $\Delta R/R$ among them under strain. (Source [15]).

Chapter 3

ITS3 Pixel Hit Sources Model

The ITS3 readout architecture operates similarly to the ITS2 continuous STROBE mode (see Section 2.5.2). In that mode, pixel hits detected within a configurable integration period are grouped and transmitted together in time-stamp packets. When the number of pixel hits detected in one or several consecutive integration periods becomes too large, the system may not be able to transmit all hits in time, producing data losses. Therefore, balancing available readout resources (e.g., memories, data links) with the expected pixel hit load is essential to ensure optimal performance at minimal cost.

As illustrated in Figure 3.1, two independent models were developed to emulate and dimension the ITS3 readout resources: the pixel hit source model and the readout model. The pixel hit source model (implemented in Python) emulates the secondary particles generated by the Pb-Pb interactions in the ALICE cavern, calculating the pixel hits in the IS3 over time. The readout model (implemented in SystemVerilog) takes as input the calculated pixel hits and tries to read them all, evaluating the performance of a candidate readout architecture by estimating the pixel-hit losses.

In this chapter, the Python-based pixel hit source model will be presented. The discussion starts with an analysis of the interactions that the lead ions can perform, followed by the Pb-Pb operation parameters at the LHC. After it, the average particle and hit densities of the ITS3 are presented. Finally, these ingredients will be combined in the Python model, which produces random pixel hits per integration period.

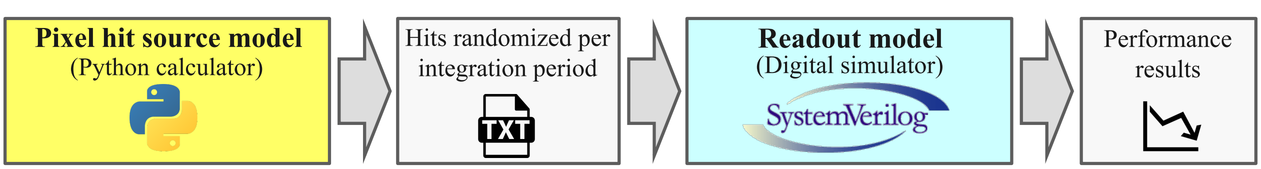
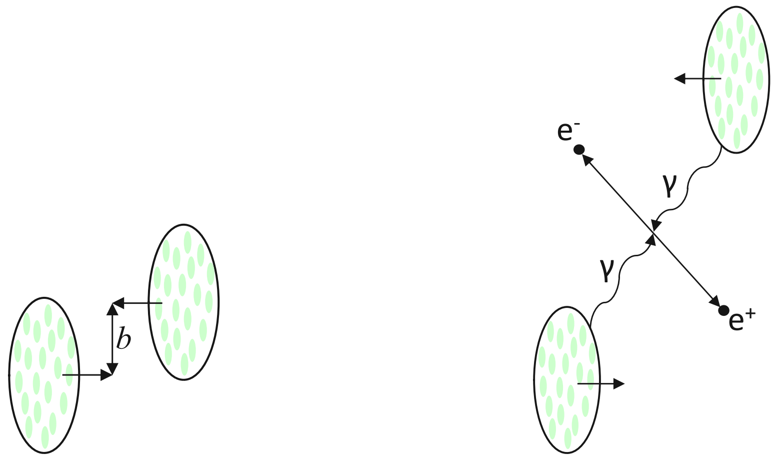


Figure 3.1: Data flow used to evaluate the ITS3 performance in the ALICE cavern. The pixel-hit source generator and the readout architecture model are implemented as two independent modules.



(a) (b)

Figure 3.2: Types of Pb-Pb interactions: (a) QCD hadronic collision, produced when the nuclei of the ions overlap with each other by an impact parameter b. (b) QED electron-positron pair, produced due to the exchange of photons between two ions that pass close to each other.

3.1 Types of Pb-Pb Interactions

The ALICE experiment is the only LHC experiment optimized for the heavy-ion physics program. During this program, Pb-Pb collisions are the densest and most common ions that collide in the LHC. The two most important interaction mechanisms between lead ions for pixel hit production are:

- **Hadronic (QCD) collisions:** when the ion nuclei physically overlap with each other (Figure 3.2(a)). The resulting number of secondary particles produced hadronically strongly depends on collision centrality, expressed in the impact parameter (b) [63].
- **Electromagnetic (QED) interactions:** when ions pass near each other interacting via their electromagnetic fields. This process leads to the production of electron-positron pairs through virtual photon exchanges as seen in Figure 3.2(b). These electron-positron pairs are commonly known as "QED electrons" and they constitute a background noise that produces pixel hits across the detector. Detailed pair production kinematics are described in Section 2.1.2.4.

The number of pixel hits that the crossing ions can generate depends on various LHC parameters, such as the energy of the ions, the frequency of the interactions, and the distance and angle between the Pb-Pb interaction and the sensor.

3.2 Properties of the Pb-Pb Run 3 at the LHC

Initially, the LHC magnets were designed for proton energies up to 7 TeV; the highest proton energy achieved so far is 6.8 TeV, corresponding to $\sqrt{s} = 13.6$ TeV for proton-proton collisions [64]. For heavy ions, the achievable nucleon-nucleon center-of-mass energy per nucleon pair depends on the lead charge-to-mass ratio Z/A:

The number of secondary particles that a Pb-Pb collision can generate strongly depends on the energy of the ions. The LHC's superconducting magnets were originally designed to accelerate protons to an energy of up to 7 TeV. However, due to design and manufacturing constraints [65], the highest proton beam energy achieved to date is 6.8 TeV, resulting in a center-of-mass energy of $\sqrt{s} = 13.6$ TeV for proton-proton collisions [64].

Lead nuclei contain both protons and neutrons. Since neutrons are electrically neutral, they cannot be directly accelerated by the LHC's RF cavity. As a result, the center-of-mass energy for lead-lead collisions is determined by the charge-to-mass ratio (Z/A) of lead nuclei:

$$\sqrt{s_{\mathrm{NN}}} = 2 \cdot (Z/A) \cdot 7\ TeV = 5.52\ TeV, \tag{3.1}$$

where:

- Z is the atomic number of lead (number of protons), which is 82.
- A is the atomic mass number of lead (total number of nucleons: protons + neutrons), which is 208.

During LHC Run 3, the maximum center-of-mass energy achieved in Pb-Pb collisions was $\sqrt{s_{\mathrm{NN}}} = 5.36$ TeV. For the design of the ITS3 pixel hit model, the Pb-Pb collision energy used will be $\sqrt{s_{\mathrm{NN}}} = 5.52$ TeV as a safety factor.

3.2.1 Pb-Pb Collisions Time Distribution

Particles (protons or ions) are injected into the LHC in bunches [66], which consist of orbit slots that can contain a group of protons or a single ion (depending on the type of run). The nominal proton bunch spacing is 25 ns (7.5 m), which defines the LHC base clock (40 MHz). For Pb ions, the filling scheme during Run 3 used a minimum spacing of 50 ns [67].

All bunches are arranged in trains that typically contain 40-56 lead bunches. Between these trains, there are some spaces with no bunches. Due to these spaces, only around two-thirds of the 27 km LHC orbit is actually filled with bunches.

The nominal average Pb-Pb collision rate experienced at ALICE is about 50 kHz (one collision every $\sim$20 μs). However, since collisions can occur only in the occupied third of the orbit, the real collision rate while a train crosses ALICE reaches 82 kHz.

In the pixel hit source model, two additional safety factors are applied:

- A safety factor 2 is applied on the average collision rate (50 kHz $\rightarrow$ 100 kHz).
- An additional assumption is that the entire ring is filled with bunches (100 kHz $\rightarrow$ 164 kHz).

These hadronic collisions occur randomly following a binomial distribution every time two Pb ions cross the ALICE experiment. This means that two Pb-Pb collisions can be separated by 50 ns, or by hundreds of microseconds. As the readout process hits in integration windows

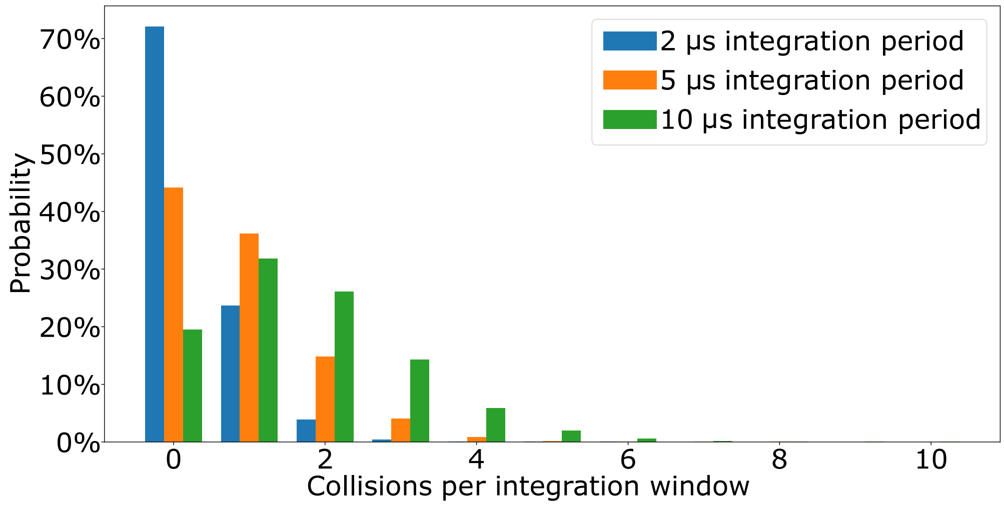


Figure 3.3: Pile-up of collisions per integration window of different periods. The collision rate is 164 kHz.

that are read and packed together, data from several consecutive Pb-Pb collisions can be shipped out together in what are called pile-up collisions. In Figure 3.3, the number of Pb-Pb collisions pile-up in integration windows of different periods is shown. As seen, approximately 72% of the 2 μs integration windows are empty, 23% have one Pb-Pb collision, and 5% contain multiple pile-up collisions.

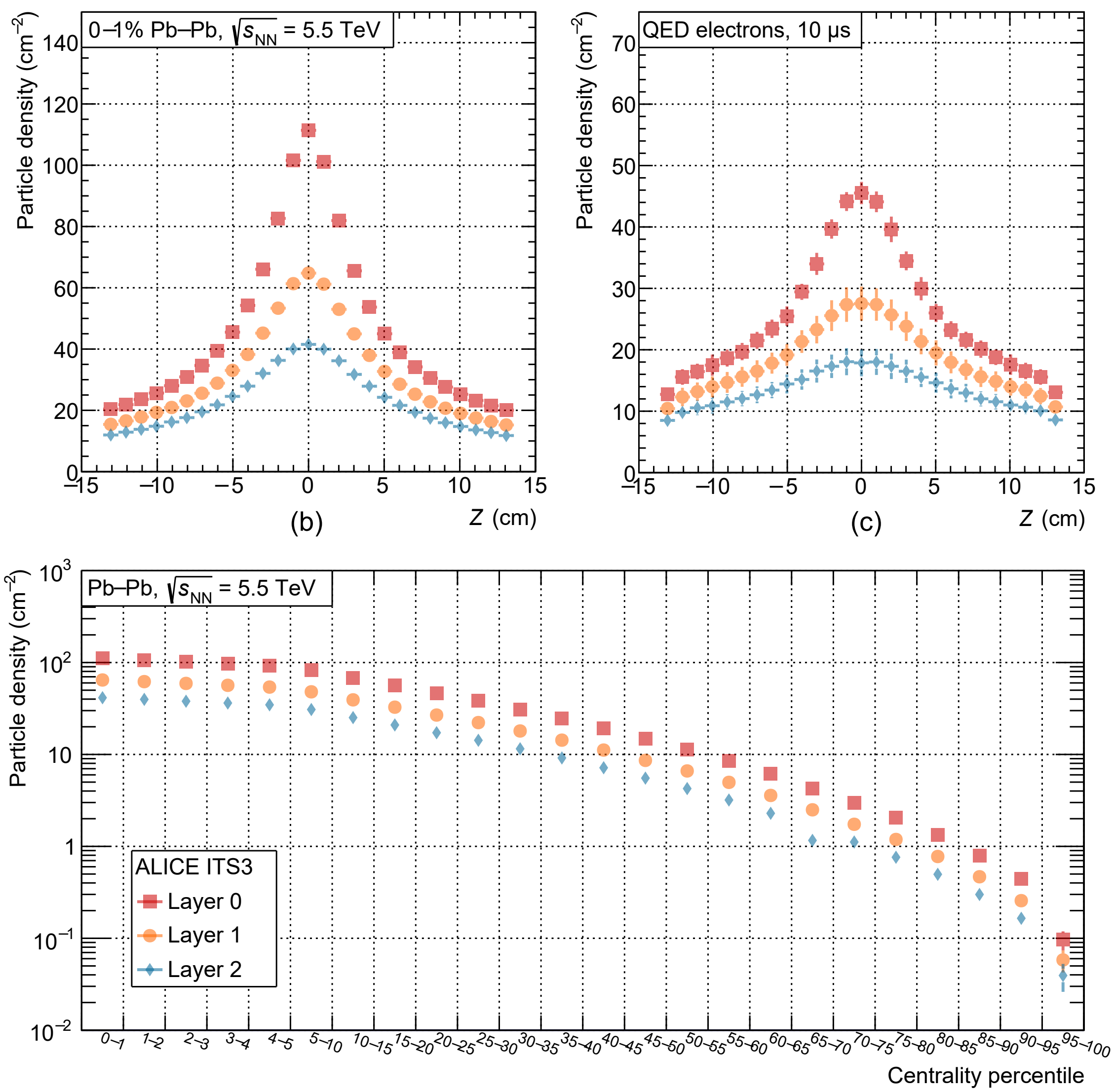
3.2.2 Pb-Pb Collisions Spatial Distribution

Collisions occur on average at the ALICE central point $(x, y, z) = (0, 0, 0)$. However, since the Pb-Pb ions move at almost the speed of light in the z direction, the collisions and QED interactions are Gaussianly distributed over z with a standard deviation of $\sigma = 6$ cm.

3.3 Particle Densities in the ITS3

The number of secondary particles crossing the detector depends on the distance from the origin of the Pb-Pb interaction. In this way, the ITS3 layer 0, will have a higher particle density than layers 1 and 2. This can be seen in Figures 3.4, where the particles crossing the different areas of the detector for Pb-Pb collisions and QED electrons are shown. In these graphs, all Pb-Pb interactions occur in $z = 0$ cm and have $\sqrt{s_{\mathrm{NN}}} = 5.52$ TeV.

Figure 3.4(a) shows the average particle density in $z = 0$ cm for Pb-Pb collisions of different centrality percentiles. As seen, the top 1 % most central collisions produce several orders of magnitude more particles than the peripheral ones. Figure 3.4(b) shows the average local particle density as a function of z for a single top 1% central collision. Here, it can be observed how the particle density decreases as Δz increases. Figure 3.4(c) shows the QED electron density along z for QED interactions originating at $z = 0$ during 10 μs. As seen, particle density also decreases with Δz, but in a more moderate way than in Pb-Pb hadronic collisions.



(a)

Figure 3.4: Particle densities on the three ITS3 layers: (a) By a Pb-Pb collision under different centrality percentiles in $z = 0$ cm. (b) By a 0-1% central collision (hadronically) in $z = 0$. (c) By QED electrons originated from $z = 0$ cm integrated over a 10 μs interval (source [9]).

3.4 Pixel Hits Clustering

A single charged particle may produce one or more pixel hits. The number of pixels fired by an incident particle is referred to as the *cluster size*, and depends on the energy that the particle can deposit by ionization and on its incident angle.

The cluster size of a minimum-ionizing particle (MIP) was studied in the ER1 run [68], [69]. To do so, a beam test was performed, and it confirmed that the cluster size for an MIP orthogonal particle was 1.2 hits per particle. However, not all particles coming from Pb-Pb interactions are MIP. To calculate the average cluster size of secondary particles from Pb-Pb

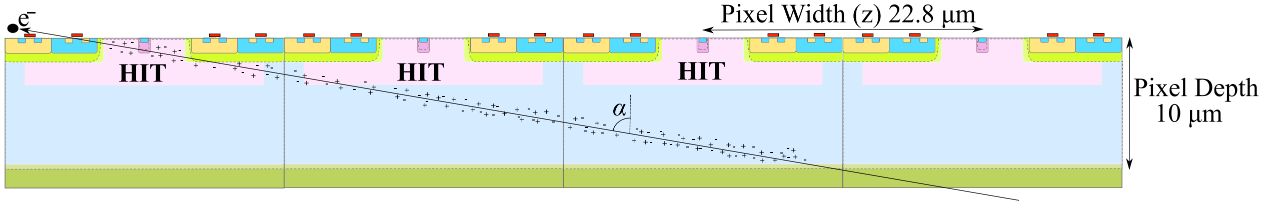


Figure 3.5: A charged particle creating three pixel hits due to its incident angle α.

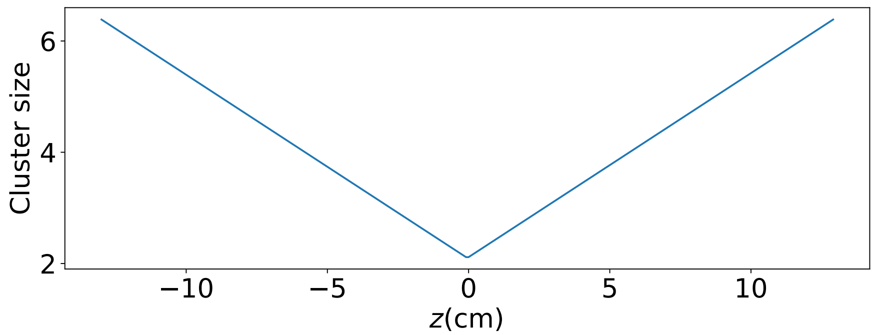


Figure 3.6: Cluster size as a function of z for particles originating at z=0 cm. Orthogonal particles yield 2.1 hits on average.

interactions, an extra factor 1.75 must be applied based on the cluster data extracted by the ITS2 in the ALICE cavern [9]. By combining these two values, the ITS3 cluster size for the average orthogonal particle was estimated as 2.1 hits per orthogonal particle.

If the particle path is not orthogonal, it can deposit energy in epitaxial layers of different pixels (Figure 3.6), triggering hits in several pixels. As the ITS3 layers are placed very close to the interaction point, it can be assumed that the pixels travel in a straight line from the Pb-Pb interaction point. Due to this, the cluster size of a particle that originated from a collision at $z = 0$ cm will increase linearly with Δz as shown in Figure 3.5.

3.5 Pixel Hit Densities in the ITS3

Figure 3.7 compares the average particle and pixel-hit densities integrated over a 10 μs period as a function of z for layer 0 of the ITS3. As seen, the particle density peaks sharply at $z = 0$ cm, having more than twice the particle density than in the detector edges at $|z| \approx 13$ cm. On the other side, Figure 3.7 shows a pixel-hit density curve notably flatter, where the increase in cluster size with Δz partially compensates for the lower particle density at the detector edges, producing a more uniform distribution of pixel hits across z.

Additionally, it can also be observed that the QED electron particles and pixel hits contributions are higher than the hadronic ones. This can be explained due to the small particle density that medium and low central collisions generate, as observed in Figure 3.4(a). If an integration window records a central collision, the number of hadronic pixel hits in that integration window will be significantly larger than that generated by QED electrons.

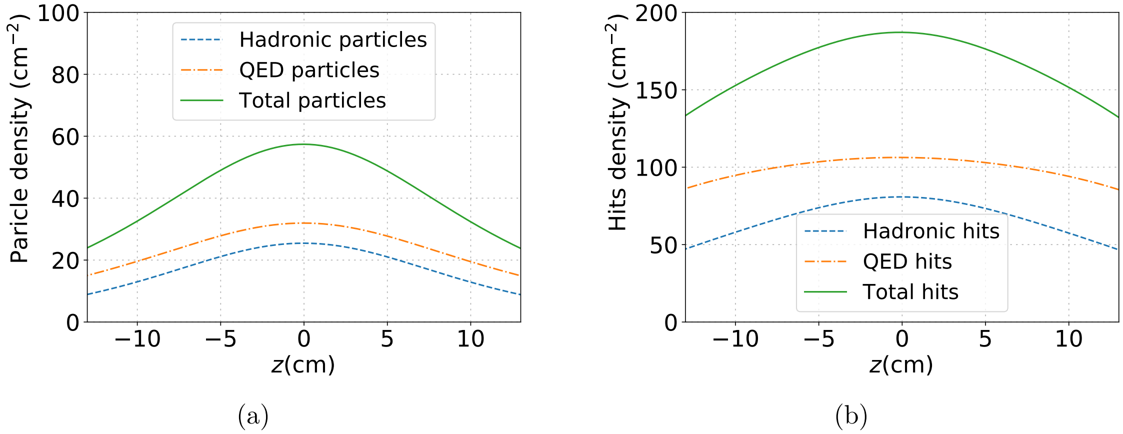


Figure 3.7: Average particle density (a), and pixel hit density (b) per unit area over z for the ITS3 layer 0 during a 10 μs integration period.

3.6 ITS3 Monte Carlo Pixel Hit Model

Particle and pixel hit averages are very useful to elaborate fast calculations of the readout performance. However, in a real scenario, the pixel hits are not uniformly distributed in time, accumulating more pixel hits in certain integration periods.

To fully emulate the pixel hit load, a Monte Carlo generator implemented in Python [70] was elaborated. This model simulates the Pb-Pb interactions that occur during an integration period and calculates the amount of pixel hits they will generate.

The key average values of the Monte Carlo pixel hit are summarized in Table 3.1.

3.6.1 Monte Carlo Algorithm

The Monte Carlo pixel hit source model proceeds as follows:

1. **Collision generation:**
    - Sample the number of Pb-Pb collisions pile-up during an integration period. This sample uses a Binomial distribution based on the interaction rate of 164 kHz and the integration period.
2. **For each collision simulated:**
    - Sample the collision centrality b as a uniformly distributed random value between 0-100%.
    - Sample the z position from a Gaussian distribution (σ=6 cm, mean = 0).
    - For each pixel matrix region:
        - Compute Δz between the collision and the pixel matrix region.
        - Use Δz and b to obtain the expected average particle density and cluster size.

Table 3.1: Key average values used in the data model.

PARAMETERS	VALUE	CONDITIONS
Particle flux (hadronic)	2.55 MHz/cm^2	over z
Particle flux (QED)	3.2 MHz/cm^2	over z
Total particle flux	5.7 MHz/cm^2	over z
Cluster size	2.1	orthogonal particles
Cluster size	3.06	over z
Particle fluency per collision (hadronic)	15.9 cm^{-2}	over z
Particle fluency per int. period (QED)	6.4 $^{-2}$	over z, int. period 2 μs
Particle fluency per int. period (QED)	16 cm^{-2}	over z, int. period 5 μs
Hit density per collision (hadronic)	59.5 cm^{-2}	over z
Hit density per int. period (QED)	21.3 cm^{-2}	int. period 2 μs
Hit density per int. period (QED)	53 cm^{-2}	int. period 5 μs
Occupancy (Pixel hit/total pixels)	$1.8 \cdot 10^{-4}$	int. period 2 μs
Occupancy (Pixel hit/total pixels)	$4.4 \cdot 10^{-4}$	int. period 5 μs

  - Sample the particle count using the expected average particle density from a Poisson distribution.
  - Multiply by the cluster size to get the pixel hits.

3. **QED electron generation:**
   - For each integration period, sample a QED interaction z-origin using the same Gaussian distribution as the one used for the collisions. This is performed because the QED interaction between ions not only occurs in z=0 cm, but also all over z.
   - Repeat the above steps using the QED particle density and cluster size model to sample the QED pixel hits on all the pixel matrix regions.
4. **Repeat for each integration period**

The pixel hits from several consecutive integration windows are used as input by the readout model explained in Chapter 4, outputting its readout performance under that load.

To validate that the Monte Carlo algorithm aligns with the provided physics data, over 100,000 integration windows were generated. From this dataset, the averages of particle and hit densities were calculated over z. These results matched the theoretical average values derived from the physics data, confirming that the input provided by the physicists is successfully modeled.

3.7 Conclusions

In high-energy physics, simplified models such as Poisson or Gaussian distributions are often used to simulate detector data fluxes. However, for heavy-ion experiments like those at ALICE, such approximations fail to capture the temporal and spatial complexity of real PbPb collisions. Despite having similar average hit rates, realistic collision events generate hit distributions that differ substantially from Poisson-based models.

A physics-driven data model is, therefore, essential for accurately dimensioning readout architectures in such environments. In the case of the MOSAIX, this model was created and used not only to dimension the architecture parameters, but also to guide the high-level architectural design of the MOSAIX chip.

Chapter 4

MOSAIX Readout Model Design and Analysis

To meet the stringent design constraints of the ITS3 project, the on-chip readout architecture must achieve minimal area usage, low power consumption, and high performance. Achieving an optimal balance between these requirements involves tuning various design parameters, including the number and depth of on-chip FIFOs, the number of serial links to the Left Endcap (LEC), and their bandwidth.

To explore these trade-offs, a behavioral model of the ITS3 on-chip readout architecture was developed using SystemVerilog [71]. This model was designed to be cycle-accurate and optimized for simulation performance, providing both functional fidelity and fast simulation times.

To evaluate the performance of this architecture, the model was first tested using a simple Poisson-based random hit generator. Subsequently, it was validated using the more realistic physics-based data model introduced in Chapter 3.

4.1 Readout Behavioral Model

Figure 4.1 illustrates the initial architectural concept of the MOSAIX chip. The only predefined constraint was that the chip would be segmented, with each segment containing multiple Repeated Sensor Units (RSUs). These RSUs are read out serially via multiple links directed toward the LEC, where data is processed and transmitted externally. At this stage, the number of tiles, their dimensions, and the number of serial links were not yet defined.

The behavioral model was created with two main purposes: (1) to estimate resource usage and optimize performance before RTL implementation, and (2) to support architectural decisions, including the number and layout of tiles, RSUs, and serial links.

As the project evolved, the segment dimensions were fixed at 1.85 cm in width and 26.6 cm in length. The left and right endcaps were allocated 0.45 cm and 0.15 cm, respectively. Initially, 10 RSUs per segment were planned; however, this was later increased to 12 to improve

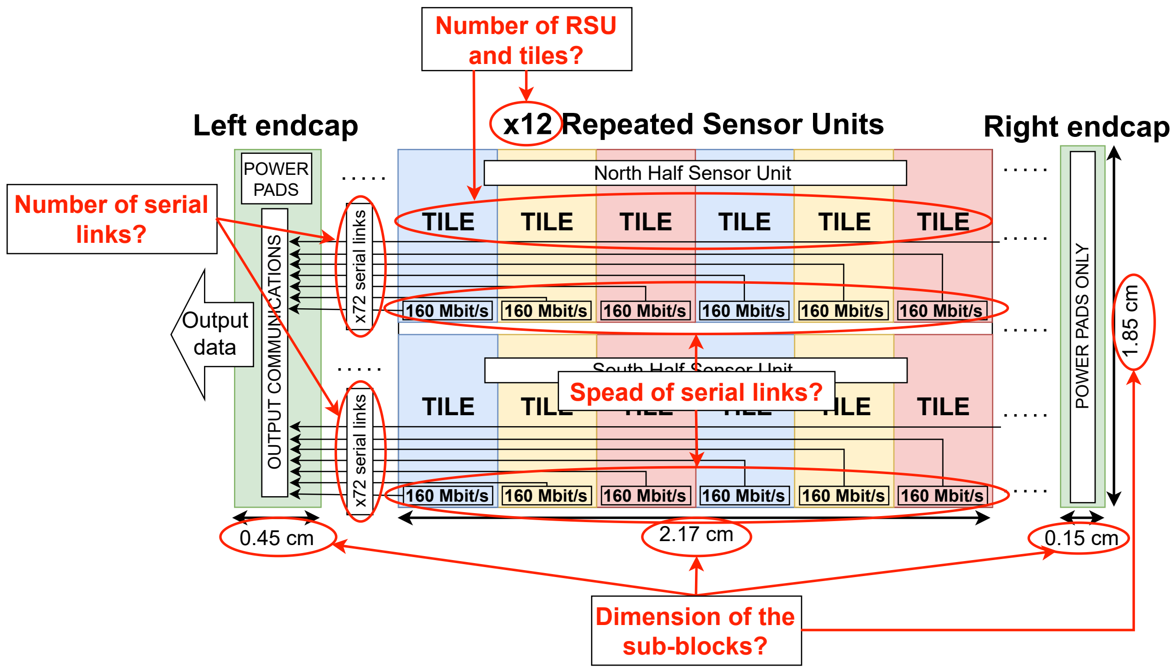


Figure 4.1: MOSAIX segment schematic showing endcaps RSU and tiles. The tiles are read out on-chip through several serial links to the LEC, where the data is shipped outside of the chip. At the start of the MOSAIX project, the dimensions and number of tiles and links were not defined.

bandwidth and accommodate analog design requirements while keeping the number of links at a manageable number. Each RSU contains 12 tiles of 3.58 mm $\times$ 9.782 mm each, enhancing power domain granularity for isolating potential shortcuts.

As illustrated in Figure 4.2, each tile is divided into two main parts: the pixel matrix and the periphery. The pixel matrix constitutes the sensitive area of the detector, while the periphery implements the readout circuitry that transfers hits from the matrix to the LEC. To reduce dead area, the periphery must be as short as possible, leaving space for the pixel matrix. The final area allocated to the periphery was 1.4 mm^2, leaving 33 mm^2 to the pixel matrix.

The readout of the pixel hits is performed using a frame-based approach, grouping the pixel hits within configurable integration windows (minimum duration: 2 μs). Hits are temporarily stored in the in-pixel memory registers.

Two digital pixel architectures were proposed. An architecture that has two in-pixel registers and an architecture that has three in-pixel registers. Due to the limited improvement in performance offered by the three in-pixel registers and the constraints on pixel size, the configuration with two in-pixel registers was selected. The selected digital pixel architecture is presented in Section 5.4.

The number of pixels within the pixel matrix of a tile also varied during the early design phase of the ITS3 due to changes in the dimensions of the pixel and the tile. After several iterations, the pixel matrix dimensions were fixed to 444 $\times$ 156 pixels, organized in 78 double

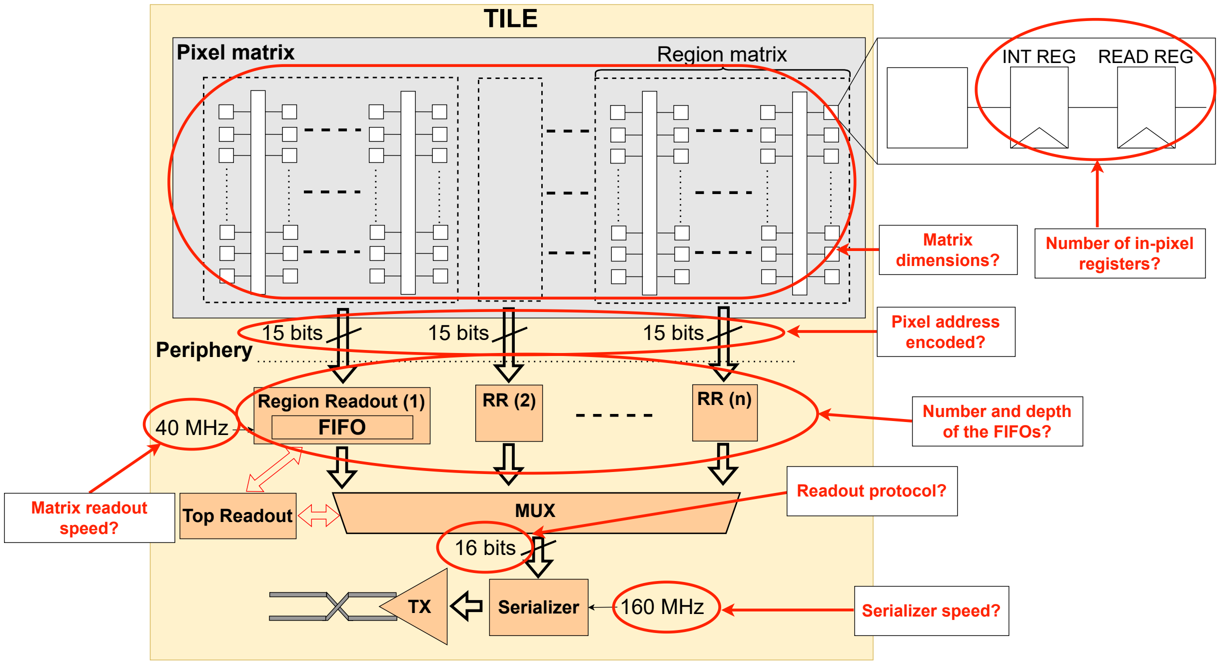


Figure 4.2: Tile-level view of the on-chip readout architecture. At the start of the MOSAIX project, several design parameters, such as the number of regions and FIFOs, and the number and speed of the serial links, needed to be defined.

Table 4.1: Bits required to encode a pixel hit within a region for different pixel matrix partitions in a tile.

	Number of regions in a tile					
	1	2	3	4	5	6
bits	17 bit	16 bit	15 bit	15 bit	14 bit	14 bit

columns.

To improve throughput, the matrix is subdivided into regions that are read in parallel into local FIFOs in the periphery. Data from the FIFOs is then serialized and transmitted on-chip to the LEC as time-stamped packets.

The communication links to the LEC adopted a 16-bit word protocol. This imposed a constraint on the address encoding, as a hit must be encoded in 15 bits or fewer (1 bit reserved for identifying the word type). However, the number of bits needed to encode a pixel hit is:

$$\lceil \log_2(444) \rceil + \lceil \log_2(156) \rceil = 17 \text{ bit}$$

To meet the 15-bit encoding constraint, pixel hits are grouped into region packets, each preceded by a region header. As shown in Table 4.1, using at least three regions per tile enables encoding within the 15-bit limit.

Table 4.2: Maximum number of pixel hits transferable from the periphery to the endcap in a 2 μs window, using 16-bit words, for different serializer configurations.

f_{MATRIX} \ n_{REGIONS}	3	4	5	6
20 MHz	120	160	200	240
40 MHz	240	320	400	480
80 MHz	48	640	800	960

Table 4.3: Maximum pixel hits transferred from the periphery to the endcap during a 2 μs window for different serializer configurations. The protocol words are encoded with 16 bit.

f_{SER} \ n_{SER}	1	2
80 MHz	10	20
160 MHz	20	40
320 MHz	40	80

4.2 Methodology and Readout Performance

The number of pixel hits that can be transferred from the pixel matrix to the periphery during a single integration window T_{INT} depends on the number of readout regions n_{REGIONS} and the matrix clock frequency f_{MATRIX}, as described by:

$$\text{Max. hits} = T_{\text{INT}} \cdot f_{\text{MATRIX}} \cdot n_{\text{REGIONS}}$$

Table 4.2 reports the number of hits that can be read out from the pixel matrix to the regional FIFOs during a 2 μs window. Frequencies are chosen as multiples or divisors of 40 MHz, matching the LHC bunch crossing frequency described in Section 3.2.1.

Once hits are stored in the region FIFOs, they are serialized and transferred to the LEC. The transfer capacity depends on the number of serial links n_{SER}, the serializer clock frequency f_{SER}, and the pixel hit encoding width (16 bits). The maximum number of hits that can be transmitted during one integration window is:

$$\text{Max. hits} = \frac{T_{\text{INT}} \cdot f_{\text{SER}} \cdot n_{\text{SER}}}{\text{encoded\_bits}}$$

Table 4.3 summarizes the serialization bandwidth in terms of protocol words that can be sent under different configurations. Comparing these numbers with the ones presented in Table 4.2, it is clear that the bandwidth from the matrix to the periphery significantly exceeds the bandwidth from the periphery to the LEC. To absorb short-term data bursts waiting in the tile to be shipped out, a region FIFO is placed to sort the pixel hits for each of the matrix regions.

All hits detected during an integration window must be shipped out together in one integration window packet. This means that, if any hit remains in the matrix after the integration window

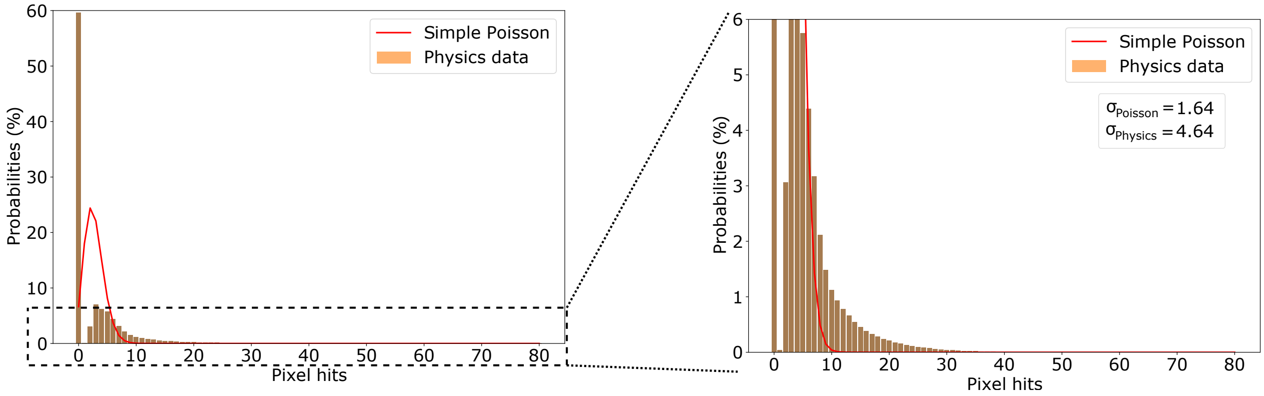


Figure 4.3: Pixel hit distribution comparison: Poisson model vs. physics-based model in a 3.56×9.24 mm region during a 2 μs integration window.

ends, the following integration window must be sacrificed to complete the readout of the previous frame.

For a successful reconstruction of an event within an integration window, all the hits from all tiles are necessary. Since each tile uses an independent link to the LEC, failure in just one tile renders the entire integration window unusable. Due to this, in this chapter, the performance of the readout will be measured in terms of integration windows incomplete in a MOSAIX segment, i.e., at least one of the tiles in the segment dropped the data corresponding to that integration window.

In a segment, there are 144 tiles. However, the possibilities of these tiles dropping integration windows are correlated to the Pb-Pb collisions, as all the tiles are observing the same physical events. This means that a high collision pile-up or a high central collision will generate lots of hits in all tiles of the ITS3 as shown in Chapter 3. If this happens, several tiles can fall into the same integration windows.

4.3 Model Simulation and Analysis

Initial simulations used a Poisson-based hit generator with a particle flux of 5.75 MHz/cm^2. Each particle was assumed to generate a cluster of three hits.

Figure 4.3 compares the hit distributions from the Poisson and physics-based models. Although both matched in mean occupancy, they differ substantially in shape. The physics-based model shows that approximately 60% of integration windows are empty, compared to only 8% in the Poisson case. Furthermore, the physics model exhibits a long tail, reflected in a standard deviation (σ) of 4.64 against the 1.64 of the Poissonian model. This standard deviation is caused by high-occupancy integration windows caused by high-centrality collisions or the pile-up of multiple collisions.

When analyzing the MOSAIX readout using the Poisson model, the FIFO depth has minimal impact on the losses, as temporal storage of data bursts is unnecessary. However, when analyzing the MOSAIX readout using the physics model, large FIFOs are required to cope with the variability of the data due to the large tails in the pixel hit distribution.

Unless otherwise stated, all simulation results presented in this chapter are reported as the number of integration windows in which at least one tile fails to transfer all hits. Each scenario is based on three simulation runs of 100,000 μs each, using the physics-based model described in Chapter 3. The error bars in the plots represent the standard deviation across these three runs.

4.3.1 Protocol Selection

The serializer protocol required to transfer integration window packets from each tile to the LEC is a critical specification in the design of the readout architecture. In addition to pixel hit addresses, each integration window packet *must* include a timestamp and region delimiters to ensure that every pixel hit can be encoded within 16 bits. Optionally, the protocol may also include features such as checksums or status flags to support error reporting or provide auxiliary information to the backend system.

The protocol defines three fundamental types of integration window packets:

- **Normal** packets: Sent when at least one pixel hit is detected and successfully read out.
- **Empty** packets: Sent if no hits are detected within an integration window.
- **Drop** packets: Sent when a packet is dropped due to incomplete readout during the allocated window (e.g., overflow requiring the next window to be sacrificed).

Empty and drop packets are both encoded using a single protocol word. The structure of normal integration window packets can vary depending on which required or optional fields are included. Based on these considerations, four protocol variants were proposed and evaluated, which will be described next.

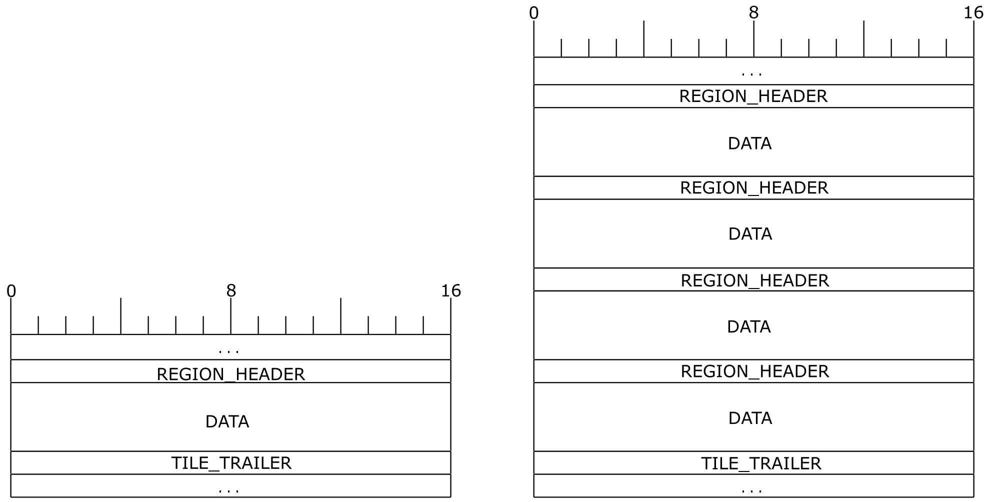
4.3.1.1 Protocol 1: `RegionHeader` and `TileTrailer`

An example of integration window packets from protocol 1 with hits in one and four regions can be seen in Figure 4.4. As seen, each integration window packet starts directly with a `RegionHeader` word followed by the data corresponding to that region. Depending on the number of regions with hits, it can be followed by additional `RegionHeader` words and their corresponding `Data` words. After all data are transmitted, the packet concludes with a `TileTrailer`.

In this protocol, the `RegionHeader` word contains the time-stamp information and the `TileTrailer` a checksum and some status flags.

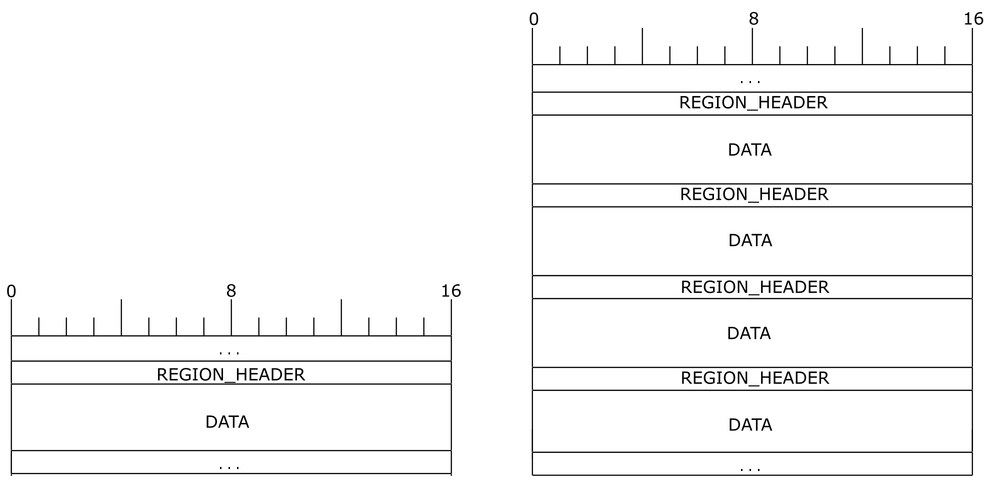
4.3.1.2 Protocol 2: `RegionHeader` Only

Protocol 2 is a simplified variant of Protocol 1 that omits the `TileTrailer` to reduce overhead and improve link occupancy. As shown in Figure 4.5, integration window packets begin with `RegionHeader` fields (only for active regions), followed directly by the corresponding pixel data.



(a) (b)

Figure 4.4: Protocol 1 integration window packets: (a) One region with hits. (b) All four regions with hits.



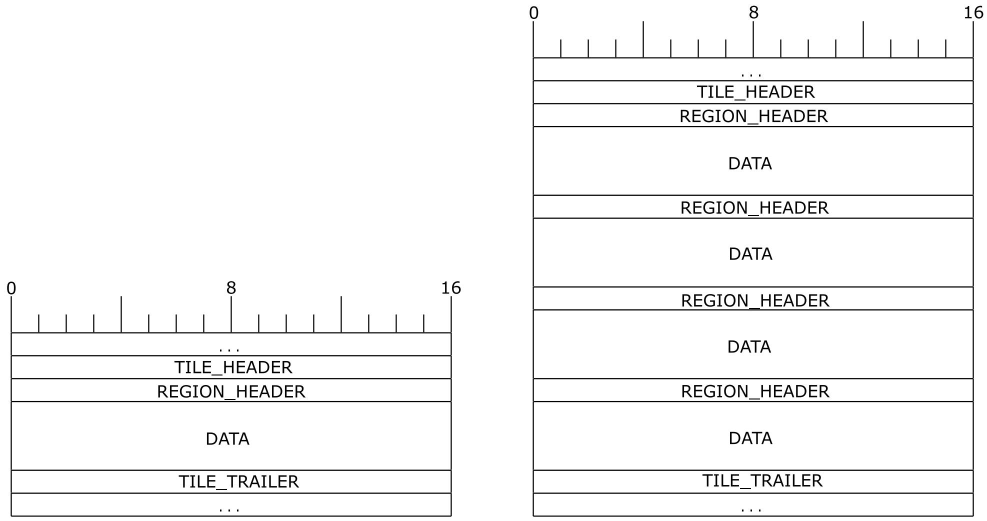
(a) (b)

Figure 4.5: Protocol 2 integration window packets: (a) One region with hits. (b) All four regions with hits.

This protocol improves link occupancy at the expense of removing checksum, status flags, and not having a special protocol word that closes the integration window packet.

4.3.1.3 Protocol 3: `TileHeader`, `RegionHeader`, and `TileTrailer`

Protocol 3 (Figure 4.6) builds upon Protocol 1 by adding a `TileHeader` at the start of each packet. This header includes the tile address, providing redundancy and allowing early recognition of the packet by the receiver. This redundancy can help detect a misbehavior and gives the backend more time to allocate resources for processing the incoming data.

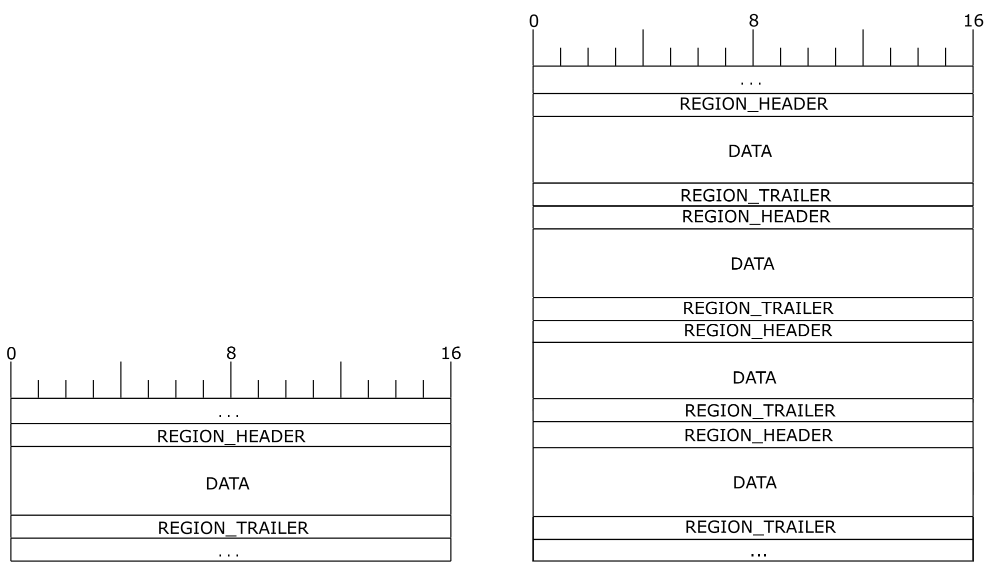


(a) (b)

Figure 4.6: Protocol 3 integration window packets: (a) One region with hits. (b) All four regions with hits.

4.3.1.4 Protocol 4: `RegionHeader`, and `RegionTrailer`

In Protocol 4 (Figure 4.7), each region with data is treated as an independent packet, enclosed by a `RegionHeader` and a `RegionTrailer`. This design allows for complete independence of regional processing and simplifies the readout RTL, as each region can be read directly from the FIFO.



(a) (b)

Figure 4.7: Protocol 4 integration window packets: (a) One region with hits. (b) All four regions with hits.

4.3.1.5 Protocol Comparison and Discussion

Figure 4.4 shows a comparison of the four proposed protocols in terms of the percentage of occupancy of a 160 Mbit/s link with the LEC for the worst-case tile in z=0 cm. As seen, short integration periods (2 μs) present big occupancy differences among protocols, while the difference among large integration windows (10 μs) is almost negligible. This comes from the fact that in large duration integration windows, the majority of the regions will have hits in all the packets. Additionally, the overhead due to an extra word (e.g., a `TileTrailer`) per integration window will become less significant.

A summary of the advantages and limitations of each protocol is presented below:

- **Protocol 1**
  - Includes a packet-level checksum.
  - Provides good link occupancy.
- **Protocol 2**
  - Omits checksum to maximize link occupancy.
  - Offers the best performance, but also reduces robustness.
- **Protocol 3**
  - Adds a redundant `TileHeader` at the packet start.
  - Has a fixed data word for the start and the end of the integration window packet.

Table 4.4: Comparison of the four protocol variants based on occupancy of a 160 Mbit/s link for the worst-case tile at $z = 0$ cm. Comparison performed mathematically.

Comparison of Protocols		
% of 160 Mbits/s link occupancy		
Integration period (µs)	Number of regions	
	3	4
	Protocol 1	
2 µs	74%	75%
5 µs	68%	69%
10 µs	65%	66%
	Protocol 2	
2 µs	70%	71%
5 µs	66%	67%
10 µs	64%	65%
	Protocol 3	
2 µs	79%	80%
5 µs	70%	71%
10 µs	66%	67%
	Protocol 4	
2 µs	84%	90%
5 µs	72%	75%
10 µs	67%	69%

– Slightly worse link occupancy, but facilitates backend synchronization.

- **Protocol 4**
  - Sends redundant headers and trailers for each region.
  - Enables complete independence between regions.
  - Results in the worst link occupancy.

Protocol 2 was rejected due to the lack of a checksum. Protocol 4 was also excluded due to poor link efficiency and minimal added value compared to simpler alternatives.

Protocols 1 and 3 were the final candidates. They differ primarily in the placement of the tile address: Protocol 3 begins with a header that explicitly indicates it, while in Protocol 1, this information is not explicitly indicated in a protocol word. Although Protocol 1 showed slightly better occupancy, the MOSAIX backend team favored Protocol 3 due to its easier synchronization and error detection capability.

Protocol 3 was therefore selected as the baseline for the MOSAIX on-chip readout architecture. All simulation and performance studies presented in this chapter are based on this protocol. A detailed RTL-level specification of the protocol is provided in Chapter 6.

4.3.2 Effect of Integration Period on Readout Performance

Beyond the readout protocol, performance is also affected by the duration of the integration window. Figure 4.8 compares the readout performance across different integration periods and FIFO depths.

As shown, an integration window of 1 μs yields poor performance, with approximately 1%

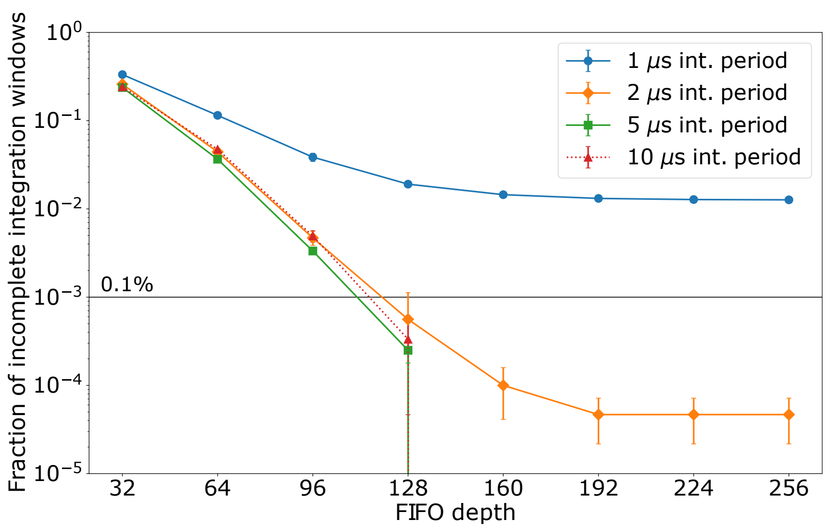


Figure 4.8: Readout performance at the segment level for various integration window periods and FIFO depths for three runs of 100,000 μs each. The error bars in the plots represent the standard deviation across these three runs. The studied configuration uses four regions and a single 160 Mbit/s link.

Table 4.5: Number of integration windows of 1 μs period with losses due to matrix readout bottleneck for different numbers of regions and matrix readout speeds. Data extracted from simulations of 100,000 μs for each number of regions.

	Number of regions in tile					
Matrix readout speed	3	4	5	6	7	8
40 MHz	2.41 %	1.32 %	0.695 %	0.354 %	0.186 %	0.085 %
50 MHz	1.03 %	0.22 %	0.065 %	0.026 %	0.015 %	0.006 %
60 MHz	0.231 %	0.027 %	0.005 %	0.002 %	0.0 %	0.0 %
70 MHz	0.042 %	0.004 %	0.001 %	0.001 %	0.0 %	0.0 %
80 MHz	0.015 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %

of integration windows presenting losses even at the highest FIFO depth, ten times higher than the ITS3 requirement indicated by the horizontal line. Notably, this loss rate remains unchanged even when increasing the FIFO depth from 160 to 256 words, indicating that the bottleneck is not the FIFO capacity, but the limited time available to transfer hits from the pixel matrix to the FIFOs.

For a 1 μs window, the maximum number of hits transferable from a region is:

$$T_{\mathrm{INT}} \cdot f_{\mathrm{MATRIX}} = 1\ \mu\mathrm{s} \cdot 40\ \mathrm{MHz} = 40\ \mathrm{hits}.$$

To detect 40 hits in a region is relatively common, occurring in at least one tile per region for 1% of the integration windows. This creates a hard upper limit on the performance for this configuration. A similar limitation appears with a 2 μs integration period with no performance improvements for FIFO depths higher than 192 words. However, in this case, the losses are significantly lower, below the 0.01% ITS3 requirement. For longer integration periods (5 μs and 10 μs), such transfer limits are rarely reached, exhibiting negligible losses when the FIFOs are at least 160 words in depth.

When the process of reading from the pixel matrix into the FIFOs is not dominant, the performance increases exponentially with the FIFO depth. This improvement does not vary much with the integration periods as the performance for 2, 5, and 10 μs integration periods are identical for FIFO depths between 32 to 128. However, at a FIFO depth of 160, losses for the 5 and 10 μs integration periods become negligible, while losses of 2 μs integration period become constrained by matrix readout limits.

Overall, this analysis confirms that an integration period of 2 μs is the minimum period with less than 0.1% integration windows packets with losses. If a shorter integration period were required, only two options remain available: increase the number of regions into which the pixel matrix is partitioned, or increase the pixel-matrix readout speed.

Table 4.5 summarizes the number of integration windows of 1 μs period with losses under different configurations. To satisfy the target requirement of losses below 0.1%, the number of regions would need to be doubled from four to eight. As an alternative to increasing the region count, the matrix readout speed could instead be raised from the nominal 40 MHz to approximately 60 MHz while keeping four regions.

Since the primary purpose of the Inner Tracking System is precise hit positioning, while timing precision is delegated to other ALICE detectors such as the TPC and ToF [72], an integration window of 2 μs is considered sufficient for ITS3. This avoids unnecessary logic while meeting performance requirements.

4.3.3 Selection of Readout Design Parameters

Figure 4.9 presents the fraction of integration windows with data losses plotted against FIFO depth for different combinations of matrix regions and serializers. The number of FIFOs matches the number of regions.

As seen, four configuration options that satisfied the <0.1% loss requirement were studied in detail. To compare them correctly, early post-implementation RTL power and area estimates were performed for each of them. In this analysis, Option 4 was discarded due to its significantly higher power and area consumption compared to the other options.

The model presented in Chapter 3 could be modified if the LHC conditions change. To further differentiate Options 1, 2, and 3, their robustness to variations in cluster size and interaction rate was evaluated. Table 4.6 summarizes performance under various scenarios.

As seen, option 3 stands out for its performance under high-cluster and high-interaction rate scenarios. For example, it meets the ITS3 requirement even at an interaction rate of 123 kHz with a cluster size of 2.6. As a result, after considering performance, power, and area trade-offs, the selected configuration contains 4 matrix regions, each with a 160-word FIFO, and a single 160 Mbit/s link from tile to LEC.

4.3.4 Simulation Extension to the Full ITS3 Detector

Since high-centrality collisions simultaneously affect all tiles, the likelihood of data losses increases with the number of tiles analyzed. Figure 4.10 shows how the fraction of incomplete

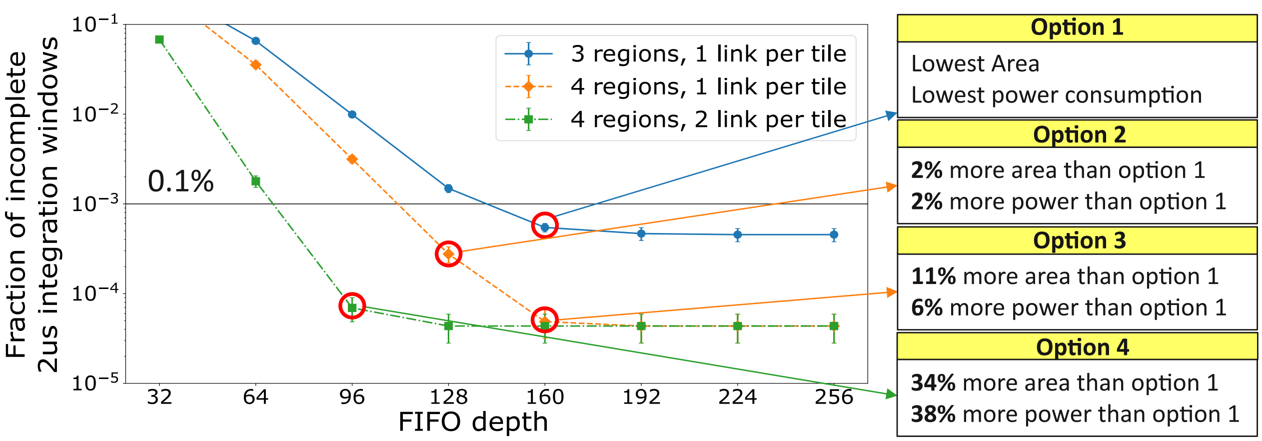


Figure 4.9: Performance comparison for different readout configurations across FIFO depths for three runs of 100,000 μs each. The error bars in the plots represent the standard deviation across these three runs. Each configuration varies in the number of regions and serializers to the LEC. The horizontal line marks the 0.1% ITS3 losses requirement.

Table 4.6: Percentage of incomplete collisions for the configuration options shown in Figure 4.9, for different combinations of cluster size of orthogonal particles and interaction rates.

	Option 1			Option 2			Option 3		
Int. rate \ cluster	2.6	2.1	1.6	2.6	2.1	1.6	2.6	2.1	1.6
82 kHz	0.073%	0.004%	0.0%	0.004%	0.0%	0.0%	0.0%	0.0%	0.0%
123 kHz	0.103%	0.043%	0.0%	0.098%	0.003%	0.0%	0.023%	0.003%	0.0%
164 kHz	1.613%	0.057%	0.002%	4.207%	0.028%	0.0%	1.937%	0.005%	0.0%
205 kHz	38.2%	1.667%	0.004%	64.1%	8.54%	0.002%	64.1%	5.17%	0.0%

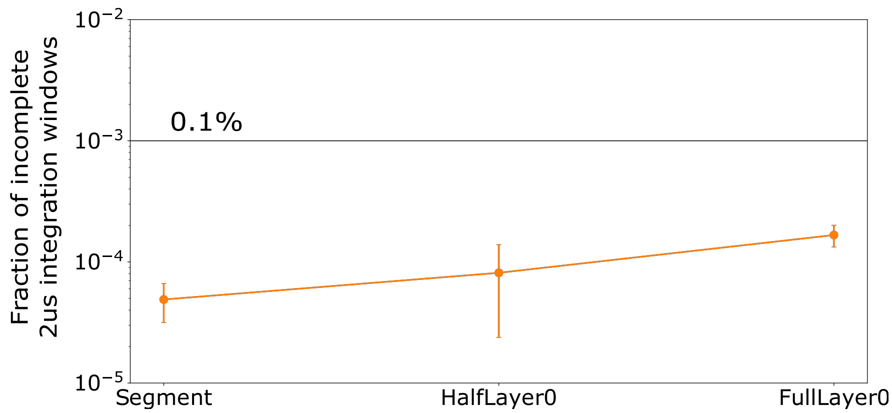


Figure 4.10: Fraction of integration windows with dropped data for different sensor areas: a Segment, a full MOSAIX chip (half ITS3 layer 0), and two MOSAIX chips (entire ITS3 layer 0). Configuration: 4 regions with 160-word FIFOs each. Three runs of 100,000 μs each were performed for each sensor area. The error bars in the plots represent the standard deviation across these three runs.

integration windows changes when scaling from a Segment to a full MOSAIX chip (half ITS3 layer 0), and to two MOSAIX chips (the entire ITS3 layer 0).

As seen, even with an increased sensor area, the losses remain below the requirement of 0.1%. This analysis demonstrates the robustness of the chosen readout architecture at the full detector scale.

4.3.5 Readout Model Validation

To validate the behavioral model, identical input data was injected into both the simulated readout and the RTL implementation described in Chapter 6.

Although the RTL simulation exhibited slightly higher losses, these discrepancies were attributed to low-level latency details, such as synchronization delays and serialization overhead, that were not defined by the time the behavioral model was designed. Despite these minor deviations, the results were highly correlated, confirming the validity of the model as a reliable predictor of the RTL design performance.

4.4 Conclusions

In this chapter, an in-depth analysis of the ITS3 readout performance is presented under the expected ITS3 data flux presented in Chapter 3 for different combinations of design parameters. From this analysis, an optimal configuration was obtained, that includes one 160 Mbit/s serial link, four pixel matrix regions per tile, and four 160-word FIFOs that are read in parallel at 40 MHz. The minimum integration period was set at 2 μs, as shorter durations (e.g., 1 μs) resulted in unacceptable data losses.

The simulation results confirm that this design meets the ITS3 requirement of less than 10^{-3} incomplete integration windows, both at the segment level and across the entire ITS3 Layer 0. This validates the architecture proposed in Chapter 6.

Chapter 5

The MOSAIX Pixel Matrix

After the model analysis presented in Chapter 4, the pixel matrix in each tile was divided into four regions, which are read out in parallel at a frequency of 40 MHz. To be able to perform this readout, the pixels and pixel matrix logic must have:

- A pixel sensor to detect the charge particles crossing the pixels.
- Logic to encode the address of each pixel within the matrix.
- Logic to route a clearing pulse that erases a hit from a single pixel at 40 MHz.
- A global `frame` pulse to capture the status of the matrix at the start of each integration window.
- A global `hold` signal to stop hit integration if the readout system cannot sustain the hit rate.
- The ability to inject both analog pulses (charge injection into the pixel sensor) and digital pulses (forced hits in the digital pixel) using a pulse command signal (`pulse_cmd`).
- Circuitry to configure and mask the pixels.

These functions are implemented between the in-pixel logic and the pixel matrix, and are operated by the readout architecture. In this chapter, a detailed implementation of the pixel and pixel matrix logic will be presented.

Although the author did not contribute to the design of the pixel sensor, he was heavily involved in the development of the digital pixel, which plays a central role in the readout architecture.

5.1 Pixel Matrix Addressing

The full pixel matrix in a tile contains 444×156 pixels, and is divided into four regions. For an easier operation, pixels are grouped into double columns of (444×2 pixels), resulting in 78 double columns in the entire matrix. Since 78 is not divisible by four, the number of double columns per region is not the same, with two regions containing 20 double columns and the

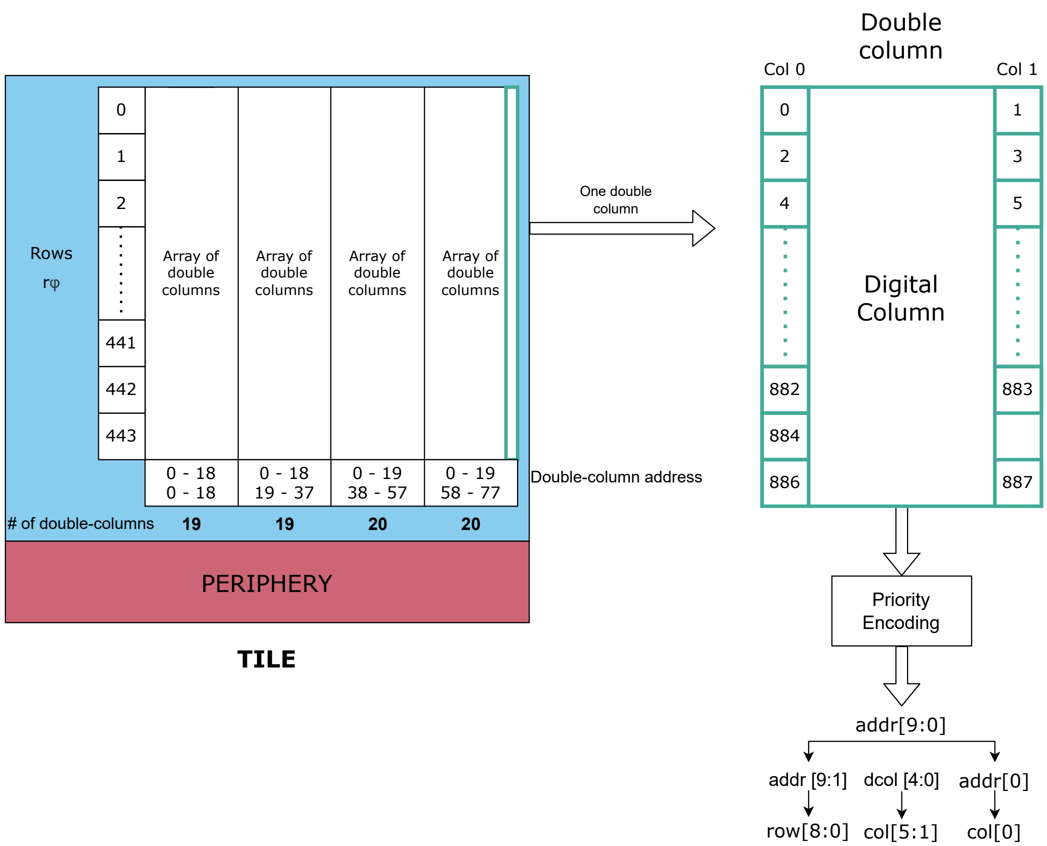


Figure 5.1: Pixel address and region encoding.

other two containing 19.

Each double column includes re-buffering of the global signals `frame`, `hold` and `pulse_cmd`, and a Priority Encoder (PE). The PE is a zero-suppression fully combinational logic that outputs the encoded address (`addr`) of one pixel hit in the double column at a time in a pre-configured priority order. Additionally, it outputs a bit that indicates if there is any valid hit in that double column or not (`hit_valid`). The PE also routes the clear pulse (`clr_pix`) back to the selected pixel, enabling sequential clearing during readout.

The address outputted by the PE has 10 bits, which correspond to the address of the pixel within the double column. To encode the address of the double column within the region, five additional bits must be added, using the address scheme presented in Figure 5.1.

The basic double column does not include the tree to configure each pixel. To route these configuration signals without losing sensitive area, they were embedded into two special double columns: the global-signals double column and the decoder double column. These special

Table 5.1: Delay comparison across digital column flavours.

From	To	basic		column decoder		global signals	
		Max [ns]	Min [ns]	Max [ns]	Min [ns]	Max [ns]	Min [ns]
clr_pix	addr	15.491	0.868	17.628	0.891	15.999	0.810
clr_pix	hit_valid	13.082	0.864	14.638	0.818	13.501	0.742
frame	hit_valid	14.349	0.623	15.870	0.629	14.929	0.583
frame	addr	16.889	0.732	18.296	0.749	17.462	0.720
hold	INT_REG	6.466	1.242	7.905	1.203	6.679	1.702
pulse_cmd	INT_REG	8.199	0.027	8.455	0.254	8.067	0.125

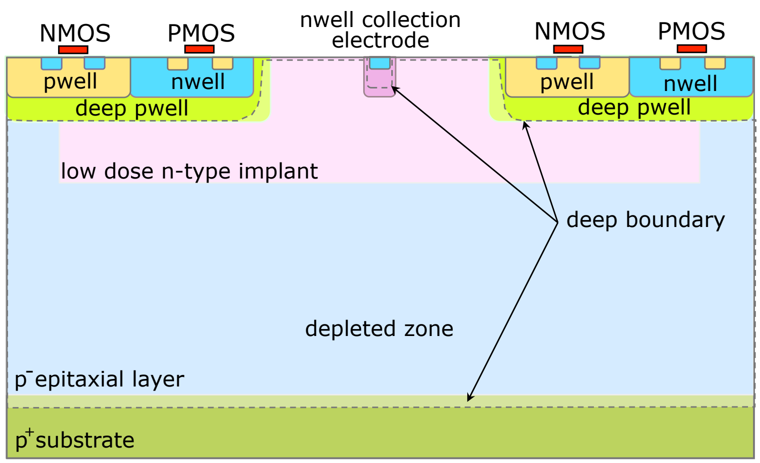


Figure 5.2: Schematic cross-section of the ITS3 pixel sensor well structure.

double columns have the same functionality as the basic double columns, but include extra logic in the same area, resulting in a degradation of the timing arcs as shown in Table 5.1. Despite this, all the double column flavours share the same pixel sensor and logic, presented in the following sections.

5.2 Pixel Sensor

During the MLR1 prototype run of the ITS3 project, various pixel sensor architectures were simulated and tested [73]–[77]. Among these, the *modified with gap* variant demonstrated the best performance and was selected for the MOSAIX design.

This variant of pixel sensor, illustrated in Figure 5.2, features a low-dose n-type implant in the epitaxial layer. This allows for full depletion of the epitaxial layer while using a small collection electrode, which reduces power consumption while improving radiation tolerance. A distinguishing feature of this sensor variant is the *gap* in the n-type implant at the pixel boundaries. This gap enhances the lateral electric field at the edges of the pixels, leading to faster charge collection and reduced charge sharing between neighboring pixels.

5.2.1 Back Bias

For the MOSAIX sensor, minimizing power consumption is essential to keep IR drops in the central region of the chip under control. In MAPS sensors, the analog power consumption scales approximately with the square of the capacitance of the collection electrode [40]. For a parallel-plate capacitor, the capacitance is inversely proportional to the distance between the plates. In a MAPS collection electrode, this effective distance corresponds to the thickness of the depletion region, which acts as the dielectric. Increasing the depletion depth therefore reduces the capacitance of the sensor and, in turn, the analog power consumption. This can be achieved by enlarging the voltage difference between the collection electrode and the surrounding p-well/substrate.

Some pixel detectors accomplish this by electrically isolating the power domains of the collection node and the front-end circuitry [78]. In MOSAIX, however, such decoupling would require an additional capacitor exceeding the intrinsic sensor capacitance, negating the advantages of the low-capacitance electrode. Instead, MOSAIX adopts a back-biasing scheme: the p-well and substrate (bulk) are biased to $-1.2\,\mathrm{V}$, while the collection electrode is held at up to $1.2\,\mathrm{V}$ and connected directly to the front-end input. As a result, all nMOS transistors in the pixel matrix operate with a back bias of $-1.2\,\mathrm{V}$, rather than the conventional $0\,\mathrm{V}$.

Beyond reducing the sensor capacitance, this back bias introduces three key effects on all nMOS transistors in the pixel matrix:

- Reduced leakage power consumption.
- Increased propagation delay.
- Decreased radiation tolerance (explored in Chapter 7).

To account for these effects, the pixel-matrix standard-cell library was re-characterized at $V_{\mathrm{BS}} = -1.2\,\mathrm{V}$ for accurate timing and power analysis.

5.3 Pixel Front-End

The pixel front-end circuitry plays a crucial role in converting the charge collected by the sensor into a digital signal that indicates the presence of a hit. Figure 5.3 shows the front-end architecture used in ITS3.

Each pixel includes a test injection capacitor to emulate hits during characterization, a front-end amplifier to shape and amplify the signal, and a discriminator to determine whether the signal corresponds to a valid particle hit based on a configurable threshold. The discriminator output is asserted only when the amplified signal exceeds this threshold. The ITS3 pixel does not store energy information; it simply detects whether the energy exceeds a configurable threshold, indicating whether the hit is valid or not.

The performance of the front-end is tailored for the detection of Minimum Ionizing Particles (MIPs) and other relevant charged particles in the ALICE experiment. The key performance parameters of the ITS3 pixel are summarized in Table 5.2.

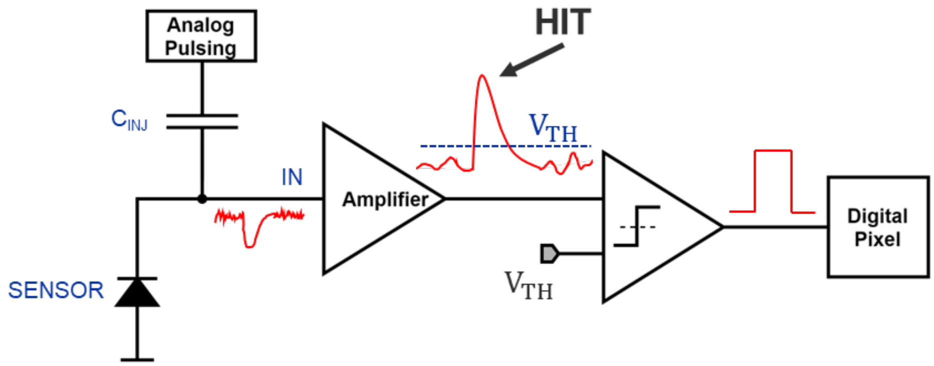


Figure 5.3: Schematic of the pixel front-end of the ITS3.

Table 5.2: ITS3 front-end parameters.

Pixel Width (z-axis pitch)	22.8 μm
Height (r-ϕ axis pitch)	20.8 μm
Pixel Depth (epitaxial layer depth)	10 μm
Recovery time	$< 10\ \mu s$
Time over Threshold	$5 - 24\ \mu s$
Peaking time	$< 1\ \mu s$
Time-walk	$< 200\ ns$
Threshold (configurable)	$\approx 150\ e^-$

5.4 In-Pixel Logic, the Digital Pixel

Once a hit is detected by the pixel sensor and processed by the front-end circuitry, it must be temporarily stored until the readout architecture can handle it. This requires in-pixel logic that provides local, short-term storage, which can be accessed from the tile readout. This logic is implemented using standard cells and is synthesized via a digital design flow, receiving the name of "digital pixel".

5.4.1 Basic Functionality and Requirements

The digital pixel is operated by the readout architecture using predefined time intervals called *integration windows*. During each integration window, the pixel accumulates pixel hit information (integrating) from the discriminator output. All hits collected within an integration window are packaged together into a time-stamped *integration window packet*, which includes both the hits and the associated timestamp (with granularity equal to the integration window period).

To design the digital pixel, several requirements were used. Among them, the most relevant are:

- The digital pixel must respond to the rising edge (not by level) of the discriminator output, to prevent a single hit from being sent in multiple integration window packets due to its Time over Threshold (ToT).

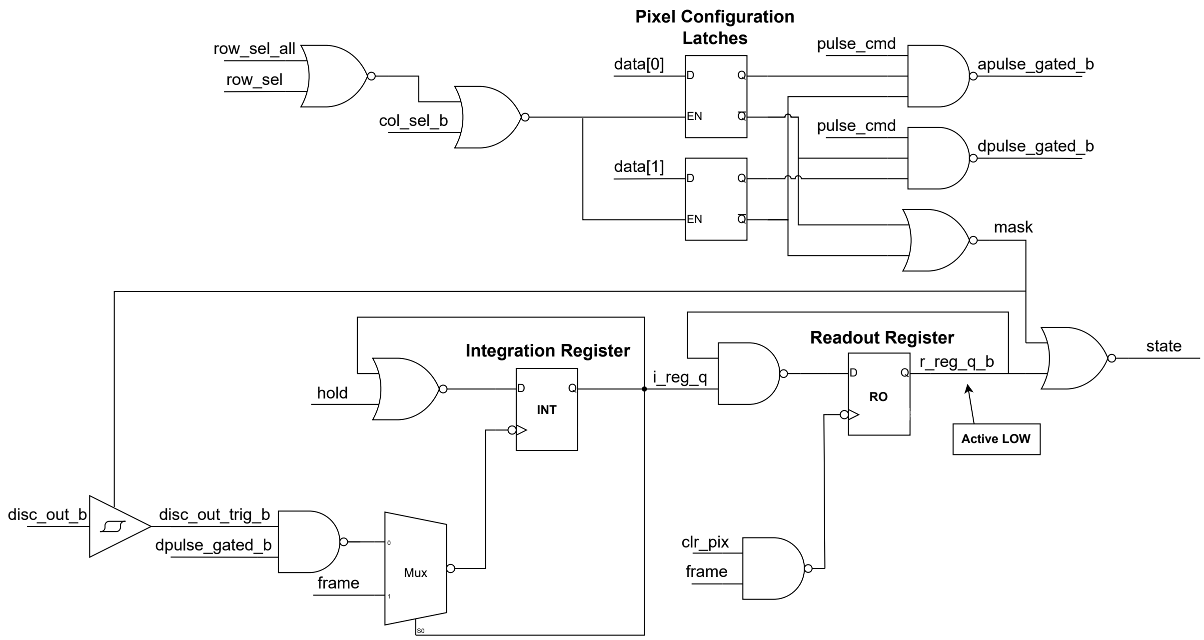


Figure 5.4: Schematic of the MOSAIX digital pixel logic.

- It shall be possible to disable the front-end integration and freeze the current state of the registers.
- The architecture must support the transmission of complete integration window packets, regardless of the number of hits. This is crucial for physics analyses at ALICE, particularly for high-centrality Pb-Pb collisions, which generate the most interesting physics events.
- At least two consecutive integration windows must be readable without loss. This is required to account for the *time-walk* effect in the front-end, where hits from the same collision may be delayed by up to 200 ns.
- In scenarios where the readout architecture cannot process all hits within a single window, it must prioritize the complete readout of the current and next integration windows, even at the expense of subsequent windows. This ensures full capture of high centrality Pb-Pb collisions.

5.4.2 Proposed Digital Pixel Architecture

Figure 5.4 shows the full digital schematic of the digital pixel, while Figure 5.5 illustrates a simplified schematic showing the different interconnections between the digital pixel, the front-end of the pixel, and the double column logic.

As seen, the digital pixel reads the hits detected in the discriminator output (output low) `disc_out_b` and temporarily stores them into two flip-flops called:

- **Integration Register (INT_REG):** In charge of monitoring the status of the

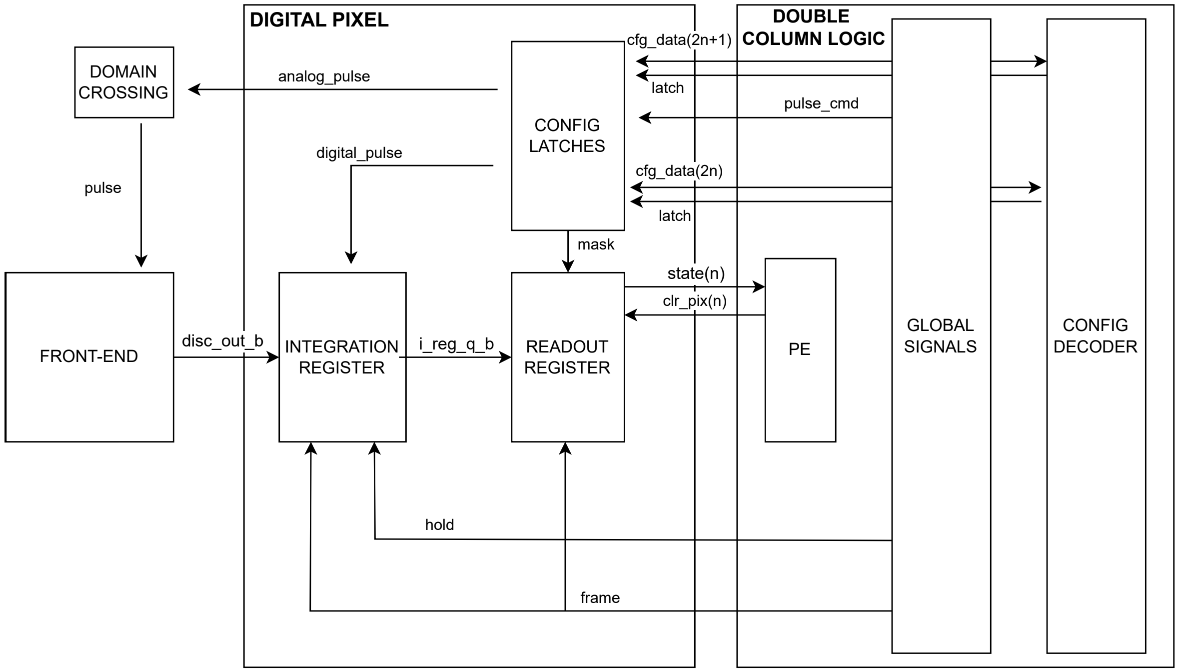


Figure 5.5: Schematic of the ITS3 digital pixel interconnections with the pixel front-end and the double column logic (PE, global signals, and the configuration decoder).

discriminator output, capturing the pixel hit when it occurs.

- **Readout Register (RO_REG):** Holds the pixel hit information from the previous integration window to be read by the on-chip readout logic. At the same time, the INT_REG stores the pixel hits corresponding to the current integration window.

Apart from these two memory elements, two configuration latches are included to set the pixel's operating mode: *normal*, *analog pulse*, *digital pulse*, and *masked*.

The signal `frame` is a periodic global pulse that defines the time boundaries of the integration and readout intervals (integration windows). It acts with the rising edge on the Integration Register and the Readout Register simultaneously, transferring the content of the INT_REG to the RO_REG.

The `hold` is a global signal that disables the integration of the front-end. When asserted, the integration registers will keep their value regardless of the `disc_ out` rising edge. It holds negative values by assigning '0' to the D port of the flip-flop and positive values because the CK port of the flop will not look to `disc_ out` but to `frame`, which will act as a global reset for the integration registers as long as the `hold` is asserted.

The INT_REG is controlled by the following signals:

- Set by a rising edge on `disc_out`.
- Alternatively, set by a rising edge on `dpulse_gated`.

- Level sensitive `hold`, disables the integration of the front-end.
- Cleared on the rising edge of `frame`.

The RO_REG is controlled by the following signals:

- Loaded from the Integration Register on the rising edge of `frame`.
- Cleared on the rising edge of `clr_pix`.

The operation of the pixels is complex and requires a lot of dedicated logic in the periphery. Appendix A presents various scenarios of the operation of the pixel matrix by the readout presented in Chapter 6.

5.5 Conclusions

This chapter presented the architecture and operation of the MOSAIX pixel matrix developed for ITS3. It introduces the overall matrix organization and addressing scheme, the structure and behavior of the pixel sensor, the pixel front-end, and the in-pixel digital logic.

The designed ITS3 pixel has a pitch of 22.8×20.8 μm, integrating the sensor, analog front-end circuitry and two digital memory elements that store hit information from the current and previous integration windows respectively.

The pixels are organized into double columns, each equipped with a fully combinatorial zero-suppression PE. The double-column encodes the addresses of the pixel hits, and routes clear pulses to the selected pixel, enabling a sequential readout at 40 MHz. The double-column outputs are then forwarded to the periphery, where data are processed, packed into integration windows, and delivered to the LEC.

Chapter 6

MOSAIX Tile Readout Architecture Design and Implementation

Within each MOSAIX tile, pixel hits are transferred from the pixel array to the periphery, grouped into timestamped window packets, and then transmitted to the Left Endcap (LEC). The readout architecture responsible for this process is implemented in the tile periphery, following the results from the readout model described in Chapter 4.

This chapter presents the design and implementation of the readout architecture in a 65 nm CMOS commercial imaging technology. It begins with an overview of the complete MOSAIX system and the top-level constraints. Then, the readout scheme is introduced, followed by detailed implementation and power-reduction techniques.

6.1 MOSAIX Structure

The MOSAIX chip is divided into 12 RSUs, each containing 12 tiles, for a total of 144 tiles. Each tile measures 3.591 mm $\times$ 9.782 mm and operates as an independent sensor block with its own power domain, control interface, and dedicated data link to the LEC. As discussed in Chapter 4, all tiles function autonomously, reading their pixel matrices and transmitting data to the LEC via dedicated serial links.

Figure 6.1 illustrates the schematic of the bottom half of an RSU, including six tiles, two service nodes, and two stitched backbones (SBBs). To ensure a synchronized operation, all tiles must share a common timestamp. This is achieved using a global synchronization signal (`sync`), which coordinates the start and end of integration windows across tiles. The `sync` and other signals common across tiles, such as the `clock` and `reset`, are distributed globally from the LEC.

The MOSAIX clocking scheme is multi-domain. Most LEC modules operate at 320 MHz, but they also include eight high-speed output channels that support up to 10.24 Gb/s. A 160 MHz clock is distributed from the LEC to all tiles. Inside each tile, this clock is divided by four to obtain a 40 MHz base clock. Inside the periphery, the 40 MHz base clock is generally

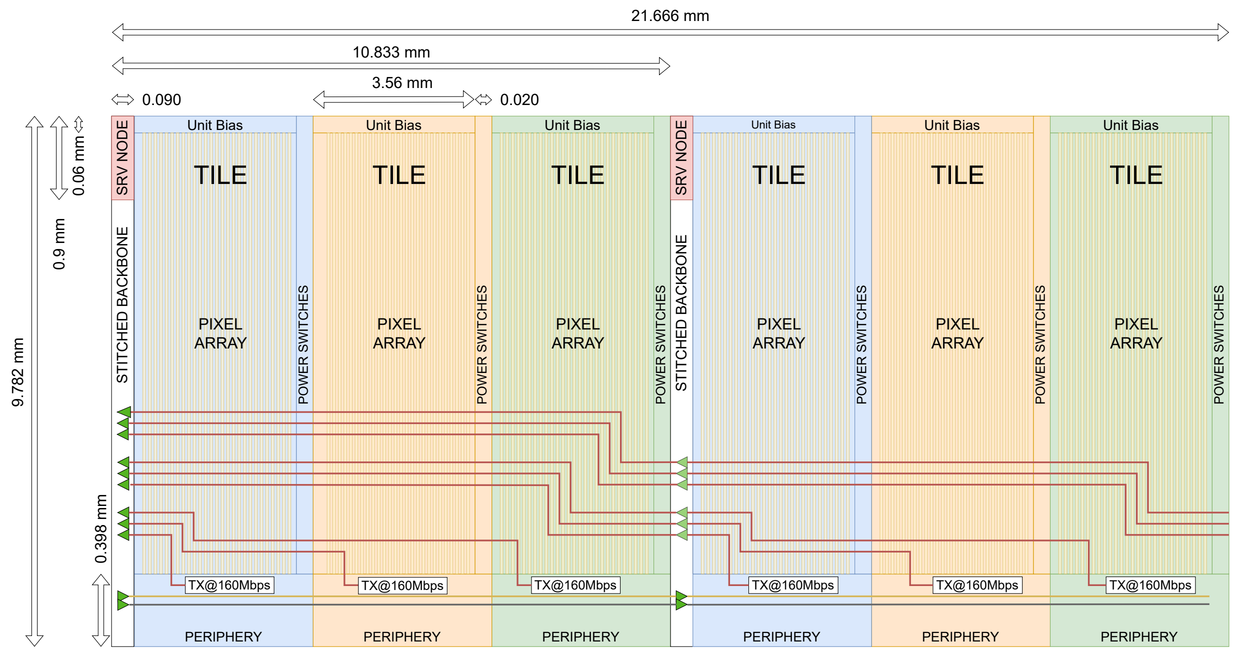


Figure 6.1: High-level block diagram of half RSU, composed of six tiles, two SBBs, and two service nodes (SRV NODE).

used. However, the serializer responsible for data transmission operates at 160 MHz, requiring clock-domain crossing (CDC) dedicated FIFOs and synchronizers to communicate the two clock domains.

The maximum distance between the LEC and a tile can be up to 266 mm. To maintain signal integrity in the signals between LEC and tiles, periodic buffers are required. For asynchronous signals such as the `sync` and the `reset`, normal digital buffers are placed in the tiles with minimal effort. In contrast, the 160 MHz clock and 160 Mb/s data signals demand a more complex buffer scheme. For the clock signal, a custom full-swing buffer scheme is used. For data signals, full-swing buffers could not be used because these paths are routed on top of the pixel arrays. Driving high-speed full-swing lines over the pixel array risks coupling noise into the collection electrodes, generating false hits. To address this, a custom low-swing buffer was developed. Both clock and data are repeated in the SBB, placed on every three tiles (approximately 10.8 mm).

Given the unprecedented size of MOSAIX, some tiles may fail in fabrication. Each tile is therefore powered through its own local domain, which can be independently disabled via a dedicated power switch controlled by the service node. Service nodes are placed every three tiles, following the same distribution as the SBBs.

The control and configuration of tiles and service nodes are handled through a 5 Mb/s slow-control network of the LEC. Due to the low data rate, simple buffers suffice to maintain signal integrity. This network transmits configuration commands and can also read back the status of nodes and tiles.

Regarding the tile, it is divided into four main submodules:

- **Pixel Matrix** Contains the active sensing area of the chip, i.e., the pixels. In addition, it also includes all the logic for pixel readout and configuration. A detailed description of the components of the pixel matrix is provided in Chapter 5.
- **Unit Bias** Analog block responsible for configuring the various bias currents required by the analog front-end of each pixel, such as the discriminator thresholds or the pulse injection currents.
- **Power Switches** Switches that control the local power domain supply of each tile.
- **Periphery**: Contains the logic for control and data readout. It receives global and control signals (via the SBB), configures the unit bias, reads the pixel matrix, and assembles the integration window packets sent to the LEC.

6.1.1 MOSAIX Top Implementation

The RSU, LEC, and REC are implemented and fabricated separately, each using independent design flows. These sub-chips are later combined through stitching, where additional metal layers are added across their boundaries to form a MOSAIX segment, as illustrated in Figure 1.6.

The implementation of the RSU follows a digital-on-top methodology. Within it, the SBBs, the service nodes and the tiles are implemented separately and then instantiated at the top level as shown in Figure 6.1. The tile itself is also implemented using a digital-on-top flow, with independent sub-flows for the pixel matrix, unit bias, power switches, and periphery.

The MOSAIX implementation presents several challenges:

- **Yield:** A MOSAIX chip can reach dimensions of up to 266 mm × 93 mm. Due to the large area of the MOSAIX, there is a high likelihood of shorts in the circuit.. Therefore, a DFM robust design is required at both the standard cell library and top levels.
- **Power Consumption:** Each MOSAIX segment measures 26.6 cm, and it is powered only from its two edges (LEC and REC). This power scheme causes a large IR drop in the center of the MOSAIX segment, expecting to go from 1.3 V to 1.1 V as shown in

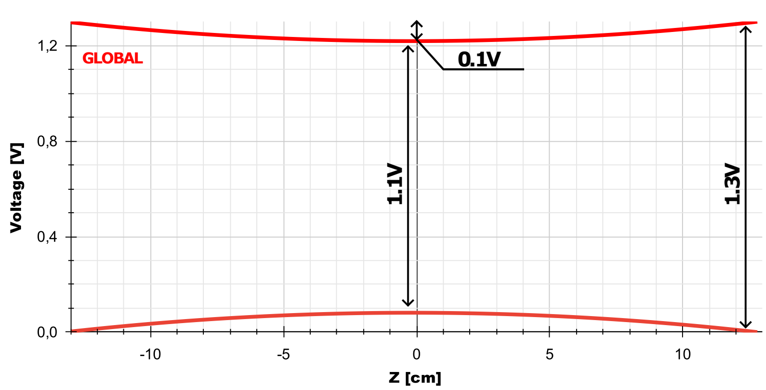


Figure 6.2: Expected worst case IR drop in the MOSAIX over z. At the center, a 0.2 V drop is expected, resulting in a 1 V supply voltage. (Source: [79])

Table 6.1: Analysis corners used for the tile power and timing analysis. Corners are based on voltage, Layout Parasitic Extraction (LPE), Quantus QRC tech file (resistance and capacitance extraction), and temperature.

corner name	LPE corner	QRC tech file	voltage (V)	temperature (C)
TYP	tt	typical	1.20	27
MAX	ss	CWORST, RCWORST	1.08	65
MIN	ff	CBEST, RCBEST	1.32	-20
SKEW1	ff	CWORST, RCWORST	1.32	65
SKEW2	ss	CBEST, RCBEST	1.08	-20

Figure 6.2. These maximum, minimum, and typical voltages (with a 0.02 V margin) are used in the delay corners used in the physical implementation of the tile. If the IR drop in the central tiles is higher than estimated, the timing analysis performed under these corners will no longer be valid. To avoid this problem, the power consumption from all the modules must be reduced as much as possible.

- **Area Constraints:** The area allocated for the periphery logic follows a T-shaped footprint measuring 3.71 mm $\times$ 0.4 mm. All digital logic must be placed and routed within this area. Increasing the footprint would increase the non-sensitive (dead) area of the chip.
- **Timing Closure:** The periphery must support reading the pixel matrix at 40 MHz (25 ns period).

These challenging design constraints drive the design and implementation of all MOSAIX modules. In the case of the pixel matrix, it requires the generation of a full custom standard cell library to reduce the leakage current and enhance the DFM robustness. More information on this low-leakage DFM robust standard cell library can be found in the Appendix D.

From the perspective of the periphery, the most challenging constraint was the timing closure of the arcs between the periphery and the pixel matrix. This requires clearing a pixel signal, propagating it over a 9.2 mm priority encoder, and returning another 9.2 mm back to the periphery, all within the 25 ns window. The system must be ready to handle the next read cycle immediately, without dead time. An in-depth analysis of these timing arcs is presented in Section 6.3.3.

Additionally, the periphery requires several optimizations to adapt to the allocated area, enhance DFM robustness, and reduce the power consumption.

6.2 Periphery Structure

Figure 6.3 details the internal components of the periphery, including:

- **SBB Tile Bridge**: Interfaces between the tile and the SBB. Since the tile and SBB belong to separate power domains (local vs. global), this bridge ensures a correct signal connection.

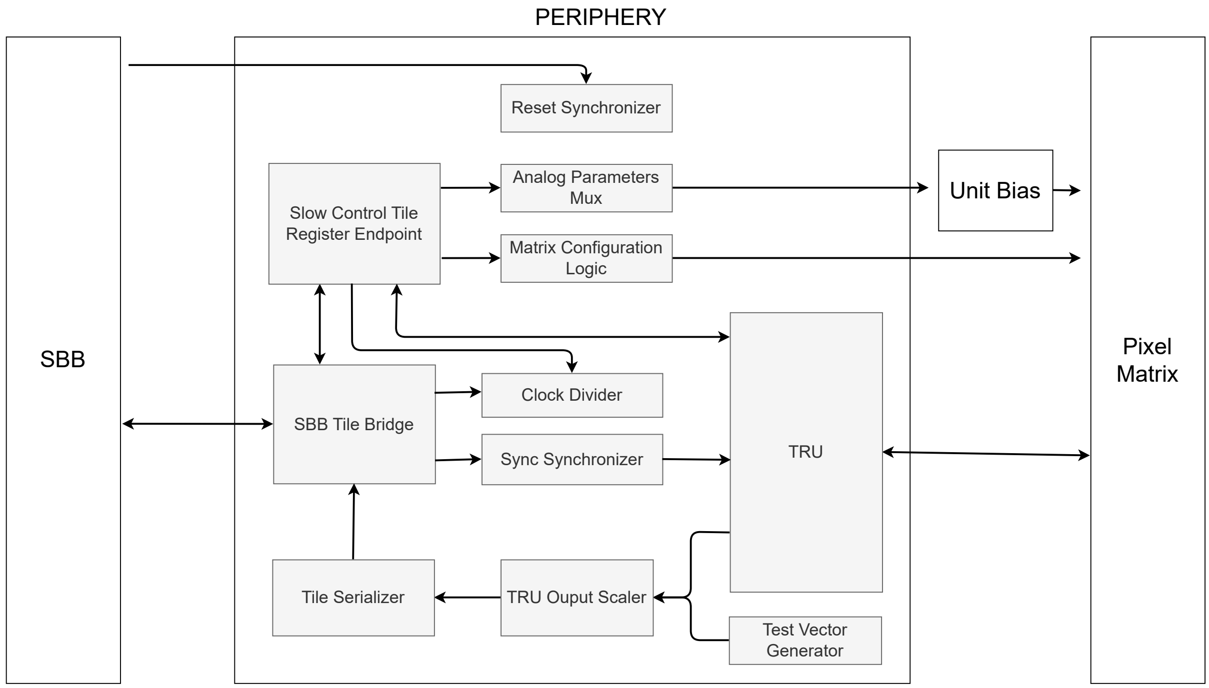


Figure 6.3: High-level block diagram of the tile periphery components.

- **Clock Divider**: It takes the 160 MHz clock signal from the SBB and divides it by 4, distributing the resulting 40 MHz signal to all components except the Tile Serializer.
- **Slow Control Tile Register Endpoint**: Receives slow control commands from the LEC and stores them in internal configuration registers.
- **Analog Parameters Mux**: Routes analog biasing configurations from the slow control registers to the Unit Bias.
- **Matrix Configuration Logic**: Adapts and routes configuration data from slow control to the Pixel Matrix.
- **Sync Synchronizer** A two-stage synchronizer that aligns the global `sync` signal with the 40 MHz periphery clock.
- **Reset Synchronizer**: Similar to the Sync Synchronizer, but it aligns the `reset` signal.
- **Top Readout Unit (TRU)**: This central block is responsible for reading data from the Pixel Matrix and assembling it into packets. Operates at 40 MHz and is controlled by slow control signals and the global `sync`.
- **Test Vector Generator**: Generates a pseudo-random binary sequence (PRBS) for link testing, replacing real data from the TRU when activated.
- **TRU Output Scaler**: Converts 16-bit parallel TRU or Test Vector Generator into chunks of 4-bit suitable for the Tile Serializer.
- **Tile Serializer**: Serializes the 4-bit chunks from the TRU Output Scaler and sends it

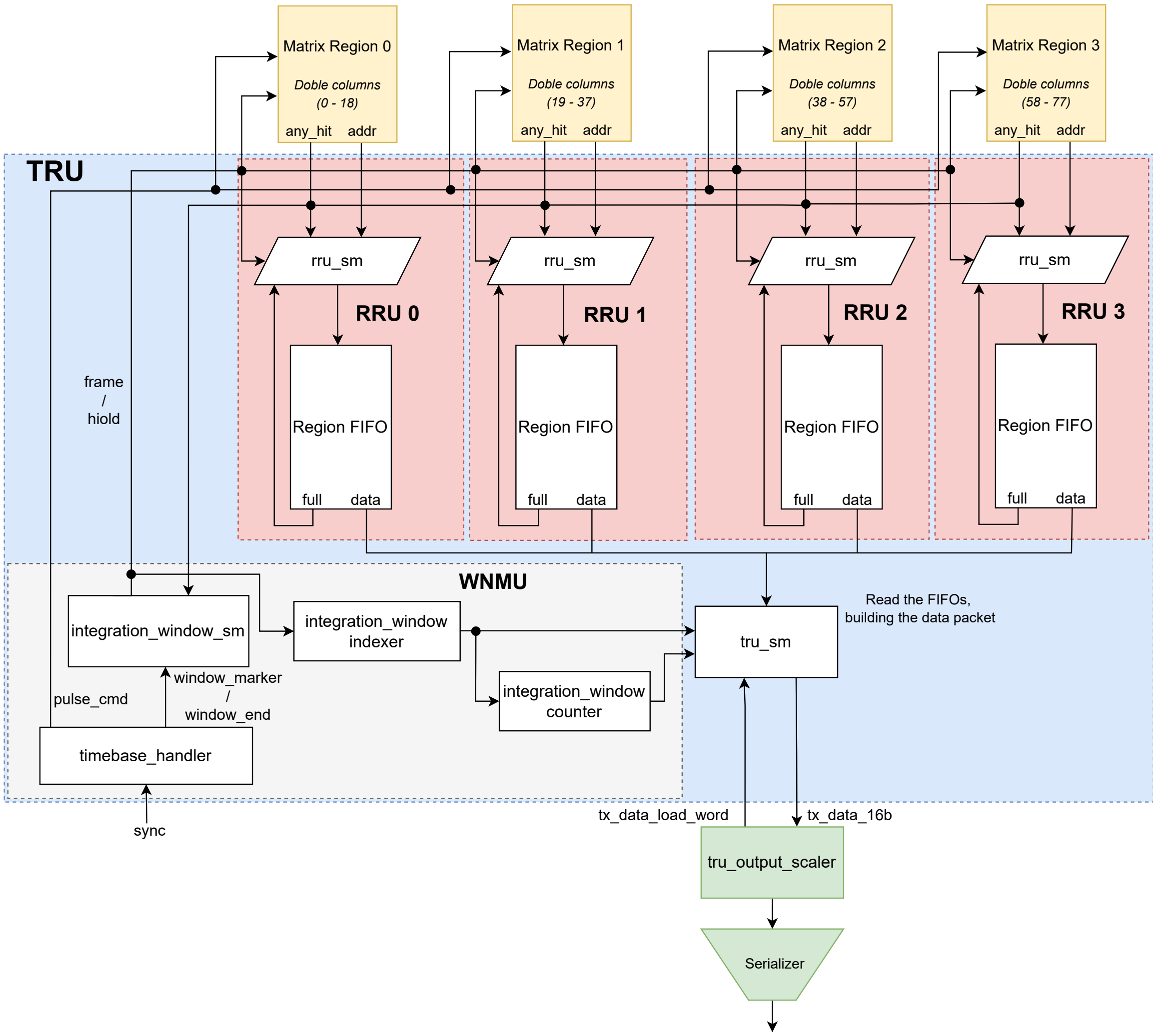


Figure 6.4: High-level block diagram of the TRU.

to the SBB. It includes a Clock Domain Crossing (CDC) FIFO to convert data from the 40 MHz tile domain to the 160 MHz SBB domain.

The TRU is the heart of the readout system and forms the core of this thesis. Its design is based on the theoretical readout model presented in Chapter 4, which analyzes both the protocol design and the hardware specifications.

6.2.1 Top Readout Unit (TRU)

The TRU is responsible for reading data from the pixel matrix and assembling it into output integration window packets. The design and implementation of the TRU followed a set of requirements presented in the Appendix B. A high-level block diagram of the TRU architecture is shown in Figure 6.4.

As illustrated, the TRU is made up of several submodules: The Window Management Unit (WNMU), four Readout Region Units (RRUs), and the `tru_sm`.

The WNMU is responsible for time management, receiving the `sync` pulse, and generating the integration windows. Itself, the WNMU is subdivided into:

- **timebase_handler:** Receives the `sync` pulse and generates the integration window delimiters (`window_marker`, `window_end`), and the pulse command to the pixel matrix (`pulse_cmd`) with respect to the `sync` pulse.
- **integration_window_sm:** Receives the integration window delimiters (`window_marker`, `window_end`) and controls the pixel matrix, propagating `frame` pulses or asserting the `hold` when needed.
- **integration_window_indexer:** Tracks the number of integration windows that remain to be processed in the FIFOs, keeping track of which ones were dropped and which ones were not.
- **integration_window_counter:** Generates the timestamp to be embedded in the data packets transmitted.

The four RRUs divide the pixel matrix into four spatial regions, which are read through the `rru_sm` finite state machine into a local `region FIFO`. Since the 78 double columns of the matrix are not evenly divisible by 4, the distribution of double columns across RRUs is non-uniform:

- Region 0: 19 Double columns (0 - 18)
- Region 1: 19 Double columns (19 - 37)
- Region 2: 20 Double columns (37 - 57)
- Region 3: 20 Double columns (58 - 77)

The TRU top finite state machine (`tru_sm`), aggregates the data and constructs the output packet, which is then forwarded to the Tile Serializer for transmission.

A detailed explanation of the RTL design of the TRU is presented in the Appendix C.

6.2.2 Speed Throughtling

As the timing closure for the MOSAIX project was one of the main challenges, an alternative mode of operation was included to limit this risk. This mode allows the readout speed of the pixel matrix and SBB to be throttled to half in the case that the timing could not be closed:

- **Pixel Matrix:** Could be operated at 40 MHz or 20 MHz. Regardless of the pixel matrix readout speed, the clock tree will propagate a 40 MHz clock. To throttle the speed to 20 MHz, the `rru_sm` will toggle only half of the cycles.
- **SBB:** Bandwidth can be 160 Mbps or 80 Mbps. The slower mode duplicates bits via the output scaler, e.g., transmitting `1100_0011` instead of `1001`, effectively halving the data rate.

This approach simplifies implementation by avoiding complex clocking schemes while ensuring that the MOSAIX can be operated even if the timing could not be met.

6.3 Periphery Physical Implementation

As shown, the periphery integrates several blocks, such as the Slow Control Tile Register Endpoint, the Tile Serializer, or the Top Readout Unit (TRU), which are physically implemented together within one flow. To improve the timing, power consumption and DFM robustness of the design, several techniques were applied in the various stages of the physical implementation.

6.3.1 Floorplan

The periphery is allocated in a rectangular area of 3.56 mm $\times$ 0.371 mm. However, two squares of 0.31 mm $\times$ 0.27 mm are cut from the horizontal edges to accommodate extra power pads, resulting in a T-shaped floorplan. These pads will only be used for testing the MOSAIX in the case that the IR drop or any other power problem does not allow powering from the REC and the LEC. These paths will be removed in the final ITS3 iteration as they cannot be used in the final ITS3 installation.

To support timing closure, three types of special boundary constraints can be applied during floorplanning that assist the tool in cell placement. These boundaries are:

- **Guide:** A soft placement hint. When a module is moved or placed within a guide area, its cells are **preferentially** placed inside. However, cells from **other modules may still be placed** inside if the area is underutilized, or placed outside if overutilized.
- **Region:** A moderately strict constraint. All cells of the specified module **must be placed inside** the defined region. Cells from **other modules may still be placed inside** if space permits, but cells from the defined module cannot be placed outside.
- **Fence:** The strictest constraint. **Only** the specified modules cells can be placed inside the fence, and they must remain entirely within its boundaries. No external cells are allowed in.

Figure 6.5 shows the floorplan of the tiles periphery. As seen, all three boundary types are used to improve the physical implementation output. Guides and regions helped improve timing closure between the pixel matrix and the periphery by placing the periphery logic that drives the pixel matrix arcs as close as possible to the pixel matrix in-out pins (logic described in Section 6.3.3). Guides were also used to place the region FIFOs away from the pixel matrix interface. A fenced region was reserved at the bottom-center of the layout for the SBB tile block, which was implemented as an analog macro.

6.3.2 DFM strategies

The pixel matrix is implemented using a 13-track DFM, low-leakage standard cell library presented in the Appendix D. In the periphery, a modified sub-variant of this library is

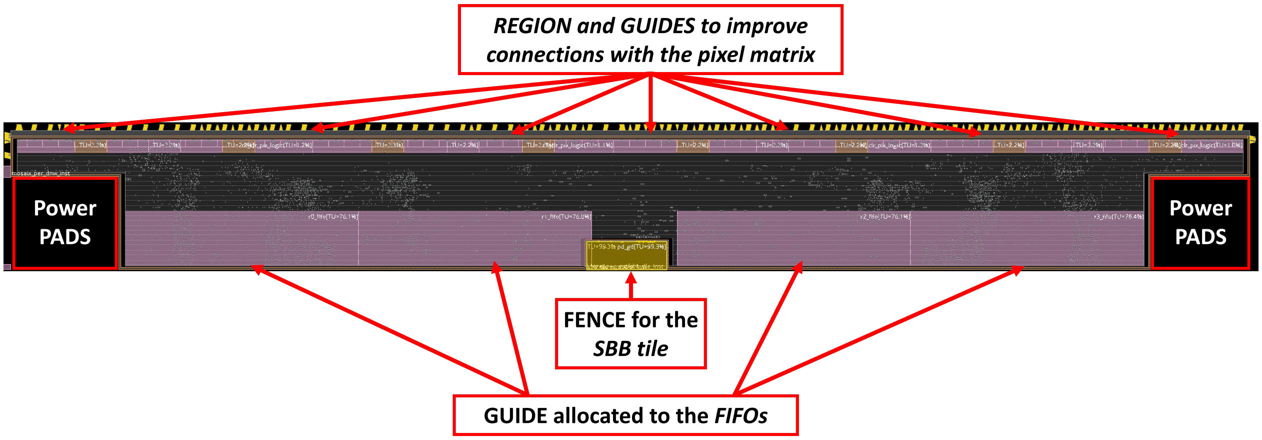


Figure 6.5: Tile periphery floorplan. Guide boundaries are marked in pink, region boundaries in light yellow, and fenced boundaries in bright yellow.

Table 6.2: Distribution of single-cut and multi-cut across layers.

Layer (cut)	Single-cut	Multi-cut	Total
M1-M2 (V1)	138,525 (45.8%)	164,155 (54.2%)	302,680
M2-M3 (V2)	4,415 (1.5%)	288,951 (98.5%)	293,366
M3-M4 (V3)	1,183 (1.7%)	70,372 (98.3%)	71,555
M4-M5 (V4)	0 (0.0%)	1,834 (100.0%)	1,834
M5-M6 (V5)	4 (15.4%)	22 (84.6%)	26
Total	**144,127 (21.5%)**	**525,334 (78.5%)**	**669,461**

employed.

This variant presents:

- The gate length is set to the minimum value, since leakage is not a concern in this region.
- The nMOS substrate is tied to *VSS*.
- Width and spacing rules are slightly less DFM-optimized than those of the pixel matrix library, but still significantly larger than in the original 12-track standard cell library.

To further enhance the DFM robustness of the periphery, while limiting the impact on area utilization, two additional strategies were applied.

On the one hand, the recommendations from the foundry emphasized the importance of DFM, limiting the use of single-cut vias. As making all vias multi-cut vias required a lot of area as well, the usage of multi-cut vias was prioritized for the vias that connect the metals higher than Metal 2, as shown in Table 6.2.

On the other hand, the trunk of the clock tree (the top parts of the clock tree, which are not directly connected to any cell) has twice the minimum metal width. Additionally, the space between the routing of the clock tree and any other metals was extended to be twice the minimum.

6.3.3 Closure of Pixel Matrix Delay Arcs

The periphery interacts with the pixel matrix through multiple signals. Several are configuration signals, while six are used during normal operation:

- From periphery to matrix: `frame`, `clr_pix`, `hold`, `pulse_cmd`
- From matrix to periphery: `addr`, `hit_valid`

Each periphery output is connected independently to its corresponding double column within the pixel matrix, as illustrated in Figure 6.6.

As seen, `hold` and `pulse_cmd` (in green) have relaxed timing constraints, as they only have to be routed one way of the 9.2 mm of double column until the integration registers of the digital pixel.

On the other side, `frame` and `clr_pix` (in red) are part of critical paths. These signals are combinatorially linked through the pixels readout registers and, therefore, must reach the pixel, clear it, and return to the periphery through the `addr` and `hit_valid`.

As seen, several combinatorial paths exist between the periphery and the pixel matrix:

- `frame` ⇒ `addr`
- `frame` ⇒ `hit_valid`
- `clr_pix` ⇒ `addr`
- `clr_pix` ⇒ `hit_valid`

To reduce the timing arcs, the `addr` and `hit_valid` inputs are pipelined immediately after the input pins. This can be seen with the 78 top pink guide boundaries located below each double column in Figure 6.5.

The `frame` pulse is propagated to all pixels in the matrix region. To improve its timing, the fanout is reduced by half, having two independent flip-flop origins located near the pixel matrix connections. These two flops are located between the group of double columns to `frame` using the light yellow region boundaries highlighted in Figure 6.5.

The `clr_pix` output is the hardest arc to close. Its critical path does not start from the `clr_pix` mechanism explained in Appendix C.4. Instead, the critical path starts in the *hit_valid* pipeline flop as it must be used to propagate the `clr_pix` pulse to the first double column that has a valid hit. This column encoder mechanism was also partitioned into two paths to improve its timing arc. All these mechanisms are also placed inside a guide boundary near the pixel matrix connections, as seen in Figure 6.5.

With all these optimizations and time-aware design, the timing arcs between the pixel matrix and the periphery were successfully closed in less than 24 ns for the worst arc in the worst analysis corner, giving a nice 1 ns extra margin in the case that something was not calculated correctly.

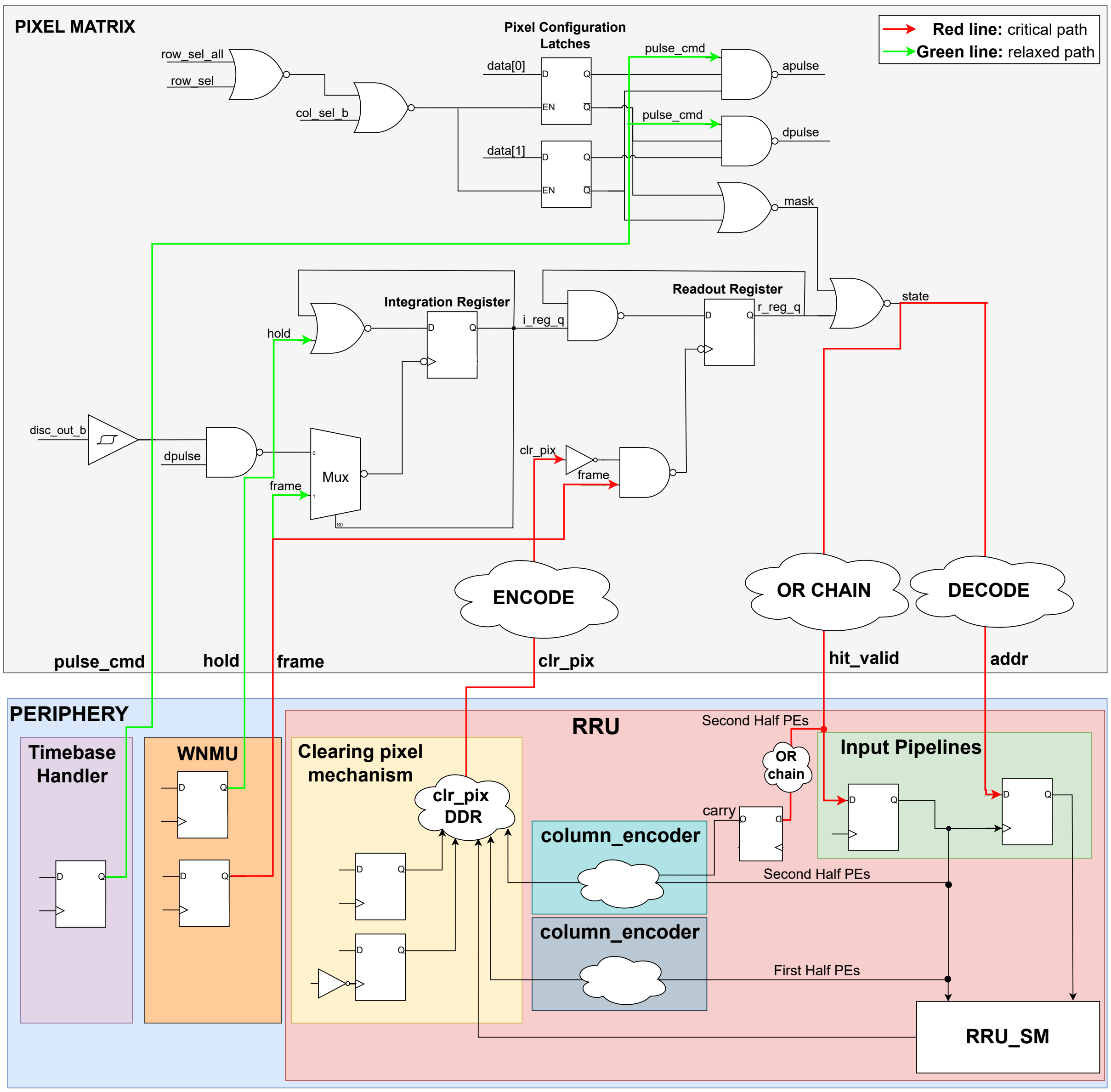


Figure 6.6: Signal connections between the pixel matrix and the periphery.

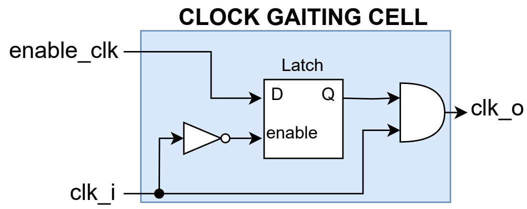


Figure 6.7: Schematic of the clock-gating cell used for power reduction.

6.3.3.1 SDC of Matrix Connections

As the periphery and the pixel matrix have independent implementation flows, these timing arcs are not correctly analyzed or balanced by the Innovus EDA tool.

Since Innovus does not detect these loops, the reg-to-reg paths must be manually constrained by breaking the arc into two parts:

- **Launch path:** from the flop in the periphery until the readout register in the pixel.
- **Capture path:** to the readout register in the pixel until a pipeline flop in the periphery.

To constrain these paths, the setup and hold analysis must account for the delay introduced on one side, while manually balancing on the other side.

6.3.4 Power Reduction Techniques

To meet the stringent power constraints of the MOSAIX chip, clock-gating strategies were applied in several modules of the TRU. This clock-gating was implemented using the clock-gating cell shown in Figure 6.7, composed of a D latch and an AND gate.

One of the clock-gated modules is the `tru_sm`. This FSM toggles only once every 4 or 8 clock cycles, depending on the serializer speed, as it just provides words to serialize when the serializer has finished the previous one. As a result, clock gating reduces dynamic power consumption by up to 75%. Similarly, the WNMU logic operates only when receiving a `window_marker`, FIFO full conditions, or indexer events, making it an ideal candidate for aggressive clock gating based on its internal status and input conditions.

In the region FIFO, clock gating was applied such that only a single memory cell receives a propagated clock signal at a time. Additionally, the FIFO was implemented using latches instead of flip-flops, as the latch-based design demonstrated significantly lower area and power usage.

The `rru_sm` was not gated to avoid introducing any power-dependent behavior based on the number of hits to be read from the matrix.

A major contributor to dynamic power consumption was the `addr` pipelines described in Section 6.3.3. Each `addr` of each PE is 10 bits wide. Multiplying it by the 78 PEs of the pixel matrix, we get 780 flip-flops.

To reduce their power consumption, only 2 columns are clocked during the readout of any given region: the currently active PE and the next PE to be read. This selective activation strategy reduces the number of active flip-flops from 780 to just 80 (10 bits × 4 regions × 2 PEs active), achieving nearly a 10× reduction in dynamic power without introducing power dependencies with the number of pixel hits to be read.

6.3.5 Periphery Implementation Results

After applying all the proposed implementation optimizations, the periphery successfully met the timing closure at 40 MHz, with a timing margin exceeding 1 ns at all corners. The implementation area fits within the allocated region of 3.591 mm × 0.371 mm shown in Figure 6.1.

In the typical corner, the peripherys average power consumption was measured at 6.24 mW under a high hit rate 2 μs integration period scenario, as described in Chapter 3. This corresponds to a power density of 3.59 mW/mm^2, which remains within the MOSAIX power budget. Additional scenarios with low and zero hit density were also evaluated. The reported power differed by approximately 5 % across these cases, as clock gating was designed to be independent of hit rate.

6.3.6 Conclusions

In this chapter, the RTL and physical implementation of the MOSAIX readout periphery were presented. The adopted approach was a digital-on-top, fully modular solution in which smaller-scale designs are replicated multiple times to cover the full wafer area. Due to the complexities of the ITS3 design being the first wafer-scale detector ever deployed in the field of HEP, several challenges required dedicated attention.

Given the wafer-scale dimensions, yield was a primary concern. To maximize yield, multi-cut vias, inter-metal spacing, and metal path widths were increased at both the standard-cell level and the EDA implementation flow. Moreover, because these enhancements alone may still not guarantee fully functional wafer-scale detectors, the MOSAIX allows for tiles to be switched off on demand. This enables a power granularity capable of isolating defective tiles without compromising the entire wafer.

As the chip is powered only from its short edges, IR drops can be significant at the wafer center, leading to malfunctions of the chip. To mitigate this, power consumption was thoroughly analyzed and reducedfocusing on leakage mitigation in the pixel matrix through a dedicated set of low-leakage, DFM robust standard cells, and on sequential power reduction in the periphery using clock gating.

As a result, the MOSAIX chip is expected to achieve acceptable yield and power consumption, aspects which will be evaluated in detail during the first tests, scheduled in early 2026.

Chapter 7

ITS3 Technology Characterization under Radiation

In the ALICE experiment, pixel detectors are subjected to high levels of radiation. To determine whether a specific semiconductor technology can tolerate this irradiation, controlled experiments must be conducted in which transistors are exposed to similar radiation conditions, allowing their behavior to be analyzed.

For the 65 nm technology that ITS3 uses, an initial irradiation test was performed to assess its radiation tolerance [80]. However, this test did not account for a critical operating condition, the reverse back bias. Within the pixel matrix, all logic circuitry is biased with an nMOS bulk voltage below 0 V. This back-biasing creates an electric field that could influence the transistor's response to irradiation.

To investigate the impact of back bias under irradiation, this chapter presents results from three independent radiation experiments conducted in three Transistor Test Structures (TTS) fabricated in the 65 nm technology of the ITS3. The devices were irradiated under the following conditions:

- **Irradiation 1:** $V_{\mathrm{BS}} = 0\,\mathrm{V}$ at room temperature (25 °C)
- **Irradiation 2:** $V_{\mathrm{BS}} = -2\,\mathrm{V}$ at room temperature (25 °C)
- **Irradiation 3:** $V_{\mathrm{BS}} = -2\,\mathrm{V}$ at low temperature (−30 °C)

Each irradiation test was performed in incremental steps, with ID–VG measurements taken after each step under both biased and unbiased conditions. These measurements enabled detailed tracking of the device's behavior as the radiation dose increased. The results show that the bias voltage applied during measurement significantly affects the behavior of charges trapped in the Shallow Trench Isolation (STI). In addition, devices irradiated under back-bias conditions experienced a notable increase in leakage current.

7.1 Principles of Radiation Effects in Semiconductor Devices

Semiconductor devices can malfunction when exposed to interactions with certain particles [81], [82]. These interactions are collectively referred to as radiation effects in electronics. The radiation effects in semiconductor devices can be classified as either cumulative or stochastic.

7.1.1 Cumulative Effects

Cumulative effects occur due to the gradual degradation of semiconductor devices resulting from repeated exposure to radiation. These effects result from the accumulation of small, individual interactions and can be caused by either ionizing or non-ionizing radiation.

7.1.1.1 Non-Ionizing Effects

Non-ionizing effects, also known as displacement damage, occur when energetic particles (typically neutrons) displace atoms from their original lattice positions within the semiconductor crystal. Displacement damage primarily affects the minority carrier lifetime in silicon substrates, electrons in p-doped substrates, and holes in n-doped substrates. This degradation is particularly critical in bipolar devices, which rely heavily on minority carrier transport. In contrast, MOS devices are less sensitive to minority carrier lifetime, making displacement damage a secondary concern compared to ionizing radiation effects.

7.1.1.2 Ionizing Effects

As explained in Section 2.1, when a charged particle or a photon traverses a material, it can generate electron-hole pairs. To quantify the cumulative ionizing radiation that a silicon device is subjected to, we use the metric called Total Ionizing Dose (TID). The TID measures the energy deposited through ionization, expressed in grays (Gy = J/kg) or radiation absorbed dose (1 Gy = 100 rad). Depending on the application, the electronic components must tolerate different TID levels: Several krad for satellites, Mrad for deep-space probes or in the ALICE experiment at the LHC, or up to 1 Grad for the high-luminosity upgrades of ATLAS and CMS experiments [83].

In CMOS transistors, the electron-hole pairs generated within the silicon substrate have temporary effects (explained in Section 7.1.2) but no long-term effects. However, modern transistors also include insulating layers made of silicon dioxide (SiO_2) or silicon nitride (Si_3O_4), which are sensitive to long-term ionization effects as seen in Figure 7.1. The three most important silicon insulators for ionization effects in commercial CMOS transistors are:

- **Gate oxide**: Separates the polysilicon gate from the silicon substrate.
- **Spacers**: Used during fabrication to define lightly doped drain (LDD) extensions. The LDD extension mitigates short-channel effects, reduces hot carrier injection, and improves reliability.
- **Shallow Trench Isolation (STI)** or **Local Oxidation of Silicon (LOCOS)**: Used

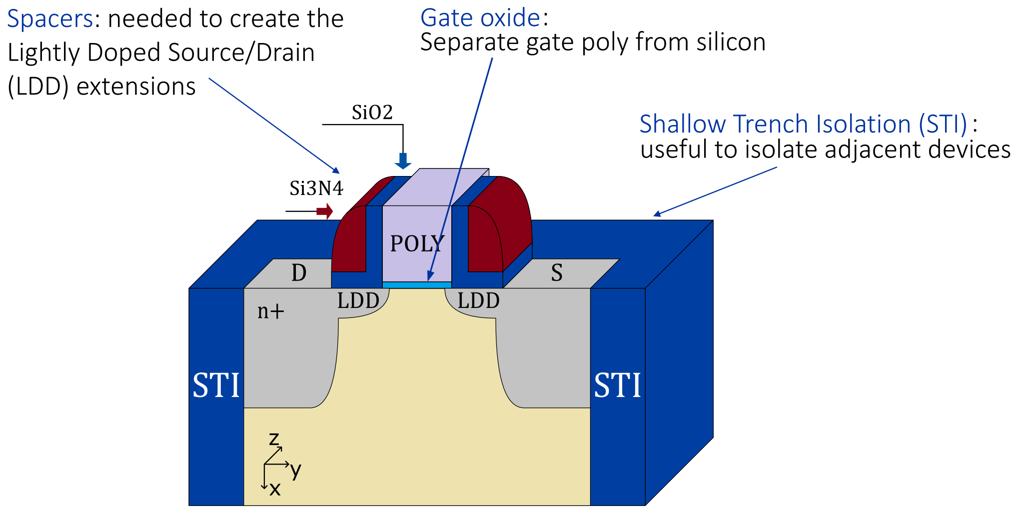


Figure 7.1: Location and types of silicon insulators in planar CMOS transistors (Source [84]).

to electrically isolate adjacent devices.

Exposure to ionizing radiation in these oxide regions leads to the creation of charge traps, which fall into two categories: oxide traps and interface traps. Oxide traps can be further classified into immobile oxide charge and mobile ions. Immobile oxide charges arise from structural defects or contamination, such as oxygen vacancies in Si-Si bonds that can trap positive charges, breaking this Si-Si bond. In contrast, mobile ions are generated when hydrogen atoms (injected during the fabrication of the chip) capture a hole. These ions are not static, and they can move around under the influence of electric fields or thermal energy.

As the mobile ions migrate toward the Si/SiO_2 interface, they can reach pre-existing bonding imperfections and de-passivate them. The nature of these interface traps, whether they capture electrons or holes depends on the doping type of the underlying silicon and the electric field they are submitted to. In the interfaces between SiO_2 and p-doping silicon and a positive voltage in the oxide with respect to the silicon, the interface traps capture electrons. Contrary, in interfaces with n-doping silicon and a negative voltage in the oxide with respect to the silicon, they trap holes.

When a transistor is irradiated, the radiation-induced effects at the oxide traps are the first to dominate. However, mobile ions moved over time until depasivating interface traps, generating a second type of radiation effect that overlaps with the oxide traps effect.

The extent and evolution of these effects depend on multiple environmental and operational parameters. The total number of non-recombined electron-hole pairs generated by radiation is influenced by the dose rate and the material properties of the oxides. A high electric field within the oxide promotes the separation of these pairs and suppresses their recombination, increasing the probability of charge trapping. Temperature also plays a dual role. It enhances the mobility of charge carriers and ions, facilitating their movement toward sensitive regions such as interfaces. However, a high temperature can also increase recombination rates, reducing the pair production.

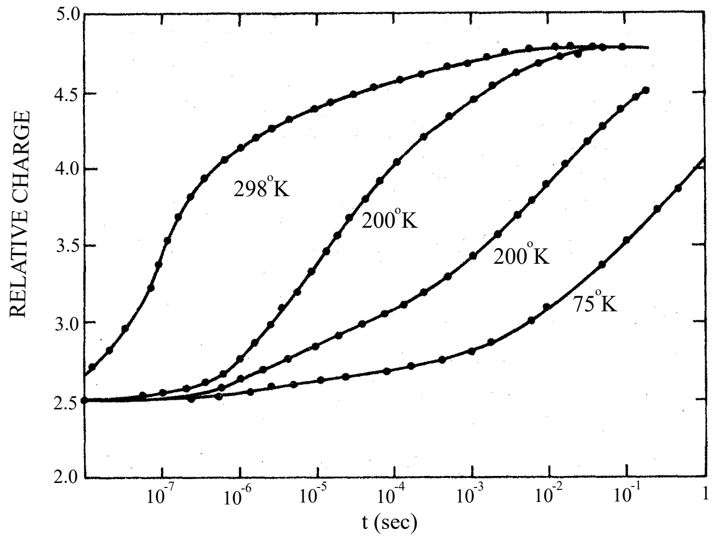


Figure 7.2: Integrated current over time in logarithmic scale following a 3 ns X-ray exposure for a 370 nm dry oxide under an electric field of 4.7 MV/cm at four different temperatures. The data are normalized to the same initial yield (Source [85]).

In general, degradation due to TID is accelerated under conditions of high temperature, strong electric fields, and low dose rates. These factors collectively increase the mobility and lifetime of charge carriers and mobile ions, thus enhancing the likelihood of charge and interface trap de-passivation [86]. This temperature-dependent behavior is illustrated in Figure 7.2. In it, it can be seen how higher temperatures (298 K) lead to faster hole transport, shortening the time needed for charge redistribution and damage manifestation.

Post-irradiation, device characteristics may continue to evolve. Both annealing (recovery) and degradation can occur, depending on temperature, electrical bias, and continued exposure to radiation.

The impact of TID varies significantly between technologies, manufacturers, process lots, and even between transistors on the same die [87]. Hence, care must be taken in characterizing and mitigating these effects.

Initially, most of the studies on ionization effects in CMOS devices focused on the gate oxide. However, with technology scaling, STI and spacer oxides have become dominant in TID sensitivity due to their increased thickness and relatively poor quality compared to gate oxides.

The accumulation of radiation-induced charges in the STI oxides and their associated interfaces has distinct consequences in NMOS and PMOS transistors. The STI effects are particularly relevant in large and thin devices (big L and small W).

In NMOS transistors, the primary effect is a radiation-induced drain-to-source leakage current. Under normal operation, NMOS devices are activated by applying a positive gate voltage, which attracts electrons to form a conductive channel between the drain and source. However, when exposed to ionizing radiation, positive charges become trapped in the STI oxides. These charges can induce parasitic conduction paths by effectively turning on unintended parasitic

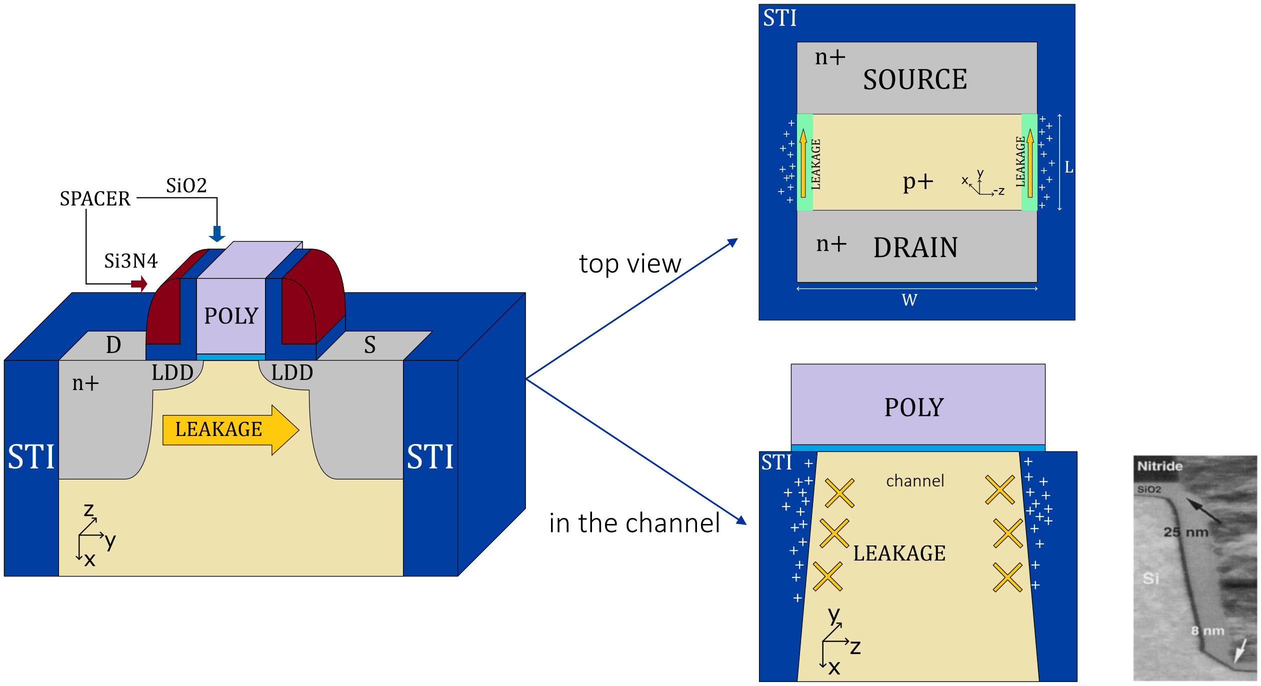
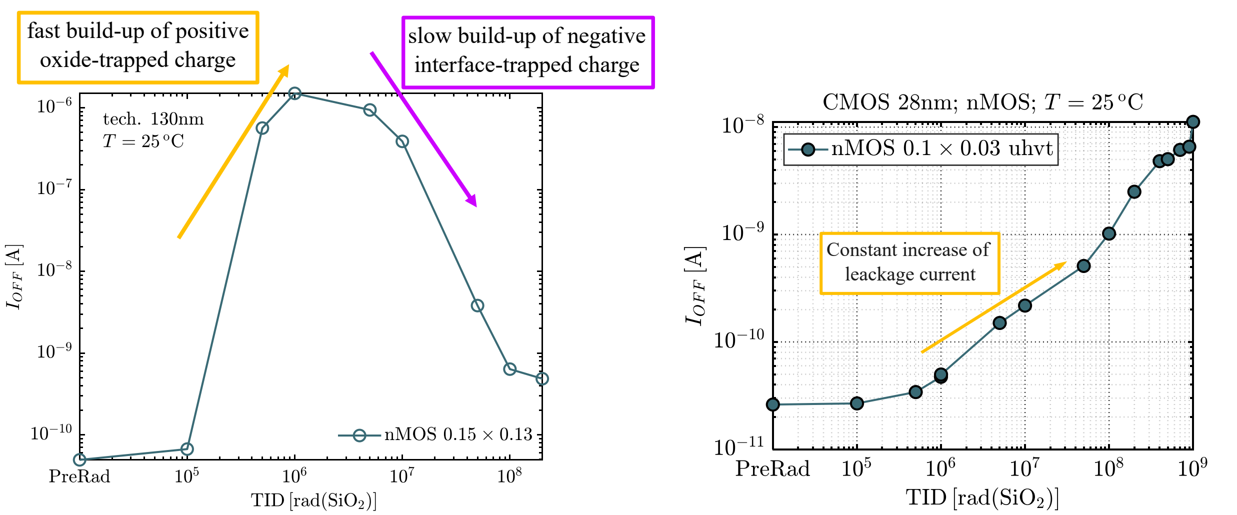


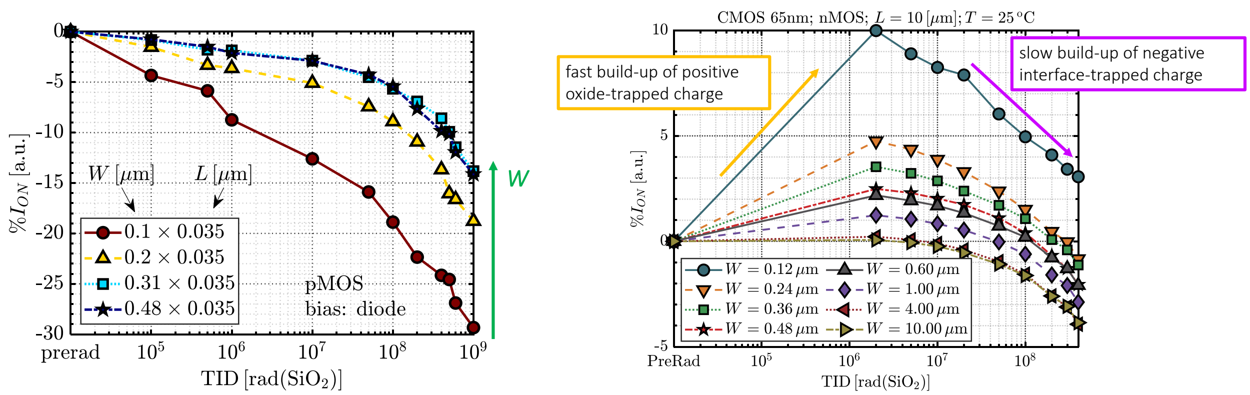
Figure 7.3: STI oxide traps open parasitic transistors in an NMOS (Sources [84], [88]).



(a) (b)

Figure 7.4: Leakage current over TID for NMOS devices at room temperature a) 130 nm technology, b) 28 nm technology (Sources [84], [87]).

transistors along the STI sidewalls, as illustrated in Figure 7.3. This leads to a rise in the off-state current. Over time, hydrogen ions may migrate toward the oxide interfaces and depassivate interface traps, which in NMOS devices capture negative charges. The accumulation of these negative charges can neutralize the effect of the trapped positive charges in the oxide, resulting in a subsequent decrease in leakage current. This dynamic behavior leads to the characteristic bell-shaped leakage profile seen in NMOS of devices under ionization irradiation,



(a) PMOS (28 nm) (b) NMOS (40 nm)

Figure 7.5: Variation of activation current (I_{ON}) over Total Ionizing Dose (TID) for both (a) 28 nm PMOS transistors and (b) 40 nm NMOS transistors. In PMOS devices, the current decreases monotonically over TID, while in NMOS devices the I_{ON} initially increases due to positive oxide charge accumulation and later decreases as interface traps capture negative charges (Source [84]).

as depicted in Figure 7.4(a). However, this behavior does not appear in all technologies. On one hand, technologies such as 130 nm or higher tend to present this bell-shaped curve. On the other hand, smaller technologies such as the 28 nm usually experience a monotonic constant increase in leakage as illustrated in Figure 7.4(b).

PMOS devices, on the other hand, exhibit a phenomenon known as the Radiation-Induced Narrow Channel Effect (RINCE) [89]–[91]. In PMOS both oxide and interface traps capture positive charges. This positive trapped charges repel holes (the majority carriers in PMOS), effectively narrowing the channel and reducing the transistor activation current. As shown in Figure 7.5(a), this effect is also more pronounced in narrow and thin devices.

RINCE effects can also be observed in NMOS devices, but in a different way. Initially, the accumulation of positive oxide charges increases the activation current by reducing the threshold voltage of the device. Over time, this behaviour is compensated by the de-passivation of the interface traps capturing negative charges, leading to a decrease in the activation current. This behavior can be observed in Figure 7.5(b), which shows a similar rise-and-fall trend as the one observed for the leakage current in 7.4.

The spacers, located above the lightly doped drain (LDD) extensions, play a crucial role in controlling short-channel behavior. Under irradiation, positive charges can accumulate in the oxide material of the spacers, especially in PMOS transistors. This modifies the electrostatics of the LDD regions, effectively increasing their resistance and thereby reducing the on-current of the device. In NMOS transistors, the effect of trapped charges in spacers is generally less severe. If oxide traps dominate, the positive charges may reduce resistance and temporarily enhance performance; if interface traps dominate, the resulting negative charge can increase resistance as seen in Figure 7.6. This behavior is known as Radiation-Induced Short Channel Effect (RISCE) [89], [92]–[94].

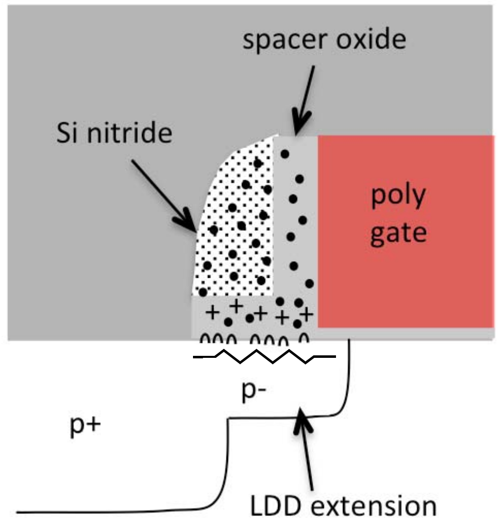


Figure 7.6: A schematic view of the source/drain area during the first irradiation stage. Initially, holes got trapped in the oxide traps (represented as +), and hydrogens got ionized, represented as dots. These hydrogen atoms can reach the oxide interface and depassivate the interface traps, represented as curved lines (Source [92]).

7.1.2 Stochastic Effects

The stochastic effects, also known as Single Event Effects (SEEs), refer to the malfunction of a circuit caused by the interaction of a single energetic particle with semiconductor material. These effects typically occur when a charged particle passes through the depletion region of a reverse-bias p-n junction, as illustrated in Figure 7.7.

When this particle traverses the junction, it generates charge carriers that induce a transient current, which the anode and cathode of the diode may collect. Depending on the location and energy of the particle strike, three categories of outcomes are possible: non-destructive effects, including Single Event Upsets (SEUs) and Single Event Transients (SETs); and destructive effects, such as Single Event Latch-up.

7.1.2.1 Single Event Upsets (SEUs)

An SEU occurs when a charged particle strikes a memory element, such as a latch or flip-flop, potentially altering its stored state. This can result in a bit flip, changing a logical '0' to a '1', or vice versa.

Mitigation of SEUs relies primarily on redundancy. This can be implemented through techniques such as Triple Modular Redundancy (TMR), where flip-flops are triplicated and majority voting logic is used at their inputs or outputs.

An SEU may occur when a charged particle crosses a memory element such as a latch or a flip-flop. When this happens, the stored value in these memory elements might switch, transforming from a 0 into a 1 or from a 1 into a 0. Alternatively, error-correcting codes (ECCs), such as Hamming[96] or Reed-Solomon codes[97], may be employed to detect and

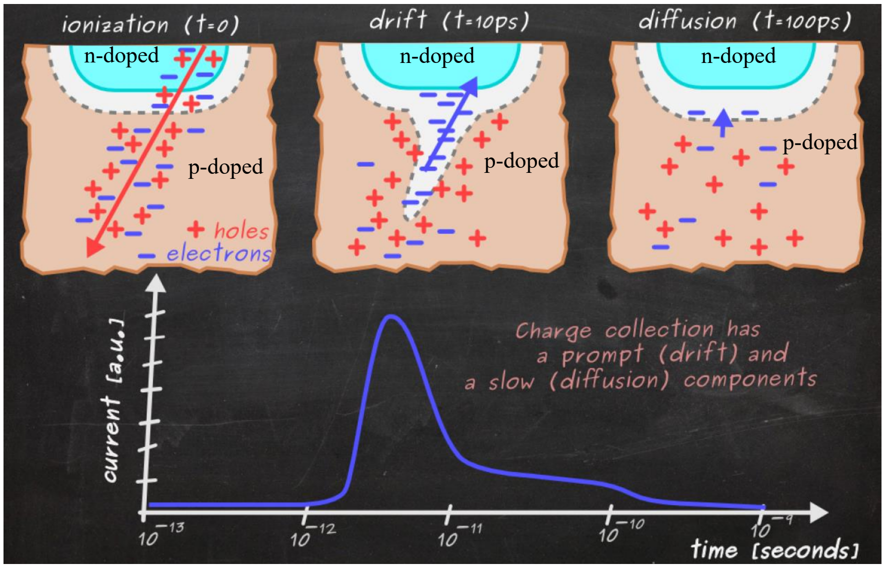


Figure 7.7: Effect of an incident charged particle crossing a reverse-bias p-n junction. An initial current spike is generated rapidly due to the drift of charge carriers in the depletion region. This is followed by a slower diffusion current, resulting in a pulse with a characteristic tail [95].

correct bit flips in a more optimized way than TMR.

7.1.2.2 Single Event Transients (SETs)

A SET occurs when a charged particle affects a combinatorial logic element, such as an inverter or buffer. The resulting transient current can cause a temporary glitch at the output, with the duration and amplitude depending on the particle's energy and the affected circuitry.

Similar to SEUs, SETs can be mitigated using logic triplication combined with majority voting. However, additional care must be taken for SETs, as they can affect clock trees, reset networks, and other critical control signals. Effective protection against SEEs requires triplication of combinatorial paths, clock trees, reset trees, and memory elements, with dedicated voter circuits placed appropriately throughout the design.

Even with such protective measures, a sufficiently energetic particle or a non-perpendicular incident particle can simultaneously affect multiple elements. This may result in a multi-bit upset, where two or more of the triplicated elements are corrupted simultaneously, potentially leading to incorrect voter outputs. To address this, physical separation between redundant elements is crucial, ensuring that a single particle cannot affect more than one copy of a triplicated circuit.

7.1.2.3 Destructive Effects: Latch-Up

In certain semiconductor technologies, an energetic particle can trigger a positive feedback loop that results in a high, sustained current that can lead to severe damage or destruction of the chip. The most common of these destructive SEEs is the Single Event Latch-Up (SEL).

SELs have been observed in 130 nm CMOS technologies, such as those used in the ALICE and LHCb experiments [98], [99].

SEL can be mitigated by incorporating tap cells throughout the design, minimizing the distance between the n-well and p-well bulk ports. These measures reduce the likelihood of parasitic thyristor activation, thereby preventing latch-up conditions.

7.2 Experiment Details

To evaluate the irradiation effects of back bias in the ITS3 commercial 65 nm planar CMOS imaging technology, we utilized custom test structures fabricated in this process. These structures contain a wide variety of nMOS transistors with different gate lengths (L) and widths (W), denoted as $W \times L$. Two threshold voltage (V_{TH}) flavors are included: standard V_{TH} (SVT) and super low V_{TH} (SLVT). In these structures, all nMOS transistors share a common bulk and source connection. Half of the devices share one gate contact, while the other half share a second, distinct gate. Each transistor has an independent drain terminal. Due to the presence of ESD protection circuits, the maximum supported bulk voltage was limited to $V_{\mathrm{BS}} = -2\,\mathrm{V}$.

Irradiation experiments were conducted using a Seifert RPI149 X-ray machine equipped with a tungsten target, producing X-rays with an energy of 10 keV. All reported radiation doses are referenced to $\mathrm{SiO_2}$. The dose rate during all the irradiations performed was 7.9 $\mathrm{Mrad(SiO_2)/h}$

Three irradiation campaigns were performed under different back-bias and temperature conditions: Irradiation 1 ($V_{\mathrm{BS}} = 0\,\mathrm{V}$ at 25 °C), Irradiation 2 ($V_{\mathrm{BS}} = -2\,\mathrm{V}$ at 25 °C) and Irradiation 3: ($V_{\mathrm{BS}} = -2\,\mathrm{V}$ at (−30 °C).

During exposure, all devices were biased in diode configuration, meaning with $V_{\mathrm{DS}} = V_{\mathrm{GS}} = V_{\mathrm{DD}} = 1.2\,\mathrm{V}$, as this configuration is known to induce the largest TID damage in 65 nm CMOS technologies [89], [92], [94].

Each irradiation was carried out in incremental dose steps. After each step, $I_D V_G$ characteristics were measured under both $V_{\mathrm{BS}} = 0\,\mathrm{V}$ and $V_{\mathrm{BS}} = -2\,\mathrm{V}$ for all nMOS transistors. This strategy enables the separation of irradiation-induced effects from those arising during measurement.

7.3 Results and Discussion

Figure 7.8 shows the $I_D V_G$ characteristics of an SVT 0.2 μm × 6.0 μm nMOS transistor, both pre-irradiation and after a TID of 1 $\mathrm{Grad(SiO_2)}$. The left panel corresponds to irradiation and measurement at $V_{\mathrm{BS}} = 0\,\mathrm{V}$, while the right shows results for $V_{\mathrm{BS}} = -2\,\mathrm{V}$.

The plots display $I_D V_G$ on both logarithmic (solid lines, left axis) and linear (dashed lines, right axis) scales. The pre-irradiation differences under the two bias conditions arise from the body effect, which modifies the threshold voltage [100]. Post-irradiation, differences are observed in leakage current, subthreshold swing, and maximum current, attributable to the effect of different V_{BS} during irradiation and measurement. Accordingly, we separately

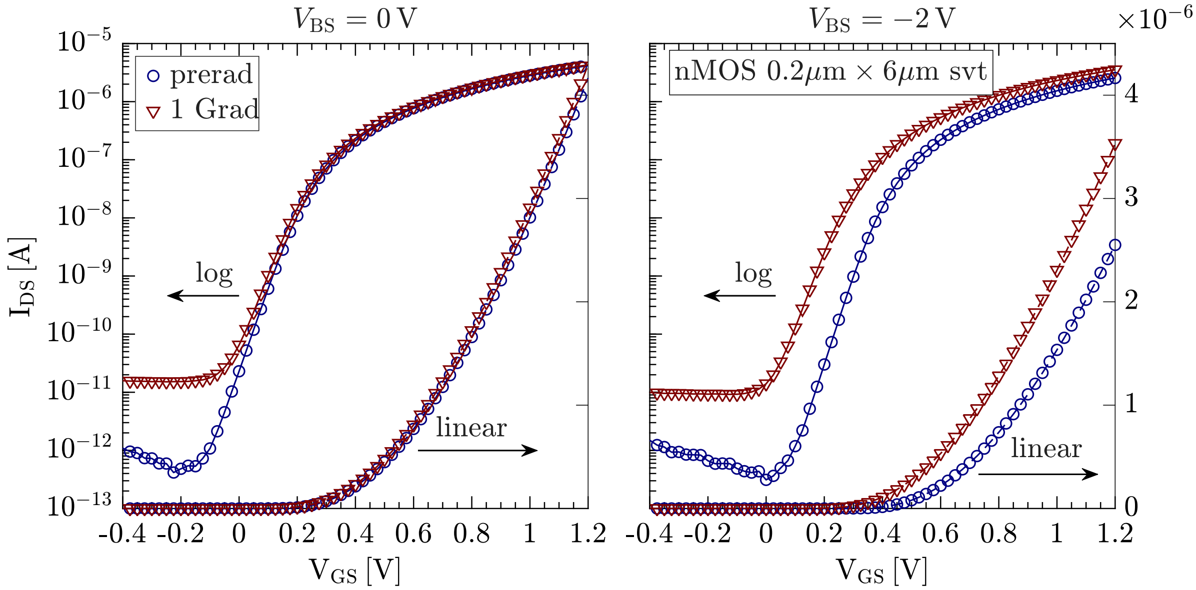


Figure 7.8: $I_D V_G$ characteristics of a $0.2\,\mu$m $\times$ $6.0\,\mu$m SVT nMOS transistor at various TID levels. Left: $V_{\mathrm{BS}} = 0\,$V; right: $V_{\mathrm{BS}} = -2\,$V. Curves are shown on logarithmic (solid lines) and linear (dashed lines) scales.

analyze the influence of back bias during irradiation and the effect of back bias during the measurement of the I_D-V_G curves.

7.3.1 Radiation Effects Due to Back Bias During Measurement

Figure 7.9 presents the variation in I_{ON}, defined as the drain current at $V_{\mathrm{DS}} = V_{\mathrm{GS}} = 1.2\,$V, as a function of TID for nMOS transistors with different dimensions. Devices were irradiated with $V_{\mathrm{BS}} = 0\,$V at room temperature and measured with $V_{\mathrm{BS}} = 0\,$V (left) and $V_{\mathrm{BS}} = -2\,$V (right).

At $V_{\mathrm{BS}} = 0\,$V, the radiation response matches expectations for 65 nm CMOS [80], [87], [89], [92]–[94]. Narrow-channel devices show a characteristic non-monotonic I_{ON} behavior, experiencing an initial increase followed by a drop when the TID exceeded 1–10 Mrad(SiO_2). The influence of W is especially noticeable when comparing the $0.2\,\mu$m $\times$ $6.0\,\mu$m nMOS with the $5.0\,\mu$m $\times$ $6.0\,\mu$m one, as a large L highlights the effect of accumulated charges trapped in the STI oxides. These effects are therefore caused by the accumulation of charges in the STI regions [90], [91], consistent with the RINCE mechanism [91].

Additionally, the radiation response varies according to the channel length due to RISCE effects, with short-channel devices exhibiting more pronounced degradation. Comparing the $0.2\,\mu$m $\times$ $0.1\,\mu$m nMOs transistor with the $0.2\,\mu$m $\times$ $6.0\,\mu$m one, it can be observed that a reduction in channel length provokes an increase in radiation sensitivity, specially when the TID dose exceeds 10 Mrad(SiO_2).

The effect of reducing V_{BS} to $-2\,$V (right plot in Figure 7.9), is more pronounced in narrow-channel devices. This indicates that the back bias modifies the effects that the charges

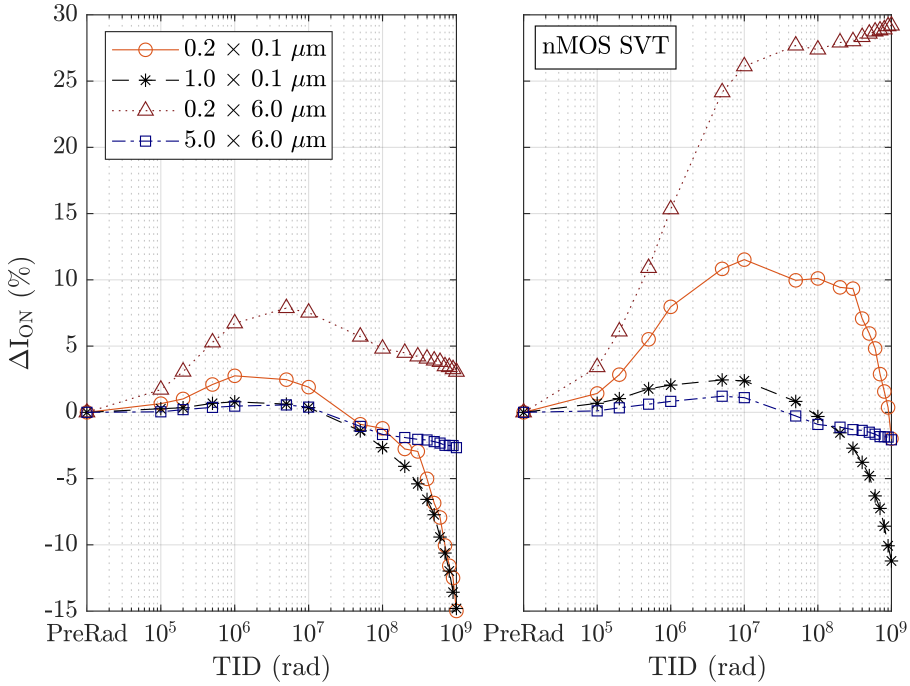


Figure 7.9: Relative I_{ON} variation vs. TID for SVT nMOS transistors of varying W and L. Devices were irradiated with $V_{\mathrm{BS}} = 0\,\mathrm{V}$ at room temperature and measured at $V_{\mathrm{BS}} = 0\,\mathrm{V}$ (left) and $V_{\mathrm{BS}} = -2\,\mathrm{V}$ (right).

trapped in the STI have in the conduction channel. On the contrary, the current reduction in narrow transistors seems the same independently of the back bias applied, suggesting that the effects of the charges trapped in the spacers are not affected by the variations in V_{BS} during measurements. These two behaviors make the arrow and long $0.2\,\mu\mathrm{m} \times 6.0\,\mu\mathrm{m}$ transistor the one showing the largest sensitivity to V_{BS} among all transistors tested.

Figure 7.10 shows the I_{ON} variation over TID for two $0.2\,\mu\mathrm{m} \times 6.0\,\mu\mathrm{m}$ nMOS device with different V_{TH} flavors. As shown, standard V_{TH} devices present a larger increase in I_{ON} compared to super low V_{TH} ones, at both V_{BS} voltages. This behavior reflects the dominant role of threshold voltage shift in RINCE-induced I_{ON} variation. In this case, a device with a lower initial V_{TH} (SLVT) exhibits a smaller relative I_{ON} degradation than one with a higher initial V_{TH} (SVT).

Additionally, the I_{ON} curves over TID in Figure 7.10 present a completely different behavior when measured with different V_{BS} voltages. These different behaviors come from a different ΔV_{TH} curve over TID under both conditions. This V_{TH} shift can originate from either a higher effect of the positive charges oxide traps, or due to a reduced effect of the negative interface traps. To decompose the V_{TH} shift, we applied the McWhorterWinokur method [101], which separates oxide-trapped charge contributions (ΔV_{OX}) from interface trap effects (ΔV_{IT}).

The results from applying this method to a $0.2\,\mu\mathrm{m} \times 6.0\,\mu\mathrm{m}$ SVT transistor are shown in Figure 7.11. As seen, while the total ΔV_{TH} (circles) varies significantly among the two bias conditions, ΔV_{IT} (triangles) remains nearly constant. In contrast, the ΔV_{OX} contribution

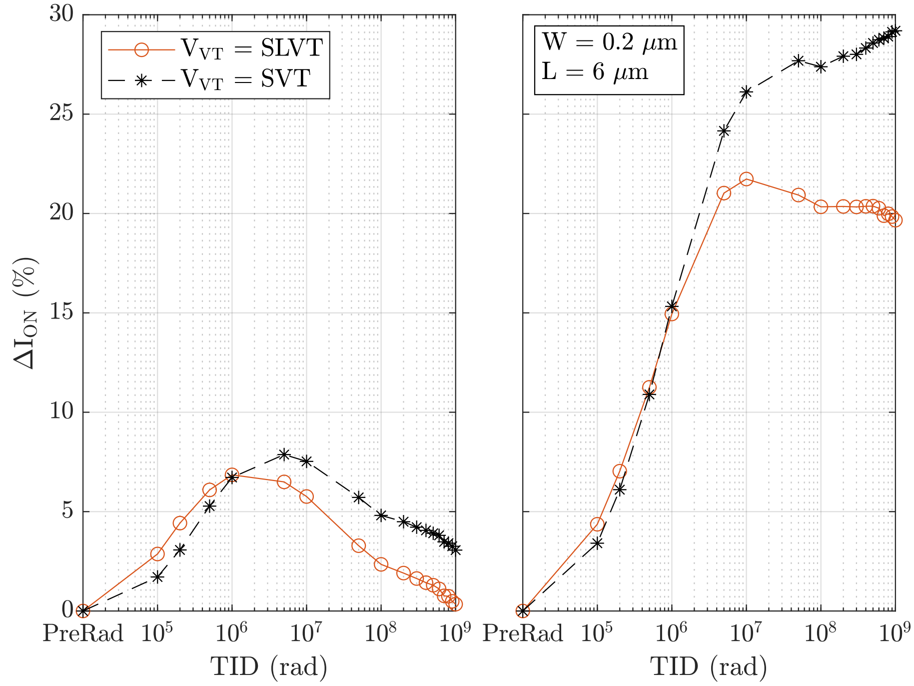


Figure 7.10: Relative I_{ON} vs. TID for two $0.2\,\mu m \times 6.0\,\mu m$ nMOS transistor with SVT and SLVT V_{TH} flavors respectively. Both transistors are Irradiated with $V_{BS} = 0\,V$ and measured at $V_{BS} = 0\,V$ (left) and $V_{BS} = -2\,V$ (right).

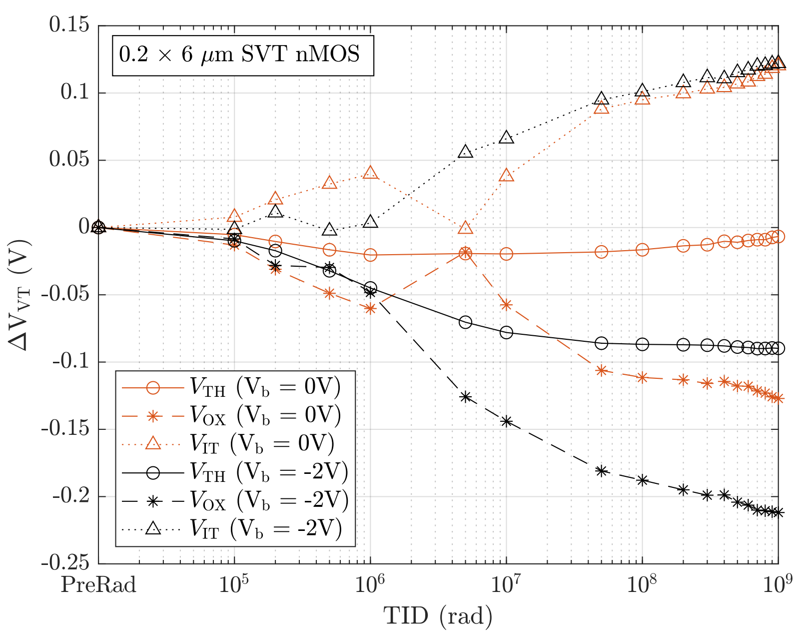


Figure 7.11: Threshold voltage shift ΔV_{TH} vs. TID for a $0.2\,\mu m \times 6.0\,\mu m$ SVT nMOS transistor irradiated with $V_{BS} = 0\,V$ and measured with $V_{BS} = 0\,V$ and $V_{BS} = -2\,V$. The contributions to the total ΔV_{TH} shift are splitter in oxide traps (ΔV_{OX}, stars) and interface traps (ΔV_{IT}, triangles).

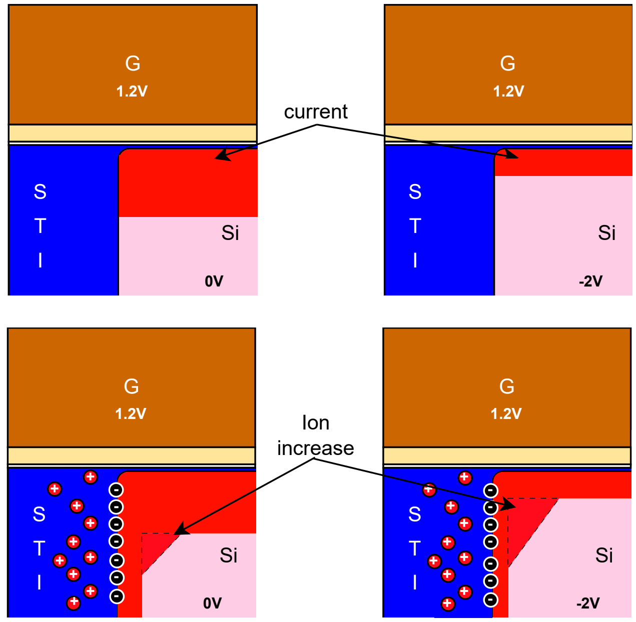


Figure 7.12: Schematic cross-sections of an nMOS transistor in bias configuration. Left: $V_{\mathrm{BS}} = 0\,\mathrm{V}$; right: $V_{\mathrm{BS}} = -2\,\mathrm{V}$. Top: pre-irradiation; bottom: post-irradiation. A higher back bias reduces the conduction channel width, increasing the impact of STI-induced edge channels after irradiation.

(stars) is significantly larger when measured at $V_{\mathrm{BS}} = -2\,\mathrm{V}$ than when doing it at $V_{\mathrm{BS}} = 0\,\mathrm{V}$. These results confirm that the large increase in I_{ON} current in devices measured with $V_{\mathrm{BS}} < 0\,\mathrm{V}$ is driven primarily by an enhanced oxide charge effect rather than a reduced effect of the charges trapped in the interface.

This enhanced sensitivity may result from the narrowing of the conduction channel near the STI edges under high back bias. Figure 7.12 presents cross-sectional schematics of an nMOS transistor before and after irradiation under both $V_{\mathrm{BS}} = 0\,\mathrm{V}$ and $V_{\mathrm{BS}} = -2\,\mathrm{V}$. The high back bias compresses the central conduction channel, making the STI-induced "edge channel [91] more dominant in the total I_{ON} current compared to the $V_{\mathrm{BS}} = 0\,\mathrm{V}$ case.

7.3.2 Radiation Effects due to Back Bias during Irradiation

To evaluate the influence of irradiation conditions on RINCE, Figure 7.13 presents I_{ON} versus TID for a $0.2\,\mu\mathrm{m} \times 6.0\,\mu\mathrm{m}$ SVT nMOS transistor irradiated under different V_{BS} voltages. Measurements were conducted with $V_{\mathrm{BS}} = 0\,\mathrm{V}$ and $V_{\mathrm{BS}} = -2\,\mathrm{V}$, corresponding to dashed and solid lines respectively. As seen, a higher V_{BS} applied during irradiation leads to greater degradation due to the increased electric field in the STI region. This enhanced electric field promotes electron-hole pair separation, increasing charge yield [85], [86], reducing recombination, and ultimately worsening radiation-induced damage.

TID also impacts the off-state leakage current (I_{OFF}). Figure 7.14 illustrates the evolution of I_{OFF} with TID for a $0.2\,\mu\mathrm{m} \times 0.1\,\mu\mathrm{m}$ SVT nMOS transistor under three different irradiation conditions. All measurements were taken at $V_{\mathrm{BS}} = 0\,\mathrm{V}$ and $V_{\mathrm{DS}} = 1.2\,\mathrm{V}$, and at the same

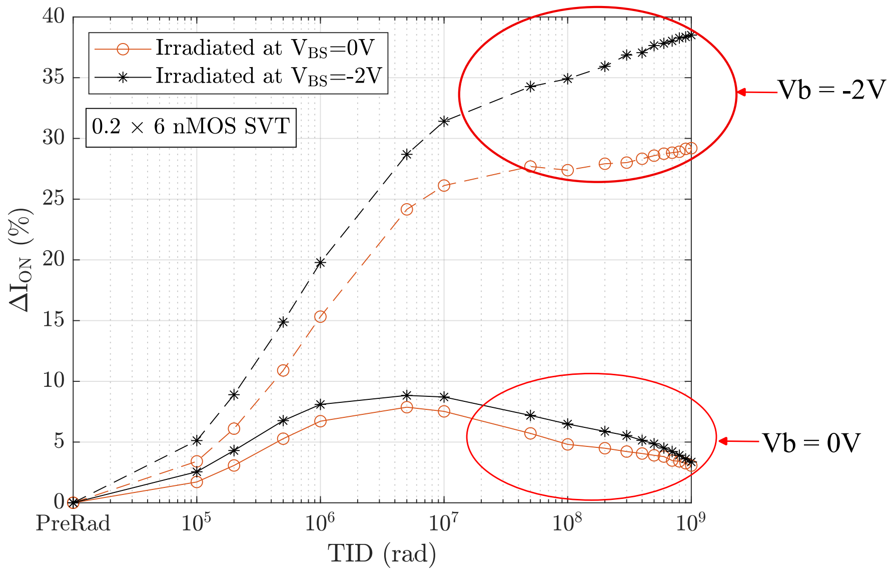


Figure 7.13: Relative I_{ON} as a function of TID for a $0.2\,\mu$m $\times$ $6.0\,\mu$m SVT nMOS transistor measured with $V_{\mathrm{BS}} = -2\,$V (solid line) and $V_{\mathrm{BS}} = 0\,$V (dashed line). Irradiations results are presented for both $V_{\mathrm{BS}} = 0\,$V and $V_{\mathrm{BS}} = -2\,$V at 25 °C.

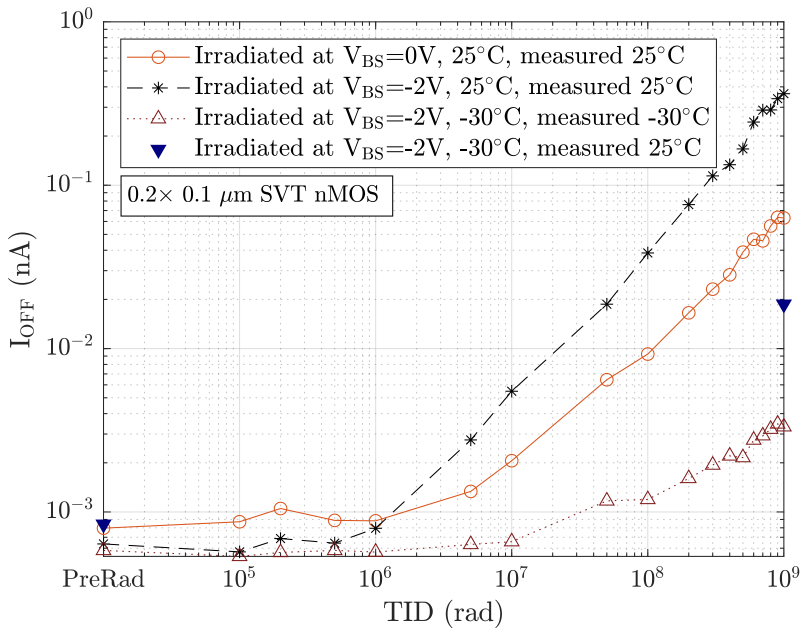


Figure 7.14: Parasitic currents (I_{OFF}) versus TID for a $0.2\,\mu\mathrm{m}\times 0.1\,\mu$m SVT nMOS transistor under three different irradiation conditions. Measurements were performed at $V_{\mathrm{BS}} = 0\,$V, $V_{\mathrm{GS}} = -0.4\,$V and $V_{\mathrm{DS}} = 1.2\,$V.

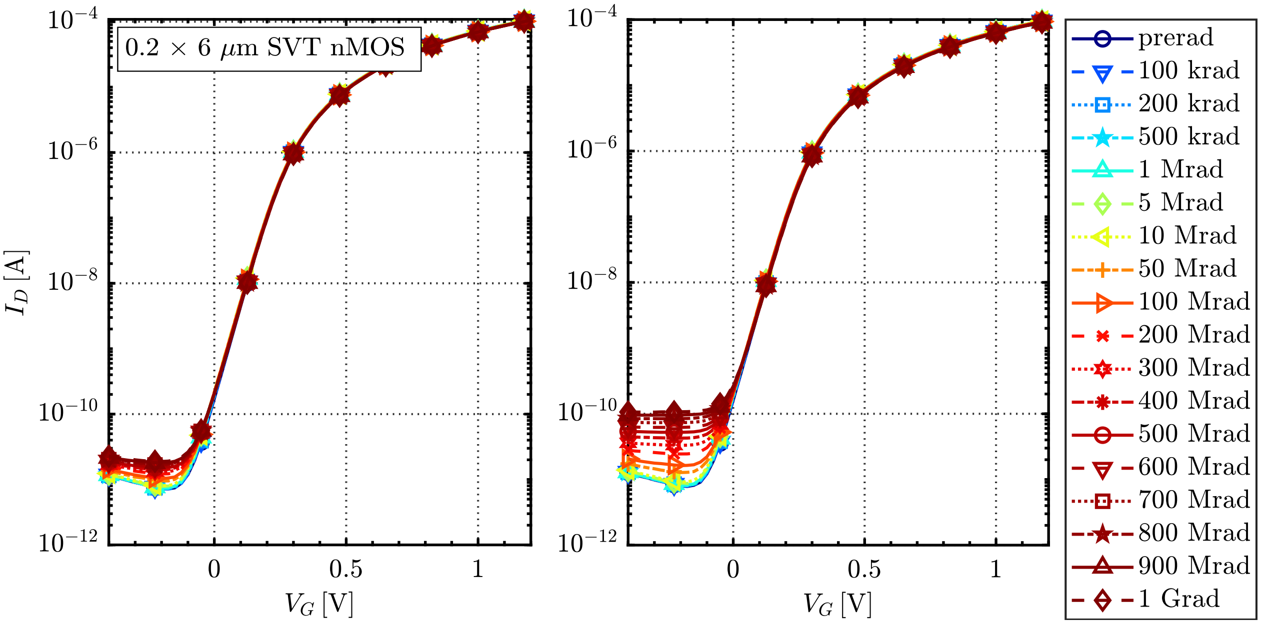


Figure 7.15: I_DV_G characteristics of a 0.2 μm × 6.0 μm SVT nMOS transistor at various TID levels measured with $V_{\mathrm{BS}} = 0$ V. The left plot shows the response after irradiation with $V_{\mathrm{BS}} = 0$ V, while the right plot corresponds to irradiation with $V_{\mathrm{BS}} = -2$ V.

temperature as the one set for irradiation (25 °C or −30 °C respectively). For the transistor irradiated at −30 °C, additional measurements before and after irradiation at 25 °C are shown as blue inverted triangles. To isolate the leakage current originating from parasitic paths from the threshold voltage shift, I_{OFF} was measured at $V_{\mathrm{GS}} = -0.4$ V.

As seen, the three curves present a constant increase in the leakage current for TID levels higher than, behaving similarly to the 28 nm example curve in Figure 7.4(b). For the two irradiations at 25 °C, higher leakage currents are observed when a $V_{\mathrm{BS}} = -2$ V voltage is applied during exposure. However, when irradiation is carried out at −30 °C, the leakage current measured after 1 Grad and at 25 °C is significantly lower than the one seen in the irradiations performed at 25 °C.

This behavior is consistent across other transistor geometries. Figure 7.15 shows I_DV_G curves for a 0.2 μm × 6.0 μm SVT nMOS transistor irradiated at 25 °C with $V_{\mathrm{BS}} = 0$ V (left) and $V_{\mathrm{BS}} = -2$ V (right). At 1 Grad, leakage current increases by a factor of 3 with $V_{\mathrm{BS}} = 0$ V, and by a factor of 10 with $V_{\mathrm{BS}} = -2$ V.

The dependence of radiation-induced parasitic current on bias and temperature suggests a possible Enhanced Low-Dose-Rate Sensitivity (ELDRS) effect [102], [103]. ELDRS leads to greater degradation at lower dose rates [104], posing challenges for hardness assurance since high dose rate tests might underestimate damage under actual operational conditions [102], [104]–[113]. ELDRS effects on leakage have been reported in 350 nm, 250 nm, 180 nm [114], and 28 nm [115] planar CMOS nodes. Bias [103] and temperature [105], [106], [108], [111] conditions during irradiation may also enhance the impact of dose-rate-dependent mechanisms. In this context, the degradation observed under high dose rates combined with high bias and/or elevated temperature may indicate an underlying dose-rate sensitivity. The data in Figure 7.14 support this theory, revealing an enhancement of the degradation under higher bias and temperature conditions, consistent with ELDRS behavior.

7.4 Conclusions

This study analyzed TID effects in a 65 nm commercial CMOS imaging process, focusing on the influence of back bias (V_{BS}) and temperature under high TID conditions. In this study, two primary effects were identified.

First, the effect of $V_{\mathrm{BS}} < 0\,\mathrm{V}$ **during measurement** alters the I_{ON} behavior associated with the Radiation-Induced Narrow Channel Effect (RINCE). Specifically, I_{ON} increases more rapidly at low TID levels, while the decrease at higher TID levels is reduced or even suppressed. This behavior is due to an enhanced contribution of the positive oxide traps in the STI, with the interface traps effect remaining constant regardless of the bias, as confirmed by the threshold voltage decomposition analysis. This effect was traced to induced conductive edge channels under back bias that dominate the I_{ON} increase.

Second, applying $V_{\mathrm{BS}} < 0\,\mathrm{V}$ **during irradiation** leads to a stronger increase in leakage current with TID. Conversely, performing irradiation at lower temperatures mitigates this increase. These trends are consistent with the Enhanced Low-Dose-Rate Sensitivity (ELDRS) behavior, where high electric fields and temperatures, or lower dose rates, reduce recombination and increase charge trapping in the STI, forming additional leakage paths.

In summary, applying negative back bias can improve I_{ON} robustness under TID but also leads to enhanced leakage degradation, especially under high temperature and/or low dose rates. These trade-offs must be carefully considered in the design of future back biased pixel detectors such as ITS3.

Chapter 8

Conclusions

Monolithic pixel detectors [40] have long dominated commercial imaging technologies. In recent years, they have attracted growing interest in high-energy physics (HEP) due to several advantages: significantly lower cost, reduced power consumption, minimal material budget (lower radiation length), and simplified assembly processes. Despite these benefits, their relatively low signal-to-noise ratio and limited radiation tolerance have restricted their use to low-radiation HEP experiments, such as ALICE at CERN.

In ALICE, the Inner Tracker System 2 (ITS2), based on Monolithic Active Pixel Sensors (MAPS), has been operational since 2021. During the LHCs Long Shutdown 3, planned from 2026 to 2029, the ITS2 three innermost layers will be upgraded to the ITS3.

The ITS3 is a novel detector composed of three cylinders, each built from two wafer-scale monolithic pixel detectors fabricated using commercial CMOS imaging technology. Thanks to a fabrication technique known as **stitching**, sensors of up to $27\,\mathrm{cm} \times 9\,\mathrm{cm}$ can be produced. These sensors can be bent around the beam pipe, enabling a purely cylindrical geometry and a significant reduction in material budget.

The ITS3 development was divided into several Engineering Runs (ER), each validating specific engineering concepts before the final chip installation.

This thesis contributed to ER2, the last R&D run before the fabrication and installation of the final ITS3 chips. The ER2 chip, named the **Monolithic Stitched Active Pixel (MOSAIX)**, was designed to incorporate all features and capabilities required for ER3. This approach minimized changes between ER2 and ER3, thereby reducing the risk of failures.

Within this context, the thesis addressed the following main topics:

- Development of a complete MOSAIX readout architecture model, based on realistic physics inputs. This model guided both early high-level architecture decisions and the fine-tuning of readout units and parameters such as the in-chip memory depth, readout link count and speed, or the transmission protocol.
- Design of the full readout architecture, from the pixel front-end to the data transmission links.

- Radiation testing of the 65 nm CMOS technology under various temperatures and substrate bias voltages.
- Implementation of the complete readout architecture, from standard-cell design until post-implementation power and timing analysis.

All of these activities were essential to the successful submission of the MOSAIX chip in July 2025. Due to the complexity of this work, several key challenges had to be addressed, yielding important lessons.

8.1 Challenges

Two challenges proved particularly significant during the MOSAIX readout design: The engineering metrics trade-off and the timing closure.

Achieving an optimal **trade-off** among engineering metrics was key to the success of the MOSAIX chip. These metrics are:

- Yield
- Power consumption
- Area
- Readout Performance
- Functionality and configurability
- Engineering time

It is relatively straightforward to design a simple circuit with low performance and minimal functionality that excels in area and power efficiency. However, the target performance was defined early by the readout model and had to be maintained. Over time, additional functionality was introduced (extra operating modes, safety features, extended slow-control status information...), each bringing substantial area and power costs.

The readouts area and power budgets also had to be balanced against other chip subsystems. By the end of the design, power consumption in the readout logic was higher than expected due to these added features. Significant reductions were achieved through extensive clock gating and the removal of non-essential functions. Ultimately, some extra power headroom became available when other blocks consumed less than anticipated.

Moreover, additional performance and functionality require additional verification. Every mode and extra feature must be validated across all configuration corners. This increases both the complexity of the verification environment and its required simulation runtime.

The second major challenge was achieving timing closure within 25 ns for the timing arcs between the periphery and the pixel matrix. This required:

- Identifying and optimizing the slowest paths, allowing the pixel matrix team to improve them at the expense of reducing other non-critical paths.

- Designing specialized pixel matrix drivers, including a custom Double Data Rate (DDR).
- Inserting pipeline stages to improve timing, while ensuring they could be clock-gated without affecting readout behavior or introducing hit-dependent power variations.
- Since the pixel matrix and periphery were implemented in independent design flows, inter-block delay information was explicitly exchanged via constraint files to enable accurate cross-boundary timing closure.
- Defining custom floorplan boundaries for certain modules to aid the tool in placing the critical modules in the most optimal spots.

Thanks to these efforts, a balance between timing, power, and area was achieved. All timing arcs were closed with more than 1 ns of margin in the worst analysis corner, while keeping power within the budget.

8.2 Lessons Learned

In HEP detector design, simplified statistical models such as Poisson or Gaussian distributions are often used to simulate data flows. For heavy-ion experiments like ALICE, these models fail to capture the temporal and spatial complexity of real Pb-Pb collisions. Although average hit rates may match, the actual event-to-event distributions differ significantly from those predicted by Poisson processes.

For MOSAIX, a physics-driven data model was therefore essential and future Pb-Pb detectors must as well perform a physics-accurate hit analysis.

Another key lesson concerns the verification of highly configurable chips. Adding extra operation modes or configurability may be straightforward at the RTL level, and important when re-using HEP detector chips in other applications (beam tests, medical imaging, or space missions). However, this versatility dramatically increases the verification time, requiring more development effort and longer simulation runtimes.

For future designs, configurability should be approached with caution: every additional mode of operation can exponentially increase verification complexity, even if it appears trivial to implement in RTL.

8.3 Future Work

The next step beyond this thesis is to evaluate the MOSAIX chip once it returns from fabrication. Initial “smoke test” results are expected in December 2025, followed by more extensive testing throughout 2026. These results will be crucial for identifying any final adjustments required in the fabrication of the ITS3, including possible modifications to the readout architecture.

If no major issues are observed in MOSAIX, the same readout design will be adopted for the final ITS3 chip, which is scheduled for installation in the ALICE cavern during the upcoming three-year long shutdown (2026-2029)

With regard to radiation tests, the mechanisms underlying the two observed effects (effects of TID when measuring under reverse bulk bias, and effects of the TID when irradiating under back bias) merit further investigation. Future irradiation campaigns at varying dose rates and temperatures may provide a deeper understanding of the ELDRS effect in this technology. Additional experiments using different V_B values could also refine the conclusions presented in this thesis.

The ALICE ITS3 represents the first curved wafer-scale pixel detector deployed in particle physics. Its advantages are significant, particularly the reduction of material budget and the improvement in the chip integration. This ambitious R&D project has opened the door to wafer-scale chips for a wide range of applications. Other LHC and HEP experiments may explore similar technologies, while fields such as space instrumentation and medical imaging could leverage the lessons learned from ITS3 to develop their own wafer-scale chips, thereby reducing cost and weight in future devices.

Appendix A

Pixel Matrix Operation

This Appendix presents several scenarios in which the on-chip readout architecture operates the pixel matrix.

A.1 Pixel Matrix Initialization

Upon chip power-up, the pixel matrix is in an undefined state. The initialization of the pixel matrix is shown in Figure A.1. Here, the `hold` is asserted, and two consecutive `frame` pulses are sent to initialize the matrix. The first pulse clears the INT_REG, and the second clears the RO_REG. The first integration window begins when `hold` is de-asserted, and ends when a `frame` pulse is sent.

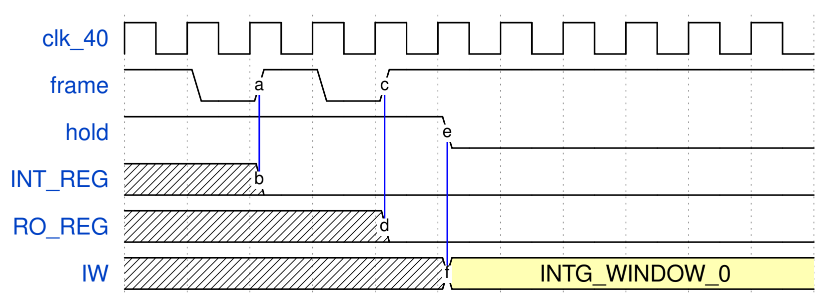


Figure A.1: Pixel matrix initialization and beginning of the first integration window.

A.2 Pixel Hit During a `frame` Pulse

If a pixel detects a hit (rising edge on `disc_out`) while `frame` is low (during a pulse), the hit is recorded as part of the **current** integration window (Figure A.2).

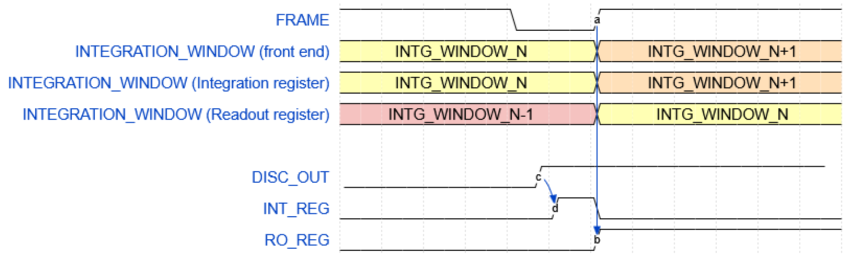


Figure A.2: Pixel hit triggered during frame pulse.

A.3 Multiple Pixel Hits Before a `frame` Pulse

If a pixel is hit multiple times within the same integration window, only the first hit is registered, as illustrated in Figure A.3.

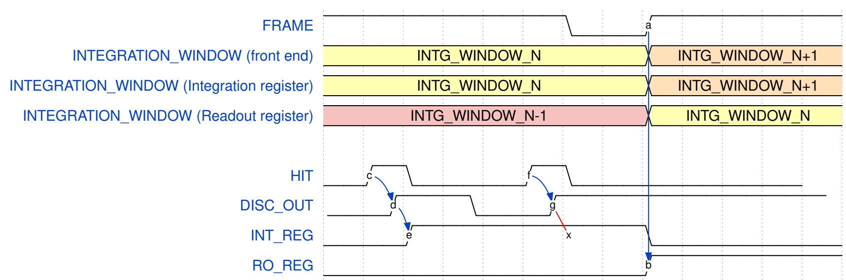


Figure A.3: Two pixel hits recorded before a frame pulse. Only one of them is recorded.

A.4 Time-over-Threshold of the Discriminator Output

The discriminator output may remain high for several microseconds due to ToT. While high, additional hits are blocked, even across frame pulses as depicted in Figure A.4.

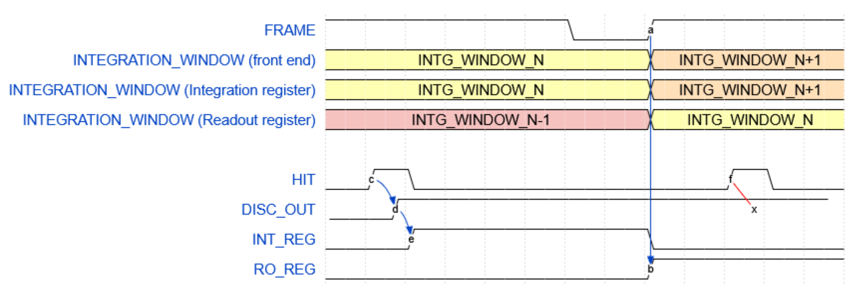


Figure A.4: ToT of the discriminator output blocks subsequent hits.

A.5 Pulse and pixel hit

The integration logic uses a NAND gate to combine `disc_out` with the pulsing signal. Therefore:

- If a digital pulse is active, the discriminator output is ignored (Figure A.5).
- Conversely, if `disc_out` is high, digital pulses are ignored (Figure A.6).

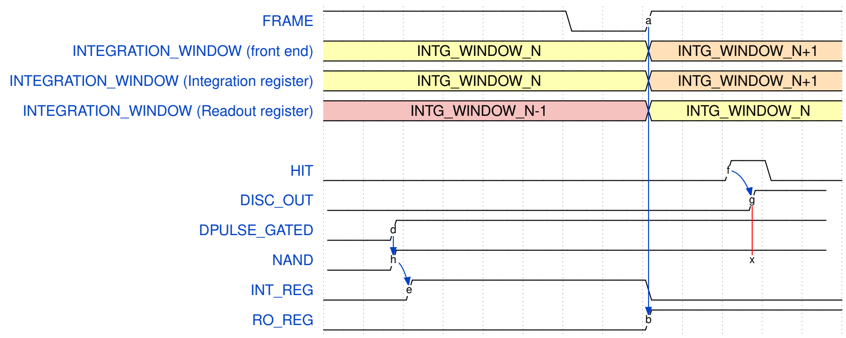


Figure A.5: Digital pulse blocking the detection of additional hits in the pixel.

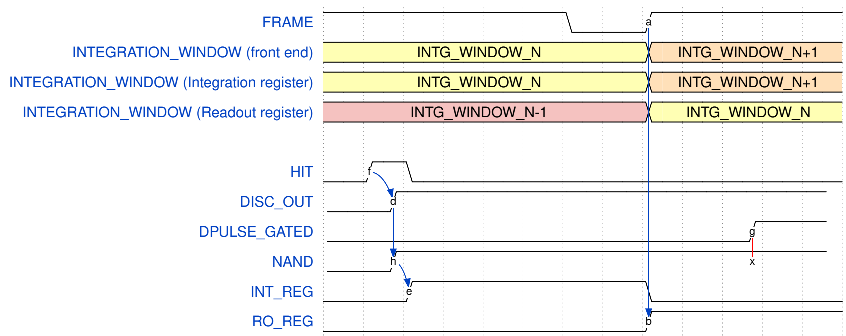


Figure A.6: Discriminator output blocking the assertion of the digital pulse.

A.6 Extending Readout Time Allocation using `hold`

If **not all** the Readout Registers of a pixel region **are fully read** at the end of the integration window, we will need to allocate extra time to complete the reading. To do so, the `hold` is asserted, freezing the Integration Registers. While `hold` is active, new pixel hits are discarded as depicted in Figure A.7.

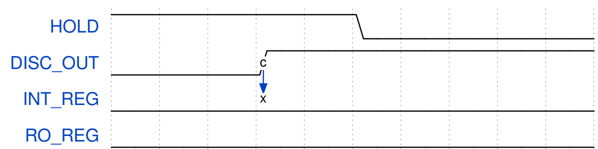


Figure A.7: Asserting `hold` freezes the Integration Registers and blocks new hits.

As a result, the boundaries of the integration window are defined by a **rising edge** of `frame` or by **any edge** of `hold`.

A.7 Readout of One Pixel Hit

Figure A.8 shows the readout of one pixel hit (i) in the priority encoder *PE_A*.

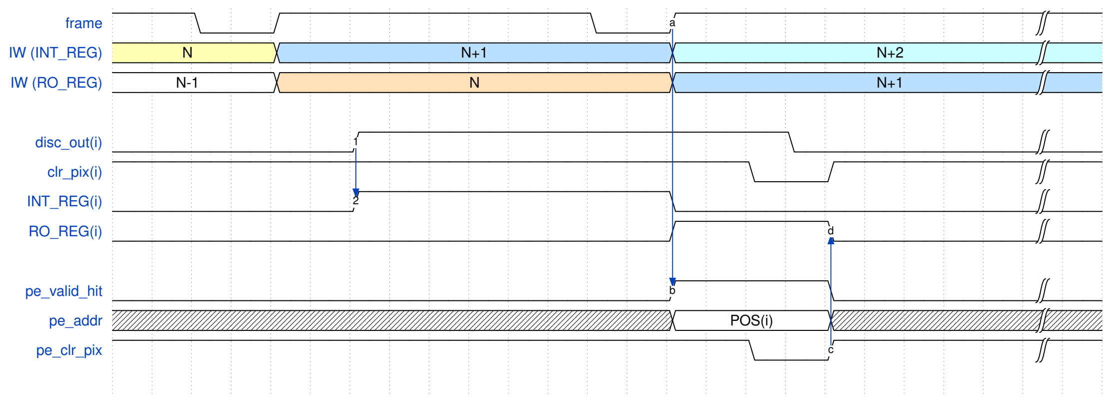


Figure A.8: Readout of one pixel hit through a priority encoder.

A.8 Multiple Pixel Hits in a Single PE

If multiple pixels in the same double column are hit, the PE processes them sequentially, one per cycle. The readout of three pixel hits (i, j and k) in *PE_A* can be seen in Figure A.9.

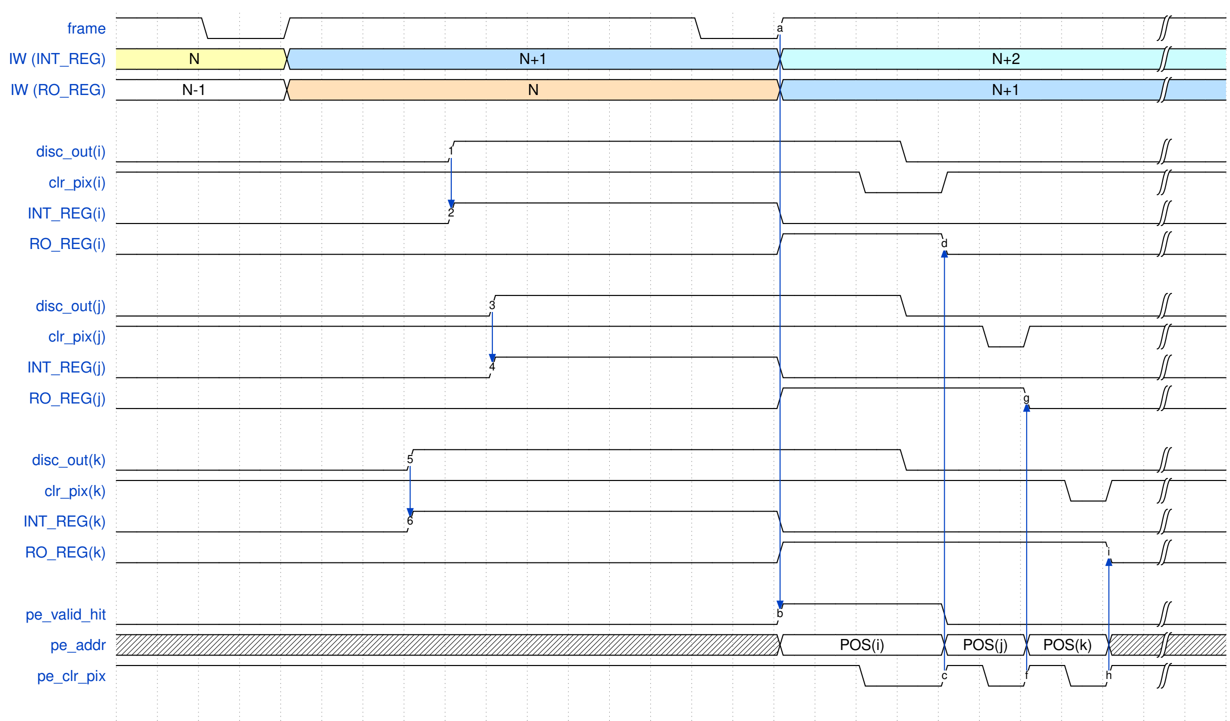


Figure A.9: Readout of three pixel hits through a single priority encoder.

A.9 Readout of Hits in Multiple PEs

To achieve an uninterrupted readout of the PEs in a region, each PE's `clr_pix_b` signal is pre-armed to '0' immediately after the frame pulse. This ensures that the readout architecture can change from PE to PE without any waiting cycle during the readout.

Figure A.10 shows the readout of 3 pixel hits (i, j and k) in three different priority encoders (PE_A,PE_B and PE_C).

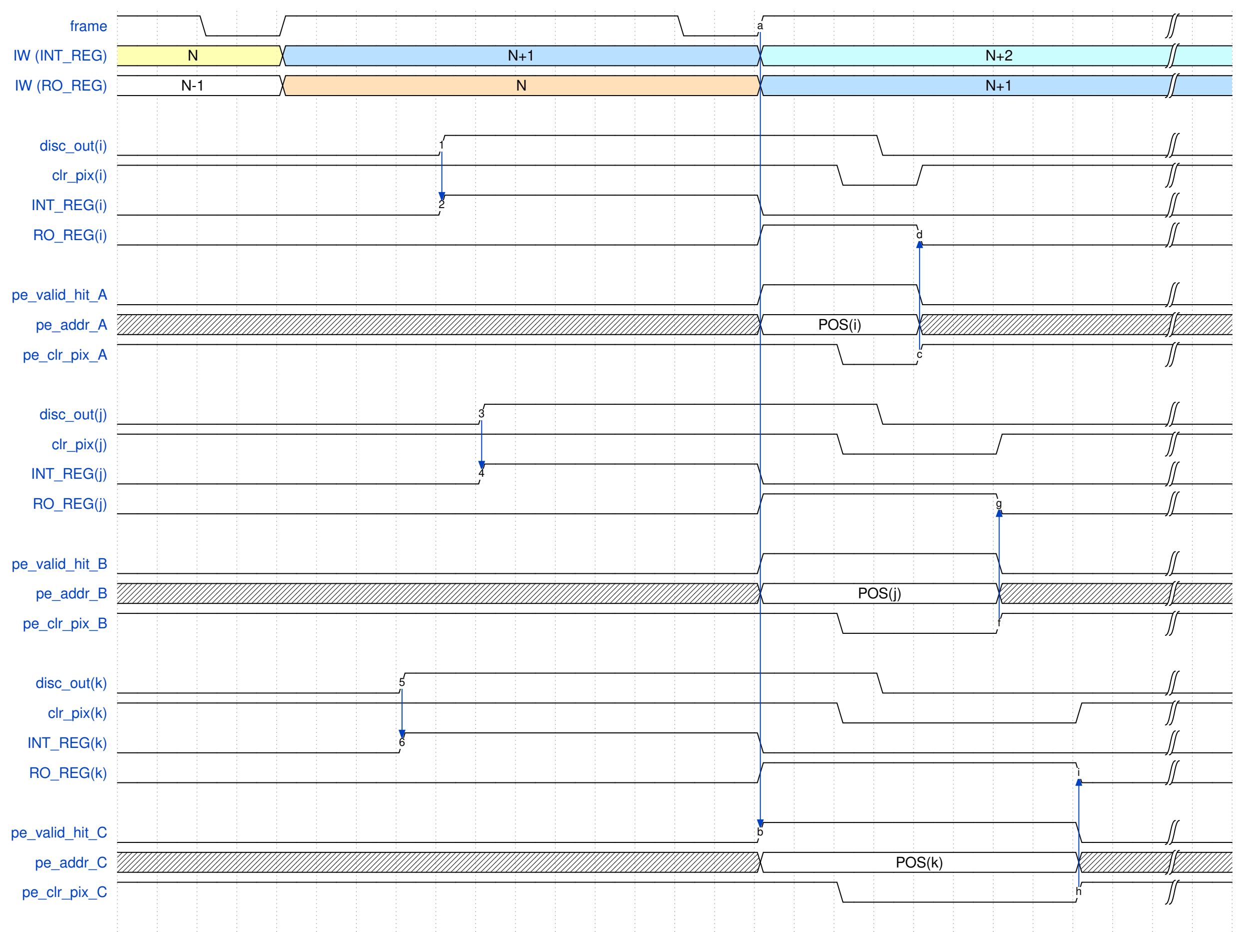


Figure A.10: Readout of three pixel hits in different priority encoders in the same region.

A.10 Readout Behavior During `hold` Assertion

If `hold` is asserted, the value of the Integration Registers is kept frozen, even if the `disc_out` is toggling.

Figure A.11 shows the impact of the `hold` signal on two pixels. `hold` is asserted from *integration_window n+1* to *integration_window n+m*. While `hold` is asserted, the Integration Registers keep their previous value, a '1' for pixel i and a '0' for pixel j. When a `frame` pulse is sent, the value of the Integration Registers is copied to the Readout Registers. The Integration Register resumes its operation when `hold` is de-asserted.

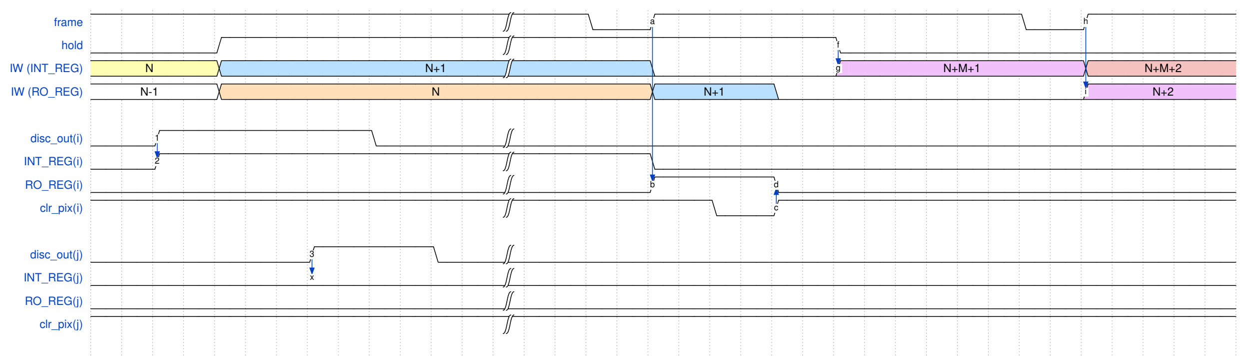


Figure A.11: Impact of `hold` on the readout process: Integration Registers are frozen and hits are ignored.

Appendix B

Tile Readout Requirements

To simplify understanding, the readout requirements are divided into three categories: timing behavior (Table B.1), protocol rules (Table B.2) and general functionality (Table B.3). These requirements guide the design and implementation of the RTL architecture.

Table B.1: Tile Readout timing requirements.

REQ ID	Description
REQ04-2-01	N $\times$ `sync` command pulses **shall** generate N-1 integration windows.
REQ04-2-02	The external `sync` command **shall** reset an internal timebase that all other pixel control signals are relative towards.
REQ04-2-03	An integration window **shall** begin relative to the internal timebase, but with a configurable latency.
REQ04-2-04	By default, there **shall** be no gap between integration windows, i.e., no dead time where pixel hits are disregarded.
REQ04-2-05	Every pixel hit **shall** be associated with a **single** integration window.
REQ04-2-06	The timing of the integration window and `pulse` signals **shall** be agnostic to each other, but deterministic and relative to the internal `sync` command.
REQ04-2-07	A `pulse` command **shall** be executed relative to the internal timebase, but with a configurable latency.
REQ04-2-08	The duration of a `pulse` command **shall** be configurable relative to the internal timebase.
REQ04-2-09	The option to introduce a configurable gap between integration windows **shall** be possible.

Table B.2: Tile Readout protocol requirements.

REQ ID	Description
REQ04-3-01	The matrix data, i.e., the pixel hits, **shall** be sent as packets in which the hits occurring during an integration window are combined.
REQ04-3-02	The packets **shall** be transmitted in order.
REQ04-3-03	The data from an integration window **shall** be sent as a single packet.
REQ04-3-04	Each integration window **shall** be acknowledged to the upstream system via the readout link.
REQ04-3-06	A lock packet **shall** be sent periodically relative to the 40 MHz clock.
–	The lock packet **shall** include the full 16-bit integration window counter.
REQ04-3-07	The upstream system **shall** be informed if any internal memory is full or was full during the last integration window.
REQ04-3-08	The upstream system **shall** be informed when internal memory is nearing capacity (e.g., 90%).
REQ04-3-09	The readout protocol **shall** use only 16-bit words.

Table B.3: General readout requirements.

REQ ID	Description
REQ04-4-01	High occupancy and `sync` rate **shall** never cause integration window `sync` loss.
REQ04-4-02	The readout **shall** ship full packets regardless of the number of hits.
REQ04-4-03	It **shall** be possible to complete packet readout even if it exceeds the integration window period.
REQ04-4-04	Dropping integration windows is accepted if readout is busy with the previous packet.
REQ04-4-05	The integration window counter **must not** be interrupted during the entire run (e.g., 12 hours).
REQ04-4-06	A user-initiated readout reset **shall** not reset the integration window counter.
REQ04-4-07	The integration window counter **shall** be resettable via slow control.
REQ04-4-08	It **shall** be possible to pause/unpause the readout using slow control.
REQ04-4-09	The integration counter **shall** continue during pause to maintain sync.
REQ04-4-10	The first packet sent **shall NOT** contain stray hits from before the first `sync`.
REQ04-4-11	The PE readout rate **shall** be configurable (e.g., halved).
REQ04-4-12	Emulating `sync` command via slow control **shall** be supported.
REQ04-4-13	It **shall** be possible to read out at least two consecutive integration windows without loss, regardless of hit count.

Appendix C

TRU Design Details

This Appendix presents the design details of the TRU, showing the different FSMs, logic and protocol in high detail.

C.1 Window Management Unit (WNMU)

The Window Management Unit (WNMU) serves as the timing control core of the TRU. It receives the global MOSAIX timestamp marker (the `sync` pulse) and processes it to manage the pixel matrix accordingly. Its main responsibilities include:

- Segmenting pixel hits into integration windows based on time.
- Allocating additional time to finish the read of a previous integration window by dropping upcoming pixel hits.
- Tracking the integration window status across all region FIFOs.
- Controlling the analog and digital pulses applied to the pixel matrix.

As seen in Figure 6.4, the WNMU is divided into several submodules, which will be explained next.

C.1.1 Timebase Handler

The timebase handler ensures timing coordination of the TRU to the global `sync` pulse. It has a local 16-bit counter, known as the `internal timebase`, which is reset upon receipt of each sync pulse. After reset, the internal timebase increments every clock cycle until reaching `16h'FFFF`. At a configurable latency with respect to the internal timebase, the timebase handler generates:

- `window_marker`: Pulse that defines the start of an integration window.
- `pulse_cmd`: Triggers a digital or analog pulse in all pixels of the pixel matrix with the analog or digital pulse configuration (see Chapter 5). The duration of `pulse_cmd` is configurable.

An example of the timebase handlers normal operation with a pixel pulse is shown in Figure C.1, with the following configuration:

- `WINDOW_MARKER_LATENCY`: 2 clock cycles.
- `PULSE_CMD_LATENCY`: 4 clock cycles.
- `PULSE_CMD_DURATION`: 4 clock cycles.

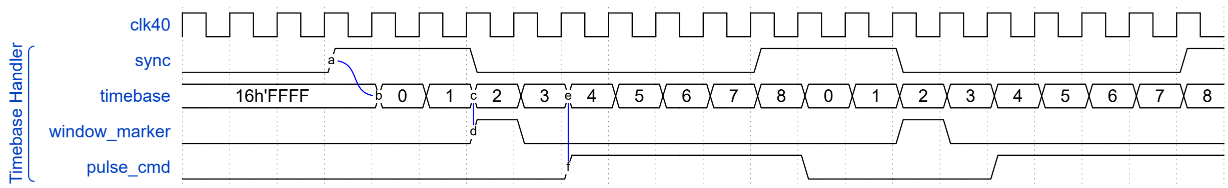


Figure C.1: Timebase handler example: integration window with an injected pulse command (`pulse_cmd`).

In some applications, such as frontend characterization, it is desirable to introduce a time gap between integration windows. During this gap, the pixel activity is ignored. To support this feature, the timebase handler can be configured with a fixed duration of the integration window, generating a `window_end` signal that explicitly closes the current window before the arrival of the next `sync`.

Figure C.2 shows a case with a 6-clock-cycle gap, using:

- `sync` interval: 9 clock cycles.
- `WINDOW_MARKER_LATENCY`: 2 clock cycles.
- `INTEGRATION_WINDOW_DURATION`: 3 clock cycles.
- `gap` duration: 9 - 3 = 6 clock cycles.

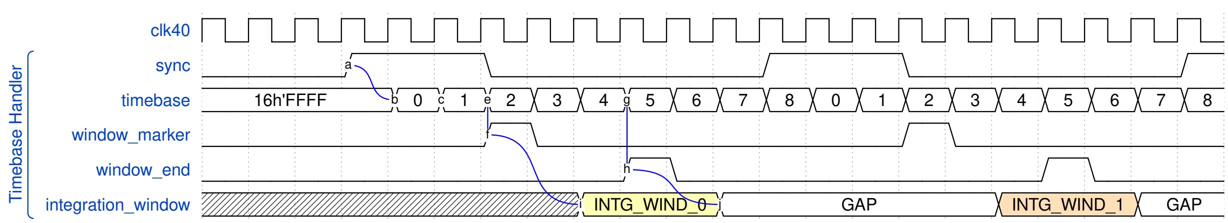


Figure C.2: Integration window with gap, generated by the timebase handler.

In addition to using external `sync` pulses, the timebase handler supports an internal sequencer mode. In this mode, internally generated `sync` pulses mimic the behavior of the external ones. This is primarily used for testing or pixel characterization and is not enabled during normal LHC operation. The sequencer can be initialized by either a global `sync` or a slow control command.

The sequencer mode works in sequencer runs, which are composed of a configurable number of `sync` pulses at a configurable frequency. At the end of a sequencer run, the timebase handler generates a `sequencer_end` pulse to indicate this state to the rest of the TRU circuitry.

C.1.2 Integration Window Indexer

The integration window indexer tracks the amount and status of the integration windows that were propagated to the matrix but have not yet produced a packet. It performs the following tasks:

- Keeps track of the status of each integration window packet. This status can be:
  - **accepted:** pixel hits recorded and transferred successfully.
  - **dropped:** pixel hits were waived using `hold`.
- Keeps the information in memory until the corresponding integration window packet is produced/processed.
- Allows data packets to be tagged with correct timing information.

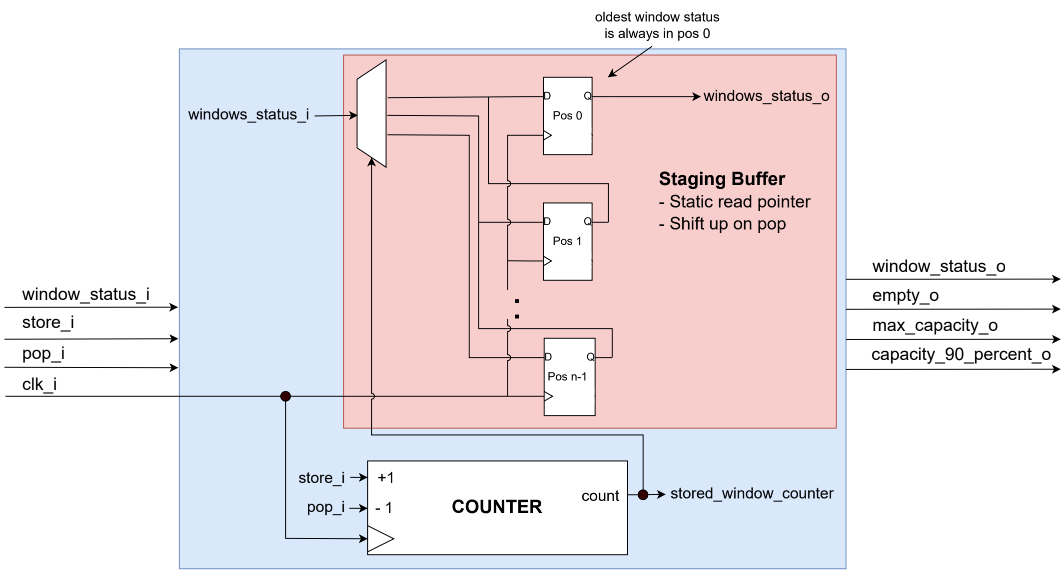


Figure C.3: Integration window indexer digital architecture, composed of a staging buffer and a counter.

Figure C.3 shows the schematic of the indexer architecture. As seen, it consists of two components: a staging **buffer** and a **counter**. The staging buffer is a shift register that operates as a queue, where each bit represents the status of an integration window. The oldest window is stored at position 0. A `store` or `pop` pulse shifts the shift register accordingly.

The counter tracks the number of entries stored in the indexer and generates status flags such as:

- `empty`.
- `90_percent_capacity`.
- `max_capacity`.

C.1.3 Buffer Depth Calculation

To determine the optimal depth of the indexer buffer, a worst-case *black event* (i.e., all pixels firing in one integration window) was initially considered. The time to transmit such an event is:

$$n_{\text{black event}} = (444 \times 156) \times \frac{16 \text{ bits}}{160 \text{ MHz}} = 3463 \approx 8192\ \mu\text{s}$$

8192 μs means a total of 4096 2μs integration windows, each of them having a status flag to be temporally stored in the integration window indexer. Storing that amount of status flags is impractical due to area and power constraints. Therefore, a realistic indexer depth was chosen using simulations from the model in Chapter 4. Simulating 750,000 integration windows (2 s each), the indexer reached a peak depth of 36 stored status.

As a result, the indexer depth was dimensioned to 48 positions, providing margin while remaining resource-efficient.

C.1.4 Readout Status Counters

To keep track of the status of the TRU, several counters are placed at different stages of the readout flow. These counters can be read through the slow control. In case any of these counters overflows, it will start counting back from 0. These status counters are:

- **Window Marker Counter:** Counts the number of `window_marker` pulses produced by the `timebase_handler`.
- **Window Marker Dropped Counter:** Counts the number of `window_marker` pulses produced by the `timebase_handler` whose status was stored as *dropped* inside the integration window indexer.
- **Integration Window Counter:** Counts the number of integration windows that were read by the `tru_sm` and produce its own integration window packet. The value of the integration window counter is used as a timestamp for the next integration window packet generated by the `tru_sm`.
- **Processed Integration Window Counter:** Counts the number of non-dropped integration windows that were read by the *tru_sm*.

C.1.5 Integration Window FSM

The `integration_window_sm` controls the `frame` and `hold` signals of the pixel matrix. These signals delimit integration windows and control their status. Details of their function are provided in Chapter 5.

These states are grouped into operations. The different operations the `integration_window_sm` can perform are described in the following subsections C.1.5.1, C.1.5.2and C.1.5.3.

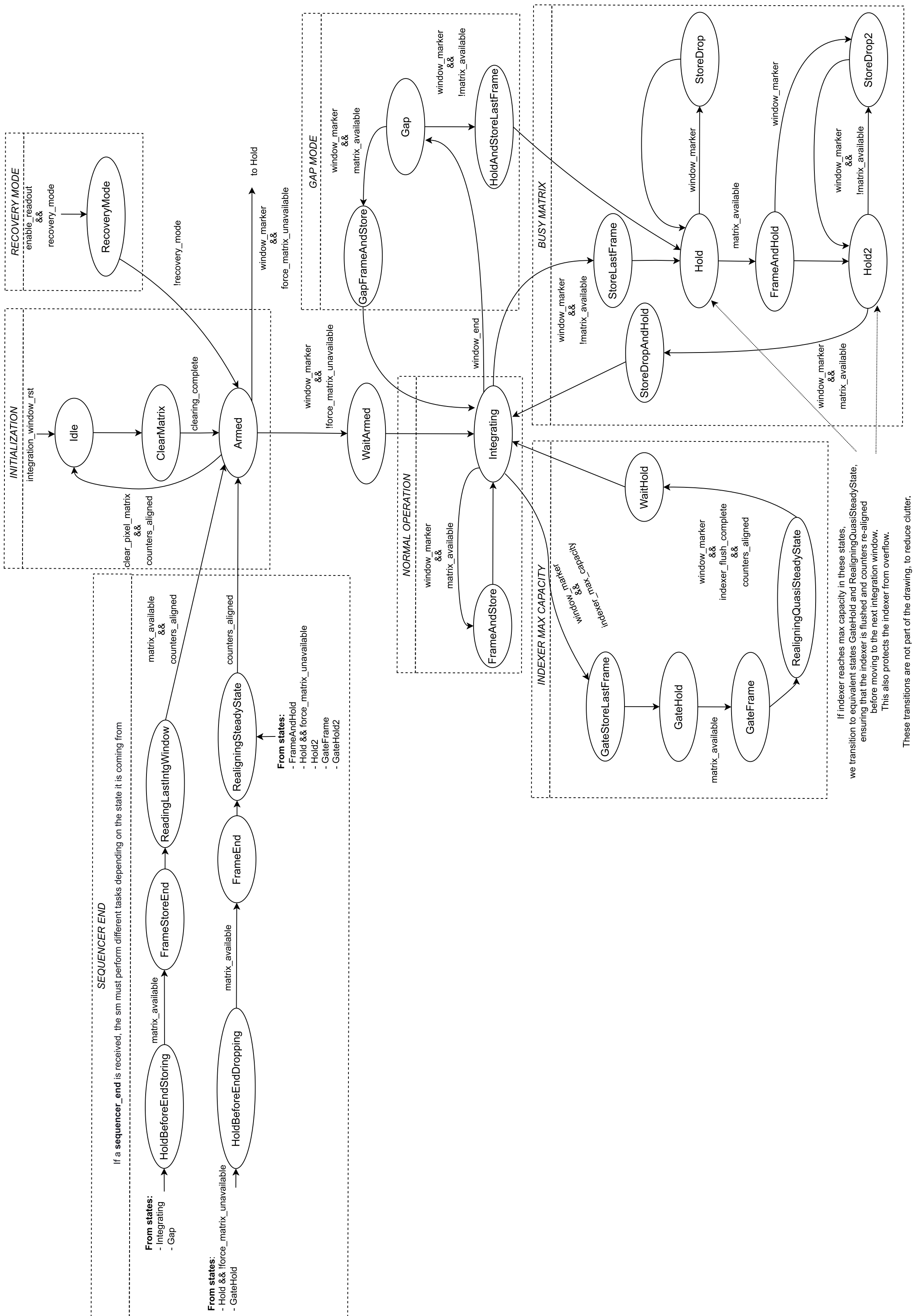


Figure C.4: Simplified Finite State Machine diagram of the `integration_window_sm`.

C.1.5.1 Initialization and Recovery Mode

The `integration_window_sm` first task is to initialize the pixel matrix following the process shown in Figure C.5. Once reset is released, the FSM clears the pixel matrix and enters in `ARMED` state, awaiting a `window_marker`.

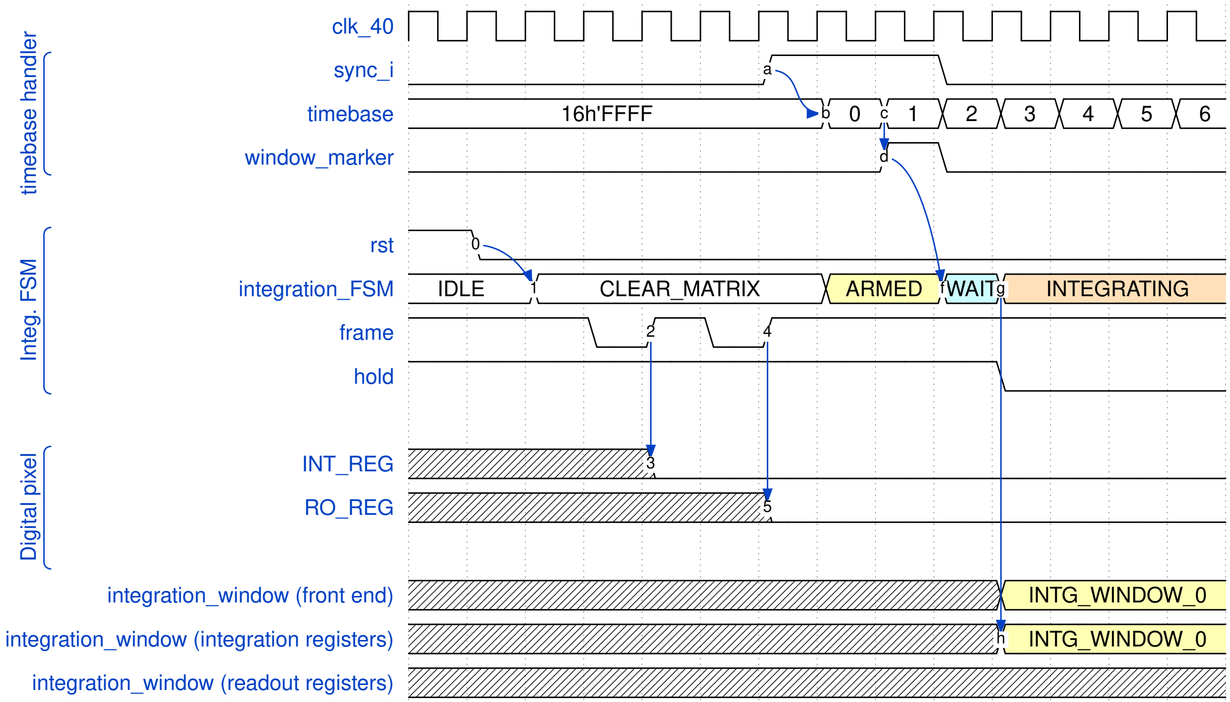


Figure C.5: Pixel matrix initialization by the `integration_window_sm`.

This FSM also supports a *recovery mode* in which the pixel matrix can be cleared upon request from the slow control. The procedure for clearing the pixel matrix is the same as for initializing it.

C.1.5.2 Normal Operation and Gap

Once the first integration window begins, `window_marker` signals are translated into `frame` pulses that delimit the integration window's pixel hit storage, marking the end of the previous integration windows and the beginning of the next. This normal mode operation can be seen in Figure C.6.

In certain applications, it is desired to add a gap in the integration window. During this gap, no hits will be recorded by the integration registers. To make this gap, the signal `window_end` must close the current integration window without opening a new one.

To do this, `hold` is asserted 1 clock cycle after the `window_end` is received. Once a new `window_marker` is received, the hold will be de-asserted, and a `frame` pulse will be sent. This will transfer the pixel hits from the integration registers to the readout registers while opening a new integration window. Figure C.7 shows the generation of integration windows with a gap of three clock cycles.

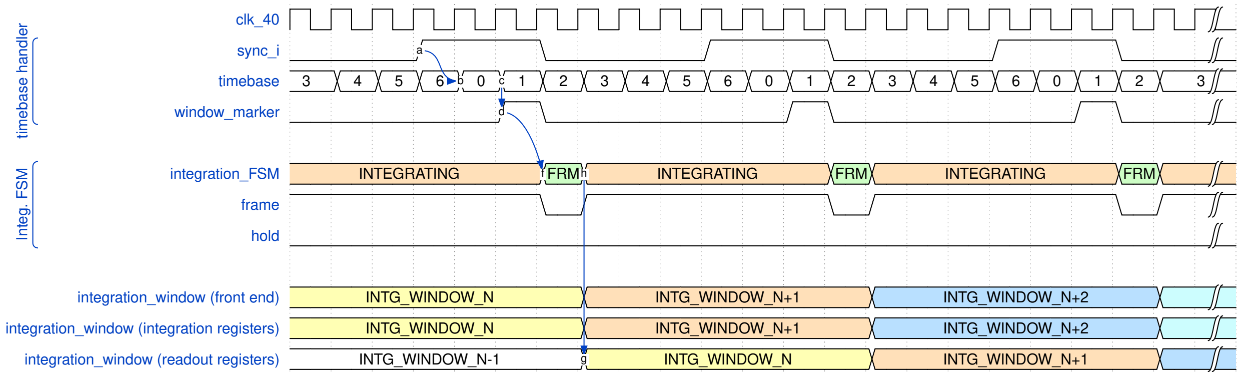


Figure C.6: Normal operation of the `integration_window_sm`.

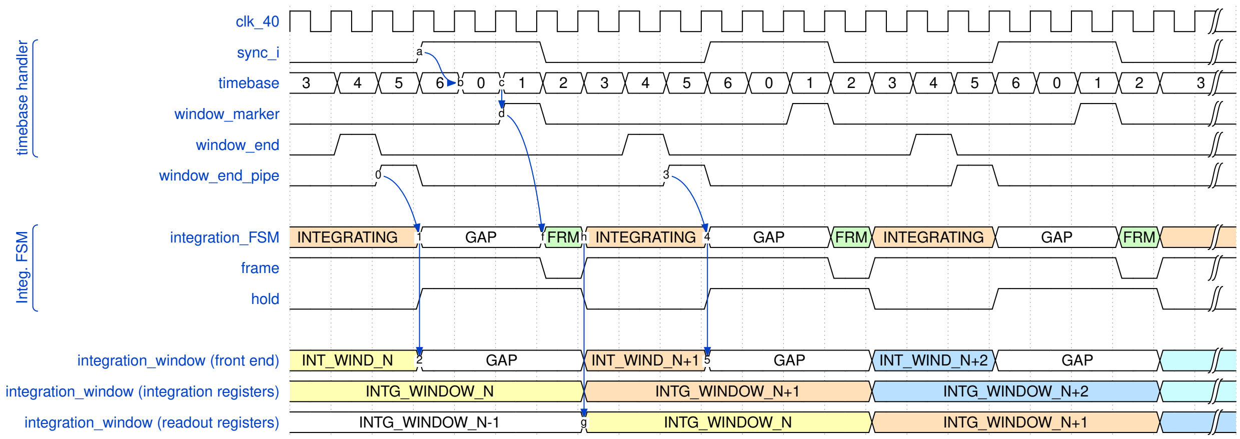


Figure C.7: `integration_window_sm` operating with integration window gaps.

C.1.5.3 Busy Matrix

We must complete the readout of all the hits of an integration window (REQ04-4-02). If the matrix is still busy when a `window_marker` arrives, `hold` is asserted, discarding all upcoming hits. While `hold` is asserted, the `window_marker` pulses will not be propagated as `frame` pulses.

When all Readout Registers are read, a frame pulse will be sent while `hold` is stay asserted. This frame pulse is not aligned with the `window_marker`, but will transfer the values from the Integration Register to the Readout Register without opening a new integration window. `hold` will only be de-asserted when all the pixel hits stored in the Integration and Readout Registers are fully read.

Figure C.8 illustrates two consecutive dropped integration windows due to a busy matrix.

C.1.6 Indexer Max Capacity

When the Indexer reaches maximum capacity, the TRU is no longer able to handle new requests for integration windows. Reading out data from older integration windows is prioritized above producing new data. To recover from this situation, the `integration_window_sm` follows

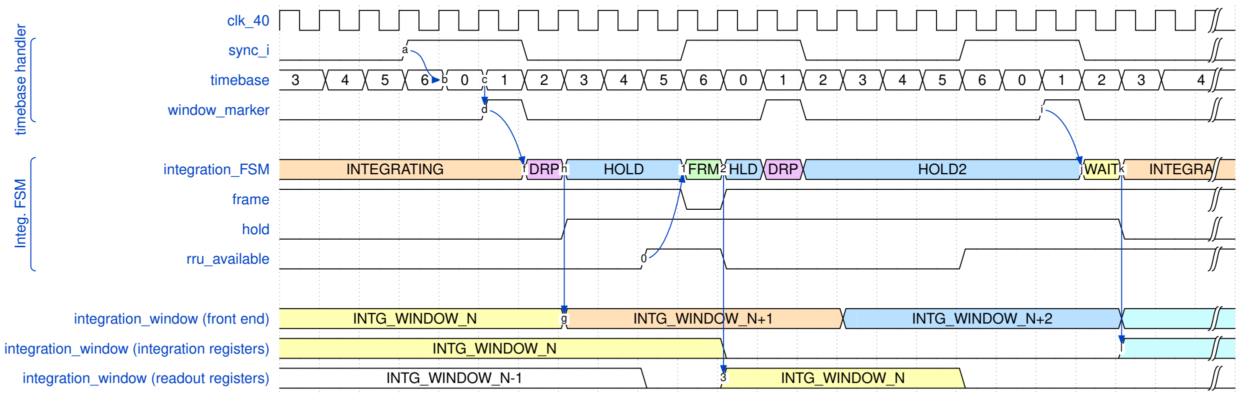


Figure C.8: Dropped integration windows due to matrix pixel hits remain in the matrix longer than an integration window.

this sequence:

- `integration_window_sm` stops propagating `window_markers` and asserts `hold`.
- All incoming window markers are marked as dropped, the window marker counter continues incrementing.
- The existing indexer contents are processed normally by the `tru_sm`.
- When the Indexer eventually is empty, the `tru_sm` continues transmitting dropped packet until the window marker counter matches the integration window marker counter.
- Upon counter match, transmission of dropped packets stops, as the TRU is back in sync with the `sync` pulse.
- The next valid `window_marker` resumes the integration process.

C.1.6.1 Sequencer End

In sequencer mode, the timebase handler emits a `sequencer_end` pulse at the end of a sequencer run. When this occurs, the `integration_window_sm` must:

- Wait for all data left in the pixel matrix, region FIFOs, and indexer to be processed.
- Return to the `ARMED` state, ready for the next sequencer run.

C.2 Region Readout Unit (RRU)

The Region Readout Units (RRUs) are responsible for reading the pixel hits from the four matrix regions, storing them temporarily in the FIFOs of the region until they are read out by the `tru_sm` FSM.

C.2.1 RRU FSM

The `rru_sm` controls the readout of the pixel matrix. Its state diagram is shown in Figure C.9. It begins in the `IDLE` state and transitions to `ARMED` when a `frame` pulse is sent to the pixel matrix.

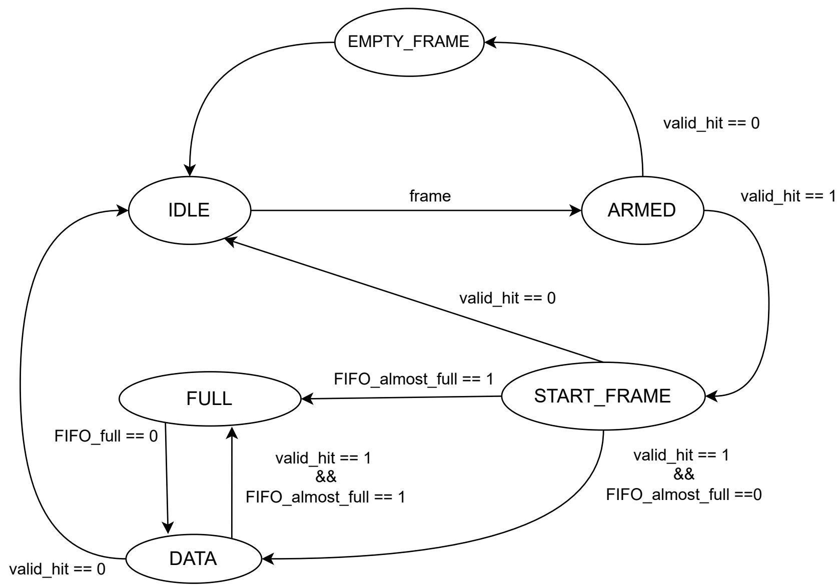


Figure C.9: States and transitions of the `rru_sm`.

If no hits are detected in the region (`any_valid_hit` == '0'), the FSM writes an empty packet to the FIFO. If at least one hit is present, the FSM writes an integration window start word, and starts transferring the pixel addresses of the hit pixels from the matrix into the FIFO. Once all hits are read, the FSM returns to the `IDLE` state, waiting for the next integration window.

If the FIFO becomes full, the FSM transitions to a `FIFO_FULL` state and halts further writes until space is available.

C.2.2 Region Address Ordering Correctness Mechanism

Duplicate hits from the same pixel are not expected in a single integration window, but may occur due to single-event effects (SEEs) or stuck pixels. If an out-of-order address is detected, the FSM skips the corresponding double column for the remainder of the window and logs the error in the FIFO. This ensures robustness and prevents a deadlock status in the TRU.

C.2.3 Region FIFO

Each region FIFO has a depth of 160 words (as defined by the analysis in Chapter 4) and a width of 17 bits. Each word is structured as follows:

```
[window_start][empty_frame][15-bit pixel hit address]
```

where:

- `window_start` bit marks the beginning of a region packet.
- `empty_frame` bit indicates that the region contains no pixel hits.
- The 15-bit field encodes the pixel address.

If a region contains no hits during a given integration window, a single word is written with both `window_start` and `empty_frame` set to '1'.

The end of an integration window is inferred either by detecting the next `window_start` or by the absence of additional hits in both the FIFO and the pixel matrix.

C.3 TRU FSM

The `tru_sm` is responsible for reading the region FIFOs and integration window indexer, constructing integration window packets. The FSM protocol supports four packet types: three integration window packets (Data, Empty and Dropped), and a Lock packet. These packets are built using words from the protocol shown in Table C.1. The FSM state diagram is shown in Figure C.10.

Table C.1: Readout protocol supported by the `tru_sm`.

<table>
<tr><th colspan="17">READOUT PROTOCOL</th></tr>
<tr><th>Word type</th><th>15</th><th>14</th><th>13</th><th>12</th><th>11</th><th>10</th><th>9</th><th>8</th><th>7</th><th>6</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>0</th></tr>
<tr><td>IDLE</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>1</td><td colspan="8">IDLE config [7:0]</td></tr>
<tr><td>LOCK</td><td>1</td><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td colspan="8">IDLE config [7:0]</td></tr>
<tr><td>READOUT_STATUS_1</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td><td colspan="8">Integration Window Counter [15:8]</td></tr>
<tr><td>READOUT_STATUS_2</td><td>1</td><td>1</td><td>1</td><td>1</td><td colspan="4">IWC Checksum [3:0]</td><td colspan="8">Integration Window Counter [7:0]</td></tr>
<tr><td>TILE_HEADER</td><td>1</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td colspan="8">Tile Address [7:0]</td></tr>
<tr><td>REGION_HEADER</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td><td>1</td><td colspan="2">Region Address [1:0]</td><td colspan="8">Integration Window Counter [7:0]</td></tr>
<tr><td>DATA</td><td>0</td><td colspan="9">Pixel Address Row [8:0]</td><td colspan="6">Pixel Address Column [5:0]</td></tr>
<tr><td>TILE_TRAILER</td><td>1</td><td>0</td><td>1</td><td>1</td><td colspan="4">TRU Status Flags [3:0]</td><td colspan="8">Packet Checksum [7:0]</td></tr>
<tr><td>EMPTY_INTEGRATION_WINDOW</td><td>1</td><td>1</td><td>1</td><td>0</td><td colspan="4">TRU Status Flags [3:0]</td><td colspan="8">Integration Window Counter [7:0]</td></tr>
<tr><td>DROPPED_INTEGRATION_WINDOW</td><td>1</td><td>1</td><td>0</td><td>1</td><td colspan="4">TRU Status Flags [3:0]</td><td colspan="8">Integration Window Counter [7:0]</td></tr>
<tr><td>RECOVERY</td><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td colspan="8">'0</td></tr>
<tr><td>RESERVED</td><td>1</td><td>0</td><td>0</td><td>0</td><td colspan="12">RESERVED</td></tr>
</table>

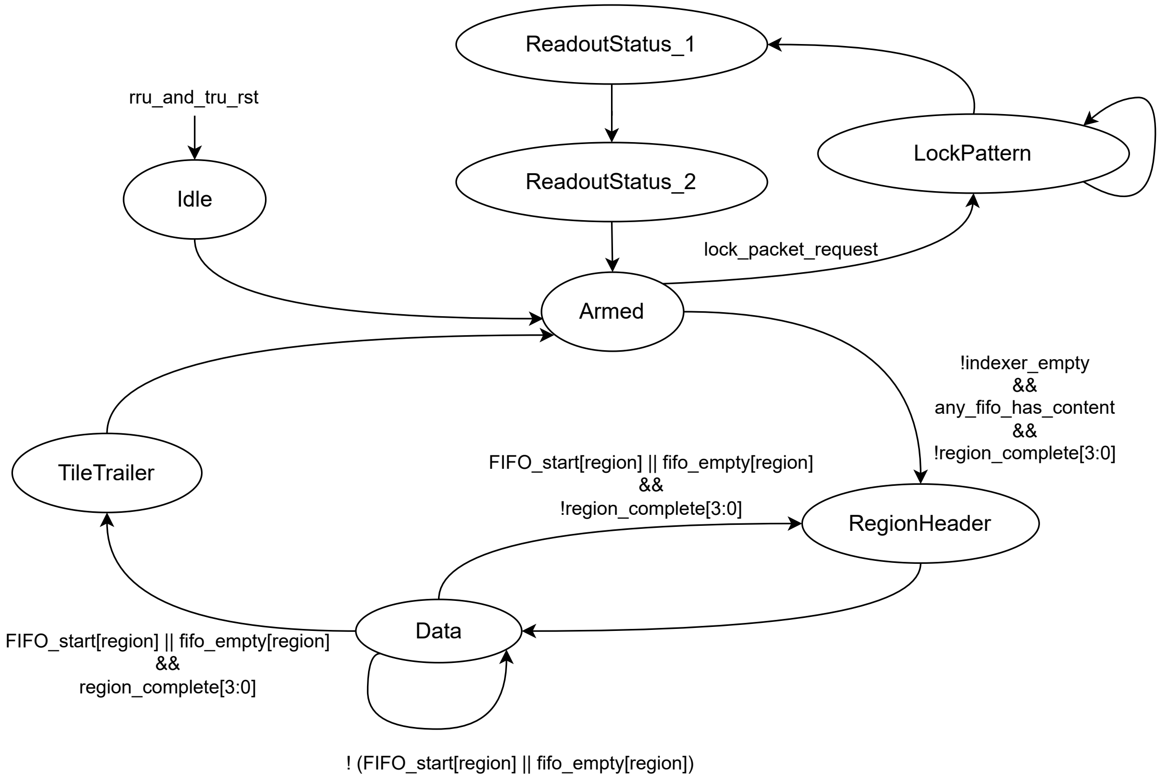


Figure C.10: TRU FSM states and transitions. It reads the integration windows status from the indexer, the pixel hits addresses from the region FIFO, and builds the packets according to the protocol words.

C.3.1 Integration Window Packet Types

Data packets are generated if there is at least one hit in the corresponding integration window. The packet format is based in the study presented in Section 4.3.1 and consists of:

- `TILE_HEADER`
- One or more `REGION_HEADER`s (only for regions with hits)
- Corresponding pixel `DATA` words inside that region
- `TILE_TRAILER`

If no hits are recorded, a single-word `EMPTY_INTEGRATION_WINDOW` packet is transmitted. If an integration window request was dropped, a single-word `DROPPED_INTEGRATION_WINDOW` packet is sent.

C.3.2 Lock Packet

A lock packet is a standalone packet composed of 6 `LOCK` words and the *READOUT_STATUS_1* and *READOUT_STATUS_2* words. It can be sent at any time, as long as it does not interrupt another packet.

These packets synchronize the backend system with the incoming data stream, ensuring both word and time alignment. Lock packets are sent periodically and before the first data packet after initialization. They include the full integration window counter for the **next** data packet and a checksum.

C.3.3 Protocol Details

The TRU continuously sends protocol words to the LEC. If no packets are ready for transmission, the TRU outputs `IDLE` words. The number of `IDLE` words between packets depends on packet size and timing.

The MOSAIX data protocol includes features for robustness and debug support:

- **TRU Status Flags:** Indicate the region FIFO and WINMU status.
- **Checksums:** Validate both data and synchronization integrity.
- **Recovery Mode:** User-enabled mode to halt normal transmission and allow recovery from unforeseen faults.
- **Debug Mode:** Allows manual stepping through data for troubleshooting.

C.3.3.1 Status Flags

The TRU Status Flags contain information about the current state of the TRU.

Flag bits [3:2] (Table C.2) contain information about the Integration Window Indexer. The flags [1] and [0] (Table C.3) describe the status of the region FIFOs.

Table C.2: Description of FLAGS[3:2] and corresponding events.

FLAGS[3:2]	Priority	Event
2'b11	1	Indexer empty after max capacity. Re-aligning the integration window counter with the window marker counter.
2'b10	2	Indexer reached max capacity. This flag will remain high until all the content from the indexer is read out.
2'b01	3	Indexer at 90% capacity.
2'b00	4	Indexer is healthy.

C.3.3.2 Checksums

Checksums are used to verify data integrity during transmission. Each data packet and readout status packet contains a checksum. The data packet checksum is calculated as:

Table C.3: Description of FLAGS[1:0] and associated events.

FLAG BIT POS	Event
[1]	At least one Region FIFO has reached max capacity during the readout of the integration window.
[0]	At least one Region FIFO has reached 90% capacity during the readout of the integration window.

$$
\begin{aligned}
Checksum[0] &= protocol\_word[0] \oplus protocol\_word[1] \oplus previous\_checksum[0] \\
Checksum[1] &= protocol\_word[2] \oplus protocol\_word[3] \oplus previous\_checksum[1] \\
Checksum[2] &= protocol\_word[4] \oplus protocol\_word[5] \oplus previous\_checksum[2]
\end{aligned}
$$

Where $\oplus$ represents an XOR operation.

The integration window checksum is calculated as:

$$
\begin{aligned}
Checksum[0] &= \oplus\ integration\_window\_counter[0:3] \\
Checksum[1] &= \oplus\ integration\_window\_counter[4:7] \\
Checksum[2] &= \oplus\ integration\_window\_counter[8:11] \\
Checksum[3] &= \oplus\ integration\_window\_counter[12:15]
\end{aligned}
$$

Where $\oplus$ before a vector represents the boolean XOR between all the bits of the vector.

C.3.3.3 Special Modes and reserved words

Apart from the TRU normal operation, the TRU can be set into recovery or debug mode on demand by the user.

When the recovery mode is activated, the TRU transmits only `RECOVERY` words. In this mode, the TRU allows the clearing and restarting of all FSM states, the indexer, the FIFOS, the counters and the pixel matrix.

When in debug mode, the `tru_sm` do not change the sending word every time the serializer is available. Instead, it repeats the same until manually stepped via slow control.

Additionally, a `RESERVED` word is defined in the protocol, allowing the backend system to insert diagnostic data into the MOSAIX data stream.

C.3.4 Pixel Address Encoding

Each pixel address has two components:

- Row (rϕ): 9 bits (from the Priority Encoder)
- Column (z): 8 bits total:
  - Region Header, upper 2 bits (03)
  - Priority Encoder output, lower 6 bits (063)

The region information is sent in order: |1-2-3-4|1-2-3-4|1-2-3-4|...

C.4 Pixel Clearing Mechanism

The periphery is capable of reading the pixel matrix at either 40 MHz or 20 MHz. As the `clr_pix` pulses at 40 MHz must be generated from the general 40 MHz clock, both rising and falling edges must be comprised within a single clock cycle. This is achieved using the Double Data Rate (DDR) mechanism illustrated in Figure C.11.

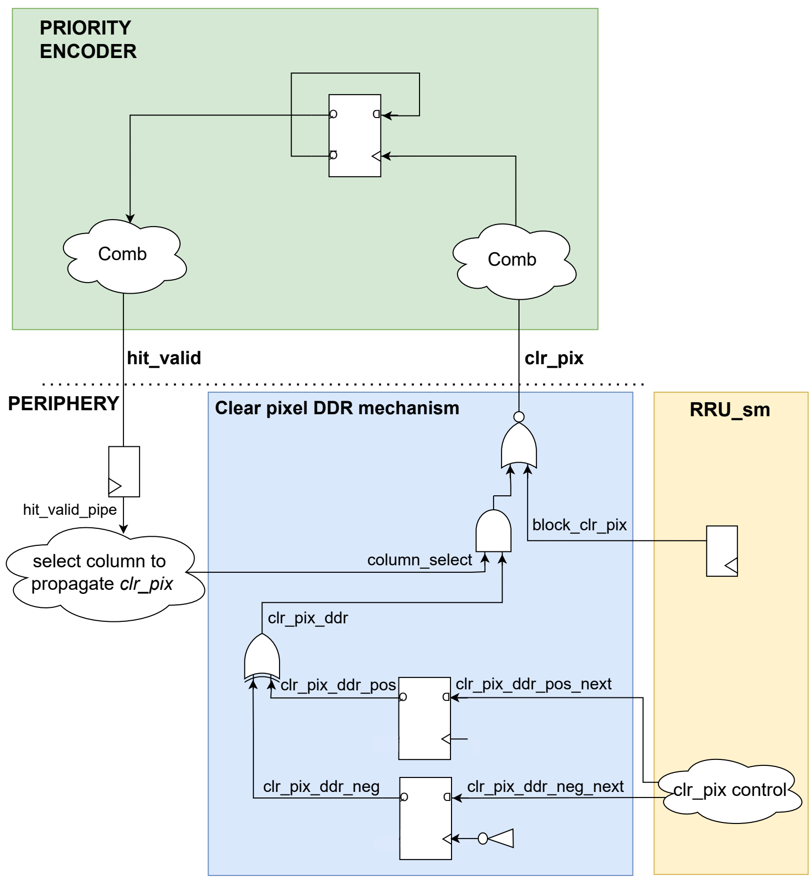


Figure C.11: Clear pixel DDR mechanism from the periphery to the pixel matrix.

As seen, the 40 MHz pulses are generated by XORing rising-edge and falling-edge flip-flops. This DDR XOR output (`clr_pix_ddr`) is then ANDed with a column selection signal, ensuring that the `clr_pix` signal is propagated only to the first double column following a pre-defined clearing priority.

The `clr_pix` signal can also be gated by the `block_clr_pix` signal, which is controlled by the `rru_sm` state machine. This gating is due to the selected digital pixel architecture (Chapter 5.4.2), which only allows the propagation of the `frame` pulse if `clr_pix` is '0'.

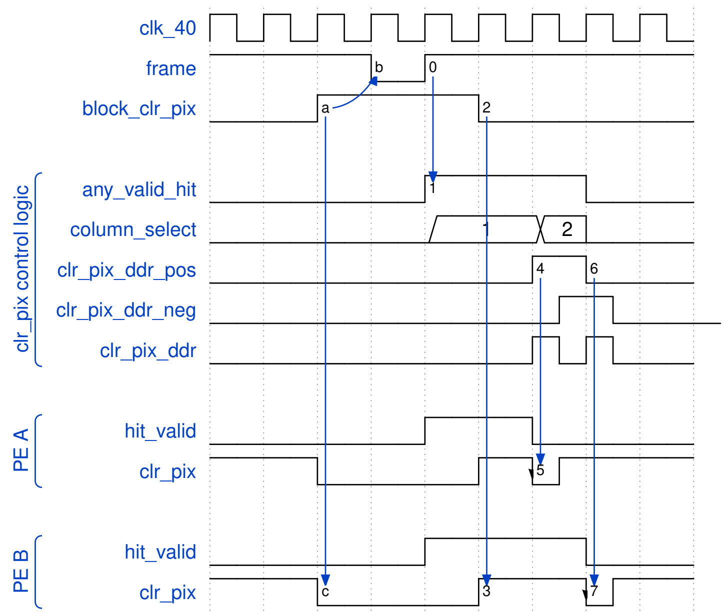


Figure C.12: Waveform example showing the readout of 2 PEs, each with one pixel hit, at 40 MHz.

Figure C.12 illustrates a waveform example involving two pixel hits in *PE_A* and *PE_B*, read at 40 MHz. In this waveform, we can observe:

- Pixel hits detected during the previous integration window
- (a-b-c) `block_clr_pix` asserted to allow `frame` to propagate to all in-pixel readout registers
- (0-1) `frame` pulse loads values from the integration registers into the readout registers, `hit_valid` asserted
- (2-3) `block_clr_pix` de-asserted; all `clr_pix` lines armed
- (4-5) First pixel hit cleared
- (6-7) Second pixel hit cleared

The reduction of the `clr_pix` frequency to 20 MHz is performed by disabling the falling-edge flip-flop. This way, only the rising-edge flip-flop toggles, producing a single transition per clock cycle. In Figure C.13, a waveform with the readout of one pixel hits in *PE_A* and another in *PE_B* is shown.

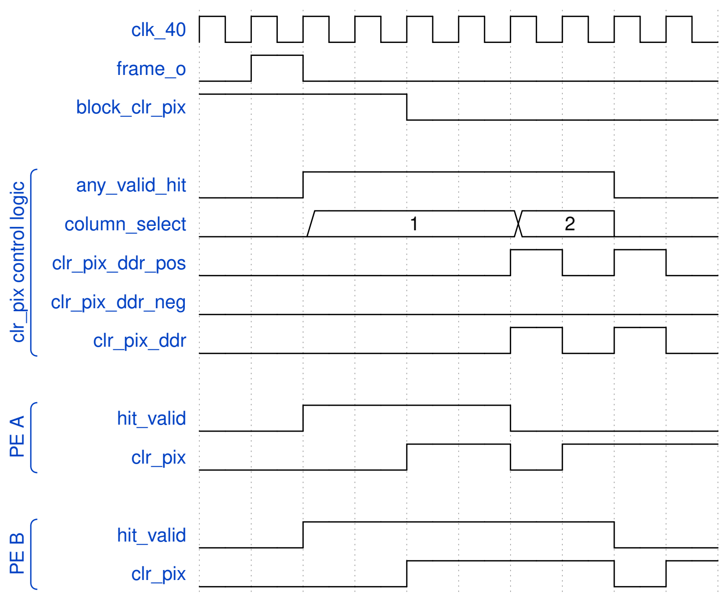


Figure C.13: Waveform example showing the readout of 2 PEs, each with one pixel hit, at 20 MHz.

Appendix D

Low-Leackage, DFM Standard Cell Library

The pixel matrix contains the majority of transistors in the MOSAIX chip. However, given the low occupancy of pixel hits ($1.8 \cdot 10^{-4}$ hits every 2 μs), the dynamic power consumption of these cells is minimal. The dominant contributor to power consumption in the pixel matrix is therefore the leakage. Leakage can be significantly reduced by increasing the gate length of transistors; however, this also increases the cell's delay. Since connections between the pixel matrix and periphery are timing-critical arcs, a trade-off between these two characteristics must be made.

A study of leakage and delay in inverters with varying gate lengths was conducted. Results are shown in Figure D.1.

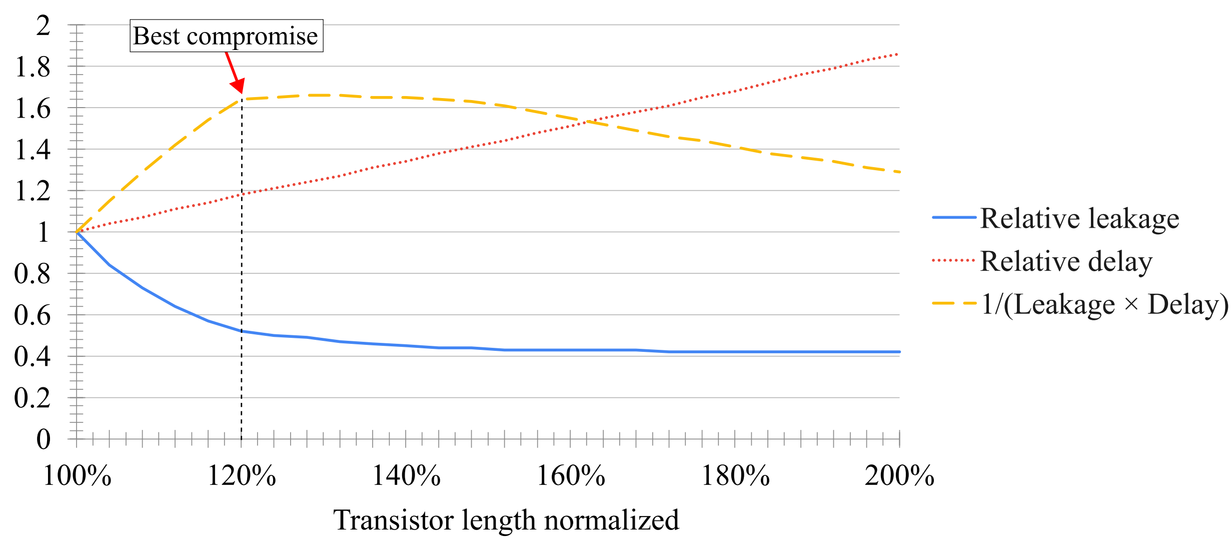


Figure D.1: Leakage and delay normalized over nMOS transistor length. Delay increases linearly with length, while leakage drops rapidly until approximately 20 % increase, and then flattens.

As shown, the delay (red, dotted) increases linearly with gate length, while the leakage (blue, straight) decreases sharply from minimum size to 20 % increase, tapering off beyond that. The combined figure of merit (yellow, dashed) suggests that a 20 % length increase offers the best compromise, halving leakage while incurring only an additional 20% delay. Therefore, a gate length 20 % larger was selected for both nMOS and pMOS transistors in all standard cells used in the pixel matrix.

In addition to leakage reduction, the MOSAIX design must be highly robust from a DFM perspective. Given the large chip area, the risk of random defects such as shorts is elevated.

To enhance DFM robustness, the new custom library incorporates:

- Increased via count, avoiding the usage of single vias
- Wider metal tracks
- Greater metal spacing

These enhancements could not fit within the original 12-track (12T) height of the foundry-provided cells. Therefore, the new standard cells were built with a 13-track (13T) height. Moreover, as described in Chapter 5.2.1, the NMOS substrates in the pixel matrix are biased at -1.2 V rather than connected to VSS. Thus, VSS-substrate contacts are removed, and dedicated TAP cells connect the substrate to -1.2 V are placed all over the pixel matrix. All these modifications come at the price of area, with an increase of approximately 30 % per cell.

Table D.1: Custom low-leakage, DFM standard cell library content used in the pixel matrix physical implementation.

TYPE	**Count**
BUFFERS	7
INVERTERS	8
DELAY BUFFERS	2
NAND	4
NOR	2
AND	1
OR	2
XOR	1
MUX	1
CUSTOM GATES	6
LATCH	1
FLIP-FLOP	4
TIE-HIGH	1
TIE-LOW	1
FILLER	11
ENDCAP	1
ANTENA	1
TAP CELL	1
TOTAL	**55**

This design process was repeated for a wide range of logic cells to create the complete CERN standard cell library used in the pixel matrix. The cell distribution is summarized in Table D.1.

References

[1] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version* (Physics textbook). New York, NY: Wiley, 2008. [Online]. Available: https://cds.cern.ch/record/111880.

[2] M. Slupecki, "The Fast Interaction Trigger for the ALICE Upgrade", Jyvaskyla U., 2020. [Online]. Available: http://cds.cern.ch/record/2741462.

[3] ALICE Collaboration, *The ALICE experiment – A journey through QCD*, 2022. arXiv: 2211.04384 [nucl-ex].

[4] A. Lopes, "Recreating Big Bang matter on Earth", Nov. 13, 2020. [Online]. Available: https://home.cern/news/series/lhc-physics-ten/recreating-big-bang-matter-earth.

[5] L. Evans and P. Bryant, "LHC Machine", *JINST*, vol. 3, no. 08, S08001, Aug. 2008. DOI: 10.1088/1748-0221/3/08/S08001. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/3/08/S08001.

[6] R. Bruce, T. Argyropoulos, H. Bartosik, *et al.*, "HL-LHC operational scenarios for Pb-Pb and p-Pb operation", CERN, Geneva, Tech. Rep., 2020. [Online]. Available: https://cds.cern.ch/record/2722753.

[7] R. Bruce, R. Alemany Fernandez, T. Argyropoulos, *et al.*, "First results of running the LHC with lead ions at a beam energy of 6.8 Z TeV", *J. Phys. Conf. Ser.*, vol. 2687, no. 2, p. 022 001, Jan. 2024. DOI: 10.1088/1742-6596/2687/2/022001. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/2687/2/022001.

[8] L. Musa, *et al.* (ALICE Collaboration), "Letter of Intent for an ALICE ITS Upgrade in LS3", CERN, Geneva, Tech. Rep., 2019. [Online]. Available: https://cds.cern.ch/record/2703140.

[9] ALICE collaboration, "Technical Design report for the ALICE Inner Tracking System 3 - ITS3; A bent wafer-scale monolithic pixel detector", CERN, Geneva, Tech. Rep., 2024. [Online]. Available: https://cds.cern.ch/record/2890181.

[10] G. Aglieri, "Monolithic active pixel sensor development for the upgrade of the ALICE Inner Tracking system", *JINST*, vol. 8, no. 12, p. C12041, Dec. 2013. DOI: 10.1088/1748-0221/8/12/C12041. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/8/12/C12041.

[11] P. V. Leitao, "Development of a Stitched Monolithic Pixel Sensor prototype (MOSS chip) towards the ITS3 upgrade of the ALICE Inner Tracking system", *JINST*, vol. 18, no. 01, p. C01044, Jan. 2023. DOI: 10.1088/1748-0221/18/01/C01044. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/18/01/C01044.

[12] A. Fant, "The Art of CMOS Stitching: Challenges and Solutions", Sep. 1, 2024. [Online]. Available: https://www.isdicmos.com/news-events/2024/10/30/the-art-of-cmos-stitching-challenges-and-solutions.

[13] B. Blidaru, on behalf of the ALICE collaboration, "Testbeam performance results of bent ALPIDE monolithic active pixel sensors in view of the ALICE Inner Tracking System 3", *Journal of Instrumentation*, vol. 17, no. 09, p. C09006, Sep. 2022. DOI: 10.1088/1748-0221/17/09/C09006. [Online]. Available: https://doi.org/10.1088/1748-0221/17/09/C09006.

[14] A. Andronic, P. Becht, M. B. Blidaru, *et al.*, "Detection efficiency and spatial resolution of Monolithic Active Pixel Sensors bent to different radii", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1083, p. 171 131, 2026. DOI: https://doi.org/10.1016/j.nima.2025.171131. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900225009337.

[15] M. Rossewij, E. Okkinga, H. Naqvi, R. Barthel, S. Alving, and on behalf of the ALICE collaboration, "Electrical/piezoresistive effects in bent Alpide MAPS", *Journal of Instrumentation*, vol. 19, no. 04, p. C04057, Apr. 2024. DOI: 10.1088/1748-0221/19/04/C04057. [Online]. Available: https://doi.org/10.1088/1748-0221/19/04/C04057.

[16] M. Buckland, *et al.* (ALICE Collaboration), "First measurements with monolithic active pixel test structures produced in a 65 nm CMOS process", *JINST*, vol. 19, no. 02, p. C02017, Feb. 2024. DOI: 10.1088/1748-0221/19/02/C02017. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/19/02/C02017.

[17] M. Selina, *et al.* (ALICE Collaboration), "Exploring unique design features of the Monolithic Stitched Sensor with Timing (MOST): yield, powering, timing, and sensor reverse bias", *JINST*, vol. 20, no. 07, p. C07032, Jul. 2025. DOI: 10.1088/1748-0221/20/07/C07032. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/20/07/C07032.

[18] A. Tauro, "3D ALICE Schematic RUN3 - with Description", Jun. 19, 2019. [Online]. Available: https://alice-figure.web.cern.ch/node/11220.

[19] D. Dobrigkeit Chinellato, "ALICE event displays and geometry rendering of lead-lead ion collisions during Run 3", Nov. 23, 2022. [Online]. Available: https://cds.cern.ch/record/2841865.

[20] S. Fenkart, "CERNs neutrino odyssey", Sep. 3, 2023. [Online]. Available: https://cerncourier.com/a/cerns-neutrino-odyssey/.

[21] A. Falcone, "Deep underground neutrino experiment: DUNE", *Nucl. Instrum. Methods Phys. Res.*, vol. 1041, p. 167 217, 2022. DOI: https://doi.org/10.1016/j.nima.2022.167217. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900222005721.

[22] R. D. Evans, *The Atomic nucleus*. Krieger, 1982.

[23] D. I. Kaiser, "Physics and Feynman's diagrams", *Am. Sci.*, vol. 93, pp. 156–165, 2005. DOI: 10.1511/2005.2.156.

[24] S. Hayakawa, *Cosmic ray physics*. John Wiley & Sons: Wiley Interscience, 1969.

[25] Particle Data Group, "Review of Particle Properties", *Phys. Lett. B*, 1990.

[26] Particle Data Group, "Review of Particle Properties", *Phys. Rev. D*, 1992. [Online]. Available: https://cds.cern.ch/record/788099.

[27] Particle Data Group, "Review of Particle Properties", *Phys. Lett. B*, 2004.

[28] B. B. Rossi, *High-energy particles*. New York, USA: Prentice-Hall, 1952.

[29] C. Grupen, "Electromagnetic Interactions of High Energy Cosmic Ray Muons", *Fortschritte der Physik*, vol. 24, no. 3, pp. 127–209, 1976. DOI: https://doi.org/10.1002/prop.19760240302. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/prop.19760240302. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/prop.19760240302.

[30] H. Kolanoski and N. Wermes, *Particle Detectors: Fundamentals and Applications*. Oxford, UK: Oxford University Press, 2020.

[31] C. Grupen and B. Shwartz, *Particle Detectors* (Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology). Cambridge, UK: Cambridge University Press, 2008.

[32] O. e. a. Kodolova, "Heavy Ion Physics with CMS". arXiv: 0907.4492. [Online]. Available: https://cds.cern.ch/record/1204545/plots#0.

[33] Brookhaven National Laboratory, "The Omega Group". [Online]. Available: https://www.bnl.gov/omega/history.php.

[34] W. B. Fowler and N. P. Samios., "The Omega-Minus Experiment", *Scientific American*, vol. 211, no. 4, pp. 36–45, 1964. [Online]. Available: http://www.jstor.org/stable/24931659 (visited on 11/05/2024).

[35] L. Rossi, P. Fischer, T. Rohe, and N. Wermes, *Pixel detectors: from fundamentals to applications* (Particle acceleration and detection). Berlin: Springer, 2006. DOI: 10.1007/3-540-28333-1. [Online]. Available: http://cds.cern.ch/record/976471.

[36] ATLAS Collaboration, "ATLAS pixel detector electronics and sensors", *JINST*, vol. 3, P07007, 2008. DOI: 10.1088/1748-0221/3/07/P07007. [Online]. Available: https://cds.cern.ch/record/1119279.

[37] L. Gonella, "The ATLAS ITk detector system for the Phase-II LHC upgrade", *Nucl. Instrum. Methods Phys. Res. A*, vol. 1045, p. 167 597, 2023. DOI: https://doi.org/10.1016/j.nima.2022.167597. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900222008890.

[38] E. Rodríguez Rodríguez, "The LHCb VELO detector: Design, operation and first results", *Nucl. Instrum. Methods Phys. Res., A*, vol. 1065, p. 169 469, 2024. DOI: 10.1016/j.nima.2024.169469. [Online]. Available: https://cds.cern.ch/record/2900026.

[39] K. Nash, on behalf of the CMS collaboration, "The Phase-2 upgrade of the CMS outer tracker", *Nucl. Instrum. Methods Phys. Res. A*, vol. 1058, p. 168 788, 2024. DOI: https://doi.org/10.1016/j.nima.2023.168788. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900223007799.

[40] W. Snoeys, "Monolithic pixel detectors for high energy physics", *Nucl. Instrum. Methods Phys. Res. A*, vol. 731, pp. 125–130, 2013, PIXEL 2012. DOI: https://doi.org/10.1016/j.nima.2013.05.073. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900213006840.

[41] J. Kemmer and G. Lutz, "New detector concepts", *Nucl. Instrum. Methods Phys. Res. A*, vol. 253, no. 3, pp. 365–377, 1987. DOI: https://doi.org/10.1016/0168-

9002(87)90518-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0168900287905183.

[42] W. Snoeys, "A process modification for CMOS monolithic active pixel sensors for enhanced depletion, timing performance and radiation tolerance", *Nucl. Instrum. Meth. A*, vol. 871, pp. 90–96, 2017. DOI: 10.1016/j.nima.2017.07.046.

[43] C. Marinas and M. Vos, "The Belle-II DEPFET pixel detector: A step forward in vertexing in the superKEKB flavour factory", *Nucl. Instrum. Methods Phys. Res. A*, vol. 650, no. 1, pp. 59–63, 2011, International Workshop on Semiconductor Pixel Detectors for Particles and Imaging 2010. DOI: https://doi.org/10.1016/j.nima.2010.12.116. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900210028962.

[44] A. Dorokhov, G. Bertolone, J. Baudot, *et al.*, "High resistivity CMOS pixel sensors and their application to the STAR PXL detector", *Nucl. Instrum. Methods Phys. Res. A*, vol. 650, no. 1, pp. 174–177, 2011, International Workshop on Semiconductor Pixel Detectors for Particles and Imaging 2010. DOI: https://doi.org/10.1016/j.nima.2010.12.112. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900210028925.

[45] L. Greiner, E. Anderssen, H. Matis, *et al.*, "A MAPS based vertex detector for the STAR experiment at RHIC", *Nucl. Instrum. Methods Phys. Res. A*, vol. 650, no. 1, pp. 68–72, 2011, International Workshop on Semiconductor Pixel Detectors for Particles and Imaging 2010. DOI: https://doi.org/10.1016/j.nima.2010.12.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900210027439.

[46] F. Reidt, "Upgrade of the ALICE ITS detector", *Nucl. Instrum. Methods Phys. Res. A*, vol. 1032, p. 166 632, Jun. 2022. DOI: 10.1016/j.nima.2022.166632. [Online]. Available: http://dx.doi.org/10.1016/j.nima.2022.166632.

[47] ALICE Collaboration, "Technical Design Report for the Upgrade of the ALICE Inner Tracking System", *J. Phys. G*, vol. 41, no. 8, p. 087 002, Jul. 2014. DOI: 10.1088/0954-3899/41/8/087002. [Online]. Available: https://dx.doi.org/10.1088/0954-3899/41/8/087002.

[48] G. Aglieri Rinella, "The ALPIDE pixel sensor chip for the upgrade of the ALICE Inner Tracking System", *Nucl. Instrum. Methods Phys. Res. A*, vol. 845, pp. 583–587, 2017, Proceedings of the Vienna Conference on Instrumentation 2016. DOI: https://doi.org/10.1016/j.nima.2016.05.016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900216303825.

[49] M. Mager, "ALPIDE, the Monolithic Active Pixel Sensor for the ALICE ITS upgrade", *Nucl. Instrum. Methods Phys. Res. A*, vol. 824, pp. 434–438, 2016, Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors. DOI: https://doi.org/10.1016/j.nima.2015.09.057. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900215011122.

[50] A. Szczepankiewicz, "Readout of the upgraded ALICE-ITS", *Nucl. Instrum. Methods Phys. Res. A*, vol. 824, pp. 465–469, 2016, Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors. DOI: https://doi.org/10.1016/j.nima.2015.10.056. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900215012681.

[51] *ALPIDE Operations Manual*, ALICE ITS ALPIDE development team, Jul. 2016. [Online]. Available: https://pct.wiki.uib.no/images/f/fc/Alpide_manual_ver3.pdf.

[52] D. Mallick, "An AI Chip With Unprecedented Performance To Do the Unimaginable - Cerebras", 2021. [Online]. Available: https://www.cerebras.ai/blog/an-ai-chip-with-unprecedented-performance-to-do-the-unimaginable.

[53] S. Lie, "Cerebras Architecture Deep Dive: First Look Inside the HW/SW Co-Design for Deep Learning : Cerebras Systems", in *2022 IEEE Hot Chips 34 Symposium (HCS)*, 2022, pp. 1–34. DOI: 10.1109/HCS55958.2022.9895479.

[54] J.-B. Kim, S.-K. Hong, and O.-K. Kwon, "A High-Speed Wafer-Scale CMOS X-Ray Detector With Column-Parallel ADCs Using Oversampling Binning Method", *IEEE Transactions on Electron Devices*, vol. 62, no. 3, pp. 888–895, 2015. DOI: 10.1109/TED.2014.2386533.

[55] R. Turchetta, N. Guerrini, and I. Sedgwick, "Large area CMOS image sensors", *JINST*, vol. 6, no. 01, p. C01099, Jan. 2011. DOI: 10.1088/1748-0221/6/01/C01099. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/6/01/C01099.

[56] J. N. Burghartz, "You can't be too thin or too flexible", *IEEE Spectrum*, vol. 50, no. 3, pp. 38–61, 2013. DOI: 10.1109/MSPEC.2013.6471057.

[57] Y. Chen, F. Liu, B. Lu, Y. Zhang, and X. Feng, "SkinLike Hybrid Integrated Circuits Conformal to Face for Continuous Respiratory Monitoring", *Advanced Electronic Materials*, vol. 6, p. 2 000 145, Jun. 2020. DOI: 10.1002/aelm.202000145.

[58] P. Jin, J. Fu, F. Wang, *et al.*, "A flexible, stretchable system for simultaneous acoustic energy transfer and communication", *Science Advances*, vol. 7, no. 40, eabg2507, 2021. DOI: 10.1126/sciadv.abg2507. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.abg2507. [Online]. Available: https://www.science.org/doi/abs/10.1126/sciadv.abg2507.

[59] W. Liu, Y. Liu, Z. Q. Yang, *et al.*, "Flexible solar cells based on foldable silicon wafers with blunted edges", *Nature*, vol. 617, pp. 717–723, May 2023. DOI: 10.1038/s41586-023-05921-z.

[60] J. Richter, "Piezoresistivity in Microsystems", English, Ph.D. dissertation, May 2008.

[61] N. Doble, L. Gatignon, E. Wilson, and K. Hübner, "WSP: The Super Proton Synchrotron (SPS): A Tale of Two Lives", *Adv. Ser. Direct. High Energy Phys.*, vol. 27, pp. 135–177, 2017.

[62] R. Diener, J. Dreyling-Eschweiler, H. Ehrlichmann, *et al.*, "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. DOI: https://doi.org/10.1016/j.nima.2018.11.133. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900218317868.

[63] ALICE Collaboration, "Centrality determination of Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV with ALICE", *Phys. Rev. C*, vol. 88, p. 044 909, 2013. DOI: 10.1103/PhysRevC.88.044909. arXiv: 1301.4361. [Online]. Available: https://cds.cern.ch/record/1500579.

[64] T. C. collaboration, "First measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s}$= 13.6 TeV", *J. High Energy Phys.*, Aug. 2023. DOI: 10.1007/JHEP08(2023)204.

[65] L. Rossi, "Superconductivity: its role, its success and its setbacks in the Large Hadron Collider of CERN", *Supercond. Sci. Technol.*, vol. 23, no. 3, p. 034 001, Feb. 2010. DOI: 10.1088/0953-2048/23/3/034001.

[66] R. Bailey and P. Collier, "Standard Filling Schemes for Various LHC Operation Modes", CERN, Geneva, Tech. Rep., 2003. [Online]. Available: https://cds.cern.ch/record/691782.

[67] R. Bruce, T. Argyropoulos, H. Bartosik, *et al.*, "HL-LHC operational scenarios for Pb-Pb and p-Pb operation", CERN, Geneva, Tech. Rep., 2020. [Online]. Available: https://cds.cern.ch/record/2722753.

[68] A. Villani, "Recent results from MAPS prototypes for ITS3", *PoS*, vol. VERTEX2023, p. 046, 2024. DOI: 10.22323/1.448.0046. [Online]. Available: https://cds.cern.ch/record/2913546.

[69] A. Trifiro, "ALICE Inner Tracking System upgrade, ITS3: characterization of first chips fabricated in 65 nm CMOS technology", p. 076, Apr. 2025. DOI: 10.22323/1.468.0076.

[70] M. Viqueira Rodríguez, on behalf of the ALICE collaboration, "Model and analysis of the data readout architecture for the ITS3 ALICE Inner Tracker System", *JINST*, vol. 19, no. 02, p. C02013, Feb. 2024. DOI: 10.1088/1748-0221/19/02/C02013. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/19/02/C02013.

[71] C. Spear and G. Tumbush, *SystemVerilog for Verification, A Guide to Learning the Testbench Language Features*. Springter, 2012. DOI: 10.1007/978-1-4614-0715-7. [Online]. Available: https://app.dimensions.ai/details/publication/pub.1025992835.

[72] E. Scapparone, "Particle identification with the ALICE detector", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 598, no. 1, pp. 152–155, 2009, Instrumentation for Collding Beam Physics. DOI: https://doi.org/10.1016/j.nima.2008.08.074. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900208012138.

[73] W. Snoeys, G. Aglieri Rinella, H. Hillemanns, *et al.*, "A process modification for CMOS monolithic active pixel sensors for enhanced depletion, timing performance and radiation tolerance", *Nucl. Instrum. Methods Phys. Res. A*, vol. 871, pp. 90–96, 2017. DOI: https://doi.org/10.1016/j.nima.2017.07.046. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016890021730791X.

[74] M. Munker, M. Benoit, D. Dannheim, *et al.*, "Simulations of CMOS pixel sensors with a small collection electrode, improved for a faster charge collection and increased radiation tolerance", *JINST*, vol. 14, no. 05, p. C05013, May 2019. DOI: 10.1088/1748-0221/14/05/C05013. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/14/05/C05013.

[75] G. Aglieri Rinella, *et al.*, "Characterization of analogue Monolithic Active Pixel Sensor test structures implemented in a 65 nm CMOS imaging process", *Nucl. Instrum. Methods Phys. Res. A*, vol. 1069, p. 169 896, 2024. DOI: https://doi.org/10.1016/j.nima.2024.169896. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900224008222.

[76] A. Villani, "Characterization results of MAPS digital prototypes for the ALICE ITS3", *Nucl. Instrum. Methods Phys. Res. A*, vol. 1071, p. 170 032, 2025. DOI: https:

//doi.org/10.1016/j.nima.2024.170032. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900224009586.

[77] CERN EP R&D, "Development of monolithic sensors for high energy physics in TPSCo 65nm ISC technology", Jun. 13, 2023. [Online]. Available: https://agenda.infn.it/event/36206/contributions/202641/attachments/106800/150589/FEE_2023_WS.pdf.

[78] K. Gautam and A. Kumar, "Characterisation of analogue MAPS fabricated in 65 nm technology for the ALICE ITS3", *Nucl. Instrum. Methods Phys. Res. A*, vol. 1068, p. 169 787, 2024. DOI: https://doi.org/10.1016/j.nima.2024.169787. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900224007137.

[79] G. Aglieri Rinella, S. Bugiel, D. Dobrijevic, C. Lemoine, W. Snoeys, and P. Vicente Leitao, "Power distribution over the wafer-scale Monolithic pixel detector MOSAIX for ALICE ITS3", *JINST*, vol. 20, no. 02, p. C02015, 2025. DOI: 10.1088/1748-0221/20/02/C02015. [Online]. Available: https://cds.cern.ch/record/2927149.

[80] A. Dorda Martin, R. Ballabriga, G. Borghello, *et al.*, "Measurements of total ionizing dose effects in TPSCo 65 nm and influence of NMOS bulk bias", *JINST*, vol. 18, no. 02, p. C02036, Feb. 2023. DOI: 10.1088/1748-0221/18/02/C02036. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/18/02/C02036.

[81] T. P. Ma and P. V. Dressendorfer, *Ionizing Radiation Effects in MOS Devices and Circuits*. Wiley, 1989.

[82] F. Larin, *Radiation effects in semiconductor devices*. Wiley, 1968.

[83] B. Schmidt, "The High-Luminosity upgrade of the LHC: Physics and Technology Challenges for the Accelerator and the Experiments", *J. Phys. Conf. Ser.*, vol. 706, no. 2, p. 022 002, Apr. 2016. DOI: 10.1088/1742-6596/706/2/022002. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/706/2/022002.

[84] G. Borghello, "Ionizing radiation effects in nanoscale CMOS technologies exposed to ultra-high doses", Presented 08 Mar 2019, Udine U., 2018. [Online]. Available: https://cds.cern.ch/record/2680840.

[85] R. C. Hughes, "Time-resolved hole transport in $a - \mathrm{SiO_2}$", *Phys. Rev. B*, vol. 15, pp. 2012–2020, 4 Feb. 1977. DOI: 10.1103/PhysRevB.15.2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.15.2012.

[86] G. Borghello, F. Faccio, G. Termo, *et al.*, "Effects of bias and temperature on the dose-rate sensitivity of 65-nm CMOS transistors", *IEEE Trans. Nucl. Sci.*, vol. 68, no. 5, pp. 573–580, May 2021. DOI: 10.1109/TNS.2021.3062622.

[87] G. Termo, G. Borghello, F. Faccio, S. Michelis, A. Koukab, and J.-M. Sallese, "Fab-to-fab and run-to-run variability in 130 nm and 65 nm CMOS technologies exposed to ultra-high TID", *JINST*, vol. 18, no. 01, p. C01061, Jan. 2023. DOI: 10.1088/1748-0221/18/01/C01061. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/18/01/C01061.

[88] P. Sallagoity, F. Gaillard, M. Rivoire, M. Paoli, M. Haond, and S. McClathie, "STI process steps for sub-quarter micron CMOS", *Microelectron. Reliab.*, vol. 38, no. 2, pp. 271–276, 1998. DOI: https://doi.org/10.1016/S0026-2714(97)00166-2. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0026271497001662.

[89] F. Faccio, S. Michelis, D. Cornale, A. Paccagnella, and S. Gerardin, "Radiation-Induced Short Channel (RISCE) and Narrow Channel (RINCE) Effects in 65 and 130 nm MOSFETs", *IEEE Trans. Nucl. Sci.*, vol. 62, no. 6, pp. 2933–2940, 2015. DOI: 10.1109/TNS.2015.2492778. [Online]. Available: https://cds.cern.ch/record/2161121.

[90] F. Faccio and G. Cervelli, "Radiation-induced edge effects in deep submicron CMOS transistors", *IEEE Trans. Nucl. Sci.*, vol. 52, no. 6, pp. 2413–2420, 2005. DOI: 10.1109/TNS.2005.860698.

[91] S. Bonaldo, G. Borghello, F. Faccio, M. Bagatin, S. Mattiazzo, and S. Gerardin, "TID effects in the lateral STI oxide of planar CMOS transistors", *IEEE Trans. Nucl. Sci.*, vol. 72, no. 8, pp. 2264–2275, 2025. DOI: 10.1109/TNS.2024.3517512.

[92] F. Faccio, G. Borghello, E. Lerario, *et al.*, "Influence of LDD Spacers and H+ Transport on the Total-Ionizing-Dose Response of 65-nm MOSFETs Irradiated to Ultrahigh Doses", *IEEE Trans. Nucl. Sci.*, vol. 65, no. 1, pp. 164–174, Oct. 2018. DOI: 10.1109/TNS.2017.2760629.

[93] S. Bonaldo, S. Gerardin, X. Jin, *et al.*, "Charge Buildup and Spatial Distribution of Interface Traps in 65-nm pMOSFETs Irradiated to Ultrahigh Doses", *IEEE Trans. Nucl. Sci.*, vol. 66, no. 7, pp. 1574–1583, Mar. 2019. DOI: 10.1109/TNS.2019.2903020.

[94] G. Borghello, E. Lerario, F. Faccio, *et al.*, "Ionizing radiation damage in 65 nm CMOS technology: Influence of geometry, bias and temperature at ultra-high doses", *Microelectron. Reliab.*, vol. 116, p. 114 016, Jan. 2021. DOI: https://doi.org/10.1016/j.microrel.2020.114016. [Online]. Available: %7Bhttps://www.sciencedirect.com/science/article/pii/S0026271420309033%7D.

[95] S. Kulis, "Digital synthesis for rad-hard components, Single Event Upsets mitigation techniques with TMRG tool", May 14, 2017. [Online]. Available: https://tmrg.web.cern.ch/tmrg/tmrg_kulis_in2p3.pdf.

[96] R. W. Hamming, "Error detecting and error correcting codes", *Bell Labs Tech. J.*, vol. 29, no. 2, pp. 147–160, 1950. DOI: 10.1002/j.1538-7305.1950.tb00463.x.

[97] I. S. Reed and G. Solomon, "Polynomial Codes Over Certain Finite Fields", *SIAM J. Appl. Math.*, vol. 8, no. 2, pp. 300–304, 1960. DOI: 10.1137/0108018. eprint: https://doi.org/10.1137/0108018. [Online]. Available: https://doi.org/10.1137/0108018.

[98] S. M. Mahmood and K. Roeed, "Investigation of Single Event Latch-up effects in the ALICE SAMPA ASIC", *PoS*, vol. TWEPP2018, p. 023, 2019. DOI: 10.22323/1.343.0023. [Online]. Available: http://cds.cern.ch/record/2710375.

[99] F. Faccio, "ASIC survival in the radiation environment of the LHC experiments: 30 years of struggle and still tantalizing", *Nucl. Instrum. Methods Phys. Res. A*, vol. 1045, p. 167 569, 2023. DOI: https://doi.org/10.1016/j.nima.2022.167569. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900222008610.

[100] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, 4th. USA: Addison-Wesley Publishing Company, 2010.

[101] P. J. McWhorter and P. S. Winokur, "Simple technique for separating the effects of interface traps and trappedoxide charge in metaloxidesemiconductor transistors", *Appl. Phys. Lett.*, vol. 48, no. 2, pp. 133–135, Jan. 1986. DOI: 10.1063/1.96974. eprint:

https://pubs.aip.org/aip/apl/article-pdf/48/2/133/18455983/133_1_online.pdf. [Online]. Available: https://doi.org/10.1063/1.96974.

[102] G. I. Zebrev and M. S. Gorbunov, "Modeling of Radiation-Induced Leakage and Low Dose-Rate Effects in Thick Edge Isolation of Modern MOSFETs", *IEEE Trans. Nucl. Sci.*, vol. 56, no. 4, pp. 2230–2236, Aug. 2009. DOI: 10.1109/TNS.2009.2016096.

[103] A. H. Johnston, R. T. Swimm, and T. F. Miyahira, "Low Dose Rate Effects in Shallow Trench Isolation Regions", *IEEE Trans. Nucl. Sci.*, vol. 57, no. 6, pp. 3279–3287, Dec. 2010. DOI: 10.1109/TNS.2010.2085452.

[104] R. L. Pease, R. D. Schrimpf, and D. M. Fleetwood, "ELDRS in Bipolar Linear Circuits: A Review", *IEEE Trans. Nucl. Sci.*, vol. 56, no. 4, pp. 1894–1908, Aug. 2009. DOI: 10.1109/TNS.2008.2011485.

[105] D. M. Fleetwood, S. L. Kosier, R. N. Nowlin, *et al.*, "Physical mechanisms contributing to enhanced bipolar gain degradation at low dose rates", *IEEE Trans. Nucl. Sci.*, vol. 41, no. 6, pp. 1871–1883, Dec. 1994. DOI: 10.1109/23.340519.

[106] R. D. Schrimpf, R. J. Graves, D. M. Schmidt, *et al.*, "Hardness-assurance issues for lateral PNP bipolar junction transistors", *IEEE Trans. Nucl. Sci.*, vol. 42, no. 6, pp. 1641–1649, Dec. 1995. DOI: 10.1109/23.488761.

[107] D. M. Fleetwood, L. C. Riewe, J. R. Schwank, S. C. Witczak, and R. D. Schrimpf, "Radiation effects at low electric fields in thermal, SIMOX, and bipolar-base oxides", *IEEE Trans. Nucl. Sci.*, vol. 43, no. 6, pp. 2537–2546, Dec. 1996. DOI: 10.1109/23.556834.

[108] S. C. Witczak, R. D. Schrimpf, D. M. Fleetwood, *et al.*, "Hardness assurance testing of bipolar junction transistors at elevated irradiation temperatures", *IEEE Trans. Nucl. Sci.*, vol. 44, no. 6, pp. 1989–2000, Dec. 1997. DOI: 10.1109/23.658978.

[109] R. J. Graves, C. R. Cirba, R. D. Schrimpf, *et al.*, "Modeling low-dose-rate effects in irradiated bipolar-base oxides", *IEEE Trans. Nucl. Sci.*, vol. 45, no. 6, pp. 2352–2360, Dec. 1998. DOI: 10.1109/23.736454.

[110] S. C. Witczak, R. C. Lacoe, D. C. Mayer, D. M. Fleetwood, R. D. Schrimpf, and K. F. Galloway, "Space charge limited degradation of bipolar oxides at low electric fields", *IEEE Trans. Nucl. Sci.*, vol. 45, no. 6, pp. 2339–2351, Dec. 1998. DOI: 10.1109/23.736453.

[111] J. Boch, F. Saigne, R. D. Schrimpf, D. M. Fleetwood, R. Cizmarik, and D. Zander, "Elevated temperature irradiation at high dose rate of commercial linear bipolar ICs", *IEEE Trans. Nucl. Sci.*, vol. 51, no. 5, pp. 2903–2907, Oct. 2004. DOI: 10.1109/TNS.2004.835055.

[112] J. Boch, F. Saigne, R. D. Schrimpf, J.-R. Vaille, L. Dusseau, and E. Lorfevre, "Physical Model for the Low-Dose-Rate Effect in Bipolar Devices", *IEEE Trans. Nucl. Sci.*, vol. 53, no. 6, pp. 3655–3660, Dec. 2006. DOI: 10.1109/TNS.2006.886008.

[113] N. L. Rowsey, M. E. Law, R. D. Schrimpf, D. M. Fleetwood, B. R. Tuttle, and S. T. Pantelides, "Mechanisms Separating Time-Dependent and True Dose-Rate Effects in Irradiated Bipolar Oxides", *IEEE Trans. Nucl. Sci.*, vol. 59, no. 6, pp. 3069–3076, Dec. 2012. DOI: 10.1109/TNS.2012.2222669.

[114] S. C. Witczak, R. C. Lacoe, J. V. Osborn, J. M. Hutson, and S. C. Moss, "Dose-rate sensitivity of modern nMOSFETs", *IEEE Trans. Nucl. Sci.*, vol. 52, no. 6, pp. 2602–2608, Feb. 2005. DOI: 10.1109/TNS.2005.860709.

[115] G. Borghello, G. Termo, F. Faccio, *et al.*, “ELDRS in a commercial 28nm CMOS technology”, *IEEE Trans. Nucl. Sci.*, pp. 1–1, 2025. DOI: 10.1109/TNS.2025.3543658.