

ÉCOLE DOCTORALE DE PHYSIQUE ET CHIMIE-PHYSIQUE

INSTITUT PLURIDISCIPLINAIRE HUBERT CURIEN, UMR 7178

THÈSE présentée par :

Jean SOUDIER

soutenue le : 12 décembre 2024

pour obtenir le grade de : Docteur de l'université de Strasbourg

Discipline/ Spécialité : microélectronique

Etude d'architectures de lecture asynchrone intégrée
pour capteurs à pixels CMOS

THÈSE dirigée par :

M. BAUDOT Jérôme

M. UHRING Wilfried

Professeur, université de Strasbourg

Professeur, université de Strasbourg

RAPPORTEURS :

Mme VILELLA-FIGUERAS Eva

M. FESQUET Laurent

UKRI Research fellow, University of Liverpool

Maître de conférence, Université Grenoble Alpes

AUTRES MEMBRES DU JURY :

M. SICARD Gilles

M. LLOPART CUDIE Xavier

M. MOREL Frédéric

M. BITAR EL Ziad

Ingénieur chercheur, CEA-LETI

Ingénieur chercheur, CERN

Ingénieur de recherche, CNRS/IPHC

Directeur de recherche, CNRS/IPHC

“In the world of microelectronics, it’s not about the quantity of information you possess but the precision with which you wield it in the creation of innovative technology.”

Acknowledgements

I would like to thank first **Jérôme BAUDOT** my director's thesis to supporting me through all this work and trials over this 3 years. I would thank him for his trust in me, it was great.

I would like to thank also **Fédéric MOREL** my advisor during this thesis for all his good advice, support and his time over all these 3 years.

A big thank to **Wilfried UHRING** to proposed me this thesis at first and provides materials to work with a lot of support.

I will thank in advance the **entire jury of this thesis** to accept reviewing my thesis work and giving the best advice they could.

I thank all the **C4PI** team for the welcoming, help and good mood during all my thesis. I thank also the **ICube** and **IRFU** teams which whom I collaborate a lot and get a lot of advice.

I would like to thank the **CERN ALICE ITS3** team specifically **Walter SNOEYS** and **Gianluca AGLIERI RINELLA** for their trust in me and for letting me propose a chiplet in the submission.

I thank the **Test** team at IPHC where we exchange to improve the testability of my chip.

I thank the **Physicist** team which gave me a lot of inputs and advice to do relevant models to test my chip.

I would like to thank all **PhD students** and **apprentice** of the IPHC team with whom I exchanged various subjects and advice.

I thank **Sandrine COURTIN**, the director of the IPHC laboratory for letting me do this thesis.

I would like to thank the **CSDDD** committee for supporting me and guiding me to achieve the thesis.

I thank the entire **IPHC** laboratory for the welcome.

I would like to thank the **doctoral School 182 Physic and Physic-Chemistry** and the **doctoral School 269 Mathematics, Science of Information and Engineering** for making this thesis possible.

I thank the **QMat** committee for supporting me and helping me share my work over various conferences.

I thank the **MI2I** group for supporting me and letting me present my work.

Finally, I would like to thank my **Wife**, my **Family** and **every person** I met in the past 3 years who helped me grow my knowledge and perfect my work.

This thesis was of considerable interest to me, yielding a wealth of new results and avenues for future development. A significant amount of work has been completed, but the final circuit must still be tested before the project can be considered complete.

Contents

Acknowledgements	iii
Introduction	1
1 State of the art for particle physics tracking sensors based on pixels	3
1.1 Brief presentation of open questions in particle physics	4
1.1.1 Particle physics context	4
1.1.2 Tracker sensors	5
1.2 Generic requirements for particle tracking sensors	8
1.2.1 Spatial Resolution	8
1.2.2 Speed & Time Resolution	9
1.2.3 Power density	9
1.2.4 Material Budget	10
1.2.5 Area & Dead zone	11
1.2.6 Radiation hardness	11
1.2.7 Trade-offs	11
1.3 Different pixels sensors	12
1.3.1 Hybrid Pixel Detectors	12
1.3.2 DEPFET	13
1.3.3 SOI	14
1.3.4 MAPS	15
1.3.5 Summary	16
1.4 Specific tracker requirements for some collider experiments	17
1.4.1 ALICE	17
1.4.2 CMS	18
1.4.3 ATLAS	19
1.4.4 LHCb	20
1.4.5 CBM	20
1.4.6 Belle II	21
1.4.7 FCCee	22
1.4.8 Other experiments: Quantum physics	23
1.4.9 Summary	23
1.5 Conclusion	25
2 State of the art for readout in particle physics pixel tracker sensors	29
2.1 Readout by pixels or columns	31
2.1.1 Analog readout	31
2.1.2 Rolling shutter	31
2.2 Zero suppression readout	34
2.2.1 Priority encoder	34
2.2.2 With token ring	35
2.2.3 Pulse width encoding	36
2.2.4 Fixed/Dynamic Priority Arbiter (Asynchronous)	37

2.3	Summary	40
3	Asynchronous architecture design and knowledge	45
3.1	State of the art in asynchronous circuits	46
3.1.1	Types of asynchronous circuits	46
3.1.2	Two or four phases protocol	46
3.1.3	Types of synchronization	47
3.1.4	Know how in asynchronous CMOS design	48
3.2	Asynchronous digital Flow	50
3.2.1	Classical flow steps	50
3.2.2	Timings inside the flow	51
3.2.3	Flow modifications	53
3.2.4	Timing constraints	55
3.3	Proposal of a structure for the readout architecture	57
3.3.1	Adaptation to the particles physics requirements	57
3.3.2	Working principle	57
	Decision part	58
	Data management and synchronization part	59
	Global FPA circuit	59
3.3.3	Timings for the proposal	60
3.4	Conclusion on asynchronous readout	62
4	Implementation and performance of the proposed asynchronous readout logic	65
4.1	Simulation environment	66
4.1.1	The TOWER 65nm technology	66
4.1.2	The Cadence tools	66
4.1.3	The particle physics simulation	67
4.2	Early comparison	69
4.2.1	Area usage	69
4.2.2	Power and timing results	70
4.2.3	Summary	72
4.3	Exploring the limits of the architecture	73
4.3.1	Single column block over different sizes	73
4.3.2	Simulation for very small pitch	79
4.3.3	Radiation hardness tolerance	79
4.3.4	Simulation with stitched sensors	81
4.4	Summary	82
5	Prototype Sensor Pixel Asynchronous Readout CMOS (SPARC)	85
5.1	Design of SPARC	86
5.1.1	Goals and context	86
5.1.2	The pixels design	87
5.1.3	Asynchronous gate	88
5.1.4	The readout of the matrix	90
5.1.5	Memory for the datas	91
5.1.6	Timestamping	91
5.1.7	Design of SPARC	92
5.1.8	Expectations	94
5.2	Testing strategy	99
5.2.1	Electrical tests	99

5.2.2	Laboratory tests	99
5.2.3	Under beam tests	100
5.3	Conclusion	102
6	Conclusion	107
6.1	Work already done	108
6.1.1	Asynchronous flow	108
6.1.2	Performance of the proposed asynchronous readout architecture	108
6.1.3	Demonstrator prototype design	109
6.2	What's next ?	110
6.2.1	Further architecture optimization	110
6.2.2	Foreseen applications of the asynchronous architecture	112
A	Müller gate	115
B	Constraints modes	117
C	Arbiter priority function	119
D	Slow control registers	121
E	PAD list	123
F	Complet view of the SPARC design	125
	Bibliography	127

List of Figures

1.1	Schematic views of a tracker arrangement with multi detection layers taken from [2].	6
1.2	Representation of the circuitry inside a MAPS diode detector.	7
1.3	Cross section of the material budget of the ALICE ITS2 tracker [3]. . . .	10
1.4	Cross section view of the basic element of an hybrid pixels detector [5].	12
1.5	Cross section view of the pixel in a DEPFET detector [7].	13
1.6	Cross section view of the pixel in a SOI detector [10].	14
1.7	Cross section view of pixel in a MAPS detector [15].	15
1.8	Schematic of the ALICE experiment [19].	18
1.9	Photography of the ALICE experiment.	18
2.1	Classical topology of a matrix for monolithic pixel sensors.	30
2.2	A representation of the rolling shutter readout [32].	32
2.3	Schematic of the Priority Encoder.	34
2.4	Chronogram of the Priority Encoder.	35
2.5	View of the token ring readout.	36
2.6	Schematic view of the pulse width readout [44].	37
2.7	Topology of a tree made with 2 to 1 arbiters	38
2.8	Topology of a tree made with one 512 to 1 arbiter	38
2.9	View of the priority arbiter topology and the chronogram in a collision (double request) [46].	39
2.10	Comparison of the different types of readout on their possible range. .	41
3.1	The two- and four-phases routine [51].	47
3.2	The Muller gate diagrams with reset [46].	48
3.3	View of the whole classic flow.	50
3.4	Schematic time flow comparison of synchronous and asynchronous logics.	52
3.5	View of the whole asynchronous flow [50].	54
3.6	Arbiter pixel selection schematic.	58
3.7	Multiplexer selection schematic.	59
3.8	Structure of the fixed priority arbiter.	60
3.9	Functional structure of a tree composed of fixed priority arbiters (FPA). .	60
4.1	Algorithmic flow to generate particle hits.	68
4.2	Evolution of the area usage with the controller size, expressed in power of 2 (3 indicates a $2^3 = 8$ to 1 controller).	69
4.3	Evolution of the time to read all pixel with the controller size, expressed in power of 2 (3 indicates a $2^3 = 8$ to 1 controller).	71
4.4	Evolution of the number of pixels read in 100 ns with the controller size, expressed in power of 2 (3 indicates a $2^3 = 8$ to 1 controller). . . .	71

4.5	Evolution of (a) the power and (b) the pixel mean reading time for a hit rate of 300 MHz/cm ² rates, with the controller size, expressed in power of 2 (3 indicates a $2^3 = 8$ to 1 controller).	72
4.6	Evolution of (a) the power and (b) the pixel mean reading time for a hit rate of 3 GHz/cm ² rates, with the controller size, expressed in power of 2 (3 indicates a $2^3 = 8$ to 1 controller).	72
4.7	Power density versus the hit rate for all hit rates in the continuous case (solid lines) and clocked case (dashed lines).	75
4.8	Energy without leakage per hit versus the hit rate in the continuous case (solid lines) and the clocked case (dashed lines).	75
4.9	Violin plot comparing the distribution in time of the arrivals of the fired pixel addresses for three different combinations of pitch and controller size: 24 μ m and 2 $\rightarrow$ 1 (blue), 24 μ m and 512 $\rightarrow$ 1 (orange), 30 μ m and 512 $\rightarrow$ 1 (violet).	76
4.10	Violin plot comparing the cumulated distribution in time of the arrivals of the fired pixel addresses for three different combinations of pitch and controller size: 24 μ m and 2 $\rightarrow$ 1 (blue), 24 μ m and 512 $\rightarrow$ 1 (orange), 30 μ m and 512 $\rightarrow$ 1 (violet).	77
5.1	Diagram of the double column readout arrangement for the SPARC circuit.	85
5.2	Schematic of pixel wiring to the double column with the digital pixel block.	88
5.3	Schematic of the Muller gate D1 at the transistor level.	89
5.4	Distribution of the various controller sizes implemented in the pixel matrix readout of the SPARC circuit.	90
5.5	TOP view of the functional blocs of the SPARC sensor.	94
5.6	Mask of SPARC.	94
5.7	Comparison of the fired pixel address time distributions for the different controller sizes implements in the SPARC sensor.	96
5.8	Comparison of the cumulated distribution of the arrival times for the fired pixel addresses, for the different controller sizes implements in the SPARC sensor.	97
6.1	Possible architecture for a very small pitch of 10 μ m.	111
F.1	TOP view of SPARC.	125

List of Tables

1.1	Summary of the requirements on tracking sensors for some selected experiments.	24
2.1	Summary of MAPS sensors and some readout ASIC (* use zero suppression) and their readout matrix architecture: PE=Priority encoder, PW=Pulse width, TR=Token ring, RS=Rolling shutter	40
4.1	Percentage of GCells used.	70
4.2	Results over different sizes for density of cells, total metals and vertical metals usage	74
4.3	Timing results from simulations for different hit rates and column configurations in the continuous case: mean time to read a pixel/time to read 99.9% of the pixels. (TW i.e. Time-Walk)	78
4.4	Timing results from simulations for different hit rates and column configurations in the clocked case: mean time to read a pixel / time to read 99.9% of the pixels.	78
4.5	Maximal hit rates allowing a maximal hit loss of 1/1000th.	78
5.1	Attribution of the 24 bits word in the FIFO.	91
5.2	VCO table	91
5.3	Results in terms of area, power and timing for the 5 different architecture implemented.	95
5.4	Comparison of the timing over address distribution.	96
A.1	Timing arc for the Müller gate	115
A.2	Leakage arc for the Müller gate	115
B.1	Logic table of the priority function	117
C.1	Differents modes summary	120
D.1	Control bits	121
E.1	Control bits	123

List of Abbreviations

IPHC	I nstitut P luridisciplinaire H ubert C urien
C4PI	C entre de C ompétences de C apteurs CMOS à P ixels I ntégrés
ICube	laboratoire des sciences de l'Ingénieur, de l'Informatique et de l'Imagerie
SMH	S ystèmes et M icrosystèmes H étérogènes
CSDDD	C omité de S uivi D es D octorants du D RS
QMat	Q uantum M aterials
IN2P3	I nstitut N ational de P hysique nucléaire et de P hysique des P articules
MI2I	M icro électronique des 2 I nfinis
IRFU	I nstitut de R echerche sur les lois F ondamentales de l' U nivers
CERN	C onseil E uropéen de R echerche N ucléaire
ALICE	A L arge I on C ollider E xperiment
CMS	C ompact M uon S olenoid
LHC	L arge H adron C ollider
CBM	C ompressed B aryonic M atter
ATLAS	A T oroidal L HC A pparatu S
FCC	F uture C ircular C ollider
HEP	H igh E nergy P hysic
ITS	I nnner T racking S ystem
OTS	O uter T racking S ystem
HGTD	H igh G ranularity T iming D etector
ITk	I nnner T racker
UT	U pstream T racker
MT	M ighty T racker
MVD	M icro V ertex D etector
VXD	V erte X D etector
VTX	V er T e X detector
MIMOSA	M inimum I onizing P article MOS A ctive pixel sensor
MIMOSIS	MIMOSA for FAIR-SIS experiment
ALPIDE	ALICE P ixel D etector
FEI	P ixel F ront- E nd C hip
MOSS	M onolithic S titched S ensor
MOST	M onolithic S titched T iming
MOSAIX	M onolithic S titched A ctive P ixel
MATLA	M onolithic from ALICE to ATLAS
SPARC	S ensor P ixel A synchronous R eadout CMOS
ASIC	A pplication- S pecific I ntegrated C ircuit
MOS	M etal- O xide- S emiconductor
CMOS	C omplementary M etal- O xide- S emiconductor
VLSI	V ery L arge S cale I ntegration

CCD	Charge Coupled Device
RTL	Register Transfer Level
LGAD	Low Gain Avalanche Diode
SEE	Single Event Effect
SEU	Single Event Upset
XSEU	Cross Section SEU
TOT	Time Over Threshold
TOA	Time Of Arrival
PE	Priority Encoder
FPA	Fixed Priority Arbiter
DPA	Dynamic Priority Arbiter
QDI	Quasi Delay Insensitive
BD	Bundle Data
LCS	Local Clock Set
PVT	Process Voltage Temperature
CTS	Clock Tree Synthesis
AOCV	Advanced On Chip Variations
MMMC	Multi Modes Multi Corners
DMMMC	Distributed Multi Modes Multi Corners
FIFO	First In First Out
VCO	Voltage Controlled Oscillator
TDC	Time to Digital Converter
SPI	Serial Peripheral Interface
ADC	Analog to Digital Converter
FPGA	Field Programmable Gate Array
SOI	Silicon On Insulator
SiPM	Silicon PhotoMultiplier
MAPS	Monolithic Active Pixel Sensors
DMAPS	Depleted Monolithic Active Pixel Sensors

Introduction

The scientific field of High Energy Physics (HEP) is currently planning future experiences. As expected for instruments targeting to advance scientific knowledge, they are required to deliver higher performances than their predecessors. This thesis focuses on the sensors equipping sub-detectors devoted to reconstruct the trajectories of charge particles produced in high energy physics experiments, namely the vertex detectors and trackers. More specifically, monolithic active pixel sensors, or MAPS fabricated in the technology of integrated circuits, are considered. They indeed offer native features, such as small pixels and light materials, which are prevalent advantages for tracking particles.

The technology of MAPS applied to HEP have progressed considerably since it was proposed more than twenty years ago. Nevertheless new experiments still require improvements in terms of position and time resolutions, hit rates, power dissipated and in some cases radiation tolerance.

To match such requirements, R&D on various parts of monolithic sensors should be undertaken. The collection system and front-end analogue as well as digital treatments are primarily concerned since they drive most key performances from the detection efficiency to the resolutions.

Nevertheless, one should not forget that pixel sensors often feature a huge number of channels, typically a few 100,000 channels per cm^2 , which information should be driven away of the sensor for further analysis. This is a task for the matrix readout. In order to preserve the potential high performance of the collection and treatment systems and handle large hit rates, the readout should operate swiftly but with adding a minimum of resources. The architecture should consume neither much area to limit the pixel size and the spatial resolution, neither much power.

A key point for achieving all the goals of the readout is to extract from the matrix only the information of the pixels fired by particles traversing the sensor. This is the sparse readout technique, which performance depends strongly on its technical implementation.

Due to the random nature of the distribution of the particle hits over the surface sensor, synchronous logic is not the most efficient way to implement a sparse readout. On the contrary, asynchronous logic seems to be very well adapted to the problematic of reading randomly spread hits.

However synchronous logic is the most widely spread design methodology for integrated circuit design in HEP and most current MAPS relies entirely or partially on it. The goal of this thesis is to explore the performance and feasibility of a fully asynchronous readout architecture for MAPS in HEP.

The work has been conducted in the C4Pi (Centre de compétences pour les capteurs CMOS) facility of IPHC (Institut Pluridisciplinaire Hubert Curien), which is a pioneer and leading group for the development of MAPS. The idea for the asynchronous architecture proposed in this thesis come however from another scientific

domain, photonics, and especially from a sensor development team located at the ICube laboratory, neighboring IPHC.

Two main objectives drive the organization of this work: first, to ascertain the viability of an asynchronous architecture in HEP, and second, to delineate the means of creating such an architecture through an automated design flow.

The initial chapter, designated as Chapter 1, establishes the context of the HEP trackers and how the requirements impact the ambitions of this thesis. The chapter starts with an explication of the HEP, a delineation of the tracker's properties and their requisite specifications. This chapter presents a state-of-the-art overview of the various possibilities for CMOS sensing technologies. In conclusion, the experiments are presented, along with an analysis of the variations in requirements between the different scenarios.

The second chapter, designated as Chapter 2, presents the requirements and actual possibilities of the readout architecture employed in the HEP. It elucidates the lengthy process of transitioning from analog to sparse readout behaviors. All of the aforementioned architectural approaches ultimately rely on the implementation of zero suppression algorithms, with some even employing data-driven techniques.

The chapter, referenced as Chapter 3, introduced the concept of asynchronous processing, which enables a data-driven approach to reading out the data. The fundamental principles of asynchronous design are elucidated, along with the methodologies employed in its construction. The presentation of an automated design flow for use in HEP is explained in detail. The text discusses the difficulties and solutions to developing circuits based on asynchronous logic, focusing on the elementary blocks needed to realize a readout architecture.

The fourth chapter, designated as Chapter 4, presents the findings of the simulation of the layout, with the objective of deepening the comprehension of the flow and the proposed architectural design. Subsequently, the chapter presents the implementation of the proposed asynchronous architecture in a specific CMOS process and explores its performance with detailed simulations. The performance of the system is evaluated in relation to the various parameters of the implementation, taking into account the requirements outlined in the initial chapter 1.

The fifth chapter, referenced as Chapter 5, provides an overview of the SPARC sensor prototype design. The SPARC circuit features small pixel matrices whose readouts adhere to the asynchronous architectural design, thereby facilitating experimental validation. The following sections delineate the architectural and experimental choices that will be employed to characterize the chip.

The concluding chapter 6 provides a summary of the work presented in this thesis and reexamines some of its achievements. It also discusses prospects to apply the proposed asynchronous readout architecture for future instruments in the field of particle physics.

Chapter 1

State of the art for particle physics tracking sensors based on pixels

This first chapter intends to introduce the interest of pixel sensors in particle physics and presents their requirements for a number of future projects. The initial section provides a very short introduction to the goals of particle physics and its anticipated future experimental program. Subsequently, we discuss briefly the generic aspects of the large instruments used in such programs before focusing on the central topic of this thesis, i.e. the sub-detectors devoted to tracking charged particles or trackers. In particular the main parameters driving their performance are listed. The third section of the chapter presents an overview of the various types of sensor technologies utilized in such trackers. The final section of the chapter is devoted to a comparison of the specifications of a number of future trackers.

1.1 Brief presentation of open questions in particle physics

This section presents an overview of the historical branch of science known as particle physics, outlining its fundamental objectives and some of its perspectives. A subsequent subsection delves into the role of sub-detectors named trackers for this field.

1.1.1 Particle physics context

Particle physics is the branch of physics that studies the fundamental or elementary particles of the universe and the forces that govern their interactions. Its roots trace back to the early 20th century when scientists first explored the structure of the atom and other phenomena to discover the proton (1918), the photon (1923), the neutron (1932), the muon (1937) and the pion (1946).

During the mid-20th century, scientists added a range of new particles using continuously more powerful colliders, leading to the need for a classification system. Shortening a long and rich story in a few words, this gave rise to the Standard Model in the 1970s, which organizes particles into quarks, leptons, and gauge bosons, as force carriers for each interaction at the elementary levels: the photon, the W and Z bosons, and the gluons.

One of the most significant results in the recent years was the discovery of the Higgs boson in 2012 [1] at the Large Hadron Collider (LHC) located at CERN. The Brout-Englert-Higgs mechanism is vital for explaining how elementary particles acquire mass and the discovery of the eponymous boson kind of complete the experimental confirmation of the Standard Model.

However, the Standard Model of particle physics, while highly successful, has notable limitations.

First, it can not be considered as a "theory of everything" since it does not include gravity. While general relativity describes gravity on a macroscopic scale, it is hardly compatible in the current theoretical framework describing relativistic quantum mechanics, meaning that physicists cannot describe gravitational interactions at the quantum level, such as those that occur in black holes or during the early moments of the universe's existence. A quantum theory of gravity, which would unite these two frameworks, remains elusive and poses a significant theoretical challenge.

Another major gap in the Standard Model is its inability to explain dark matter and dark energy, which make up about 95% of the universe's total mass-energy content. Dark matter interacts gravitationally but does not interact via the electromagnetic, weak or strong forces described by the Standard Model. Similarly, dark energy, which drives the accelerated expansion of the universe, is not accounted for by any known particles or interactions in the current theory.

The universe's matter-antimatter asymmetry presents another unsolved problem. The Big Bang should have produced equal amounts of matter and antimatter, but today the universe is overwhelmingly made of matter. While the Standard Model includes some mechanisms, like CP violation, that cause matter and antimatter to behave slightly differently, these effects are too weak to explain the large imbalance observed. Solving this puzzle requires deeper insight into why these symmetry violations occur and how they relate to the universe's early conditions.

In view of these limitations, a quite unsettling situation in nowadays particle physics, is that there are no experimental observations at the elementary level that would clearly put the Standard Model at fault, in terms of particle content or theoretical structures (conserved symmetries for instance).

Researchers are searching for new particles or interactions that are beyond the reach of the Standard Model, so called New Physics, and that might give hints to answer the open questions listed above. So far, with no success.

The search for New Physics often, but not only, requires particle accelerators operating at extremely high energies and/or luminosity.

The Large Hadron Collider (LHC) at CERN is the most powerful accelerator currently in operation and sets the energy frontier at 13.6 TeV. The LHC may not be energetic enough to reveal phenomena like supersymmetric particles or the extra dimensions predicted by string theory. Exploring further the energy frontier entails new hadron colliders, such as the proposed Future Circular Collider (FCC).

The SuperKEKB, operating e^+e^- collisions at KEK in Japan, sets the current instantaneous luminosity frontier at $4.7 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ and expect to overcome $10^{35} \text{ cm}^{-2}\text{s}^{-1}$ in the coming years.

A future high-energy e^+e^- collider is also planned with various machine types, linear such as ILC in Japan or circular such as FCCee in Europe or CEPC in China.

These new machines will host new large experiments. Their task is to provide data that allow reconstructing as robustly as possible the final quantum state produced by the collisions. That entails individualizing the particles emitted and measuring their properties. Large detectors recording and observing such collisions are hence equipped with various sub-detectors each devoted to a specific task.

Tracking systems include vertex detectors and trackers. They are meant to reconstruct the trajectories of charged particles, estimate their point of origin and measure their momentum. Calorimeters are tasked with the measurement of the energy of the emitted particles. Other systems complement these two broad categories to identify the particles. Triggering the data recording when an interesting collision has happened, usually exploits the fastest sub-detectors already mentioned and specific fast detectors.

The next section introduces in some more details the world of trackers, since this thesis focuses on the sensors they are build of. The requirements set by future projects on these sensors will be discussed at the end of this first chapter.

1.1.2 Tracker sensors

A particle tracker detector is a device used in particle physics experiments to trace the paths of charged particles as they move through a detector, see figure 1.1. It works by measuring the precise position of a particle at different points along its trajectory. Typically, these detectors are made up of layers of sensors arranged in concentric rings around the particle collision point.

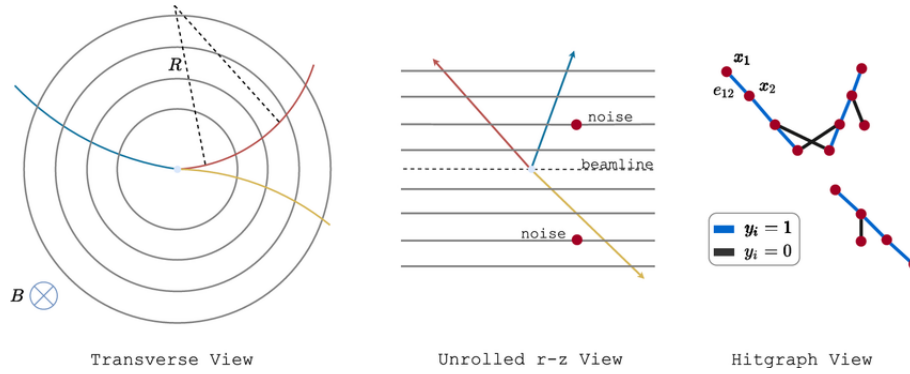


FIGURE 1.1: Schematic views of a tracker arrangement with multi detection layers taken from [2].

When a charged particle, such as a proton or electron, passes through the detector, it ionizes the material in the sensor layers, leaving a trail of signals. These signals are collected by the sensors, which may be made of materials like silicon. By analyzing the positions of the hits across different layers, the detector can reconstruct the particle's path. This information allows scientists to determine the particle's momentum, velocity and charge, often in the presence of a magnetic field, which curves the particle's path based on its momentum and charge.

Particle tracker detectors are highly precise and are essential for identifying particles produced in high-energy collisions, such as those at the Large Hadron Collider (LHC). They provide high-resolution data, crucial for distinguishing between different particle types and studying rare events. However, the performance of these detectors can be limited by factors such as radiation damage, signal to noise, and the need for fast, efficient data processing in high-energy environments where billions of particles are tracked per second.

The IPHC laboratory's principal focus was the advancement of tracking technology. The objective is to identify novel architectural solutions that can achieve optimal performance and facilitate the advancement of physics. The primary objective is to ascertain the spatial resolution, which is essential for determining the origin of the collision. The IPHC laboratory primarily utilizes MAPS sensors, which are regarded as the optimal choice in terms of detector thickness. The MAPS is a fully integrated system comprising both the electronic components and the sensing apparatus. The sensing component is comprised of a doped P (positively) substrate and a well doped N (negatively). The imager process enables the introduction of a deep doping implant surrounding the diode, thereby providing protection for the electronic components as a shield. Indeed, when a particle traverses the substrate, the deposited energy is approximately one hundred electrons, with no loss of energy allowed. A technique exists whereby the substrate can be depleted entirely, thus allowing the electrons to be captured rapidly and particles to be detected. The subsequent step is to amplify and digitize the low signal, see figure 1.2. Designing a low-noise amplifier and establishing a threshold for differentiating between noise and particle detection represents a significant challenge. The following sections will present a detailed examination of these aspects.

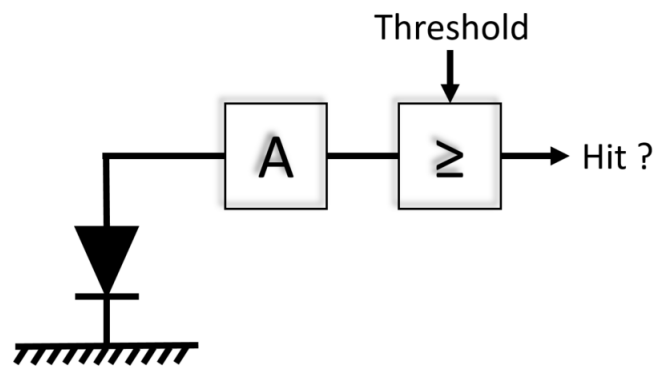


FIGURE 1.2: Representation of the circuitry inside a MAPS diode detector.

1.2 Generic requirements for particle tracking sensors

This section will resume the various requirements for a particle sensor. The detector that is the subject of this thesis are trackers, and in particular, semiconductor trackers. Other types of detectors exist, such as those used to detect gases, but they lack the necessary spatial resolution and material budgets to be effective.

The nature of the phenomenon being observed and the environment in which it occurs impose significant constraints on the use of charged particle sensors. Indeed, particles are minute, rapid, and capable of producing considerable radiation doses. This section will provide an explanation of the different requirements, their importance, and their impact. This chapter will focus on MAPS (Monolithic Active Pixel Sensor) for tracking purposes. Each section provides a detailed account of a specific requirement, accompanied by a reference to the other relevant sections. This is because each requirement represents a trade-off with other requirements. The presented requirements are those of spatial resolution, time resolution, power density, material budget, pixel pitch, and radiation hardness.

1.2.1 Spatial Resolution

The spatial resolution is of particular interest, as it represents the primary measurement employed in particle physics. Each sensor is arranged in layers, with the objective of aligning the particle trajectory. In order to ensure the optimal measurement, it is necessary for each layer to determine the point of intersection between the particle and the sensor.

In the initial approximation, the spatial resolution of a square pixel is given by $S_r = \frac{P}{\sqrt{12}}$, where P is the pitch of the pixels. The pitch is defined as the distance between the collection diode of 2 pixels. The actual spatial resolution is approximately $5\text{ }\mu\text{m}$, with an objective of reducing this to $3\text{ }\mu\text{m}$. This would result in a pitch of $18\text{ }\mu\text{m}$ to $10\text{ }\mu\text{m}$, respectively. A reduction in pixel size will also result in a reduction in the area available for the electronics and the readout circuit. There are several methods for placing the readout outside the matrix, but these approaches result in significantly slower data acquisition and wider dead zones.

In particle physics, a single particle can fire multiple pixels, which are collectively referred to as a cluster. Such clusters may be extensive, thereby creating the potential for the storage of a greater quantity of information. One potential optimization is to reduce the size of the cluster by optimizing the sensing process of the pixel. This can be achieved by firing the minimum number of pixels necessary. As multiple pixels are fired, the energy or a method for identifying the center of the cluster can assist in reducing the spatial resolution without altering the size of the pitch.

Both methods are constrained in terms of space for the design, in one case by the density and in the other by the necessity of adding a huge block creating a dead zone. Additionally, both methods necessitate the transmission of more data, as they encompass a greater number of pixels or power-related information. This resulted in a compromise between speed and spatial resolution. This is the reason why there is a growing interest in the development of new fast-reading architectures.

1.2.2 Speed & Time Resolution

The speed of the circuit is defined as the ability to read every pixel in a given time. In the aforementioned experiment, the speed was quantified between two triggers, resulting in a frame. The capacity of a sensor to read all pixels in a given frame is highly valued, as it facilitates the reconstruction of the track and avoids the need to process data across multiple frames.

A significant number of circuits are required to read pixels, and the majority of contemporary architectures are based on synchronous reading. Consequently, the rate of reading can be increased in proportion to the speed of the clock. The design constraint is the speed of the clock, which is necessary for computations to be performed between two edges of the clock. Furthermore, the power consumption will increase significantly as the clock speed is increased. In fact, CMOS technologies are based on capacitors, and, to a first approximation, doubling the speed is equivalent to the square of the power.

In contrast, other circuits employ asynchronous reading, which is activated only when it is beneficial and conserves power when it is not. The asynchronous circuit design must be adapted to accommodate the additional functions, but this necessitates greater care during the design process. This type of design is promising and represents the central thesis of this paper.

The time resolution is a method for annotating pixels in time. This can be accomplished at the pixel level, which necessitates the utilization of considerable space and data transfer resources. Alternatively, it can be conducted at the end of the column, which may result in a less precise outcome. In this manner, the design is not constrained by the necessity of frames, as each impact is afforded an opportunity for reconstruction at a designated time. This method is most commonly employed in untriggered experiments, as the space required for the time count is considerable. The speed limits are determined by the consumption and cannot be increased to the extent required.

1.2.3 Power density

The power density is calculated based on the number of pixels, ensuring accurate comparisons between the various pixel pitches. The power can be increased to the extent that the cooling system is capable of maintaining the necessary temperature.

Some systems, such as the ATLAS experiment, employ liquid helium as a cooling agent, with the temperature of the circuit being reduced to -200°C . In certain experiments, it is necessary to achieve a high degree of spatial resolution in order to accurately measure the trajectory of a particle, which is analogous to an image of the particle itself. Some particles have a lifetime of a few centimeters and cannot be measured by tracking systems. However, they create other particles at the end of their lifetime, which can be measured. It is therefore essential to measure the resulting particle with high precision and to reconstruct the track of the interest particle. The introduction of high-precision measurements inevitably leads to a reduction in noise. Noise can manifest in the sensor itself or as a deviation of the track from a collision in the sensor itself during measurement. To mitigate this phenomenon, it is

crucial to consider the circuit's thickness and, in some instances, the removal of the cooling system may be necessary.

The necessity of creating smaller pitch increase the density of pixels and transistors so the power too. Some experiments, like ALICE, do not feature any power cooling system and must ensure a low power density to keep the circuit cool. This requirement comes from the fact that the first layer is so close to the pipe that the cooling system can be placed and the material budget is too critical.

1.2.4 Material Budget

The material budget is a numerical value calculated based on the thickness of the circuit and its composition. For instance, a substantial atomic nucleus will be more susceptible to colliding with the particle. Consequently, materials are assigned coefficients based on their atomic numbers. The material budget is the product of the thickness and the coefficient and must not exceed a threshold imposed by the physicist.

Some proposed solutions aim to constrain the material budget by employing a stitching technique that will enable the fabrication of a wide sensor with minimal attachment, as well as a bending technique that will facilitate the reduction of the surrounding structure of the pipe. View of a cross-section for the ALICE inner tracker version 2, see figure 1.3.

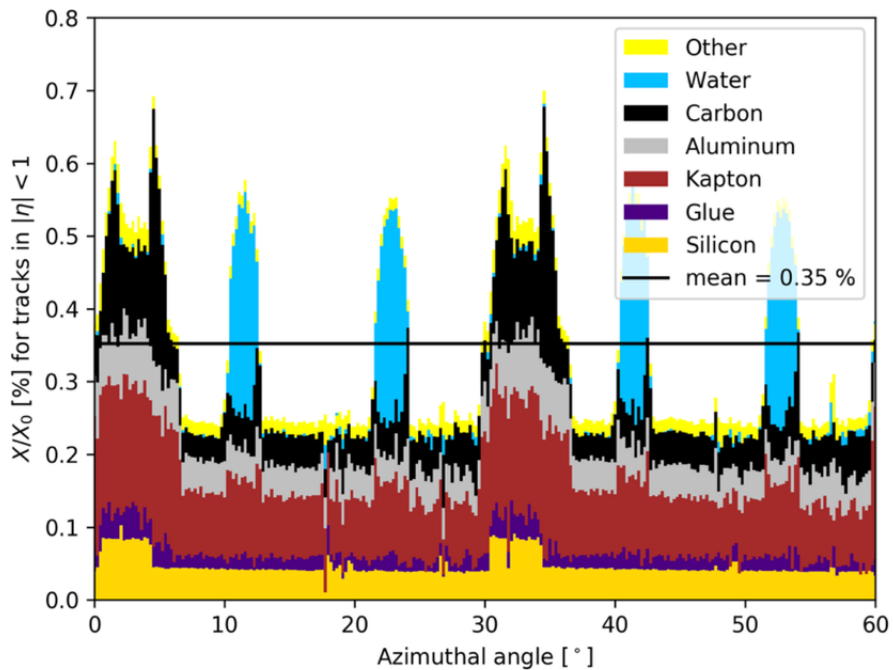


FIGURE 1.3: Cross section of the material budget of the ALICE ITS2 tracker [3].

1.2.5 Area & Dead zone

The pixel area determines the potential techniques that can be employed. To illustrate, a small pitch precludes the possibility of measuring time or power within the pixel. For pixels with a particularly small area, even the readout circuit is constrained by the need to process a large matrix.

In order to compensate for the presence of dense pixels, matrix designers may create a dead zone in which common features, such as powering or transmission, may be placed. It is necessary to contain these defective zones, which may exceed a few percent of the sensing portion.

1.2.6 Radiation hardness

The final limitation is the radiation hardness. The phenomenon of particle collision can result in the deposition of radiation on the circuit, which can subsequently lead to damage. There are two forms of radiation: ionizing and non-ionizing. The two forms of radiation have distinct effects on the circuitry and may be present simultaneously in a single experiment.

Two distinct side effects of radiation are bit flip, which alters memory values, and doping changes. The bit flip can be rectified by implementing a design with a reduced number of memories or by tripling the memory capacity. This ultimately resulted in an expansion of the circuit area and power consumption. The other effect, namely the doping changes, will affect the characteristics of the sensing part of the chip. In fact, the device will become less sensitive to particles, but this can be rectified by increasing the tuned powering. In terms of the numeric aspects, an increase in power consumption can be observed, which in turn leads to an increase in the need for power cooling.

1.2.7 Trade-offs

The majority of the requirements presented below are mutually exclusive and can be considered trade-offs from one another. The material budget is linked to the thickness of the chip, as well as the necessity for power dissipation. The radiation hardness is correlated with the utilized area. A method to mitigate bit flip errors is to triple the design. The relationships between speed, power, and spatial resolution are direct trade-offs.

The objective of the chip design is to achieve a balance that will facilitate the desired particle physics operations. The subsequent section addresses the technology utilized for particle collection. It is crucial to be aware of the potential for developing an optimal sensor. In fact, a new technology for a smaller diode or amplifier will allow more degrees of freedom for the readout.

1.3 Different pixels sensors

This section of the text is devoted to an examination of the various types of pixels utilized in matrix sensing. Other technologies, such as those employing stripes or gas detectors, also exist. All of the aforementioned types of detectors are CMOS detectors with varying properties and characteristics. All semiconductor architectures exhibit comparable power consumption, as they are constructed around a diode, amplified, and readout circuits. The material budget is also comparable, with the exception of the hybrid detectors, which has double the material budget of the others.

1.3.1 Hybrid Pixel Detectors

Hybrid pixels [4] represent a promising advancement in tracker technology, combining the precision of traditional pixel-based tracking with the versatility of hybrid systems. These pixels integrate both charge-integrating and photon-counting capabilities, offering advantages in dynamic range, energy resolution, and temporal resolution. This hybrid approach allows for improved performance in various tracking applications, including medical imaging, astronomy, and particle physics. The hybrid pixels detector is built by assembling 2 layers with bump solder as presented in the figure 1.4.

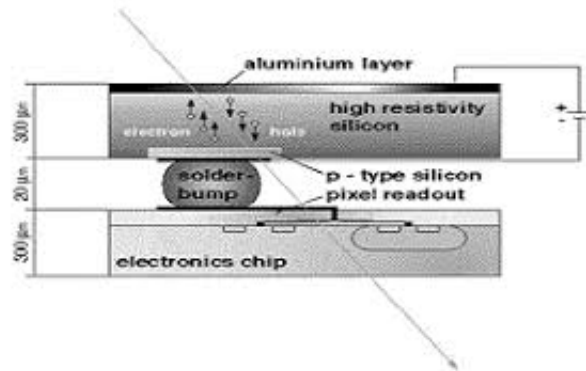


FIGURE 1.4: Cross section view of the basic element of an hybrid pixels detector [5].

One of the main advantages of hybrid pixels is their ability to accurately measure high-energy particles while maintaining sensitivity to low-energy signals, enhancing the overall efficiency of particle detection. Additionally, their ability to precisely localize events in both time and space makes them valuable for applications requiring detailed spatial resolution and temporal accuracy.

However, hybrid pixels also face constraints, notably in terms of fabrication complexity and cost. The integration of multiple functionalities within each pixel requires sophisticated manufacturing processes, which can increase production expenses. Furthermore, optimizing the performance of hybrid pixels often entails trade-offs between parameters such as energy resolution, timing resolution, and pixel density, presenting challenges in achieving an optimal balance for specific tracking tasks.

In conclusion, while hybrid pixels offer significant advantages in terms of performance and versatility for tracking applications, their implementation requires careful consideration of design trade-offs and manufacturing challenges. With ongoing advancements in fabrication techniques and optimization algorithms, hybrid pixel technology holds great promise for future developments in precision tracking systems.

1.3.2 DEPFET

Depleted Field Effect Transistor (DEPFET) [6] represents cutting-edge technology in the realm of pixel detectors, particularly in applications like high-energy physics experiments and X-ray imaging. The DEPFET sensor consists of a sensitive volume where charge carriers are generated by incident particles or photons. This volume is integrated with a field-effect transistor structure, allowing for both signal amplification and charge readout within each pixel. The DEPFET detection is done by adding a bulk contact coming from a deep doping process 1.5

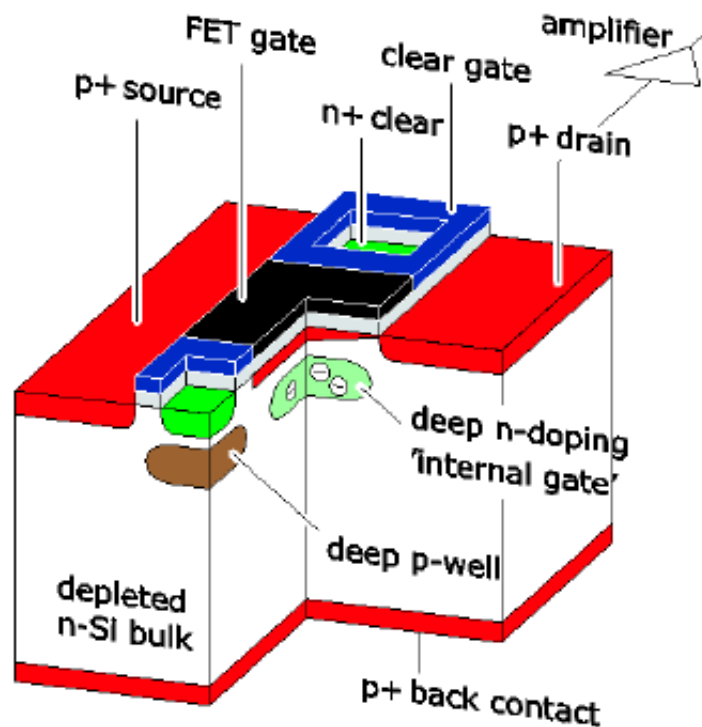


FIGURE 1.5: Cross section view of the pixel in a DEPFET detector [7].

One of the significant advantages of DEPFET lies in its low noise characteristics, enabling high-resolution imaging even at low signal levels. The depletion region within the sensor efficiently collects charge carriers, resulting in superior energy resolution and spatial accuracy compared to conventional pixel detectors. Moreover, DEPFET sensors offer fast readout capabilities, making them suitable for high-rate data acquisition scenarios.

However, DEPFET technology also comes with challenges, notably in terms of fabrication complexity and cost. The manufacturing process requires precise control over doping profiles and thin-film deposition, which can increase production expenses. Additionally, the integration of readout electronics within each pixel necessitates sophisticated fabrication techniques, limiting the scalability of DEPFET arrays.

In summary, DEPFET sensors offer unparalleled performance in terms of sensitivity, spatial resolution, and speed, making them ideal for demanding imaging applications. Despite the challenges associated with fabrication and cost, ongoing research efforts aim to address these limitations, further enhancing the capabilities of DEPFET technology for future scientific and medical advancements.

1.3.3 SOI

Silicon-On-Insulator (SOI) [8] [9] technology revolutionizes semiconductor manufacturing by introducing a layer of insulating material, typically silicon dioxide, between the silicon substrate and the active silicon layer. This innovation offers several advantages over conventional bulk silicon technology. The SOI pixels are created by adding an insulator between the doping and metal layers in figure 1.6

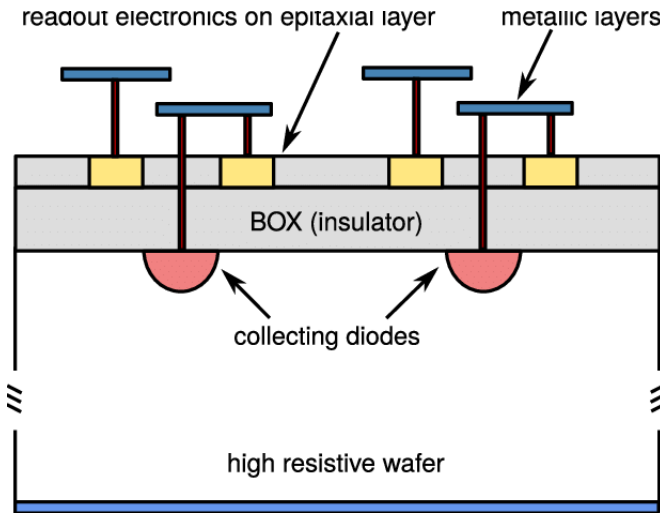


FIGURE 1.6: Cross section view of the pixel in a SOI detector [10].

One key benefit of SOI is reduced power consumption due to decreased parasitic capacitance and leakage currents. The insulating layer isolates the active silicon layer from the substrate, minimizing unwanted electrical interactions and improving transistor performance. This results in faster switching speeds and lower operating voltages, making SOI particularly suitable for high-performance and low-power applications in fields like mobile computing and Internet of Things (IoT) devices.

Moreover, SOI technology enhances device reliability by reducing susceptibility to radiation-induced soft errors and latch-up phenomena. The insulating layer provides a barrier that helps prevent charge accumulation and transient disturbances, improving the robustness of integrated circuits in harsh environments.

Additionally, SOI enables the integration of advanced device architectures, such as fully depleted silicon-on-insulator (FDSOI) transistors, which offer further performance enhancements, including enhanced electrostatic control and reduced short-channel effects.

Despite these advantages, SOI technology presents challenges in terms of fabrication complexity and cost. The additional insulating layer requires additional processing steps and specialized techniques, which can increase manufacturing expenses. Moreover, the adoption of SOI technology may require adjustments in design methodologies and tooling, adding further complexity to the development process.

In conclusion, Silicon-On-Insulator technology offers compelling advantages in terms of performance, power efficiency, and reliability, making it a preferred choice for a wide range of semiconductor applications. As fabrication techniques continue to evolve and economies of scale improve, SOI technology is poised to play an increasingly prominent role in shaping the future of integrated circuit design and manufacturing.

1.3.4 MAPS

Monolithic Active Pixel Sensors (MAPS) [11] [12] [13] [14] are an advancement in sensor technology, particularly in the realm of particle detection and imaging applications. These sensors integrate both the sensing element and the readout electronics within each pixel, offering numerous advantages over traditional pixel detectors. The MAPS detector is based around a standard process with deep doping in the full depleted form at figure 1.7.

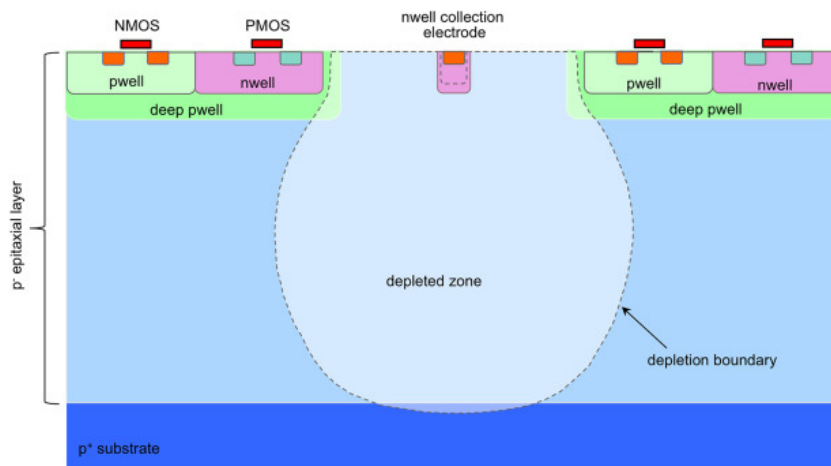


FIGURE 1.7: Cross section view of pixel in a MAPS detector [15].

One of the key benefits of MAPS is its compactness and scalability. By integrating all necessary components into a single silicon substrate, MAPS eliminates the need for bulky external readout circuits, allowing for densely packed pixel arrays with high spatial resolution. This feature makes MAPS ideal for applications requiring fine-grained imaging, such as particle tracking in high-energy physics experiments.

and medical imaging modalities like computed tomography (CT) and mammography.

Moreover, MAPS offers excellent radiation hardness, thanks to their monolithic structure and low material thickness. This resilience to radiation-induced damage makes MAPS suitable for use in high-flux environments, such as particle colliders and space missions, where traditional detectors may degrade over time.

Additionally, MAPS provides fast readout capabilities, enabling real-time data acquisition and processing. This rapid readout speed is advantageous for dynamic imaging applications, including time-resolved spectroscopy and high-speed microscopy.

Despite these advantages, MAPS also poses challenges, such as increased fabrication complexity and cost. The integration of sensitive electronics within each pixel requires advanced manufacturing techniques and precise process control, which can elevate production expenses. Furthermore, optimizing the performance of MAPS often involves trade-offs between parameters like spatial resolution, readout speed, and power consumption, necessitating careful design considerations.

The MAPS also offers the possibility to create a wafer scale sensor that reduces the cable management, increases the speed and simplifies the integration. This type of chip is called VLSI (Very Large Scale Integration) [16].

In conclusion, Monolithic Active Pixel Sensors represent a significant leap forward in sensor technology, offering unparalleled performance and versatility for a wide range of imaging and detection applications. As research and development efforts continue to refine fabrication processes and address technical challenges, MAPS are poised to play an increasingly prominent role in advancing scientific discovery and technological innovation.

1.3.5 Summary

The primary experiment conducted at the IPHC is the ALICE experiment, which is primarily concerned with the material budget and spatial resolution. For these reasons, the group is developing MAPS and attempting to enhance its functionality. It is evident that the readout speed of trackers for ALICE must be accelerated, given that they are situated within matrices. The following section presents an analysis of the physical processes observed in a series of experiments, along with a discussion of the implications of these observations for the requisite conditions.

1.4 Specific tracker requirements for some collider experiments

In particle physics, there are two main types of detectors: calorimeters and trackers. Calorimeters are employed to quantify the energy of particles and are the sole detectors capable of detecting neutral particles. The trackers reconstruct the trajectories of charged particles, thereby enabling the measurement of their momentum and origin. This thesis is focused on sensors dedicated to trackers. This section outlines the necessity of trackers in various experiments.

1.4.1 ALICE

The ALICE (A Large Ion Collider Experiment) [17] is a prominent particle physics experiment at the Large Hadron Collider (LHC) at CERN, specifically designed to study the physics of strongly interacting matter at extreme energy densities. Its primary focus is on understanding the properties of quark-gluon plasma (QGP), a state of matter thought to have existed just after the Big Bang. ALICE aims to investigate how this primordial matter evolves and to explore the fundamental aspects of Quantum Chromodynamics (QCD), the theory describing the strong interaction.

The ALICE experiment has several critical scientific objectives. It seeks to characterize the quark-gluon plasma by studying the behavior of strongly interacting matter at high temperatures and densities. The experiment also aims to understand the mechanisms of particle production, the nature of deconfinement and chiral symmetry restoration, and the dynamics of heavy-ion collisions. Additionally, ALICE investigates the properties of high-density nuclear matter and the modifications of particle properties in the medium.

Central to achieving these objectives is the precise tracking and identification of particles produced in high-energy heavy-ion collisions at the LHC. The ALICE detector includes a sophisticated array of tracking systems, which are essential for reconstructing particle trajectories and identifying their origins. The primary tracking system in ALICE consists of the Inner Tracking System (ITS) and the Time Projection Chamber (TPC).

The ALICE ITS2 tracker is the largest MAPS-based sub-detector created. The ALICE ITS3 with bended stitched MAPS and even the ALICE3 aims to build trackers mostly around MAPS [18]. The ALICE experiment's Inner Tracking System (ITS) must endure radiation levels up to 270 krad per year and 1×10^{13} neq/cm². The pixel detector power density is around 0.3 W cm⁻². ITS supports readout rates up to 50 kHz for Pb-Pb collisions (Plomb-Plomb) and 200 kHz for pp collisions (proton-proton), with a hit rate of about 100 hits/cm²/s. Spatial resolution is approximately 5 μm to 10 μm, and the system contains around 12.5 million pixels. These specifications ensure precise particle tracking and identification under extreme collision conditions at the LHC.

The ALICE experiment at CERN (see figure 1.9 and figure 1.8) is a leading initiative in particle physics aimed at exploring the properties of quark-gluon plasma and the behavior of strongly interacting matter at extreme conditions. With its advanced tracking systems, particularly the Inner Tracking System and the Time Projection Chamber, ALICE is well-equipped to investigate a wide range of phenomena related

to heavy-ion collisions and the fundamental aspects of Quantum Chromodynamics.

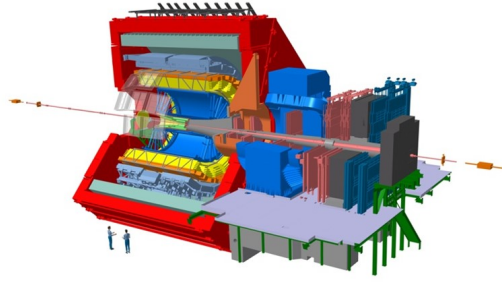


FIGURE 1.8: Schematic of the ALICE experiment [19].



FIGURE 1.9: Photography of the ALICE experiment.

1.4.2 CMS

The CMS (Compact Muon Solenoid) [20] [21] experiment is one of the largest and most comprehensive particle physics experiments conducted at the Large Hadron Collider (LHC) at CERN. It is designed to explore a wide range of physics, including the search for the Higgs boson, the discovery of new particles, and the investigation of the fundamental forces that govern the universe. The CMS detector is a general-purpose detector with a broad scientific agenda, making significant contributions to our understanding of particle physics.

The CMS experiment has several key scientific objectives. It aims to study the properties of the Higgs boson, investigate the nature of dark matter, explore the possibility of extra dimensions, and search for signs of supersymmetry (SUSY). Additionally, CMS examines the strong force that binds quarks and gluons together within protons and neutrons, tests the validity of the Standard Model, and looks for deviations that might indicate new physics.

Central to achieving these objectives is the precise tracking and identification of particles produced in high-energy proton-proton collisions at the LHC. The CMS detector includes a sophisticated array of tracking systems, which are crucial for reconstructing particle trajectories and identifying their origins. The primary tracking system in CMS consists of the silicon tracker, which is located close to the interaction point and surrounded by other sub-detectors, including the electromagnetic calorimeter (ECAL) and the hadron calorimeter (HCAL).

The CMS trackers do not feature any MAPS by could have been has the ATLAS experiment since they feature close requirements. The CMS experiment's tracking systems, including the Silicon Tracking System (STS) and the Micro Vertex Detector (MVD), must withstand radiation levels up to 1 Mrad and 1×10^{14} neq/cm². The power density for the STS is around 0.5 W cm⁻². The readout rate supports up to 1 MHz, with hit rates of approximately 10 000 hits/cm²/s. The spatial resolution for the STS is about 30 μm, and the MVD features a resolution of 5 μm. The total number of pixels in the MVD is around 10 million. These specifications enable precise tracking and vertex reconstruction in the high-density environments studied by the

CMS experiment.

The CMS experiment at CERN is a leading initiative in particle physics aimed at exploring the fundamental forces and particles that make up the universe. With its advanced tracking systems, particularly the silicon tracker, CMS is well-equipped to investigate a wide range of physics phenomena and search for new physics beyond the Standard Model.

1.4.3 ATLAS

The ATLAS (A Toroidal LHC ApparatuS) [22] experiment is one of the four major experiments at the Large Hadron Collider (LHC) at CERN. It is designed to explore a broad spectrum of physics, from the search for the Higgs boson to the discovery of new particles and the investigation of fundamental interactions. The ATLAS detector is a general-purpose detector with a wide-ranging scientific agenda, contributing significantly to our understanding of particle physics and the fundamental forces of nature.

The ATLAS experiment has several critical scientific objectives. It aims to study the properties of the Higgs boson, investigate the nature of dark matter, explore potential extra dimensions, and search for evidence of supersymmetry (SUSY). Additionally, ATLAS examines the strong interaction that binds quarks and gluons within protons and neutrons, tests the predictions of the Standard Model, and searches for deviations that might indicate new physics phenomena.

Central to achieving these objectives is the precise tracking and identification of particles produced in high-energy proton-proton collisions at the LHC. The ATLAS detector features a sophisticated array of tracking systems, essential for reconstructing particle trajectories and identifying their origins. The primary tracking system in ATLAS consists of the Inner Detector (ID), which includes the Pixel Detector, the Semi-Conductor Tracker (SCT), and the Transition Radiation Tracker (TRT).

The ATLAS experiment has done an R&D prospection for trackers build around MAPS for the ITk [23] but the MAPS technology was not chosen. The ATLAS experiment's tracking systems, including the Inner Detector (ID) with the Pixel Detector, Semi-Conductor Tracker (SCT), and Transition Radiation Tracker (TRT), must withstand radiation levels up to 500 krad and 1×10^{15} neq/cm². The power density for the Pixel Detector is approximately 1 W cm⁻². The readout rate supports up to 40 MHz, with hit rates around 1 MHz cm⁻². The spatial resolution for the Pixel Detector is about 10 μm, and the SCT provides a resolution of 17 μm. The total number of pixels in the Pixel Detector is around 80 million. These specifications ensure high precision in particle tracking and identification under the extreme conditions of proton-proton collisions at the LHC.

The ATLAS experiment at CERN is a leading initiative in particle physics aimed at exploring the fundamental forces and particles that constitute the universe. With its advanced tracking systems, particularly the Inner Detector, ATLAS is well-equipped to investigate a wide range of physics phenomena and search for new physics beyond the Standard Model.

1.4.4 LHCb

The LHCb (Large Hadron Collider beauty) [24] experiment is a prominent particle physics project designed to study the differences between matter and antimatter and to explore phenomena that lie beyond the Standard Model of particle physics. Located at CERN, the European Organization for Nuclear Research, the LHCb is one of the four main experiments at the Large Hadron Collider (LHC). Its primary focus is on the behavior of particles containing bottom (or beauty) quarks, which are crucial for understanding CP violation and the matter-antimatter asymmetry in the universe.

The LHCb experiment has several critical scientific objectives. It aims to investigate the properties and decays of heavy flavor hadrons, particularly those containing bottom and charm quarks. By studying these particles, LHCb seeks to uncover new sources of CP violation, which could help explain why the universe is dominated by matter rather than antimatter. Additionally, the experiment looks for rare decays and new particles that could provide insights into physics beyond the Standard Model.

Central to achieving these objectives is the precise tracking and identification of particles produced in high-energy proton-proton collisions at the LHC. The LHCb detector is designed to detect forward particles, those produced at small angles relative to the beamline, which is where particles containing bottom quarks are most likely to be found. The detector includes a sophisticated array of tracking systems, which are essential for reconstructing particle trajectories and identifying their origins.

The LHCb experiment are interested in MAPS detector for their future trackers UT and MT [25]. The LHCb experiment's tracking systems, including the Vertex Locator (VELO), Tracker Turicensis (TT), Inner Tracker (IT), and Outer Tracker (OT), must withstand radiation levels up to 300 krad and 1×10^{14} neq/cm². The power density for the VELO is around 0.4 W cm⁻². The readout rate supports up to 40 MHz, with hit rates of approximately 500 khits/cm²/s. The spatial resolution for the VELO is about 10 μ m, and the IT provides a resolution of 50 μ m. The total number of pixels in the VELO is around 40 million. These specifications enable precise vertexing and particle tracking in the forward region of high-energy proton-proton collisions at the LHC.

The LHCb experiment at CERN is a leading initiative in particle physics aimed at unraveling the mysteries of CP violation and the matter-antimatter asymmetry in the universe. With its advanced tracking systems, particularly the VELO, TT, IT, and OT, LHCb is well-equipped to explore the properties of particles containing bottom quarks and to search for new physics beyond the Standard Model.

1.4.5 CBM

The Compressed Baryonic Matter (CBM) [26] experiment is a pioneering project in nuclear physics, designed to explore the properties of dense baryonic matter. This state of matter is thought to exist in the cores of neutron stars and during the early stages of heavy-ion collisions. The experiment is one of the flagship projects at the

Facility for Antiproton and Ion Research (FAIR) in Darmstadt, Germany. Its primary aim is to investigate the phase diagram of Quantum Chromodynamics (QCD) at high baryon densities and moderate temperatures.

The CBM experiment has several key scientific objectives. It aims to study the Equation of State (EOS) of nuclear matter under extreme conditions, explore the onset of chiral symmetry restoration and its effects on hadron masses, investigate the properties of exotic forms of matter such as strange quark matter, and understand the mechanisms of particle production and transport in high-density environments.

Central to achieving these objectives is the use of high-energy heavy-ion collisions, which create the conditions necessary to study high-density nuclear matter. The experiment relies on a sophisticated array of detection systems to track and identify particles produced in these collisions. The most critical components of the detection system are the trackers, which include the Silicon Tracking System (STS) and the Micro Vertex Detector (MVD).

The Silicon Tracking System (STS) is essential for the precise tracking of charged particles. It provides high-resolution spatial measurements that allow researchers to reconstruct particle trajectories with great accuracy. This system is designed to operate in the high-radiation environment typical of heavy-ion collisions, ensuring reliable performance under challenging conditions.

The MAPS trackers has been chosen with the MIMOSIS chip in the MVD and are considered for the future STS upgrade [27]. The CBM experiment's tracking systems, including the Silicon Tracking System (STS) and the Micro Vertex Detector (MVD), must withstand radiation levels up to 1 Mrad and 1×10^{14} neq/cm². The power density for the STS is around 0.5 W cm⁻². The readout rate supports up to 1 MHz, with hit rates of approximately 10 khits/cm²/s. The spatial resolution for the STS is about 30 μ m, and the MVD features a resolution of 5 μ m. The total number of pixels in the MVD is around 10 million. These specifications enable precise tracking and vertex reconstruction in the high-density environments studied by the CBM experiment.

The CBM experiment at FAIR is a groundbreaking initiative aimed at exploring the fundamental properties of dense baryonic matter. With its advanced tracking systems, particularly the STS and MVD, it is poised to make significant contributions to our understanding of the high-density phase of QCD and the behavior of matter under extreme conditions.

1.4.6 Belle II

The Belle II [28] experiment is a globally recognized research endeavor hosted at the SuperKEKB accelerator facility in Tsukuba, Japan. It focuses on the study of rare decays and phenomena involving B mesons and tau leptons, aiming to shed light on the fundamental principles of particle physics.

At its core, the Belle II experiment seeks to explore the mysteries surrounding matter-antimatter asymmetry, also known as CP violation, and to scrutinize the

properties of rare decays and particles that could offer insights beyond the Standard Model of particle physics.

The tracker used at KEK are currently DEPFET and strips but the upgrade truly considered to do it with MAPS for the VTX [29]. The Belle II experiment's tracking systems, including the Vertex Detector (VXD) and the Central Drift Chamber (CDC), must withstand radiation levels up to 10 krad and 1×10^{12} neq/cm². The power density for the VXD is around 0.2 W cm⁻². The readout rate supports up to 30 kHz, with hit rates of approximately 500 hits/cm²/s. The spatial resolution for the VXD is about 10 μ m, and the CDC provides a resolution of 100 μ m. The total number of pixels in the VXD is around 8 million. These specifications ensure high precision in vertexing and particle tracking in the complex environment of electron-positron collisions at the SuperKEKB accelerator.

In conclusion, the Belle II experiment represents a cutting-edge pursuit in particle physics, aiming to unravel the mysteries of matter-antimatter asymmetry and probe the frontiers of the Standard Model. Through international collaboration and innovative technology, Belle II seeks to advance our understanding of fundamental particles and their interactions.

1.4.7 FCCee

The Future Circular Collider electron-positron (FCCee) [30] experiment is a pioneering international research initiative poised to advance our understanding of particle physics. Located at the proposed Future Circular Collider facility at CERN, this experiment aims to explore fundamental particle interactions with unprecedented precision and energy levels.

At its core, the FCCee experiment seeks to address key questions in particle physics and cosmology, including the nature of dark matter, the existence of new particles beyond the Standard Model, and the origins of the universe's matter-antimatter asymmetry.

To achieve these ambitious goals, the FCCee experiment will utilize state-of-the-art particle detectors and accelerator technologies. These include high-precision tracking detectors, electromagnetic and hadronic calorimeters, particle identification systems, and advanced data acquisition systems.

The FCC-ee experiment's tracking systems are considering MAPS detector as an initial solution for the vertexing and at least for the first tracking layers. They include the Vertex Detector (VXD) and the Silicon Tracker, must withstand radiation levels up to 1 Mrad and 1×10^{14} neq/cm². The power density for the VXD is around 0.1 W cm⁻². The readout rate supports up to 100 kHz, with hit rates of approximately 1000 hits/cm²/s. The spatial resolution for the VXD is about 3 μ m, and the Silicon Tracker provides a resolution of 10 μ m. The total number of pixels in the VXD is around 10 million. These specifications ensure precise vertexing and particle tracking in the high-luminosity environment of electron-positron collisions at the Future Circular Collider (FCCee).

In conclusion, the FCCee experiment represents a groundbreaking endeavor in particle physics, poised to push the boundaries of our understanding of the universe's fundamental constituents and interactions. Through international collaboration and cutting-edge technology, FCCee aims to unlock the secrets of the universe's origin and composition.

1.4.8 Other experiments: Quantum physics

The Quantum Physics Experiment is a forefront research initiative aimed at exploring the fundamental principles and phenomena of quantum mechanics. Conducted by leading research institutions worldwide, this experiment seeks to unravel the mysteries of quantum phenomena and harness their potential for transformative technologies. One of the primary applications of this technology is in the field of quantum computers, which will facilitate the completion of significant tasks in a relatively short period of time. The detector utilized for data acquisition is a particle detector.

In a quantum computer reading tracker system, the quantum bits (qubits) must endure radiation levels up to 10 Mrad and 1×10^{14} neq/cm². The power density for the qubit control electronics is around 0.2 W/cm². The readout rate supports up to 1 MHz, with hit rates of approximately 1000 hits/cm²/s. The spatial resolution for the qubit readout is about 10 μ m. The total number of qubits in the system is around 10 million. These specifications ensure accurate and reliable qubit readout in the high-radiation and high-data environment of a quantum computer.

In conclusion, the Quantum Physics Experiment represents a pioneering effort to unlock the potential of quantum mechanics for scientific discovery and technological innovation. Through international collaboration and interdisciplinary research, this experiment aims to revolutionize our understanding of the quantum world and shape the future of technology.

1.4.9 Summary

In all of these experiments, the primary focus is on either the material budget and power consumption or the hit rate and radiation hardness. The experiments in question are ALICE, LHCb, Belle II, FCCee, and the quantum computer. With regard to the material budget, they are representative of the relevant aspects. ATLAS, CMS, and CBM are representative of the higher rates. It would be beneficial to develop a readout architecture that can be adapted to all experiments. The subsequent chapter will develop the existing readout technologies for MAPS. However, this thesis aims to propose a solution that can achieve high bandwidth or low bandwidth with minimal power consumption, regardless of whether the asynchronous approach is employed. The table 1.1 shows the requirements from an experiment to another. The "time figure" shows sometimes the expected timestamp of the experiment or a time to differentiate the bunch crossing.

Experiment	ALICE	CMS	ATLAS	LHCb	CBM	Belle II	FCCee
Sub-detector	ITS3	HGTD	ITk	UT	MVD	VXD	VTX
Expected year	2029	2026	2029	2035	2026	2028	2040
Spatial resolution	5	100-200	5	10	5	10	3
Hit rate [MHz/cm ²]	9	1-2	40	160	15-70	100	20
Time figure [ns]	5x10 ³		25	O(1)	5x10 ³	100	100-1000
Power density [mW/cm ²]	20	100-200	500	100-300	100-200	200-300	200-400
Material budget [%X/X ₀]	0.05	1-2	1.5	1	0.3	0.15	0.15
Tracker sensing area [m ²]	10	10	13	4.5	2.5	1	1
Non-ionising [MeV · n _{eq} /cm ²]	3 · 10 ¹²	1 · 10 ¹⁵	1 · 10 ¹⁵	3 · 10 ¹⁵	10 ¹⁴ /y	5 · 10 ¹³	5 · 10 ¹¹

TABLE 1.1: Summary of the requirements on tracking sensors for some selected experiments.

1.5 Conclusion

This chapter showed that CMOS MAPS present very interesting features as sensors for trackers in high energy physics. Hence, they are considered for future sub-detectors in a number of experiments (ALICE, Belle II, CBM, LHCb, FCCee). The aforementioned project requirements exhibit a considerable degree of variability. However, from the point of view of design resources, it would be desirable to develop a single solution matching this variability. This is probably unreachable but it remains a good research direction for a new sensor architecture. The subsequent chapter elucidates the manner in which these requirements impose constraints on the sensor design and more specifically the read-out architecture.

Résumé

Chapitre 1 : État de l'art des capteurs de suivi en physique des particules basés sur les pixels

Le Chapitre 1 de cette thèse est consacré à l'examen des capteurs de suivi utilisés dans les expériences de physique des particules, en particulier les technologies basées sur les pixels. Ce chapitre vise à fournir un cadre théorique solide en introduisant les concepts fondamentaux et en explorant les développements récents dans le domaine des capteurs de particules. Les capteurs de suivi jouent un rôle critique dans les expériences de physique des hautes énergies, car ils permettent de détecter et de reconstruire les trajectoires des particules générées lors des collisions.

Les capteurs à pixels sont particulièrement précieux dans ce contexte en raison de leur capacité à fournir une résolution spatiale élevée, une caractéristique essentielle pour reconstruire les trajectoires des particules avec précision. Le chapitre commence par une introduction générale à la physique des particules et aux détecteurs utilisés, avant de plonger dans les détails des capteurs à pixels.

1.1 Introduction à la physique des particules et aux détecteurs

Les expériences de physique des hautes énergies, telles que celles menées au Grand collisionneur de hadrons (LHC), nécessitent des détecteurs extrêmement précis pour suivre les trajectoires des particules produites lors des collisions. Ces détecteurs doivent être capables de résister à des environnements de radiations intenses tout en fournissant des mesures précises à des vitesses élevées. Les détecteurs de particules sont donc conçus pour combiner robustesse et précision, deux caractéristiques souvent en tension.

Les capteurs à pixels se sont imposés comme une solution de choix pour ces applications, car ils permettent une détection précise sur des surfaces étendues, ce qui est crucial pour les expériences nécessitant une couverture spatiale importante. Le développement de ces technologies a été motivé par les besoins croissants en matière de résolution spatiale et de rapidité de lecture, des facteurs qui ont conduit à l'évolution des capteurs à pixels des premières versions hybrides vers les capteurs à pixels actifs monolithiques (MAPS).

1.2 Capteurs à pixels hybrides vs. MAPS

Le chapitre compare ensuite deux grandes familles de capteurs à pixels : les capteurs hybrides et les MAPS. Les capteurs à pixels hybrides, bien que robustes, sont caractérisés par une séparation entre le capteur et l'électronique de lecture, ce qui peut entraîner des défis en termes de complexité et de coût de fabrication. Ces capteurs consistent en une matrice de pixels connectée à une électronique de lecture via des micro-boules de soudure, une architecture qui permet une grande flexibilité mais à un coût élevé en termes de production et d'intégration.

En revanche, les MAPS intègrent à la fois le capteur et l'électronique de lecture sur une seule puce de silicium, ce qui permet de réduire la complexité du système tout en offrant une résolution spatiale proche. Cette intégration monolithique permet également de diminuer les coûts de fabrication, rendant les MAPS plus attractifs pour une utilisation à grande échelle. Cependant, cette intégration pose des défis techniques, notamment en matière de dissipation thermique et de gestion des radiations, qui sont discutés en détail dans ce chapitre.

1.3 Exigences des capteurs en physique des particules

L'un des aspects les plus critiques des capteurs de suivi est leur capacité à fonctionner efficacement dans des environnements de radiation intense, une caractéristique indispensable pour les détecteurs de particules dans des expériences telles que celles du LHC. Le chapitre aborde les méthodes pour améliorer la tolérance aux radiations des capteurs, notamment à travers des techniques de durcissement du silicium et l'optimisation des circuits intégrés.

La rapidité de lecture est un autre facteur clé, car les expériences de physique des particules génèrent d'énormes volumes de données en un temps très court. Les capteurs doivent donc être capables de traiter ces données à une vitesse extrêmement élevée pour éviter les pertes d'information. Ce besoin de rapidité a conduit à l'adoption d'architectures de lecture asynchrone, qui permettent de minimiser la latence en traitant les signaux dès qu'ils sont reçus, sans attendre un signal d'horloge global.

Enfin, la dissipation thermique est un problème majeur pour les capteurs à haute densité de pixels, car la chaleur générée peut dégrader les performances du capteur et limiter sa durée de vie. Le chapitre explore différentes approches pour gérer la dissipation thermique, y compris l'utilisation de matériaux avancés et de conceptions optimisées des circuits.

1.4 Défis et perspectives futures

Le chapitre se termine par une discussion sur les défis à venir dans le développement des capteurs à pixels pour la physique des particules. L'amélioration de la résolution spatiale reste une priorité, mais elle doit être équilibrée avec d'autres considérations telles que la réduction des coûts et l'amélioration de la robustesse des capteurs dans des environnements extrêmes. Les directions futures incluent également l'intégration de technologies avancées, telles que les capteurs 3D, qui pourraient offrir des avantages significatifs en termes de compacité et de performances.

La recherche continue dans ce domaine vise à repousser les limites de ce qui est possible en matière de détection de particules, avec l'objectif de fournir des outils toujours plus précis pour les expériences futures. Les avancées dans les capteurs à pixels pourraient également avoir des applications au-delà de la physique des particules, dans des domaines tels que l'imagerie médicale et les systèmes de surveillance industrielle, où des capteurs à haute résolution et à grande vitesse sont également nécessaires.

Chapter 2

State of the art for readout in particle physics pixel tracker sensors

This chapter provides an overview of the different architectures implemented to read out pixel matrices, having in mind the requirements exposed in chapter 1. The final section provides a summary of some sensors developed for particle physics in various experiments, primarily for MAPS, illustrating the evolution of the readout architecture and their performances.

Comparing these architectures intends to assess the potential benefit of a fully asynchronous logic to approach sensor the requirements when the hit occupancy over the matrix is actually low, less than a few percent.

The architectures presented here are mostly designed for use with matrix CMOS pixel sensors. The architectures are presented in the order of their chronological development, with the oldest one presented first because it was also the simpler one to design. The initial architecture does not incorporate digital reading, and the third architecture is the first to include discrimination inside the pixels. The final architectures benefit from a novel approach to information reading, namely zero suppression. The objective is to read only the pixels that have fired, thereby reducing the required bandwidth, so called data driven readout.

In order to comprehend the functioning of a MAPS tracker sensor, it is essential to gain an understanding of its classical architectural design. The sensor is a diode that is commonly attached to an amplifier, a pulsing circuit that is used to create the polarization and a discriminator circuit that is used to set a level for pixel firing. All of the aforementioned components are integrated into the pixel structure and are designed in an analog manner to a digital output. The readout is typically a double-column structure attached to pixels on both sides, with glue logic employed to mask pixels and test features. The aim of the readout is to provide the address of the fired pixels outside the matrix. Subsequently, a peripheral circuit is present to control all functions, including slow control, data storage, output communication, and, on occasion, readout operations. The figure 2.1 presents a classical topology of this architectural configuration with the pixels represented with empty squares and the readout in between.

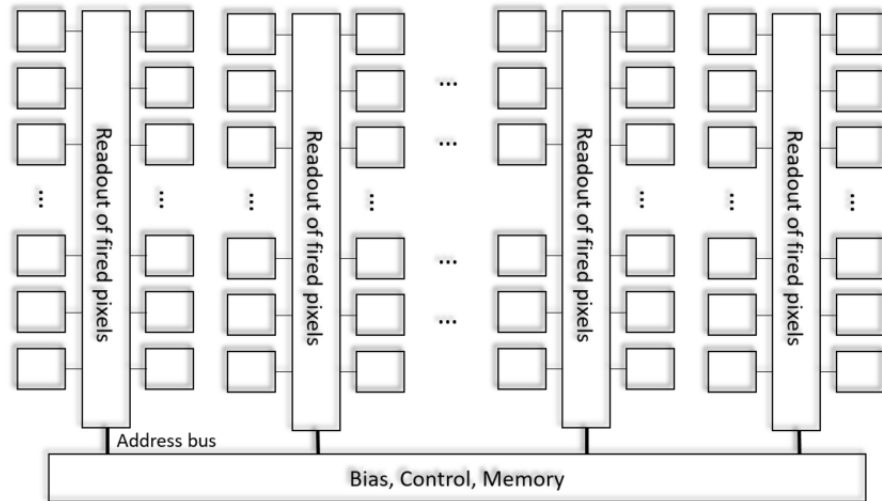


FIGURE 2.1: Classical topology of a matrix for monolithic pixel sensors.

2.1 Readout by pixels or columns

The initial readout presented is a pixel-by-pixel or column-by-column readout with the rolling shutter. This type of readout scan the whole matrix every time and read all pixel. It is inefficient and was subsequently superseded by the advent of thinner technologies that facilitate the integration of intelligence within the pixel. This intelligence allows to read only fired pixels and will be presented in the next section.

2.1.1 Analog readout

The analog readout was the initial method for reading a matrix of pixels. The pixel information is stored in a capacitor and can be read via an analog output. To restrict the number of outputs, a single output is implemented in conjunction with a selector, such as a snake or analog multiplexer. The snake strategy is employed for a limited matrix, constrained by the current for selection. It scans to matrix in a pixel by pixel in a snake shape. The analog multiplexer introduces noise into the system. It is evident, therefore, that not all solutions are optimal. This type of readout was implemented in a lot of circuit and in particle physics our laboratory build the circuits MIMOSA 1 to 4. The MIMOSA 1 [11] is a proof of concept with four analog outputs, a small pitch (20 μm), and 16,000 pixels. The MIMOSA 5 [31] is the inaugural reticle prototype sensor with one million pixels and a pitch of 17 μm . Both circuits were designed using AMS 0.6 technology.

This type of reading is easy to implement in a circuit but present a lot of disadvantages like the consumption to read a pixel and transmit the information outside the matrix, the low reading speed because pixels are read one after the other and the information quantity is huge. All this reasons make this architecture not suitable for particle detection.

2.1.2 Rolling shutter

The rolling shutter represents an enhancement to the readout process, offering the potential for utilisation in both analog and digital readout applications. It is a method of rearranging the output canals in order to enhance the bandwidth. The rolling shutter is a pervasive mechanism in digital imaging devices, including cameras and smartphones, for capturing images or videos figure 2.2. In contrast to a global shutter, which exposes the entire sensor simultaneously, a rolling shutter exposes different parts of the sensor sequentially, typically from the top to the bottom or vice versa.

The application of this sequential exposure technique results in the generation of a range of distinctive visual effects. One noteworthy consequence of this phenomenon is the distortion of images in rapidly moving scenes. The sensor captures different parts of the image at slightly different times, which can result in the appearance of skewed or distorted moving objects, a phenomenon known as the rolling shutter effect. This effect is particularly noticeable in videos of fast-moving objects or when the camera is panned rapidly.

Furthermore, the rolling shutter can result in the introduction of temporal artifacts in videos, such as wobbling or jello-like distortions, particularly when the camera or the subject is in motion. These artifacts are the result of the asynchronous scanning of the sensor, which leads to temporal inconsistencies in the captured frames.

Despite these drawbacks, the rolling shutter design offers several advantages, including lower power consumption and cost compared to global shutter designs. Furthermore, it permits the attainment of higher resolution and faster readout speeds, rendering it an optimal choice for a multitude of consumer-grade imaging applications.

In applications where precise temporal synchronization and minimal distortion are critical, such as professional videography or high-speed imaging, global shutter sensors are preferred. It is of paramount importance to gain an understanding of the characteristics and limitations of the rolling shutter in order to utilise digital imaging devices in a variety of contexts in an effective manner.

The reduction of the transistor length led to the introduction of digital intelligence in chips, a development that revolutionized the field of electronics. For the prototypes, the digital component could not be integrated directly into the pixels, but rather positioned outside the matrix. The size of the pitch was insufficient to accommodate a significant number of logic gates, and the design of the diode limited the use of P-MOS transistors. The primary objective was to repurpose the existing analog multiplexers and integrate a digital circuit to differentiate the values by reading the entire matrix in an automated manner.

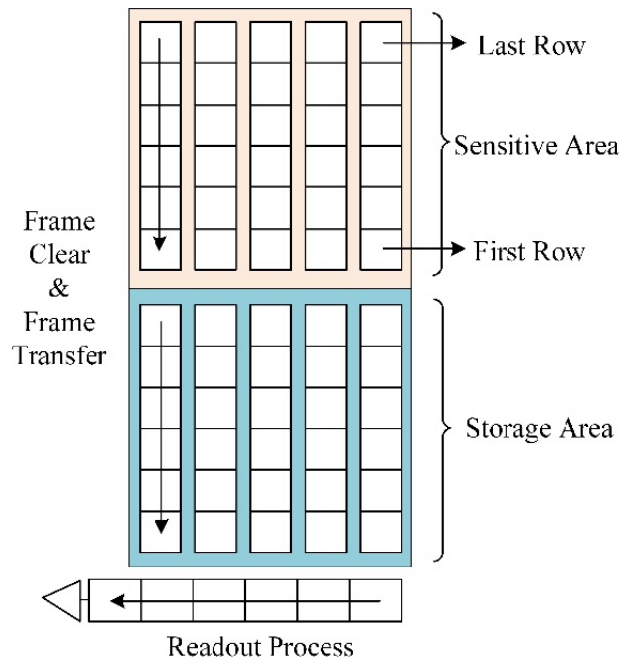


FIGURE 2.2: A representation of the rolling shutter readout [32].

The readout is notably slow due to the time required to read each pixel, which is a function of the number of pixels. A novel approach was introduced to enhance the speed of the process by reading the entire row in units with the use of multiple

acquisition systems. In this instance, the readout time is multiplied by the number of columns only. The primary advantage of this readout is its simplicity, which is achieved through the use of a shift register. However, this approach necessitates a significant amount of space within the matrix, and it can also result in a considerable reduction in the readout speed when the matrix is composed of one million pixels or more.

The consumption is linked to the low number of gates due to the small architectural constraints, as well as the activity on those gates and the frequency of operation. For illustrative purposes, consider a frame with a duration of $1\text{ }\mu\text{s}$ and a matrix size of 1000 pixels by 1000 pixels. In order to read every pixel in one microsecond, the frequency must be set to one-tenth of a picosecond, or one terahertz. The analysis reveals an exceedingly high frequency, which is accompanied by a correspondingly high power consumption.

The circuit MIMOSA 6 [33] to MIMOSA 28 [34] were developed with the rolling shutter for the experiment EUDET (MIMOSA-26 [35] [36]) and START PXL [34] [37]. The versions from 6 to 25 were constructed with a conventional digital output, using a discriminator inside the pixel. The pixel pitch was approximately $20\text{ }\mu\text{m}$, with matrices approaching the megapixel range.

2.2 Zero suppression readout

The zero suppression algorithm is designed to read only those pixels that have been fired. The following section presents the various interpretations of this concept. By only reading the fired pixels, the output bandwidth is reduced, as is the consumption. The rationale behind this methodology is that the advent of thinner technologies has enabled the integration of greater logic capacity within a single device. This approach involves the incorporation of two key components: firstly, a discriminator with a digital output [33], and secondly, a dedicated readout circuit that facilitates the selective reading of pixels.

2.2.1 Priority encoder

The priority encoder is a pseudo-asynchronous circuit that employs a zero suppression algorithm figure 2.3 [38]. The fundamental principle is to encode the fired pixels in binary with a priority. For the sake of argument, let us assume that the priority encoder is a four-to-one small tile, that the inputs are pixels 0 to 3, and that the highest priority is the 3rd pixel. In the event that pixels one and three have fired, only the address of the third pixel is encoded and subsequently read and hidden. The subsequent phase is identical, with the exception that pixel one, which is currently the one with the highest priority, is considered. A diagrammatic representation is provided below figure 2.3, as well as a chronogram figure 2.4.

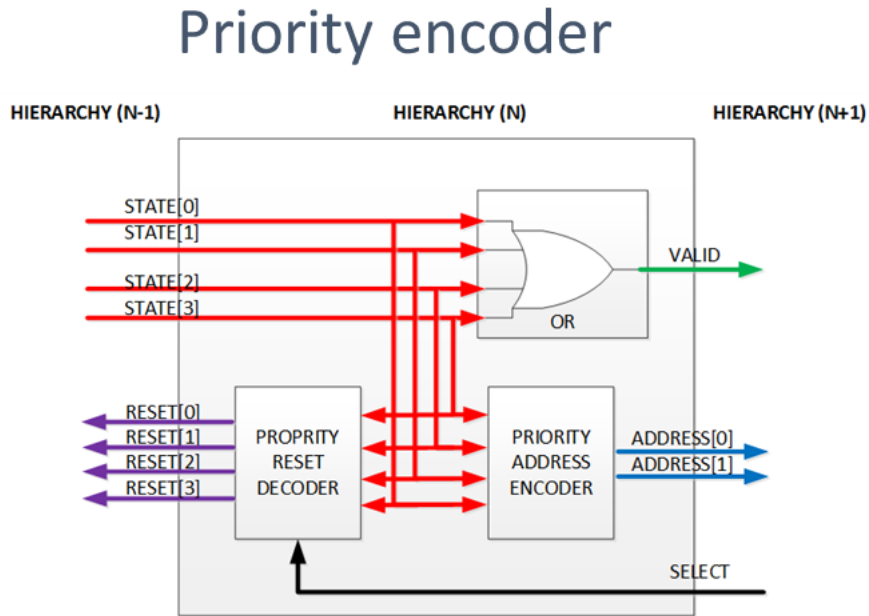


FIGURE 2.3: Schematic of the Priority Encoder.

The "states" signals are the pixels that have been hit, the "resets" signals are to acknowledge the correct pixel, the valid act as a flag to know if an address has to be read, and the select is a propagation over levels of the reset signal, see figure 2.4.

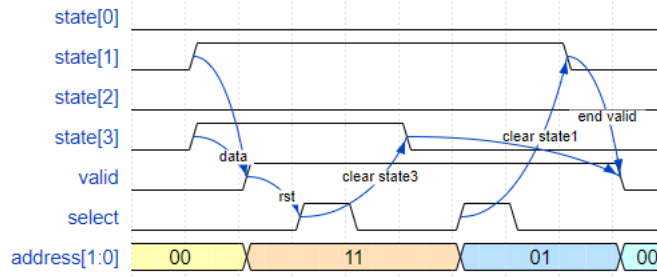


FIGURE 2.4: Chronogram of the Priority Encoder.

The dimensions of this small tile can vary, and it can be cascaded to create a tree-like shape. This approach allows for the identification of potential trade-offs between size and timing, which may be necessary to achieve certain optimizations. The IPHC laboratory typically employs a 4 to 1 small tile ratio and modifies the tree to align with the requisite physical constraints. This appears to be the optimal compromise in terms of size and reading speed.

The priority encoder features a compact and combinatorial schematic, which is beneficial in a matrix with a high density of pixels. Furthermore, the circuit exhibits excellent radiation hardness, as it does not employ any registers so it is not susceptible to bit-flips regardless of the doses. This type of circuit is faster than a CCD, but it is still relatively slow because it can read one pixel at a time across the entire tree. In order to store the data at the end of the column, an additional circuit is required and must be synchronously clocked. The circuit reads one address per clock cycle so it can be adjusted with a trade-off on the consumption. Classically, the IPHC uses a 25 ns clock period with a consumption around 10 mW/cm². The data regarding the timing is unavailable since the interval between the subsequent clock edge and the activated pixel is undetermined. An asynchronous reading can be conducted; however, it has not yet been developed and is not an optimal solution for the asynchronous behavior. The 25 ns clock is derived from the frequency of the bunch crossing that creates frames and limits power, as there is no need to operate at faster speeds.

The priority encoder is a commonly utilized component in particle physics, exhibiting favorable characteristics for experiments such as ALICE. The first in-pixel readout circuit, ALPIDE [13] [39], which features a pitch of 28 μm and 500,000 pixels, was equipped with this encoder. The MIMOSIS 0 [26] to 2 [40] circuit employs a priority encoder as a readout circuit with a medium pitch of 30 μm and 500,000 pixels. The MOSS stitched sensor [16], developed for the ALICE collaboration, comprises 6.72 million pixels with a small pitch of 18 μm .

2.2.2 With token ring

The token ring is a synchronous circuit that implements a zero suppression algorithm figure 2.5. In this instance, a token is placed in the column and transferred pixel by pixel. In the event that a pixel is firing and the token is present, the pixel transmits its data and then passes the token to the subsequent pixel, thereby completing a loop around the column. The pixel on top pass the token to the pixel on the bottom.

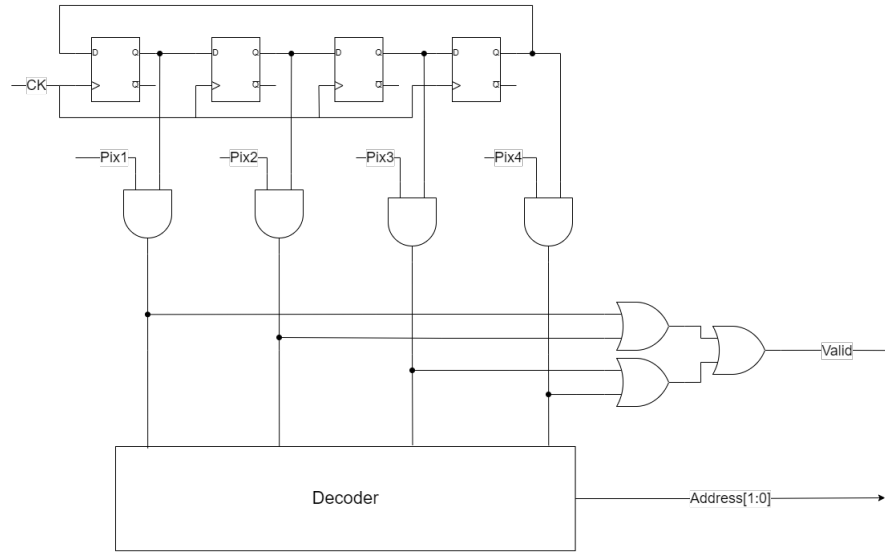


FIGURE 2.5: View of the token ring readout.

This circuit is compact due to its incorporation of a shift register, yet it is still utilized as a synchronous circuit with a clock to facilitate the transfer of the token. The circuit does not make use of the available bandwidth, as it is capable of transmitting a single pixel per clock cycle.

The token ring readout is a commonly utilized technology in particle physics for experiments such as ATLAS, where the circuits FEI3 and FEI4 [41] are employed. These circuits feature a large pitch of approximately 250 μm and a low pixel count of 30,000. The Timepix circuits were also based on a token ring and feature a much more reasonable pitch of 55 μm and a higher pixel count of 65,000. This type of architecture is also implemented in TJ-Monopix2 [42] and the OBELIX [43] chip that re-uses parts of the Monopix. A hit-OR is added to increase the time resolution in the OBELIX chip.

2.2.3 Pulse width encoding

The pulse width encoder employs pulses to transmit an address over a common line, see figure 2.6. The width of the pulse encodes the address ID. The circuit requires the use of oscillators in each pixel, which must be synchronized. Additionally, a time counter must be placed at the end of each column.

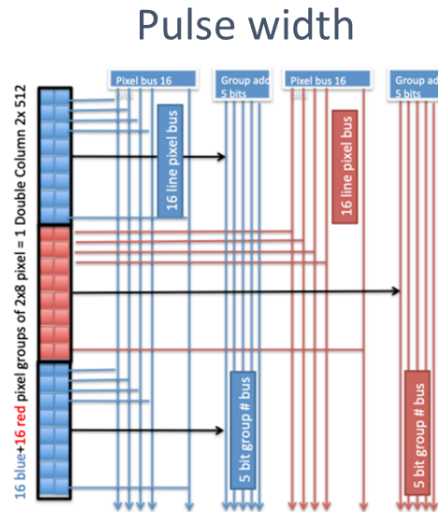


FIGURE 2.6: Schematic view of the pulse width readout [44].

This type of circuit is characterized by its high speed, as it is capable of transmitting pulses in close proximity to one another. Additionally, it does not reconstruct the address in the pixel readout, but rather performs this operation outside the matrix. Conversely, the readout is technology-dependent, as it can be designed in an analog manner or must be at least verified with analog tools. Furthermore, the pulse width encoder occupies a significant portion of the available design space. The ability to achieve a high degree of precision in timing is a significant advantage of this architectural approach. The signal can be subjected to superposition of hits, which can then be reconstructed at the end. In this manner, there is no integration of time and the timing is contingent upon the resolution bit of the address.

The pulse width encoder is also employed in particle physics in experiments such as CMS with the circuit MALTA [44]. However, it was retained as a prototype with 1M pixels and a pitch of approximately 20 micrometers. The MOST chip [45] is based on the concept of the pulse width, which is digitized via a serial bus. The circuit in question features 225,000 pixels and a pitch of 18 micrometers.

2.2.4 Fixed/Dynamic Priority Arbiter (Asynchronous)

This proposition represents a significant contribution to the field of thesis work, initially presented by the ICube laboratory ¹. The chip is based on photon detection and has been adapted to the particle physics environment, in addition to which a digital flow optimization has been incorporated.

The underlying principle differs from those previously presented, which employ zero suppression. Rather than utilising positional data, this approach relies on the time taken to arbitrate. This allows for a significant enhancement, as there is no integration time. The time integration refers to the interval between the occurrence of the firing event and the point at which the architectural configuration is prepared to start the data acquisition. To illustrate, the priority encoder read one pixel with each clock cycle, thereby establishing the maximum time integration as the clock period.

¹ICube web site

However, this value remains unknown. The absence of time integration enables accelerated data acquisition, as the information is conveyed with greater rapidity to the output. Additionally, the capability to timestamp the pixels is afforded. The alternative proposal employs a bottom circuit that interrogates the readout. This proposal is triggered exclusively by the pixels and operates in true asynchronous fashion.

The construction of the architecture is first based on 2 to 1 time arbiters placed in a pyramid shape, see figure 2.7. This thesis work want to present different controller sizes to make different pyramid topologies with the same number of pixels, for example one big controller of 512 to 1, see figure 2.8. The construction of the address uses the previous data correctly selected and adds one bit if the request comes from top or bottom, thus adding one bit at each level in the 2 to 1 case.

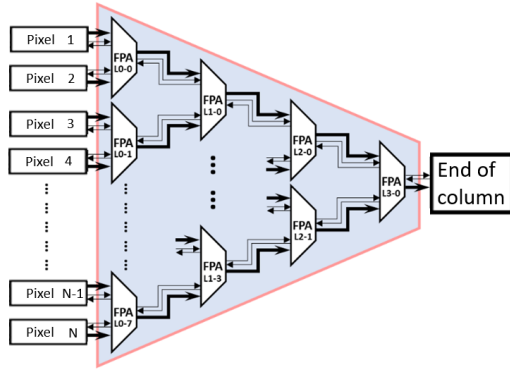


FIGURE 2.7: Topology of a tree made with 2 to 1 arbiters

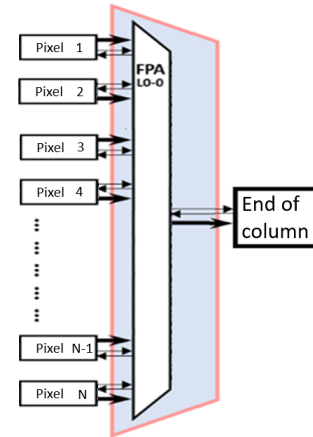


FIGURE 2.8: Topology of a tree made with one 512 to 1 arbiter

The priority arbiter is an asynchronous arbiter that is synchronized with two inputs to arbitrate figure 2.9 on the left. The arbitration is initiated when both signals arrive within a closed time window of hundreds of picoseconds. The arbitration can then be fixed and prioritize one input consistently or be dynamic and change fairly. In the figure 2.9, the arbitration is on the top (request 1). In order to implement dynamic arbitration, it is necessary to allocate additional memory to store the input that was prioritised previously. Subsequently, the aforementioned process can be cascaded to form a tree, as illustrated in the accompanying diagram figure 2.7. The request signal is generated by the upper block to request the saving of data, and the acknowledge signal is used to terminate the transmission. It should be noted that input 2 is currently awaiting treatment.

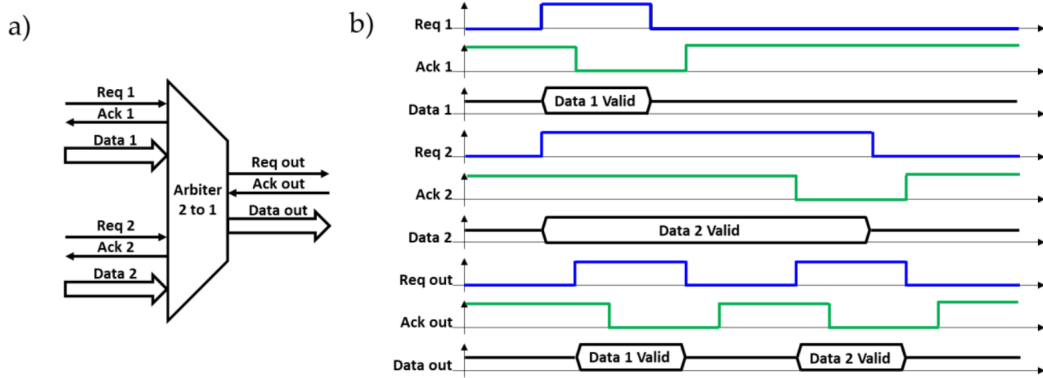


FIGURE 2.9: View of the priority arbiter topology and the chronogram in a collision (double request) [46].

The circuit is compact, does not require a global signal, and the time to cross the controller is solely dependent on the timings of the gates, allowing for a high degree of speed. The design can be modified to accommodate additional inputs for arbitration. A circuit designated as EDWARD was subjected to testing with 1,000 pixels and a large pitch of 100 micrometers. This thesis presents a circuit of this kind, which is inspired by the circuit Eyesun 2 [46] and is used for light detection to feature a few 10th pixels.

2.3 Summary

A summary of the aforementioned circuits is presented in the table below 2.1. The majority of these circuits are MAPS for particle physics. The previous chapter 1 has presented a comparative analysis of the requirements and specifications, with a view to evaluating their compatibility with the proposal set forth in this thesis.

Circuit name	Year	Matrix	Pitch (μm)	Technology	Readout	Experiment
MIMOSA 1	1999	16k	20	AMS 0.6	Analog	
MIMOSA 6	2002	4k	28	AMIS 0.35	RS	
FEI3	2003	2.9k	50x400	0.25	TR *	ATLAS
Timepix	2006	65k	55	0.25	TR *	ATLAS
MIMOSA 26	2008-09	660k	18.4	AMS 0.35	RS *	EUDET
MIMOSA 28	2011	890k	20.7	AMS 0.35	RS *	STAR
FEI4	2011	26.9k	50x250	0.13	TR *	ATLAS
Timepix 3	2013	65k	55	0.13	TR *	ATLAS
ALPIDE	2016	500k	28	TOWER 0.18	PE *	ALICE
MALTA v1	2016	1M	25	TOWER 0.18	PW *	ATLAS
MIMOSIS 1	2019	500k	30	0.18	PE *	CBM
MONOPIX	2019	100k	36x40	TOWER 0.18	TR *	ATLAS
MONOPIX	2019	5k	250x50	LFoundry 0.15	TR *	ATLAS
MALTA v2	2020	1M	18.3	TOWER 0.18	PW *	ATLAS
EDWARD	2021	1k	100	65nm	Async *	
Eyesun 2	2022	1k	SPAD	STM 28nm	Async *	
MIMOSIS 2	2023	500k	30x26	0.18	PE *	CBM
MOSS	2023	6.72M	18/22.5	TOWER 65n	PE *	ALICE
MOST	2023	225k	18	TOWER 65n	PW *	ALICE

TABLE 2.1: Summary of MAPS sensors and some readout ASIC (* use zero suppression) and their readout matrix architecture: PE=Priority encoder, PW=Pulse width, TR=Token ring, RS=Rolling shutter

The figure 2.10 shows the different readouts with their respective performance ranges in terms of speed to read fire pixel addresses, power dissipated and pixel pitch (or area required for implementation). Indeed, the design of an architectural structure is inherently a compromise between the three aforementioned constraints. The design may therefore prioritize one constraint over the others, or a combination thereof. In this manner, one architecture is not represented by a single position but rather by an interval corresponding to the possible range the architecture can reach. It should be noted that a readout can be made with all constraint values at their lowest possible levels, and the presented sensors are represented with a mark inside the interval.

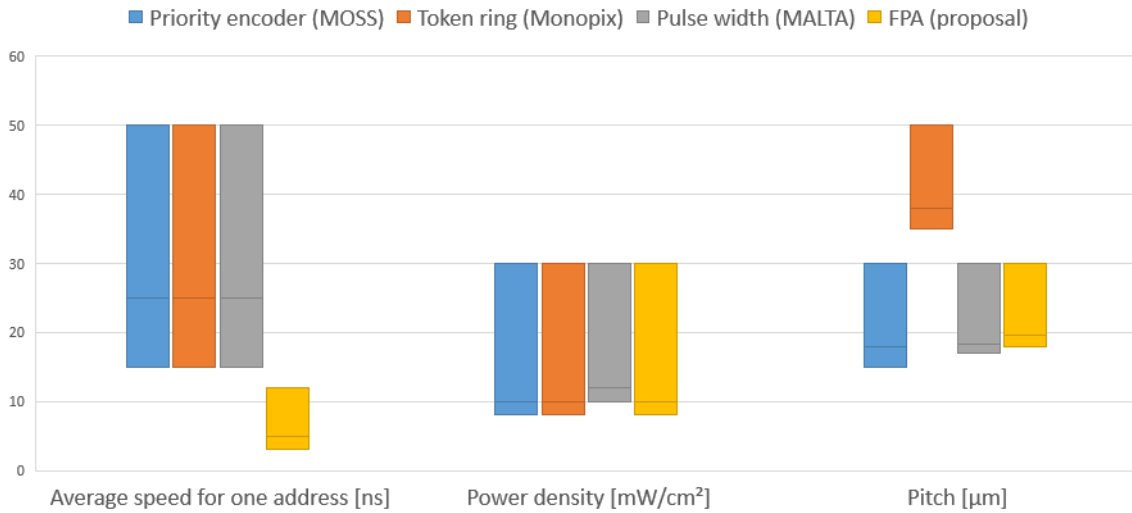


FIGURE 2.10: Comparison of the different types of readout on their possible range.

The most prevalent architectural paradigm currently employed sensors designed at IPHC laboratory is the priority encoder, which is pseudo-asynchronous in nature. Synchronous or asynchronous circuits are based on the presence of a global clock, which is related to registers. The priority encoder in the column did not utilize any registers, making it challenging to draw any definitive conclusions. Nevertheless, the data at the output is read synchronously with a bus of registers, and some signals are synchronously driven in the column. The next chapter 3 will show a true asynchronous design to brought a new type of readout that is really data driven and is expected to save power consumption, space as it is more locally wired and increase the reading speed.

Other asynchronous architectures exist as applied to photonics [47] [48] or outside matrix hybrid detector [4] but they are not compatible with the radiation tolerance or the material budget expected for the targeted experiments and thus are not presented in this chapter.

This thesis proposes an architecture that is truly asynchronous as it presents a lot a advantages and maybe can achieve better requirements because it truly follow the principle of zero suppression. It contains registers that are controlled locally by asynchronous controllers and allows independence inside the tree. The FPA (Fixed Priority Arbiter) is a two-to-one arbiter controller that manages the priority in time for the two inputs. In the event of quasi-simultaneous firing, the priority is fixed arbitrarily on the upper side. The next chapter 3 will explain how to develop an asynchronous circuit.

Résumé

Le chapitre 2 de la thèse présente un panorama des différentes technologies de lecture matricielle utilisées dans les capteurs CMOS dédiés à la physique des particules. L'objectif principal est d'identifier les solutions les plus efficaces pour la lecture des pixels, tout en tenant compte des contraintes spécifiques déjà abordées dans le chapitre 1. Les capteurs MAPS (Monolithic Active Pixel Sensors), qui nécessitent une lecture par conception, ont vu leur technologie évoluer au fil des années. Le chapitre se concentre notamment sur le développement d'une méthode de lecture asynchrone, motivée par le faible taux d'occupation des impacts (moins de quelques pour cent) observé au fil du temps. Cette section conclut par un résumé des capteurs développés pour la physique des particules dans divers contextes expérimentaux, principalement pour les MAPS, illustrant les spécifications requises et l'évolution de ces capteurs.

Architectures de lecture des capteurs de pixels

Diverses architectures de lecture ont été développées pour les capteurs de pixels en physique des particules, principalement pour des capteurs de type CMOS matriciel. Les architectures sont présentées dans l'ordre chronologique de leur développement, avec une attention particulière portée aux innovations qui ont permis d'améliorer les performances des systèmes de détection.

Lecture analogique

La première architecture de lecture développée ne comportait pas de numérisation des signaux, ce qui la rendait relativement simple mais limitée en termes de fonctionnalités et de performances.

Lecture avec discrimination intégrée

Avec l'avènement de nouvelles technologies, des architectures intégrant une discrimination à l'intérieur des pixels ont été développées. Cela a permis une meilleure gestion des signaux, notamment en termes de réduction du bruit et de précision de la détection.

Lecture avec suppression des zéros

Une innovation majeure a été l'introduction de la suppression des zéros. Cette méthode consiste à ne lire que les pixels ayant détecté un événement (les pixels activés), réduisant ainsi la bande passante nécessaire pour la transmission des données. Ce type de lecture est qualifié de "lecture pilotée par les données" (data-driven readout).

Conception classique des capteurs MAPS

Pour comprendre le fonctionnement d'un capteur MAPS, il est crucial de saisir les principes de base de son architecture. Un capteur MAPS est essentiellement constitué d'une diode reliée à un amplificateur, un circuit pulsé qui crée les courants de polarisation, et un circuit de discrimination qui définit le seuil d'activation des pixels. Tous ces composants sont intégrés dans la structure du pixel et sont conçus de manière analogique.

Le schéma classique d'une architecture de capteur MAPS se compose généralement d'une structure à double colonne, où chaque colonne est associée à des pixels de chaque côté, schéma 2.1. Une logique de contrôle est souvent utilisée pour masquer certains pixels ou pour tester les fonctionnalités du capteur. En périphérie, un circuit de contrôle gère l'ensemble des fonctions du capteur, y compris le contrôle

lent, le stockage des données, la communication de sortie, et parfois les opérations de lecture.

Évolution des architectures de lecture

La première méthode de lecture développée, appelée "lecture par pixels" ou "lecture par colonnes", reposait sur un mécanisme de type "rolling shutter". Ce type de lecture, bien que simple, s'est avéré inefficace avec l'augmentation de la complexité des capteurs et l'exigence de performances accrues.

Lecture asynchrone

Face aux limites des architectures de lecture précédentes, la thèse propose le développement d'une méthode de lecture asynchrone. Ce type de lecture est motivé par le faible taux d'occupation des impacts dans les capteurs, voulant optimiser une lecture synchrone de chaque pixel. La lecture asynchrone permet une gestion plus efficace des ressources, notamment en termes de consommation d'énergie et de traitement des données, tout en maintenant une haute réactivité aux événements de détection.

Conclusion

Le chapitre 2 de la thèse offre une analyse approfondie des différentes architectures de lecture développées pour les capteurs de pixels en physique des particules. Il met en lumière l'évolution des technologies de lecture, depuis les architectures analogiques simples jusqu'aux systèmes de lecture asynchrone avancés. Cette évolution est motivée par la nécessité de répondre aux contraintes strictes des expériences de physique des particules, tout en maximisant l'efficacité et la précision de la détection. La méthodologie développée dans cette thèse, notamment l'introduction de la lecture asynchrone, constitue une avancée significative dans le domaine, promettant une amélioration des performances des capteurs MAPS dans les applications futures.

Chapter 3

Asynchronous architecture design and knowledge

This chapter starts with an overview of the state of the art for asynchronous circuits, how they can be designed and tuned to fit requirements from experiments viewed in the chapter [1](#). A second section describes the approach to timing constraints, how to make a robust design to work properly, to be reusable and automated. The last section details the technical choices made to implement the proposed readout architecture, which performance are later studied in this thesis.

3.1 State of the art in asynchronous circuits

This section explains the various types of asynchronous circuits and their trade-offs. It also presents the main component of an asynchronous circuit, the synchronizer gate. A special gate must be used in order to create the "rendez-vous" function.

3.1.1 Types of asynchronous circuits

Asynchronous circuits may be classified according to their robustness to timings, quasi-delay insensitivity or bundled data. The circuit's robustness is demonstrated by its ability to successfully transmit and save all the data without any loss.

The quasi-delay-insensitive (QDI) circuit is not significantly influenced by the delay of cells, as the timing constraints are fixed by construction with the circuit function [49]. All signals are processed in accordance with their intended trajectory, and the circuit provides an output that is a combinatorial function of its inputs. It generally uses decision gates that take two inputs and provide two outputs correctly prioritized.

The locks are equipped with dual or multi-rail encoding. In order to transmit a single bit, two wires are employed, with a three-state usage that indicates the presence or absence of a signal. This type of circuit is particularly resource-intensive due to the number of gates required.

The bundled data (BD) is based on timings and, as a result, is less robust. Consequently, it must be constrained with great care. However, it occupies less space, consumes less power and can be designed in a robust manner [50].

In the context of particle physics, the space and power consumption requirements are of paramount importance. This is why the bundle data is particularly well-suited for this application. Both circuits employ a routing mechanism to facilitate data transfer between controllers, a process that is further elucidated in the subsequent subsection.

3.1.2 Two or four phases protocol

The fundamental structure of an asynchronous circuit is comprised of two principal phases: a request phase and an acknowledgment phase. During the request phase, a data save is requested, and the acknowledgment phase is initiated by the saver, which then responds with an acknowledgment to terminate the transaction.

The routing can be conducted in a two- or four-phase protocol. Figure 3.1 illustrates the two potential scenarios. The phases are contingent upon the design of a function in the controller that is intended to await a particular outcome. Consequently, the two-phase approach involves a wait of twice the duration, while the four-phase approach necessitates a wait of four times the duration. In the two-phase approach, the request to send an acknowledgment is only permitted once the function has been completed. Furthermore, the four-phase approach also involves waiting for the acknowledgment in order to reset the request by function. The two

phases operate under the assumption that the delay associated with the request and acknowledgment routing is fixed and known.

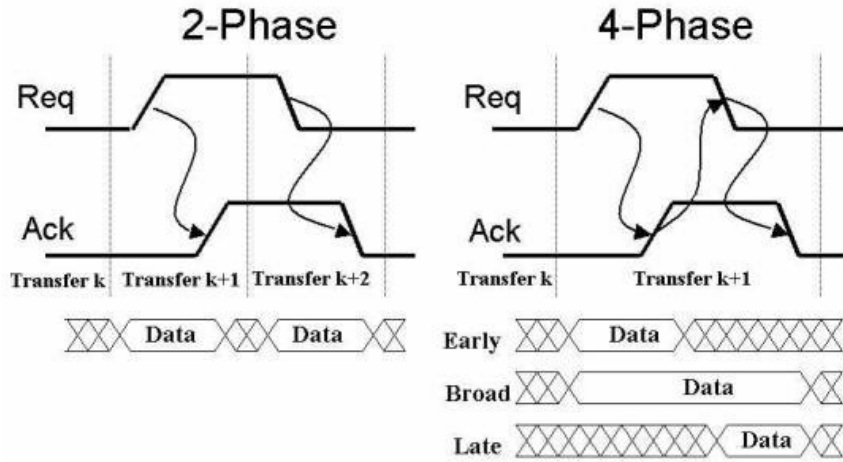


FIGURE 3.1: The two- and four-phases routine [51].

In particle physics, the arbiters are 2 to 1 controllers, which can remain pending for an undetermined time rather than classical asynchronous units. Consequently, the two-phase approach cannot be employed as data can be lost. It should be noted that the four phases require more complex logic, which in turn necessitates more space and power. The waiting by function is done with a special cell that is described in the next subsection.

3.1.3 Types of synchronization

It is necessary for both the QDI and BD circuits to have a cell that serves to synchronize the circuit at some point to respect the phases. There exist a variety of circuit types designed to synchronize asynchronous circuits internally. The Muller gate (or C-element) and the T-flop are two of the most commonly used gates for this purpose.

The Muller gate [46] may be considered a kind of a RS memory function figure 3.2. The Muller gate is also known as the "Rendez-vous" function, as both signals must be in the same state in order for the output to be switched, see the chronogram 3.2. A reset port can be incorporated to establish a starting state.

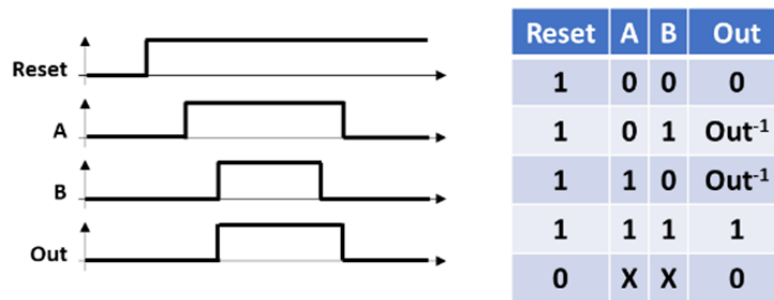


FIGURE 3.2: The Muller gate diagrams with reset [46].

The T-flop is a D-flop with the complementary output wired at D, resulting in a flop that toggles on every clock edge. The path delay of the flop can be readily adjusted through the use of a delay constraint, thereby enabling the same functionality as that of the Muller gate. The architectural design is referred to as the "click-element" [52] [53], which transmits pulses between controllers in response to combinatorial functions.

The T-flop and the Muller gate employ the same operational principle. However, the Muller gate is a fully integrated device, which, once designed, eliminates the need for additional constraints and simplifies the design. Furthermore, the T-flop requires the limitation of the minimum pulse width, as the Muller gate utilizes fixed states, which are less radiation-dependent on the wires around. The radiation hardness inside the gate will remain close as the layouts and schematics are close.

3.1.4 Know how in asynchronous CMOS design

In order to constrain the bundled data circuit in time, a number of methods exist. Two methods are presented for consideration. One is relatively straightforward to implement but may require a longer application period. The other is the reverse.

The initial approach entails establishing a maximum and minimum time frame for the dissemination of information. Consequently, the aforementioned constraints are applicable to all paths. The method is relatively straightforward to comprehend, yet it can be somewhat intricate when applied to extensive circuits. Moreover, it is PVT (Process Voltage Temperature) dependent. This necessitates an adjustment for each corner, thereby requiring five times the work for the chosen technology.

The LCS (Local Clock Set) method [50] is predicated on the premise that local clocks, when set to act in a synchronous manner, can be utilized as a means of establishing a synchronized temporal reference. This approach enables the tool to perform the timing analysis and address all paths in an automated manner. The method is complex to understand, and the placement of the clocks presents difficulties due to the sensitivity of the result to the clock start point. The LCS algorithm is not dependent on PVT, and thus, it is designed to be applicable to all corners.

The LCS method is more flexible than the previous method's min/max delay, as changing the code outside the asynchronous controller would not affect the clock's

start point or the constraints. Furthermore, this methodology is more efficacious since the delays are assessed on the actual path and not estimated based on anticipated values. The LCS method can easily be implemented in particle physics due to the repetitive nature of the pyramid shape.

3.2 Asynchronous digital Flow

This section concerns the design of a truly asynchronous circuit. The design flow employed (Cadence Layout Suite) is similar to other flows based on synchronous circuit design, as it is more robust and historically easier to design. In the present era, there exist methodologies for the design of asynchronous circuits with such tools, which are then subject to a few minor alterations to the design process. The basic flow is presented, then all modifications are presented in the Flowtool Cadence tool, which allows for some flexibility with the LCS method.

3.2.1 Classical flow steps

The Cadence common tools are Genus, Innovus, Voltus and Tempus. They all are connected to provide a flow to design digital circuits. Cadence decided to link them with another tool called Flowtool. The user can parameterize, add or remove steps of the flow. The common base of the flow is presented in the figure 3.3.

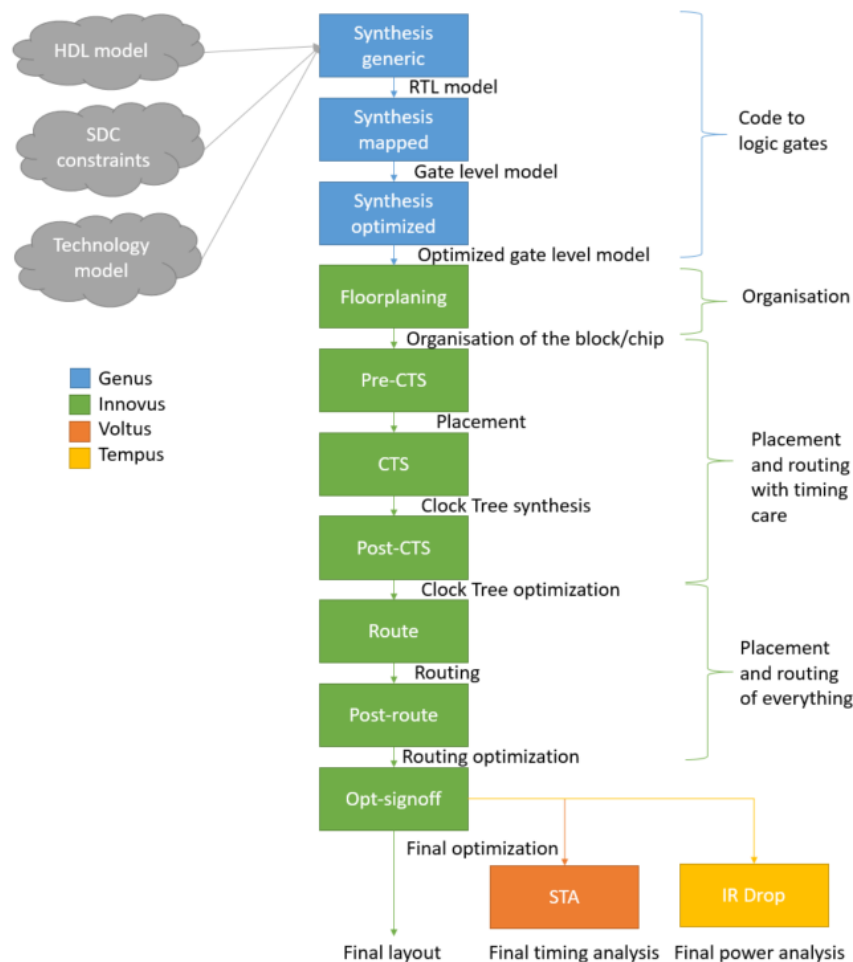


FIGURE 3.3: View of the whole classic flow.

The synchronous flow is comprised of three distinct phases: synthesis, place and route, and verification (see figure 3.5). The synthesis comprises three sub-phases: generic, mapped, and opt (optimized). The generic sub-phase takes the code to be transcribed and transforms it into a logic function. The mapped sub-phase then

takes this logic function and transforms it into a set of gates, which are subsequently optimized. The place and route phase is comprised of three sub-phases: floor planning, CTS (Clock Tree Synthesis), and routing. The CTS is responsible for placing the gates and routing the clock tree. The route sub-phase is responsible for routing all elements of the design, with the exception of the clock tree. The final phase of the process is dedicated to the verification of all parameters, including timing, drawing errors, power violations, and other potential issues. This is the stage of the process designated as the sign-off.

It is important to note that the flow takes great care at each stage of the process to ensure precise timing. From the synthesis stage to the opt-signoff phase, the timing is estimated every time more precisely. Indeed, a chip with erroneous timing will be entirely inoperable. Upon completion of the layout phase, a number of tasks are accomplished, thereby enabling the circuit to be simulated in an appropriate manner. During the synthesis phase, the circuit lacks knowledge of the utilized gates, preventing the application of a timing model. Instead, an average gate estimate is employed to fulfill the intended functionality. At the CTS step, the circuit is not fully routed, and thus, the net loads are unknown. Consequently, timing is extrapolated from the gate distance. In conclusion, as the circuit progresses through each phase, precision in timing is incrementally incorporated, prompting the tool to anticipate potential degradation within the user-defined margins on timings [54].

3.2.2 Timings inside the flow

This sub section present the differences of timing from a synchronous to an asynchronous circuit and how the tool calculates the timings. The figure 3.4 presents the two aspects. It exist two types of logic gates, combinatorial and sequential. The difference is that a sequential gate with the same input vector can provide a different output vector values. In other words, a memory can store data and a logic function will always output the same result. From the construction of the memories, there is an unstable phase due to an inside loop that makes the memory state. This transition step is critical when storing data and must be controlled with great care. It force the user to prepare the data a bit before the clock edges (the edge is the saving signal) and to keep the data valid a bit after the clock edge. Both timings are respectively known as setup and hold time.

In a synchronous circuit, the clock is the same for all memories and there are combinatorial gates between the data sender (launch register) and the data receiver (capture register). In asynchronous, the controller is made with combinatorial logic and then the clock net is not the same for all memories figure 3.4.

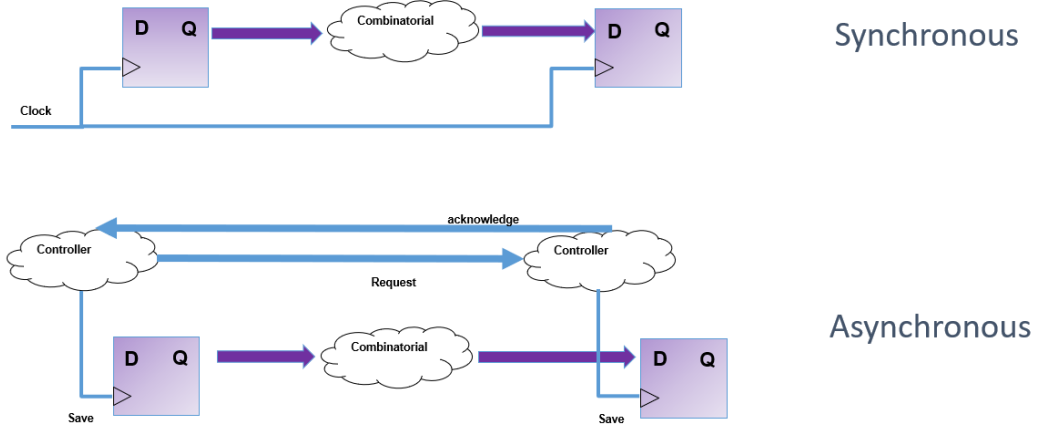


FIGURE 3.4: Schematic time flow comparison of synchronous and asynchronous logics.

For a synchronous circuit, the clock is computed with a slack time which is the time margin and must be positive. For the setup time, it is the next formula:

$$\begin{aligned}
 t_{required} &= t_{period} - t_{setup} \\
 t_{arrival} &= t_{launch\ register} + t_{combinatorial} \\
 t_{slack} &= t_{required} - t_{arrival} > 0
 \end{aligned} \tag{3.1}$$

For the hold time, the period is not incident to the delay, only the time of the combinatorial logic counts as we maintain the clock a bit after the edge:

$$\begin{aligned}
 t_{required} &= t_{hold} \\
 t_{arrival} &= t_{launch\ register} + t_{combinatorial} \\
 t_{slack} &= t_{arrival} - t_{required} > 0
 \end{aligned} \tag{3.2}$$

This means that if the data path is too short compared to the hold time, there is a violation (hold) and if the data path is too long compared to the clock period (setup), there is also a violation. For an asynchronous circuit, there is a point of divergence where the request and the save of the launch register separate inside the controller. The upper formulas are still valid equation 3.1 and equation 3.2 but the delays of the gates of the controllers must be added. The second point is that the time period is 0, as the request and the launch are done at the same time asynchronously equation 3.3 and equation 3.4. Note that, it can be possible with a synchronous circuit as the setup time will be violated.

$$\begin{aligned}
 t_{required} &= t_{request} - t_{setup} \\
 t_{arrival} &= t_{launch\ register} + t_{combinatorial} \\
 t_{slack} &= t_{required} - t_{arrival} > 0
 \end{aligned} \tag{3.3}$$

$$\begin{aligned}
t_{required} &= t_{hold} \\
t_{arrival} &= t_{acknowledge} + t_{launch\ register} + t_{combinatorial} \\
t_{slack} &= t_{arrival} - t_{required} > 0
\end{aligned} \tag{3.4}$$

The equation equation 3.3 shows that the request time must be greater than the data path. So slowing that controller if the data takes too much time. The equation of the hold equation 3.4 presents a path that will not be violated. In fact, in synchronous, sometimes the t_{hold} and $t_{launch\ register}$ are close but do not satisfy the slack and the data path can be a wire so 0 timing. In asynchronous, the delay of the acknowledge must be taken and thus add at the minimum the gate delay of the Muller gate, so it is hardly violated.

3.2.3 Flow modifications

The asynchronous design is quite tricky as there is no global clock but small ones and local [55]. As previously seen, the tool places only the clock on the CTS phases, but more than half the design is a clock when dealing with asynchronous. In fact, it is not a single net, but the whole controller that acts as a clock with the request/acknowledge routine. The second point is that from the synthesis generic to the pre-CTS steps the clock is considerate as ideal. This means that the net is not adding delay, all repeater timings are hidden (for the synchronous design) and only the timing of the data path is taken. The asynchronous timing can only be positive when taking in account the $t_{request}$ and $t_{acknowledge}$, which are set to 0 in the early phases. This flow works for the synchronous as the clock period is known and then just the data path has to be estimated to fit between two clock edges. As seen in the asynchronous formulas, the clock period is not present equation 3.3 and then the clock path has to be estimated. By construction, the tools does not allow to estimate the clock path before the CTS. The first modification is to ensure that the tools will not create bad optimizations as it correct invalid timings checks before the CTS phase.

The second aspect is how to constrain the delays. In fact, in a synchronous design, the clock comes from outside, on a port and work for the whole circuit but in asynchronous it is distributed from inside. There exists two methods the constrain an asynchronous circuit, the min/max delays and the LCS (Local Clock Set). The min/max delays fix minimum and maximum delays on combinatorial logic to ensure positive slacks. The user has to know the setup/hold time a the given register and the request/acknowledge time to fix the remaining margins for the data. This method is quite simple but can be fastidious as it is PVT depends, the timings change with the temperature, voltages and the work has to be done more than once. The LCS creates clocks locally and lets the tools compute the formula itself. This method is complex because the user has to take great care of the data path fund be the tool and ensure that the clock is in fact local and will not access to other registers.

When dealing with trackers, the number of pixels can be huge (see the sensible area, in table 1.1). The aim of this thesis is to provides a reusable flow, automated and agile. The LCS method allows to constrains the data regardless of the logic around it. Also, the number of pixels and the pyramid shape allow the user to script the LCS method for particle physics. An adaptation of the flow is proposed with a

script generator on the timing constraints figure 3.5 that creates a database of SDC files (constrains timing).

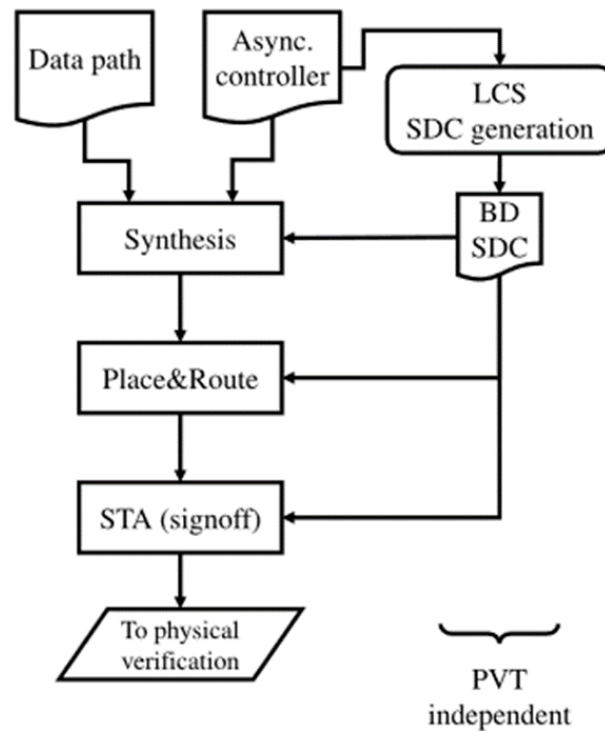


FIGURE 3.5: View of the whole asynchronous flow [50].

In the case of the asynchronous flow, the synthesis is not significantly different, but a gate must be added (C-element or Muller gate), which serves as a "Rendez-vous" function and is rarely present in the foundry gates set. In some instances, it is possible to instantiate and mark a gate as "don't touch" in order to prevent certain optimizations that might otherwise be performed due to the specific form of the asynchronous controller. Furthermore, they are given a fixed name in order to facilitate their identification within the SDC files. The SDC files are TCL scripts for timing constraints. It allows the tool to know where are the clocks, the waveform and input/output loads.

The CTS in an asynchronous flow is markedly distinct from other designs due to its lack of clock tree synthesis. Instead, it performs only the placement of gates (pre-CTS state). The circuit's timing will not be optimized. In fact, at the placement state, the circuit is not in clock propagation mode and will correct timing violations that are not as accurate as those of the clock tree (which lacks timing information). In order to prevent the occurrence of unreliable skew optimization, no CCOpt command (CTS state) is performed. The skew is an optimization in the tool to align closely every register clock. As the circuit is asynchronous, it does not need to align in time a controller to another. This optimization has to be removed. A further step is added (post-CTS) to optimize the setup time prior to the hold analysis with an opt_design command. In fact, as the optimization is not done inside the pre-CTS and the CTS, no setup optimization has been done. The tools need first a setup optimization and

second a hold optimization to ensure good convergence.

In addition, certain settings must be adjusted to achieve the desired skew target (no skew present) and CPPR (clock pessimism propagation removal) values. The CPPR is an algorithm to compensate the pessimism of the timing estimation for common paths inside controllers. In fact, the tool can add pessimism on the clock path regarding the data path for the same gate to ensure good margin. As the path is common on both sometimes, this pessimism is not relevant and must be compensated.

The tool can detect clock gating functions on the clock path due to the presence of logic gates. It is not relevant as it is the controller logic and not proper clock gating. It must be disabled and the analysis must be conducted in AOCV mode. The AOCV mode is a mode of analysis for the delay computation. It is named Advanced On-Chip Variability and allows the tool to make a better estimation of the delay for the real circuit depending on the corners (PVT).

The route and sign-off sub-phase are identical to those of a synchronous flow, with the exception of certain constraints that must be enforced due to the tool change from Innovus to Tempus. In fact, the Innovus tool is used to place and route the design and the Tempus is used to provide a better estimation of the timings through the STA (Static Timing Analysis). The Tempus tools re-compute all timing and then need to be set for asynchronous with the clock propagation.

3.2.4 Timing constraints

The timing constraints pertain to files, which are TCL code unified for tools to describe the location of the clocks and their performance (SDC). In a synchronous circuit, a single clock is typically provided by one input. Constraints are typically defined in terms of transition time, loads, and frequency.

In an asynchronous circuit, the clock is disposed on the point of divergence of the request and the clock of the launching register. This way, the CPPR adjustment is limited and limits the interaction to other registers or false paths with the command "set_false_path". By construction, the controllers make a combinatorial loop figure 3.4. The tool cannot measure timing as it can spin forever inside and add infinite delay. To prevent it, the tool automatically cuts the loop at a decided point which is not always relevant. The user has to take care of this manually before to tools do it. The easiest way is to cut the output of the acknowledge gate on the setup analysis and the output of the request gate on the hold analysis. The command is "set_disable_timing".

Creating a clock in the tool is a point of reference, all clocks must be separated inside groups to ensure no interaction as the tool will consider a data path from a clock and a clock path from an other and lead to wrong timing analysis. The other aspect is to ensure that only the wanted register is addressed by a certain clock and this is done with the false path command. It is also important to separate different clock start points in modes. Modes are one effective approach to describing the clock is to categorize it according to its operational modes, with each mode representing a potential shared data path. It is not possible for the tool to differentiate between each

case correctly, and it may identify a path that is function-locked. In fact, a clock defined inside the clock path of another will be erased and lead to a missing clock path.

In particle physics, the pyramid shape of the readout is easy to script with a double loop and the false path can be made with wild-care using the hierarchy of the three. The modes can be separated by type of check and by layer of the pyramid (analysis modes are limited to 256 in Innovus). The common constraints can be consolidated into a single file for all modes.

Each SDC file is described in this order as follows: a disabled timing to cut the combinational loop of the controller, the clock declaration, the false path on each input and output, unconcern registers disabling, and between each clock. Subsequently, clock groups may be constructed with transition, capacitive, and load constraints. Finally, a multicycle command to 0 and the propagation of all clocks are added. The multicycle explains to the tool that the clock and the data are considered on the edge number 0, so the same one (not the edge 1 which is after one clock period in synchronous).

Finally, due to the presence of all clocks, the tool can face longer computing time as moving one gates can affect a lot of clocks and paths. The increasing time computing is effective when dealing with a bigger number of pixels and is most focused on the synthesis part. One can face a computation time ten times higher in asynchronous.

3.3 Proposal of a structure for the readout architecture

This section provides an explanation of the controller that has been selected for use in this thesis project. It demonstrates the operational principles and the implications of the design choices for particle physics.

3.3.1 Adaptation to the particles physics requirements

Firstly, an asynchronous circuit is composed of controllers with three lines: request, acknowledge, and data. In a synchronous circuit, there are only two lines: the clock and the data figure 3.4. In synchronous, it is essential that the designer ensures that the data arrives in a timely manner before the next clock edge, as the clock is periodically reset. In an asynchronous circuit, the request signal instructs the circuit to take an action and to refer to the second clock cycle (also known as the capture clock). The acknowledgment signal ensures that the falling edge of the request occurs at the appropriate time. Two distinct approaches exist for the execution of the routine, namely two-phase and four-phase. In the four-phase system, each request is contingent upon the previous one. In two phases, the first two sequences are in a state of waiting, while the last two are in a state of delay. The two-phase approach utilizes fewer gates and is therefore potentially faster, but if the next block is pending for an uncertain period of time, data may be lost. In particle physics, the loss of data is acceptable at a ratio of one in a thousand. Consequently, the two-phase protocol is unsuitable, given that data accumulation is a frequent occurrence.

The blocking of a circuit by function necessitates the presence of a memory mechanism. The controller is in a state of either being active or inactive. The memories are the Muller gate (or C-element), which is small and well-suited to particle physics applications where the area used is critical. The gate has been designed at the transistor level and has been characterized to be used as a numeric standard cell figure 5.3.

The presented structure is a fixed-priority arbiter. The priority is fixed and determined by the request wired on the D port of the selection latch. A design with dynamic arbitration exists, but it is more space-consuming and therefore unsuitable for our application.

3.3.2 Working principle

The FPA (Fixed Priority Arbiter) is a time arbiter with a fixed priority in the case where both inputs arrive in a closed window. This window is relatively small and mostly depends on the locking system of the priority. In the proposal case, it is the setup time of the selection flop, so about 20ps in the given technology. This priority can be dynamic in DPA with another flop to store the previous input with the highest priority. The FPA is composed of combinational logic for the priority, a 1-bit memory to store the priority, a multiplexer to assign the previous data of the correct controller at the output and the Muller gate to ensure the request acknowledge phases.

Decision part

The controller is based on a circuit presented by the ICube Research Institut. The ICube's circuit employs an RS memory to store the priority. This type of memory is less expansive than a conventional D-flop, yet it is not capable of functioning on edges but rather on states. Consequently, if the priority in question undergoes a change, the states in question may also undergo a change, thereby compromising the data in question. One method for maintaining a fixed priority is through the use of D-flops, which employ edge-saving techniques. The arbiter two-to-one is composed of a first memory and an OR gate on the clock, with the arbiter itself being a small tile figure 3.6. The clock is initiated by both inputs, with only the upper request being retained. Thus, there are five cases to consider:

- the upper pixel arrived, the memory saved 1
- the downer pixel arrived, the memory saved 0
- both pixels arrived the downer first from 100 ps to 0 in advance, the memory save 1 (because the data path is shorter)
- both pixels arrived the downer first from 100 ps to ∞ in advance, the memory save 0 (the downer pixel will trigger the latch before the upper arrived)
- both pixels arrived at the upper first, the memory saved 1

The arbiter is permitted to prioritize the first pixels that fire, with the exception of the third case. In any case, when both pixels arrive, both must be treated. Therefore, the order is not a matter of concern, only for the reconstruction of the timing path. However, it seems complex to reconstruct the timing path from the addresses in order to achieve a good time resolution.

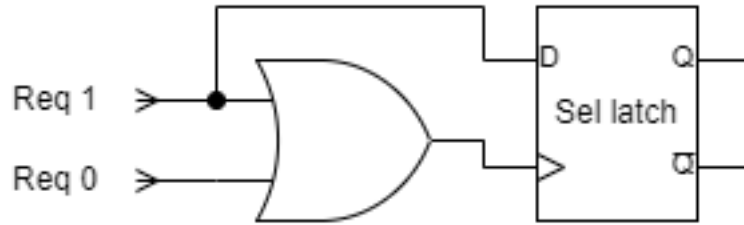


FIGURE 3.6: Arbiter pixel selection schematic.

In conclusion, if both pixels fired, the circuit does not care about the choice since both have to be treated. Once both requests have been initiated, the first one to be executed will have stored the selection in memory, preventing the second request from accessing the same data using the OR function. A more complex function must be applied using the following table. The function is given by $S = (R0|A0) \& (R0|R1) \& (R1|A1) = (R0 \& A1) | (R1 \& A0)$. The "X" represents an unknown state where both acknowledgments are received table B. This function ensures a reset phase when both requests arrive with the acknowledgment to perform the second request as with the previous OR function, the clock stays high.

Data management and synchronization part

The subsequent section comprises a multiplexer gate, which is under the control of a selection D-flop and a Muller gate, which is used to synchronize the controller figure 3.7. Synchronization is achieved by coordinating the acknowledgment of the next arbiter with a potential request from the previous arbiter. The sharing of this part may result in reduced area and power consumption, although this may be at the expense of bandwidth results presented in the next chapter 4.

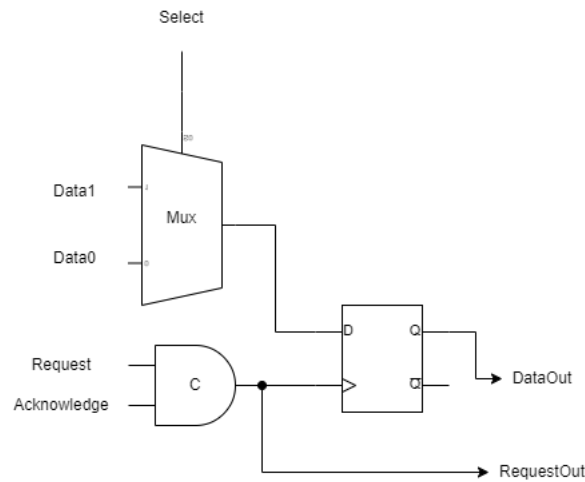


FIGURE 3.7: Multiplexer selection schematic.

Global FPA circuit

The final schematic is presented below for your perusal figure 3.8. In order to create the acknowledgments, NAND gates are added to the circuit, with the request out and the selection state.

The decision part is mandatory but the data management can be shared by decision parts. A controller of 4 to 1 can be done by using 3 times the decision part (OR gate) and adding just one data management that will be bigger. This way, the number of gates can be reduced but as there is only one Muller gate to synchronize. The controller can only treat one data from 4 inputs (instead of 2) and the bandwidth will be reduced. In the next chapter 5, the so called most shared controller is one big controller with the pixel number to one (N to 1) and the less shared controller is basically a 2 to 1 controller cascaded as shown in figure 3.9.

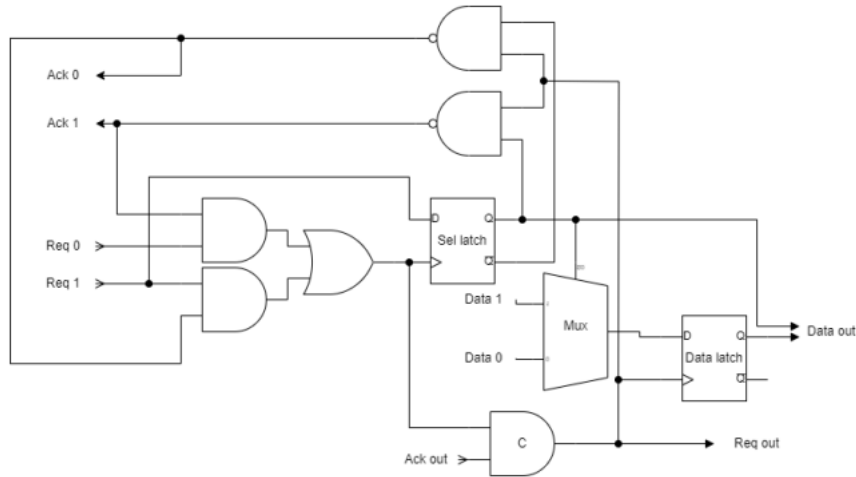


FIGURE 3.8: Structure of the fixed priority arbiter.

The combination of all arbiters results in the formation of a tree between the pixels, with a single output directed to the FIFO memory. The following diagram illustrates this process figure 3.9.

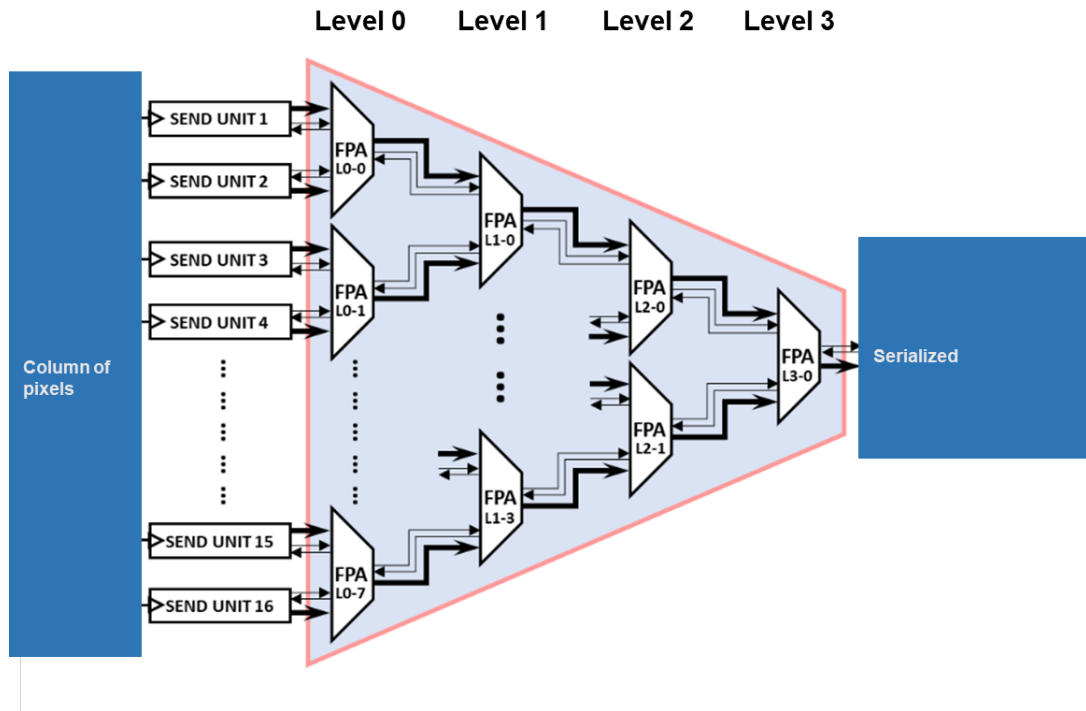


FIGURE 3.9: Functional structure of a tree composed of fixed priority arbiters (FPA).

3.3.3 Timings for the proposal

The disable timing for the setup analysis is the NAND gates on the acknowledges. The disabled timing for the hold analysis is the path inside the controller for the

Muller gate and the acknowledged input of the OR gate.

The proposal disposed of two types of register, the select and the data. The divergence point of the select register on the setup is the request 1 input (Select Mode). The divergence points for the data register on the setup analysis are the output of the OR decision on the upper (Data Select mode) or inside (Internal mode) the controller or the Muller gate output (Data mode) in the upper controller. The divergence point on the hold analysis for the select register is the OR gate (Select mode). The point of divergence on the data register for the hold analysis is the output of the Muller gate (Internal Data Select mode). All points are summarized in table [C](#).

3.4 Conclusion on asynchronous readout

This chapter presented a comprehensive overview of asynchronous behaviors. Among the various readout strategies, the asynchronous approach appears to hold considerable promise [56], offering a novel avenue for particle physics research. In light of the complexity of particle physics requirements, a decision must be made regarding the optimal balance between readout robustness and minimal size, with a focus on power consumption and speed. The optimal approach to guaranteeing the aforementioned outcomes is to implement a controller that incorporates bundled data, Muller gates and a four-phase routine design with the LCS method. The next Chapter 4 will present the simulation results of various asynchronous circuits and the characterization of each constraint present in the chapter 1 with the design flow presented in this chapter.

Résumé

Le chapitre 3 de la thèse se concentre sur la conception asynchrone et ses implications dans le contexte des capteurs de particules, en particulier ceux utilisés dans les capteurs de pixels CMOS. La conception asynchrone est explorée en profondeur, mettant en avant ses avantages potentiels par rapport aux architectures synchrones traditionnelles, notamment en termes de consommation d'énergie, de robustesse et de performances dans des environnements exigeants comme la physique des particules.

Introduction à la Conception Asynchrone

La conception asynchrone est une approche où le circuit fonctionne sans une horloge globale, ce qui le distingue des systèmes synchrones. Cette approche présente de nombreux avantages, notamment une meilleure efficacité énergétique et une flexibilité accrue dans la gestion des délais. Le chapitre commence par une introduction aux concepts fondamentaux de la conception asynchrone, expliquant comment cette méthode peut être utilisée pour répondre aux exigences strictes des expériences de physique des particules.

Arbitres et Sélection de Pixels

Une partie importante de la conception asynchrone discutée dans ce chapitre concerne l'arbitrage des signaux provenant des pixels dans un capteur. L'arbitre joue un rôle crucial dans la gestion des requêtes simultanées provenant de différents pixels, en décidant lequel sera traité en premier. Le texte décrit en détail les schémas de sélection des pixels et les fonctions logiques utilisées pour assurer un traitement efficace des signaux. Par exemple, il est expliqué que dans certains cas, l'ordre de traitement des pixels n'est pas critique tant que toutes les requêtes sont satisfaites, ce qui simplifie la conception du circuit.

Propositions pour la Structure Asynchrone

Le chapitre propose une structure détaillée pour un circuit asynchrone, illustrant comment les différents éléments comme les arbitres à priorité fixe et les portes de Muller peuvent être intégrés pour créer un système efficace. Les schémas proposés montrent comment les arbitres peuvent être organisés en arbre pour gérer les signaux provenant de multiples pixels et les acheminer vers la mémoire FIFO (First-In-First-Out) pour un traitement ultérieur. Cette section est essentielle pour comprendre les choix de conception qui permettent de minimiser la consommation d'énergie tout en maintenant une haute performance de lecture.

Synchronisation et Méthodes de Conception

La synchronisation des circuits asynchrones est un défi majeur, car l'absence d'horloge globale nécessite des techniques spécifiques pour garantir que les différentes parties du circuit fonctionnent en harmonie. Le chapitre détaille l'utilisation des contrôleurs locaux et des méthodes de synchronisation telles que la méthode LCS (Local Clock Set) pour assurer une coordination précise entre les différentes parties du circuit. Ces méthodes permettent de gérer les contraintes temporelles de manière flexible, améliorant ainsi la robustesse et la résilience du circuit face aux variations de processus et de température.

Conclusion sur la Conception Asynchrone

Le chapitre se termine par une analyse des avantages potentiels de la conception

asynchrone dans le contexte des capteurs de physique des particules. Il est souligné que bien que cette approche présente des défis, notamment en termes de complexité de conception, elle offre des perspectives prometteuses pour améliorer les performances des capteurs dans des environnements exigeants. La conclusion propose que la conception asynchrone pourrait devenir une avenue importante pour la recherche future en physique des particules, en particulier pour les applications nécessitant une lecture rapide et une faible consommation d'énergie.

Conclusion Générale du Chapitre

En résumé, le chapitre 3 explore en profondeur les concepts, les défis et les solutions liés à la conception asynchrone pour les capteurs de pixels en physique des particules. Il met en avant les avantages potentiels de cette approche par rapport aux conceptions synchrones, tout en fournissant des propositions concrètes pour l'implémentation de circuits asynchrones efficaces. Ce chapitre constitue une base solide pour les travaux futurs dans ce domaine, en proposant des méthodes innovantes pour répondre aux exigences toujours croissantes des expériences de physique des particules.

Chapter 4

Implementation and performance of the proposed asynchronous readout logic

The previous chapters showed that asynchronous architecture could allow to reach a new performance optimization for the readout of pixel matrices. This chapter presents a detailed exploration of a fully implemented design with digital design tools. The first section reviews the simulation environment that uses the described method in the previous chapter [3](#). The second section shows early results obtained by simulating the architecture that was presented at the [18th workshop of Trento \[57\]](#). The results are presented at an early stage, as they have been extracted from a CTS step of the flow. At this juncture, no routing is conducted for the data, which are instead simulated as ideal. Due to time constraints, the simulation was not extended to yield initial results regarding the direction of future investigation. The last section describes a full implementation of the circuit in a 65 nm CMOS process and presents the expected performance from detailed simulations with realistic inputs. In particular, the various implementation options are evaluated and compared in terms of area (impacting the pixel size), speed and power dissipation requirements presented in the chapter [1](#). This work has been presented with a poster at the PISA conference and in a paper [\[58\]](#).

4.1 Simulation environment

This section describes the environment of simulation used for both analyses, the early and final ones. It explains how the measurements are taken in the software and how to interpret them.

4.1.1 The TOWER 65nm technology

TPSCo, previously known as TOWER Group, is an Israeli company specializing in the manufacture of integrated circuits. TPSCo employs a range of advanced process technologies, including SiGe, BiCMOS, Silicon Photonics, SOI, mixed-signal, RFC-MOS, CMOS image sensors, non-imaging sensors, power management (BCD), and non-volatile memory (NVM), in addition to MEMS capabilities.

The proposal entails the implementation of a camera technology based on a 65nm node. The implementation of a back-biased power supply facilitates the accumulation of charges within the diode. Furthermore, it offers a comprehensive doping profile, which serves to safeguard the digital electronics from the collection process.

The node is of a thin construction, thereby enabling the incorporation of a substantial quantity of digital intelligence within the pixel. CERN has a specific affiliation with the TOWER Foundry, which facilitates the advancement of technology for use in particle physics. The technology has evolved to a 7 metals layer routing.

4.1.2 The Cadence tools

The Cadence tools have been employed for the implementation and simulation of the circuits. The use flow tool employs a conventional flow structure, augmented with additional steps for simulation. The Genus and Innovus tools are employed for the generation of the layout. At the conclusion of the aforementioned process, a file containing the cumulative timing data for all gates is generated and utilized as an input for the Xcelium tool. The Xcelium tool generates an activity file pertaining to the particle physics simulation, along with certain metrics pertaining to the timing performance. The activity file is utilized by the Voltus tool for the purpose of conducting a power analysis based on a genuine activity.

The Xcelium tool provides all the timings of all address reads during the simulation. These data can be used to determine the minimum, average, maximum and standard deviation time of reading. In addition, the number of addresses read correctly can be known. The Voltus tool can provide the average consumption and the voltage drop on the power wires.

The simulation in the early phase encompasses all aspects of a particle sensor for a single double column, including the area, consumption, and speed. The column was designed with a resolution of 512 pixels, which is approximately half the resolution of the detectors that will be used. The pixel pitch under consideration is 24 micrometers, with a rectangular analog area of 136 micrometers squared. The analog area is somewhat limited in size, but it is consistent with the actual dimensions of

particle detection and comprises a rectangle of $24 \times 5.67 \mu\text{m}^2$. All results were transferred to the post-CTS state, which did not consider data routing but solely the clock. A clock set has been designed using the LCS method to constrain the input/output buffers. The number of metal layers utilized is four (considered for the sub design in addition with 2 layers for the powering and top routing), which is a considerable increase over the number typically employed in classical designs. The simulations were conducted on three controller sizes: 512 to 1, 8 to 1, and 2 to 1.

The area is evaluated in terms of the percentage of standard cells and the percentage of GCells that are fully populated. A GCell is a software-based grid for routing that comprises five route peer layers, with each tile representing a single cell. The power is demonstrated by means of density. The timings are presented for two cases: a functional case and a realistic case. The objective of the functional test is to fire all pixels and measure the timing of all gates in the simulation. This involves determining the time required to read all pixels and the number of pixels read in 100 ns. The realistic simulation represented a particle physics application and sought to estimate the performance of the circuit. The system employs an ALICE ITS2 dictionary to define the shape of the cluster and an algorithm that incorporates random numbers to generate a pixel-fired map. The used dictionary and algorithm are presented in the next subsection. The mean time to read a pixel is extracted from the realistic simulation.

4.1.3 The particle physics simulation

An algorithm was developed to simulate the actual usage of the chip, its flow is illustrated by figure 4.1. The algorithm generates two cases: a continuous case that simulates hits arriving randomly with time and a clocked case that simulates bunch of hits produced by bunch crossings spaced by 25 ns. In order to simulate the effects of noise, a uniform distribution of single pixel firings is added to the simulation, with a rate of 10^{-7} hits per pixel per microsecond. In order to prevent any unintended consequences, the selected double column is 200 and 201 in the matrix (close to the center). The cluster is associated with an ID in a dictionary, with a probability of occurrence assigned to it. Additionally, it is linked to a shape in the pixels matrix. The clusters are generated with varying angles to encompass a broad range of scenarios in an Au-Au collision, with a pitch of 27 micrometers. The dictionary employs an ITS2 model that is not fully depleted and comprises clusters of considerable size. All data are stored in a file and the output file is a Verilog code with instructions for pixel on and wait delays.

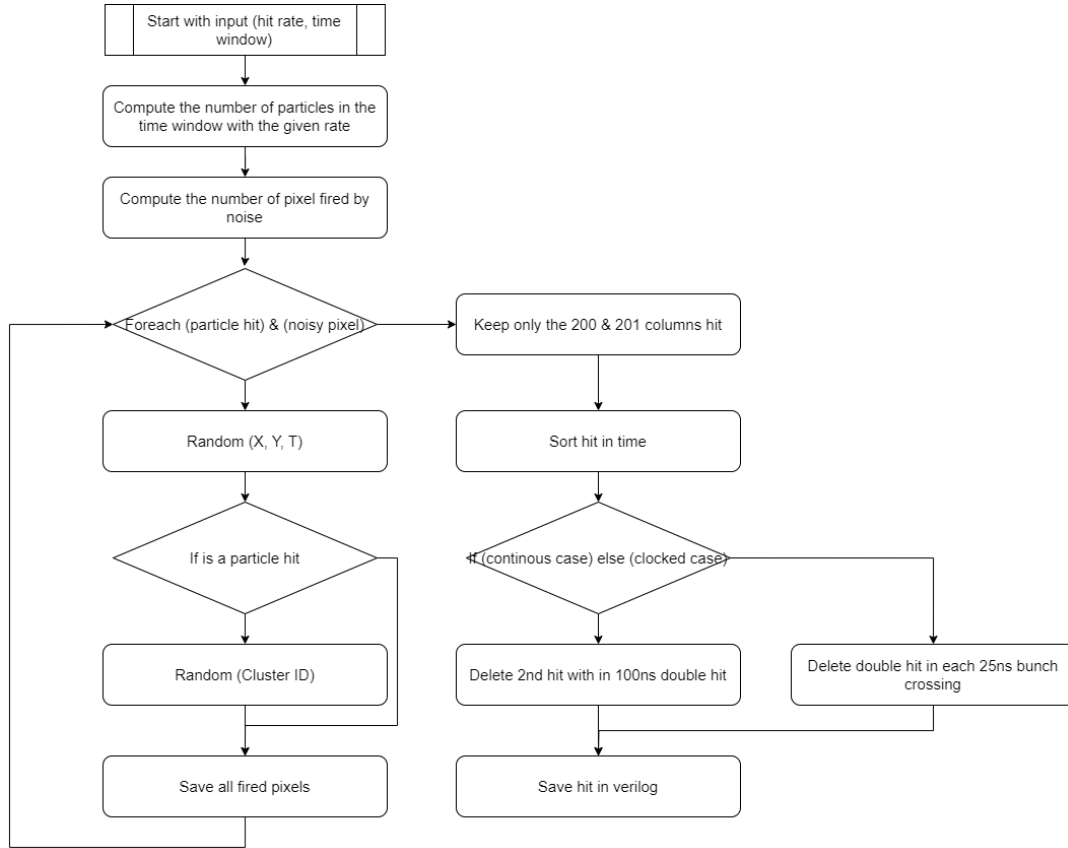


FIGURE 4.1: Algorithmic flow to generate particle hits.

Two distinct cases are presented for consideration: a continuous case and a clocked case. The simulation is identical when the clocked use employs the LHC 40 MHz bunch crossing (25 ns) timestamp, with all hits occurring in the same frame. In the continuous case, the hits were processed with a time step of 1 ns.

4.2 Early comparison

An initial comparison of the proposed asynchronous circuit to the state-of-the-art laboratory was conducted to gain an understanding of the potential and feasibility of particle physics. The aforementioned results were presented at the [18th workshop of Trento](#), where they were merely indicative of the pre-layout steps.

The percentages are extracted from the Cadence tools, it computed from the area of the cells used over the area of the block. The priority encoder data are extracted from the MOSS circuit [16] and a ratio for the same pixel numbers. The initial findings indicate that the area designated for the gates is a viable option for particle physics. It should be noted that the PE (or Priority Encoder) is better in that it lacks memory, but it is anticipated to be less efficient in time reading. The controller size is 2 to 1, 8 to 1, and 512 to 1, respectively, for 2^1 , 2^3 and 2^9 , see figure 4.2.

4.2.1 Area usage

A further simulation is conducted to ascertain whether the circuit can be accommodated within a smaller pixel, specifically a 20-micrometer pitch with an analog area of 180 micrometers squared and four metal layers. The size employed is an 8 to 1 controller size. The limitation in pitch is closely related to the shape of the circuit, which is a rectangle with dimensions equal to twice the pitch and the number of pixels times the pitch. The outcome is a lengthy rectangle, with all outputs exhibiting a reduced scale, necessitating a complex wiring configuration. The number of metal layers and the width of the small size are of critical importance. If the pitch is exceedingly small, it may prove challenging to position the gates. However, the primary concern is the number of metal layers.

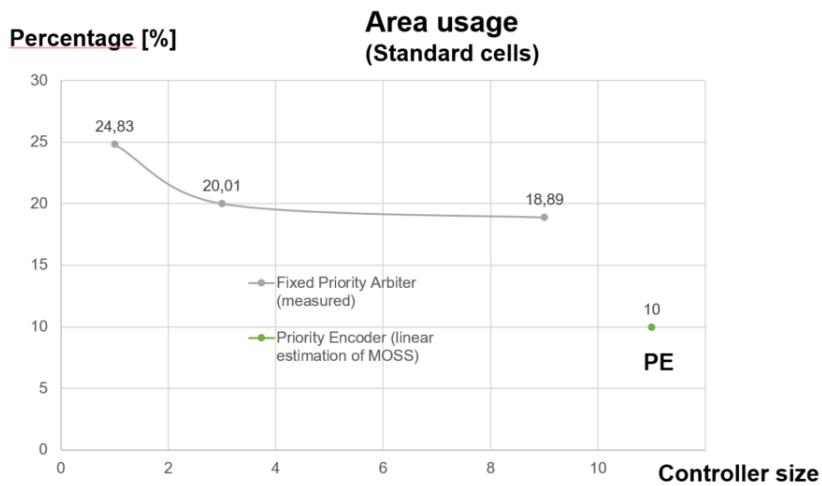


FIGURE 4.2: Evolution of the area usage with the controller size, expressed in power of 2 (3 indicates a $2^3 = 8$ to 1 controller).

The GCell is a square that can hold 5 routes with the minimum width size and spacing. The percentage is computed from the number of GCells full of 5 routes over the number of total GCell in the block area. The subsequent result is the percentage of metal layer utilized. In the 8 to 1 and 512 to 1 cases (3 and 9), the GCell is not congested; however, in the 2 to 1 case (1), 44% of the cells are full as seen in table 4.1. The

limiting parameter is more the cell density than the GCells used because the routing can be increased up to 80% instead of the cell that is preferably less than 60%. This is not a significant issue, but the pitch can be significantly reduced for this case.

Controller size	Percentage of GCells full (routing cells)
1 (9 levels of 2 to 1)	44.31%
3 (3 levels of 8 to 1)	0%
9 (1 level of 512 to 1)	0%

TABLE 4.1: Percentage of GCells used.

4.2.2 Power and timing results

The simulations encompass two tests whose results are depicted by figures 4.3, figure 4.4, the first of which is presented here and is the functional test. The goal of the test is to fire every pixel at the same time and count the time to read everything.

The priority encoder data is an estimation with a clock period of 25 ns, as used in ALICE. Therefore, the period to read all pixels at 25ns is $T = 25 \cdot 512 = 12.8 \mu\text{s}$. It should be noted that this computation does not take into account the integration time of the tree. The integration time is the time between the first pixels hit and the first edge of the clock, which may result in an overly optimistic estimation.

It is evident that the asynchronous circuit exhibits a markedly higher bandwidth. The number of pixels read in 100 nanoseconds for the priority encoder is also computed as follows: $N = 100/25 = 4$. It should be noted that the size 512 to 1 is higher than the size 8 to 1 for the time to read pixels in 100ns. This discrepancy can be attributed to the use of round numbers.

Additionally, the anticipated reduction in bandwidth has not been observed. The 2 to 1 case has a greater number of controllers that can be operated in parallel, as the 512 to 1 case has only one controller. The 512 to 1 case does not include a multiplexer component, as all components are shared. This implies that half of the register must be crossed. The controller is designed to await the selection register and then the data memory. The outcome is a size 512 to 1 configuration that completes the first pixel in half the time required by the size 2 to 1 arbiter.

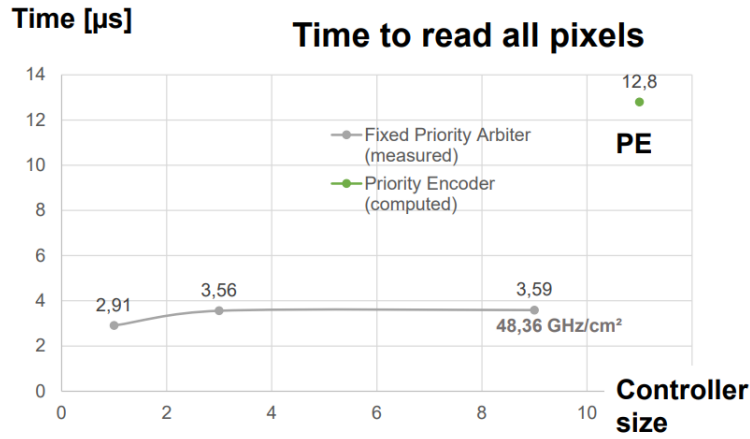


FIGURE 4.3: Evolution of the time to read all pixel with the controller size, expressed in power of 2 (3 indicates a $2^3 = 8$ to 1 controller).

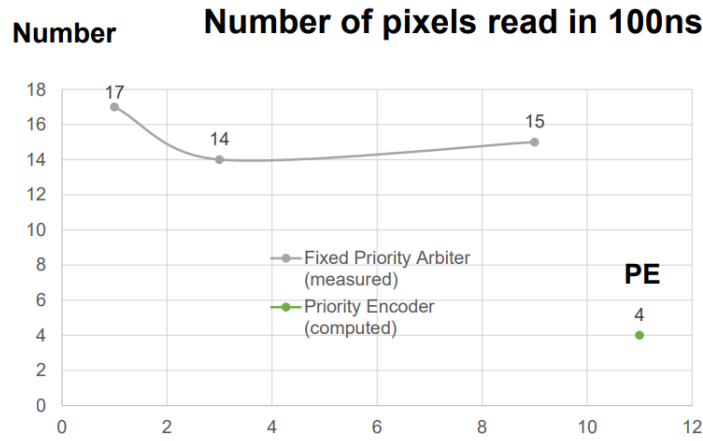


FIGURE 4.4: Evolution of the number of pixels read in 100 ns with the controller size, expressed in power of 2 (3 indicates a $2^3 = 8$ to 1 controller).

The next plots present the power consumption results and the pixel mean reading time (average over all pixels of the times to read their address) for the different configurations. Two hit rate cases are presented at 300 MHz/cm^2 and at 3 GHz/cm^2 rates. The results demonstrate a clear potential for faster processing than that of the priority encoder. It should be noted that the timing for the priority encoder is fixed, without consideration of the integration time. No data was available for the power consumption of the priority encoder, as no implementation had been carried out for a pitch of 18 micrometers and 512 pixels. Two rates are presented: one that is close to ALICE's actual uses, see figures 4.5 and a 3 GHz/cm^2 high rate to demonstrate the limits figures 4.6.

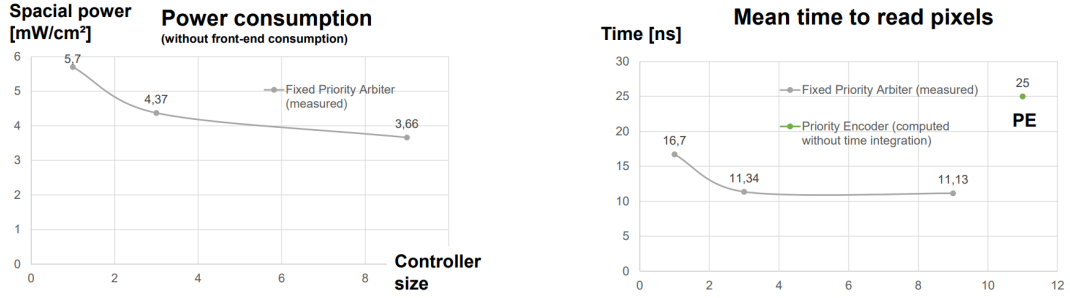


FIGURE 4.5: Evolution of (a) the power and (b) the pixel mean reading time for a hit rate of 300 MHz/cm² rates, with the controller size, expressed in power of 2 (3 indicates a $2^3 = 8$ to 1 controller).

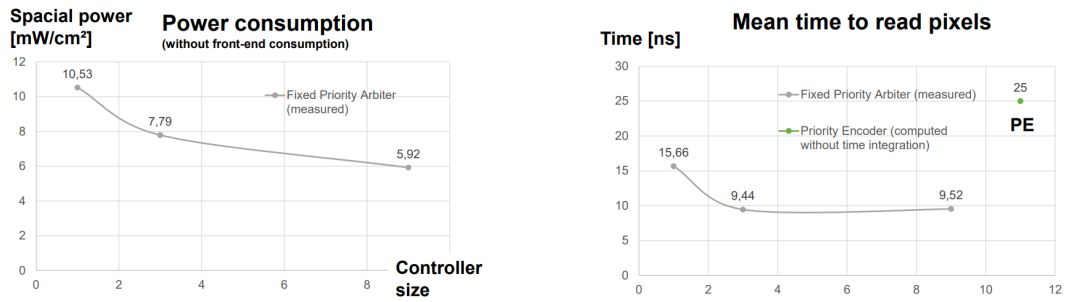


FIGURE 4.6: Evolution of (a) the power and (b) the pixel mean reading time for a hit rate of 3 GHz/cm² rates, with the controller size, expressed in power of 2 (3 indicates a $2^3 = 8$ to 1 controller).

4.2.3 Summary

In conclusion, it appears that the proposed asynchronous circuit is compatible with an 18-micrometer pitch, offering a substantial analog reserved area and exhibiting enhanced performance when utilized as a priority encoder. One aspect that requires further investigation is the comparison of power consumption. Based on the available data, it can be concluded that the circuit in question is below the acceptable limits of the requirements for the particle physics experiments, which is 10 mW/cm² and performs best in time to read even for the most shared case (512 to 1).

4.3 Exploring the limits of the architecture

This section is devoted to elucidating the constraints of the architectural framework for a standard-sized circuit with a height of approximately 1 cm and to examining the specifications that have been met. The initial discussion pertains to the variability of requirements for oversize. Subsequently, other pertinent considerations are addressed, including the implications of a minimum pitch, the necessity for radiation resistance, and the potential for integrating the sensor into a larger assembly. All results have been presented with a poster at the 16th PISA meeting and a paper in a special edition of NIM-A [58]. The previous section present early results that are extended for a definitive layout results in this section.

4.3.1 Single column block over different sizes

A variety of reasonable sizes were discussed, and three pixel pitches were ultimately retained based on two considerations: practicality (matrix composed of several pixels, which is the power of two) and a size approximately 1 cm in length for a double column. The pitch is equally spaced with 18 μm , 24 μm , and 30 μm , resulting in the following number of pixels: 1024, 512, and 512 (respectively, 9.22 mm, 6.14 mm, and 7.68 mm). The size of the arbiter is also a power of two for reasons of practicality, being 2 to 1 and MAX (1024 to 1 or 512 to 1). The parameters displayed include the standard cell density (%), metal density (%), vertical metal density (%), and wire length (cm), see table 4.2.

The objective of this work is to demonstrate the upper limits of the ratings that can be achieved with the proposed architectural configuration. A decision has been made to reduce the number of developments and sizes. Modifying the pitch of a controller is a relatively straightforward process, as it entails only a change in floor-planning scripts. However, it can potentially lead to timing convergence issues due to the necessity of additional repeaters, which alter the global circuit. When changing the size of the controller, the timing constraints files have to be rewritten and tested, which takes long run times. For reasons of time constraints, only the minimum and maximum controllers have been developed, when feasible.

As previously discussed, middle controllers are inherently a compromise between the minimum and maximum sizes if the controller size is the same everywhere. In fact, with the same size, it is only a matter of scale, and a larger controller has a shared part, so it consumes a bit less and its timing is linked to the propagation time. As we deal with asynchronous systems, there is not much in the way of repeaters, as all signals are local and thus provide a good indication of the scale applied. This section aims to anticipate the maximum ratings of the proposed architecture and subsequently identify the optimal and sub-optimal specifications. This approach will facilitate the design of the minimum and maximum controllers, while the remaining controllers will be evaluated as trade-offs between the aforementioned results to find the suited one for the targeted experiment. The results will not be linear with the controller sizes but they will be framed.

In the event of non-uniformity among the controllers, the resulting estimates will be of limited precision. However, they will not deviate significantly from the values obtained in this study. The objective of this work is to identify the optimal

optimization strategy for one application when applied to another. The design of non-uniform controllers is a challenging endeavor, primarily due to the intricate linkages introduced by non-uniform controllers, which significantly complicate the timing modes. The simulation results for the SPARC chip indicate that the non-uniform controller 16 to 1 exhibits superior performance. It can be postulated that a pyramid configuration with a wider controller on the pixels would benefit from increased power and reduced area, while a smaller controller on the output would enhance speed. The objective of this work was to explore the potential of uniform controllers and to identify avenues for further investigation in non-uniform ones.

Ctrl size	Pitch [μm]	Pixel [nb of]	Cells [%]	Metal [%]	Vertical Metal [%]	wire length [cm]
1024 to 1	18.0	1024	34.64	16.86	17.80	37.42
2 to 1	24.0	512	25.30	10.32	7.23	21.11
512 to 1	24.0	512	11.75	9.41	7.41	21.26
2 to 1	30.0	512	13.71	8.03	4.81	25.33
512 to 1	30.0	512	6.38	7.58	4.99	26.16

TABLE 4.2: Results over different sizes for density of cells, total metals and vertical metals usage

The percentage of metalization is calculated as the mean value of the total percentage of each layer from M1 to M4. This is done by designating columns as blocks. The vertical metalization value is derived exclusively from layers M2 and M4. A reduction can be observed in most cases from the total metalization to the vertical one. Moreover, the density of standard cells is greater than that of the metalization. These arguments demonstrate that the limiting parameters are the cells and not the routing, as vertical routing is not a common practice. The reduction in column width, and thus the smallest accessible pitch, is derived from vertical routing or standard cell width. In the case of the priority encoder, it is the vertical metalization that constrains the width of the columns, given the prevalence of global nets. The proposed circuit is constrained by the width of the cells, which is approximately $12.4 \mu\text{m}$ for the smallest flop. It is theoretically possible for the double column to fit within the dimensions of $12.4 \mu\text{m}$, with a reserved area of $184.5 \mu\text{m}^2$ for the analog part and a width of $11.8 \mu\text{m}$, so a pitch of $18 \mu\text{m}$.

The algorithm described in section 1 was employed to compute power activities and timing results. The selected hit rates are 1, 10, 100, 200 MHz/cm². The power measurements are presented in the figure 4.7. The energy per hit, exclusive of leakage power, is presented in the figure 4.8.

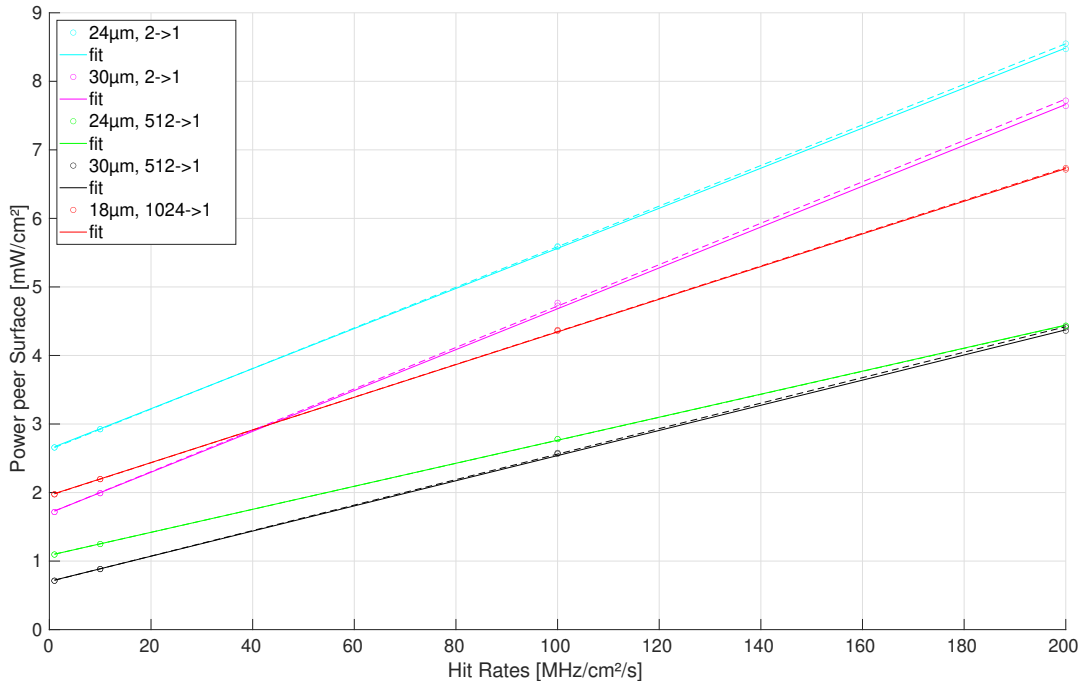


FIGURE 4.7: Power density versus the hit rate for all hit rates in the continuous case (solid lines) and clocked case (dashed lines).

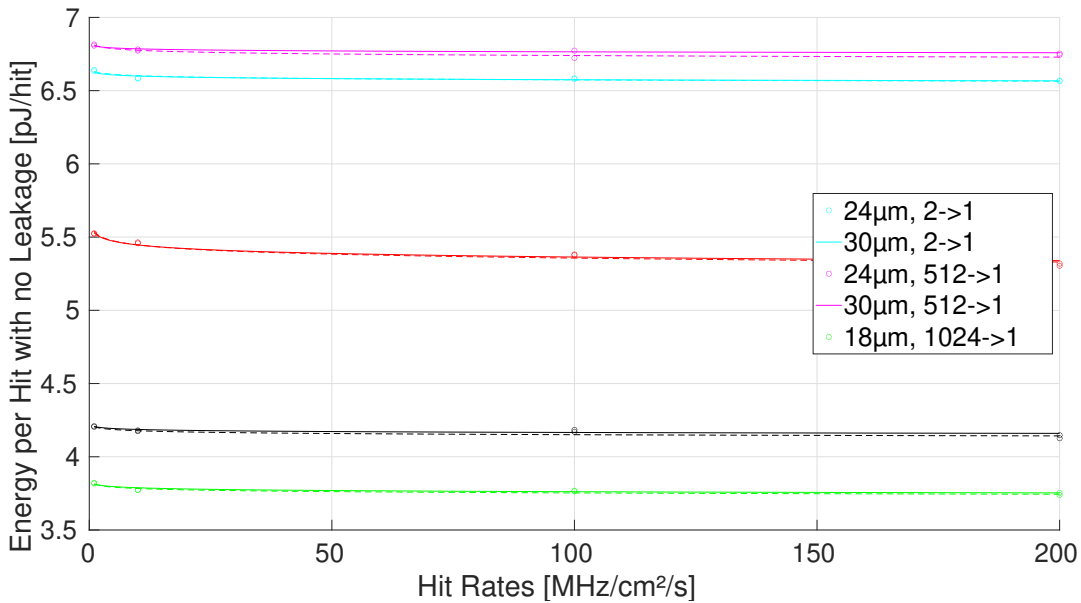


FIGURE 4.8: Energy without leakage per hit versus the hit rate in the continuous case (solid lines) and clocked case (dashed lines).

The initial plot illustrates a linear relationship between consumption and hit rates. This is to be expected for a truly asynchronous circuit, and it is reassuring to note that this circuit can be adapted to a wide range of experiments, depending on whether the hit rate or the consumption is given priority. It should be noted that the consumption did not start at zero, as leakage consumption represents a kind of offset. The two cases exhibited minimal variation in consumption at different rates. The pitch differences do not change the consumption and then confront the fact that

the repeaters are not present to buffer global signals. The second plot depicts a quasi-constant energy per hit. The observed variation may be attributed to the sharing of activity or collisions within the arbiters, which result in a redistribution of activity. It should be noted that all simulations were extracted from the typical corner, with a temperature of 27°C and a voltage of 1.2V.

The final results are presented in the form of timings, as illustrated in figure 4.9 and figure 4.10. These figures depict the data as a violin graph. The violin graph is a histogram rotated vertically and placed back-to-back with another for a more comprehensive comparison. The two graphics utilize the same data set, but the second is presented as a summed percentage.

The diagram in figure 4.9 depicts the number of pixels read over time at varying levels, including 69%, 90%, 99.9%, and the mean time. The plots were extracted from an activity at a frequency of 10 megaHz/cm². The left-hand comparison illustrates the differences between arbiter sizes, while the right-hand plot depicts the differences in pitch. It should be noted that the orange curve is identical in all instances. It can be observed that the most shared controller (referred to here as the 512 to 1 controller) reads the first pixels at a faster rate than the other controllers, as it features a smaller number of gates in the critical path. This controller exhibits a slightly higher limit at 99.9%, as it is capable of reading only one pixel at a time, in contrast to the most shared controller (aka 2 to 1), which is susceptible to pile-up in the final controller within the reading tree. The observed differences between pitches appear to be a shift in time, with the highest pitch being the slowest. This phenomenon can be explained by the necessity for a higher pitch to accommodate a larger buffer for signal reshaping, which is then followed by a delay in the critical timing path.

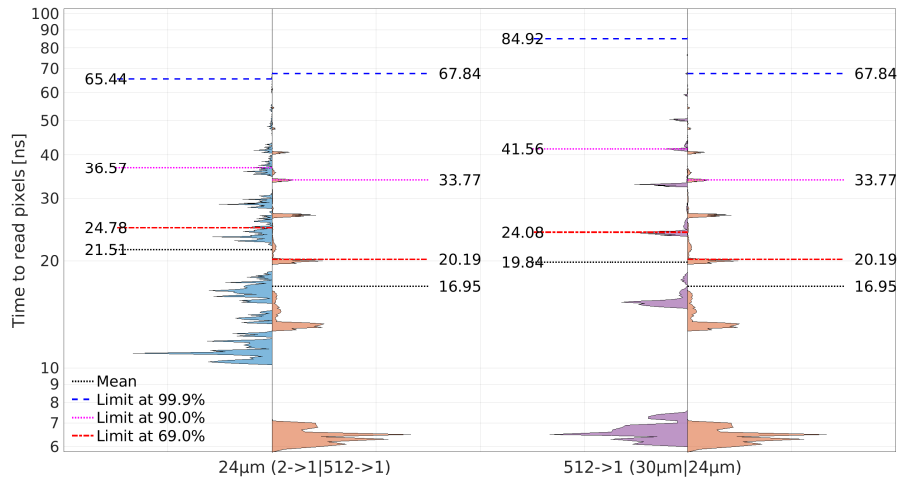


FIGURE 4.9: Violin plot comparing the distribution in time of the arrivals of the fired pixel addresses for three different combinations of pitch and controller size: 24 μm and 2 → 1 (blue), 24 μm and 512 → 1 (orange), 30 μm and 512 → 1 (violet).

The plot in reference to the figure 4.10 illustrates the percentage of pixels read over time, with the same comparisons. It is evident that the reading time exhibits

distinct increments, each corresponding to a reader address timing. For the controller that is most shared, the reading time encompasses the entire readout, with a prolonged tail extending beyond the initial pixels. The less shared controller features smaller steps, as only the last controller is involved.

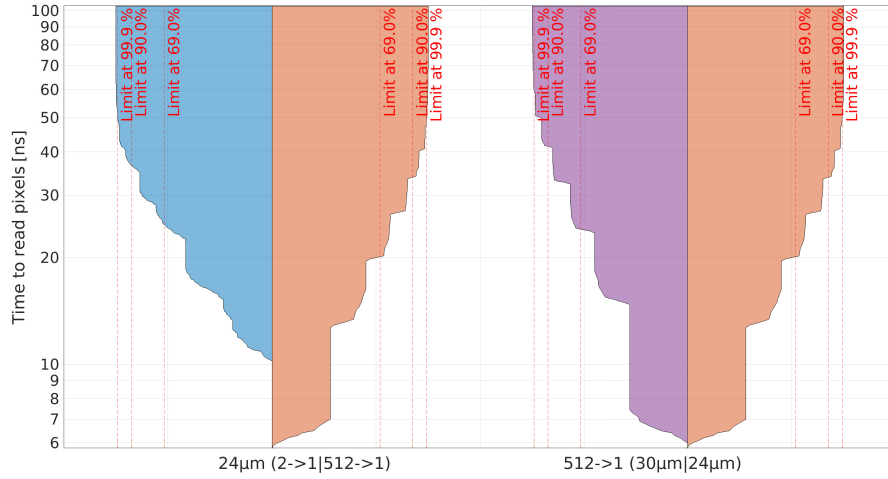


FIGURE 4.10: Violin plot comparing the cumulated distribution in time of the arrivals of the fired pixel addresses for three different combinations of pitch and controller size: 24 μm and 2 $\rightarrow$ 1 (blue), 24 μm and 512 $\rightarrow$ 1 (orange), 30 μm and 512 $\rightarrow$ 1 (violet).

Tables 4.3 and 4.4 illustrate the mean and maximum times required to read 99.9% of the pixels, respectively. The two tables present the results for the continuous case with a step of 1 nanosecond and a diode hiding of 100 nanoseconds, as well as the clocked case with a 25 nanosecond bunch crossing. The violin plots demonstrate that the mean time for the most shared controller is lower than that of the other controllers. However, the primary observation is that all controller sizes achieve 99.9% of the pixels reader in the 100 nanosecond double column. The probability of observing a double cluster in this time-lapse is essentially zero, and all clusters can be considered to be separate in time. The asynchronous logic does not require an integration time, as the pixel triggers the readout rather than a global clock. This opens a new possibility for the timestamping of all pixels at the conclusion of the double column without the necessity of incorporating an in-pixel circuit with a precision inferior to 5 nanoseconds, given the dispersion of the address over time. A reconstruction can be made with the knowledge of the address and the gate path, which allows for the estimation of the real time at which the cluster appeared.

Ctrl \ Rate	1 MHz/cm ²	10 MHz/cm ²	100 MHz/cm ²	200 MHz/cm ²
18 μ m, 1024:1	17 ns / 102 ns	26 ns / 106 ns	28 ns / 114 ns	29 ns / 114 ns
24 μ m, 2:1	20 ns / 63 ns	22 ns / 65 ns	22 ns / 65 ns	22 ns / 65 ns
24 μ m, 512:1	15 ns / 67 ns	17 ns / 67 ns	17 ns / 67 ns	17 ns / 67 ns
30 μ m, 2:1	20 ns / 60 ns	21 ns / 61 ns	21 ns / 61 ns	21 ns / 61 ns
30 μ m, 512:1	18 ns / 84 ns	20 ns / 84 ns	20 ns / 85 ns	20 ns / 84 ns

TABLE 4.3: Timing results from simulations for different hit rates and column configurations in the continuous case: mean time to read a pixel/time to read 99.9% of the pixels. (TW i.e. Time-Walk)

Ctrl \ Rate	1 MHz/cm ²	10 MHz/cm ²	100 MHz/cm ²	200 MHz/cm ²
18 μ m, 1024:1	17 ns / 102 ns	26 ns / 114 ns	29 ns / 115 ns	29 ns / N/A
24 μ m, 2:1	20 ns / 64 ns	22 ns / 65 ns	22 ns / 65 ns	22 ns / 65 ns
24 μ m, 512:1	15 ns / 63 ns	17 ns / 68 ns	17 ns / 68 ns	17 ns / N/A
30 μ m, 2:1	20 ns / 59 ns	21 ns / 60 ns	21 ns / 61 ns	21 ns / 61 ns
30 μ m, 512:1	18 ns / 84 ns	20 ns / 85 ns	20 ns / N/A	21 ns / N/A

TABLE 4.4: Timing results from simulations for different hit rates and column configurations in the clocked case: mean time to read a pixel / time to read 99.9% of the pixels.

The final table 4.5 presents the maximum achievable hit rate when a loss of 0.1% of the hits occurs. The loss of a single hit in a thousand is not significant and can be reconstructed as the cluster reappears with holes. Furthermore, it should be noted that no particle diode is capable of achieving a recovery time of 100 ns, and thus the data presented are in a worst-case scenario for the readout. Notwithstanding these considerations, in the continuous case, the circuit can achieve a rate of approximately 5 gigahertz per centimeter squared, which is considerably higher than the actual requirements of particle physics.

Ctrl \ Rate	Continuous case	Clocked case
18 μ m, 1024:1	3.5 GHz/cm ²	150 MHz/cm ²
24 μ m, 2:1	7.5 GHz/cm ²	3.5 GHz/cm ²
24 μ m, 512:1	7 GHz/cm ²	100 MHz/cm ²
30 μ m, 2:1	5.5 GHz/cm ²	2.5 GHz/cm ²
30 μ m, 512:1	3 GHz/cm ²	75 MHz/cm ²

TABLE 4.5: Maximal hit rates allowing a maximal hit loss of 1/1000th.

The results demonstrate that a double column of a reasonable size can be achieved and fit even for a smaller pitch. The devices in question exhibit a lower power consumption below the requirements of 10 mW/cm², and a fast timing readout with a possible acceptable resolution in the periphery.

A simulation for a 2048 pixel with 24 μ m is in progress to report large double column height performance. The fact that the arbiters are close to each other and not globally piloted allows no limit in the matrix size as other architectures. In fact, the priority encoder has a clock to distribute and in final the construction time of the address limits the maximum frequency. This architecture is faster and really sizable

and could be suitable for stitching.

I did not expect results far from those once presented before and that will show the size of the architecture. Some tests can be done in the future with progressive arbiter size, like 128 to 1, then a layer of 16 to 1 and finally a 2 to 1 which will accelerate the progressive readout to the congested area. The large simulation takes a lot of time due to the increasing number of clocks to manage and the number of modes. A new design flow needs to be implemented to do user pruning to select a corner to design and a DMMMC (Distributed Multi-Mode Multi-Corners) flow to verify the complete design in Innovus with more than the limit of 256 modes.

4.3.2 Simulation for very small pitch

Particle physics requires the highest possible spatial resolution. The spatial resolution is contingent upon the pitch size; thus, the smaller the pitch, the higher the spatial resolution. The actual sensing component is not capable of accommodating such a small pitch. However, recent research has identified LGAD (Low Gain Avalanche Diode) as a promising alternative, as it features no pre-amplifier. Consequently, the diode will be more compact, and the pitch can be reduced. For the purposes of this analysis, it is assumed that the readout circuit represents half of the pixels. Consequently, the pitch explored here is 11 micrometers. This implies a diode well of $5 \times 5 \mu\text{m}^2$, a discriminator, and a flop in another $5 \times 6 \mu\text{m}^2$, and the readout in a $6 \times 11 \mu\text{m}^2$, which is $84 \mu\text{m}^2$. From the aforementioned table, the most economical configuration is the "MAX," which features a cell density of 40% and a size of $7.5 \times 18 = 135 \mu\text{m}^2$. Upon calculation, the density is estimated to be approximately 85% of the standard cell density.

The chip's consumption is $5.4 \text{ mW}/\text{cm}^2$ for the continuous case and $10 \text{ mW}/\text{cm}^2$ for the clocked one in a $10 \text{ MHz}/\text{cm}^2$ hit rate. The mean reading time is 23.52 nanoseconds for the continuous case and 23.61 nanoseconds for the clocked case, which is nearly identical to the previous results. The maximum time required to read 99.9% of the hits is 110 nanoseconds. Finally, the density of the area utilized is 74.23%, while the metal density is 24.57% and 25.42% in the vertical direction. The maximum achievable hit rates are 9 gigahertz per centimeter squared in the continuous case and 5.5 gigahertz per centimeter squared in the clocked case. The pitch of $11 \mu\text{m}$ appears to yield comparable results to those observed in previous studies, which is a promising indication.

4.3.3 Radiation hardness tolerance

One of the most crucial requirements for particle sensors is their radiation hardness. Radiations have the potential to destroy or deregulate the sensor. It has a multitude of effects, including SEE (Single Event Effect) and SEU (Single Event Upset). SEE represents a gain in energy that is jailed within the sensor. This phenomenon is considerably more dependent on the technology than the schematic. SEU is a bit-flip effect, whereby some nets are susceptible to gain or loss of energy due to particle crossings. In most cases, the net that is most susceptible to radiation damage is the one that is responsible for the construction of memory. A solution to this problem

does exist and will be discussed in detail.

The protection against SEU can be achieved through the removal of the register (other schematic), the use of dynamic flops, or the triplication of the registers. It is not feasible to modify the schematic, and the dynamic flop must be capable of maintaining a maximum operational time, which is uncertain in asynchronous systems. Finally, the area does not permit to triplicate flops. In this asynchronous design, two types of memory are present: selection and data. A bit flip on the data memory will result in the corruption of only the data. A bit-flip on a selection memory can result in the circuit becoming permanently inoperable, as the acknowledgment may have been received by the incorrect controller. Furthermore, the pending request will not be reset. In this discussion, we will focus on the selection memory, as data corruption is a tolerable consequence in certain circumstances. The number of selection memories is equal to the number of pixels minus one, regardless of the size. The SEU must occur during the reading phase and on the register-sensitive node. Two cases will be examined in detail: the most common controller and the least common one.

The most shared controller is composed of one N to 1 controller on an N pixels column. First, we assume a total occupation as the synchronous behavior and compute the SEU probability for the used flop (S2CDDFFHQD1). The probability is given by the product of the SEU rate, the flux, and the bits per second: $P_{SEU} = X_{SEU} \cdot Flux \cdot Bits = (5 \cdot 10^{-14}) \cdot (10^6) \cdot 1023 = 5.11 \cdot 10^{-5}$ Hz. The probability is relatively low, but it is now possible to compute a percentage of activity for all flops. The mean time to read an address is 6 ns, indicating that regardless of occupation for all registers with a cluster of 4 pixels, the mean occupation can be computed by applying the following formula: The total time required to read all bits is given by the following equation: $T_{tot} = C \cdot T \cdot (\sum_{i=1}^R i) = 4hits \cdot 0.6ns \cdot 55bits = 132$ ns, with C the cluster size, T the mean time in a controller and R the number of registers. Each pixel fired cross the tree and needs to be maintained, so crossing the first controller is one flop sensible to SEU, crossing the second is 2 flops sensible and so on. So the formula for one hit is the sum of the integer to the output size multiply by the number of hit and the time were one flop is activated. The rate is approximately 200 megahertz per centimeter squared, with the area, the number of particles hit, being 66,355.2 hertz. The occupation is defined as the rate multiplied by the mean time to bit or $Occ = Rate \times T_{mean}/bits$. In this case, the occupation is 663552×132 ns = 8.75%. The probability of a SEU is defined as $P_{SEUact} = P_{SEU} \cdot Activity = 4.48 \cdot 10^{-6}$ Hz.

The less shared controller features a cross-section that is identical to that of the SEU, utilizing the same flops, yet with a lower level of activity. Each register will not be blocked for an extended period of time. Consequently, the total occupation time with a mean time to read of 10 ns is given by the following equation: $T_{tot} = C \cdot T \cdot R = 4 \cdot 0.6 \cdot 10 = 24$ ns with no collisions. Each pixel only activate one controller at the time, so this is the product of the size of the tree, the time and the cluster size. $T_{tot} = T \cdot (C \cdot R + (\sum_{i=C-1}^1 i)) = 0.6 \cdot (40 + 6) = 27.6$ ns. Each collision makes the other pixels wait with one less each time. For a given hit rate, the percentage of occupation is given by $Occ = 663552 \cdot 27.6ns = 1.83\%$. The actual probability of a SEU is estimated to be approximately $P_{SEUact} = 9.35 \cdot 10^{-7}$ Hz.

In both cases, the probability of experiencing an SEU event is estimated to occur

every five days. The dead time is the interval required for the system to reset, which is approximately 1 microsecond. The off-time is demonstrably not equal to $1/1000$. The flux is calculated using the worst-case expected value on ALICE ITS3, which is equal to 1 megahertz per centimeter squared.

The probability of an SEU on the Muller gates cannot be estimated, as the XSEU (Cross section of SEU) is not known. However, it may be assumed to be similar to the flop one, given the high degree of similarity between the two designs. It is also important to note that the SEU on the Muller gates will not result in the system being blocked, but rather in the loss of an event. The flop on the address storage has not been considered previously; it is also a method of losing data rather than blocking the system.

In conclusion, the most shared architecture is more susceptible to bit-flip errors, yet the error rate remains within an acceptable range.

4.3.4 Simulation with stitched sensors

The technique of stitching is a method for replicating the reticule on the wafer and performing a complete wafer circuit [59]. This method entails the use of wider wires and a very low-power design, which serves to limit voltage drops. Some design rules were established by the ALICE collaboration for the ITS3 project. The most critical aspect is the implementation of a double-spacing isolation and a double-width wire configuration, along with wider cells, in order to effectively limit the leakage and increase the yield.

The double spacing width can be estimated by doubling or tripling the metal density from the previous results. It appears that all sizes are capable of passing with a maximum density of 60% of the metal density, which is deemed acceptable. In fact, the arbiter tree is not constrained by metal routing but rather by the density of the cells. The cells have undergone a redesign to reduce leakage power consumption, resulting in an enlargement to 13 tracks instead of 12 and an increase in width to provide additional space between lower metals within the cells. The ratio is increasing by approximately 20% for all cells. The increase in density in the worst case scenario (i.e., 18 μm with the 1024 to 1 arbiter) is 42% cell density, which is deemed acceptable for the design. It is possible to stitch this circuit and maintain closer results.

4.4 Summary

The results presented in this chapter demonstrate the feasibility of designing an asynchronous circuit for a tracker constructed with MAPS and following the requirements of the chapter 1. Furthermore, the preliminary simulations of the investigation indicate that the most shared controllers are well-suited for the particle physics application. The final simulations confirms this hypothesis with encouraging results as a 10 mW/cm^2 , reading speed of 10 ns and pitch of $20 \text{ }\mu\text{m}$. The following chapter 5 presents a prototype that serves to demonstrate the feasibility of designing an asynchronous circuit in a safe manner and achieving the desired results as presented in the previous chapter 3.

Résumé

Le chapitre 4 de cette thèse, intitulé "Implementation et performances d'un circuit asynchrone", explore les perspectives et les enjeux liés à l'utilisation des circuits asynchrones dans la conception des capteurs pour la physique des particules, en particulier ceux basés sur les pixels. Ce chapitre se concentre principalement sur l'analyse des contraintes, des performances attendues, et des innovations potentielles offertes par les circuits asynchrones, en réponse aux défis posés par les capteurs de particules dans les environnements à haute énergie.

Contexte et justification

La conception des capteurs de particules pour la physique des hautes énergies, comme ceux utilisés dans les expériences au CERN, exige des solutions techniques robustes capables de gérer des flux de données extrêmement élevés et de résister à des niveaux élevés de radiation. Les capteurs doivent être non seulement rapides et précis, mais aussi économes en énergie et en espace. La technologie asynchrone est proposée comme une solution potentielle à ces défis, offrant des avantages tels que la réduction de la consommation d'énergie et l'amélioration de la fiabilité par rapport aux approches synchrones traditionnelles.

Les avantages des circuits asynchrones

Les circuits asynchrones fonctionnent sans horloge globale, ce qui permet une réduction significative de la consommation d'énergie, car les parties du circuit ne sont activées que lorsqu'elles sont nécessaires. De plus, l'absence d'horloge élimine les problèmes liés aux variations de l'horloge (comme le jitter), ce qui peut améliorer la robustesse du circuit face aux fluctuations de température et aux variations de tension, des conditions courantes dans les environnements de détection de particules.

Défis et inconvénients

Cependant, la conception de circuits asynchrones n'est pas sans défis. L'un des principaux inconvénients est la complexité accrue du design, notamment en ce qui concerne la gestion de la synchronisation entre les différentes parties du circuit. En l'absence d'une horloge commune, les concepteurs doivent s'assurer que les données circulent de manière fluide et sans conflit, ce qui nécessite une ingénierie méticuleuse. De plus, le manque de standards dans la conception asynchrone par rapport aux circuits synchrones bien établis peut représenter un frein à l'adoption généralisée de cette technologie.

Études de cas et applications spécifiques

Le chapitre analyse plusieurs études de cas et exemples d'application des circuits asynchrones dans le domaine de la physique des particules. Un exemple clé est l'utilisation de ces circuits dans les Monolithic Active Pixel Sensors (MAPS), qui bénéficient grandement de la réduction de la consommation d'énergie et de la compacité offerte par les circuits asynchrones. Les MAPS sont particulièrement adaptés aux expériences où la minimisation de l'épaisseur est essentielle pour réduire les perturbations sur les particules détectées.

Perspectives d'avenir

Le chapitre se conclut en discutant des perspectives d'avenir pour les circuits asynchrones dans le domaine de la détection de particules. Il est noté que bien que

les défis techniques restent importants, les bénéfices potentiels de cette technologie justifient des efforts de recherche continus. En particulier, l'intégration de circuits asynchrones dans les futurs détecteurs de particules pourrait ouvrir la voie à des systèmes plus efficaces et robustes, capables de répondre aux exigences toujours croissantes des expériences en physique des hautes énergies.

En résumé, ce chapitre met en lumière les avantages potentiels des circuits asynchrones pour la détection de particules tout en reconnaissant les défis associés à leur conception et à leur mise en œuvre. Les perspectives offertes par cette technologie, notamment en termes de réduction de la consommation d'énergie et d'amélioration de la robustesse, en font une avenue prometteuse pour la recherche future dans le domaine des capteurs de particules.

Chapter 5

Prototype Sensor Pixel Asynchronous Readout CMOS (SPARC)

The SPARC prototype represents a circuit that is intended to demonstrate the viability of implementing an asynchronous readout in a matrix of MAPS. The asynchronous component of the circuitry is designed to facilitate the interconnection of the pixels with the external memory, which is accessible from outside the circuit. The functional diagram of the circuit is depicted in the following diagram figure 5.1. The first section enters in the detail of the design, while the second section explores the testing strategy.

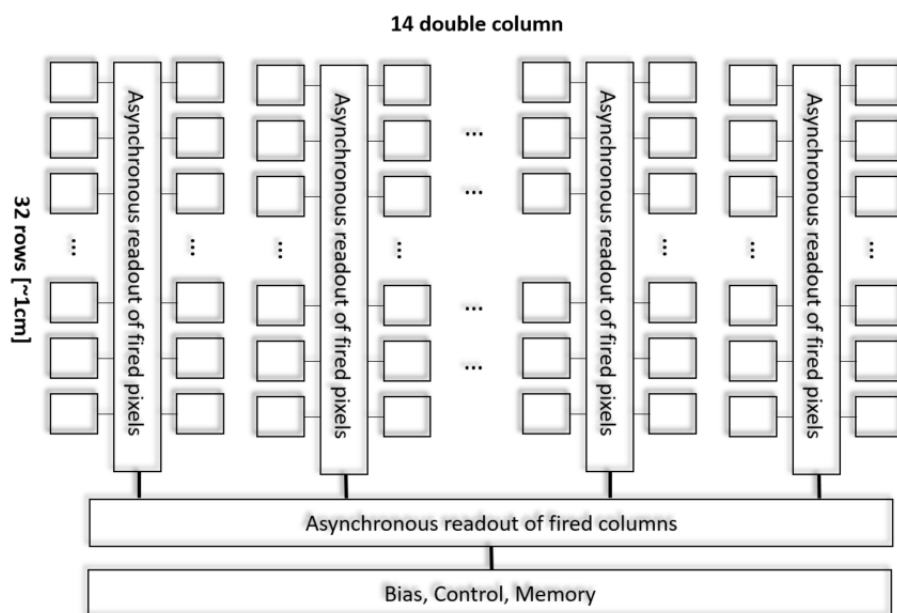


FIGURE 5.1: Diagram of the double column readout arrangement for the SPARC circuit.

5.1 Design of SPARC

This part is intended to elucidate the evolution of a compact chip designed to assess the asynchronous architecture previously discussed. It will delineate the decisions made and the enhancements incorporated for the purpose of the tests. The technology of the chip is based on a 65nm TSPCo with 12 tracks standards cells and low Vt.

5.1.1 Goals and context

The proposed asynchronous architecture has been thoroughly simulated with digital tools as described in the previous chapter. Nevertheless, the implementation in a physical sensor including pixels collecting charge is a compulsory step. Using the architecture on a large scale requires validation based on the test of such a sensor with real particles. This is the main goal of the SPARC prototype.

The objective of the SPARC circuit is to demonstrate the feasibility of implementing an asynchronous readout in a matrix. The design is exclusively created using digital tools for the asynchronous part, thereby allowing for flexibility in the design variations. A variety of controller sizes are presented for consideration within the readout, with the objective of identifying an optimal configuration. The circuit is designed to test the feasibility of timestamped hits with ToT (Time Of Threshold) and the resilience of the asynchronous circuit to SEU.

The circuit has been designed using TOWER 65nm imager technology [60], which was recently adopted by CERN and may be used for the next 10 years. The submission has been made in the ER2 of the EP-R&D WP1.2 of ALICE ITS3 [16]. The pixel has been reused from the DPTS test circuit of ER1 in order to test the architecture under test beam. The allowed space for the R&D circuits prototypes is 1.5x1.5 mm².

The ALICE collaboration and related experiments, such as FCCee, are striving to achieve greater position resolution by employing smaller pixel pitches. The IPHC laboratory is interested in pursuing smaller pixel pitches in the future, and the SPARC project aims to provide a readout with a smaller size. The tests of such structures are conducted at reticule sizes of approximately 2 cm for the matrix. The next objective of SPARC-V2, will be to demonstrate the viability of a full-scale prototype in normal conditions.

This section provides a comprehensive overview of the elements and design choices present in the SPARC prototype. The objective of this section is to provide an overview of the circuit and to explain the rationale behind the inclusion of specific blocks and their implementation.

The dimensions of the chip are 1.5 x 1.5 mm², which allows for the fabrication of small matrices with a usable space of 800 x 800 μm². The circuit is composed of a matrix of pixels, which are grouped together to form a diode, an amplifier circuit, a discriminator, a pulsing block, a digital part that links the output of the pixel to the readout (with masking and tuning), a column readout, and the line readout. In addition, the absence of digital-to-analog converters (DACs) is notable. However, current mirrors, which are directly attached to the PAD inputs, are reused from the DPTS circuit. A memory is employed to store the hits with timestamping. This

memory is a FIFO (First In First Out) double input asynchronous device, which allows no links between read and write. Serializers have been incorporated into the FIFO in order to restrict the number of PADs. Given the presence of timestamping, a digital counter has been developed with the use of digital ring oscillators. Finally, a control circuit is reused with the SPI protocol.

5.1.2 The pixels design

The front end of SPARC is identical to that of the DPTS chips. The CERN was kind enough to agree to embed the DPTS pixel, amplifier, pulsing, and biasing circuit inside SPARC. The layout is composed of a rectangle with dimensions of 8.6 micrometers by 15 micrometers. The diode has been slightly modified to accommodate a height of 16 micrometers by lengthening the bias. The resulting pitch for SPARC is $24.1 \cdot 16 \mu\text{m}$.

The DPTS specifications include a pixel of $15 \times 15 \mu\text{m}$ with 1024 pixels, a typical consumption of 120 nW, high radiation hardness (up to 10^{15} Mev/cm^2), a relatively low fake hit rate of 10/pixel/s, and an efficiency of 99% [61].

The analog portion of the pixel is derived from the DPTS circuit and slightly modified to accommodate the new metal stack nomenclature. Some nets have been omitted as they are not obligatory. The row and column signals have been linked together, as have the MD, MH, and MV signals (Mask: Direct, Horizontal, Vertical), see figure 5.2. The layout has been relocated to the minimum extent possible. The biasing circuit, previously utilized in the DPTS, has been integrated into a single block and configured in accordance with the relevant wiring specifications.

A digital circuit has been designed to perform the masking, pulsing, and ToT functions. A tie high is incorporated into the design to ensure that the pixel with the MHDV input is selected for masking, as this operation is performed within the digital block. Two memories are present within the digital pixel, one for the ToT and the other for the request and acknowledge function of the readout.

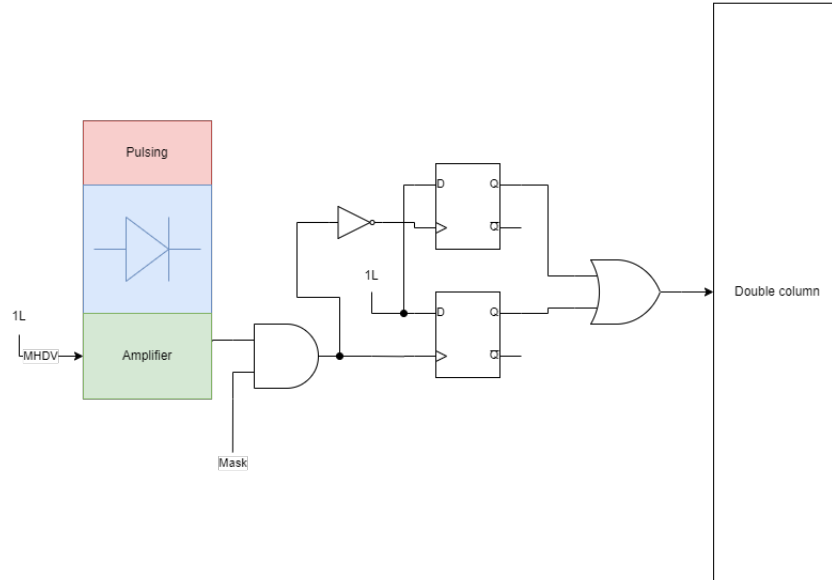


FIGURE 5.2: Schematic of pixel wiring to the double column with the digital pixel block.

5.1.3 Asynchronous gate

In addition to the circuit's development, certain components must be constructed as Muller gates figure 5.3. The "Rendez-vous" function is not included in a typical set of standard gates. Subsequently, the circuit was drawn for different output drive strengths, namely D1, D2, D4, and D8. The designation is S2CLKCGATE2RDx, which stands for S2C (a classical standard set of LVT 12T track TOWER 65nm), CLK, which denotes a transistor size that facilitates a close rise and fall time, CGATE, which is the Muller gate or C-element, 2R, which signifies the 2 reset transistor to ensure optimal leakage and drive strength. The drive strength is derived from other standard buffer cells, but the primary component is as follows.

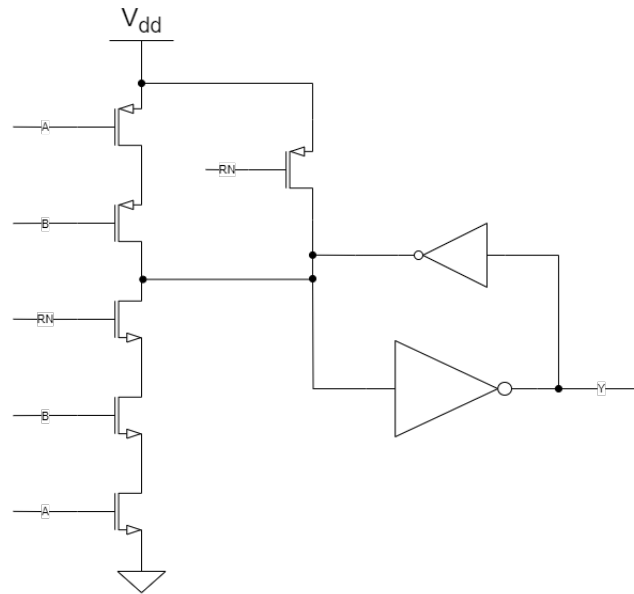


FIGURE 5.3: Schematic of the Muller gate D1 at the transistor level.

The Muller gate is set when A and B are high and reset when both are low, other states are memory ones. An asynchronous reset use. It is important to note the two reset transistors, which are connected in a bottom-to-top configuration. This configuration is used to prevent shorts. Two inverters are positioned side by side to implement the memory function.

The gates have been identified and characterized using the Liberate tool in conjunction with the assistance of Cadence support. A set of Liberty files has been generated to facilitate the use of the cell with digital tools. The liberty file is a text file that stores 2 or 3D tables of the timings of a gate to accelerate computations with a model. The following delays and leakage arcs have been identified in the appendices table [A.1](#) and table [A.2](#).

A comprehensive layout has been created for all cells with extracted analog simulation, with the objective of ensuring optimal functionality. A balance is established between the transistor sizes and is replicated from the actual cell sets. It is then anticipated that the transistor will exhibit the same behavior and compatibility with the other cells. Due to the configuration of the two inverters side by side, the gate is relatively slow (more than 500ps up to 1.8ns). This could be improved regardless of the consumption, but in the used controller, this delay helps to meet the setup timing constraint. The cell is also somewhat wide, with dimensions of 2.4 μm (12 tracks) times 3.2 μm (D1), 3.6 μm (D2), 5.2 μm (D4) and 7.4 μm (D8). It should be noted that the cell is susceptible to single-event upsets (SEUs) due to radiation exposure. However, its transistor sizes are comparable to those in the design kit, which should result in a equal degree of vulnerability than in other memory types.

5.1.4 The readout of the matrix

A 32x30 pixels matrix of a small pitch of $24.1 \times 16 \mu\text{m}^2$ with a matrix size of $674.8 \times 512 \mu\text{m}^2$ (60% of the area of the core) was selected. The core measure $788 \times 760 \mu\text{m}^2$. This configuration allows for sufficient space to accommodate the FIFO memory and smaller blocks. The second advantage is that matrices with suitable sizes for the column are available, including 64 to 1, 16 to 1 (with a 4 to 1 at the end), 4 to 1, and 2 to 1 arbiter sizes. All double columns are composed of 2 sides of 32 pixels. The sole exception is the matrix, which features four distinct columns and a line with a 2 to 1 arbiter size, thereby enhancing bandwidth. All sizes are powers of two, simplifying the design, except for the 16 to 1 size. Two additional sizes can be created as 8 to 1 and 32 to 1 (with a 2 to 1). In an initial simulation, both extreme sizes were tested as 64 to 1 and 2 to 1. The 4 to 1 size was tested as a compromise between space usage and bandwidth, with the objective of achieving a balance between the 2 to 1. The 16 to 1 with a 4 to 1 configuration represents a choice that exhibits a certain degree of acceleration in the bandwidth. In fact, slower blocks are situated in close proximity to the pixels where less bandwidth is required, while the circuit features a 4 to 1 ratio closer to the output. The 32 to 1 ratio will be closer to the 16 to 1 ratio, but for personal preference, the acceleration will not be sufficient. The 8 to 1 ratio is a medium case, and while it may be satisfactory in certain contexts, it is not the optimal choice in all situations.

The line will experience a slight reduction in speed, yet will not be constrained by bandwidth limitations. The ratio of the line to the space is 2 to 1, as the space available outside the matrix is less constrained. The investigation of varying dimensions will be conducted on four columns for each size, subsequently providing eight columns for each size to be subjected to testing on the test beam. The following table presents the details of the column figure 5.4.

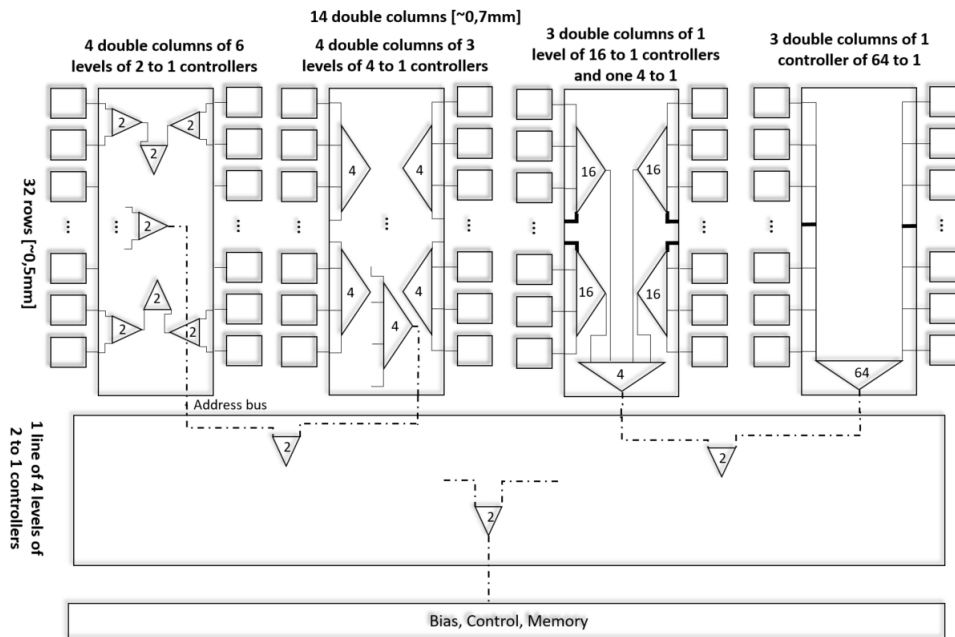


FIGURE 5.4: Distribution of the various controller sizes implemented in the pixel matrix readout of the SPARC circuit.

5.1.5 Memory for the datas

The FIFO (First In First Out) memory is a dual asynchronous clock provided by Cadence (CW_fifo_s2_sf) and features a 24-bit (see table 5.1) by 32-word configuration. The words are divided into 10 bits low address and 13 bits high for time-stamping, with one additional bit depending on the type of time-stamping, presented in the next table. The address is derived from the number of pixels (1024), which is equivalent to 10 bits. The timestamp is based on a 2 ns step, allowing for the capture of frames at 8.2 μ s in length. Given the area reserved, the 32 words occupy a power of two, with 70% of the remaining area utilized. Three serializers are incorporated, dividing the words into 8-bit segments at a frequency of 125 MHz.

Fast OR selection	Timestamp (gray)	Address
1 bit	13 bits	10 bits

TABLE 5.1: Attribution of the 24 bits word in the FIFO.

5.1.6 Timestamping

A digital counter has been incorporated into the apparatus to facilitate the annotation of timing data associated with the hits. The TDC (Time to Digital Counter) features a resolution of 2 nanoseconds, which is sufficient for the readout. In fact, the mean time to read the first address is 8 ns and the 2nd address is readed at 3 ns. The period of the counter may be controlled by a PAD or a VCO, which is fully digital. The bandwidth of the PADs is limited to 250 megahertz (4 nanoseconds). The VCO is constructed with inverter gates arranged in a ring configuration for a 2 ns period.

The VCO has been designed to be strictly at 2 ns period independent of the corner (aka the temperature, voltage, process). Four ring oscillators have been designed with a multiplexer to select the good one. The following table 5.2 summarizes the ring selection.

Selection	Number of inverters	Corner	Temperature	Process	Voltage
0	14	MAX	-40 125	slow	1.08
1	23	TYP	27	typical	1.20
2	30	unused			
3	38	MIN	-40 125	fast	1.32

TABLE 5.2: VCO table

A block is added, functioning as an asynchronous arbiter to annotate the time. The circuit permits the timestamp annotation of all pixels as addresses are read, the annotation of only the first pixel of a cluster with a fast OR signal on the matrix, or the annotation of both the fast OR and each pixel. The final bit (23rd) of the FIFO is set to "1" to indicate a timestamp derived from the fast OR. The objective of the final case is to reconstruct the pixel's timestamp by utilizing the first hit with a deterministic fast OR and the difference between pixels to calculate the ToA.

An output of the circuit from a PAD wired to the FIFO clock push signal is incorporated. The objective is to monitor the output of the matrix in order to ascertain

whether the circuit is locked in the event of a potential SEU. The reset of the matrix can be performed independently of the other blocks, if necessary. As previously stated, the PAD is unable to output the frequency of the matrix, which is in the vicinity of 500 megahertz. However, once the matrix is locked, the signal remains stable and can be detected.

The pixel trigger is a rising edge, but it is also possible for it to be triggered by a falling edge. In this case, it is possible to implement a ToT functionality. This may result in an increase in the bandwidth from double, although this could be achieved by the architecture in any case. The rising and falling edges are unknown, in that they would require the transmission of an additional bit within the matrix, as well as the use of additional registers that are not present.

5.1.7 Design of SPARC

The design of SPARC is relatively straightforward, with the TDC, Slow-Control, VCO and FIFO blocks being designed as synchronous classic flows with specifications developed prior to the commencement of the project. This section provides further insight into the enhancements made to the column, line, and top for the asynchronous components.

The column is comprised of four distinct types, each exhibiting identical register transfer level (RTL) orientation and varying specifications with respect to the SDC parameter. A 2 to 1 controller is considered next for the line. There are two types of memories: the select memory and the bus of address. The SDC files are generated through the following process. The following steps are required to create the SDC files:

- Global variables, including the period and uncertainties
- Declaration of the clocks
- False path on the I/O
- False path on the not-concern registers from/to
- Declaration of the clock groups
- Declaration of the uncertainties, if necessary
- Setting of the fanout
- Setting of the transition on I/O and clocks
- Setting of the capacitance values
- Setting of the load values

The Innovus software employs a multi-frequency or behavior in the circuit, which is defined by a series of modes. The modes serve as a workspace wherein the tools analyze the same circuit with disparate timing constraints, operating in a wholly independent manner. In asynchronous operation, the modes are employed to segregate the clocks from one another, rather than monitoring their functionality. In this manner, the clocks on the circuit points will not erase one another.

Each type of register is analyzed in a distinct mode at each level to prevent interference between the clock and speed up the analysis. This implies the existence of a mode for the setup of the Select register (mode S), a mode for the setup of the address registers from the Select register (mode I), a mode for the setup of the address registers from the upper address registers (mode D), and a mode for the setup of

the address registers from the upper Select register (mode DS). The aforementioned modes are then designated as S (Select), D (Data), DS (Data from Select), and I (Internal of the controller). Additionally, there are other modes, designated OUT, which are employed to constrain the I/O in the blocks but they all are parts of modes S, I, D, or DS in the top. In the hold timing verification, the I, D, and DS modes are merged because the clocks originate from the same point, and only two modes exist: IDS and S. A table is in the appendix to summarize all modes table C.

It should first be noted that the variable 'n' represents the level of study, while 'm' represents the sub-level in the case of a controller that is larger than 2 to 1 (i.e., more than one Sel_reg level). Secondly, it should be noted that the existence of certain modes is not universal. The I, DS, and D modes are not available with an N to 1 register with N pixels, as no Temp_reg exists (address register). The DS and D modes are not available at level 1; they are OUT modes. In the column, the modes D and DS do not exist on the final level because no data is sourced from the pixels in this design, in contrast to the line where the column address is provided.

The architectural shape of the pin names allows for their variation and scripting, which is facilitated by the ease with which the hierarchy can be altered. The SDCs are scripts that contain the majority of the time loops, which serve to reduce the code. However, different files are created for different controller sizes. It should be noted that a controller size that is not a power of two from the number of pixels can be complex to detail and script. The hierarchy is derived from the pixels to the output, with the exception of those that are not a power of 2, which are accompanied by negative level numbers. The setup.yaml file, which encompasses all modes and corners, is scripted with variables that can be modified.

In the D and DS modes, an additional degree of uncertainty is introduced to account for the decision on the Sel_reg. For the sake of argument, let us assume that the least-requested item arrives and clocks the Sel_reg, but the register is not fully locked. At this point, the most-prioritized request arrives and toggles the flop. The clock signal that is propagated will be the one from the least request, and it will be in advance of the real data that is stored (the most prioritized). This approach introduces a significant degree of pessimism, as it assumes a considerable delay on the AND3_CLK for both clocks, in order to ensure that the data is prepared with sufficient advance. The uncertainty is 100 ps for the typ corner and 180 ps for the skew2 and max corners.

The top design incorporates only the D, DS, and S constraints at the input/output (I/O) of each block, with the objective of verifying the delays. It should be noted that Innovus does not permit more than 256 analysis modes, which precludes the testing of modes within a block. There are D and DS modes between the columns and line, between the line and the FIFO, or between the arbiter and the annoter block. The S mode is verified between the DIP (Digital In Pixel) and the Column, as well as between the Columns and the Line. Two additional modes are created for synchronous blocks and the VCO.

The circuit arrangement will be represented in a top view as illustrated in the accompanying graphic figure 5.5. The matrix is positioned at the top of the circuit arrangement, surrounded by PADs. The Slow-Control is situated on the bottom left,

while the FIFO is located on the right. A list of the PAD is accessible in the appendices table E as well as a complete view of the whole circuit figure F.

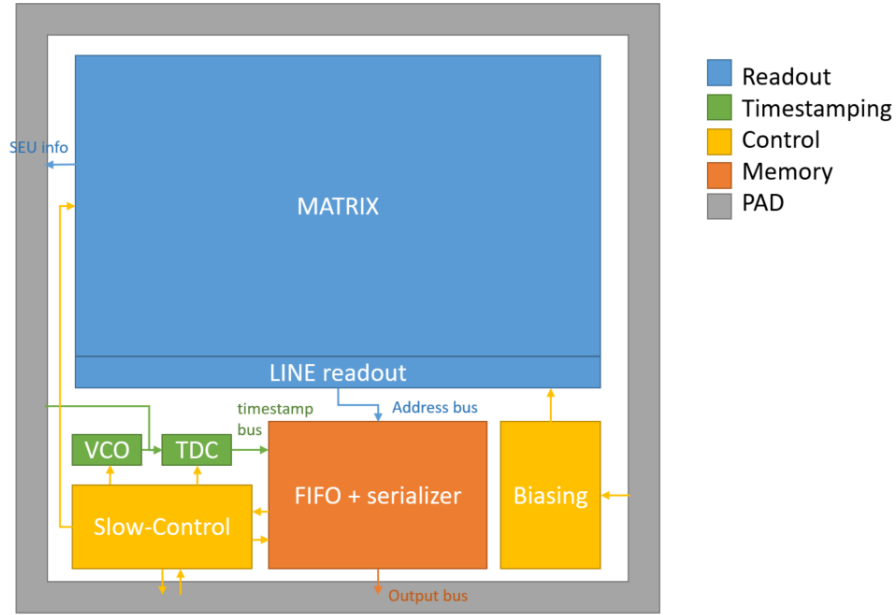


FIGURE 5.5: TOP view of the functional blocs of the SPARC sensor.

The masking scheme is quite complex and reflects some congestion in the design. In order to limit the number of wires, the mask on the columns is for a double column and thus is in the number of 16. The mask for the lines is then different on the same line as it must differentiate double columns and is in the number of 64. The details on this masking scheme is illustrated with figure 5.6.

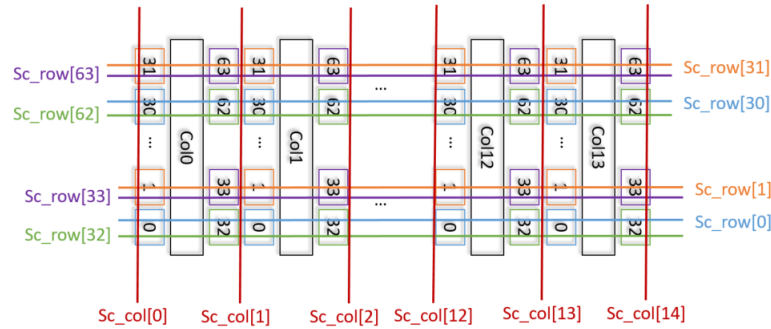


FIGURE 5.6: Mask of SPARC.

5.1.8 Expectations

The objective of this circuit is to demonstrate the feasibility of an asynchronous digital flow and an effective particle readout circuit. The primary objective is to enhance the bandwidth of the readout circuit. It is therefore expected that the FIFO will be capable of supporting the required data rate, that the TDC will be as fast as necessary, and that the readout will be in close agreement with the simulations.

Firstly, columns are subjected to post-layout simulations which the results are summarized in table 5.3. The power density and the max time to read a pixel are extracted from a simulation of the circuit with a hit rate of 40 MHz/cm^2 close to ALICE ITS3 in a continuous manner. The density of standard cells is computed by the tools from the area of the used cells over the area reserved for the readout. The wire length represent the sum of all wire lengths inside the readout block and give an idea of the possible congestion inside such circuit or at least an idea of the usage. The power density is taken from the whole simulation and is derived from the area of the pixels. Finally, the max time to read a pixel is extracted from a real simulation with ITS2 cluster sizes. The simulations are the same as the chapter 4. The measure is taken from the time when the signal arrive to the readout block and when the readout provide the address.

It is anticipated that larger controllers would have shared more parts, resulting in a reduction in the number of cells and wire length. The line readout employs a 2 to 1 controller architecture to enhance bandwidth, comprising a block of 16 inputs instead of 64 for the columns. Consequently, the line exhibits a density of 43.2%. It is noteworthy that the line incorporates additional input registers to store the column address. The power density appears to be comparable to that of the cells, as both are subject to the same activity. The line readout is not consuming a significant amount of power, with a threefold reduction compared to the column with the same controller. However, it has a ninefold maximum time to read pixels, which can be attributed to the fact that the entire matrix is simulated, rather than just a double column.

Controller Size	# inputs, levels	Standards cells [%]	Wire length [mm]	Power density [mW/cm ²]	Max time to read [ns]
2 to 1	64, 6	70.1	18.7	5.2	119.8
4 to 1	64, 3	46.6	17.4	2.8	193.3
16 to 1 + 4 to 1	64, 2	42.5	17.3	2.6	116.3
64 to 1	64, 1	39.7	17.1	2.3	100.7
Line 2 to 1	16, 4	43.2	16.7	1.8	928.4

TABLE 5.3: Results in terms of area, power and timing for the 5 different architecture implemented.

It should be noted that the area of each column is $24.1 \cdot 16 \cdot 64 = 24678.4 \text{ } \mu\text{m}^2$, while the area of the line is $24.1 \cdot 16 \cdot 14.4 = 5552.64 \text{ } \mu\text{m}^2$. The columns are composed of 64 pixels, and the line joins the 16 double columns. Other metrics are extracted with a 40 MHz/cm^2 ALICE-ITS2 cluster rate as input. They include the dispersion of times to read a fired pixel address in figure 5.7 and the readout time in figure 5.8. For the sake of simplicity, the timing extracted from the line readout uses a 64 to 1 columns attached before to every inputs.

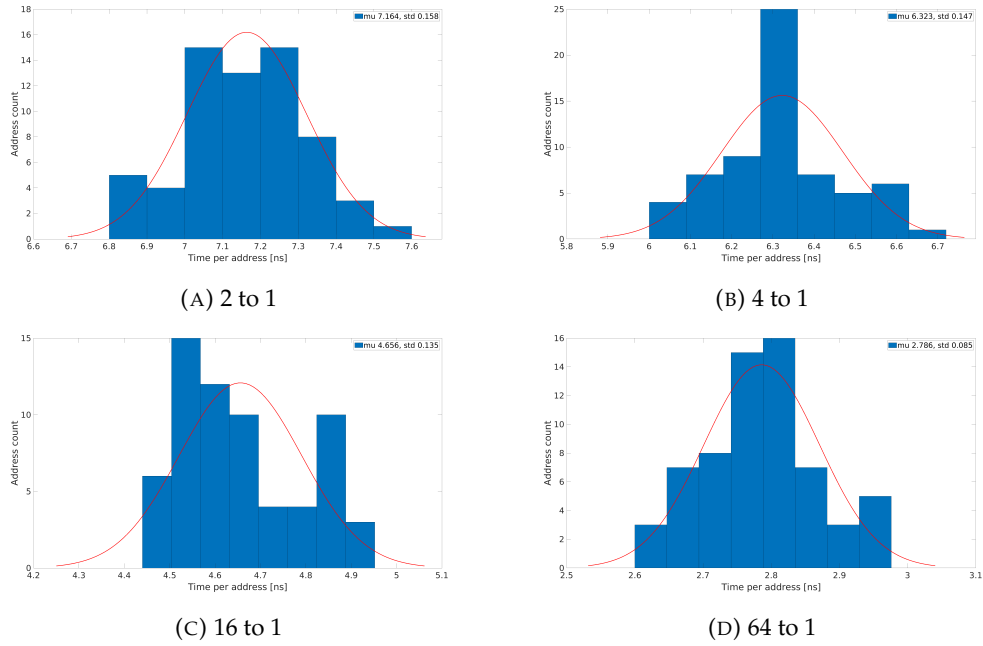


FIGURE 5.7: Comparison of the fired pixel address time distributions for the different controller sizes implements in the SPARC sensor.

The results present histograms of the reading time for each distinct address. The expected shape is that of a Gaussian distribution or, at the very least, a shape that is centered and grouped. This is indeed the case for the 64 to 1 and 2 to 1 ratios, but the 4 to 1 and 16 to 1 ratios exhibit a gap, resulting in some addresses requiring a longer time to traverse the readout block. While this is not an emergency, maintaining a grouped form ensures that the first-in-first-out working order is maintained.

The table 5.4 shows numbers obtained on the plot figure 5.7, the timing results for all column sizes are presented as minimum, mean and maximum time to read a pixel. A standard deviation and a ratio of the mean time over the number of levels are computed.

Controller Size, # levels	Min [ns]	Mean [ns]	Max [ns]	Standard deviation [ns]	Mean/level [ns]
2 to 1, 6	6.7	7.0	7.2	0.2	1.2
4 to 1, 3	5.3	5.6	6.3	0.3	1.9
16 to 1 + 4 to 1, 2	3.7	4.0	4.4	0.2	2.6 ; 1.3
64 to 1, 1	2.5	2.7	2.9	0.1	2.7
Line 2 to 1, 4	6.9	7.1	7.5	0.2	1.8

TABLE 5.4: Comparison of the timing over address distribution.

The most shared controller aka 64 to 1 has less gates on the critical timing path and then benefit from a faster address reading but it can only treat one pixel after the other so the bandwidth is reduced. The mean time over level can give an idea of the time to read the second pixel, it is computed by dividing over the number of levels and pondering for the 16 to 1 case. The expected column to perform the best is the 16 to 1 as the mean time for the first pixel is one of the best as the time for the second pixel. It use the best of the 2 to 1 and 64 to 1 architectures. Note that the

line readout has to accommodate all architectures and then seems to be quite slow in addition it has to take as an input the constructed address of the columns which could be time-consuming.

The figure 5.8 presents the summed percentage of readed pixels over time for the 40 MHz/cm² that uses the same algorithm as the realistic simulation presented in the chapter 4. It is presented as a sum normalized with as a percentage of every timing of a reader pixel for each column to the line. The limit is defined as the timing value at which 95% of the hits are read. The plots' shapes indicate a starting point that is not zero and a form that resembles a negative exponential. Obtaining the last pixels requires a significant time investment because of huge clusters; however, smaller controllers exhibit a steeper slope than bigger ones. The slopes come when the final controller is busy and does not provide new pixels.

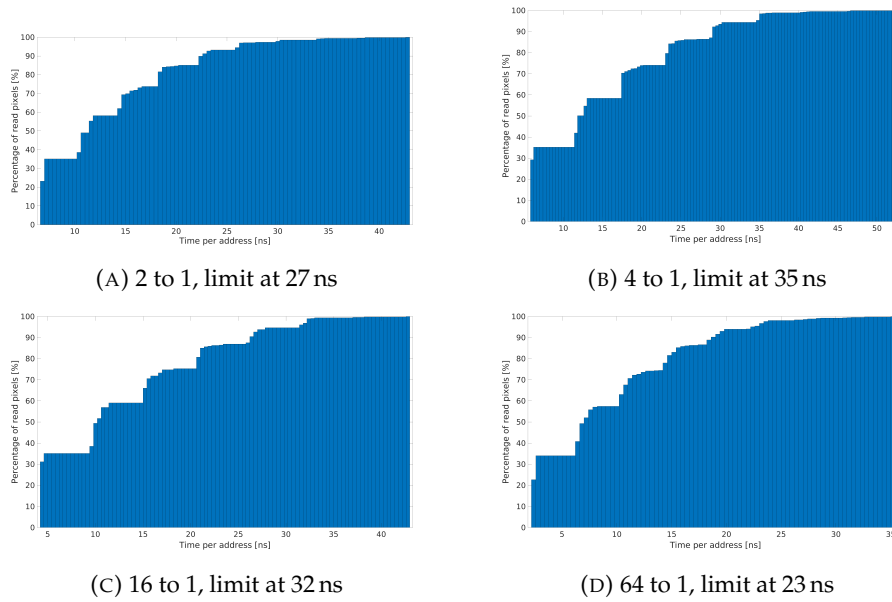


FIGURE 5.8: Comparison of the cumulated distribution of the arrival times for the fired pixel addresses, for the different controller sizes implements in the SPARC sensor.

Regarding all these results timestamping can be used with the help of the TDC block. A map can be constructed for all pixels, thereby enabling the determination of the time path of each address. For instance, the address 6 has a time of 14.3 nanoseconds to traverse the tree. This is accurate when only the pixel 6 has fired and all registers are in their initial state following a reset. In fact, the state of each register can affect the timing of the rising or falling edges, as the timing transitions may not be identical. Furthermore, the occurrence of congestion within the tree will result in the postponement of a hit to another for an undefined period of time. Timing is also dependent on PVT, which encompasses the powering scheme, temperature, and local process effects. Consequently, the map is only pertinent to the initial pixels fired in each crossing and can be deducted from the arrival time to determine the pixel's firing time.

By correlating the fast OR signals and the addresses, one can estimate the number of collisions of each hit and add the uncertainty of the decision circuit (40 ps

TYPICAL, see figure 3.6). Consequently, the map could be employed for subsequent hits with a marginal reduction in precision. It should be noted that the counter's step time is 2 ns, which represents the limiting factor in this context.

Note that the exact value where not simulated since the circuit was not well advanced during this report. The time walk is also not monitored, but the ToT can be known depending on the collisions and with lower precision in time.

5.2 Testing strategy

The circuit will be fabricated after the completion of this thesis. Consequently, we can only discuss the testing strategy here and obviously no results can be reported. The tests will be conducted in two phases. The first phase will be an electrical test to ascertain whether the circuit functions as intended. The second phase will be a beam test to quantify the circuit's performance.

5.2.1 Electrical tests

The initial electrical test involves interaction with the Slow-Control unit, which communicates over SPI (SCK, MISO, MOSI, CSEL). The registers are organized according to the table D. This test can also be implemented in the Protium platform to develop the acquisition board early in the submission. The Protium FPGA will emulate the Slow Control and the FIFO of SPARC as well as a mimic block of the readout to test everything

The test registers will be instrumental in the development of the protocol of communication. Subsequently, the Reg10 bits 4, 5, and 7 with the setting of registers 0 to 9 will be employed to test manual hit generation. The procedure begins with setting the manual mode, selecting the desired columns and rows to mask, and applying a clock signal. Another potential area for further investigation could be the TOT.

The TDC can be tested independently of the Reg10 bit 3. The Reg10 bits 1 modify the source clock as the VCO or the PAD, while the output PAD from bit 1 to bit 3 (clock division by 4) is affected by the Reg10 bit 0. Bit 3 will fill the FIFO with only TDC data, while other data are not reliable. The FIFO word is divided into three distinct parts: the address (0-9), the timestamp (22-10), and the bit (23) for the address type (F_OR or matrix request). The Reg11 bit 7 serves to mask the clock push on the FIFO when it is full.

The matrix tests will initially identify a hit in the FIFO, and then utilize the Reg11 bits 0 and 1 to select the requisite delay to be added in the event that the addresses are corrupted. The acknowledgment is subjected to a delay in order to impede the output of the matrix. The cases are as follows: 00, no delay; 01 or 10, a delay of 1 nanosecond; and 11, a delay of 2 nanoseconds. The Reg10 bit 3 allows the user to tag every hit or only the first one. The Reg10 bit 2 permits the tagging of data with the Fast OR and the matrix, with the assistance of an additional controller. The Fast OR has precedence over the matrix.

5.2.2 Laboratory tests

The SPARC circuit should be tested initially at the laboratory to confirm the correct functioning of the address reading in conjunction with the asynchronous behavior. A mask can be applied to the hidden half of the matrix to ascertain whether the addresses are in alignment. A laser can be employed to trigger different pixels and create a pattern, thereby also verifying the correct mapping of the addresses. The laser will operate at a wavelength of 1060 nm with a timing resolution of 100 ps.

An additional test could be conducted by placing the sensor on a known source with another known sensor in order to characterize the spatial resolution and the time required for reading. The measurement can be made in differential mode, with two pixels fired at a known delay, in order to characterize the time-stamping resolution of SPARC. It may be beneficial to characterize the triggering function on both the hit-OR and address settings, with the potential for identifying enhancements. The timing resolution of the laser may assist in characterizing the ToT measurement on the chip. The laser is highly sensitive and capable of triggering a minimum level of non-inhibition.

5.2.3 Under beam tests

The objective of the test beam is to validate the architectural design and quantify its operational limits. The objective is to ascertain the maximum bandwidth, radiation tolerance and the efficacy of specific functions as the ToT.

The most basic test is to obtain the same detection efficiency observed from the test of the DPTS sensor [61]. This requires a standard beam telescope and a beam of minimum ionizing particles generating a low hit rate of a few kHz/cm². Additionally, the data accumulated can be used to confirm the relative time of arrival for individual pixels within the same cluster predicted by simulations. Especially, the delay for neighboring pixels in the same column, should be reproduced. The same data shall be used to check the possibility of exploiting the time over threshold, since a proper Landau distribution (convoluted by a Gaussian distribution for the threshold dispersion) should be obtained for the cluster signal reconstructed from the ToT information.

With the addition of the timing information coming from the trigger or a dedicated fast detector (providing time resolution better than 1 ns) to the same telescope, the absolute time-stamping precision can be calibrated and compared to results obtained with the laser in the laboratory.

The maximal bandwidth estimation should be approached with a high-intensity beam. Of course, the data bandwidth translates into a hit rate depending on the average cluster size generated by particles. Cyclotrons are excellent machines to explore high hit rates since they reach beam intensity of μA . An example is the CYRCé cyclotron at IPHC-Strasbourg, delivering 24 MeV protons. Such protons generate much larger clusters than MIPs however and it is expected that the maximum hit rate will be limited by that effect. Synchrotrons providing electrons at a few GeV, such as MAMI (Mainz) or eALBA (Barcelona) can be an option to investigate the hit rate sustainable for MIP-like cluster sizes. The evaluation of the SEE sensitivity shall be done in dedicated installation in addition to the output request signal that indicates if the circuit is stuck.

At last, a SEU test may be conducted using the Request output of the matrix. Indeed, should a single-event upset (SEU) occur and lock the circuit, it can be detected with an FPGA. This is because the output request of the matrix will remain high for an extended period of time (approximately 20 nanoseconds). The FPGA could be utilized to reset the matrix without erasing the configuration in the slow-control and acquire new data. Additionally, it could be employed to count the number of SEUs

and provide an estimation of the chip's sensitivity.

5.3 Conclusion

SPARC is a sensor prototype intending to validate the concept of the asynchronous matrix read-out architecture developed in this thesis. Despite the relatively small pixel size of $20\text{ }\mu\text{m}$, SPARC will allow us to investigate the performance of four different controller sizes. In particular, we discussed a test strategy in order to verify the time-stamping ability at a few tens of nanoseconds and the maximal hit rate the architecture can cope with, at least beyond several hundred MHz/cm^2 .

Assuming tests will indeed demonstrate the efficacy of the proposed asynchronous architecture, this readout circuit becomes an excellent candidate for pixel sensors devoted to tracking systems where low power density is a strong requirements as well as coping with large hit rates, which is the case for projects like ALICE3, Belle II, FCCee, LHCb.

Résumé

Le chapitre 5 de cette thèse se concentre sur les résultats expérimentaux obtenus suite à la mise en œuvre des concepts théoriques et des designs de circuits introduits dans les chapitres précédents. Ce chapitre a pour objectif principal de démontrer l'efficacité des approches proposées pour la conception de circuits asynchrones dans le contexte des capteurs de particules utilisés en physique des hautes énergies, en particulier pour les détecteurs à pixels. Le chapitre commence par rappeler le contexte de l'étude, soulignant les défis techniques associés à la conception de capteurs capables de fonctionner efficacement dans des environnements aussi exigeants que ceux des expériences de physique des particules, notamment au CERN.

La physique des hautes énergies, qui constitue le cadre expérimental de cette thèse, nécessite des dispositifs capables de détecter et de traiter des quantités massives de données en temps réel, tout en maintenant une faible consommation d'énergie et une précision élevée. Les capteurs utilisés dans ces expériences doivent être en mesure de capturer des événements qui se produisent à des échelles de temps extrêmement courtes, souvent de l'ordre de la nanoseconde, et de fonctionner dans des conditions où les niveaux de radiation sont très élevés. Dans ce contexte, on se propose d'explorer les circuits asynchrones comme une alternative prometteuse aux architectures synchrones traditionnelles.

Les circuits asynchrones se distinguent par l'absence d'une horloge globale pour synchroniser les opérations, ce qui permet de réduire la consommation d'énergie et d'améliorer la robustesse du système face aux variations de température et de tension, qui sont courantes dans les environnements de détection de particules. Cependant, l'absence d'horloge pose également des défis significatifs, notamment en matière de conception et de gestion de la synchronisation entre les différentes parties du circuit. Le chapitre 5 vise à valider ces propositions théoriques par des expériences pratiques et des simulations afin de démontrer les avantages potentiels des circuits asynchrones dans ce contexte particulier.

La méthodologie adoptée pour valider les performances des circuits asynchrones comprend plusieurs étapes. Tout d'abord, les circuits sont modélisés à l'aide de logiciels de simulation avancés, permettant de prédire leur comportement dans diverses conditions opérationnelles. Ces simulations sont cruciales pour identifier les points forts et les éventuelles faiblesses des designs proposés avant de passer à la phase de fabrication des prototypes. Une fois les simulations terminées et les résultats analysés, des prototypes de circuits asynchrones sont fabriqués pour être testés dans des conditions réelles. Ces prototypes intègrent les innovations proposées dans les chapitres précédents, telles que l'architecture asynchrone pour la lecture des pixels et la gestion des flux de données.

Les tests expérimentaux réalisés sur ces prototypes sont destinés à évaluer plusieurs aspects critiques des circuits, notamment la vitesse de traitement des données, la précision des détections, et l'efficacité énergétique. La vitesse de traitement est un critère particulièrement important dans les détecteurs de particules, car elle détermine la capacité du circuit à capturer des événements qui se produisent à des intervalles de temps très courts. La précision, quant à elle, est essentielle pour assurer que chaque particule détectée est correctement identifiée et que les données correspondantes sont traitées sans erreur. Enfin, l'efficacité énergétique est un paramètre clé, car une faible consommation d'énergie permet de réduire les coûts opérationnels

et de prolonger la durée de vie des capteurs, ce qui est particulièrement important dans les expériences à long terme.

Les résultats expérimentaux obtenus au cours de ces tests montrent que les circuits asynchrones présentent plusieurs avantages significatifs par rapport aux architectures synchrones traditionnelles. En termes de vitesse de traitement, les circuits asynchrones se sont avérés capables de gérer des flux de données élevés avec une grande rapidité, répondant ainsi aux exigences des expériences de physique des particules. Cette performance est en grande partie attribuée à l'absence d'horloge globale, qui permet aux différentes parties du circuit de fonctionner de manière plus indépendante et flexible, réduisant ainsi les délais de traitement.

En ce qui concerne la consommation d'énergie, les circuits asynchrones ont montré une réduction significative par rapport aux circuits synchrones. Cette réduction est particulièrement importante dans le contexte des détecteurs de particules, où la minimisation de la consommation d'énergie est cruciale pour maintenir les capteurs opérationnels sur de longues périodes sans avoir besoin de fréquentes interventions pour le remplacement. Les tests ont également montré que ces circuits asynchrones étaient capables de maintenir une précision élevée dans la détection des particules, même dans des environnements à haute densité d'événements, ce qui est essentiel pour assurer la qualité des données collectées.

L'analyse des résultats met également en évidence que, malgré les nombreux avantages des circuits asynchrones, leur conception et leur mise en œuvre présentent des défis notables. L'un des principaux défis est la complexité accrue du design, en particulier en ce qui concerne la gestion de la synchronisation entre les différentes parties du circuit. En l'absence d'une horloge commune, il est crucial de s'assurer que les données circulent de manière fluide et sans conflit entre les différentes unités du circuit, ce qui nécessite une ingénierie méticuleuse et des méthodes de conception sophistiquées. De plus, le manque de standards établis pour la conception de circuits asynchrones par rapport aux circuits synchrones pose également des obstacles à une adoption plus large de cette technologie.

Le chapitre se termine par une discussion sur les perspectives d'avenir pour les circuits asynchrones SPARC dans le domaine de la détection de particules. On y souligne que, malgré les défis techniques, les bénéfices potentiels de cette technologie justifient des efforts de recherche continus. Les circuits asynchrones offrent des perspectives prometteuses pour l'amélioration des performances des capteurs de particules, en particulier dans des environnements où la vitesse de traitement des données et l'efficacité énergétique sont des critères essentiels. En outre, l'intégration de nouvelles technologies de semi-conducteurs pourrait permettre de surmonter certains des défis actuels et d'améliorer encore les performances de ces circuits.

En conclusion, le chapitre 5 démontre que les circuits asynchrones représentent une alternative viable et prometteuse aux architectures synchrones traditionnelles dans le domaine des détecteurs de particules en physique des hautes énergies. Les résultats expérimentaux confirmeront les avantages théoriques de cette approche, notamment en termes de rapidité, d'efficacité énergétique, et de précision. Bien que des défis subsistent, les perspectives d'avenir pour cette technologie sont encourageantes, ouvrant la voie à de nouvelles innovations qui pourraient transformer la manière dont les capteurs de particules sont conçus et utilisés dans les expériences

scientifiques futures.

Chapter 6

Conclusion

This chapter provides a comprehensive summary and conclusion of the research conducted on the asynchronous readout design. It delineates the subsequent steps necessary to integrate this technology within an accelerator and to optimize its performance. A section is devoted to an enumeration of potential improvements that could be made in the next iteration of the design.

6.1 Work already done

This sections reviews the three main achievements described in chapters 3, 4, 5.

6.1.1 Asynchronous flow

As previously outlined in chapter 3, an asynchronous flow was developed utilizing the standard design tools to ensure flexibility and make it reusable in future circuits. This work constituted one of the major endeavors of my three year doctoral research. For a new circuit, the RTL and the SDC must of course be rewritten to match the new specifications. To simplify this task, this report provides a detailed description of the process for doing so. The SDC could have been scripted, but given the different controller sizes involved, it was deemed too complex to create a single script and writing a script for each size did not appear to be a viable option. Finally, the RTL used for the arbiter employs multiplexers scripted in TCL to perform every size analysis, since the implementation in Verilog was estimated too complex.

In order to accommodate the largest circuit, the flow must be adapted to perform a double analysis. Innovus does not support more than 256 analysis views with 5 setup corners and 9 hold corners, and thus the user is limited to 18 modes because each mode is performed in all corners. ($\lfloor \frac{256}{5+9} \rfloor = 18$). A DMMMC analysis in each steps will allow to have distributed modes on servers and allow infinite number of modes but it will be very slow. The controller N to 1 (with N the power of 2 of the number of pixels) has N modes of S in hold and N OUT, N S in setup analysis. AS seen in the equation 6.1, the maximal power of 2 accessible is 13 so a maximal number of pixel of 8192.

$$N \cdot 9 + (N + N) \cdot 5 = 19 \cdot N \quad (6.1)$$

The 2 to 1 controller has N-1 IDS modes, N S modes in hold and N-2 D, N-1 DS, N-1 I, N S, 2 OUT modes in setup. The number of pixels accessible for the 2 to 1 controller is provided in the equation 6.2 and is a power of 2 equal to 7, so a maximal number of pixels in the standard flow of 128.

$$[(N - 1) + N] \cdot 9 + [(N - 2) + (N - 1) + (N - 1) + N + 2] \cdot 5 = 38N - 19 \quad (6.2)$$

This can be overpasses in Tempus, the timing tool or by choosing the critical modes manually to optimized in Innovus. It is expected that following the recipe described in chapters 3 and 4, the design of an asynchronous logic can be fast and focused on optimizing the circuit performance rather than the design flow.

6.1.2 Performance of the proposed asynchronous readout architecture

The proposal for the asynchronous readout shows possibility to achieve a pitch of 20 μm with columns height of around 1cm. The consumption is below the expected requirements as 10mW/cm² and the time to read is ways faster than the state of the

art as 2 to 3 times faster, see chapter 4.

A proposed chip is expected to achieve the same power consumption, fit in a $24.1 \times 16 \mu\text{m}$ pitch (so $19.64 \mu\text{m}$ square equivalent) which is only limited by the 2 to 1 controller and can be reduce on the readout to maybe a third. The simulated readout will be a bit slow compared to the simulation because of security for the first time. In fact, an output delay on the matrix can be adjusted in case of failure. The chapter 5 shows the results.

6.1.3 Demonstrator prototype design

A prototype sensor named SPARC has been designed to demonstrate the efficacy of the asynchronous readout architecture, see chapter 5. SPARC belongs to a number of circuits included in an engineering run mainly dedicated to MOSAIX, the MAPS to equip the future ITS3 of the ALICE experiment. The submission schedule follows the MOSAIX design constraint and the fabrication is expected to happen after the completion of my doctoral thesis.

The prototype area allowed was very small but it allows to explores various options. Four controller sizes are implemented, allowing to compare with the predictions from simulations for the timing performance of each possibility of tuning the chip. Pixel timestamp can be performed over three modes, fast-OR timestamping, first pixels of the cluster, every pixel timestamp. A recovery port to encompass the SEU effect is present.

The SPARC circuit has been designed with a great deal of margin in the asynchronous parts, which can be removed if necessary, thus allowing for a potential gain of approximately 10 to 20% in speed, power, and area. The thorough tests planned for SPARC should allow to fully evaluate the implemented asynchronous architecture, which then could serve as the basis for a new readout system for MAPS and offer a powerful alternative to existing architectuer for new experiments.

6.2 What's next ?

Results obtained in this work supports the application of the developed asynchronous architecture in larger sensor (1M pixels), following the SPARC demonstrator and targeting specific experiments. There are also paths to further optimize the architecture itself. This section investigates briefly these future developments.

6.2.1 Further architecture optimization

The time constraint of a PhD thesis naturally limits the options to be followed in the developments. The implementation described in Chapter 4 is a result of these choices, which allowed an initial evaluation of the expected performance for the readout architecture. We consider here some other options that could be pursued in the future and their potential benefits.

A version of the controller can be made with no reset signal inside the selection memory to gain in area even more and could be useful when the pixel pitch should be minimized. The reset is not needed but assure a fix start state and also no request due to a powering scheme that are not hits. This option was not pursued in SPARC to limit the risk, but could be considered once the first implementation of the architecture has been thoroughly tested and well understood.

Currently, the Priority Encoder (PE), described in chapter 2, constitutes the principal pixel matrix readout deployed by the C4Pi at IPHC, see the MIMOSIS [40] and MOSS [16] sensors. The circuit is designed to operate in an asynchronous mode, with all signals permitted to move freely within the circuit, and can be seen as an asynchronous 4 to 1 controller. However the external component that reads the PE are undoubtedly synchronous, typically reading one address every 25 nanoseconds. Still, one could focus on using the internal part of the PE as an alternative implementation of an asynchronous controller compared to the implementation realized in this thesis. The priority encoder has faster result when providing the address but it need a freezing signal to ensure that the address will not change while saving it. A key difference is that in all arbiter designs presented in the thesis, a memory of the decision is present, and two inputs are prioritized. In the PE, there is no memory, and therefore, the address can change even during the acknowledgment and erase the wrong pixel. Consequently, when the fast OR is initiated, a freeze signal must be transmitted to the pixels in order to halt the output and await the address propagation. This is followed by the reading, acknowledgment, and unfreezing of the pixels. The routing is highly complex, which is why I chose to pursue my thesis with a different controller. Alternatively a memory can be added to the pixel to freeze data in a critical area. A work to evaluate the possibility of creating asynchronous routing and extract the performances of a true asynchronous priority encoder will be a good start in the future to see if it can overpass or not the priority arbiter.

The requirements on the pitch are strong and will impose a 10 μ m pitch witch is the size of the diode in DPTS. The only way to do so is to put the readout outside the matrix and read it with an selector and a tree. The asynchronous proposal can read pixels very fast and may allow this possibility to be competitive with the wanted hit rates but with the cost of a dead zone. A proposition can be to mix a priority encoder horizontally to select the columns to read and a priority arbiter to read them

fast figure 6.1.

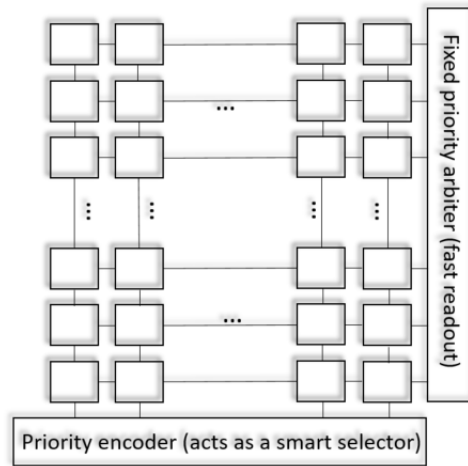


FIGURE 6.1: Possible architecture for a very small pitch of 10µm.

Besides options beneficial to a smaller pixel pitch, the time and resources to develop and implement a readout architecture for a given application could also be important considerations in a project.

A parametric arbiter could be constructed, with the objective of creating a 2 to 1 controller with multiplexers to yield a 4 to 1 and larger controller sizes. Such a solution would allow to design a matrix readout with very complex involvement in coding the implementation but will be the same layout for an inner, outer tracker or to address different experiments. It was not initially conceived as a dynamic system; rather, it was designed to accommodate a relatively slow control system, with the objective of creating a circuit for each tracker layer (ITS/OTS). This approach ensures a high bandwidth closer to the beam and low power consumption at a distance from the beam pipe. The SDC must be mixed for each size, and case analysis must be conducted on the selection of controller sizes. This could result in a complex design process. While such parametric circuits may be useful in certain contexts, the results in chapter 4 indicate that there is not a significant difference in bandwidth between the controller sizes. It seems reasonable to posit that there is an optimal solution for each experiment, and even a unique optimal solution for a large scale of experiments.

Finally, the present circuit design is identical in the line readout and the column. The sole distinction between the two is that the line incorporates input data in addition to the acknowledgment of the request. In some cases, the pixel signal is digitized with an ADC or with the Time Over Threshold techniques, necessitating the transmission of data concurrent with the fired pixel address. The arbiter developed in the thesis can be utilized to directly obtain data from the pixels during the column readout. The cost of this will be arbiters with additional memory proportional to the number of bit of the ADC at each levels, so a increase of the area, consumption and maybe a slightly degradation on the timing as more data has to be synchronized.

6.2.2 Foreseen applications of the asynchronous architecture

The asynchronous architecture proposed in this work and implemented in the SPARC demonstrator is part of the R&D activities on MAPS for the ALICE ITS3 project [59] and connected with the CERN Experimental Physics division R&D roadmap, work package 1.2 [CERN](#). The short timeline of the ITS3 project lead to the choice of the already mentioned Priority Encoder architecture for the pixel matrix readout, implemented in the MOSAIX circuit [18] to be submitted in the coming month.

Based on the simulated performance of the proposed asynchronous architecture, in terms of time stamping, hit rate range and power dissipation, applications are now considered in R&D projects with the ECFA [DRD3](#) and [DRD7](#) collaborations.

A first [project](#) targets a MAPS demonstrating as primary goal the position resolution of 3 micrometers required by future $e^+ + e^-$ colliders like FCCee and matching the requirements in terms of hit rates and speed as discussed in Chapter 1. The strong requirement on the position resolution will lead to further work to make the architecture more compact in order to adapt to the necessary small pitch of collecting diodes. An open question is whether pixel of about 15 micrometers can be reached or not.

A second project focus on a pixel matrix that can be versatile for various trackers (LHCb upstream tracker, ALICE 3, FCCee). Here the strength of our proposed architecture is the wide range of hit rates that can be handled and the fact that the power dissipated evolves linearly with this rate. The SPARC demonstrator described in chapter 5 already offers pixel size around 25 micrometers fitting tracker needs and hence can be seen as a good seed for such projects.

Résumé

Le chapitre 6 de cette thèse constitue la conclusion générale des travaux de recherche menés sur la conception et l'application de circuits asynchrones dans les détecteurs de particules, en particulier ceux utilisant des capteurs à pixels pour la physique des hautes énergies. Ce chapitre synthétise les principaux résultats obtenus, évalue l'impact des contributions apportées et propose des perspectives pour les recherches futures dans ce domaine.

Au cours des chapitres précédents, l'exploration en profondeur les défis et les opportunités liés à l'utilisation de circuits asynchrones dans des environnements exigeants tels que les détecteurs de particules utilisés dans les expériences au CERN. Les motivations initiales de cette recherche étaient centrées sur la nécessité d'améliorer la vitesse de traitement des données, de réduire la consommation d'énergie et d'augmenter la fiabilité et la précision des systèmes de détection face à des taux élevés de radiation et à des volumes massifs de données à traiter en temps réel.

Dans cette conclusion, on y récapitule les avancées théoriques et pratiques réalisées tout au long de cette étude. Les travaux ont démontré que les circuits asynchrones offrent des avantages significatifs par rapport aux architectures synchrones traditionnelles. En éliminant la nécessité d'une horloge globale, les circuits asynchrones permettent une réduction notable de la consommation d'énergie, ce qui est essentiel pour les détecteurs opérant dans des conditions où la dissipation thermique doit être minimisée. De plus, cette approche contribue à améliorer la robustesse du système en réduisant la sensibilité aux variations de température et de tension, facteurs critiques dans les environnements de physique des hautes énergies.

Les résultats expérimentaux présentés ont confirmé que les designs de circuits asynchrones élaborés sont capables de gérer efficacement des flux de données très élevés, tout en maintenant une haute précision dans la détection et le traitement des événements. Ces performances ont été validées à travers des simulations détaillées et des prototypes testés dans des conditions représentatives des applications réelles. On y souligne que ces résultats sont encourageants et ouvrent la voie à une adoption plus large de la conception asynchrone dans les futures générations de détecteurs de particules.

En outre, le chapitre discute des défis rencontrés au cours de cette recherche et des solutions apportées pour les surmonter. La conception de circuits asynchrones présente une complexité accrue, notamment en ce qui concerne la gestion de la synchronisation locale et la coordination entre les différentes unités fonctionnelles du système. On y propose et valide des méthodologies de conception innovantes, incluant l'utilisation de protocoles de communication efficaces et de structures modulaires qui facilitent l'intégration et la scalabilité des systèmes. Ces contributions méthodologiques constituent un apport significatif à la littérature existante et fournissent un cadre solide pour les futurs travaux dans ce domaine.

Le chapitre 6 aborde également les implications pratiques des recherches menées. Les avancées réalisées dans la conception de circuits asynchrones ont le potentiel de transformer la manière dont les détecteurs de particules sont conçus et utilisés, en particulier dans le contexte des expériences de physique des hautes énergies qui exigent des performances toujours plus élevées. Les améliorations en termes d'efficacité énergétique et de vitesse de traitement peuvent conduire à des détecteurs

plus compacts, plus fiables et capables de fournir des données de meilleure qualité, ce qui est crucial pour approfondir notre compréhension des phénomènes physiques fondamentaux.

Enfin, la conclusion propose plusieurs axes de recherche futurs pour poursuivre et approfondir les travaux initiés dans cette thèse. Parmi ces perspectives, on y suggère d'explorer l'intégration de matériaux et de technologies semi-conductrices avancées pour améliorer encore les performances et la résilience des circuits asynchrones. De plus, l'étude de nouvelles architectures et de méthodes de conception assistée par ordinateur pourrait faciliter le développement et la mise en œuvre de systèmes encore plus complexes et performants. On y souligne également l'importance de collaborations interdisciplinaires et internationales pour partager les connaissances et accélérer les progrès dans ce domaine en constante évolution.

En résumé, le chapitre 6 conclut que les recherches menées ont apporté des contributions significatives à la conception et à l'application de circuits asynchrones dans les détecteurs de particules, démontrant leur potentiel pour répondre aux défis actuels et futurs de la physique des hautes énergies. Les résultats obtenus offrent une base solide pour de futures explorations et innovations, et renforcent l'idée que l'adoption de technologies asynchrones peut jouer un rôle clé dans l'avancement des sciences et des technologies liées à la détection et au traitement de l'information à haute performance.

Appendix A

Müller gate

Timing arc	Prevector (A, B)	Vector (A, B, RN, Y)
A hidden	0, 0	R, x, x, x
A hidden	1, 1	F, x, x, x
B hidden	0, 0	x, R, x, x
B hidden	1, 1	x, F, x, x
RN hidden	none	0, 0, R, x
RN hidden	none	0, 0, F, x
A to Y rising	0->0, 0 -> 1	R, 1, 1, R
A to Y falling	1->1, 1 -> 0	F, 0, 1, F
B to Y rising	0 -> 1, 0 -> 0	1, R, 1, R
B to Y falling	1 -> 0, 1 -> 1	0, F, 1, F
RN to Y rising	none	1, 1, R, R
RN to Y falling	none	1, 1, F, F

TABLE A.1: Timing arc for the Müller gate

Leakage arc	Prevector (A, B)	Condition
B rising	0, 0	$\neg A \& B \& \neg RN$
A rising	0, 0	$A \& \neg B \& \neg RN$
A falling	1, 1	$\neg A \& B \& \neg RN$
B falling	1, 1	$A \& \neg B \& \neg RN$
Reset	none	$\neg A \& \neg B \& RN$
Reset	none	$\neg A \& B \& RN$
Reset	none	$A \& \neg B \& RN$
Reset	none	$A \& B \& RN$
Y falling	none	$\neg A \& \neg B \& \neg RN$
Y rising	none	$A \& B \& \neg RN$
Mean	none	none

TABLE A.2: Leakage arc for the Müller gate

Appendix B

Constraints modes

Ack 0	Ack 1	Req 0	Req 1	Out
0	0	0	0	X
0	0	0	1	X
0	0	1	0	X
0	0	1	1	X
0	1	0	0	0
0	1	0	1	0
0	1	1	0	1
0	1	1	1	1
1	0	0	0	0
1	0	0	1	1
1	0	1	0	0
1	0	1	1	1
1	1	0	0	0
1	1	0	1	1
1	1	1	0	1
1	1	1	1	1

TABLE B.1: Logic table of the priority function

Appendix C

Arbiter priority function

Timing type	Mode	Start point	Crossing register	End register
Setup	Select (S)	$C_GATE[n+1] \mid AND3_CLK[n+1]$	none	Sel_reg[n]
Setup	Internal (I)	$AND3_CLK[n][m]$	Sel_reg[n][m]	Temp_reg[n]
Setup	Data from Select (DS)	$AND3_CLK[n+1][m]$	Sel_reg[n+1][m]	Temp_reg[n]
Setup	Data (D)	$C_GATE[n+1]$	Temp_reg[n+1]	Temp_reg[n]
Setup	Output (OUT)	$C_GATE[0] \mid AND3_CLK[0][m]$	Temp_reg[0]	DOUT[*]
Hold	Select (S)	$AND3_CLK[n]$	Sel_reg[n]	Sel_reg[n] or Ack[*]
Hold	Internal (IDS)	$C_GATE[n]$	Sel_reg[n][m]	Temp_reg[n]
Hold	Data from Select (IDS)	$C_GATE[n]$	Sel_reg[n+1][m]	Temp_reg[n]
Hold	Data (IDS)	$C_GATE[n]$	Temp_reg[n+1]	Temp_reg[n]

TABLE C.1: Differents modes summary

Appendix D

Slow control registers

Register name	Function	Bits
Reg0	Row selector	0-7
Reg1	Row selector	15-8
Reg2	Row selector	23-16
Reg3	Row selector	31-24
Reg4	Row selector	39-32
Reg5	Row selector	47-40
Reg6	Row selector	55-48
Reg7	Row selector	63-56
Reg8	Column selector	0-7
Reg9	Column selector	15-8
Reg10	TDC output to PAD	0
Reg10	TDC source VCO/PAD	1
Reg10	TDC test mode	2
Reg10	TDC select first/every hit	3
Reg10	Clk selected pixels	4
Reg10	Enable mask	5
Reg10	Enable ToT mode	6
Reg10	Enable stop on FIFO full	7
Reg11	Matrix output delay selector	1-0
Reg11	VCO corner selector	3-2
Reg11	Enable VCO	4
Reg11	TDC is latch	5
Reg12	FIFO pop flags (error, full, af, hf, ae, empty)	5-0
Reg13	FIFO push flags (error, full, af, hf, ae, empty)	5-0
Reg14	Test register	0xF1
Reg15	Test register	0xE2

TABLE D.1: Control bits

Appendix E

PAD list

PAD name	Type	Count	Function / speed
SUB	Substrate	1	Substrate
DVDD	Power	3	digital power
DVSS	Ground	3	digital ground
AVDD	Power	1	analog power
AVSS	Ground	1	analog ground
FIFO_Data	Digital output	3	125 MHz
FIFO_Clk	Digital input	1	125 MHz
FIFO_full	Digital output	1	flag
FIFO_Empty	Digital output	1	flag
FIFO_SER_FIRSTB	Digital output	1	flag
SC_Clk	Digital input	1	40 MHz
SC_SCK	Digital input	1	4 MHz
SC_MISO	Digital output	1	4 MHz
SC_MOSI	Digital input	1	4 MHz
SC_CSEL	Digital input	1	4 MHz
TDC_Clk	Digital input	1	250 MHz
TDC_out	Digital output	1	32 / 250 MHz
RN	Digital input	1	reset
MAT_RN	Digital input	1	matrix reset
AB_Clk	Digital output	1	flag 500 MHz
PULSE_CMD	Digital input	1	bias
IBIAS	Analog input	1	bias
IBIASN	Analog input	1	bias
IRESET	Analog input	1	bias
IDB	Analog input	1	bias
VCASB	Analog input	1	bias
VCASN	Analog input	1	bias
VH	Analog input	1	bias

TABLE E.1: Control bits

Appendix F

Complet view of the SPARC design

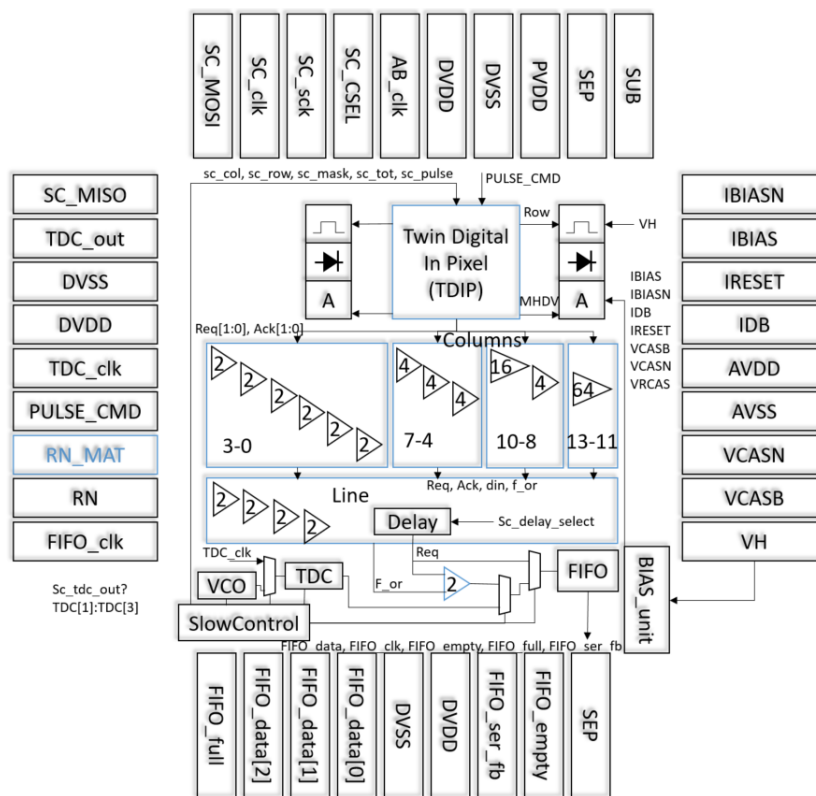


FIGURE F.1: TOP view of SPARC.

Bibliography

- [1] G. Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 17, 2012), pp. 1–29. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020). URL: <https://www.sciencedirect.com/science/article/pii/S037026931200857X> (visited on 09/06/2024).
- [2] Gage DeZoort et al. "Charged Particle Tracking via Edge-Classifying Interaction Networks". In: *Computing and Software for Big Science* 5 (Dec. 1, 2021). DOI: [10.1007/s41781-021-00073-z](https://doi.org/10.1007/s41781-021-00073-z).
- [3] Felix Reidt. *Upgrading the Inner Tracking System and the Time Projection Chamber of ALICE*. Feb. 2, 2020.
- [4] Gianluca Aglieri Rinella et al. "The TDCpix ASIC: High rate readout of hybrid pixels with Timing Resolution Better than 200 ps". In: *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*. 2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC). Oct. 2013, pp. 1–4. DOI: [10.1109/NSSMIC.2013.6829432](https://doi.org/10.1109/NSSMIC.2013.6829432).
- [5] A.R. Faruqi and G McMullan. "Electronic detectors for electron microscopy". In: *Quarterly reviews of biophysics* 44 (Apr. 28, 2011), pp. 357–90. DOI: [10.1017/S0033583511000035](https://doi.org/10.1017/S0033583511000035).
- [6] P. Kodyš. "DEPFET beam test results — Pixel properties studied at micron level resolution". In: *IEEE Nuclear Science Symposium Medical Imaging Conference*. IEEE Nuclear Science Symposium Medical Imaging Conference. Oct. 2010, pp. 1021–1024. DOI: [10.1109/NSSMIC.2010.5873920](https://doi.org/10.1109/NSSMIC.2010.5873920).
- [7] Ladislav Andricek et al. "Advanced testing of the DEPFET minimatrix particle detector". In: *Journal of Instrumentation* 7 (Jan. 27, 2012), p. C01101. DOI: [10.1088/1748-0221/7/01/C01101](https://doi.org/10.1088/1748-0221/7/01/C01101).
- [8] J. Duvernay et al. "Development of a self-aligned pnp HBT for a complementary thin-SOI SiGeC BiCMOS technology". In: *2007 IEEE Bipolar/BiCMOS Circuits and Technology Meeting*. 2007 IEEE Bipolar/BiCMOS Circuits and Technology Meeting. Sept. 2007, pp. 34–37. DOI: [10.1109/BIPOL.2007.4351833](https://doi.org/10.1109/BIPOL.2007.4351833).
- [9] Y. Arai and others. "Developments of SOI monolithic pixel detectors". In: *Nucl. Instrum. Meth. A* 623 (2010), pp. 186–188. DOI: [10.1016/j.nima.2010.02.190](https://doi.org/10.1016/j.nima.2010.02.190).
- [10] Rob Bugiel et al. "Test-beam results of a SOI pixel-detector prototype". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 901 (June 1, 2018). DOI: [10.1016/j.nima.2018.06.017](https://doi.org/10.1016/j.nima.2018.06.017).

- [11] R Turchetta et al. "A monolithic active pixel sensor for charged particle tracking and imaging using standard VLSI CMOS technology". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 458.3 (Feb. 11, 2001), pp. 677–689. ISSN: 0168-9002. DOI: [10.1016/S0168-9002\(00\)00893-7](https://doi.org/10.1016/S0168-9002(00)00893-7). URL: <https://www.sciencedirect.com/science/article/pii/S0168900200008937> (visited on 12/16/2021).
- [12] Rebecca E. Coath et al. "A Low Noise Pixel Architecture for Scientific CMOS Monolithic Active Pixel Sensors". In: *IEEE Transactions on Nuclear Science* 57.5 (Oct. 2010), pp. 2490–2496. ISSN: 1558-1578. DOI: [10.1109/TNS.2010.2052469](https://doi.org/10.1109/TNS.2010.2052469).
- [13] M. Mager. "ALPIDE, the Monolithic Active Pixel Sensor for the ALICE ITS upgrade". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors 824 (July 11, 2016), pp. 434–438. ISSN: 0168-9002. DOI: [10.1016/j.nima.2015.09.057](https://doi.org/10.1016/j.nima.2015.09.057). URL: <https://www.sciencedirect.com/science/article/pii/S0168900215011122> (visited on 11/23/2021).
- [14] G. Aglieri Rinella et al. "Charge collection properties of TowerJazz 180 nm CMOS Pixel Sensors in dependence of pixel geometries and bias parameters, studied using a dedicated test-vehicle: the Investigator chip". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 988 (Feb. 2021), p. 164859. ISSN: 01689002. DOI: [10.1016/j.nima.2020.164859](https://doi.org/10.1016/j.nima.2020.164859). arXiv: [2009.10517](https://arxiv.org/abs/2009.10517). URL: <http://arxiv.org/abs/2009.10517> (visited on 03/23/2022).
- [15] W. Snoeys et al. "A process modification for CMOS monolithic active pixel sensors for enhanced depletion, timing performance and radiation tolerance". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 871 (Nov. 1, 2017), pp. 90–96. ISSN: 0168-9002. DOI: [10.1016/j.nima.2017.07.046](https://doi.org/10.1016/j.nima.2017.07.046). URL: <https://www.sciencedirect.com/science/article/pii/S016890021730791X> (visited on 08/29/2024).
- [16] P. Vicente Leitao et al. "Development of a Stitched Monolithic Pixel Sensor prototype (MOSS chip) towards the ITS3 upgrade of the ALICE Inner Tracking system". In: *Journal of Instrumentation* 18.1 (Jan. 1, 2023), p. C01044. ISSN: 1748-0221. DOI: [10.1088/1748-0221/18/01/C01044](https://doi.org/10.1088/1748-0221/18/01/C01044). URL: <https://iopscience.iop.org/article/10.1088/1748-0221/18/01/C01044> (visited on 01/31/2023).
- [17] K. Aamodt and others. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: [10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- [18] P. Dorosz, on behalf of the MOSIAX design team, and the ALICE collaboration. "Data transmission architecture of the ALICE ITS3 stitched sensor prototype MOSAIX". In: *Journal of Instrumentation* 19.4 (Apr. 2024). Publisher: IOP Publishing, p. C04050. ISSN: 1748-0221. DOI: [10.1088/1748-0221/19/04/C04050](https://doi.org/10.1088/1748-0221/19/04/C04050). URL: <https://dx.doi.org/10.1088/1748-0221/19/04/C04050> (visited on 09/06/2024).
- [19] L. Palomo. "The ALICE experiment upgrades for LHC Run 3 and beyond: contributions from mexican groups". In: *Journal of Physics: Conference Series* 912 (Oct. 1, 2017), p. 012023. DOI: [10.1088/1742-6596/912/1/012023](https://doi.org/10.1088/1742-6596/912/1/012023).
- [20] S. Chatrchyan and others. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).

- [21] Serguei Chatrchyan and others. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014). _eprint: 1405.6569, P10009. DOI: [10.1088/1748-0221/9/10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009).
- [22] G. Aad and others. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [23] Laura Gonella. "The ATLAS ITk detector system for the Phase-II LHC upgrade". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1045 (Jan. 1, 2023), p. 167597. ISSN: 0168-9002. DOI: [10.1016/j.nima.2022.167597](https://doi.org/10.1016/j.nima.2022.167597). URL: <https://www.sciencedirect.com/science/article/pii/S0168900222008890> (visited on 09/06/2024).
- [24] S. Gambetta. "The LHCb RICH detectors: Operations and performance". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 10th International Workshop on Ring Imaging Cherenkov Detectors (RICH 2018) 952 (Feb. 1, 2020), p. 161882. ISSN: 0168-9002. DOI: [10.1016/j.nima.2019.02.009](https://doi.org/10.1016/j.nima.2019.02.009). URL: <https://www.sciencedirect.com/science/article/pii/S0168900219301846> (visited on 08/13/2024).
- [25] Yiming Li. "MAPS for the Upstream Tracker in LHCb Upgrade II". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1032 (June 1, 2022), p. 166629. ISSN: 0168-9002. DOI: [10.1016/j.nima.2022.166629](https://doi.org/10.1016/j.nima.2022.166629). URL: <https://www.sciencedirect.com/science/article/pii/S0168900222002017> (visited on 09/06/2024).
- [26] M. Deveaux et al. "Observations on MIMOSIS-0, the first dedicated CPS prototype for the CBM MVD". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. Proceedings of the Vienna Conference on Instrumentation 2019 958 (Apr. 1, 2020), p. 162653. ISSN: 0168-9002. DOI: [10.1016/j.nima.2019.162653](https://doi.org/10.1016/j.nima.2019.162653). URL: <https://www.sciencedirect.com/science/article/pii/S0168900219311404> (visited on 06/06/2021).
- [27] Michael Deveaux. "Development of fast and radiation hard Monolithic Active Pixel Sensors (MAPS) optimized for open charm meson detection >with the CBM - vertex detector". PhD thesis. Université Louis Pasteur - Strasbourg I, Mar. 20, 2008. URL: <https://theses.hal.science/tel-00392111> (visited on 09/06/2024).
- [28] T. Fillinger. "Detector Simulation for a Potential Upgrade of the Vertex Detector of the Belle II Experiment". In: *Acta Physica Polonica B* 52.8 (2021), p. 909. ISSN: 0587-4254, 1509-5770. DOI: [10.5506/APhysPolB.52.909](https://doi.org/10.5506/APhysPolB.52.909). URL: <http://www.actaphys.uj.edu.pl/findarticle?series=Reg&vol=52&page=909> (visited on 03/23/2022).
- [29] Christian Wessel and Belle II VTX collaboration. *CMOS MAPS Upgrade for the Belle II Vertex Detector*. Pisa Meeting on Advanced Detectors, 2022.
- [30] Zdenek Vostrel and Steffen Doebert. "Design of an electron source for the FCC-ee with top-up injection capability". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1063 (June 1, 2024), p. 169261. ISSN: 0168-9002. DOI: [10.1016/j.nima.2024.169261](https://doi.org/10.1016/j.nima.2024.169261).

- j.nima.2024.169261. URL: <https://www.sciencedirect.com/science/article/pii/S0168900224001876> (visited on 08/13/2024).
- [31] Gabriele D'Amen et al. "Neutron detection with fast-timing LGAD". In: *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). Oct. 2019, pp. 1–4. DOI: [10.1109/NSS/MIC42101.2019.9060016](https://doi.org/10.1109/NSS/MIC42101.2019.9060016).
 - [32] Yue Pan et al. "Fast Dark Signal Measurements of SVOM VT CCDs Using the Vertical Gradient of Dark Field Images". In: *Photonics* 8 (Apr. 20, 2021), p. 132. DOI: [10.3390/photonics8040132](https://doi.org/10.3390/photonics8040132).
 - [33] G. Deptuch et al. "Monolithic active pixel sensors with in-pixel double sampling operation and column-level discrimination". In: *IEEE Transactions on Nuclear Science* 51.5 (Oct. 2004). Conference Name: IEEE Transactions on Nuclear Science, pp. 2313–2321. ISSN: 1558-1578. DOI: [10.1109/TNS.2004.835551](https://doi.org/10.1109/TNS.2004.835551). URL: <https://ieeexplore.ieee.org/document/1344330> (visited on 08/13/2024).
 - [34] I. Valin et al. "A reticle size CMOS pixel sensor dedicated to the STAR HFT". In: *Journal of Instrumentation* 7.1 (Jan. 2012), p. C01102. ISSN: 1748-0221. DOI: [10.1088/1748-0221/7/01/C01102](https://doi.org/10.1088/1748-0221/7/01/C01102). URL: <https://dx.doi.org/10.1088/1748-0221/7/01/C01102> (visited on 08/13/2024).
 - [35] C. Hu-Guo et al. "First reticle size MAPS with digital output and integrated zero suppression for the EUDET-JRA1 beam telescope". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 1st International Conference on Technology and Instrumentation in Particle Physics 623.1 (Nov. 1, 2010), pp. 480–482. ISSN: 0168-9002. DOI: [10.1016/j.nima.2010.03.043](https://doi.org/10.1016/j.nima.2010.03.043). URL: <https://www.sciencedirect.com/science/article/pii/S0168900210006078> (visited on 08/13/2024).
 - [36] J. Baudot et al. "First test results Of MIMOSA-26, a fast CMOS sensor with integrated zero suppression and digitized output". In: *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*. 2009 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC 2009). event-place: Orlando, FL, USA. IEEE, Oct. 2009, pp. 1169–1173. ISBN: 978-1-4244-3961-4. DOI: [10.1109/NSSMIC.2009.5402399](https://doi.org/10.1109/NSSMIC.2009.5402399). URL: <https://ieeexplore.ieee.org/document/5402399/> (visited on 11/23/2021).
 - [37] Giacomo Contin et al. "The STAR MAPS-based PiXeL detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. Advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn) 907 (Nov. 1, 2018), pp. 60–80. ISSN: 0168-9002. DOI: [10.1016/j.nima.2018.03.003](https://doi.org/10.1016/j.nima.2018.03.003). URL: <https://www.sciencedirect.com/science/article/pii/S0168900218303206> (visited on 08/13/2024).
 - [38] Giacomo Contin. "The MAPS-based vertex detector for the STAR experiment: Lessons learned and performance". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. Proceedings of the 10th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors 831 (Sept. 21, 2016), pp. 7–11. ISSN: 0168-9002. DOI: [10.1016/j.nima.2016.04.109](https://doi.org/10.1016/j.nima.2016.04.109). URL: <https://www.sciencedirect.com/science/article/pii/S0168900216303539> (visited on 10/18/2021).

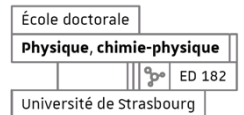
- [39] Gianluca Aglieri Rinella. “The ALPIDE pixel sensor chip for the upgrade of the ALICE Inner Tracking System”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. Proceedings of the Vienna Conference on Instrumentation 2016 845 (Feb. 11, 2017), pp. 583–587. ISSN: 0168-9002. DOI: [10.1016/j.nima.2016.05.016](https://doi.org/10.1016/j.nima.2016.05.016). URL: <https://www.sciencedirect.com/science/article/pii/S0168900216303825> (visited on 03/23/2022).
- [40] “The MIMOSIS pixel sensor”. In: ().
- [41] M. Backhaus et al. “Development of a versatile and modular test system for ATLAS hybrid pixel detectors”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. International Workshop on Semiconductor Pixel Detectors for Particles and Imaging 2010 650.1 (Sept. 11, 2011), pp. 37–40. ISSN: 0168-9002. DOI: [10.1016/j.nima.2010.12.087](https://doi.org/10.1016/j.nima.2010.12.087). URL: <https://www.sciencedirect.com/science/article/pii/S0168900210028676> (visited on 08/13/2024).
- [42] Ivan Caicedo et al. “The Monopix chips: Depleted monolithic active pixel sensors with a column-drain read-out architecture for the ATLAS Inner Tracker upgrade”. In: *Journal of Instrumentation* 14.6 (June 5, 2019), pp. C06006–C06006. ISSN: 1748-0221. DOI: [10.1088/1748-0221/14/06/C06006](https://doi.org/10.1088/1748-0221/14/06/C06006). arXiv: [1902.03679\[physics\]](https://arxiv.org/abs/1902.03679). URL: <http://arxiv.org/abs/1902.03679> (visited on 08/13/2024).
- [43] M. Babeluk et al. “The OBELIX chip for the Belle II VTX upgrade”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1067 (Oct. 1, 2024), p. 169659. ISSN: 0168-9002. DOI: [10.1016/j.nima.2024.169659](https://doi.org/10.1016/j.nima.2024.169659). URL: <https://www.sciencedirect.com/science/article/pii/S0168900224005850> (visited on 08/20/2024).
- [44] R. Cardella and others. “MALTA: an asynchronous readout CMOS monolithic pixel detector for the ATLAS High-Luminosity upgrade”. In: *JINST* 14.6 (2019), p. C06019. DOI: [10.1088/1748-0221/14/06/C06019](https://doi.org/10.1088/1748-0221/14/06/C06019).
- [45] Chun-Zheng Wang. *The ITS3 detector and physics reach of the LS3 ALICE Upgrade*. version: 1. Sept. 3, 2024. DOI: [10.48550/arXiv.2409.01866](https://doi.org/10.48550/arXiv.2409.01866). arXiv: [2409.01866\[hep-ex, physics:nucl-ex, physics:physics\]](https://arxiv.org/abs/2409.01866). URL: <http://arxiv.org/abs/2409.01866> (visited on 09/06/2024).
- [46] Enagnon Aguénonon et al. “Design and Characterization of an Asynchronous Fixed Priority Tree Arbiter for SPAD Array Readout”. In: *Sensors* 21 (June 8, 2021). DOI: [10.3390/s21123949](https://doi.org/10.3390/s21123949).
- [47] Timothé Turko et al. “An Asynchronous Fixed Priority Arbiter for High Throughput Time Correlated Single Photon Counting Systems”. In: *International Conference on Electronics, Circuits, and Systems (ICEC 2018)*. event-place: Bordeaux, France. Dec. 2018. URL: <https://hal.archives-ouvertes.fr/hal-01971018> (visited on 05/23/2022).
- [48] Tarek Al Abbas et al. “A CMOS SPAD Sensor With a Multi-Event Folded Flash Time-to-Digital Converter for Ultra-Fast Optical Transient Capture”. In: *IEEE Sensors Journal* 18.8 (Apr. 2018), pp. 3163–3173. ISSN: 1558-1748. DOI: [10.1109/JSEN.2018.2803087](https://doi.org/10.1109/JSEN.2018.2803087).

- [49] Matheus T. Moreira, Julian J. H. Pontes, and Ney L. V. Calazans. "Tradeoffs between RTO and RTZ in WCHB QDI asynchronous design". In: *Fifteenth International Symposium on Quality Electronic Design*. Fifteenth International Symposium on Quality Electronic Design. Mar. 2014, pp. 692–699. DOI: [10.1109/ISQED.2014.6783394](https://doi.org/10.1109/ISQED.2014.6783394).
- [50] Grégoire Gimenez et al. "Static Timing Analysis of Asynchronous Bundled-Data Circuits". In: *2018 24th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*. 2018 24th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC). May 2018, pp. 110–118. DOI: [10.1109/ASYNC.2018.00036](https://doi.org/10.1109/ASYNC.2018.00036).
- [51] Ying-Haw Shu et al. "XNOR-based double-edge-triggered flip-flop for two-phase pipelines". In: *Circuits and Systems II: Express Briefs, IEEE Transactions on* 53 (Mar. 1, 2006), pp. 138–142. DOI: [10.1109/TCSII.2005.855734](https://doi.org/10.1109/TCSII.2005.855734).
- [52] Ad Peeters et al. "Click Elements: An Implementation Style for Data-Driven Compilation". In: *2010 IEEE Symposium on Asynchronous Circuits and Systems*. 2010 IEEE Symposium on Asynchronous Circuits and Systems. May 2010, pp. 3–14. DOI: [10.1109/ASYNC.2010.11](https://doi.org/10.1109/ASYNC.2010.11).
- [53] Hui Wu et al. "A Design Flow for Click-Based Asynchronous Circuits Design With Conventional EDA Tools". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 40.11 (Nov. 2021), pp. 2421–2425. ISSN: 1937-4151. DOI: [10.1109/TCAD.2020.3038337](https://doi.org/10.1109/TCAD.2020.3038337).
- [54] Jordi Cortadella et al. "Narrowing the margins with elastic clocks". In: *2010 IEEE International Conference on Integrated Circuit Design and Technology*. 2010 IEEE International Conference on Integrated Circuit Design and Technology. June 2010, pp. 146–150. DOI: [10.1109/ICICDT.2010.5510273](https://doi.org/10.1109/ICICDT.2010.5510273).
- [55] Jean Simatic et al. "A Practical Framework for Specification, Verification, and Design of Self-Timed Pipelines". In: *2017 23rd IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*. 2017 23rd IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC). May 2017, pp. 65–72. DOI: [10.1109/ASYNC.2017.16](https://doi.org/10.1109/ASYNC.2017.16).
- [56] Zhiyu Li et al. "A Low-Power Asynchronous RISC-V Processor With Propagated Timing Constraints Method". In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 68.9 (Sept. 2021), pp. 3153–3157. ISSN: 1558-3791. DOI: [10.1109/TCSII.2021.3100524](https://doi.org/10.1109/TCSII.2021.3100524).
- [57] Jean Soudier. "New performance for MAPS readout using an asynchronous architecture based on priority arbiters". *Speak. Speak. FBK, Trento*, Mar. 2, 2023. URL: <https://indico.cern.ch/event/1223972/contributions/5262060/>.
- [58] Jean Soudier et al. "A versatile and fast pixel matrix read-out architecture for MAPS". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1067 (Oct. 1, 2024), p. 169663. ISSN: 0168-9002. DOI: [10.1016/j.nima.2024.169663](https://doi.org/10.1016/j.nima.2024.169663). URL: <https://www.sciencedirect.com/science/article/pii/S0168900224005898> (visited on 08/14/2024).

- [59] A. Kluge. “ALICE - ITS3 — A bent, wafer-scale CMOS detector”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1041 (Oct. 11, 2022), p. 167315. ISSN: 0168-9002. DOI: [10.1016/j.nima.2022.167315](https://doi.org/10.1016/j.nima.2022.167315). URL: <https://www.sciencedirect.com/science/article/pii/S0168900222006386> (visited on 09/06/2024).
- [60] W. Snoeys. “Monolithic CMOS sensors for high energy physics”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 924 (Apr. 2019), pp. 51–58. ISSN: 01689002. DOI: [10.1016/j.nima.2018.06.034](https://doi.org/10.1016/j.nima.2018.06.034). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0168900218307551> (visited on 05/05/2023).
- [61] Gianluca Aglieri Rinella et al. “Digital Pixel Test Structures implemented in a 65 nm CMOS process”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1056 (Nov. 2023), p. 168589. ISSN: 01689002. DOI: [10.1016/j.nima.2023.168589](https://doi.org/10.1016/j.nima.2023.168589). arXiv: [2212.08621\[physics\]](https://arxiv.org/abs/2212.08621). URL: <http://arxiv.org/abs/2212.08621> (visited on 03/18/2024).

Jean SOUDIER

Etude d'architectures de lecture asynchrone intégrée pour capteurs à pixels CMOS



Résumé

Les capteurs monolithiques à pixels CMOS sont exploités pour la trajectométrie des particules chargées en physique subatomique. Ils offrent une excellente résolution spatiale et un traitement rapide de l'information. L'évolution des expériences exige une augmentation de la vitesse de traitement tout en imposant des limites sur la consommation des capteurs. Cette thèse développe une architecture de lecture d'une matrice de pixels exploitant une logique asynchrone avec une réaction rapide et une dissipation minimale d'énergie. L'algorithme s'appuie sur la cascade de contrôleurs asynchrones qui arbitrent la transmission du signal de N entrées vers une seule sortie. Plusieurs configurations sont étudiées pour des pixels de taille variant de 18 à 30 μm et des contrôleurs de taille variant de 2 :1 à 1024 :1. Les simulations sur des signaux physiques réalistes correspondant à des taux de 100 MHz/cm² ou plus, démontrent la possibilité de reconstruire les positions sur la matrice et les temps d'arrivée en moins de 20 nanosecondes pour une consommation inférieure à 10 mW/cm². La conception d'un premier prototype est également présentée.

Résumé en anglais

Monolithic CMOS pixel sensors are used for the tracking of charged particles in subatomic physics. They offer excellent spatial resolution and rapid information processing. The evolution of experiments requires an increase in processing speed while imposing limits on sensor power consumption. This thesis develops an architecture for reading an array of pixels using asynchronous logic with rapid response and minimal energy dissipation. The algorithm is based on a cascade of asynchronous controllers that arbitrate signal transmission from N inputs to a single output. Several configurations are studied for pixel sizes ranging from 18 to 30 μm and controller sizes ranging from 2:1 to 1024:1. Simulations on realistic physical signals corresponding to rates of 100 MHz/cm² or more, demonstrate the possibility of reconstructing positions on the matrix and arrival times in less than 20 nanoseconds for a power consumption of less than 10 mW/cm². The design of an initial prototype is also presented.