



UNIVERSITÀ DI PISA

Department of Physics "E. Fermi"

Master's Degree in Physics

Study of monolithic CMOS pixel sensors in the Belle II experiment upgrade

Candidate:

Mara Stefania Caló

Supervisor:

Prof. Francesco Forti

Academic year 2022/2023

Contents

Introduction	5
1 Belle II and SuperKEKB (SKB) accelerator	8
1.1 Physics program of the B-factories	8
1.1.1 Open questions in the Standard Model	9
1.1.2 Peculiarity of asymmetric B factories	9
1.2 SuperKEKB accelerator	11
1.2.1 The facility	11
1.2.2 Nano-beam scheme	12
1.3 Belle II detector	13
1.3.1 Vertex Detector (VXD)	14
1.3.2 Central Drift Chamber (CDC)	14
1.3.3 Particle identification system (TOP e ARICH)	16
1.3.4 Electromagnetic calorimeter (ECL)	16
1.3.5 K_L Muon detector (KLM)	17
1.3.6 Trigger system	18
1.4 Current state of data taking	18
2 Belle II Upgrade	20
2.1 Purposes of the upgrade	20
2.2 Background sources and limitations in Belle II	20

2.2.1	Main background sources	21
2.2.2	Current background status and future predictions	22
2.3	Summary of possible VXD upgrade	23
2.3.1	Depleted Field Effect Transistor (DEPFET)	24
2.3.2	Thin and Fine-Pitch SVD	25
2.3.3	CMOS Monolithic Active Pixels Sensor	25
2.3.4	Silicon On Insulator (SOI)	26
3	CMOS MAPS sensors	27
3.1	Semiconductor detectors	27
3.1.1	Transport of charge carriers and signal formation in semiconductors	28
3.1.2	The pn junctions as detector	28
3.2	Hybrid and monolithic pixel sensors	29
3.2.1	Hybrid pixel detectors	30
3.2.2	Monolithic pixel detectors	31
3.3	CMOS Monolithic Active Pixel Sensors technology	32
3.3.1	MAPS pixel detectors	32
3.3.2	Depleted MAPS pixel detectors	33
3.3.3	Silicon On Insulator (SOI) technology	35
3.4	History of Monopix developments	35
3.4.1	TJ-Monopix line	36
3.4.2	TJ-Monopix2 architecture	37
4	VTX detector	42
4.1	VTX Layout and mechanical structure	42
4.1.1	iVTX	43
4.1.2	oVTX	45
4.2	Performance simulation	47
4.2.1	VTX geometries	47
4.2.2	Tracking efficiency at low momentum and impact parameter resolution	48
4.2.3	Vertexing resolution	48

4.3	OBELIX chip design	52
4.3.1	Sensor specification	52
4.3.2	Sensor implementation	53
5	TJ-Monopix2 characterization	56
5.1	Matrix and flavors	57
5.1.1	Flavors	57
5.1.2	Pixel design	58
5.2	Threshold and noise	59
5.2.1	Injection circuit	61
5.2.2	S-Curve method	61
5.2.3	Threshold, noise ad threshold dispersion results (Test Beam settings) . . .	68
5.3	TOT calibration with internal injection	68
5.3.1	TOT curves and fit	69
5.3.2	Comparison	72
5.4	Response to radioactive source and absolute calibration	75
5.4.1	^{55}Fe	76
5.4.2	^{241}Am	77
5.4.3	^{109}Cd	78
5.4.4	Conversion factor K and C_{inj} absolute calibration	80
5.4.5	Check on calibration curve TOT vs Q_{inj} with radioactive sources	80
5.5	Operation with low threshold	84
5.5.1	Registers optimization	84
5.5.2	Comparison between data and simulation	84
5.5.3	Low threshold operation, threshold dispersion and tuning	88
5.5.4	First results from threshold tuning	89
5.6	Hot pixels from cross talk issue and mitigation	91
5.6.1	Hot pixel studies	92
5.6.2	Digital signal responsible for cross talk	93
5.6.3	Cross talk signal height	96

5.6.4	Cross talk mitigation	98
5.6.5	Conclusion from cross talk studies	99
5.7	Test Beam results	100
6	Conclusions	103
	Bibliography	104

Introduction

Belle II is a particle physics experiment located at the KEK laboratory in Tsukuba (Japan). The detector is a general-purpose spectrometer to study electron-positron collisions produced by the SuperKEKB accelerator, a second generation flavor-factory which operates at the luminosity frontier, holding the world record of instantaneous luminosity with $L_{peak} = 4.7 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$.

SuperKEKB is the upgrade of the preceding facility KEKB (operational from 1998 to 2010) and it consists in a 3 km-circumference asymmetric accelerator which collides electrons and positrons beams at a center-of-mass energy near the $\Upsilon(4S)$ resonance ($\sqrt{s} = 10.58 \text{ GeV}$). It started its data taking in March 2019.

In the next decade, the collider aims to collect an unrivaled dataset of 50 ab^{-1} (x50 Belle dataset, x100 BaBar dataset) and to reach a peak luminosity of $6 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$. This will allow to study the charge-parity violation in B mesons system with more precision and to search for new hints of physics beyond the Standard Model.

To achieve these challenging targets, it will be necessary a significant upgrade of the accelerator and its main components (like the injection system and the equipment nearby the interaction region), probably requiring the installation of a new vertex detector. As a matter of fact, to the increase in luminosity corresponds not only large data collected and greater possibility to study rare processes, but also higher doses of radiation and larger backgrounds, which could undermine the integrity and the operation of the Belle II detector. In particular the subdetectors which are closest to the beam pipe are those more exposed to severe conditions, like the vertex detector (VXD), composed of the inner pixel detector (PXD, made of layers of pixels) and the outermost silicon vertex detector (SVD, made of layers of strips). They allow the reconstruction of charged particle tracks and of decay vertices with high performance. Recent studies have shown that the current detector could operate efficiently up to a luminosity of $L_{inst} = 2 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$, but safety margins are not large. Consequently, in this context, different upgrade projects have been proposed, which intend to design a new vertex detector, making it more resistant even in harsher working conditions, while the luminosity will be gradually increased.

This work focuses especially on the VerTeX Detector (VTX) proposal (the one chosen for the upgrade), replacing the whole VXD with five layers of fully pixelated sensors based on the CMOS Depleted Monolithic Active Pixel Sensor (DMAPS) technology.

The good results achieved by the ALICE experiment (LHC, CERN), which employed the same technology, have suggested this solution which has proven to be reliable. The current developments, aimed at making the chips much faster than in ALICE, are promising in maintaining low occupancy, despite the worse expected background environment, and good performance even after irradiation.

In order to fulfill the physics requirements of Belle II experiment, a new silicon sensor is being designed, called OBELIX, fabricated using the 180 nm TowerJazz Semiconductor process. Developments will ensure a faster, lighter and highly granular chip, reducing the material budget and improving tracks and vertices reconstruction.

OBELIX planning is based on studies done on previous prototypes, among which TJ-Monopix 2, whose characterization is the main topic of this thesis. Laboratory and beam tests have been conducted and are still in progress, in order to study the efficiency of the chip before and after irradiation, its power consumption, and to fully characterize its electrical characteristics. In particular, we have characterized the response of the pixel matrix, extracting important results that have allowed to interpret data taken during the Test Beam at Desy (June 2022), and that are being used in the design of the OBELIX chip. In more detail, the threshold distributions for all the different types of front-end circuits implemented in the matrix have been studied, together with their dispersion and noise distributions. The calibration of the Time Over Threshold curves (which is a time width signal processing method used in this prototype) has been done by internal injection tests. The absolute calibration of the whole matrix has been achieved, employing a ^{55}Fe radioactive source. Other radioactive sources have been used too, in order to check the trend of the TOT curves for charge values not accessible by internal injection. Additionally, different register settings have been examined in the interest of operating the matrix at low threshold, that is crucial to keep high efficiency even after irradiation. For this reason, several tests have been conducted to tune the threshold, in order to reduce the dispersion and make the threshold on the matrix as uniform as possible. During this investigation, a cross-talk issue has been discovered and therefore studied to understand its causes and possible solutions to mitigate this effect.

The measurements performed in this thesis have been used for the design of OBELIX chip, that should be submitted for fabrication in the next few months.

Chapter 1 briefly introduces some of the open questions in the Standard Model, in order to depict the background of the Belle II physics program. A short description of the SuperKEKB accelerator and Belle II detector is also given.

Chapter 2 presents the main arguments supporting the Belle II upgrade program. The primary sources of the experiment background are summarized, to understand the limitation of the detector and the accelerator, for increasingly higher luminosity values. Eventually a summary of the four main upgrade proposals for the vertex detector is presented, which are distinguished by the different type of sensors employed: Depleted Field Effect Transistor (DEPFET), Thin and Fine-Pitch strip detectors, Silicon On Insulator (SOI) and CMOS Monolithic Active Pixels Sensors.

Chapter 3 describes the principles underlying the operation of semiconductor detectors and some different type of sensors which use this technology, like the hybrid and monolithic pixel sensors. In particular the CMOS Monolithic Active Pixel Sensors technology is presented, on which the entire developments of the OBELIX chip is based. In the end, the history of the developments that led to the TJ-Monopix chip series is retraced, in order to better understand the main features of the last one, TJ-Monopix 2, which represents the starting point for OBELIX design, and whose characterization is the work of this thesis.

Chapter 4 examines in depth the VerTeX detector (VTX) upgrade program, which involves the CMOS Monolithic Active Pixel Sensors as fundamental components of the five layers of the final vertex detector. Studies and simulations are ongoing to test the performance, and some of them are shown here. The specifications and the implementation of the new chip (OBELIX) under design for this proposal are described. The innovative sensor has to fulfill the requirements of Belle II experiment, even in extreme environment due to higher doses of radiation and backgrounds.

Chapter 5 lastly shows the results obtained from laboratory measurements and tests conducted on the TJ-Monopix 2 chip. The response of the matrix has been studied in different working conditions, in order to analyze the behaviour at high and low threshold. The absolute calibration of the all front-end circuits implemented in the chip, has been done too. Moreover a cross-talk issue have been discovered and analyzed, in order to understand its causes and a

possible mitigation of this effect since it prevented from using the matrix at low threshold.

1. Belle II and SuperKEKB (SKB) accelerator

The first chapter introduces some of the main aspects of the Standard Model (SM), along with its open questions, on which the Belle II physics program is focused. A short description of the SuperKEKB accelerator and the Belle II detector's structure is also presented and in conclusion some highlights on the current state of measurements are shown.

1.1 Physics program of the B-factories

The SM is a very successful theory describing three of the fundamental forces involving elementary particles: the strong, weak and electromagnetic interactions, with the exclusion of the gravitational force. It classifies all the elementary constituents of matter in 4 main groups, as shown in Figure 1.1: quarks and leptons, constituting the matter fields; the gauge bosons, representing the interactions; and the Higgs boson, whose non-zero vacuum expectation value is needed to give mass to the otherwise massless matter fields.



Figure 1.1: Particle classification in the Standard Model.

1.1.1 Open questions in the Standard Model

Despite the undeniable success of the SM in providing predictions for all known physics phenomena, which have been experimentally verified with high precision over the years, there are many fundamental aspects of nature on which it is unable to give answers [1]. Some of them are listed in the following.

- Three generations of quarks and leptons have been discovered, but it is not known whether they are the only ones and why.
- The reasons behind the large differences in the masses of quarks and leptons (mass hierarchy) are unknown.
- Although the Higgs mechanism is able to explain the cause of elementary particles' masses through spontaneous electro-weak symmetry breaking, it is not clear whether neutrinos can get their non-zero but very small masses through the interaction with the Higgs boson.
- Another open question is the matter-antimatter asymmetry in the universe. Charge-Parity (CP) violation is necessary to explain the asymmetry, but the SM mechanism would predict a value several orders of magnitude smaller than what is needed to explain the matter domination over antimatter, which allowed the evolution of the universe as we know it today.
- The flavour-changing currents are phenomenologically described with two mixing matrices: the Cabibbo-Kobayashi-Maskawa (CKM) matrix for the quarks and the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix for the neutrinos. The SM fails to provide any explanation neither for the structure of the matrices (only charged for the quarks, only neutral for the neutrinos), nor for the values of their elements (nearly diagonal for the quarks, highly non diagonal for the neutrinos).
- Astrophysical observations indicate the existence of dark matter, but its origin and nature have not been explained yet.

All these open questions stimulate the search for new particles and processes that could provide more fundamental explanations.

At the energy frontier, experiments at the Large Hadron Collider (LHC) at CERN (CH) are looking for new particles created from the proton-proton collision with a center-of-mass energy up to 14 TeV.

At the luminosity frontier, instead, new particles and interactions are searched through precision measurements of suppressed reactions or deviations from the SM. The discrepancies could be interpreted as a clue of new physics beyond the SM. The Belle II experiment at the SuperKEKB B-Factory is following this last approach.

In particular, the experiment investigates the CP violation in the B-mesons system and searches for new physics evidence in the decays of B and D mesons, in τ leptons and in the dark matter sector (DM).

1.1.2 Peculiarity of asymmetric B factories

The SuperKEKB e^+e^- collider operates with a center-of-mass energy of $\sqrt{s} = 10.58$ GeV at the $\Upsilon(4S)$ resonance, which decays almost instantaneously into a pair of B - anti B mesons in nearly 96% of cases.

In SuperKEKB the beam energies are different, leading to a center-of-mass boost in the laboratory that allows the reconstruction of the B-mesons decay vertices, their lifetimes, and the time-dependent decay rate asymmetries.

In a symmetric beams situation, the B mesons would be produced almost at rest, decaying roughly at the same point with undetectable decay length. The investigation of CP violating processes instead, requires measuring the decay time difference of the two B mesons and its uncertainty is dominated by that of the decay vertex measurement (order of hundreds microns), as we will now see in more detail.

SuperKEKB collides an electrons beam of 7 GeV (High Energy Ring, HER) with a positrons beam of 4 GeV (Low Energy Ring, LER) with a Lorentz boost factor of the $\Upsilon(4S)$ of $(\beta\gamma)_{\Upsilon(4S)} \approx 0.28$.

The same boost is also acquired by the B-mesons, because they are produced almost at rest ($m_{\Upsilon(4S)} - m_{2B_0} \approx 19$ MeV). Moreover knowing that $\tau_B \simeq 1.5 \times 10^{-12}$ s and so $c\tau_B \simeq 450$ μm , we can compute the average flight distance travelled before decaying:

$$l = (\beta\gamma)_{\Upsilon(4S)} c\tau_B \approx 126 \mu\text{m} \quad (1.1)$$

This value must be within the vertex detector sensitivity in order to distinguish the vertex decay and as consequence to make precision measurements of lifetimes, mixing parameters and CP violation. The main task of the VerTeX Detector (VXD) is to reconstruct the production and decay vertices of the particles originated from the beam collisions. This aspect is crucial to perform time-dependent measurements, core of the Belle II physics program [1]. The six-layer VXD, discussed in section 1.3.1, can determine the position of the vertices with a precision better than 100 μm [2], allowing to reconstruct secondary vertices, i.e. the decay position of the particles coming from B decays, and also from τ leptons and D mesons.

The event kinematics is illustrated in Figure 1.2. The two B mesons are produced in an entangled quantum state, so when from the decay products of one of the two it is possible to assign its flavor (for example B^0 , identified as B_{tag}^0) accordingly can be assigned that of the other, which will be the opposite (\bar{B}^0 , called \bar{B}_{phys}^0).

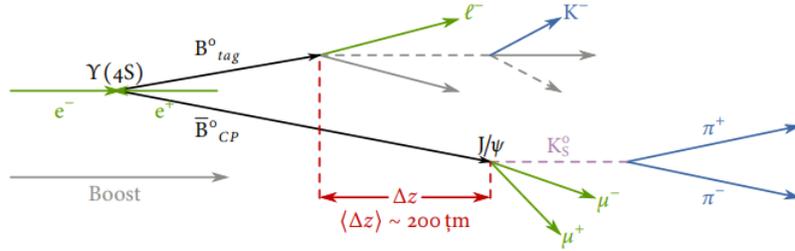


Figure 1.2: Example of the kinematics of the golden channel of Belle II experiment.

After this reconstruction, both B decay vertex positions in the longitudinal direction z_1 and z_2 are evaluated, in order to compute their difference:

$$\Delta z = z_1 - z_2 = (\beta\gamma)_{\Upsilon(4S)} c\Delta t \quad (1.2)$$

where Δt is the proper time decay difference.

Therefore this topology allows to transform a temporal information in a spatial one that we are

able to measure using a high precision vertex detector. Without the boosted center of mass none of it could be possible, and this is an essential feature for an asymmetric B-factory.

1.2 SuperKEKB accelerator

Belle II sensitivity in the precision measurements is feasible especially thanks to the extraordinary performance of the SuperKEKB accelerator which host the (almost) hermetic detector. This complex facility is the result of efforts and efficient collaboration between the researches of KEK laboratory and all the international working groups that participate to the experiment.

1.2.1 The facility

SuperKEKB[3, 4] (Figure 1.3) is an asymmetric e^+e^- collider with a circumference of 3 km and a center-of-mass energy peak equal to $\sqrt{s} = 10.58 \text{ GeV}$, which corresponds to the mass of the $\Upsilon(4S)$ resonance. Compared to its predecessor KEKB (which started its operation in 1998 and concluded it in 2010[5], reaching a peak luminosity of $2.11 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$), the current accelerator has a target luminosity of $6 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$, and already achieved a record luminosity of $4.7 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ in July 2022 [6]. This target is possible using a new scheme to accelerate and collide the beams, the so called *nano-beam scheme* (section 1.2.2).

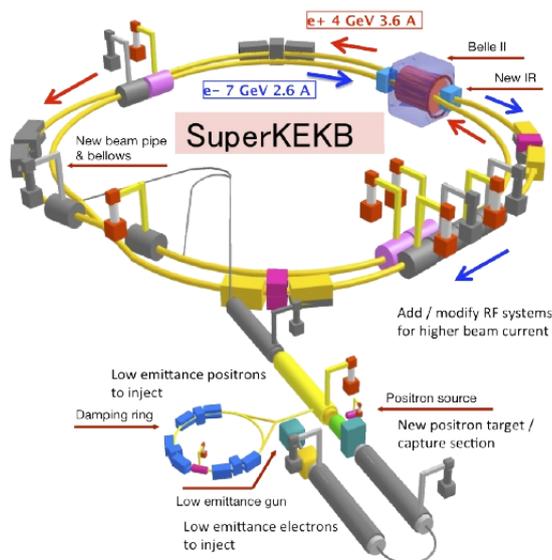


Figure 1.3: SuperKEKB accelerator structure.

Luminosity

Instantaneous luminosity is one of the key parameters of any accelerator and it represents the interaction rate per unit of cross section between colliding particles (equation 1.3). Inverting this equation is possible to obtain N , namely the number of the physical events produced in the interaction with a given luminosity:

$$L = \frac{1}{\sigma} \frac{dN}{dt} \quad \Rightarrow \quad N = \int_0^T L \sigma dt \quad (1.3)$$

where T is the duration of the experiment and σ the cross section of the physical process of interest. Luminosity is dependent from both machine and beam parameters. With respect to this, it can be expressed as:

$$L = \frac{N_- N_+}{2\pi \sqrt{\sigma_{x-}^2 + \sigma_{x+}^2} \sqrt{\sigma_{y-}^2 + \sigma_{y+}^2}} n_b f_{rev} R \quad (1.4)$$

where "±" denotes positrons and electrons beam respectively, $\sigma_{x,y\pm}$ represents the horizontal and vertical beam size, $N_{+,-}$ is the number of particles in a bunch, n_b the number of bunches, f_{rev} the revolution frequency, and R the geometrical loss factor. In the interaction region, where the beam-beam interaction becomes important, the following formula is commonly used:

$$L = \frac{\gamma_{\pm}}{2e r_e} \left(1 + \frac{\sigma_y^*}{\sigma_x^*}\right) \left(\frac{I_{\pm} \xi_{y\pm}}{\beta_y^*}\right) \left(\frac{R_L}{R_{\xi_{y\pm}}}\right) \quad (1.5)$$

where the starred parameters refer to their value at the Interaction Point (IP). I is the beam current, β_y^* the vertical beta function at the IP, $\xi_{y\pm}$ is the vertical beam parameter which includes the horizontal beta function at the IP, the horizontal emittance, the bunch length and the crossing angle between the beams. R_L and $R_{\xi_{y\pm}}$ are the reduction factors due to geometrical loss such as the hourglass effect and finite crossing.

As already mentioned, SuperKEKB holds the current world record in luminosity (with $\beta_y^* = 1.0$ mm) and in the future the target will be to reach $6 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$ (by the 2030s), by increasing currents beam and reducing their size at the IP, through the decrease of the betatron function down to $\beta_y^* = 0.3$ mm. However, the increase in charge in the bunch, causes a reduction of the Touschek lifetime and the injection system is unable to compensate for the loss [7]. This process makes also the beam-induced background increase significantly, risking deterioration and poor functioning of the detector. It has been estimated that the background should remain acceptable up to a luminosity value equal to $2.8 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$ with $\beta_y^* = 0.6$ mm [8]. Thus, the possibility to achieve higher luminosity is closely related to an upgrade plan of both the detector and the accelerator.

1.2.2 Nano-beam scheme

We have seen that the beta function at the IP (β^*) is a decisive factor to define the luminosity. To be able to ramp the luminosity up, it is necessary to reduce the value of β depending also, but not only, on the variation of the other machine parameters that appear in the equation 1.5.

This new scheme, originally designed by P. Raimondi [4], dictates that the beam bunches have to collide with sufficiently small σ_x^* and at large angle. In case of SuperKEKB the angle is equal to 83 mrad at the IP (larger with respect to the crossing angle used in KEKB) with the beam size of 50 nm in the vertical direction and 10 μm in the horizontal direction (in Figure 1.4 a simplified representation of the differences).

This strategy also helps to reduce the *hourglass effect* [3], which happens when the β^* is comparable or smaller than the bunch length, causing a decrease in luminosity. With a large crossing angle, the overlap length, which is the effective bunch length, is much shorter than the bunch length along the beam axis.

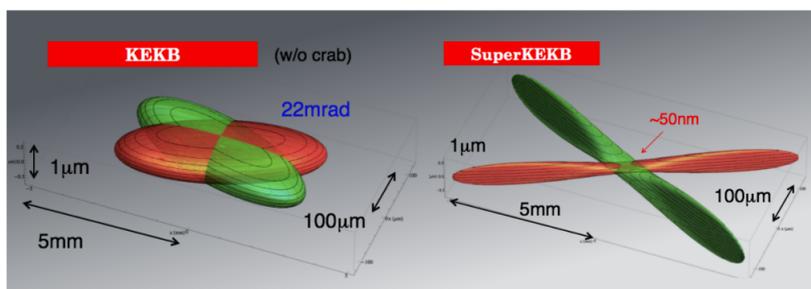


Figure 1.4: Comparison between the beam schemes used in KEKB and SuperKEKB.

Using a crossing angle large enough has other positive implications on the operation of the accelerator and its further improvements, including allowing the placement of a new focusing system at the IP (which may require more space), considering a future redesign of the interaction region.

1.3 Belle II detector

The Belle II detector is a general-purpose spectrometer which consists of a concentric sub-detectors sequence placed around the beryllium beam pipe of 10 mm radius, around the IP of the two beams. Here we will go through a brief description of the sub-detectors (Figure 1.5) going in order from the beam pipe outwards: the Vertex Detector, the Central Drift Chamber, the TOP and the ARICH, the Electromagnetic Calorimeter and the K_L and Muon detector [1, 2, 9].

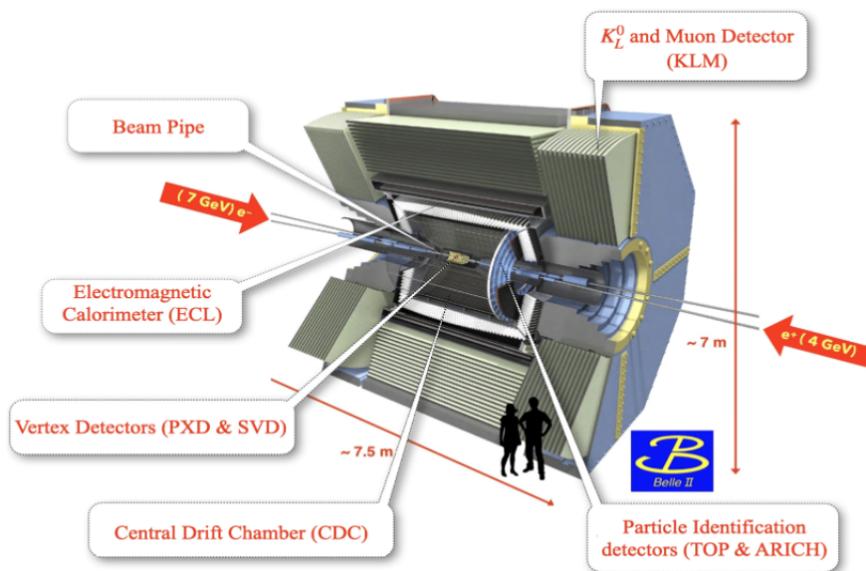


Figure 1.5: Belle II detector.

In Table 1.1 a summary of the main characteristics of all sub-detectors.

1.3.1 Vertex Detector (VXD)

The **VerteX Detector (VXD)** is composed by two devices divided into layers, the silicon Pixel Detector (PXD) and the Silicon Vertex Detector (SVD), for a total of six layers around the beam pipe.

The two inner layers of PXD (L12) consist of pixelated sensors based on the depleted field effect transistor (DEPFET) technology, realised with very thin ($< 100 \mu\text{m}$) sensors which minimise multiple scattering, thus improving the tracking resolution for low-momentum particles. They are at a radius of 14 mm and 22 mm, respectively.

The remaining four layers of SVD (L3456) instead, are equipped with double-sided silicon strip (DSSD) sensors (at 39 mm, 80 mm, 104 mm and 135 mm respectively). Since a lower background rate is expected with respect to PXD, DSSD provides similar performance with a much smaller number of readout channels. These layers are mainly used for tracking/vertexing and also for particle identification (PID), through the measurement of the energy loss (dE/dx).

We can notice in Figure 1.6 that because of the essential asymmetric configuration of the beam energies and the consequent boost of the particles produced in the collisions (section 1.1.2), the structure of the vertex detector is asymmetric along the longitudinal axis.

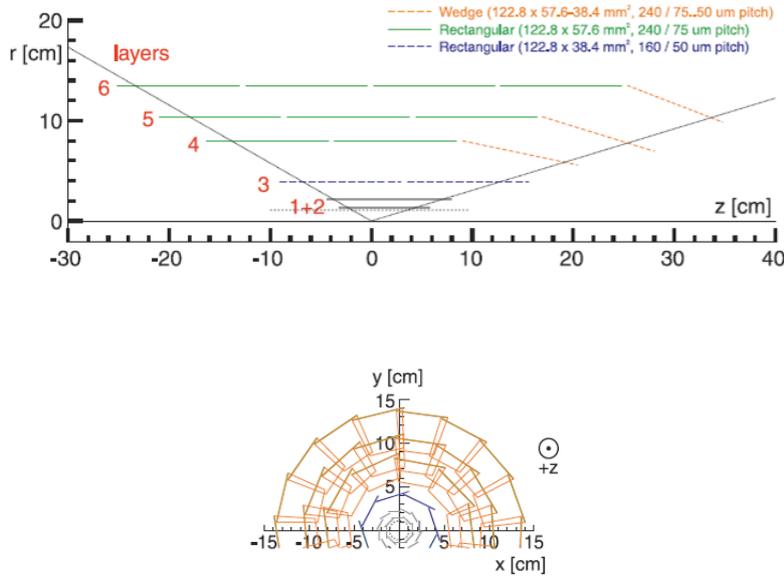


Figure 1.6: A schematic view of the Belle II vertex detector with a Be beam pipe and the six layers of PXD and SVD.

1.3.2 Central Drift Chamber (CDC)

This is the central tracking device, with a large-volume drift chamber and small drift cells. The chamber gas is a $\text{He}-\text{C}_2\text{H}_6$ (50:50) mixture with an average drift velocity of $3.3 \text{ cm } \mu\text{s}^{-1}$ and a

Table 1.1: Summary of the detector components[1].

Purpose	Name	Component	Configuration	Readout channels	θ coverage
Beam pipe	Beryllium		Cylindrical, inner radius 10 mm, 10 μm Au, 0.6 mm Be, 1 mm paraffin, 0.4 mm Be		
Tracking	PXD	Silicon Pixel (DEPFET)	Sensor size: $15 \times (L1 \ 136, L2 \ 170)$ mm ² , Pixel size: $50 \times (L1a \ 50, L1b \ 60, L2a \ 75, L2b \ 85)$ μm^2 ; two layers at radii: 14, 22 mm	10M	[17°;150°]
	SVD	Silicon Strip	Rectangular and trapezoidal, strip pitch: $50(p)/160(n) - 75(p)/240(n)$ μm , with one floating intermediate strip; four layers at radii: 38, 80, 115, 140 mm	245k	[17°;150°]
	CDC	Drift Chamber with He-C ₂ H ₆ gas	14336 wires in 56 layers, inner radius of 160mm outer radius of 1130 mm	14k	[17°;150°]
Particle ID	TOP	RICH with quartz radiator	16 segments in ϕ at $r \sim 120$ cm, 275 cm long, 2cm thick quartz bars with 4×4 channel MCP PMTs	8k	[31°;128°]
	ARICH	RICH with aerogel radiator	2×2 cm thick focusing radiators with different n , HAPD photodetectors	78k	[14°;30°]
Calorimetry	ECL	CsI(Tl)	Barrel: $r = 125 - 162\text{cm}$, end-cap: $z = -102 - +196\text{cm}$	6624 (Barrel), 1152 (FWD), 960 (BWD)	[12.4°;31.4°], [32.2°;128.7°], [130.7°;155.1°]
Muon ID	KLM	barrel:RPCs and scintillator strips	2 layers with scintillator strips and 12 layers with 2 RPCs	θ 16k, ϕ 16k	[40°;129°]
	KLM	end-cap: scintillator strips	12 layers of (7-10) \times 40 mm ² strips	17k	[25°;40°], [129°;155°]

maximum drift time of about 350 ns for a 17 mm cell size.

The CDC contains 14336 wires arranged in 56 layers either in *axial* (aligned with the solenoidal magnetic field) or *stereo* (skewed with respect to the axial wires) orientation (Figure 1.7). Combining the information from both the axial and the stereo layers it is possible to reconstruct full three-dimensional helix charged tracks and measure their momenta. It also provides information for PID by measuring the ionization energy loss, which is particularly useful for low-momentum particles.

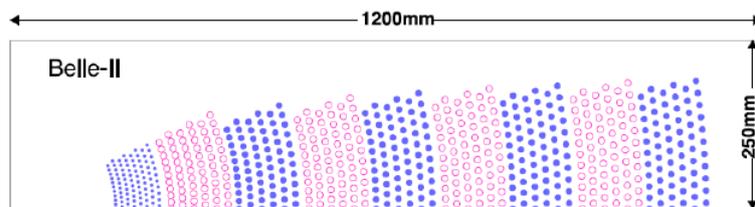


Figure 1.7: Schematic view of the CDC drift cells: blue dots represent the axial wires and the pink empty ones the stereo wires.

1.3.3 Particle identification system (TOP e ARICH)

The **TOP (Time Of Propagation)** is a special kind of Cherenkov detector used for PID in the barrel region. It employs the two-dimensional information of a Cherenkov ring image, obtained from the time of arrival and the impact position of Cherenkov photons at the photodetector at one end of a 2.6 m quartz bar. It is composed by 16 detector modules, each one consisted of a 45×2 cm quartz bar (Cherenkov radiator) with a small expansion volume (about 10 cm long) at the sensor end of the bar (Figure 1.8).

In order to achieve a single-photon time resolution of about 100 ps (required for a good PID), 16-channel microchannel plate photomultiplier tubes (MCP-PMT) are employed, specifically developed for this purpose.

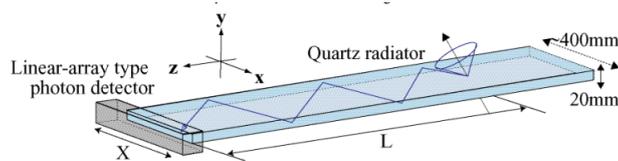
The **ARICH (Aerogel Ring Imaging Cherenkov)** completes the coverage of the PID system in the forward endcap region. It is a proximity focusing Cherenkov ring-imaging detector which adopts aerogel as Cherenkov radiator. In particular this detector employs a novel method to increase the number of detected Cherenkov photons: two 2 cm-thick layers of aerogel with slightly different refractive indices ($n_1 = 1.045$ upstream, $n_2 = 1.055$ downstream) that increase the photon yield without degrading the Cherenkov angle resolution (Figure 1.9).

Hybrid avalanche photon detectors (HAPD) are exploited as single-photon-sensitive high-granularity sensors. Here photo-electrons are accelerated over a potential difference of about 8 kV and are detected in avalanches photodiodes (APD).

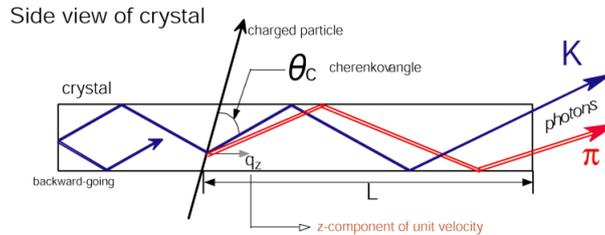
The main task of these detectors is to improve the K/π separation from about ≈ 1 GeV/c until 3.5 and 4 GeV/c of momentum, respectively.

1.3.4 Electromagnetic calorimeter (ECL)

The **ECL** is a highly segmented array of thallium-doped caesium iodide CsI(Tl) crystals assembled in a 3 m long barrel section with a radius of 1.25 m, and two endcap discs located at



(a) A schematic view of the TOP radiator.



(b) A side view of the TOP radiator.

Figure 1.8: TOP detector.

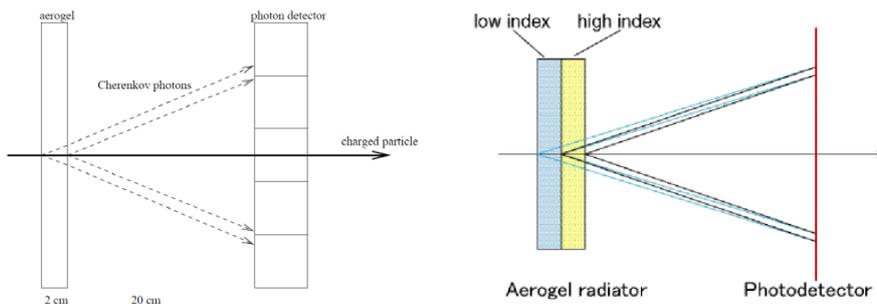


Figure 1.9: ARICH detector.

2 m (forward) and 1 m (backward). All of them are instrumented with a total of 8736 crystals, covering about 90% of the solid angle in the center-of-mass system.

This detector is used to detect gamma rays and to identify electrons separating them from hadrons, especially pions.

1.3.5 K_L Muon detector (KLM)

The K_L and Muon detector (KLM) consists of an alternating sandwich of 4.7 cm-thick iron plates and active detector elements located outside the superconducting solenoid that provides a 1.5 T magnetic field. The iron plates serve as the magnetic flux return joke for the solenoid as well as absorber for hadrons. They provide 3.9 interaction lengths or more of material, beyond the 0.8 interaction lengths of the calorimeter in which K_L^0 mesons can shower hadronically. The active detector elements have been chosen in order to cope with the reduction of the detector efficiency under the SuperKEKB background rates: resistive plate chambers (RPCs) originally installed for the KEKB/Belle experiment for the outermost active layers, while in the two innermost layers of the barrel and endcap regions, scintillator strips with wavelength-shifting fibers are used,

readout by silicon photomultipliers (SiPMs).

1.3.6 Trigger system

The trigger system of Belle II has the role to identify events of interest during data-taking at SuperKEKB, where high background rates are expected. This system is divided into two levels: a hardware-based low-level trigger (L1) and a software-based high-level trigger (HLT), implemented in the data acquisition (DAQ) system.

- **L1**: based mainly on fast track reconstruction in the CDC and on ECL energy, has a latency of $5\ \mu\text{s}$ and a maximum trigger output rate of $30\ \text{kHz}$, limited by the read-in rate of the DAQ.
- **HLT**: is a key component of the DAQ, used to fully reconstruct events that pass the L1 trigger selection. It has to reduce online event rates to $10\ \text{kHz}$ for offline storage and it must identify track regions of interest for PXD readout in order to reduce data flux. It fully reconstructs events with offline reconstruction algorithms, using all detectors information except for the PXD.

1.4 Current state of data taking

SuperKEKB accelerator reached a peak luminosity of $4.7 \times 10^{34}\ \text{cm}^{-2}\text{s}^{-1}$ and Belle II accumulated almost $428\ \text{fb}^{-1}$ before the beginning of Long Shutdown 1 (LS1) in July 2022. (Figure 1.10).

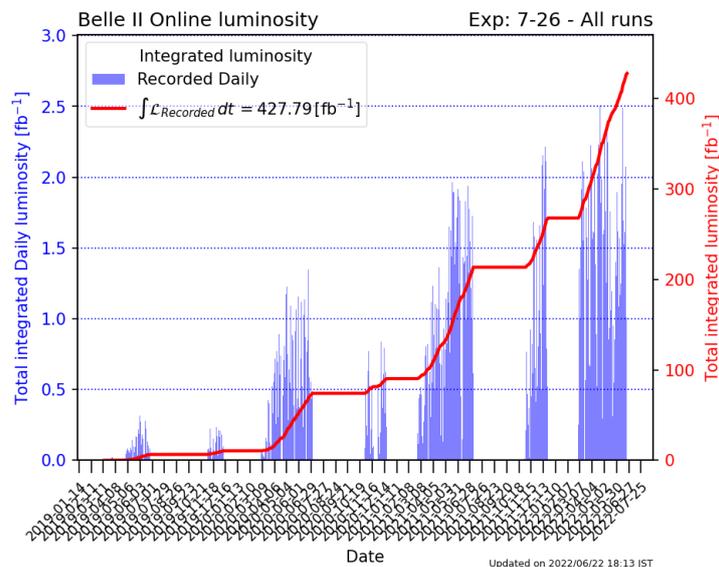


Figure 1.10: Total recorded integrated luminosity before Long Shutdown 1.

The target of SuperKEKB is to achieve a $L_{inst} = 6 \times 10^{35}\ \text{cm}^{-2}\text{s}^{-1}$ and to increase the integrated luminosity from $428\ \text{fb}^{-1}$ (current value, starting in 2019) to $50\ \text{ab}^{-1}$ (as shown in the

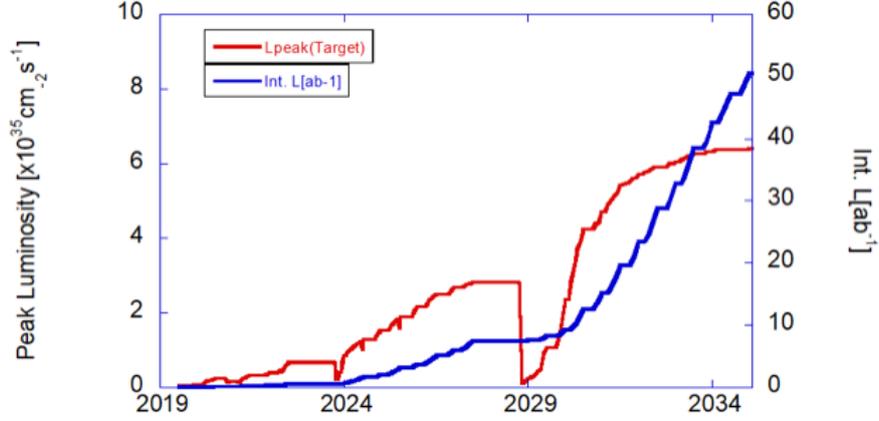


Figure 1.11: Luminosity projection plot (plan for the coming years).

projection plot in Figure 1.11).

A three-phase upgrade program has been planned [10]:

- **short term:** year 2023. Long Shutdown 1 (LS1) started in July 2022, and will be concluded at the end of 2023. The main purpose of the shutdown was the installation of a complete PXD, since in the 2019 installation only two ladders were included in layer 2. In addition, significant maintenance and improvement work has been carried out both for Belle II and SuperKEKB.
- **medium term:** approximately year 2028-29. Long Shutdown 2 (LS2) will probably be needed for the upgrade of the Interaction Region (IR) to reach a new luminosity target $L_{peak} = 6 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$. Several open questions and difficulties have triggered many studies and discussions about a possible redesign of the machine lattice during this phase. In particular it would be necessary to deal with the limitation of the optics of the machine, concerning the further increasing of the luminosity and accordingly of the backgrounds rates. A new Vertex Detector might be also required, to accommodate the new IR design, and other sub-detectors upgrades are possible.
- **long term:** years > 2032 . Studies have started to explore upgrades beyond the currently planned program, such as beam polarization and ultra-high luminosity and so possibly L_{peak} in excess of $1 \times 10^{36} \text{ cm}^{-2}\text{s}^{-1}$. While the beam polarization has a concrete proposal [11, 12], for ultra-high luminosity studies have just started.

At time of writing LS1 and the installation of the complete pixel detector (PXD) has just been completed. The restart of data taking is planned at the beginning of 2024.

2. Belle II Upgrade

This second chapter wants to address the motivation for the upgrade of Belle II. We will give an overview of the primary background sources in the experiment to understand how to mitigate them in order to achieve a better performance of the whole detector, when ramping up the luminosity. We will also introduce some of the proposals made for the improvement of the vertex detector.

2.1 Purposes of the upgrade

The physics program of the experiment requires SuperKEKB to work under challenging conditions and machine backgrounds play an important role in the safety, efficiency, and performance of data taking. Current studies [8] predict that the accelerator may reach higher luminosity targets with the existing accelerator complex, but to achieve the established final value of $6 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$, an enhancement of the interaction region is being considered.

Several improvements and modifications have been planned during two long shutdowns in order to reach the integrated luminosity target of 50 ab^{-1} : LS1 in 2022 to install a complete VXD, and LS2 in 2028 or later for the upgrade of the interaction region and the accelerator components. LS2 offers the opportunity for a significant detector upgrade, improving robustness against backgrounds and physics performance.

Belle II detector is expected to operate efficiently under the levels of background extrapolated to luminosity target, but safety margins are not large. Moreover, in case of a redesign of the interaction region large uncertainties in the background extrapolations are unavoidable.

Therefore the global upgrade program [10] is justified by many considerations, among them:

- improve detector's resistance to higher levels of background;
- make subdetectors more robust against radiation damage;
- push forward safety margins for running at higher luminosity;
- develop the technology to cope with different future paths;
- improve overall physics performance.

2.2 Background sources and limitations in Belle II

SuperKEKB is already the world's highest-luminosity collider and it aims to reach a new peak in the near future and to increase the collected statistics, to become more sensitive to rare processes

and precise measurements of Belle II physics program. But to be able to do this without losing the good functionality of the entire detector, it is necessary to understand how to reduce the beam background where possible and how to cope with the consequent challenges.

Several simulations and measurements of beam background [13] are being done in order to guess possible future machine scenarios, under new luminosity conditions. This is necessary to study the vulnerability of the subdetectors (and more generally of the machine) and thus, to design the countermeasures to adopt against the deterioration of performance and materials.

2.2.1 Main background sources

In the following some of the primary *single-beam* and *luminosity-dependent* background sources are discussed [1, 8].

Touschek effect : It is an intra-bunches scattering process, where the Coulomb scattering of two particles in the same beam bunch causes a variation of their energies, increasing the value of one of them and lowering that of the other from the nominal value. This interaction among the bunch particles is the first beam background source at SuperKEKB.

Beam-gas scattering : this represents the collision of beam particles with residual gas molecules in the beam pipe. It is the second beam background source and it can occur via two processes: *Coulomb interaction*, which changes the direction of the beam particles and *bremstrahlung scattering*, which decreases their energy.

In these two processes, the scattered particles fall out the stable orbit and hit the beam pipe while they move around the ring. This mechanism causes electromagnetic showers that can reach the detector if their origin (loss position) is near the interaction region. Collimators are in place to reduce these backgrounds by absorbing the off-orbit particles. These backgrounds are present even if the beams are not colliding and are therefore called *single-beam* backgrounds.

Radiative Bhabha scattering : It is the electron-positron scattering $e^+e^- \rightarrow e^+e^-\gamma$, an undesirable collision process at the IP which has very high cross section but only little interest for the physics studied in the experiment. The emitted photon can interact with the magnets near the detector producing a large amount of neutrons via the photo-nuclear resonance mechanism (such neutrons are the main background source for the outermost Belle II detector).

Two-photon process : it consists in the process $e^+e^- \rightarrow e^+e^-e^+e^-$. The produced electron-positron pairs can spiral around the solenoid field lines and leave multiple hits in the inner layers of the detector.

These backgrounds are called *Luminosity backgrounds* because their strength is proportional to the luminosity. They can be reduced with a careful design of the interaction region including neutron shielding.

Synchrotron Radiation (SR) : X-rays emitted from the beam when electrons and positrons pass through the strong magnetic field near the IP. The HER beam is the main source of this type of background, because SR power is proportional to the square of beam energy and magnetic field. SR can potentially damage the inner layers of the vertex detector due to an higher radiation dose in that region. Absorbing structures have been implemented inside the beam pipe to protect the detector from SR.

The rate of these processes is relatively high for small beams and high currents, leading to a beam lifetime in the order of only few minutes. This requires a continuous injection mechanism, in which particles are added to individual bunches to keep the charge constant. While this is essential for high luminosity operation, it is also the source of an additional background, since the perturbed bunches are noisy for a number of turns. The injection background is particularly hard to model and estimate.

2.2.2 Current background status and future predictions

Several monitoring devices are located all along the accelerator to measure radiation doses on the detector and delicate regions of the ring, in order to intervene as soon as possible in case too high levels are reached.

Dedicated beam background studies have been performed and for example, Figure 2.1 shows the rates measured in June 2021 [13] with a betatron function at the IP of $\beta_y^* = 1$ mm and a collision luminosity of $L = 2.6 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Current background rates in Belle II are acceptable and in most cases, well below the limits listed in Table 2.1 [8].

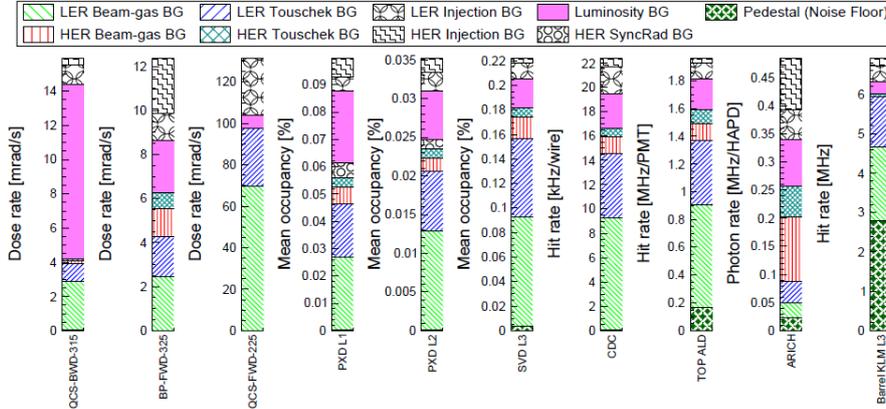


Figure 2.1: Measured Belle II background in June 2021. Each column shows different background sources for Belle II sub-detectors, for superconducting quadrupole magnet (backward and forward) and for the beam pipe.

Event though the current level is of no concern in terms of occupancy for the innermost layers of the vertex detector, a larger amount of localized SR, for example, could cause inhomogeneities in PXD modules, which would be very difficult to compensate by adjusting the operation voltages of the affected ones.

Until now it can be said that SuperKEKB and Belle II are operating stably. Beam-induced background rates are well below the limits of the detector and do not prevent from increasing further the current and hence the luminosity, as demonstrated by the predictions for the background rates *before LS2* with a known machine configuration. For what concern the predictions at $L_{inst} = 3 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$ instead, called *after LS2* operation, there are several uncertainties tied to the machine configuration. In fact at the moment, the working machine lattice to reach the target luminosity is not known and in addition, the final design of the IR and beam pipes is not concluded yet.

Detector	BG rate limit		Measured BG
Diamonds	1 rad/s to 2 rad/s		< 132 mrad/s
PXD	3 %		0.1 %
SVD L3, L4, L5, L6	4.7 %, 2.4 %, 1.8 %, 1.2 %		< 0.22 %
CDC	200 kHz/wire		22.3 kHz/wire
ARICH	10 MHz/HAPD		0.5 MHz/HAPD
Barrel KLM L3	50 MHz		4 MHz
	non-luminosity BG		
	Before LS1	After LS1	
TOP ALD	3 MHz/PMT	5 MHz/PMT	1.8 MHz/PMT
	+ luminosity BG		

Table 2.1: Background rate limits for Belle II detector sub-systems. The third column shows the total measured background rate in June 2021

Therefore an alternative solution is employed to roughly estimate the background rates. The background predicted Before LS2 phase is considered as a starting point and then different scaling factors are applied for single-beam background component, considering three different possible scenarios:

- **x2** - optimistic scenario-1 (**v1**)
- **x5** - nominal scenario-2 (**v2**)
- **x10** - conservative scenario-3 (**v3**), an arbitrary factor assuming that all single-beam backgrounds will be increased by a order of magnitude After LS2.

The luminosity background sources are instead scaled with luminosity ratio. These are then used to simulate the behaviour of the whole detector in future perspectives, as we will see in the following.

2.3 Summary of possible VXD upgrade

The Vertex Detector is particularly sensitive to machine background because it is the closest to the beam pipe and therefore subject to high doses of radiation. As we have already seen, current studies are trying to extrapolate how it could be affected by reaching the future luminosity target, but there are a lot of uncertainties due to models and still not well defined design of the interaction region. Moreover, a completely new detector might be required in the event of a considerable redesign of the IR, and the physics performance could be also improved, taking advantage of the more recent technology developments.

In particular, all different upgrade ideas of the entire Belle II detector intend to ensure its proper functioning at the higher levels of luminosity, considering also further improvements of the lattice machine and the colliding beams. The current detector configuration is not expected to maintain its performance when facing higher beam background levels or higher rates.

Concerning the Vertex Detector, all proposed improvements aim to:

- reduce occupancy level by employing fully pixelated and fast detector (CMOS technology has been chosen);

- increase robustness against tracking efficiency and resolution losses from beam background;
- improve radiation hardness to reduce detector ageing effects and performance degradation;
- reduce the inserted material budget between subdetectors in order to achieve good resolution by lessening the multiple scattering, especially important at lower momenta.

In the following we will present in a few words the four main proposals for future upgrade: Depleted Field Effect Transistor (DEPFET) pixel, thin strip sensor, CMOS Monolithic Active Pixel Sensor and SOI technology. The first two are more conservative and try to exploit as much as possible of the existing detector, making some appropriate adjustments to the sensor type, readout or mechanical structure. The last ones instead, plan to build an entirely new detector.

The reference background levels determining radiation robustness requirements of the innermost layers are the following:

- Hit rate capability: 120 MHz/cm²;
- Total Ionizing Dose: 10 Mrad;
- NIEL fluence: 5×10^{13} n_{eq}/cm².

At this time, the main effort is focused in the development of the CMOS MAPS system, with the SOI as a possible backup option, although still requiring significant R&D. The DEPFET and thin strip sensors are substantially abandoned, but we present them here to provide the full picture.

2.3.1 Depleted Field Effect Transistor (DEPFET)

This first proposal intend to minimize risks and costs of the project, preserving the general layout of the PXD system. The upgrade consists to improve the sensor to provide higher safety factor for the allowed occupancy and to prevent some issues that currently weaken the good functionality of the detector.

Some of the main improvements are listed below:

- improve signal transmission on the pixel matrix and the signal processing in the readout, in order to reduce the readout time per row from the current 100 ns to 50 ns. In this way the frame time and the background occupancy might be reduce by a factor 2, while leaving unchanged the optimized size and the number of PXD pixels;
- increase the robustness against beam losses which could make the gate lines inefficient or even inoperative, on almost all PXD modules. This reaction seems to be due to a high photocurrent on the chip because of the high instantaneous dose. It could be mitigated by adding protection circuits on-chip;
- Total Ionizing Dose (TID) effect on the chip provokes an unexpected avalanche current that does not compromise the sensor performance, but requires more power supply to provide enough current. This issue might be solved by bringing some changes in the DEPFET pixel layout.

The DEPFET improvement R&D is currently inactive because of lack of person power and dedicated funding.

2.3.2 Thin and Fine-Pitch SVD

The Thin and Fine-Pitch SVD (**TFP-SVD**) is a new detector concept that aims to improve not only SVD, but potentially also the inner part of the CDC, whose functionality could be threatened by future beam background conditions. This proposal uses Double-sided Silicon Strip Detectors (DSSDs) for the inner and middle detector volume, since a single sensor can cover large dimensions. In the current detector the DSSD technology is already used in the SVD.

The major improvement would be the reduction of the material budget. Currently SVD has about 0.7% X_0 material budget per layer. TFP-SVD instead, decreasing the sensor thickness to 140 μm , intends to reduce it to 0.41% X_0 . Moreover, a smaller sensor thickness is expected to reduce the voltage needed to reach the full depletion, even after radiation damage. The front-end developed for TFP-SVD, the SNAP128 chip, offers also a reduction of the amount of cables.

Some concerns about TFP-SVD are the feasibility and efficiency of the final sensor production and the small signal charge due to the short path length of the particles through the sensor.

A first prototype has been produced by Micron-Semiconductor Ltd (UK), with a size of $52.6 \times 59.0 \text{ mm}^2$. The characterization studies are in agreement with the expectation and also a lower full depletion voltage is confirmed. The development of the TFP-SVD option has been abandoned because of the small signal to noise ratio (10 or less) and difficulties in the chip design optimization.

2.3.3 CMOS Monolithic Active Pixels Sensor

The VTX proposal, that we will analyze in more details in the next chapter, aims to replace the entire current VXD detector using Monolithic Active Pixel Sensors (MAPS) based on CMOS (Complementary Metal-Oxide Semiconductor) technology.

In CMOS MAPS the basic commercial CMOS technology is minimally modified to allow the collection of the charge released by a charged particle traversing the epitaxial layer where the readout electronics is fabricated. There are many technological variant sharing this same concept.

The Belle II chip is called OBELIX (Optimized BELle II pIXel sensor), based on the pixel matrix of the TJ-Monopix2 chip, whose characterization is the main topic of this work (Chapter 5) and which is developed with the technology of the TowerJazz company with a minimum feature size of 180 nm. The size of the sensor is expected to be $3 \times 1.9 \text{ cm}^2$ with a pixel pitch between 30 μm and 40 μm , achieving a spatial resolution below 15 μm , required by the VTX upgrade program. The timestamp clock signal have to reach down to 25 ns, in order to deal with the target hit rate of 120 MHz/cm². All these characteristics allow to obtain a sensor with high granularity in time and space.

The VTX detector consists in 5 concentric layers of monolithic sensors with a barrel geometry. The global target thickness for the air-cooled two inner layers and for the water-cooled three outer layers, is expected to be of about 2% X_0 .

The VTX solution intends to reduce the current PXD integration time by at least two orders of magnitude introducing also other improvements:

- lower detector occupancy which allows to cope with higher background and to mitigate data-transmission bottlenecks.
- Better tracking efficiency and improved momentum resolution and impact parameter resolution at low transverse momentum.

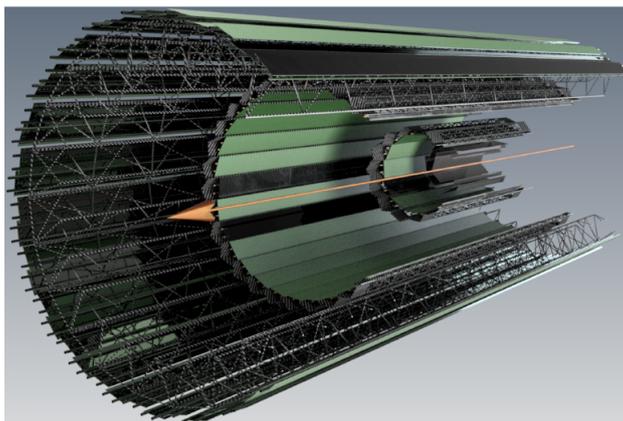


Figure 2.2: Overall VTX layout.

- Smaller data cables cross sections and less complex cooling system which might lighten the needed services.
- Simplified control and power systems due to the use of a single sensor chip for all layers.

At the current state of art intense R&D is being carried out, taking advantage from the experience of other experiments like ALICE, with the same type of sensor [14].

2.3.4 Silicon On Insulator (SOI)

An alternative proposal employs a new pixel design, called Dual Timer Pixel (DuTiP) [15], based on SOI technology (see section 3.3.3). In Silicon-On-Insulator (SOI) technology, electronics is fabricated in a silicon layer separated by a SiO_2 isolation layer from the main detector substrate. This new sensor concept has been developed to fulfill the requirements of a new vertex detector with faster readout, lower occupancy, smaller data size and smaller data transfer. In particular, it aims to store at least two hits during the Belle II trigger latency ($5 \mu\text{s}$), to avoid loss information in higher background environment.

The size of the new designed pixel is $45 \mu\text{m}$ and the sensor layer thickness is $50 \mu\text{m}$, which gives an intrinsic resolution better than $15 \mu\text{m}$. The analog part is quite standard for a binary detector and consists of a sequence of preamplifier, shaper and comparator. ALPIDE [16] was chosen as analog circuit with some modifications to adapt it to SOI technology. The quite complex digital circuit has to be assembled on each pixel, and Lapis semiconductor $2.0 \mu\text{m}$ FD-SOI CMOS technology has been chosen, based on the experience gained in the successful development of other detectors like the pixel detector for the future ILC (SOFIST) [17].

DuTiP pixel detector is designed to cover the current VXD acceptance with 7 layers. For the inner layer of the detector might be possible the cooling with airflow at room temperature; for the outer layers instead, a combination of air and water flows. The DuTiP R&D are ongoing and the chip prototypes seems to work fine, in agreement with the expectations, although there are concerns about the radiation hardness of the technology.

After this brief review of the main upgrade proposals, we can then elaborate on the VTX program in Chapter 4.

3. CMOS MAPS sensors

This chapter introduces the essential features of the semiconductor detector technology, going through the history of its advancements, which have led to the currently most promising sensors based on CMOS logic structure, the Monolithic Active Pixel Sensors (MAPS). The VTX project wants to make the most of technologies that have already proven reliable in precision measurements, obtaining fast readout and high radiation tolerance, like the TJ-Monopix development line. We will briefly present it, mentioning the peculiarities of its prototypes, to understand how they could fulfill the Belle II requirements.

3.1 Semiconductor detectors

Charged particles can be detected through the measurement of the ionization charge released during their passage through solid, liquid, or gaseous detector. Semiconductors are extensively used as ionization detectors since they are easily available and the readout can be finely segmented leading to excellent spacial resolution. This section provides the basic description of the working principle of semiconductor detectors.

Solids can be divided into three categories based on their electrical conductivity: conductors, semiconductors and insulators. In a solid state lattice, the constituent atoms have a dense periodic arrangement and the energy levels of some level groups lie energetically so dense (order of meV) that one speaks of *energy bands*, separated from each other by a *band gap*, which represents the distance in energy between them (E_G , energy gap). The electrical conduction properties of materials are determined by the two highest energy bands, which are the **valence band (VB)** and the **conduction band (CB)**; in particular, they depend on the band gap between the two levels and on the band occupation, as we can see in Figure 3.1.

Insulators are characterized by a large energy gap between the VB and CB, of the order of several eV, with the number of valence electrons matching the number of states in the VB. Since the band gap is much larger than the thermal excitation energy ($KT \approx 25$ meV at room temperature) the CB is practically empty and the resistivity becomes very large.

Semiconductors have a smaller energy gap with respect to the insulator (for example 1.12 eV in silicon). In this way electrons from VB can easily overcome the gap, moving to the CB by thermal excitation or by external electric fields. When an electron makes this transition, it leaves a hole in the VB, which could be filled in turn, by another electrons of the VB. Applying an external electric field, the free electrons in the CB and the holes in the VB start to move producing two different current flows, one with negative carriers and the other with positive carriers.

In conductors, either the conduction and valence bands overlap or the conduction band is partially filled, so transitions within the same band and between the two different bands are easy

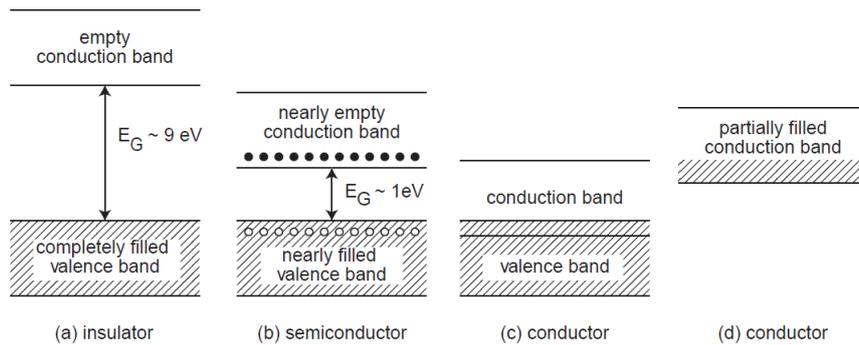


Figure 3.1: Schematic structure of the energy bands in insulators (a), semiconductors (b) and conductors (c,d).

and current conduction requires minimal energy.

3.1.1 Transport of charge carriers and signal formation in semiconductors

When a charged particle goes through the medium, it releases a certain amount of energy mainly by ionization. Most of this energy loss causes the formation of positive and negative charges. In semiconductors these carriers are the hole-electron pairs created by ionization, which start to move in opposite direction if an external electric field is applied: the holes (positive) migrate towards the *cathode* and the electrons (negative) towards the *anode*, which sense the signal induced by this movement. Their drift induces an accumulation of charges on the electrode surfaces, and it is possible to record this charge induction as a charge, current or voltage signal.

The charge carrier density of semiconductors can be modified by doping the material with specific chemical elements, causing also a modification of their conduction properties. Undoped semiconductors are called *intrinsic semiconductors*. *Extrinsic semiconductors* instead, are artificially doped with external impurities like:

- Pentavalent elements (P, As, Sb), called *donors*, added in a tetravalent material (Si, Ge) produce an excess of conduction electrons with respect to the holes (n doping, Figure 3.2 (a)).
- Trivalent atoms (B, Al, Ga), called *acceptors*, create an excess of holes (p doping, Figure 3.2 (b)).

3.1.2 The pn junctions as detector

The base material employed in semiconductor detectors is silicon, because it is stable, abundant and it has a low band gap which allows an adequate amount of charge to be produced.

When a p-doped semiconductor gets in contact with a n-doped material, a *pn junction* is formed. In the p-doped part the holes are the dominant charge carriers, called *majority carriers*, while in the n-doped part, the electrons are dominant. The presence of these excesses of opposite charge in the two parts of the junction, generates a density difference across the junction, which

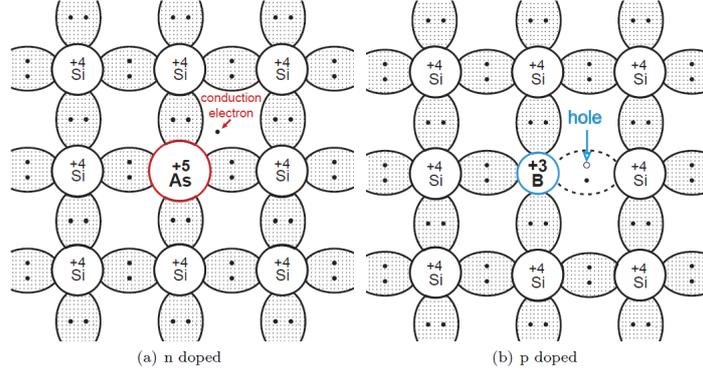


Figure 3.2: Schematic representation of atom bonding structure in n-doped and p-doped semiconductors.

causes the diffusion of the majority carriers from each part to the opposite one. At the boundary the charges recombine (that is when a conduction band electron occupies a valence band hole, losing energy), creating a zone which is free of charge carriers, called **depletion zone**. After the recombination, the atoms of this depletion region are ionized, and so it is no longer neutral, but features a *space-charge* (Figure 3.3): a positive one in the n-layer, and negative in the p-layer. Due to their opposite sign, they generate an intrinsic electric field that stops the original diffusion current.

The application of an external voltage V_{ext} between the two sections of the junction, causes a variation of the depletion region widths, depending on the size and polarity of the applied voltage. It is possible to distinguish:

- *forward bias*, $V_{ext} > 0$: a positive external voltage applied to the p side with respect to the n side, causes a reduction of the depletion region;
- *reverse bias*, $V_{ext} < 0$: if a negative voltage at the p side or positive at the n side relative to the respectively opposite side is applied, the depletion region gets wider.

The depleted region is a zone without free charge carriers, and with the presence of an external electric field (reverse bias). It is the region where the created charge can be collected and produce a measurable signal.

The basic structure of semiconductor detectors is a reverse biased p-n diode where the depleted region represents the active volume and the diode terminals are the charge collection electrodes.

3.2 Hybrid and monolithic pixel sensors

The collection electrode in semiconductor detectors can be segmented with various patterns by using the features of modern electronics processing technology. Typical detectors have 200-300 μm thickness with charge collection time of the order of 5-10 ns, and segmentation of the order of few hundred micron, organized in strips for 1D measurements or pixels for 2D measurements.

We discuss in more details the *hybrid* and *monolithic* pixel sensors, where the sensor and the readout could be separate entities or integrated on the same silicon substrate [18].

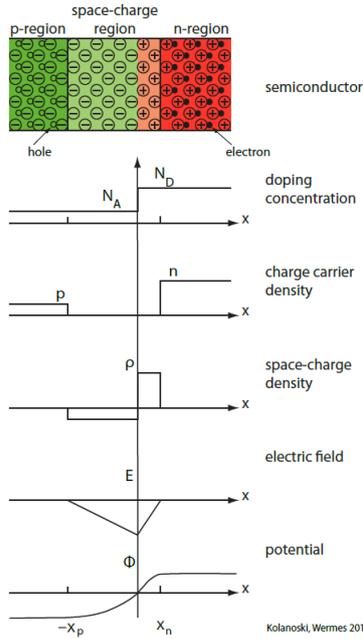


Figure 3.3: Doping concentration, charge carrier and space charge densities, electric field strength and electric potential in a pn junction.

3.2.1 Hybrid pixel detectors

Hybrid pixel detectors are composed by two different parts: a silicon layer structured in pixel cells, and the readout chips with the same cell pattern, that amplify, digitize, store and transmit the hit data. The sensor and the readout are connected at each pixel by a conducting micro-connection (called *bump bond*). In Figure 3.4 an example of a detector module is shown.

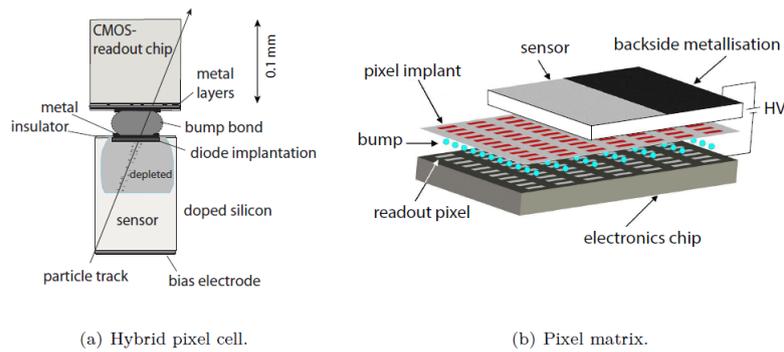


Figure 3.4: On the left (a), the structure of a single pixel cell, made by the sensor and the electronic readout. On the right (b), an exploiting representation of the entire hybrid pixel matrix arrangement.

When a particle crosses the sensor, a signal is generated on the electrodes due to the drift of

the charge carriers in the depleted region. This signal passes through the conductive bump in the readout chip, where it is amplified and processed.

The main advantage of the pixel detectors with respect to the strip sensors, is the higher granularity that translates in a lower occupancy, a better signal to noise ratio, and also a better tolerance to radiation damage. The *leakage current* increases as the radiation dose becomes higher, it depends on the volume of the sensors and it is distributed over all the electrode. In pixel detectors this current is shared among more electrodes with respect to the strip detectors, so the leakage current per electrode is smaller. Moreover, as separate entities, sensor and readout could be optimized independently.

Among the disadvantages instead, there are the cost and complexity of the implementation processes, but also larger material thickness, which worsen the track reconstruction and momentum resolution because of the multiple scattering.

3.2.2 Monolithic pixel detectors

We have seen that hybrid detectors are made by the active sensor and the passive readout chip in two separate structures, connected by micro-connections. Since both of them are made of silicon, in principle, it could be possible to build them in a monolithic unit. In this way the amount of material decreases, enhancing the tracking performance and momentum resolution.

This kind of sensors have been developed either by using dedicated technologies, or by exploiting existing technological processes with minimal modifications. An example of first type is the DEPFET pixel detector, which employs a specifically optimized fabrication process integrating one transistor in each pixel cell.

Developments of the second type have tried to exploit industrial technologies already available, like the CMOS technologies. They have to reach a large depletion zone to improve the signal-to-noise ratio, but also design electrode with small capacitance, to reduce the power consumption and the noise. High radiation tolerance is also required.

Monolithic Active Pixel Sensors (MAPS) employ CMOS technologies to include a readout circuit in their structure and we will see them in more details in section 3.3

Depleted p-channel field effect transistor (DEPFET)

In a DEPFET pixel detector, each pixel implements only one transistor. In Figure 3.5 an example of a DEPFET pixel is shown.

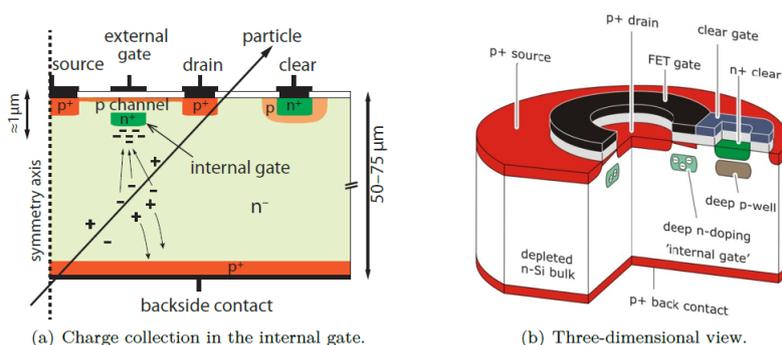


Figure 3.5: On the left (a) a cross section of a circular DEPFET pixel cell, where the charge collection is also sketched. On the right (b), a three-dimensional view of the same pixel cell.

The depletion region extends between the backside p^+ contact and the several p^+ regions near the transistor element (drain, source and a n^+ clear contact installed in a p region) and the n^- substrate. When a traversing particle releases energy, ionizing the medium, electrons drift towards the top surface, and the holes towards the backside due to the external potential. The transistor is a p-channel MOSFET that produces a hole current from source to drain, controlled by the potential on the external gate. In addition, there is a deep n^+ implant placed a few micrometers under the transistor. It is the most positive point in the pixel structure and so it is a local minimum for the electrons. This implant features an accumulation of electrons, which changes the potential, making it an *internal gate*. Electrons collected on this electrode, and the external gate influence the current flow in the DEPFET transistor channel allowing the detection of the released charge. After the measurement, they are removed applying a positive voltage on the *clear contact*. This element is embedded in a p region (*deep p-well*), not to compete as a collection electrode for the electrons. With some variations in the pixel structure, this is the type of sensors that the Belle II PXD detector uses [19].

The readout of the DEPFET matrix is sequential making it a relatively slow device. In Belle II, the PXD has an integration time of 20 μs , which requires the fast SVD (25 ns shaping time) to obtain good tracking efficiency. On the positive side, the DEPFET can be made extremely thin ($\approx 50 \mu\text{m}$).

3.3 CMOS Monolithic Active Pixel Sensors technology

First prototypes of pixel detector, which employed the CMOS technology to fabricate both the sensor and the readout circuit in the same silicon die, date back to the early 1990s. Further developments have led to the *monolithic active pixel sensors* (MAPS) with an epitaxial silicon layer for charge collection and then to the Depleted MAPS (DMAPS), where the depleted region is extended throughout the volume.

3.3.1 MAPS pixel detectors

This type of sensors have used the CMOS technologies for optical applications, consisting of an epitaxial layer with high resistivity to detect the light, and the electronic circuitry assembled on top of it. In Figure 3.6 a schematic of their structure is shown.

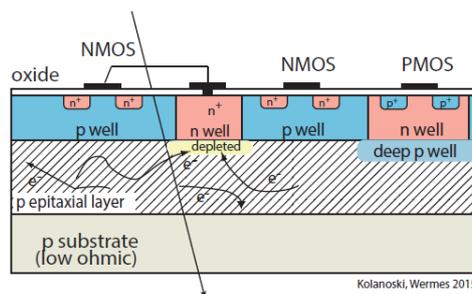


Figure 3.6: Schematic of a monolithic pixel detector (MAPS).

In this detectors, not the entire area is sensitive to the charge production. Without technology modifications the sensitive layer cannot be depleted and charge collection is mostly achieved by diffusion, leading to a suboptimal charge collection efficiency. The collection electrode is a n^+

contact in a n-well, embedded in a p-epitaxial layer placed on top of a p substrate. Other n-wells can be necessary, for example, to host PMOS transistors, and for this reason they have to be shielded by deep p-well, to prevent them from becoming competitive for charge collection. These highly doped deep layers assume a negative potential with respect to the collection electrode and hence a repulsive effect. Due to the absence of a drift field in the epi-layer, with the exception of the region immediately below the collection electrode, collection charge occurs mainly by diffusion and thus it is slow and incomplete. Typical values of the epi-layer thickness are within the range of 1 - 20 μm [18].

These MAPS detectors cannot be used in high rates experiments because they have charge collection time on the order of 100 ns, thus too slow. The signal is very small (typically $\geq 1500 e^-$) in this standard process, due to the small thickness of the epitaxial layer, not even fully depleted, and they also have limited radiation tolerance. The very small electrode capacitance allows the production of a sizeable signal even with a very small collected charge. In addition, due to Non-Ionising Energy Loss (NIEL) effects [20], radiation produces displacement damage in the sensor bulk also creating energy levels in the band gap that act as trapping centers. Electrons and holes can be trapped and then be released again after some time, producing a decrease of the signal amplitudes if the de-trapping time constant is longer than the time of signal formation.

Although their limited radiation tolerance MAPS sensors have been successfully employed for heavy ions collisions, like STAR experiment at RHIC (Relativistic Heavy Ion Collider) [21] and the ALICE upgrade at the LHC [22], which are characterized by a relatively low event rate.

3.3.2 Depleted MAPS pixel detectors

In order to obtain a fast and complete collection of the released charge, it is necessary to drift the carriers toward the electrodes by an external electric field. A full depletion of the sensitive region enhances the charge collection, producing larger signals and a better signal-to-noise ratio. The depletion depth depends on the substrate resistivity and the bias voltage, according to the relation:

$$d \propto \sqrt{\rho V} \quad (3.1)$$

For this reason, new processing techniques have been employed, allowing the application of higher voltage, and/or high-resistivity substrate wafers have been used. Both techniques are used to increase the depletion region underneath the collection electrode (typically to 25– 150 μm) thus providing a sufficiently large and fast signal.

We can distinguish two main variants of DMAPS pixel detectors [18]: with a *large* or *small* collection electrode, shown in Figure 3.7.

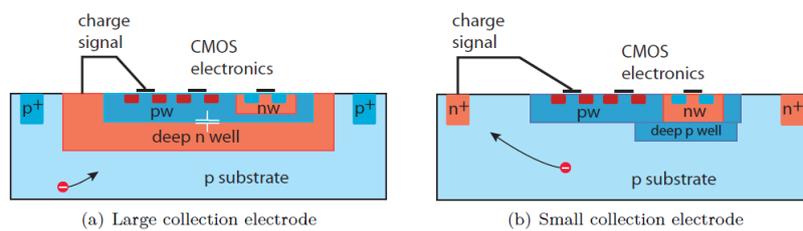


Figure 3.7: On the left (a) a schematic of the large electrode design. On the right (b) the small electrode design.

The *large collection electrode* DMAPS features a deep n-well acting as an electrode but also

as a shield for the entire CMOS readout electronics, embedded within it. With a large electrode charge is drifted vertically towards the electrode from the region underneath. This architecture improves the radiation tolerance because the reduced average drift distance of the charge carriers decreases the probability of trapping. At the same time though, the large size of the electrode implicates higher values of capacitance (several hundred fF) which increases the noise and worsens the timing performance.

The *small collection electrode* variant has a small n-well collection node, distanced from the CMOS circuitry which is embedded in p-well and deep p-well layers needed for shielding. In this design, low capacitance of about 5-20 fF can be obtained, improving noise and timing performance. Radiation tolerance instead, is more difficult to reach due to larger average distance travelled by the carriers. In addition, obtaining complete charge collection efficiency is more difficult. Smaller pixel dimensions are preferred for small electrode (to reduce the path), and therefore higher power density is accepted in exchange of increased robustness against radiation.

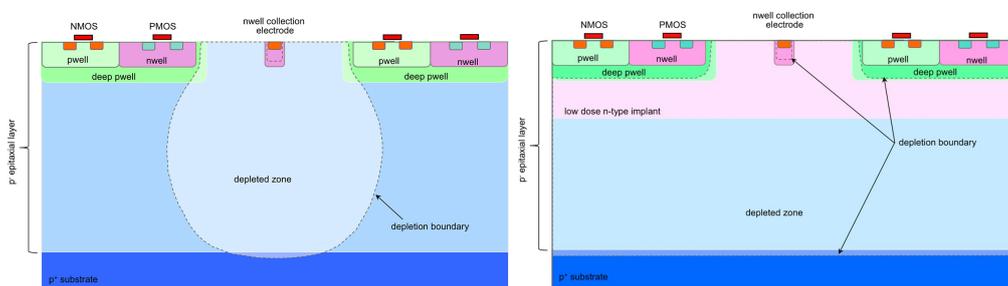


Figure 3.8: Small electrode design adopted in the ALPIDE sensor with the standard process (left) and the process modification with the addition of a low dose n-doped layer used to implement a planar junction and deplete the epitaxial layer over the full pixel area.

In the process adopted for the ALPIDE sensor with a small electrode [16], it is difficult to deplete the epitaxial layer over its full width as it is shown in Figure 3.8 left side. In order to achieve the full depletion of the sensitive layer, combined with a low capacitance collection electrode, a planar junction separated from the small collection electrode has been implemented. A low dose deep n-type implant has been used to realize a planar junction in the epitaxial layer within the pixel matrix below the wells containing circuitry, as shown in Figure 3.8 right side. The epi-layer is thus depleted through two pn junctions: the deep p-wells to low dose n-implant junction, and the n-implant to p-epitaxial layer [23]. This addition creates a potential minimum for electron collection underneath the deep p-well with a field direction towards the n collection well, thus strengthening the lateral collection of charges [20]. The epi-layer is 25 μm thick and the collection electrode is on positive potential. This modification of the technology has allowed to further improve radiation tolerance.

The most probable value (MPV) of the signal released by a Minimum Ionizing Particle (MIP) in the thin MAPS sensitive thickness ($\approx 25 \mu\text{m}$) is only about 2000 e^- .

The small charge signal achieved in the thin epi-layer (typically 1600 e^-) becomes a sizable voltage signal (about 50 mV) due to the small (≥ 5 fF) capacitance according to $dV = dQ/C$. Therefore voltage (rather than charge) amplification is employed for the readout of small electrode MAPS.

3.3.3 Silicon On Insulator (SOI) technology

Silicon on Insulator (SOI) technology represents another way to combine the sensitive region and the readout circuit in a single monolithic unit. In this architecture, the transistor is isolated by vertical trenches and is divided from the bulk by a SiO_2 layer, called *buried oxide* (BOX). An high resistivity bulk wafer allows to deplete the volume in the region below the BOX, to generate a large charge signal when a particle passes through the detector. The bulk and the CMOS electronics are connected by vertical connections, called *vias*. In Figure 3.9 a monolithic SOI pixel is displayed, with a doped volume placed between the CMOS circuitry and the BOX. This variation prevents the capacitive coupling from the substrate into the electronics, through the BOX. The SOI process is significantly more complicated than the DMAPS process, and in addition it has significant limitations in terms of radiation hardness, since under irradiation a large fixed charge accumulates in the thick oxide altering the electric fields in both the electronics and the sensor [24].

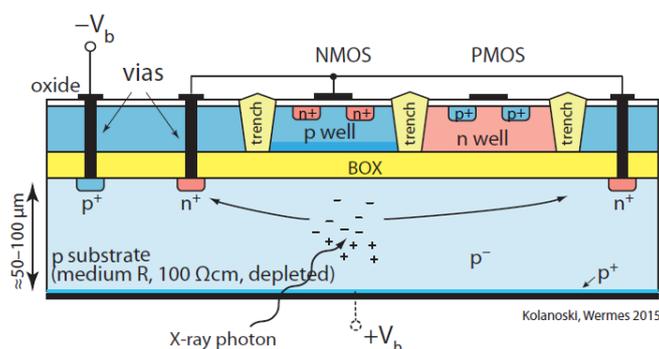


Figure 3.9: An example of a Monolithic SOI pixel.

3.4 History of Monopix developments

In recent years, requirements for high-energy physics experiments with high rates and high radiation environments have led to advances in CMOS technologies, and the development of a new generation of monolithic pixel sensors (DMAPS) with fast readout and high radiation tolerance.

Specifically two different DMAPS development lines have been followed, distinguished by different pixel architectures, using two different implementation process technologies [25]:

- **large fill factor** line: with large collection electrode and the electronics inside the charge collection well, these prototypes are indicated to experiments with high rate and high radiation conditions, because they could ensure a greater tolerance to large doses of radiation. They have been fabricated in LFOUNDRY 150 nm process [26]. In Figure 3.10 some of these chip developments are shown.
- **small fill factor** line: with small collection electrode and the electronics outside the charge collection well, these devices need a process modification to enhance the radiation hardness [23]. They are faster compared to the previous type, and due to the smaller values of total capacitance, they require much less power consumption. They are fabricated in

a TowerJazz 180 nm CMOS imaging process, modified to include the additional implant shown in Figure 3.8.

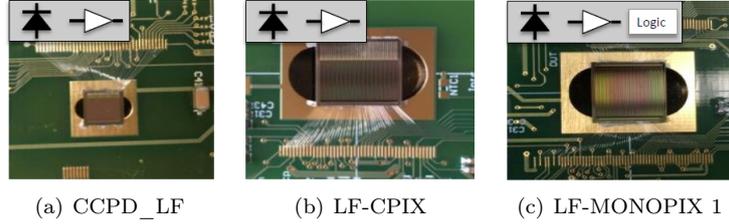


Figure 3.10: LFOUNDRY 150 nm development line.

3.4.1 TJ-Monopix line

In Figure 3.11 the TowerJazz development line is displayed, that have led to the design of the last iteration of this series, the TJ-Monopix2 chip, whose characterization results will be shown in Chapter 5.



Figure 3.11: TowerJazz 180 nm development line.

The standard TowerJazz 180 nm process has been employed to realize the ALPIDE monolithic active pixel sensor, selected for the ALICE Inner Tracking System (ITS) upgrade [22, 27].

The chip proved to be suitable for the ALICE requirements, but the standard process does not ensure the full depletion volume, which is crucial to limit the signal degradation especially after irradiation. For this reason, the aforementioned process modification have been developed by CERN in collaboration with the foundry [23], that allows full depletion of the sensitive layer. This new implementation has been tested in a dedicated chip called TJ-Investigator [28], and the results obtained have demonstrated the effectiveness of the modified process.

Therefore two large scale demonstrator chips have been realized, called TJ-Monopix1 and TJ-Malta1 [29], whose main difference lies in the readout architecture. The TJ-Malta1 chip implemented an *asynchronous* readout architecture, which eliminate the Bunch Crossing ID (BCID, timestamp), to reduce the digital power consumption. For TJ-Monopix1 instead, a *column-drain* readout architecture was chosen, which will be described in the following, as it is the same implemented in TJ-Monopix2. Both chips have been fully tested and irradiated to investigate their

functionality and efficiency, which however has decreased from 97% to 70% after irradiation with an equivalent neutron fluence of $10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$ [30].

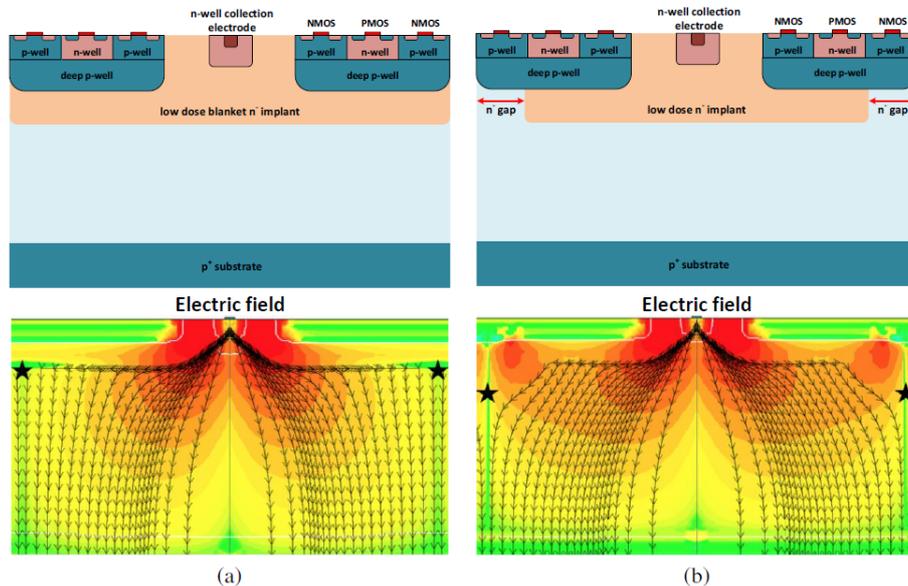


Figure 3.12: On the left, the lateral electric field for a pixel implementation in the modified process (only low dose n^- implant), compared with the modified process with the n-gap variant.

The main reason for the efficiency drop has been discovered to be related to the weak lateral electric field at the pixel edges. So the process has been further optimized to resolve this issue. Two different approaches were found to increase the lateral electric field at the pixel borders [31]: creating a gap in the deep n-implant **n-gap** variant, requiring only a mask change, or introducing an additional p-type implant at the pixel border **extra deep p-well** variant. The n-gap variant is shown in Figure 3.12 compared to the previous version.

It can be seen the significant improvement of the electric lateral field, which results in turn in faster charge collection and high efficiency even after irradiation.

Further optimization of the pixel size, which is critical to take full advantage of field shaping through process modifications and to improve charge collection, have been implemented in the last iteration of this development line: TJ-Monopix2, considered as starting point for the development of the OBELIX final chip, designed for the upgrade of the Belle II vertex detector.

3.4.2 TJ-Monopix2 architecture

We will briefly describe the chip architecture of TJ-Monopix2.

Analog front-end

The advantage of the small collection electrode concept is the small capacitance, which allows to transform the charge signal arriving on the sensing node into a high voltage signal ($dV = dQ/C$).

In Figure 3.13 a schematic of the voltage amplification and the readout stages is shown.

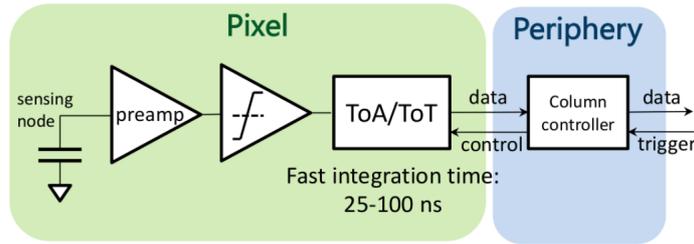


Figure 3.13: Sketch of the voltage amplification stages and the readout.

The charge signal, produced by particles traversing the sensor and collected by the collection node, is converted in a voltage signal through a small capacitance. It goes to a pre-amplifier and then to a discriminator, providing a digital output signal when the input is above the set threshold. The length of the output signal is the Time Over Threshold (TOT) and provides also information about the signal charge. The pre-amplifier and the discriminator working conditions are defined by several chip registers (as we will see in Chapter 5). The collected data are finally sent to the chip periphery in order to transmit and store them outside the chip.

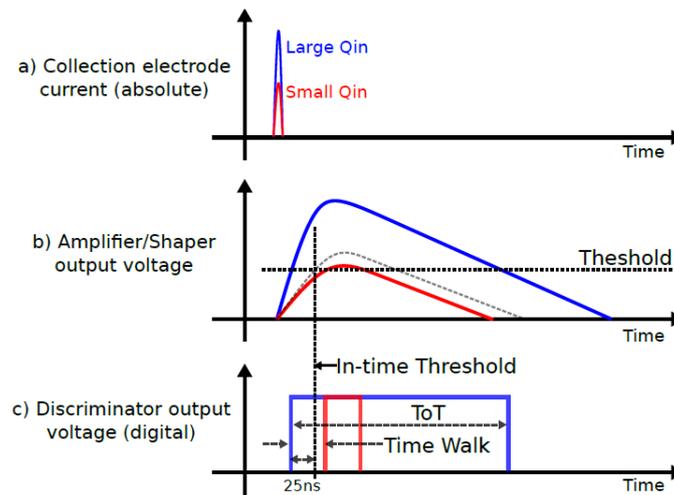


Figure 3.14: Time Over Threshold (TOT) technique.

The column-drain readout architecture

The readout architecture [32, 33] implemented in TJ-Monopix ensures fast readout by encoding the analog charge information using the standard Time Over Threshold (TOT) technique shown in Figure 3.14. This procedure exploits two timing information: the *Leading Edge* (LE) which corresponds to the hit time of arrival (when the signal value goes over the threshold), and the *Trailing Edge* (TE) that is when the signal goes below the threshold value. From the difference between the TE and the LE, the TOT can be calculated.

The *in-pixel* circuitry implements a Random Memory Access (RAM) to store the LE and TE timing info, a Read-Only Memory (ROM) to store the pixel address and the control and arbitration logic, essential to read out the hit, as described in the following.

The readout is *column-based*, so all pixels of each double column share a common column-bus which could be accessed by one pixel at a time, with a defined priority logic. The column bus includes the BCID timestamp, the data (LE, TE, address) and the control signals. The *periphery* includes the End Of Column (EoC) block which deals with the transmission and the readout of the column-bus signals, and the Digital Chip Bottom (DCB) that instead, processes the hit information. The overall scheme of TJ-Monopix2 is shown in Figure 3.15.

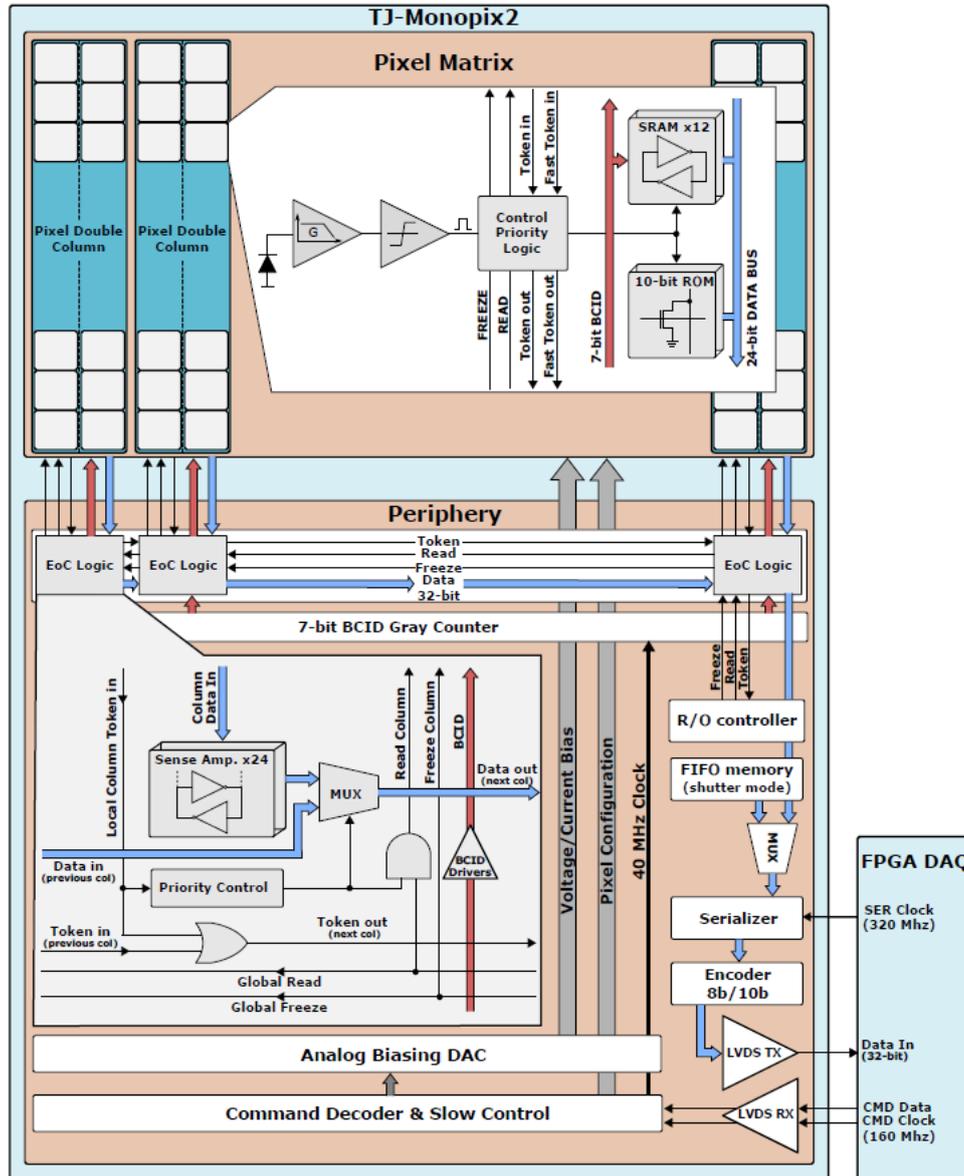
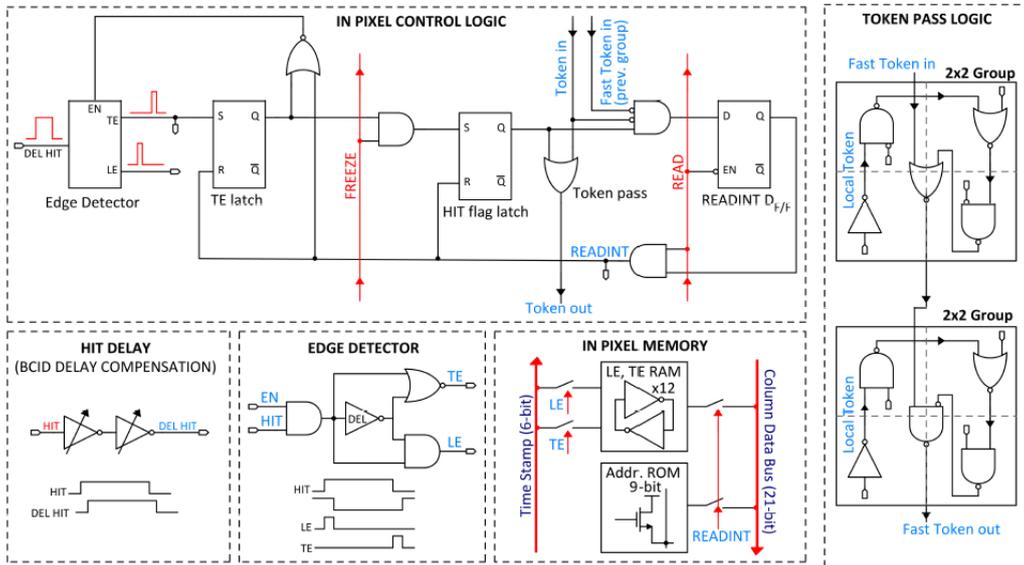
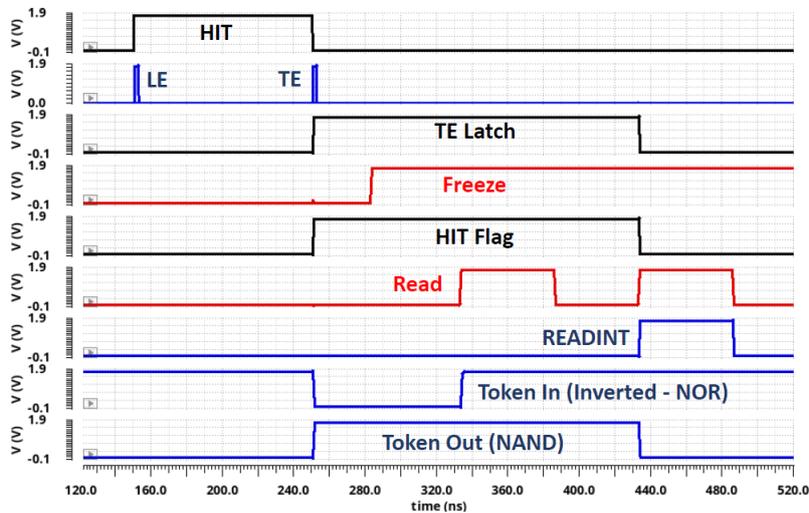


Figure 3.15: Architecture of TJ-Monopix2.

The column-drain architecture uses a TOKEN signal interpreted by the pixel control logic when data are available to be readout. Only pixels with a non-zero signal are readout, when going through a column, realizing an on-chip sparsification mechanism that allows fast readout. No local buffer is implemented, so a hit pixel remains frozen, and therefore insensitive, until it is readout.



(a) Schematic of TJ-Monopix2 in-pixel readout logic.



(b) Simulation of the in-pixel readout logic.

Figure 3.16:

The schematic and a simulation of the in-pixel readout logic are shown in Figure 3.16. The digital output signal coming from the FE is transmitted to an edge detector which generates the

TE and LE pulses. The TE and LE pulses force the storage of the current timestamp (6 bits) into a 12 bit memory cell. The TE signal is sent to an SR latch, which registers the hit and disables the edge detector, preventing new hits to be registered before reading out the previous one. If the column is not frozen, a HIT flag latch is activated, which signals the presence of the hit to be readout. This latch also activates the TOKEN signal which informs the readout controller that hits could be read. In TJ-Monopix2 two TOKEN signals are implemented: a fast one, which, taking advantage of pixel grouping in 2×2 cores, defines the higher (group) level priority and propagates across the double column; and a local one, which defines the priority within the group. When the readout controller receives the information from the TOKEN, the matrix is frozen by asserting the FREEZE signal. Then the READ signal is raised and data (TE, LE, pixel address information) are transmitted to the periphery through the column data-bus, accessible only if the TOKEN signals allow it (highest priority) and the HIT flag is active. In this case, the raising edge of the READ enables a D-latch and at the same time, resets the TE and the HIT flag latches. In the meantime, if the D-latch is set and enabled by the READ, an internal signal, called READINT is produced to control the access of the data to the column data-bus.

In conclusion the main features of the TJ-Monopix2 architecture can be summarized as follows:

- the chip is self-triggered and the hit pixels are frozen by the first pixel that completes the TOT cycle;
- each pixel stores its own address and the timestamps of the leading and trailing edges of the TOT;
- a token-passing logic is implemented to read the data from only the hit pixels without wasting clock cycles;
- while the readout is proceeding the free pixels can still store their TOT, but will be readout in a successive read cycle.

4. VTX detector

This chapter describes the vertex detector upgrade of Belle II, VTX. We will go through the VTX concept and layout, designed with a different geometry with respect VXD and with a new pixel sensors (OBELIX). We will discuss the motivations for some of the design choices and present some of the detector performance studies demonstrating the robustness against backgrounds.

4.1 VTX Layout and mechanical structure

As previously discussed (Chapter 2), reaching full luminosity will entail higher backgrounds and will probably require a re-design of the interaction region. It is therefore necessary to use a radiation-hard technology, with high granularity to reduce occupancy, and with a reduced material budget to improve the already excellent performance of VXD.

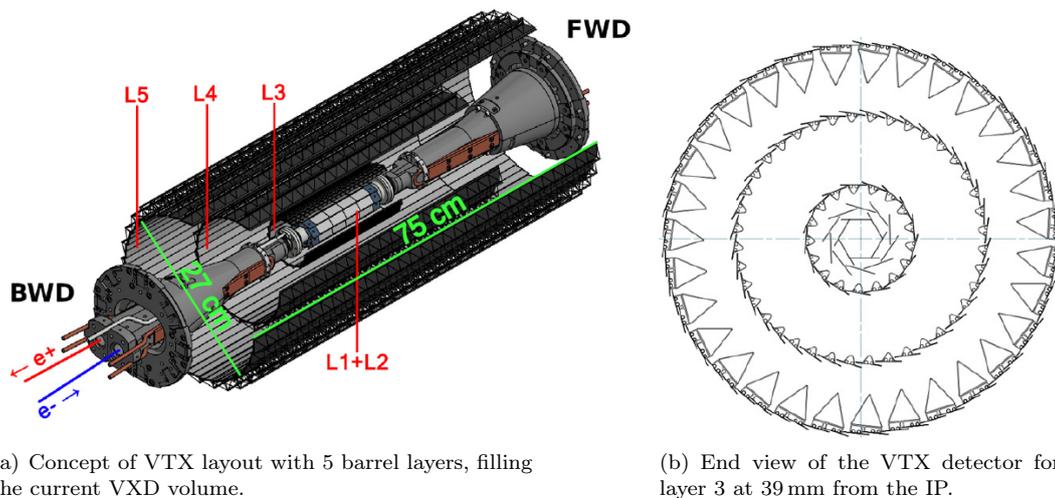


Figure 4.1: VTX detector layout.

The VTX project aims to replace the VXD with a fully pixelated detector based on Depleted Monolithic Active Pixel Sensors (DMAPS) arranged in five layers at different distance from the beam pipe (Figure 4.1) [34, 35]. The radii and the number of the layers are currently subject to several studies and simulations, in order to achieve an optimized arrangement (shown in Table 4.1). Currently two layers are planned in the innermost part (*i*VTX) and three in the outer part (*o*VTX). The active length of the ladders varies from 12 to 70 cm to cover the required acceptance of $17^\circ < \theta < 150^\circ$.

It is important to try to reduce the material budget, in order to minimize the multiple Coulomb scattering which particularly affects the very soft particles produced in Belle II, down to 50 MeV. By using the MAPS sensors, it is expected a reduction of the overall material budget down to less than 2% of radiation length, against the present 3% of VXD.

Layer	1	2	3a	4	5	Total	3b	Total
Radius (mm)	14.1	22.1	39.1	89.5	140.0		69.1	
# Ladders	6	10	17	40	31	104	30	117
# Sensors/ladder	4	4	7	16	2 × 24	2311	12	2552
Mat. Budget (X_0)	0.1	0.1	0.3	0.5	0.8	1.8	0.4	1.9

Table 4.1: The VTX detector main parameters with the two options (3a and 3b) for the radius of the middle layer. From [35].

4.1.1 iVTX

The *internal* VTX consists of the first two detector layers assembled with a self-supported air-cooled all-silicon ladder concept, where four contiguous sensors are diced out of a single wafer, thinned and interconnected with post-processed redistribution layers. They are designed to be at 14 and 22 mm respectively from the beam pipe, and target an individual material budget of about 0.1% radiation length. This is achievable because the overall surface of these layers is moderate, below 400 cm², the sensor power dissipation is expected to be low, and the connections needed for the operation to be only a few. Air cooling should be a workable solution to remove the heat produced.

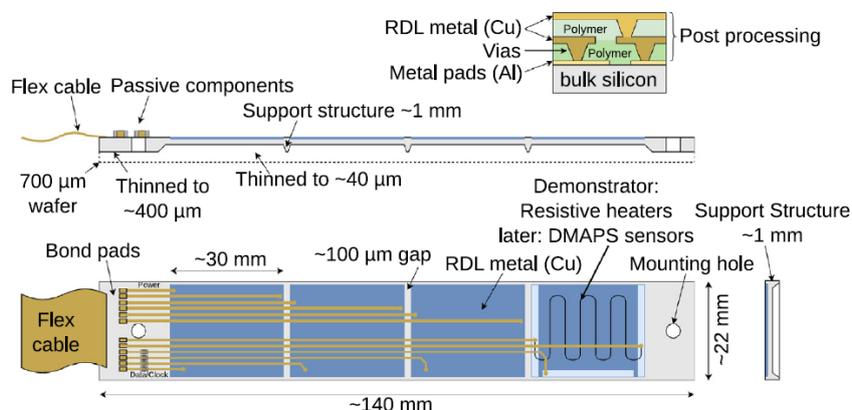


Figure 4.2: Sketch of the all-silicon ladder concept of the iVTX. Four dummy sensors are shown in blue on the silicon support in grey. The yellowish lines instead, indicate power and data transmission lines. Power is delivered to the ladder by a flex cable, which also transmits data to and from the chips in the final one.

The ladder has to be equipped with four OBELIX chips and thinned to 50 μm except in some border regions, where a few hundreds of μm are necessary to ensure mechanical stability. In order to interconnect the sensors along the ladder and provide a unique connector at the backward end, during the post-processing metal strips are etched on the redistribution layer (RDL). The

latter has the main purpose to route power and data via impedance-controlled transmission lines to a flex cable, added at the end of the ladder. After the RDL processing, the ladder is thinned as discussed above. Mounting holes will be added via laser-cutting.

Figure 4.2 shows a sketch of the iVTX demonstrator ladder (currently under production), 140 mm long and 22 mm wide (grey). Instead of the actual sensors, it is equipped with four dummy chips with a length of about 30 mm (blue), holding resistors to mimic the expected heat load to test the air cooling system and more generally to characterize the electrical, mechanical and thermal performance of the ladder. A redistribution layer for power and data is also added to the demonstrator, to connect the chips with a flex cable at the end of the ladder (yellowish lines). The wafer thickness is reduced to 400 μm , while the sensitive areas are thinned down to 40 μm with the purpose to test the mechanical integrity of the whole structure.

The R&D is ongoing and the full-silicon ladder concept is currently being assessed with industrial partners. The first thin ladders have been produced and characterised with different thicknesses and geometries.

Several tests are focused on evaluating power delivery efficiency, the quality of the signals which travel through the ladder and also the process used to fully assemble it. Figure 4.3 shows eye diagrams from simulation with a transfer rate of 640 Mbps, indicating that 320 Mbps of data rate will be possible.

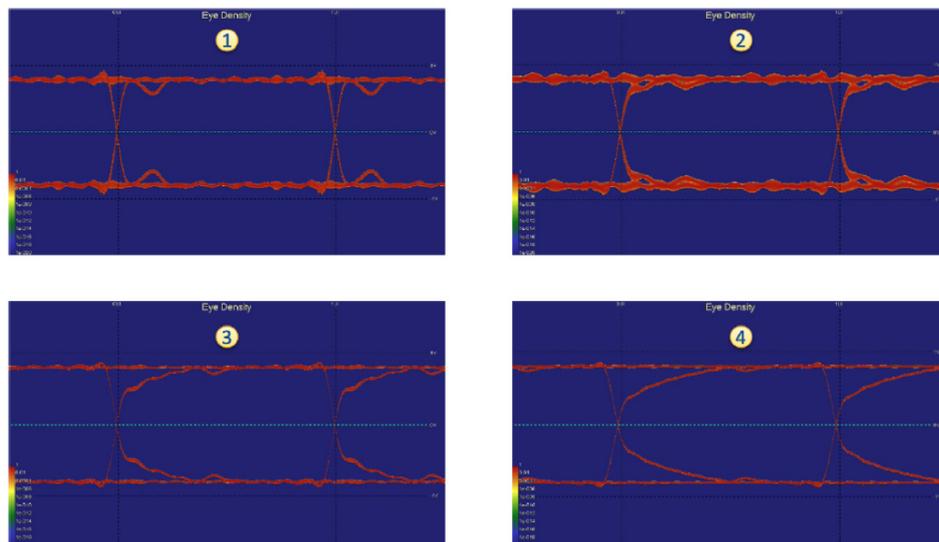


Figure 4.3: Eye diagrams of the iVTX data transmission lines at four different locations on the ladder. They represent the overlap of multiple samples of the output of transmission line triggered by the clock cycle. From [34].

Moreover, it has been demonstrated that air at the temperature of 15 $^{\circ}\text{C}$ and flowing with a speed of 10 m/s succeeds to cool a single inner module, assuming power is uniformly dissipated on the sensor surface. The maximum temperature reached is 20 $^{\circ}\text{C}$.

Through very first estimates it is expected that an equivalent section of 6 tubes with 10 mm of diameter is necessary to expel the heat from the inner layers, roughly equal to 65 W. So it is essential to design a mechanical structure which provides for the space needed to the tubes in order to bring the air at the IP and also compatible with the new interaction region.

4.1.2 oVTX

The *outer*VTX consists of three layers respectively at radii of 39 or 69, 89 and 140 mm from the interaction point. Because of the larger distance required to cover the acceptance, they cannot be self-supporting. They follow a more traditional approach, inspired by the design developed for the ALICE ITS2 [14]. Each ladder is water cooled and supported by a light carbon fiber support structure, which provides the mechanical stiffness. Its truss design is shown in Figure 4.4 : 70 cm long and 5.8 g of weight, it is able to support more than 40 sensors in two rows next to each other with a small overlap, leading to a material budget of 0.3% X_0 for the first layer, 0.5% X_0 for the second and 0.8% X_0 for the outermost one.

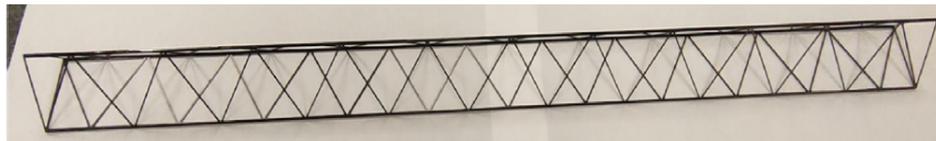


Figure 4.4: Prototype of the layer 5, called *truss*, which is the longest, made from thin carbon fibre structures.

For the cooling of the ladder a cold-plate concept is under development (Figure 4.5), on which the sensors are glued and that in turn is installed on the *truss*. For each row there is a polyimide cooling tube that runs over all the sensors and turns back at the other end, so that the cooling circuit connections are on the same side. Two flexible printed circuits are wire-bonded to the OBELIX chips and provide the connection to the readout.



Figure 4.5: A prototype of the cold-plate for cooling. One coolant tube(golden) is connected to the cold plate(black) and turns 180° on the other end (not shown) so that the coolant flows in both directions and thus leaves on the same side it starts.

For layer 3 instead, only one flex printed circuit cable in the backward side is considered, to leave more space forward for other possible services and accelerator components. For the third layer two different solutions are under study: at radius of 39 mm e 69 mm respectively. In Figure 4.6 schematic examples of some design hypotheses are displayed.

Figure 4.7 shows the several substructures described before, that constitute a ladder of the outermost layer 5. From bottom to top the carbon fibre structure, two cold-plates for the two neighbouring sensor rows (indicated as "Chips", in grey) and the flex cables for power and data transmission (green) follow one another.

The carbon fiber support structure and the flex cables have been designed and fabricated for layer 5, which is also the longest. Services for the last two layers, like electrical connections and cooling, can be provided both on forward and backward sides. A Multiline Power Bus has been realized to power each OBELIX chip along the ladder.

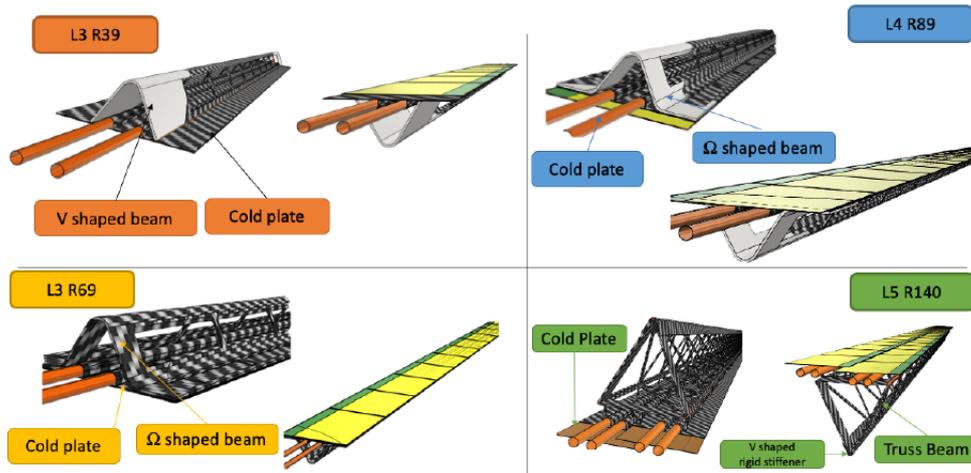


Figure 4.6: Schematic view of possible solutions for the three outermost layers.

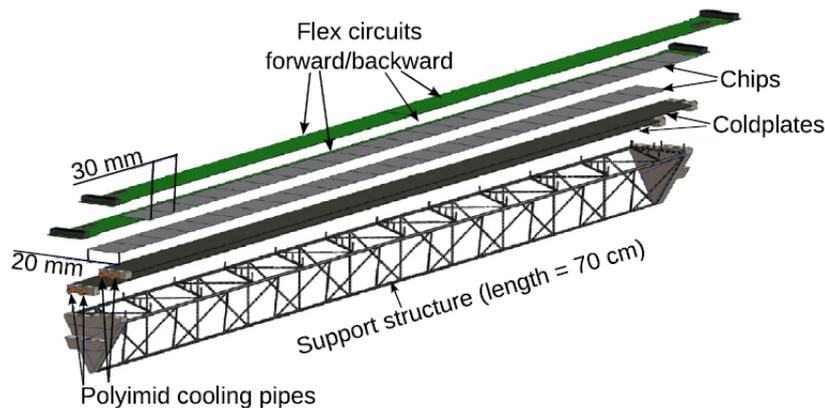


Figure 4.7: An exploding drawing of a fully assembled layer 5 ladder.

Mechanical tests have been performed showing that the first resonance frequency is at 200 Hz, which is safely far from the frequencies of the typical earthquakes in Japan. To reduce as much as possible the material budget, the transmission lines and the flex cables must be as thin as possible, but they also have to ensure safe data transmission.

In Figure 4.8 the resulting eye diagram from testing the signal integrity of one of the 35 cm long transmission lines for data transmission rate of 500 Mbps is shown. This result demonstrates that the bandwidth is large enough to allow the needed 160 Mbps for data transmission.

In addition, thermal tests have been performed for the last layer prototypes using kapton heaters to emulate the power dissipation of the chips. The coolant (demineralized water) has been set to a temperature of 10 °C at the beginning, the environment at 20 °C with a negative pressure of 0.2 bar. Results have demonstrated that for three different configurations of the flow (such as monodirectional, bi-directional and with an U-turn at one end) the average temperature

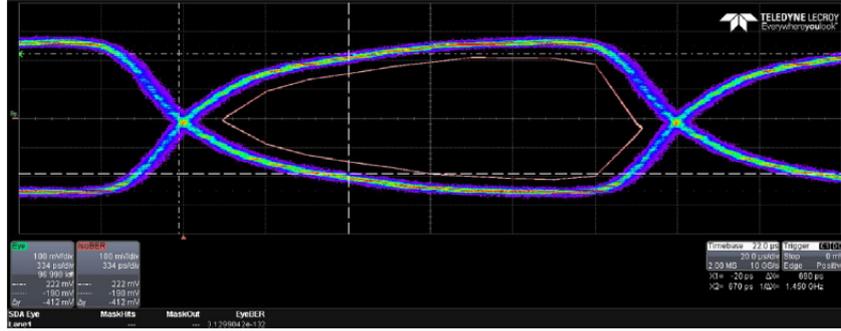


Figure 4.8: Eye diagram for the oVTX transmission line signal integrity of the layer 5 flex cable. From [34].

remains below 24°C with a maximum gradient of $\Delta T \approx 4^{\circ}\text{C}$ along the full length of the ladder.

All these investigations validate the design of the longest ladder, which is the most challenging, and therefore the possibility to operate the chips safely.

4.2 Performance simulation

As we have seen in Chapter 2, increasing the luminosity implies higher level of machine related background and larger doses of radiation, especially in the inner layers of the detector. For these reasons simulations and studies are focusing on ensuring that the main physics goals of the experiment will be achieved despite the more severe working conditions.

The VTX detector, with high granularity in both space and time and thin sensors, could bring significant improvements in tracking efficiency and resolution especially at low momentum, in the impact parameter resolution, and in the robustness against backgrounds. Moreover, better vertexing performance entails not only improved time-dependent analyses of B and D mesons, but also an enhanced capability to distinguish among different decay topologies, and a more powerful rejection of background events.

4.2.1 VTX geometries

Two different VTX geometries are currently under study, which differ only in the position of the third layer (Figure 4.6).

The *nominal* geometry is expected to maximize the track impact parameter resolution and it places the third layer at 39 mm from the IP. The *alternative* geometry instead, aims to improve the K_S^0 reconstruction efficiency and the third layer is located at 69 mm from the IP.

Simulations are ongoing to compare the performance of these two different layouts with that of the current Belle II detector (utilizing a full Geant4 simulation of the detector in the study of specific decay modes of interest) [36]. Moreover, the different machine background predictions are also simulated, to compare the detector performance under different background conditions.

4.2.2 Tracking efficiency at low momentum and impact parameter resolution

Tracking efficiency at low momentum is one of the areas where the VTX upgrade provides the most promising results, particularly for the *soft pions* originated from the decays of $D^{*\pm}$ mesons.

Studies are based on the reconstruction of the decay chain $B^0 \rightarrow D^{*-}l^+\nu$, with $D^{*-} \rightarrow \bar{D}^0\pi_{soft}^-$ and $\bar{D}^0 \rightarrow K^+\pi^-$ or $K^+\pi^-\pi^+\pi^-$. All background scenarios mentioned in section 2.2.2 are considered in the evaluation of the *nominal* VTX performance and they are compared with the nominal Belle II geometry in the intermediate (**v2**) background hypothesis.

	Belle II (v2)	VTX (v1)	VTX (v2)	VTX (v3)
Generated events	32533	32559	32559	30255
Correctly reconstructed signal	10059	16913	16848	15583
Combinatorial	28495	51375	51826	47527
Efficiency	30.9%	51.9%	51.7%	51.5%
Purity	26.1%	24.8%	24.5%	24.7%

Table 4.2: Reconstruction efficiency and purity for the the decay chain $B^0 \rightarrow D^{*-}l^+\nu$, with $D^{*-} \rightarrow \bar{D}^0\pi_{soft}^-$ and $\bar{D}^0 \rightarrow K^+\pi^-$, for the nominal Belle II detector at the intermediate background conditions (**v2**) and the nominal configuration of VTX in all three background scenarios.

In Table 4.2 the VTX reconstruction efficiency¹ in all three background hypotheses is reported, and it is improved by almost a factor 1.7 with respect to the nominal Belle II, with comparable purity². Moreover, efficiency remains practically stable in all background conditions, even in the most severe one.

This enhancement in tracking efficiency relies in particular on improved tracking efficiency for the π_{soft}^- mesons, as we can see in Figure 4.9 where the efficiency is shown as a function of momentum and polar angle.

4.2.3 Vertexing resolution

Studies on vertexing performance have been conducted using samples of one million $B^0 \rightarrow J/\psi K_S^0$ events generated and reconstructed with all the aforementioned combinations. The distributions of the decay vertex resolution σ_z (i.e. the width of the distribution obtained considering the differences between the measured and the true simulated positions) along the z axis of the B decay signal are shown in Figure 4.10.

In Table 4.3 a summary of the results that shows that the new geometries achieve a better resolution on the B decay vertex of about 35% on average and they also do not suffer of any significant degradation as the background conditions varies, unlike the nominal Belle II configuration.

Similar studies for the K_S^0 decay vertex resolution are displayed in Figure 4.11 and in the same way, the upgraded geometries reach a better vertexing resolution with respect to the nominal Belle II detector, without any significant degradation as the backgrounds increase.

¹Efficiency is defined as the ration between the number of correctly reconstructed signal events and the total number of candidates.

²In a few words, the probability that a correctly reconstructed signal is a "signal event".

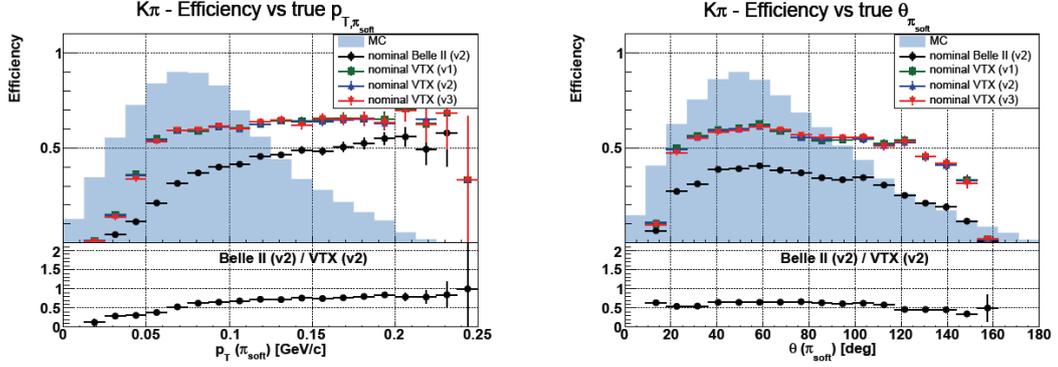


Figure 4.9: Reconstruction efficiency of $B^0 \rightarrow D^{*-}l^+\nu$ as a function of the transverse momentum of the π_{soft}^- (from $D^{*-} \rightarrow \bar{D}^0\pi_{soft}^-$) in the plot on the left and of the polar angle of the π_{soft}^- on the right. The shaded blue histograms represents the momentum spectrum of the π_{soft}^- . The nominal Belle II geometry efficiency in the intermediate background scenario (**v2**) is represented by black dots and it is compared with the nominal VTX configuration in the optimistic (**v1**, green squares), medium (**v2**, blue upward pointing triangles) and pessimistic (**v3**, red downward pointing triangles) background hypotheses. The bottom plots show the ration between nominal Belle II and nominal VTX in the **v2** background scenario. From [35].

B_{sig} z vertex resolution (μm)	Bkg (v1)	Bkg (v2)	Bkg (v3)
Belle II	21.9	23.0	24.9
Nominal VTX	14.5	14.4	14.1
Alternative VTX	14.4	14.3	14.0

Table 4.3: B_{sig} vertex resolution along the z axis for the three detector layouts and the three background scenarios.

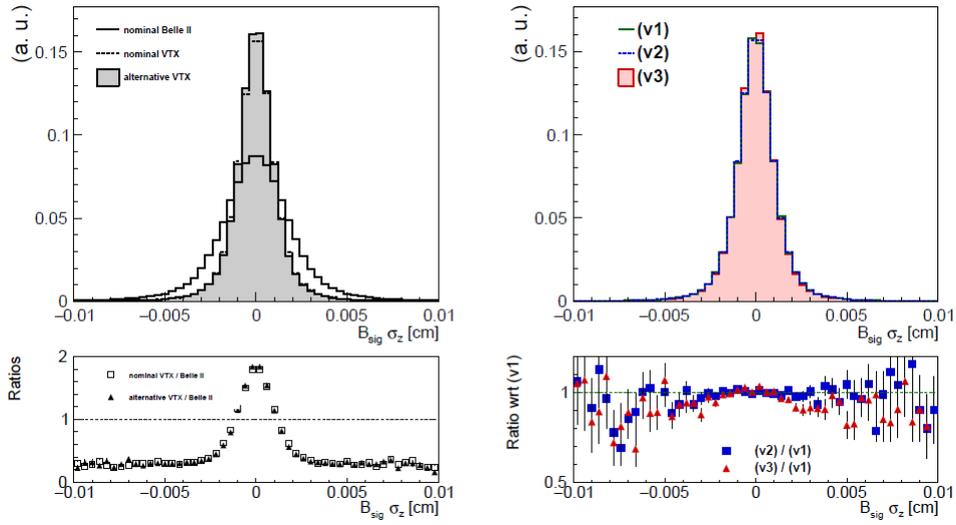


Figure 4.10: On the left: comparison of the B decay vertex resolution along the z axis in $B^0 \rightarrow J/\psi K_S^0$ events for the nominal Belle II (solid line), nominal VTX (dotted line) and alternative VTX geometry (filled grey histogram). The bottom plot shows the ratio between the VTX geometries (empty squares the nominal one and filled triangles the alternative) and nominal Belle II.

On the right: B decay vertex resolution along the z axis for the nominal VTX geometry in the three background scenarios: optimistic **v1** (green solid line), intermediate **v2** (blue dotted line) and pessimistic **v3** (red filled histogram). The bottom plot represents the ratio between the two higher background scenarios and the optimistic one. From [35].

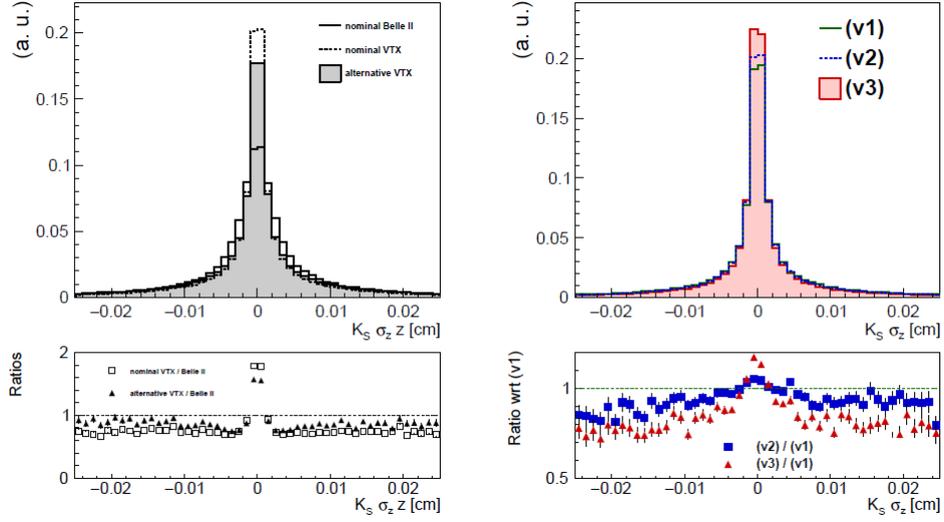


Figure 4.11: On the left: comparison of the K_S^0 decay vertex resolution along the z axis in $B^0 \rightarrow J/\psi K_S^0$ events for the nominal Belle II (solid line), nominal VTX (dotted line) and alternative VTX (filled grey histogram). The bottom plot shows the ratio between the VTX geometries (empty squares for the nominal and filled triangles for the alternative) and nominal Belle II detector. On the right: K_S^0 decay vertex resolution along the z axis for the nominal VTX in the three background scenarios: optimistic **v1** (green solid line), intermediate **v2** (blue dotted line) and pessimistic **v3** (red filled histogram). The bottom plot represents the ratio between the two higher background scenarios and the optimistic one. From [35].

4.3 OBELIX chip design

The VTX detector is designed with a single type sensor optimized for the specific needs of Belle II, called OBELIX (Optimized BELle II pIXel sensor) and currently under development, based on fast and high granular Depleted Monolithic Active Pixel Sensors (DMAPS). The sensor design comes from an evolution of TJ-Monopix2 chip, whose characterization will be discussed in Chapter 5. Both of them are fabricated in a modified TowerJazz Semiconductor 180 nm CMOS process.

4.3.1 Sensor specification

The main design specifications for the OBELIX chip are listed in Table 4.4.

Pixel pitch	30 to 40 μm
Matrix size	512 rows \times 928 to 752 columns
Time stamping	25 to 100 ns precision over 7 bits
Signal Time over threshold	7 bits
Output bandwidth	320 to 640 Mbps
Power dissipation	100 to 200 mW/cm ²
Radiation tolerance	1 MGy and 10^{14} n _{eq} /cm ²

Table 4.4: Designed specifications of the OBELIX sensor.

The pixel pitches³ are designed to be from 30 μm to 40 μm in both directions. This range is necessary to achieve a spatial resolution below 15 μm , which is required to obtain sufficient tracking performance [35].

Moreover the sensor thickness has to be below 100 μm to respect the material budget constraint of 0.2% X_0 (100 μm Si correspond to 0.1% X_0) which is possible with the DMAPS technology. To deal with the target hit rate of 120 MHz/cm², the timestamp clock signal can reach down to 25 ns, even if studies have demonstrated that a window of 100 ns (*integration time*) is enough to limit to 320 Mbps the data throughput at the same expected hit rate. The expected radiation tolerance must comply with the radiation robustness requirements mentioned in section 2.3. All characteristics inspected above allow to realize a sensor with high granularity in both space and time.

With respect to TJ-Monopix2, which is equipped with a triggerless column-drain readout without memory at the periphery (section 3.4.2), OBELIX must have a triggered readout architecture to satisfy the needs of Belle II. The latency is fixed to 10 μs and it might operate up to 30 KHz trigger rate.

Single Event Upset⁴ is a concern for future detector operation, therefore an important feature of the chip must be to ensure that the control system is able to reset the sensor registers to default operational values at least every minutes. The reset frequency will be chosen after the measurement of the SEU cross section with OBELIX and the comparison to the occurrence distribution of large energy loss in the experiment.

³The distance between the centers of two contiguous pixels.

⁴A Single Event Upset (SEU), also known as a Single Event Error (SEE), occurs if the charge released by ionizing particle, happens to be deposited close to a sensitive node of the device (RAM cell, register), causing a bit value to flip leading to corrupt information.

The power consumption is expected to be about 200 mW/cm^2 , a value which should allow air-cooling for the small areas corresponding to the two inner layers and liquid coolant for the outer ones. This relatively high value is necessary to obtain the required speed of the chip.

4.3.2 Sensor implementation

A schematic layout of the chip is shown in Figure 4.12. The sensor size is expected to be $3 \times 1.9 \text{ cm}^2$, with an active area of $3 \times 1.6 \text{ cm}^2$ and an additional part in the periphery of about $3 \times 0.3 \text{ cm}^2$, dedicated to data pre-processing and triggering.

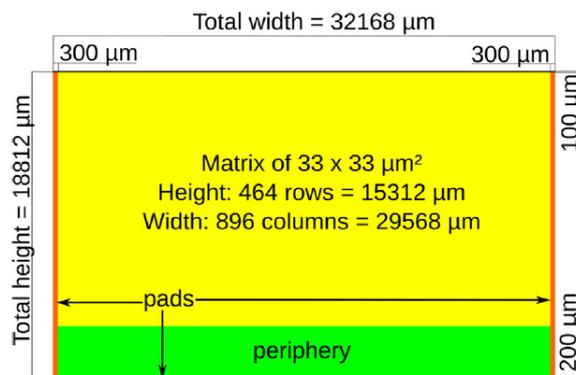


Figure 4.12: OBELIX chip layout.

As mentioned above, this new sensor is the development of TJ-Monopix2, whose characteristics fit already many of the Belle II requirements (Table 4.5).

From TJ-Monopix2 design, the pixel size of $33 \times 33 \mu\text{m}^2$ is maintained, as well as the layout of both digital and analog parts (section 3.4.1). Also the Time-Over-Threshold method to digitize the signal is preserved, with a bus width of 7 bit, together with the column-drain readout architecture implemented for pairs of columns. Other features which will be explained in depth in Chapter 5, have been maintained in the new design, like the 3-bit register dedicated to the threshold tuning, but with a larger range of correction (for the last bit). To aim at the integration time of 100 ns, the clock frequency which defines the precision of TOT and BCID (that is the timestamp), has been decreased from 40 to 20 MHz. Thus the current baseline for OBELIX timestamp precision is 50 ns.

Additionally, two new modules have been added to the implementation, related to the Belle II trigger: the Trigger Logic Unit (TRU) and the Track Trigger Transmitter (TTT).

Trigger Logic Unit (TRU)

The TRU has the task to select the fired pixel information from the matrix which are in-time with the triggers sent by the Belle II system. For this purpose it employs two stages of memory (Figure 4.13): the first stage has to store the pixel information during the trigger delay; the second memory has to compare the BCID (Bunch Crossing ID) of the fired pixel with each trigger time information buffered in a dedicated global memory. When they have a match, the pixel data is transferred to the Transmission Unit (TXU). Considering the BCID time precision, the time integration of the OBELIX sensor becomes 100 ns.

	Specification	TJ-Monopix2
Pixel pitch	$< 40\mu m$	$< 33\mu m$
Sensitive layer thickness	$< 50\mu m$	$30\mu m$ and $100\mu m$
Sensor thickness	$< 100\mu m$	-
Hit rate capability in the matrix	$> 600 \text{ MHz cm}^{-2}$	$> 600 \text{ MHz cm}^{-2}$
Hit rate capability at the sensor output	$> 120 \text{ MHz cm}^{-2}$	$\gg 100 \text{ MHz cm}^{-2}$
Trigger delay	$> 10\mu s$	-
Trigger rate	30 kHz	-
Overall integration time	$< 100 \text{ ns}$	-
(optional) Time precision	$< 50 \text{ ns}$	-
Total ionizing dose tolerance	100 kGy/year	1 MGy/year
NIEL fluence tolerance	$5 \times 10^{13} \text{ n}_{\text{eq}}/\text{cm}^2/\text{year}$	$1.5 \times 10^{15} \text{ n}_{\text{eq}}/\text{cm}^2/\text{year}$
SEU tolerance	flash configuration (min^{-1})	-
Matrix dimensions	around $30 \times 16 \text{ mm}^2$	$19 \times 19 \text{ mm}^2$
Overall sensor dimensions	around $30 \times 19 \text{ mm}^2$	$20 \times 19 \text{ mm}^2$
Powering	through voltage regulators	-
Outputs	one at $< 200 \text{ MHz}$	one at 160 MHz

Table 4.5: OBELIX sensor specifications, compared to the relevant specifications of the TJ-Monopix2 sensor.

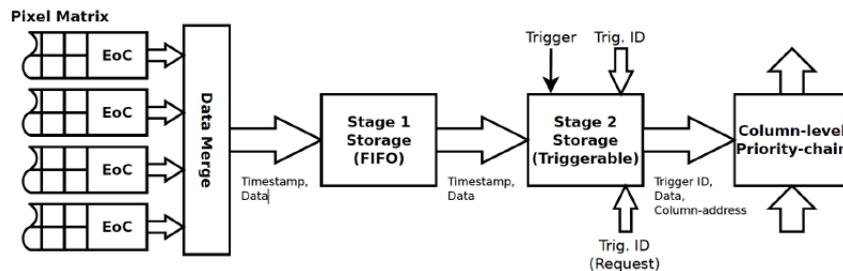


Figure 4.13: Schematic of the Trigger Logic Unit.

These components are designed to minimize power dissipation and to optimize the efficiency, even in severe operating conditions: maximum hit rate of 120 MHz/cm^2 , 30 KHz of trigger rate and $10 \mu s$ of trigger delay.

Track Trigger Transmitter (TTT)

It is under study the possibility for OBELIX to generate information to be used for a fast track reconstruction at trigger time. The TTT module divides the matrix in 8 logic regions (this value is still under study) and generates a one-byte word depending on the region firing. It is expected that this information could be transmitted to trigger system within 100 ns and along a line of transmission parallel to the main data output of the sensor. This component behaviour is still under study and it needs of further simulations in correlation with the whole VTX system.

Control Unit (CRU) and power dissipation

As well as TJ-Monopix2, the main features of the OBELIX chip operation could be configured by several registers, which allow to determine the threshold settings, masked pixels, time response of the pixels, and which also define its power consumption. The Control Unit is responsible for receiving these configuration instructions and the trigger information, and at the same time, sending out data coming from TXU module.

For what concern power dissipation, there are three main features which have the greatest impact: the biasing current flowing into the in-pixel amplifier (I_{BIAS}), the BCID clock frequency (on which the timestamping precision depends) and the hit rate. In Figure 4.14 the estimations of power dissipation varying these parameters is shown.

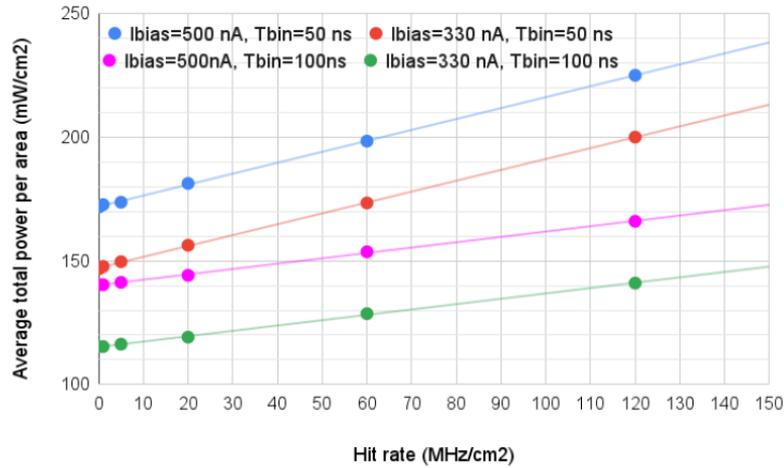


Figure 4.14: OBELIX sensor power dissipation depending on the front end current (I_{bias} , the BCID frequency (t_{bin}) and the hit rate).

As we can see, the power consumption at the maximum hit rate of 120 MHz/cm² exceeds by little more than 10% the power budget of 200 mW/cm², considering the higher precision for the timestamp of 50 ns. Therefore, it is necessary to find a compromise between the value of timing precision and the biasing current to stay within the power budget.

The first version of the sensor, called OBELIX-1, is being designed. A second improved version, OBELIX-2, will be designed based on performance studies on the first version and it is expected that it will be the final sensor needed for the experiment.

5. TJ-Monopix2 characterization

In this chapter we will go through the main features of TJ-Monopix2 designed to address efficiency degradation after irradiation, one of the main issues of its predecessor TJ-Monopix1 (section 3.4.1). The characterization of the chip is crucial in the VTX upgrade program and in the design of the next OBELIX chip.

The chip W14R12, shown in Figure 5.1, is one of the sensors tested during the first Test Beam (TB) campaign in Desy (June 2022). It has been fully characterized in Pisa and in particular several aspects have been analyzed, among which:

1. TOT calibration by internal charge injection;
2. characterization with radioactive sources and absolute calibration;
3. systematic study of different registers' settings to operate the chip at low threshold;
4. investigation of an important issue with cross-talk, due to digital signal from the readout, discovered trying to operate the matrix at low threshold (below $250 e^-$).

This detailed characterization returned crucial results (points 1, 2 above) for the Test Beam data reconstruction and the simulations of the upgraded VTX detector with CMOS MAPS devices. Furthermore the optimization of the registers to reduce the operating threshold was very useful for the preparation of the next TB (July 2023) campaign with irradiated sensors.

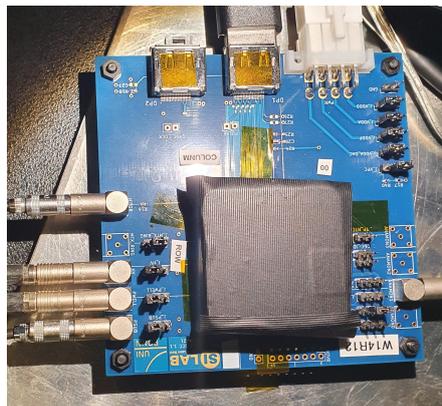


Figure 5.1: The W14R12 chip tested during the Test Beam in Desy.

5.1 Matrix and flavors

TJ-Monopix2 is a small collection electrode DMAPS prototype in TowerJazz 180 nm process. The need to create a sensor capable to maintain high efficiency even after irradiation, required improvements compared to TJ-Monopix1 in two important fields: a lower operating threshold, to keep a good efficiency with the reduced charge collected after irradiation, and a smaller pixel pitch, to increase charge collection efficiency all over its area, especially in the corners.

To achieve these goals, a different in-pixel front-end circuit was implemented and many efforts were focused on the pixel layout optimization to reduce its size, which was decreased from $36 \times 40 \mu\text{m}^2$ in TJ-Monopix1 to $33.04 \times 33.04 \mu\text{m}^2$ (pixel pitch). The pixel dimensions are critical to accomplish faster charge collection across all active area, increasing the lateral electric field (section 3.4.1) [31]. For this reason it was necessary a special effort to design and create a smaller pixel pitch but still adequate to embody the full digital readout. All this required to work at the technology density limit and to optimize the circuit design.

In order to operate with a lower threshold, TJ-Monopix2 incorporates an improved front-end circuit that reduces the noise by $\approx 40\%$ and the threshold dispersion by about 80-90% with respect to TJ-Monopix1. Furthermore, in-pixel threshold tuning has been integrated to achieve a more uniform threshold distribution across the pixel matrix, particularly after irradiation. As a result of these improvements, the operating threshold in optimized working conditions, was expected to be at $\approx 100 e^-$, three times lower than in TJ-Monopix1, the noise less than $8 e^-$, and the threshold dispersion less than $10 e^-$ (Table 5.1) [33].

Specification	Value
Threshold	$\approx 100 e^-$
Noise	$\leq 8 e^-$
Threshold dispersion	$\leq 10 e^-$

Table 5.1: Expected performance of TJ-Monopix2.

5.1.1 Flavors

The prototype is a $2 \times 2 \text{ cm}^2$ pixel matrix, $300 \mu\text{m}$ thick, and consisting of 512×512 pixels. The total active area is approximately 286 mm^2 .

As shown in Figure 5.2, the matrix is divided in four sectors, named *flavors* that implement different variations of the front-end circuit. In the first two flavors the collection electrode is directly DC-coupled to the readout electronics; the continuous baseline reset is implemented by a forward bias diode, but they differ in the pre-amplifier circuit design. The second flavor, named **Cascade FE**, includes an extra-cascade transistor that increases the pre-amplifier gain, which in turn leads to a 50% reduction of the threshold dispersion compared to the first flavor, the **Normal FE**. The other two flavors consist of AC-coupled pixels (through a metal-oxide-metal MOM capacitor) and in particular, the **HV-Cascade FE** also incorporates the aforementioned pre-amplifier variation. AC-coupling allows to apply a high positive bias voltage (HV stands for High Voltage) to the collection electrode widening the depleted region, but at the same time causing signal losses mainly due to the additional parasitic capacitance introduced into the sensitive input node.

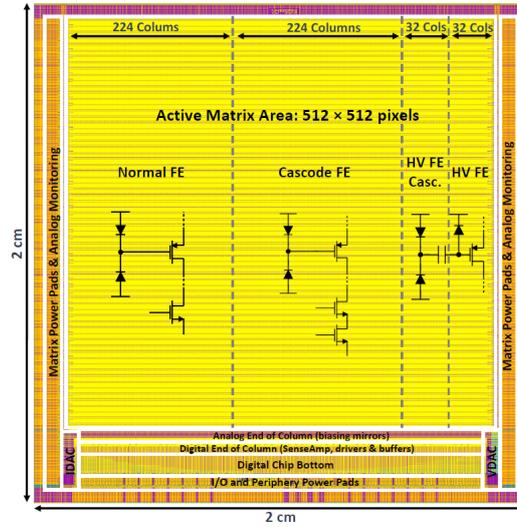


Figure 5.2: The layout of the TJ-Monopix2 prototype divided in four different flavors: Normal, Cascode, HV-Cascode and HV FE.

5.1.2 Pixel design

The 2×2 pixel core layout, shown in Figure 5.3, is fully optimized and it is designed to share as much features as possible between the four pixels. The analog area incorporates the front-end circuit, the 3-bit threshold tuning DAC and the pixel configuration registers. The digital region is composed by the 7-bit Leading Edge (LE) and Trailing Edge (TE) memory (14 SRAM cells per pixel), the 10 bit address ROM (2 bit for the pixel position inside the core and 8 for the group address) and the readout control logic.

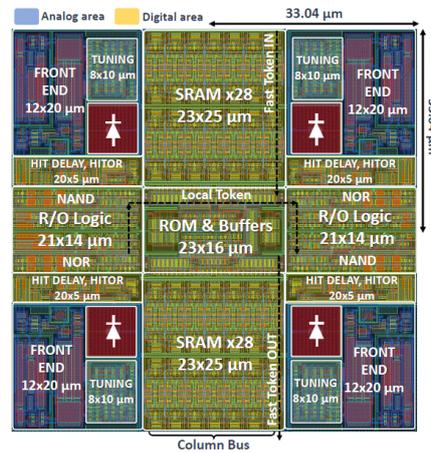


Figure 5.3: Layout of a TJ-Monopix2 2×2 pixel core. In blue the analog area and in yellow the digital one.

As described above, there are two variations of the front-end circuit (Figure 5.4), the *nor-*

mal and the *cascode* type. The latter in particular includes an extra-cascode transistor which increases the pre-amplifier gain and consequently reduces the threshold dispersion (equation 5.3 in section 5.2). The front-end operating point could be set changing the registers' values, which define the various voltages and currents of the circuit illustrated in Figure 5.4.

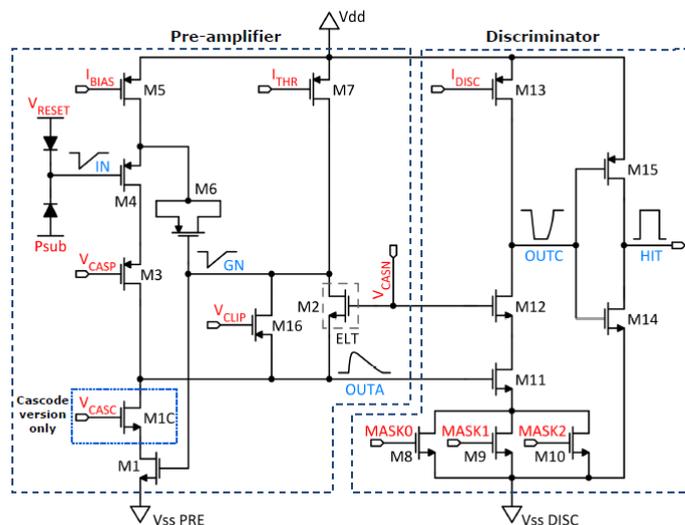


Figure 5.4: Schematic of the improved front-end circuit and its variation (extra-cascode transistor) in TJ-Monopix2.

In the last two AC-coupled flavors the same improvements have been implemented, but here the different coupling causes an important loss in the collected charge, as verified during the testing phase of TJ-Monopix1 (50% losses), due at most to additional parasitic capacitances. Thus a lot of efforts have been made to improve this aspect, working on the coupling capacitor values. A signal loss of 41.5% has been reached in TJ-Monopix2, which is a relevant enhancement with respect to its predecessor [33].

5.2 Threshold and noise

The most probable value (MPV) of the signal released by a Minimum Ionizing Particle (MIP) in the thin MAPS sensitive thickness ($\approx 30 \mu\text{m}$) is only about $2250 e^-$ (assuming $75 e^-/\mu\text{m}$ [20]). This small signal is collected by 2 or more pixels and could be further reduced after radiation damage due to trapping effects. In MAPS sensors it is then crucial to set the discriminator threshold to very low values in order to retain high detection efficiency, even after radiation damage.

The minimum value of the global threshold, compatible with a stable detector operation, depends on two important figures of merit: the noise and the threshold dispersion across the matrix.

Noise and equivalent noise charge. Electronic noise from the sensor diode and the pre-amplifier transistor (or other) devices results in time-varying voltage fluctuations at the pre-amplifier output. The Root Mean Square (RMS) of the voltage fluctuations is called noise output voltage. To quantify the noise of a readout channel the equivalent noise charge (ENC)

is introduced, defined as the corresponding charge, in electrons, at the input of the channel that gives an output signal equal to the RMS noise output voltage.

$$ENC = \frac{\text{noise output voltage (V)}}{\text{gain (V/e}^-)} \quad (5.1)$$

Operating threshold and threshold dispersion. The global detection threshold has to be set to a value as low as possible in order to maximize the detection efficiency, but not too low in order to keep the noise hits at an acceptable level. Therefore the operating threshold value is ideally located between the noise peak around the baseline and the signal distribution. The dispersion of the threshold across the pixel matrix originates from the fabrication process variations and the mismatch of the integrated electronic components. These process related effects are responsible for dispersion of the quiescent voltage at the output of the pre-amplifier (baseline dispersion) and also for dispersion of the discriminator voltage offset, both contributing to the dispersion of the detection threshold.

The threshold and its dispersion are also usually quantified reporting the corresponding charge, in electrons, at the input of the channel that gives an output signal corresponding to the threshold voltage and its dispersion:

$$THR(e^-) = \frac{\text{voltage threshold (V)}}{\text{gain (V/e}^-)} \quad (5.2)$$

$$\sigma_{thr}(e^-) = \frac{\text{threshold dispersion (V)}}{\text{gain (V/e}^-)} \quad (5.3)$$

The noise sets a lower bound to the threshold, but does not determine the stable operating threshold value which additionally depends on how the threshold varies with time and from pixel to pixel. The operating threshold lower limit is chosen to be at least approximately 5-6 times higher than the combined standard deviation of noise and threshold dispersion:

$$THR = (5 - 6) * \sqrt{ENC^2 + \sigma_{thr}^2} \quad (5.4)$$

Therefore, apart from the global threshold, the possibility to fine tune the threshold locally per pixel is usually included in order to compensate these variations and is realized by digital to analog converters (DACs). The dispersion of the threshold after tuning is inversely proportional to the number of bits of the tuning DAC (TDAC):

$$\sigma_{thr,tuned} = \frac{\sigma_{thr}}{2^{n_{dac}}} \quad (5.5)$$

Since both the ENC (equation 5.1) and the threshold dispersion (equation 5.3) are inversely proportional to the gain, a lower global threshold could be achieved increasing the gain, through an optimized choice of chip registers.

To measure the threshold and noise of the whole matrix, the response of each pixel has been characterized by internal charge injection.

5.2.1 Injection circuit

The hit injection circuit included in TJ-Monopix2, allows to produce artificial hits on each pixel through an injection capacitance C_{inj} connected at the collection electrode. The injected charge is proportional to the injection voltage pulse amplitude: $Q_{inj} = C_{inj} \cdot \Delta V_{inj}$. The injection pulse is set by two registers " V_L " and " V_H ", with $\Delta V_{inj} = V_H - V_L$, and the minimum injection step is given by the DAC resolution, with the Least Significant Bit (LSB) = 7.03 mV.

The injected signal is then often expressed in DAC units $Q_{inj}(DAC)$ and can be converted to electrons using the design value of the injection capacitance $C_{inj} = 230$ aF, the same value for all FEs implemented. The nominal conversion factor K from DAC to e^- corresponds to the injected charge given by a voltage step of 1 DAC:

$$K = C_{inj} \cdot LSB = \frac{230 \text{ aF}}{q_{e^-}} \cdot 7.03 \frac{\text{mV}}{\text{DAC}} = 1.4375 \frac{e^-}{\text{mV}} \cdot 7.03 \frac{\text{mV}}{\text{DAC}} \approx 10.1 \frac{e^-}{\text{DAC}} \quad (5.6)$$

An absolute calibration of the conversion factor K (i.e. of the injection capacitance C_{inj}) has been also performed using radioactive sources, as explained in section 5.4, obtaining results in agreement within 10% from the design value.

The conversion factor of equation 5.6 has been used to convert the information of the injected charge from DAC unit to electrons unit, useful for further analysis.

The response of each pixel to internal injection has been measured to extract their threshold and noise with the *S-curve method* explained in the next section. The four flavors have been analyzed separately to be able to study their main differences concerning their performance and features.

5.2.2 S-Curve method

The response of the pixels is measured injecting different amounts of charge into each pixel a given number of times and recording the amount of registered hits (i.e. the signal is above the discriminator threshold). For each value of the input signal we measure the occupancy, or hit probability, as the fraction of events where the pixel has registered a hit. This occupancy has the typical S-curve shape shown in Figure 5.5 as an example.

As the injected signal passes the discriminator threshold the pixel starts to register some hits, finally reaching a plateau corresponding to the total number of injected events. This behaviour produces a step function smeared by the fluctuation on the input signal due to the noise. The threshold of the pixel corresponds to the injected signal that gives 50% occupancy, while the noise influences the slope of the S-curve.

Assuming a gaussian noise distribution the S-curve can be fitted with the Cumulative Distribution Function (CDF):

$$CDF(x, \mu, \sigma) = \frac{1}{2} \cdot \left(1 + \operatorname{erf} \left(\frac{x - \mu}{\sigma \sqrt{2}} \right) \right) \quad (5.7)$$

where "*erf*" is the Gauss error function, x is the injected signal, μ and σ are the threshold and the noise of the pixel, respectively.

This method allows to measure the noise and the threshold of all pixels and also the threshold dispersion across an entire FE.

In the following the results of this study for the four flavors of the matrix are reported. The injected signal was varied, with the corresponding voltage injection registers " V_L, V_H ", from 0

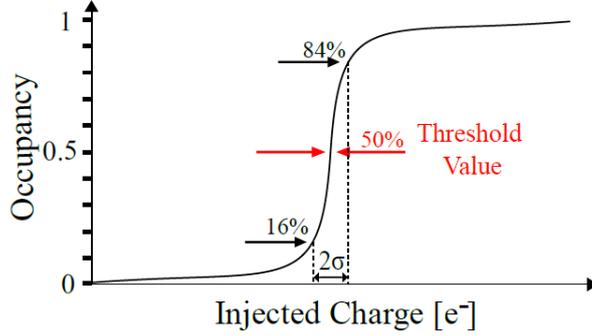


Figure 5.5: An example of the S-Curve fitted by the CDF to evaluate threshold and noise.

to 140 DAC, corresponding to a charge of $\approx 1400 e^-$, adopting the nominal conversion factor K in equation 5.6 .

Normal FE

The first flavor of the matrix is the **Normal FE**, which consists of 512 rows and 224 columns for a total of 114.688 pixels. The chip registers have been set with the same values used during the Test Beam at Desy (June 2022), which are different for the DC and AC-coupling cases. The most relevant among those responsible for the FE working point are reported in Table 5.2, where the different biasing voltages used to power up the chip are also added. At the time of this first Test Beam campaign the front-end registers were still not optimized to reach low threshold values.

Registers	Normal/Cascode FE ($P_{SUB}/P_{WELL} = -3$ V)	HV/HV-Cascode FE ($P_{SUB}/P_{WELL} = 0$ V, HV = +5 V)
I_{THR}	64	30
I_{BIAS}	50	60
V_{RESET}	143	100
I_{CASN}	0	8
I_{DB}	100	100
I_{TUNE}	53	53

Table 5.2: Settings of the main registers used for all flavors (W14R12 chip) during the Test Beam in Desy. The register meaning is explained in section 5.5.1.

In Figure 5.6 the S-curves of the all Normal flavor pixels are plotted. The width of the figure is a first indication of the threshold dispersion across the whole flavor. For each pixel the S-curve was fitted with the function in equation 5.7 to extract their threshold and noise. The threshold and noise distributions measured for all pixels of the Normal FE sub-matrix are shown in Figure 5.7, with their maps too. A fit with a gaussian distribution is used to extract the average threshold and noise values and their dispersion across the sub-matrix, and results are summarized in Table 5.3.

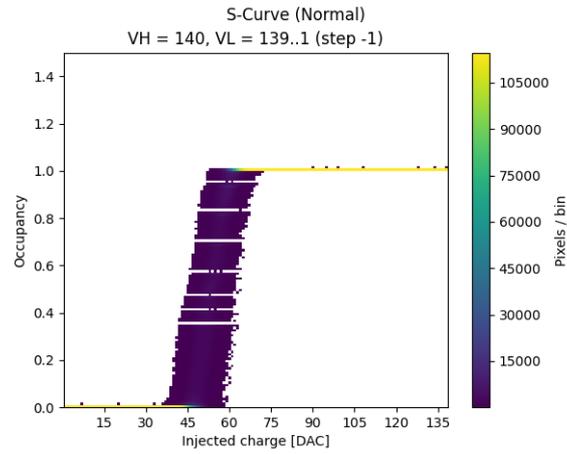


Figure 5.6: S-curves of all pixels of the Normal FE with a maximum injection pulse of 140 DAC.

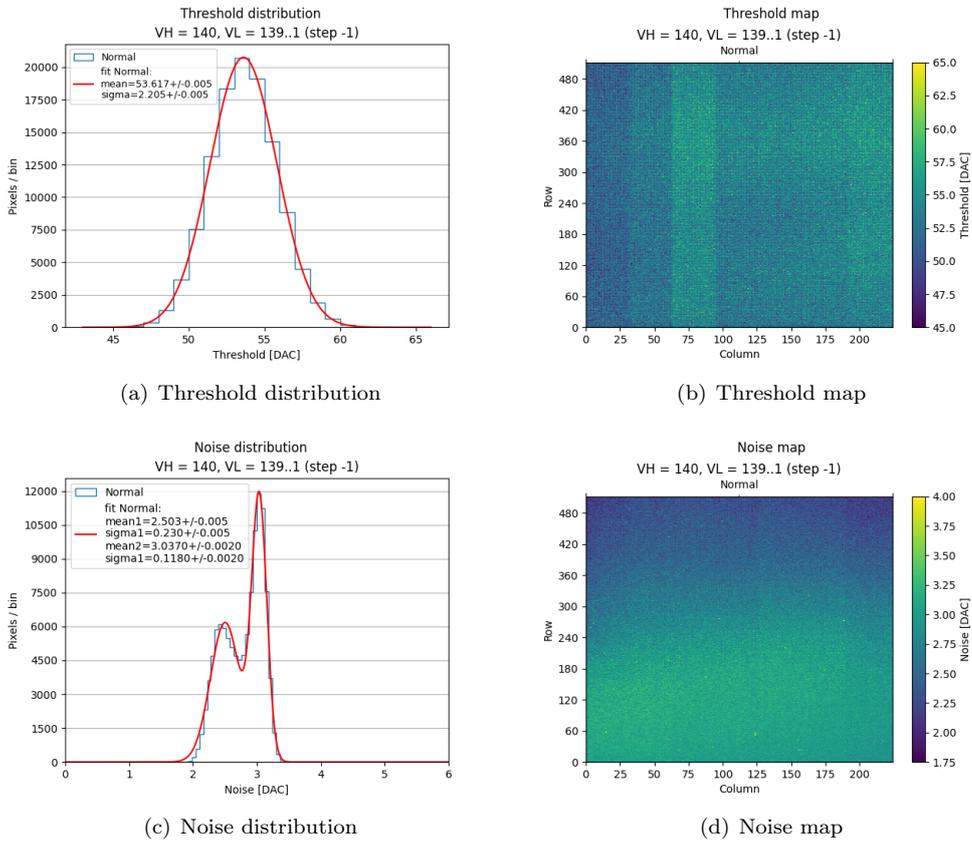


Figure 5.7: Distributions and maps of threshold (first row) and noise (second row) in the entire Normal flavor.

The striped pattern visible in the threshold map, is due to the power distribution network. In more detail, the matrix power pads are distributed along its left and right sides (Figure 5.2), therefore the voltage drop across the matrix analog power grid is compensated by using a local analog matrix supply, for groups of 32 columns [33], which determine this pattern.

The noise map shows a different behaviour in the upper part of the flavor, which has a smaller noise value on average, with respect to the lower part, where it results higher. We did not investigate further this behaviour in this phase of the characterization, although we believe the effect was eliminated by a change in the chip operating conditions.

Cascode FE

The **Cascode FE** is the second flavor and, like the previous one, it consists of 512 rows and 224 columns for a total of 114.688 pixels. For this flavor the same procedure of the Normal FE has been followed and the same registers' values (Table 5.2) have been used. In Figure 5.8 the S-curves of all pixels are shown.

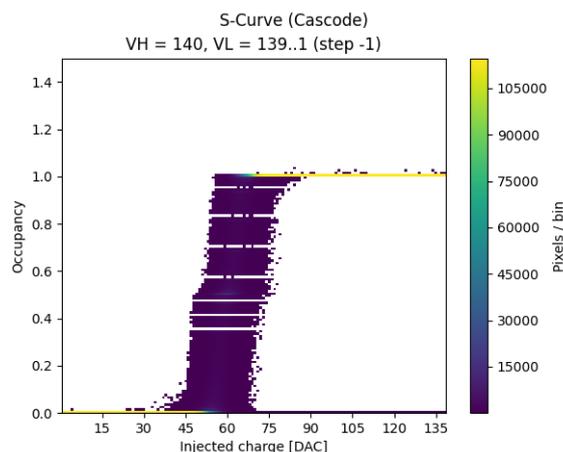


Figure 5.8: S-curves of all pixels in the Cascode flavor with a maximum injection pulse of 140 DAC.

The distributions of threshold and noise extracted with the fit to the S-curve, and their maps, are shown in Figure 5.9. The striped pattern due to the power distribution network is less evident, and the noise map is more uniform, too. Results for Cascode FE flavor are also summarized in Table 5.3.

HV-Cascode FE

The third flavor is **HV-Cascode FE** and it is composed by 512 rows and 32 columns for a total number of pixel equal to 16384. Also for these last two flavors, the main chip registers are set with the same values tested and used during the Test Beam (but different from those used for the first two flavors). They are reported in the second column of Table 5.2.

As we can see from the plot of all the S-curves in Figure 5.10, with this choice of registers there were a lot of "hot" pixels with occupancy > 1 , but at this stage of measurements they were not masked. These hot pixels with occupancy > 1 , register more hits than the number of injected events. This behaviour seems to indicate that they are stimulated, not by the charge injection

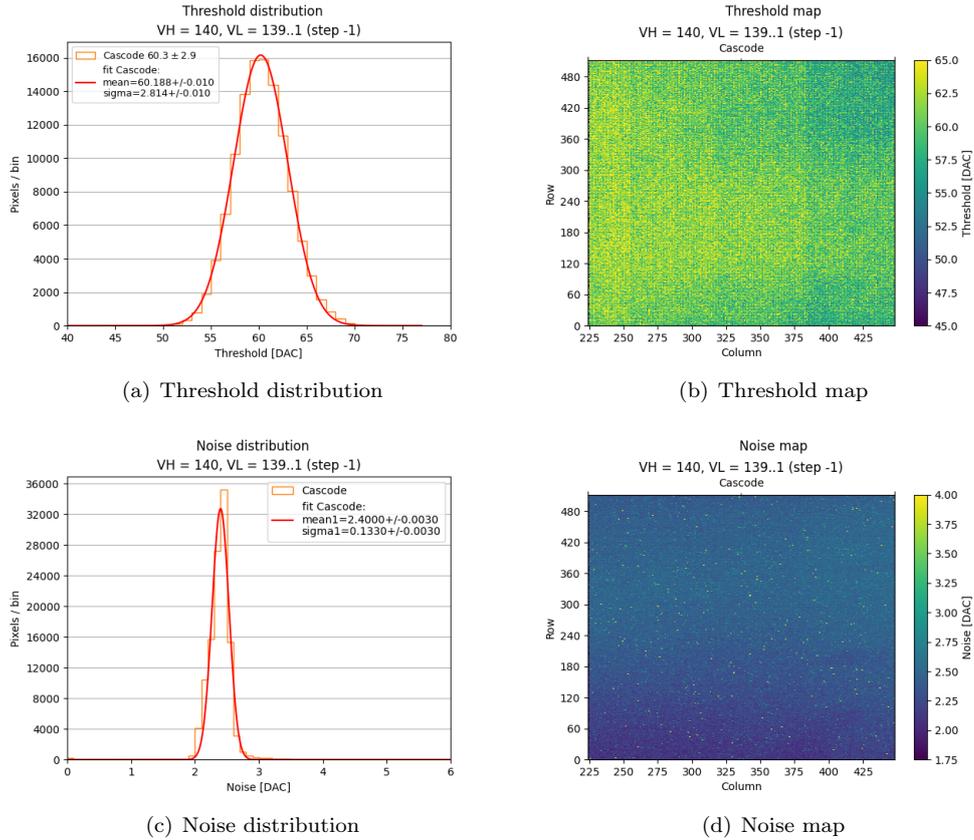


Figure 5.9: Distributions and maps of threshold (first row) and noise (second row) in the entire Cascade flavor.

itself, but by some other input, active during the readout of the matrix, causing "cross-talk". The origin of the hot pixels and cross-talk was carefully investigated later (see section 5.6) .

In Figure 5.11 the fits of the threshold and noise distributions are shown.

HV-Normal FE

The fourth and last flavor is the **HV-Normal FE** which has the same layout of the previous FE. The main registers have been set with the values reported in the second column of Table 5.2. The S-curves of all pixels in this flavor are shown in Figure 5.12 and we can notice some hot unmasked pixels, with occupancy > 1 . Moreover, the last 16 columns were not working and they returned a peak of threshold near the value 0, which is excluded from the threshold distribution plots.

Therefore in this part of the matrix, the real number of pixel studied was 8192, half of the total (visible in the maps in Figure 5.14, where the last 16 columns are dark in color, corresponding to zero threshold and noise values).

In Figure 5.13 the fits of the threshold and noise distributions oh the HV-Normal, and in Figure 5.14 the threshold and noise maps of the whole HV flavors, are shown.

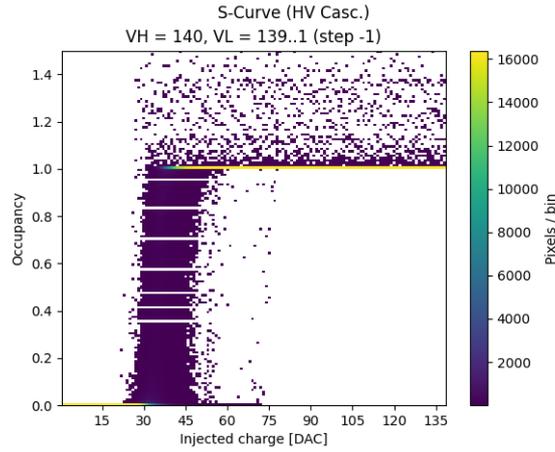


Figure 5.10: S-curves of all pixels in HV Cascode flavor with a maximum injection pulse of 140 DAC.

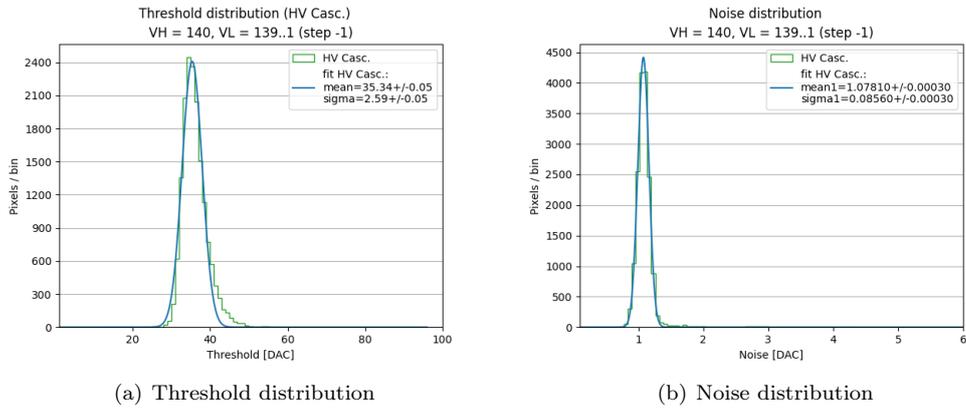


Figure 5.11: Distributions of threshold (a) and noise (b) in the entire HV Cascode flavor.

As we will see in the following (section 5.6), the atypical S-curves with many hot pixels in the HV flavors, have been the first hint of the cross-talk problem, tied to a global lower threshold in these sectors, compared with the first two thresholds measured in the Normal and Cascode sector (with the Test Beam settings).

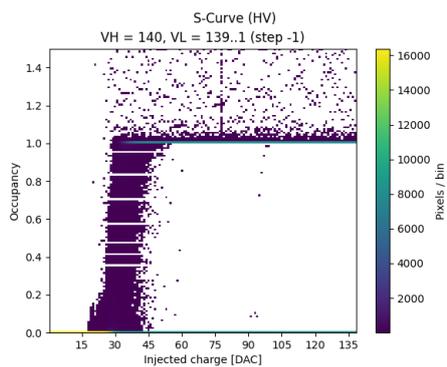
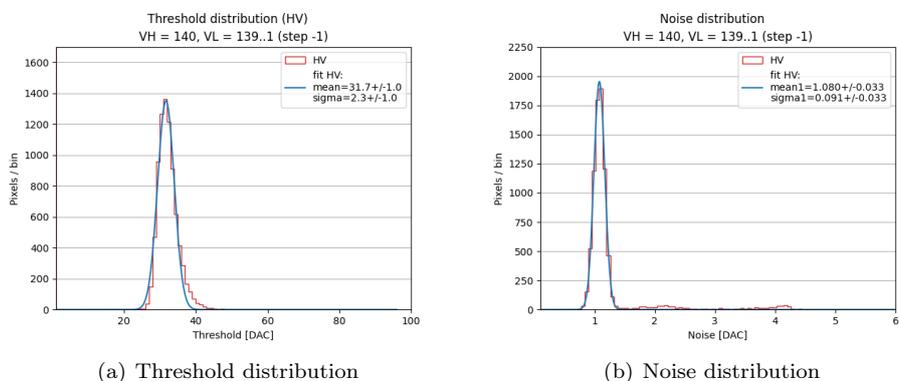


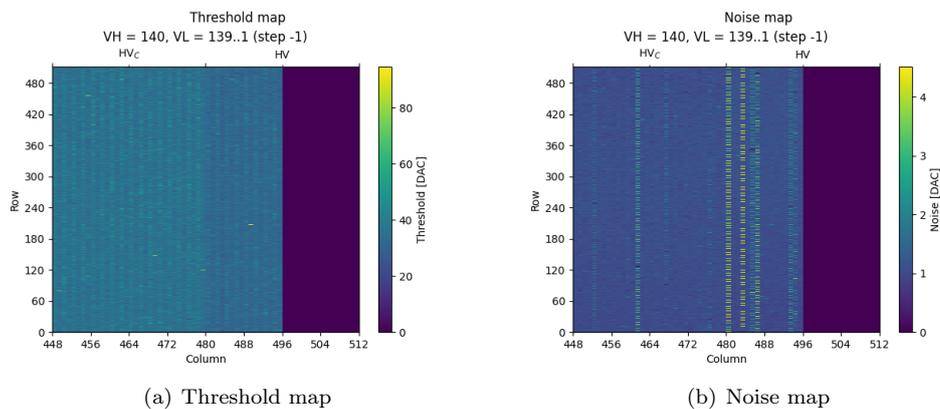
Figure 5.12: S-curves of all pixels in HV Normal FE with a maximum injection pulse of 140 DAC.



(a) Threshold distribution

(b) Noise distribution

Figure 5.13: Distributions of threshold (a) and noise (b) in the entire HV Normal flavor.



(a) Threshold map

(b) Noise map

Figure 5.14: Threshold (a) and noise (b) maps in HV Cascade (left 32 columns) and HV Normal FE (right 32 columns).

5.2.3 Threshold, noise and threshold dispersion results (Test Beam settings)

In Table 5.3 a summary of the results for threshold, noise and threshold dispersion of all the FE is reported, both in DAC and in e^- unit, using the nominal conversion factor K in equation 5.6.

FE	Threshold [DAC]	Threshold [e^-]	Threshold dispersion [DAC]	Threshold dispersion [e^-]	Noise [DAC]	Noise [e^-]
Normal	53.6	541.6	2.2	22.3	2.5 3.0	25.3 30.7
Cascode	60.2	607.9	2.8	28.4	2.4	24.2
HV - Cascode	35.3	356.9	2.6	26.2	1.1	10.9
HV	31.7	320.0	2.3	23.0	1.1	10.9
Target (Table 5.1)		≈ 100		≤ 10		≤ 8

Table 5.3: Summary table of threshold, threshold dispersion and noise values for all flavors of the W14R12 chip. The conversion from DAC to e^- unit has been done using the nominal conversion factor $K = 10.1 e^-/\text{DAC}$.

The measured values are much higher than expected under convenient working conditions (optimized values shown in Table 5.1). However, as we have pointed out in the previous, we carried out the characterization of the matrix response adopting the TB register settings, to obtain results essential for the TB data reconstruction, and not to work in optimized conditions.

In section 5.5 we will show some test results, trying to operate the chip at low threshold.

5.3 TOT calibration with internal injection

The analog information on the signal height is provided by the Time Over Threshold (Figure 5.15) digitized with a 25 ns clock (the Bunch Crossing ID (BCID) clock-frequency is 40 MHz).

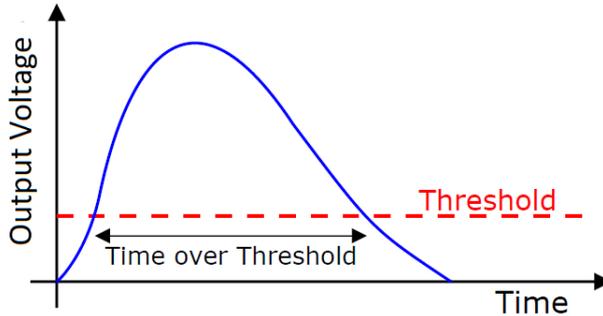


Figure 5.15: Time Over Threshold (TOT) method to encode the voltage analog information.

The choice to use a simple diode (instead of a PMOS transistor) as reset input baseline element, increases the tolerance to Total Ionizing Dose (TID) radiation, but it also implies a

difficult to model relationship between the injected charge and the TOT. For this reason, one of the goal of this analysis was to measure the calibration curve TOT vs Q_{inj} via the internal injection. The fit to the calibration curve allows to find the conversion function needed to reconstruct the signal amplitude, measuring its TOT. It also allows an absolute calibration of the injection capacitance, done comparing the results of the TOT response from the radioactive source with known released signal, with the response to the same signal through the injection circuit.

In carrying out the measurements mentioned above, we started to notice some issues with the injection circuit, which showed some saturation in the voltage pulse at high values of the registers. Due to this issue, the TOT response to internal injection could only be measured up to 170 DAC ($\approx 1700 e^-$). This was enough for the absolute calibration using the ^{55}Fe 5.9 KeV emission line ($\approx 1616 e^-$) as explained below (section 5.4).

A method has been therefore devised to obtain reliable values of threshold and TOT up to a value of 170 DAC of effective charge injected.

The calibration function adopted to describe the Q_{inj} -TOT relationship was then used to extrapolate TOT values in the region of high charge (above 170 DAC, $\approx 1700 e^-$) not accessible with the injection circuit, to compare with the emission peaks of other radioactive sources and explore a larger range (section 5.4.5).

5.3.1 TOT curves and fit

The characterization of the function that reproduces the Q_{inj} -TOT relationship is crucial to convert the TOT information returned by the chip to the actual charge collected. These measurements have also been taken using the Test Beam register settings (Table 5.2), thus allowing the reconstruction of TB data.

Given the shape of the TOT distribution depending on the injected charge shown in Figure 5.16(a), the empirical function chosen to describe the calibration curve is:

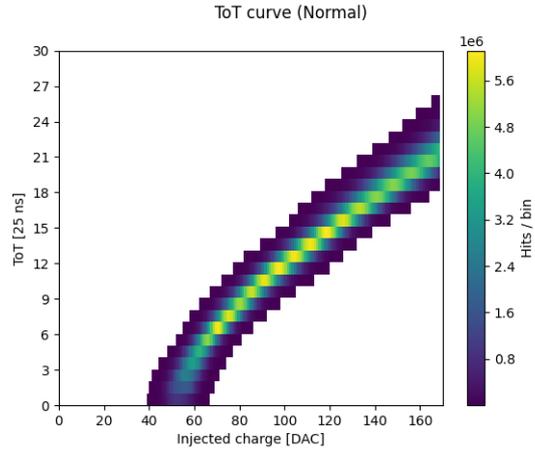
$$y(x) = a \cdot x + b - \frac{c}{x - t} \quad (5.8)$$

with a , b , c and t free parameters and where y represents the TOT corresponding to the value of collected charge, expressed by x .

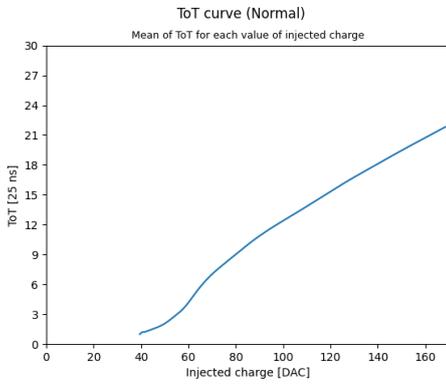
For calibration purposes it would be necessary to fit the TOT distribution of each pixel, but at this stage, we need to characterize the overall response of each FE. Therefore, we have computed the average value of the TOT distribution of all pixels in each flavor. The several steps done to improve the results are explained in the following, considering the TOT curve of the Normal FE as example.

At first the weighted average of the TOT values for each injected charge bin was calculated and the result is displayed in Figure 5.16(b). As we can see, this method does not reproduce the TOT distribution shape at low values very well, since due to the threshold dispersion on the flavour, there are low TOT values corresponding to injected charges lower than the global threshold of the FE, estimated in the previous (Table 5.3). For this reason the weighted average on TOT near the threshold does not go to zero as expected.

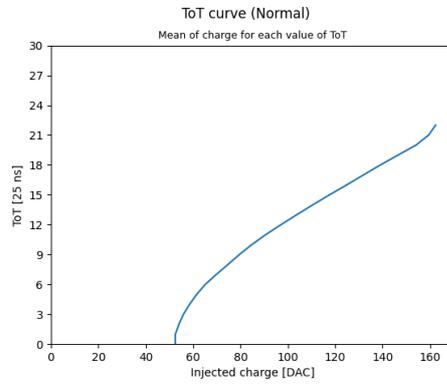
To improve the reproduction of the TOT distribution shape at low values, we made the weighted average of the charge for each TOT bin. The result is reported in Figure 5.16(c), and the shape obtained is much more in agreement with the distribution in Figure 5.16(a). Therefore we tried to fit both dataset with the function in equation 5.8, leaving all parameters free. In Figure 5.16(d) and in Figure 5.16(e) are reported the fit of the weighted average values on TOT



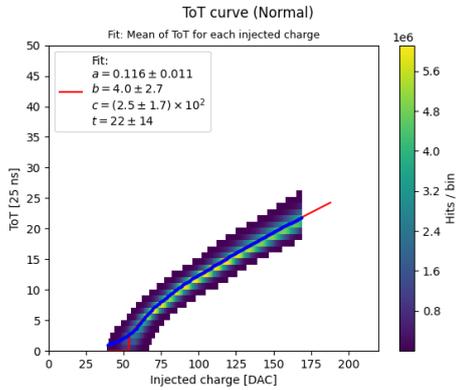
(a) TOT curve



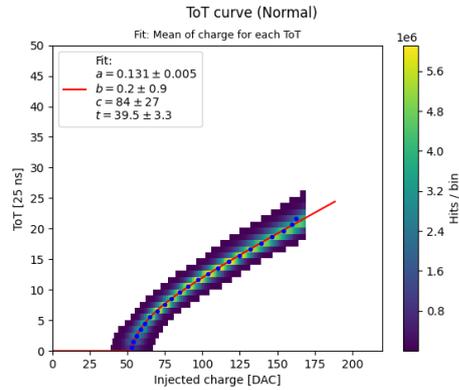
(b) Weighted average on TOT for each charge bin.



(c) Weighted average on charge for each TOT bin.



(d) Fit of the weighted average TOT values for each charge bin, using eq. 5.8, all parameters free.



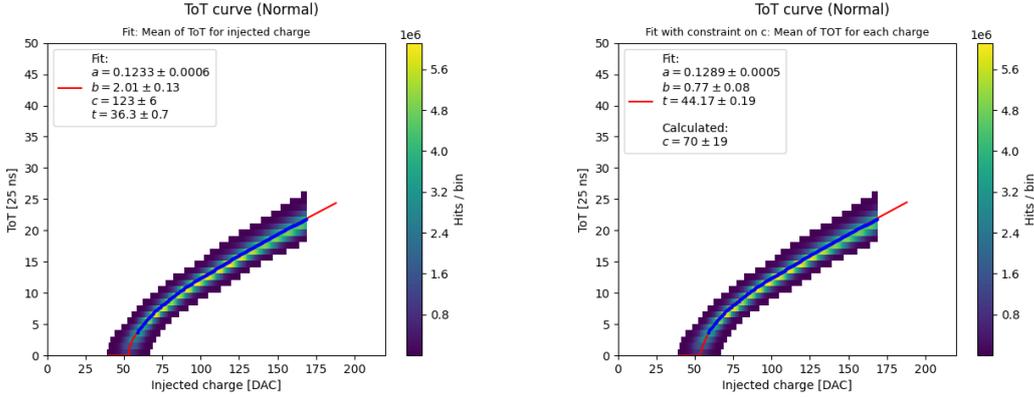
(e) Fit of the weighted average charge values for each TOT bin, using eq. 5.8, all parameters free.

Figure 5.16: Characterization of the calibration function for the Normal FE.

and on charge, respectively. To compare the fit results, we calculated the relative error for each parameter, reported in Table 5.4. The first two columns, corresponding to the results obtained in these two first steps, show that this adjustment has also improved the uncertainties on the parameters, except for b , whose value is of the same order of its uncertainty.

Although the averaged values on the charge returned smaller, but not yet acceptable errors, we tried to repeat the fit on the weighted average TOT values, cutting the low TOT points between 0 and 5 TOT unit (with a BCID clock frequency of 40 MHz, one TOT unit corresponds to 25 ns).

The parameter values obtained are shown in figure 5.17(a), and their relative uncertainties in the third column of Table 5.4. As we can see, fitting the cut data returned smaller uncertainties on the parameters.



(a) Fit of the weighted average TOT values (≤ 5) for each charge bin, using eq. 5.8, all parameters free. (b) Fit of the weighted average TOT values (≤ 5) for each charge bin, using eq. 5.10, with a , b and t left free and c computed using equation 5.9.

Figure 5.17: Characterization of the calibration function for the Normal FE.

To further enhance the characterization of the calibration function, we have exploited the information on the global threshold of the FEs (Table 5.3). The TOT distribution starts to grow near the threshold, then one of the four free parameters could be computed as a function of its value because, from equation 5.8, knowing that $y(x_{th})$ must be equal to 0 (that is the TOT value at the threshold), it can be imposed:

$$0 = a \cdot x_{th} + b - \frac{c}{x_{th} - t} \quad \Rightarrow \quad c = x_{th}^2 \cdot a + x_{th} \cdot (b - a \cdot t) - t \cdot b \quad (5.9)$$

So the function to fit the TOT curve becomes:

$$y(x) = a \cdot x + b - \frac{x_{th}^2 \cdot a + x_{th} \cdot (b - a \cdot t) - t \cdot b}{x - t} \quad (5.10)$$

with a number of parameters to fit reduced from four to three (a , b and t).

In 5.17(b) the results obtained for the Normal FE. The parameter c was calculated using the equation 5.9 and its uncertainty was evaluated using the error propagation formula, including the threshold dispersion as uncertainty of the threshold value.

The relative uncertainties on parameters are reported in the fourth column of Table 5.4. The uncertainties on parameters a and t are reduced, but those on b and c are increased (in case of c , due mostly to the threshold dispersion in its error propagation formula).

	Weighted TOT fit using eq. 5.8 a, b, c, t free	Weighted charge fit using eq. 5.8 a, b, c, t free	Weighted TOT (cutting TOT < 5) fit using eq. 5.8 a, b, c, t free	Weighted TOT (cutting TOT < 5) fit using eq. 5.10 a, b, t free c computed (eq. 5.9)
$\Delta a/a$	9.6%	3.5%	0.5%	0.4%
$\Delta b/b$	67.5%	109.2%	6.3%	10.8%
$\Delta c/c$	67.3%	38.3%	4.6%	27.7%
$\Delta t/t$	60.6%	8.1%	2.0%	0.4%

Table 5.4: Relative uncertainties of parameters obtained from the several fits of TOT curve for the Normal FE.

There is a strong correlation between the four parameters, therefore improving the result on one of them, could worsen the result on another, as we can see from Table 5.4. We have observed the same trend (with different percentages) of the uncertainties also for the other FEs.

5.3.2 Comparison

To evaluate which of the fit results could best reproduce the TOT- Q_{inj} relationship, we inverted the empirical calibration function (equation 5.8) to obtain the value of the charge (x) as a function of the TOT, a , b , c , and t :

$$x(y) = \left(\frac{t}{2} - \frac{b}{2a} + \frac{y}{2a} \right) \pm \sqrt{\left(\frac{t}{2} + \frac{b}{2a} - \frac{y}{2a} \right)^2 + \frac{c}{a}} \quad (5.11)$$

Knowing that for a zero TOT value (y), we should obtain the value of the threshold (x_{th}), we computed the uncertainty on the expression below, using the error propagation formula:

$$x(0) = x_{th} = \left(\frac{t}{2} - \frac{b}{2a} \right) \pm \sqrt{\left(\frac{t}{2} + \frac{b}{2a} \right)^2 + \frac{c}{a}} \quad (5.12)$$

Between the two possible solutions we took the one with the "+", which returns the higher value and represents our threshold. The other solution instead represents nothing of our data, so it is discarded.

Then we compared this result with the threshold dispersion extracted from the threshold distribution fit (Table 5.3). Table 5.5 shows the comparison: in the first row the threshold value and its dispersion for each FE, in the second row the same quantities estimated as described in the previous, using the parameter values (a , b , c and t) obtained by the fit method summarized in the third column of Table 5.4, while in the third row using the fit method summarized in the fourth column of the same table.

The agreement with data in the first row, is obtained with parameter values extracted from the fourth fit method, even though the parameter uncertainties (Table 5.4) are higher compared to the parameter results obtained without condition on c . It is worth to mention that in the

	Normal	Cascode	HV Cascode	HV
$x_{th} \pm \Delta x_{th}$ [DAC] from the threshold distribution fit	53.6±2.2	60.2±2.8	35.3±2.6	31.7±2.3
$x_{th} \pm \Delta x_{th}$ [DAC] from $x_{th}(a,b,c,t)$ in eq. 5.12 using a,b,c,t from 3° column of Table 5.4	51.1±0.2	64.2±0.4	35.1±2.7	31.5±2.6
$x_{th} \pm \Delta x_{th}$ [DAC] from $x_{th}(a,b,c,t)$ in eq. 5.12 using a,b,c,t from 4° column of Table 5.4	53.4±2.2	60.8±2.7	35.1±2.7	31.5±2.4

Table 5.5: In the first row the the threshold value and its dispersion for each FE (Table 5.3); in the second row the same quantities estimated by eq. 5.12 and their uncertainties, using parameter values obtained from the third fit method (third column in Table 5.4); in the third row same procedure but using parameter values from the fourth fit method (fourth column in Table 5.4).

third fit method, parameters are estimated cutting data near the threshold, which is what we want to return with this procedure, therefore we do not expected the results to be in agreement.

Finally, we extracted the expected TOT value corresponding to the charge released from the 5.9 KeV emission line of ^{55}Fe source, to compare it with the peak measured during the acquisitions with the source, shown in Figure 5.19. We computed the TOT value using equation 5.8 for parameter values obtained without condition on c parameter, and equation 5.10 for parameters obtained with the last fit method. The results are reported in Table 5.6, together with the measured peak values. To calculate the expected charge in DAC corresponding to the charge released from the source ($\approx 1616 e^-$) we used the nominal conversion factor $K=10.1 e^-/\text{DAC}$ (equation 5.6), as following:

$$Q_{exp}(\text{DAC}) = \frac{N_{e^-}}{K} = \frac{1616 e^-}{10.1 e^-/\text{DAC}} = 160 \text{ DAC} \quad (5.13)$$

	Normal	Cascode	HV Cascode	HV
TOT(160 DAC)±ΔTOT [TOT unit] from source acquisition	23.4±1.4	21.8±1.3	35.3±2.6	31.7±2.3
TOT(160 DAC)±ΔTOT [TOT unit] from $y(x,a,b,c,t)$ in eq. 5.8 using a,b,c,t from 3° column of Table 5.4	20.8±1.4	19.4±1.3	35.1±2.7	31.5±2.6
TOT(160 DAC)±ΔTOT [TOT unit] from $y(x,a,b,c,t)$ in eq. 5.10 using a,b,t from 4° column of Table 5.4	20.8±2.7	19.4±3.1	35.1±2.7	31.5±2.4

Table 5.6: In the first row the the source peak value of ^{55}Fe and its dispersion for each FE (Figure 5.19); in the second row the same quantities estimated by equation 5.8 and their uncertainties, using parameter values obtained from the third fit method (third column in Table 5.4); in the third row same procedure but using parameter values from the fourth fit method (fourth column in Table 5.4).

Although the results obtained from the fourth fit method return threshold and dispersion values more compatible with those measured via internal charge injection (Table 5.5), the estimate of the TOT value corresponding to the iron emission peak and its dispersion, seem to be slightly more similar with the measured ones, using the parameters obtained from the third fit method. Moreover in this case, the relative uncertainties of the parameters are smaller. So in the following, we decided to fit the TOT curve with equation 5.8 leaving all parameters free. In Figure 5.18 the fit results for all FEs are shown.

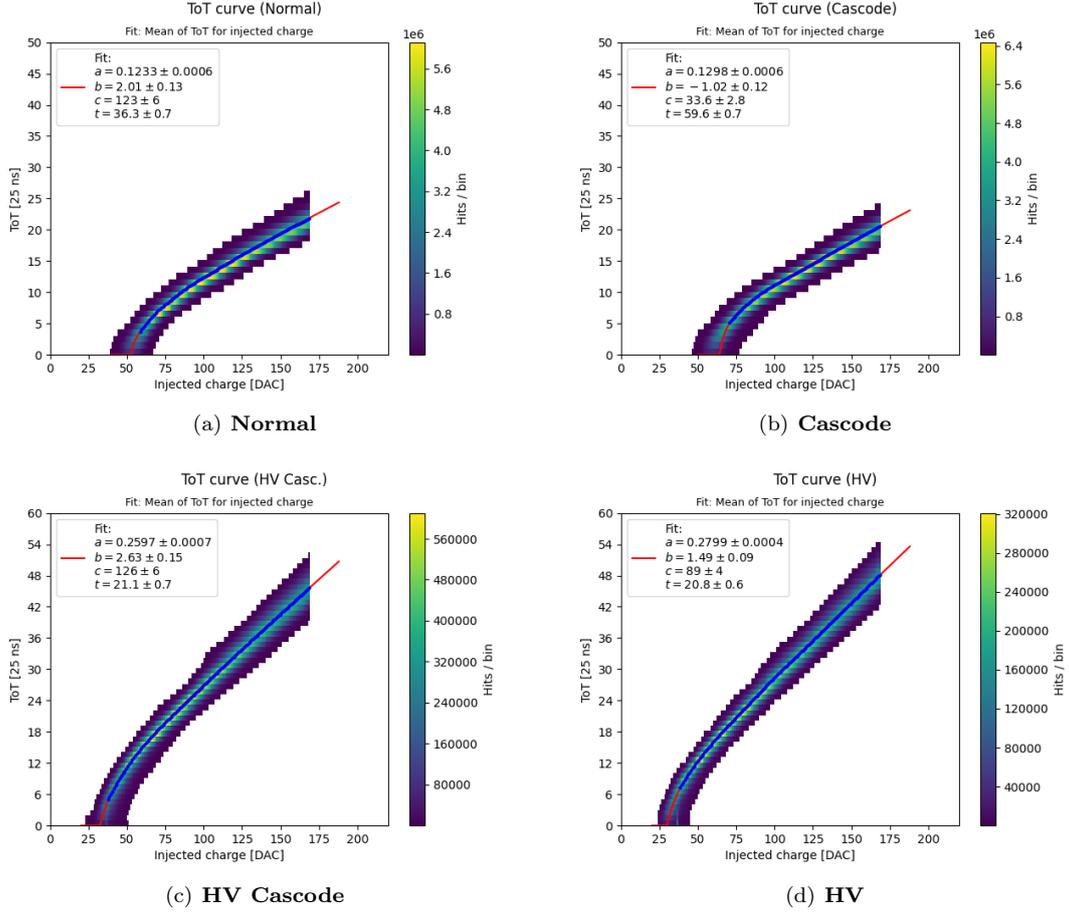


Figure 5.18: Fit of the weighted average TOT values (≤ 5) for each charge bin, using equation 5.8 to extract the values of a , b , c and t . The results for all FEs are shown.

5.4 Response to radioactive source and absolute calibration

After the characterization by the internal injection, we measured the response of the matrix to three different X-rays radioactive sources: ^{55}Fe , ^{241}Am , ^{109}Cd , with emission lines from 6 to 60 KeV, corresponding approximately to a range between 1600 and 16000 e^- signal charge released in the silicon, as shown in Table 5.7.

In this way we could perform the absolute calibration of the matrix response and extend the study of the TOT spectrum above the limit imposed by the saturation of the internal injection circuit ($\approx 1700 e^-$). This is also useful to verify the goodness of the calibration curve characterization at higher signal, that could be deposited by a MIP due to the long tail of the Landau distribution, which describes the fluctuations of the energy released by a charged particle traversing a medium.

The absolute calibration of the matrix consists in characterizing the response of each pixel to a known signal, like the emission peaks of radioactive sources, and then comparing the results with the response from the internal injection to the same amount of signal, as explained in section 5.4.4.

The absolute calibration of the conversion factor of equation 5.6, equivalent to measuring the injection capacitance C_{inj} , is performed with the 5.9 KeV peak of the ^{55}Fe source ($\approx 1616 e^-$), that is still in the range explored with the injection circuit. The other radioactive sources, with high energies, allowed to extend the TOT comparison at higher values with respect to the limit imposed by the saturation of the internal injection circuit. In Table 5.7 the energies of the γ emitted by the sources used are shown.

Considering that the average energy necessary to produce an electron/hole pair in silicon is 3.65 eV, the peak energies expressed in KeV could be converted in a mean value of electrons released, using the equation 5.14. The equivalent emission lines in electrons unit, which will be useful later, are reported in the second column of Table 5.7

$$N_{e^-} = \frac{E [eV]}{3.65[\frac{eV}{e/h \text{ pair}}]} \quad (5.14)$$

Source	Energy γ [KeV]	Equivalent charge [e^-]
^{55}Fe	5.9	1616
^{241}Am	13.9	3808
^{241}Am	17.7	4849
^{241}Am	20.7	5671
^{109}Cd	22	6027
^{241}Am	26.4	7233
^{241}Am	59.7	16356

Table 5.7: Emission lines of ^{55}Fe , ^{241}Am , ^{109}Cd sources.

Now we can go through the results obtained from three different sources: ^{55}Fe , ^{241}Am and ^{109}Cd .

5.4.1 ^{55}Fe

The ^{55}Fe source decays to ^{55}Mn by electron capture. One of the photons emitted in this transition has an energy of 5.9 KeV (K_α) and it produces an electron via photoelectric effect, which in turn deposits a ionization charge of about 1616 e^- in the sensor. All four flavors of the matrix were exposed to a ^{55}Fe source, with an activity of 18 MBq.

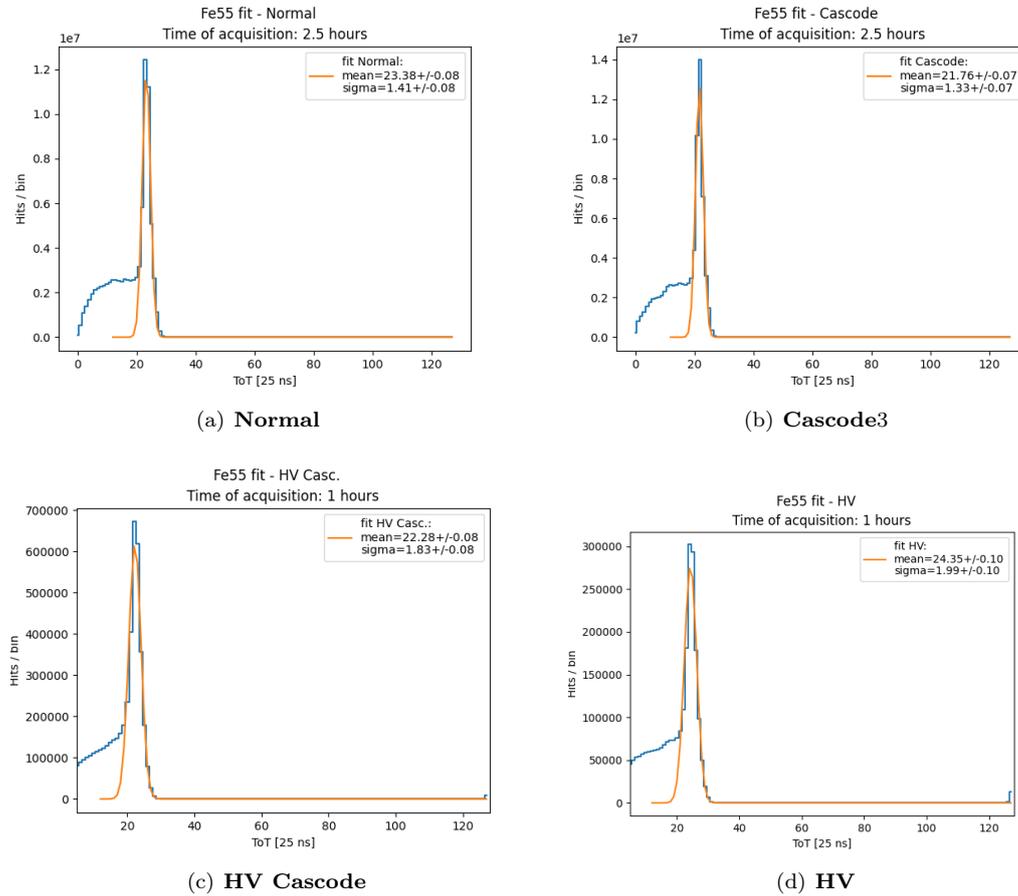


Figure 5.19: ^{55}Fe single pixels TOT spectrum for all frontends.

The TOT spectra of single pixels in each of the four flavors are reported in Figure 5.19. The peak corresponds to events where the γ interacts close to the collection diode and the entire signal is on a single pixel. The shoulder at smaller TOT is due to the charge only partially collected on a single pixel, since charge is shared among several pixels and no cluster reconstruction is done in this analysis. The TOT peak was fitted by a gaussian function, limited in the region of the peak itself, and the results of the fit are reported in the box of each plot. As it can be seen, for both the HV frontends a cut has been applied at low TOT, only to make clearly visible the emission line, since a lot of noisy pixels caused a sharp peak at zero TOT value. In those flavors there were several columns of not-functioning pixels.

The TOT peak value can be converted to signal in DAC unit, using the fitted calibration curve TOT vs $Q_{inj}[\text{DAC}]$ of equation 5.8, and then converted to signal in electrons, assuming

the nominal conversion factor K of $10.1 e^-/\text{DAC}$ of equation 5.6. In this way we could see a reasonable agreement between the expected signal $Q_{5.9 \text{ KeV}} = 1616 e^-$ and the measured one for the Normal FE ($1610 e^-$) and Cascode FE ($1611 e^-$). On the contrary, we observed a significant loss in the collected signal for the HV flavours as expected, due to the AC coupling and the effect of the parasitic capacitance at the sensitive input node. The signal collected was of about $874 e^-$ in the HV normal FE, and $843 e^-$ in the HV Cascode, corresponding to a loss of about 46% and 48%, thus improving from previous version of the chip, where higher loss ($> 50\%$) were found.

This signal loss is not obvious from the TOT spectra, since in the HV FEs the measured threshold was lower (Table 5.3) with respect to the Normal and Cascode FE. So even if the collected charge is smaller (approximately half), and the TOT signal is smaller than expected from the calibration curve (Figure 5.18), the lower threshold does not allow to make the difference evident.

5.4.2 ^{241}Am

The ^{241}Am source has a more complex spectrum (Figure 5.20) and not all of its peaks can be detected by the chip (because of the limited TOT range available). The spectrum shows other minor peaks besides the usual intense gamma peaks (59.5 and 26.3 KeV) and several characteristic L X-rays from ^{237}Np (20.7, 17.7 and 13.9 KeV) [37].

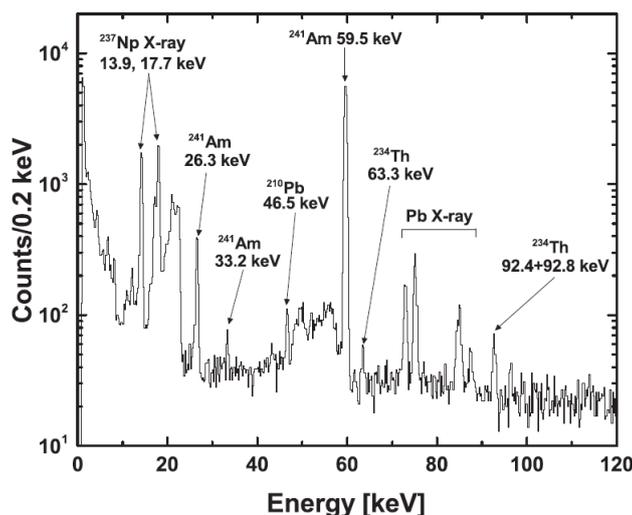


Figure 5.20: ^{241}Am γ emission spectrum. From [37].

The measured TOT spectra obtained with the ^{241}Am source emitting on the four flavours are reported in Figure 5.21. Two peaks at lower energy are clearly visible while, for higher TOT values there are larger structures. They have been fitted by a distribution given by the sum of an exponential (to characterize the background) and four gaussians for the four visible peaks of the source. Fit results are reported in the box for each plot and the peak mean values for Normal and Cascode FE, together with the peaks obtain with the other sources, are shown in Table 5.9 and Table 5.10, respectively.

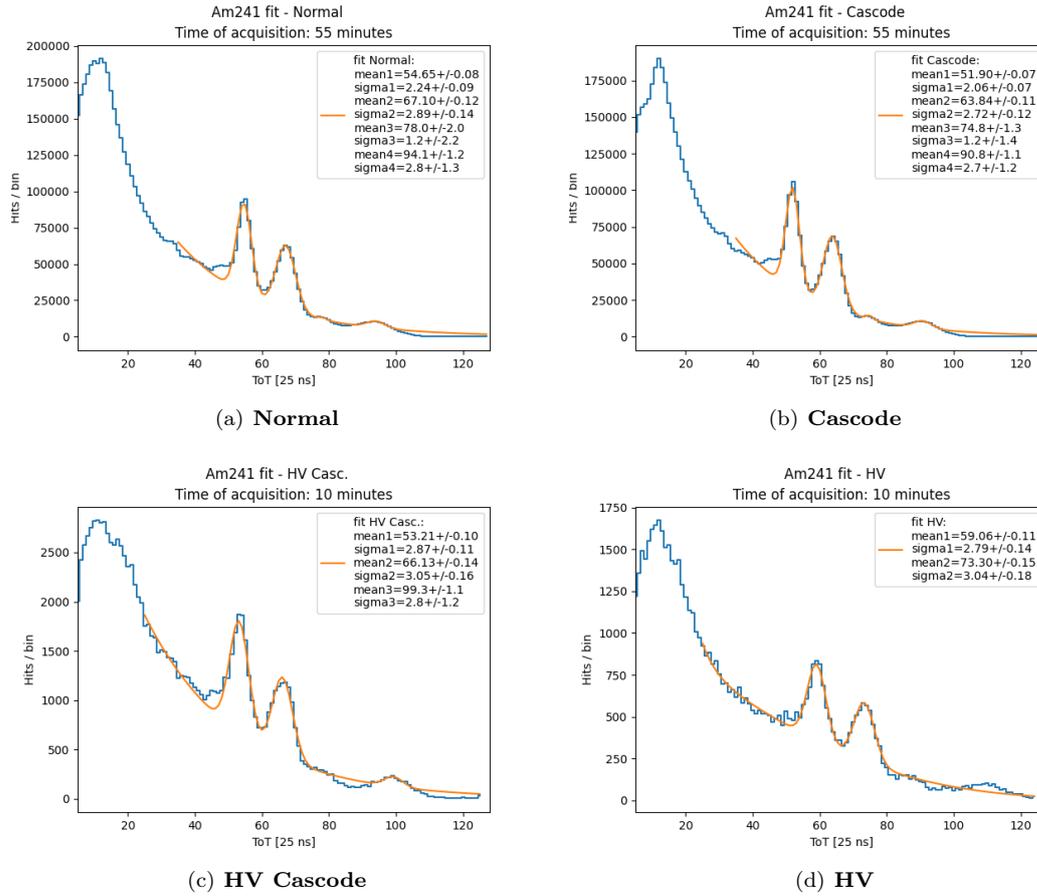


Figure 5.21: ^{241}Am single pixels TOT spectrum for all frontends.

In the case of the first two flavors, it was possible to fit four peaks of the emission lines. In case of the HV flavors instead, only three peaks for the HV-Cascode FE and two for the HV. It should be noted that in the HV flavors there is a significant (nominal 41.5%) charge loss as discussed in section 5.1.2. Nonetheless, since the threshold for the HV flavors is much smaller than the one for the Normal flavor (Table 5.3), the peaks appear at roughly the same TOT value.

5.4.3 ^{109}Cd

The third source employed was the ^{109}Cd . This isotope decays in ^{109}Ag by electronic capture, producing a photon of 22 KeV in the transition. The TOT measured spectra with ^{109}Cd source on the four flavors of the matrix are reported in Figure 5.22. They have been fitted by summing an exponential distribution to model the background, and a gaussian for the source peak. Fit results are shown in the box of each plot and the value obtained for Normal and Cascode FE are reported in Table 5.9 and Table 5.10, respectively.

As visible from the plots, even in these acquisitions, despite the loss of charge in the HVs, the peaks are measured roughly at the same TOT value due to the lower threshold in these flavors.

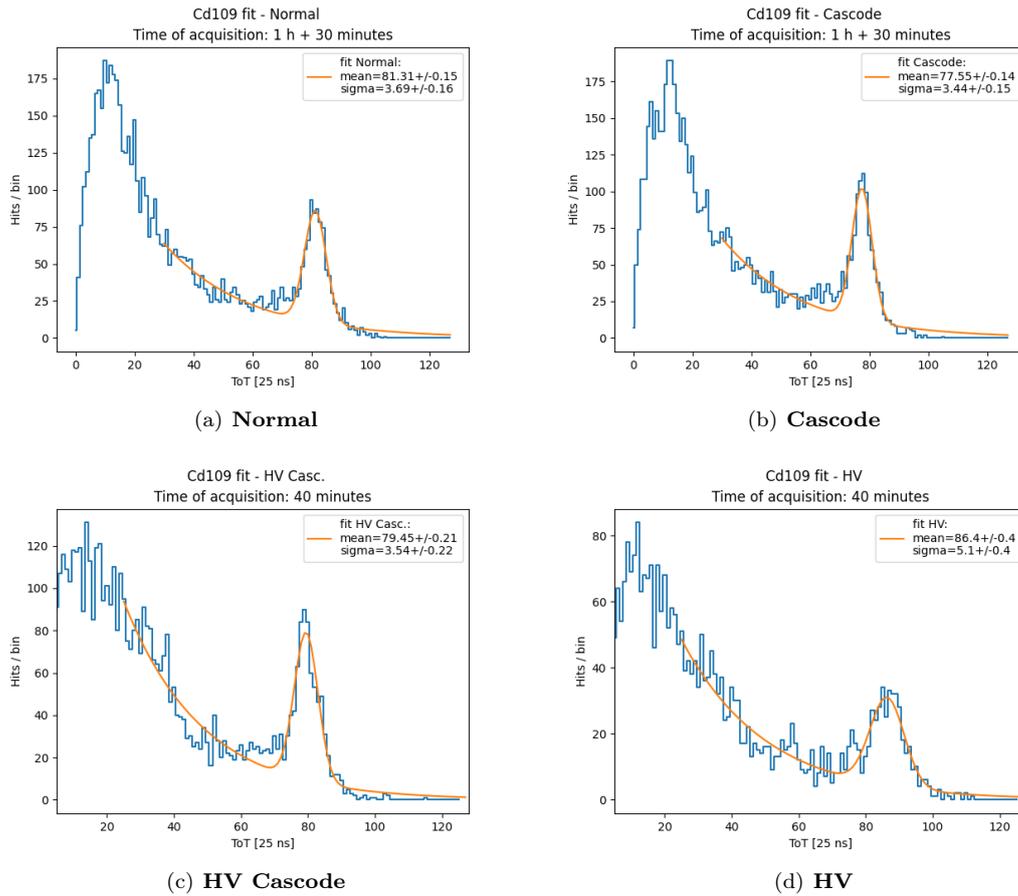


Figure 5.22: ^{109}Cd single pixels TOT spectrum for all frontends.

5.4.4 Conversion factor K and C_{inj} absolute calibration

The absolute calibration of the conversion factor K , that converts a signal charge expressed in DAC units to e^- , is performed using the data from the ^{55}Fe source. As shown in section 5.2.1, the conversion factor K corresponds to the injected charge given by a voltage step of 1 DAC ($\Delta V_{LSB} = 7.03 \text{ mV/DAC}$) through the injection capacitance C_{inj} implemented in each pixel.

$$K\left(\frac{e^-}{\text{DAC}}\right) = \frac{C_{inj}}{q_{e^-}} \cdot \Delta V_{LSB} \quad (5.15)$$

For the design value of $C_{inj} = 230 \text{ aF}$ we expect $K = 10.1 e^-/\text{DAC}$, but an absolute measurement is needed since there could be significant variations from chip to chip and also across the matrix.

As first step the TOT value of the peak of the 5.9 KeV γ line for each pixel is converted to signal charge in DAC, $Q_{5.9 \text{ KeV}}[\text{DAC}]$, using the fitted calibration curve of equation 5.8.

Specifically the fit function was inverted obtaining:

$$x(y) = \left(\frac{t}{2} - \frac{b}{2a} + \frac{y}{2a}\right) \pm \sqrt{\left(\frac{t}{2} + \frac{b}{2a} - \frac{y}{2a}\right)^2 + \frac{c}{a}} \quad (5.16)$$

where x represents the charge in DAC corresponding to the TOT labeled by y . We have only considered the function with the "+" to model the data, the other has been discarded.

As shown in Table 5.7, the charge released in the sensor by the 5.9 KeV γ corresponds roughly to $Q_{5.9 \text{ KeV}}[e^-] = 1616 e^-$. Therefore the conversion factor K for each pixel can be calculated as:

$$K\left(\frac{e^-}{\text{DAC}}\right) = \frac{1616 e^-}{Q_{5.9 \text{ KeV}}(\text{DAC})} \quad (5.17)$$

By these steps, a value of the conversion factor K was estimated for each pixel in the Normal and Cascode FE. The map of the measured values together with their distribution in the two flavors is shown in Figure 5.23, with the results of the gaussian fit. In Table 5.8 the average conversion factors found in the two flavors are compared with the design values, showing a good agreement (10%).

It is important to stress that due to the signal loss in the HV flavours, the amount of charge collected from 5.9 KeV γ is not really known thus it is not possible to perform the absolute calibration of the conversion factor for the pixels in these flavours.

Source peak	$K_{design} \left(\frac{e^-}{\text{DAC}}\right)$	$K_{Normal} \left(\frac{e^-}{\text{DAC}}\right)$	$K_{Cascode} \left(\frac{e^-}{\text{DAC}}\right)$
^{55}Fe (5.9 KeV)	10.10	8.93	9.07

Table 5.8: Average conversion factor K for the Normal and Cascode FE using the ^{55}Fe radioactive source emission line at 5.9 KeV.

5.4.5 Check on calibration curve TOT vs Q_{inj} with radioactive sources

The calibration of the TOT vs Q_{inj} (DAC) response with the injection circuit could be performed only up to ≈ 170 DAC (about $1700 e^-$). The radioactive sources can be used up to higher energy to check the agreement of the response with the fitted function in equation 5.8.

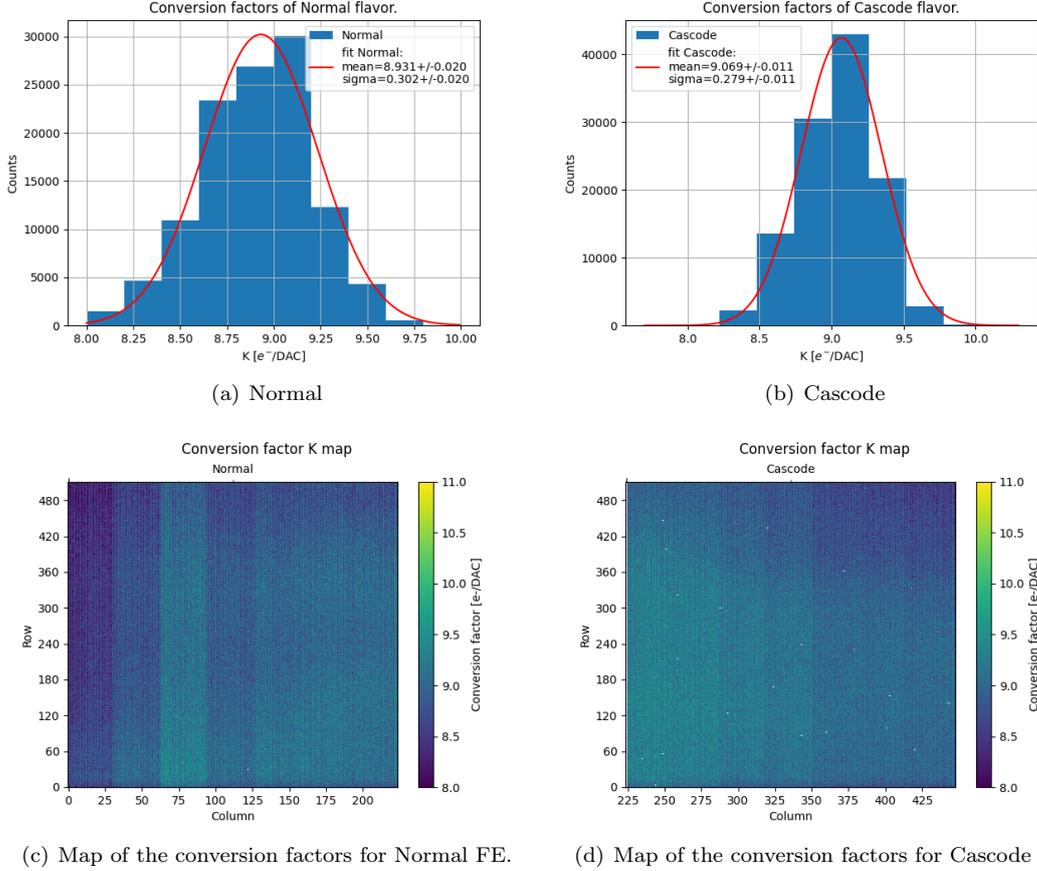


Figure 5.23: In the first row, the distribution of the conversion factor K measured for all pixels in the Normal and Cascode FE. In the second row the maps of the conversion factor values in Normal and Cascode sectors.

This comparison is shown in Figure 5.24 and Figure 5.25. In both plots the TOT vs $Q_{inj}(\text{DAC})$ fitted curve is displayed superimposing the points corresponding to each γ line from the different sources. For each γ line the point shown in the plot has the TOT measured from the source spectrum and the charge in DAC corresponding to the charge released by that photon. In Table 5.9 and in Table 5.10 the various inputs for Normal and Cascode sectors are shown. The signal charge in DAC, corresponding to the number of e^- of each γ line, has been calculated using either the nominal conversion factor equal to $10.1 \frac{e^-}{\text{DAC}}$ or the average value of the conversion factor K measured with ^{55}Fe absolute calibration (Table 5.8).

In Figure 5.24, we can see a reasonable agreement between data and TOT fitted function, with the nominal conversion factor, and in case of the Normal FE, a slightly better agreement using the absolute calibration of the conversion factor specific for this chip. Deviations from the fitted TOT response are visible, especially in the Cascode FE above 600 DAC ($\approx 6000 e^-$), and this effect could be ascribed to a deviation from the linear response of the TOT to high signal, more likely to happen in the Cascode flavor that has higher gain.

This preliminary study showed that through the calibration of the TOT with internal injection

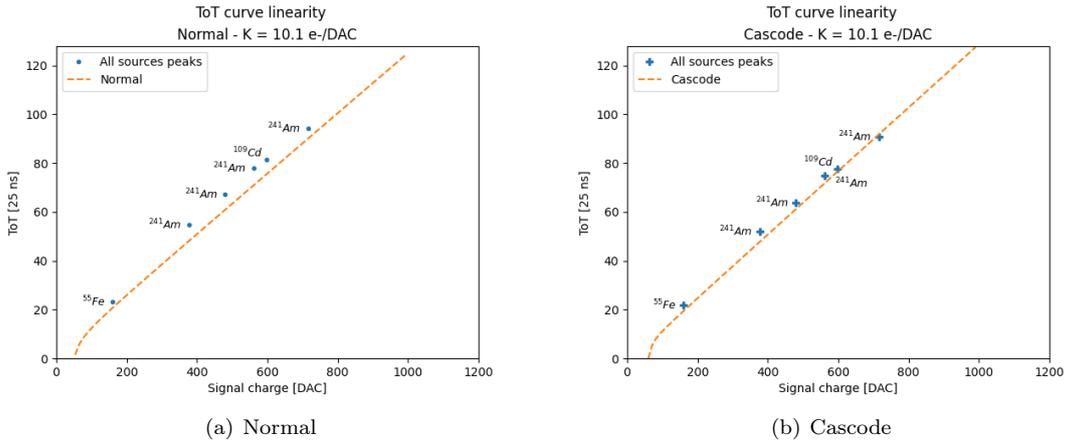


Figure 5.24: TOT vs Q calibration curve measured with the internal injection circuit compared with response from several γ lines from radioactive sources, assuming the nominal conversion factor equal to $10.1 e^-/\text{DAC}$.

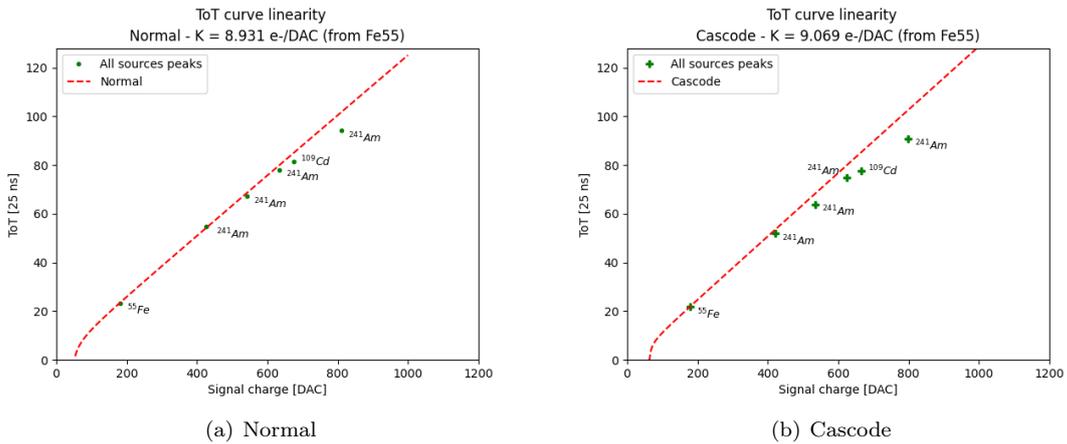


Figure 5.25: TOT vs Q calibration curve measured with the internal injection circuit compared with response from several γ lines from radioactive sources, assuming the conversion factor obtained from ^{55}Fe .

circuit, and the absolute calibration, we are now able to interpret data from the MIP, collected during the TB, with a reasonable accuracy to perform the reconstruction, which was the main purpose of this analysis.

Source	Energy γ [KeV]	$Q_{expected}$ [e^-]	$ToT_{measured}$	$Q_{measured}$ [DAC]	Q [DAC](e^-) with nominal K factor = 10.1 e^- /DAC	Q [DAC](e^-) with ^{55}Fe K factor = 8.93 e^- /DAC
^{55}Fe	5.9	1616	23.4	180	160 (1820)	181 (1610)
^{241}Am	13.9	3808	54.7	429	377 (4338)	426 (3836)
^{241}Am	17.7	4849	67.1	530	480 (5352)	543 (4733)
^{241}Am	20.7	5671	78.0	618	561 (6242)	635 (5520)
^{109}Cd	22	6027	81.3	645	597 (6512)	675 (5759)
^{241}Am	26.4	7233	94.1	748	716 (7558)	810 (6683)

Table 5.9: Emission lines of ^{55}Fe , ^{241}Am , ^{109}Cd sources for Normal frontend.

Source	Energy γ [KeV]	$Q_{expected}$ [e^-]	$ToT_{measured}$	$Q_{measured}$ [DAC]	Q [DAC](e^-) with nominal K factor = 10.1 e^- /DAC	Q [DAC](e^-) with ^{55}Fe K factor = 9.07 e^- /DAC
^{55}Fe	5.9	1616	21.8	178	160 (1795)	178 (1611)
^{241}Am	13.9	3808	51.9	408	377 (4125)	420 (3704)
^{241}Am	17.7	4849	63.8	500	480 (5053)	535 (4537)
^{241}Am	20.7	5671	74.8	585	561 (5905)	625 (5302)
^{109}Cd	22	6027	77.6	606	597 (6118)	665 (5494)
^{241}Am	26.4	7233	90.8	708	716 (7149)	798 (6419)

Table 5.10: Emission lines of ^{55}Fe , ^{241}Am , ^{109}Cd sources for Cascode frontend.

5.5 Operation with low threshold

Many experimental environments are exposed to high doses of radiations, thus one important target in sensor design is to keep the detection efficiency high even after radiation damage, that could cause a reduction of the collected signal due to trapping.

For this reason, many tests were performed to understand the chip behaviour at low threshold, that is needed to preserve a good efficiency even with a reduced signal, caused by charge sharing or charge trapping, especially in case of thin epitaxial material.

5.5.1 Registers optimization

As we have seen in section 5.1.2, there are many global registers that control the operating point of the front-end circuit and the discriminator threshold as well as the readout sequence. For this reason, it was necessary to first explore their possible settings in order to operate the chip at low threshold.

Now we will go through the main registers used for this purpose, to explain their functionality, only focusing on those that affect the threshold and its dispersion.

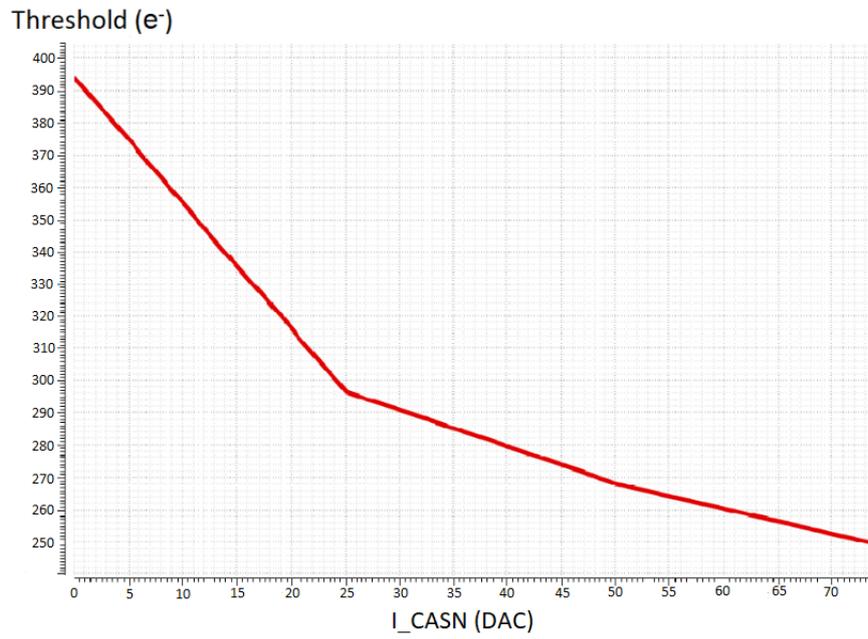
- I_{CASN} : this current is responsible of the output baseline signal. In particular it sets the baseline of the output signal that goes to the discriminator input. For a given value of the discriminator threshold, set with a different register, the higher I_{CASN} and the baseline, the lower is the effective threshold.
- I_{THR} : it controls the pre-amplifier feedback strength and speed, so it is responsible for the output reset rate. With lower I_{THR} the gain increases, so the threshold decreases, according to equation 5.3. Lowering I_{THR} also increases the time that the analog output takes to get back to the baseline. As a consequence with lower I_{THR} also the TOT response changes, increasing a lot the maximum value of the TOT. In fact it is recommended to set I_{THR} to relatively high value (e.g. 8 nA [33]) in order to avoid high TOT slope and saturation, which limits the TOT resolution.
- I_{DB} : this current represents the primary current that sets the discriminator threshold voltage. In threshold tuning (see section 5.5.3) another current is added to this, to fine tune the pixel threshold.
- I_{TUNE} : referring to the tuning equation (equation 5.18, in section 5.5.3), this is the current to multiply by the TCODE value (that is the decimal representation of the TDAC value), which is added to I_{DB} , during the tuning process.
- I_{BIAS} : this current acts on the pre-amplifier input transistor and influences the gain, thus the threshold and its dispersion. In particular increasing this value, the gain increases and the threshold and its dispersion decrease. Nevertheless it cannot be increased too much since it affects the power consumption, too.

5.5.2 Comparison between data and simulation

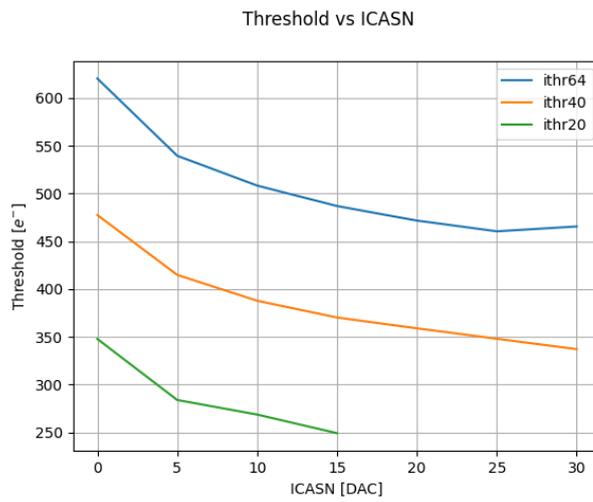
To understand how the registers' setting of the chip influences the threshold, several measurements have been taken with different configuration values. The results are compared with some simulations done by another researcher (Hung Pham, IPHC in Strasbourg), who participates in the study and characterization of TJ-Monopix2 chip, for the VTX upgrade program.

I_{CASN}

In Figure 5.26(a) the simulated behaviour of the threshold depending on the value of I_{CASN} is shown.



(a) Simulation of threshold values varying I_{CASN} .



(b) Threshold vs. I_{CASN} for $I_{THR} = 20, 40, 64$ DAC.

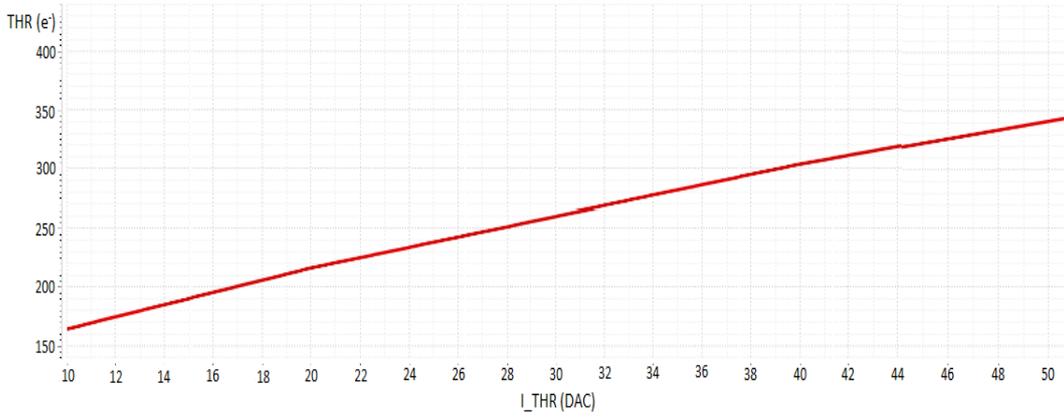
Figure 5.26: Simulation and experimental data of threshold vs. I_{CASN} , fixing I_{THR} .

To verify the trend of the threshold as I_{CASN} varies, three different acquisitions have been taken fixing $I_{THR} = 20, 40, 64$ DAC respectively, and increasing I_{CASN} from 0 to 30 DAC, with

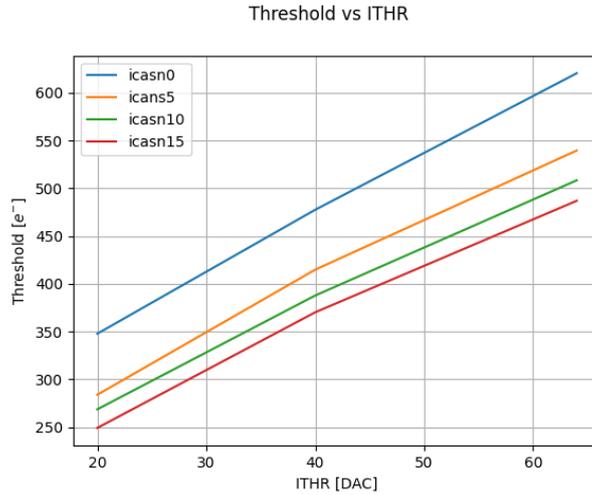
a step of 5 DAC. We have done these measurements enabling about 200 pixels in the Cascode FE. The mean thresholds of the enabled pixels have been measured by the S-curve method for each registers' setting, so the threshold distributions have been fitted with a gaussian function to estimate the average threshold value and its dispersion. Then these values have been converted in electrons unit, using the nominal conversion factor $K = 10.1 e^- / \text{DAC}$ (equation 5.6).

The results obtained are reported in Figure 5.26(b) and they seem to follow the trend of the simulation, despite the few experimental points measured. As a matter of fact, also during these measurements we noticed something strange. Lowering I_{THR} , a lot of hot pixels started to fire, not allowing a reasonable measurement. As explained in the previous (section 5.5.1), decreasing I_{THR} , also the threshold decreases, making the pixels hot, due to the aforementioned cross-talk issue, addressed in section 5.6

I_{THR}



(a) Simulation of threshold values increasing I_{THR} .



(b) Threshold vs. I_{THR} for $I_{CASN} = 0, 5, 10, 15$ DAC.

Figure 5.27: Simulation and experimental data of threshold vs. I_{THR} , fixing I_{CASN} .

Reusing the same data of the previous measurements, the trend of the threshold have been studied, varying the value of I_{THR} and fixing I_{CASN} . In this case, only I_{CASN} from 0 to 15 DAC is considered, because for higher values we could not manage to take enough threshold measurements for the cross-talk issue mentioned above. In fact increasing I_{CASN} , the threshold decreases, making some pixels hot, and invalidating the measurements.

As expected, increasing I_{THR} results to lower gain so higher threshold, and faster return to baseline. We can compare the measured trend, shown in Figure 5.27(b), with the expected one from the simulation shown in Figure 5.27(a), noting a qualitative agreement between them.

Time Over Threshold (TOT)

The last analysis done to make a comparison with the simulations, is about the trend of the TOT changing the value of I_{CASN} for a fixed value of I_{THR} and vice versa. In particular we consider the data obtained with I_{CASN} fixed to 0 DAC and I_{THR} to 64 DAC, which are the values studied and used for these registers during the Test Beam in Desy.

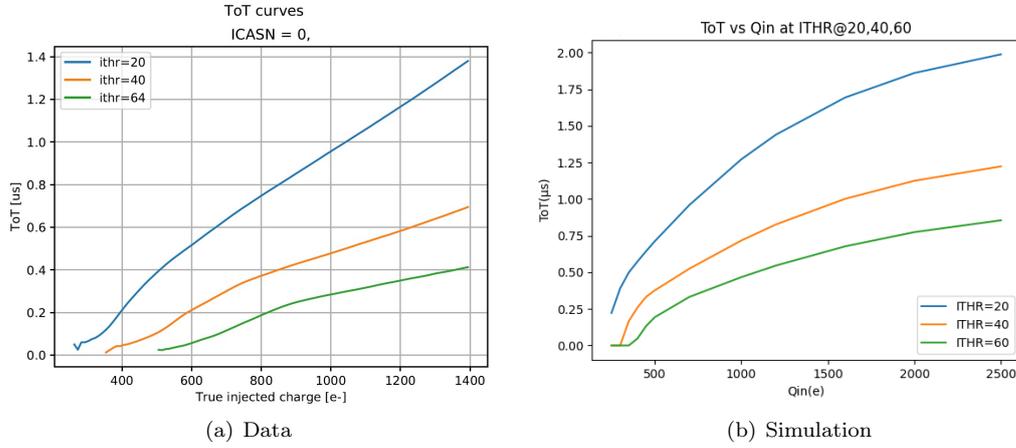


Figure 5.28: TOT vs Q_{inj} varying I_{THR} , with $I_{CASN} = 0$ DAC for Cascode FE.

The shapes of the measured TOT curves (Figure 5.28(a) and Figure 5.29(a)) are a little different than those obtained in the simulations (Figure 5.28(b) and Figure 5.29(b)), in particular at high injected charge. The range of charge explored was limited by the saturation of the charge injection circuit, previously discussed (section 5.2.1).

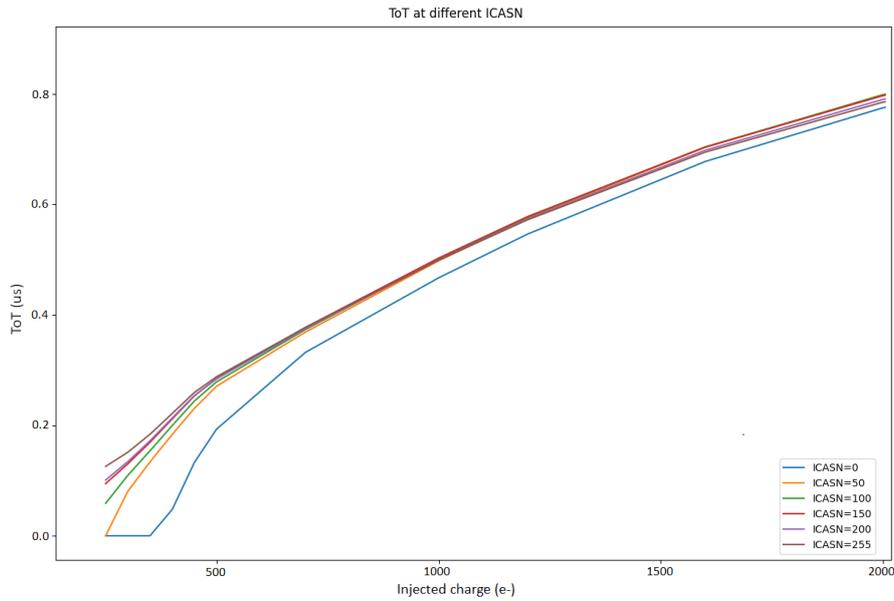
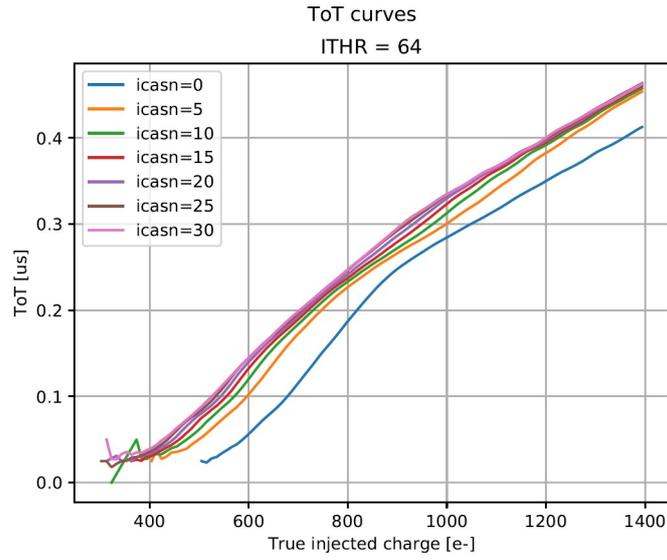


Figure 5.29: TOT vs Q_{inj} varying $ICASN$, with $I_{THR} = 64$ DAC for Cascode FE.

5.5.3 Low threshold operation, threshold dispersion and tuning

After the study on register optimization to operate the matrix at low threshold, we found a stable working point around $250 e^-$, as shown in this section. To reach this global threshold value it was also important to have a uniform threshold distribution across the pixels in the matrix.

TJ-Monopix2 is equipped with a circuit which allows the *threshold tuning*. We have already

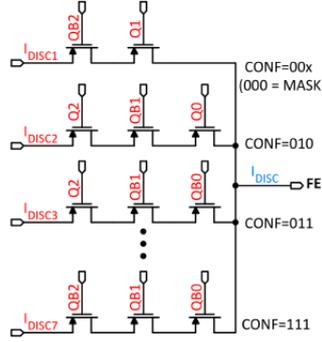


Figure 5.30: Schematic of 3-bit Tuning DAC (TDAC)

mentioned that the analog part of the in-pixel front-end (Figure 5.3) includes the 3-bit threshold tuning DAC, that can be used to adjust the discriminator threshold of each pixel with respect to the global chip threshold level thus reducing the threshold dispersion. In other words it can adjust every pixel threshold, in order to have a threshold on the matrix as uniform as possible, which is especially important to operate the matrix with low threshold, needed with the reduced collection efficiency due to radiation damage. This system has been designed to compensate for the various effects causing threshold dispersion, related to biasing non uniformity, process and temperature variations, and radiation damage.

Specifically, the Tuning DAC (TDAC) circuit shown in Figure 5.30, allows the threshold adjustment in each pixel. This component controls the discriminator active load current I_{DISC} which is partially responsible of the pixel threshold (Figure 5.4). It does not generate I_{DISC} but works as an analog multiplexer, which selects one of seven I_{DISC} lines generated by the main 8-bit biasing DAC. So the possible values of the final I_{DISC} is given by the sum of two contributions:

$$I_{DISC,TCODE} = I_{DB} + (TCODE - 1) \cdot I_{TUNE}, \quad \text{where } 1 \leq TCODE \leq 7 \quad (5.18)$$

I_{DB} is the current set by the primary value of threshold, resulting from the setting of the main registers that are responsible for it (section 5.5.1), together with I_{TUNE} , which represents the smallest possible increment to add to I_{DB} to fine tune the pixel. The other possibilities are multiples of I_{TUNE} . So $I_{DISC,TCODE}$ is the current selected by the fine tuning step (TDAC) and it depends on the 3-bit tuning code stored in the in-pixel configuration memory. "TCODE" is the decimal representation of the TDAC code.

5.5.4 First results from threshold tuning

After deciding on the target threshold, the threshold tuning algorithm tries to assign the optimized TDAC value for each pixel, in order to get as close as possible to the target. In this way, the TDAC values obtained for each single pixel are returned.

We tried to apply the fine tuning method to flatten the threshold of some pixels as much as possible. We have considered about 12.000 pixels of the Cascode FE and the results before and after the threshold trimming for the S-curves and threshold distributions are shown in Figure 5.31.

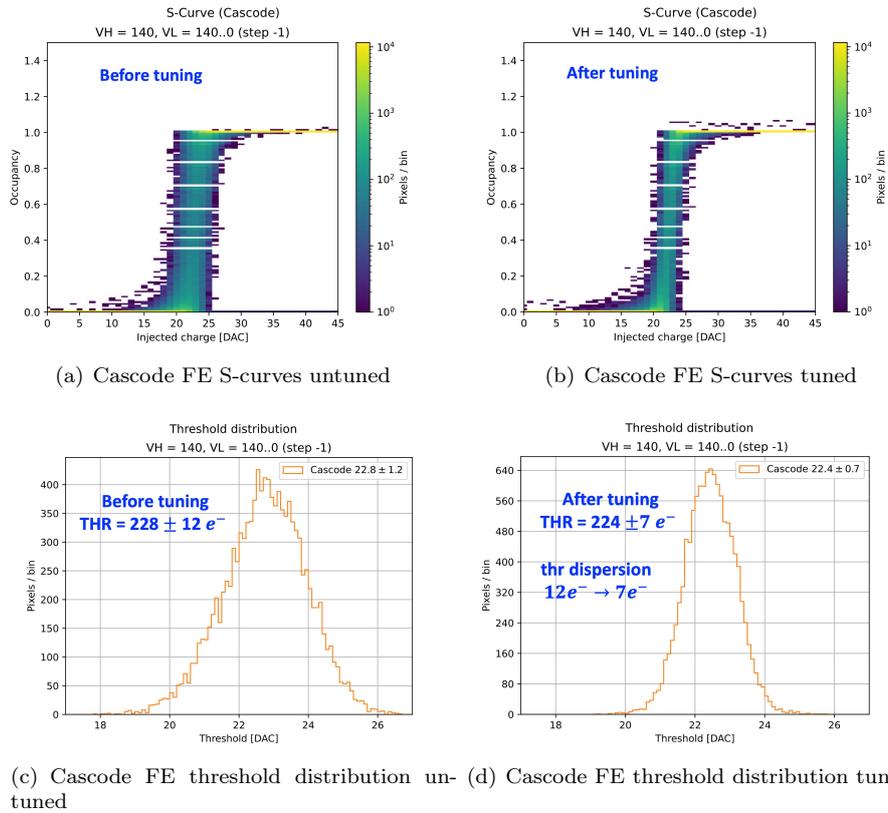


Figure 5.31: Cascade FE before tuning and after tuning comparison.

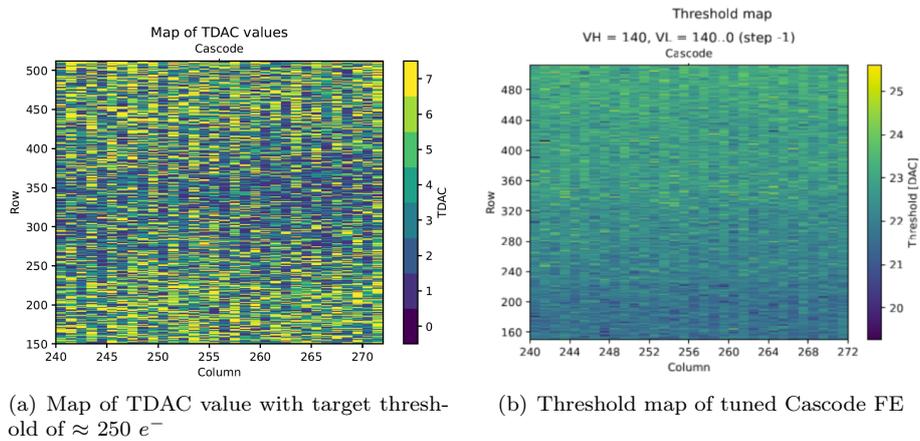


Figure 5.32: Maps of tuned Cascade FE.

As we can see the dispersion has been reduced of the 42% after the tuning and as consequence the estimation of the threshold is also more precise. Figure 5.32 displays the maps of the

threshold and of the TDAC values, such as the TCODE value assigned to each pixel, in order to obtain a threshold as uniform as possible.

5.6 Hot pixels from cross talk issue and mitigation

As previously mentioned, there was something atypical in the S-curves of the HV flavors during average threshold measurements of all FEs: some "hot" pixels started to appear, showing occupancy greater than 1, thus firing more than the number of injected events.

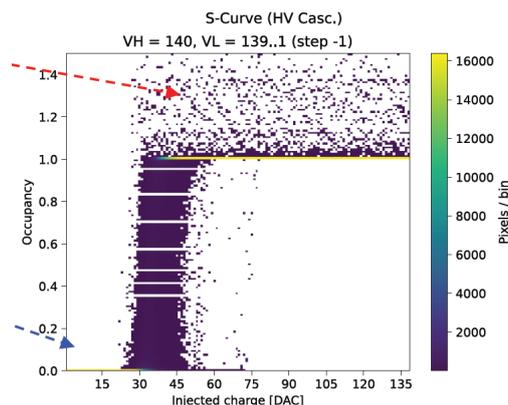


Figure 5.33: HV-Cascode S-curves showing the hot pixels.

These extra hits from hot pixels were not simply related to random fluctuations of pixels above threshold due to their noise, since these hot pixels were firing only when there was some digital activity in the matrix, during the readout sequence of other pixels. The same pixels were instead silent, with no extra hits, in case the readout sequence was not activated by other pixels.

This is visible in Figure 5.33: the region indicated by the red arrow has extra hits from hot pixels and corresponds to events in the matrix where many pixels have real hits from the injection ($Q_{inj} > THR$) and the digital readout is then active. In the region indicated by the blue arrow instead, where $Q_{inj} < THR$, there are no real hits from the injection which activates the readout, and also the hot pixels do not fire.

The observation that the hot pixels were not firing simply due to their noise fluctuation, was also confirmed with additional acquisition without injection. With a given setting of the threshold the matrix had no hits, indicating that the threshold was high enough to cut to zero the occupancy from noise fluctuation. As soon as the matrix was stimulated with real hits from a radioactive source, several hot pixels started to fire, sending hits just after the event where real hits from the radioactive source were readout. This was an additional confirmation that the spurious hits were correlated with some activity of the readout sequence.

This behavior compromises the good functionality of the overall matrix response, because these hot pixels flood the readout, giving unreliable results. Furthermore, during the systematic study of different register configurations the presence of hot pixels prevented using certain settings to reach lower global thresholds. For this reason, a detailed investigation has been conducted in order to understand the reasons why the hot pixels start to fire and how to cure them

as much as possible.

Thanks to these tests we could identify the origin of this cross-talk with some coupling of the front-end, or the input collection diode, with a specific digital signal activated in the entire matrix during the readout sequence. Using the available timing information for the hits, their Leading edge (LE), we could study the time correlation of the extra hits with the timing of the various digital signals in the readout sequence, that could be selectively moved changing some readout registers.

These tests were performed in controlled situation, as shown in the next section, and the main conclusion is the following:

- the cross talk coupling is present in all pixels of the matrix. The responsible digital signal is distributed to each pixel during the readout sequence, producing a similar signal pulse in all pixels.
- this induced or cross-talk signal is anyway producing hits only in pixels with a threshold lower than the induced signal height. For a given setting of global threshold, higher than the induced signal height, some pixels could have a significantly lower threshold, simply due to threshold dispersion or other reasons, and they behave as "hot" pixels.
- also "normal" pixels can become "hot", firing on this induced signal, if their threshold is reduced, either lowering the global threshold setting or reducing their specific threshold by the TDAC threshold tuning.

In the following section we describe some of these tests and some attempts to mitigate the issue using different settings/bias.

5.6.1 Hot pixel studies

At first, the threshold has been lowered in order to "create" hot pixels also in the first two flavors of the matrix. In fact, with the TB settings, the threshold for Normal and Cascode FE was higher (above $500 e^-$) than the threshold in the HV FEs (of about $350 e^-$) where the hot pixels were first observed, so the hypothetical induced signal did not cause spurious hits in the two first flavors. For this purpose, different settings were tried, changing some registers responsible for the threshold like those listed and explained in section 5.5.1. We confirmed that hot pixels were also appearing in the first two flavors, operating the matrix at a global threshold below $350 e^-$.

Then several tests have been performed under controlled conditions:

- one healthy (good) pixel was injected;
- one hot pixel (two or three in different tests) was enabled but not injected;
- all the matrix except these pixels was disabled.

In this way the readout cycle, shown schematically in Figure 5.34, has a known duration, and the timing of the various digital control signals (FREEZE, READ) has a known position with respect to Trailing Edge (TE) of the first hit which activates the readout. The digital readout signal timing can be moved acting on specific registers listed in Table 5.11. We studied in detail the time correlation of the hot pixels LE and TE with the digital signals.

Two different timing information have been used to study the time correlation of the induced signals with the digital signals:

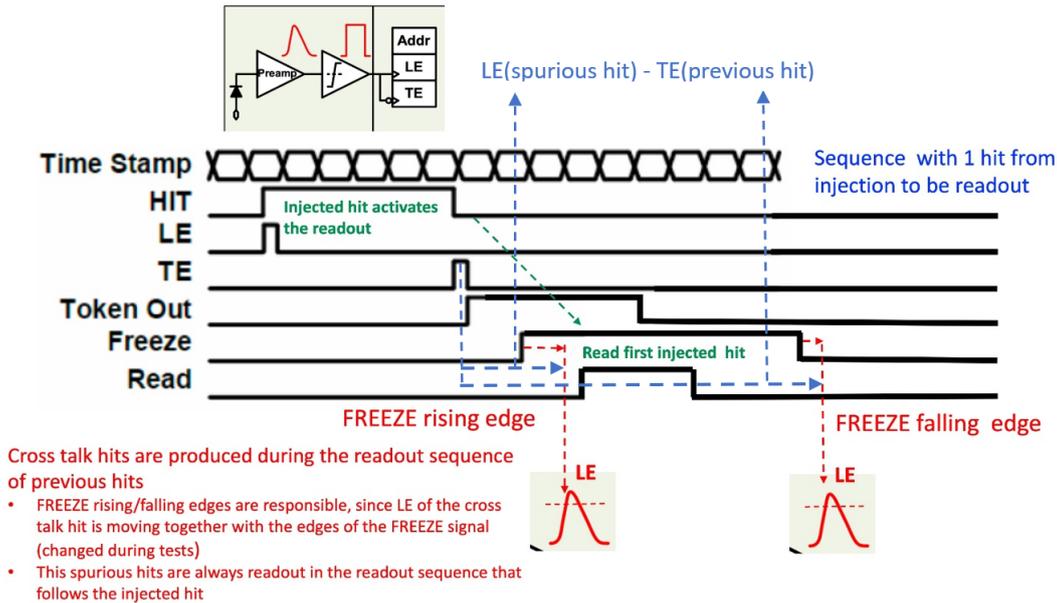


Figure 5.34: Hot pixel study: readout sequence.

- ΔTS (TimeStamp) between two consecutive hits: the timestamp is assigned by the FPGA when the TOKEN rises on the TE of the first hit to read, but only if the previous readout frame is completed. Therefore, if the hit coming from a hot pixel is after the hit from the injected one, the minimum ΔTS has to be equal to the readout time of 1 pixel and so to the duration of the FREEZE signal. This information has allowed to verify whether the hot pixel fires immediately after the good injected one or not.
- $LE(\text{hit}) - TE(\text{previous hit})$: this quantity measures the elapsed time between the start of a hit (Leading Edge) and the end of the previous one (Trailing Edge). The TE of the first hit activates the readout and also corresponds to the beginning of the entire readout sequence. All other readout control signals have a known/selectable timing position with respect to this TE. Thus the distance $LE(\text{hit}) - TE(\text{previous hit})$ allows to correlate the LE of the spurious hit, due to cross-talk, with the specific digital signal edge that produces the effect. This quantity is shown in Figure 5.34 for the two cases of hot pixels firing on the rising or falling edge of the FREEZE signal.

5.6.2 Digital signal responsible for cross talk

Referring to the readout sequence, in order to understand which digital signal could induce cross-talk, we moved the registers used to control the timing of these signals with respect to the start of the readout sequence (i.e. with respect to the TE of the first hit). Thus we expected to see the same shift in the LE-TE distance when the responsible digital signal was changed.

Table 5.11 shows an example of the settings used for these registers. With a timestamp frequency of 40 MHz, one clock cycle corresponds to 25 ns.

For example, if the register FREEZE_START_CONF, which is the rising edge of the

Register	Value (clock cycles)
FREEZE_START_CONF	10
READ_START_CONF	13
READ_STOP_CONF	15
LOAD_CONF	30
FREEZE_STOP_CONF	31
STOP_CONF	31

Table 5.11: Register values of the readout cycle.

FREEZE signal, is responsible for the cross-talk, we expect that shifting its value by a certain number of clock cycles would also delay the timing of the cross-talk induced hit by the same number of clock cycles, and this should be visible in the LE-TE position.

With this procedure, repeated for each readout register, we have confirmed that the cross-talk is related to the rising and falling edge of the FREEZE signal. In the following examples, one pixel is injected (217,140) and two pixels are enabled (218,155) and (222,188), while the rest of the matrix is disabled.

```

Injected pixel: (217, 140)
Other pixels: [(218, 155) (222, 188)]
Assuming timestamp clock = 40.00 MHz
Green = injected pixels

```

Row	Col	LE	TE	Δ LE	Δ TE	Δ TS[25ns]	TS[25ns]
140	217	12	29	124	123	5627.0000	635972.0000
140	217	8	25	124	124	5628.0000	641600.0000
155	218	60	60	52	35	35.0000	641635.0000
188	222	59	60	127	0	0.0000	641635.0000
155	218	115	115	56	55	55.0000	641690.0000
188	222	114	115	127	0	0.0000	641690.0000
155	218	42	43	56	56	55.0000	641745.0000
188	222	42	42	0	127	0.0000	641745.0000
140	217	4	21	90	107	5482.0000	647227.0000

Figure 5.35: An example of the time information used.

In Figure 5.35 an example of the time information used is shown, where the timings of the injected pixel are green coloured, that of the hot pixels are white. The quantities reported in the columns of the table are: the row and the column of the hit pixel, the LE and the TE of the hit, the difference between the LE of the hit and the LE of the previous one (Δ LE), the difference between the TE of the hit and the TE of the previous one (Δ TE), the difference between the timestamps of the same consecutive hits (Δ TS), and the timestamp of the hit. All these quantities are expressed in clock cycles. Each row shows a hit coming from the enabled pixels.

In the first two rows (green), two hits of the injected pixel (114,217) are displayed. Their timestamps are different, and the Δ TS \approx 5600 clock cycles can be attributed to the time between two injections. The timestamp is assigned by the FPGA when the TOKEN rises on the TE of the first hit to read.

The hits of the two hot pixels (155,218) and (188,22) (third and fourth row respectively), have TOT \approx 0, thus the LE coincides with the TE, and their timestamps are the same, so they fire at the same time. In particular, the Δ TE of the first hot pixel is 35 clock cycles, which means

that it fires during the readout of the previous injected hit (longer blue arrow in Figure 5.34). In this case the ΔTE corresponds to the time elapsed between the TE of the injected hit and the LE of the hit coming from the hot pixel, for which $TE \simeq LE$. For the same reason, the zero ΔTE of the hit in the fourth row corresponds to the time elapsed between the two hot pixels' hits, which means that they fire at the same time, due to the same cross-talk induced signal (due to the TE of the FREEZE). The hit in the fifth row has a different timestamp with respect to the previous one, but they differ from 55 clock cycles, which means that this hot pixel fires again during the readout of the two previous hit (specifically after the TE of the FREEZE).

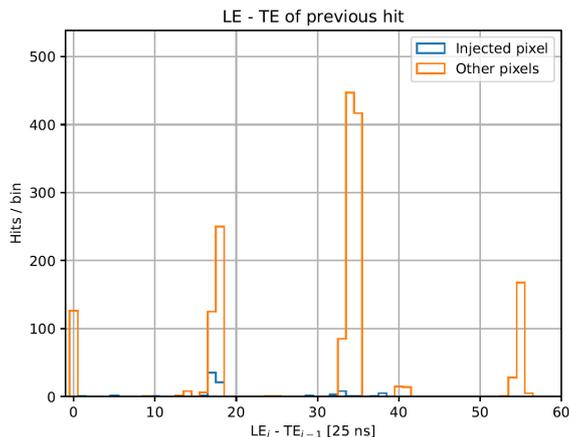


Figure 5.36: An example of the LE(hit)-TE(previous hit) histogram.

Figure 5.36 shows the histogram of the elapsed time between the leading edge LE of an hit and the trailing edge TE of the previous hit, when one pixel is injected and two additional hot pixels are enabled. When only one hit is readout, the duration of the FREEZE signal is of 21 clock cycles.

It is possible to see several peaks in the LE(hit) - TE(previous hit) distribution (referring to the readout setting reported in Table 5.11):

- a peak at 0 clock cycle, representing the event where both hits come from the two hot pixels firing simultaneously after the injection in the other pixel. This means that they are both activated by a given signal and so, it is the most important confirmation that they come from cross-talk and not from a random firing pixel signal;
- a peak at ≈ 18 clock cycles equal to the FREEZE_START_CONF value (10 clock cycles) + 8 clock cycles \rightarrow the rising edge of the FREEZE signal is responsible for the induced signal;
- a more pronounced peak at ≈ 35 clock cycles equal to the FREEZE_STOP_CONF value (31 clock cycles) + 4 clock cycles \rightarrow the falling edge of the FREEZE signal is also responsible for the induced signal with higher probability than the FREEZE rising edge;
- a peak at ≈ 55 clock cycles equal to the FREEZE falling edge (located at 51 clock cycles when 2 different hits are readout in the same sequence) + 4 clock cycles. In more detail, after the first 30 clock cycles until the first LOAD_CONF, an additional pixel reading starts and it takes another 20 cycles (LOAD_CONF - FREEZE_START_CONF) + 1 cycle to conclude the frame. Therefore, when two pixels are read, the FREEZE falls after 51 clock cycles as indicated above, and it is compatible with the last peak in the plot.

An additional confirmation of the previous studies, that indicate the FREEZE as responsible for this cross-talk, was also found by directly observing through the oscilloscope the behaviour of the analog output of a test pixel, available for debugging purpose. In Figure 5.37 an analog acquisition of the readout signals is shown, taken during the test performed in a similar way as before.

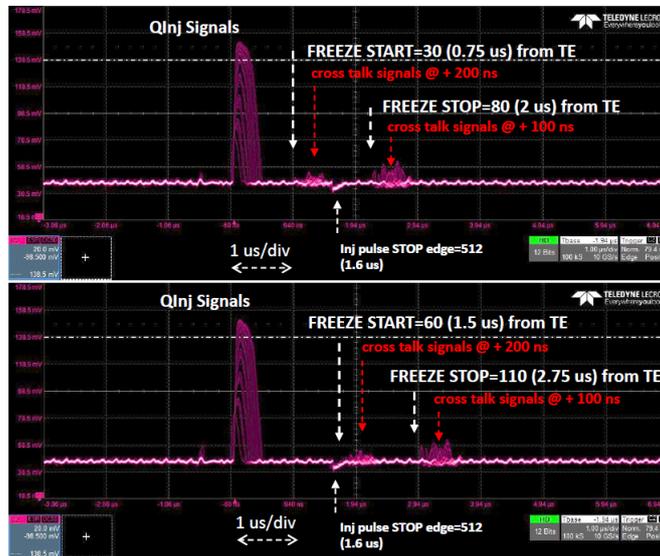


Figure 5.37: Cross-talk of the FREEZE signal on oscilloscope’s analog output, for different value of FREEZE_START_CONF register.

In these tests the analog pixel, together with another pixel in the matrix, was injected with variable signals from 0 to 140 DAC (this can be seen by the increasing signal height in the acquisition). Two different groups of spikes are also visible in the analog output: the first and smaller one represents the cross-talk effect on this pixel from the raising of the FREEZE signal during the readout sequence of the other pixel, while the second, larger one corresponds to the cross-talk from the falling edge of the same signal. Moreover, it is possible to see that in the two different pictures, the cross-talk induced signals move according to the different settings of the FREEZE START/STOP edge that was changed in the two acquisitions.

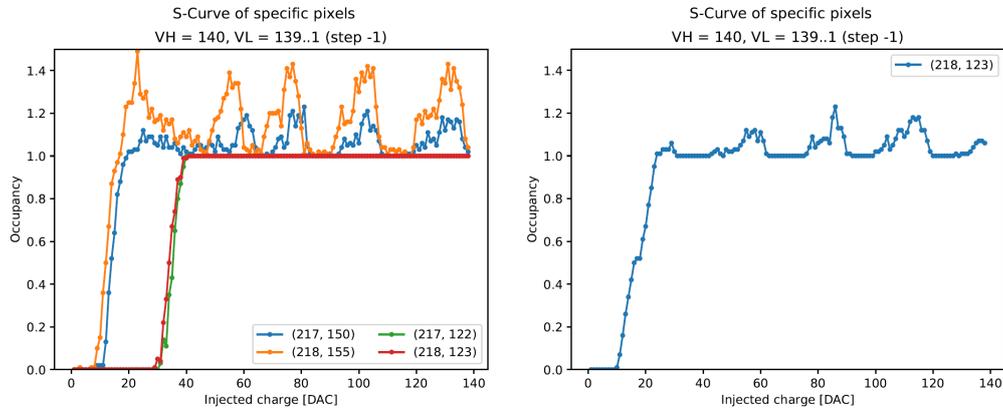
As already stated, we run several tests varying the number of pixels to read and the value of the readout registers, using different combination of hot and good pixels and also different spatial location of them in the matrix to exclude the possibility that the problem was related to particular columns. All results are in agreement with the interpretation explained above.

5.6.3 Cross talk signal height

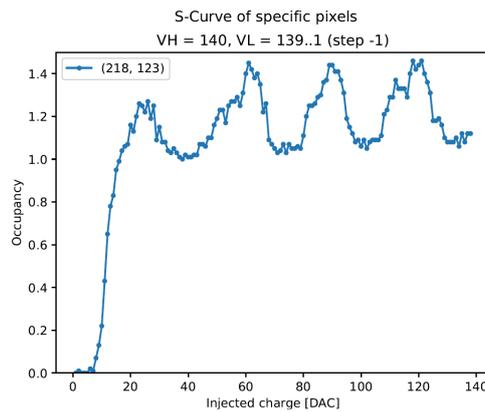
We also tried to estimate the induced signal height from the threshold of the hot pixel. For this purpose, we have tried different settings of the registers cited above, to make a pixel "hot" in order to understand when the induced signal went above the threshold. We have found that the induced signal correspond to about $100/150 e^-$ depending on the bias condition.

An example of these tests is shown in Figure 5.38. We enabled part of the matrix (the columns 217 and 218 and the rows between 120 and 220), injecting only one pixel (217, 140).

In Figure 5.38(a) the S-curves of different pixels are shown and among these we chose one that was not hot (218, 123).



(a) $I_{DB}=100$, $I_{TUNE}=53$ - Good behavior for the pixel (218,123) (the red one) with $THR > 35 \text{ DAC} = 350 e^-$, while other pixels with $THR \approx 15 \text{ DAC} = 150 e^-$ are already "hot"
 (b) $I_{DB}=60$, $I_{TUNE}=150$ - The pixel (218,123) has THR around 20 DAC and starts to misbehave



(c) $I_{DB}=55$, $I_{TUNE}=150$ - The pixel has THR below 15 DAC and becomes "hot"

Figure 5.38: S-curve of the pixel (218, 123) for different register settings, used to gradually reduce its threshold.

The S-curves of the same pixel (218, 123) are shown for different register settings used to gradually reduce its threshold. The pixel that has a normal S-curve when it has a high threshold of approximately $350 e^-$ (Figure 5.38(a)), starts to misbehave with occupancy slightly higher than 1 for threshold of about $200 e^-$ (Figure 5.38(b)), and becomes hot for threshold of about $150 e^-$ (Figure 5.38(c)). We can notice a structure in the S-curve of the hot pixels when the occupancy becomes greater than 1. We did not investigate further this behaviour.

5.6.4 Cross talk mitigation

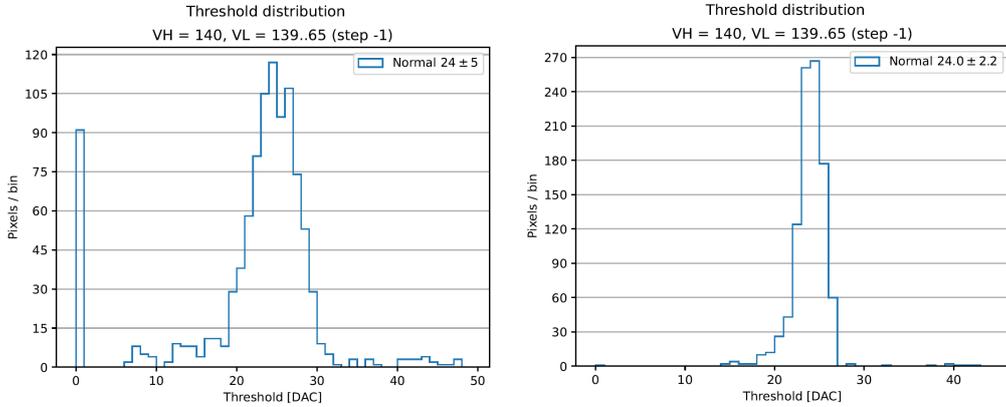
As seen in the previous, the hot pixel problem is related to the induced signal produced during the readout, which causes cross-talk. This becomes even more serious when there is a larger threshold dispersion, since although the global threshold is set well above the level of the induced signal, some pixels can still have a threshold lower than the induced signal.

Potentially every pixel could become "hot" if its threshold is lower than the height of the cross-talk signal, since the FREEZE is sent across the entire matrix.

Threshold trimming

Therefore, a possible treatment could be related to the threshold tuning, explained in section 5.5.3, which could allow to make the pixel thresholds more uniform (smaller threshold dispersion) and at the same time to target a value higher than the induced signals.

Figure 5.39 shows an example of the results obtained.



(a) Threshold distribution before tuning procedure. (b) Threshold distribution after tuning procedure.

Figure 5.39: Threshold tuning to reduce hot pixels.

The reduction of the tail in the threshold distribution is evident, in fact the dispersion is reduced by 56%. The number of the hot pixels also decreases from 18% to 1.2% of the total pixels studied. We can also notice that the peak at 0 threshold disappears, which comes from those hot pixels with low thresholds, whose S-curve is so distorted that its threshold evaluation is 0. The hot pixels are those with occupancy greater than 1 and an example of a hot pixel (218,155) with threshold 0 is shown in Figure 5.40. After the tuning their thresholds are increased, preventing them from misbehaving.

Bias Voltage

Moreover, we have tried to increase the bias voltage of the matrix to study how it might affect the cross-talk. All previous tests have been performed with P_{WELL}/P_{SUB} set to $-3V$. This value was increased to $-6V$ and there were some improvements. In fact increasing the bias, we expected a decrease of the diode capacitance thus higher gain and lower threshold dispersion. In addition the coupling with the cross-talk signal seemed reduced too and so the induced signal height.

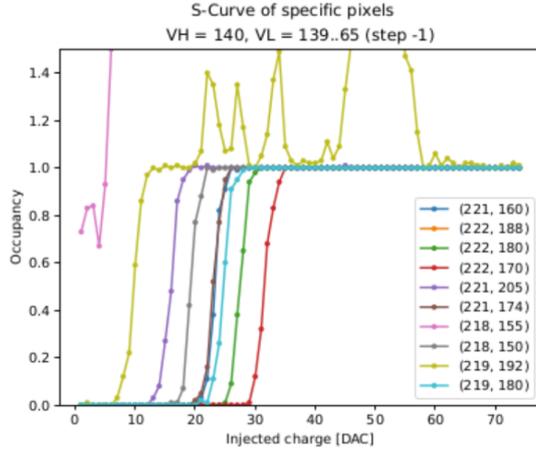


Figure 5.40: An example of a hot pixel (218,155) with threshold 0.

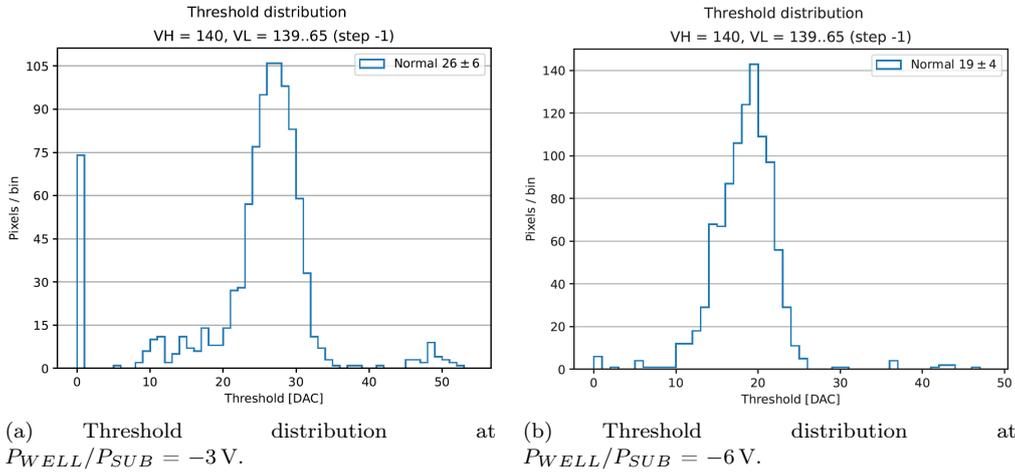


Figure 5.41: A comparison between the threshold distributions resulted with different biasing voltage, without tuning.

Figure 5.41 shows a comparison between the threshold distributions respectively at $-3V$ and $-6V$, with the same register setting and without tuning.

At higher bias voltage not only the threshold is lower (higher gain), but also its dispersion is, as expected. And despite that, there are fewer hot pixels: 1.3% at $-6V$ against 17% at $-3V$. Also here, the reduction of the threshold distribution tail is evident.

5.6.5 Conclusion from cross talk studies

The results obtained with both the threshold tuning and a bias voltage of $P_{WELL}/P_{SUB} = -6V$ are shown in Figure 5.42, where we can see the decrease in threshold dispersion with the number of the hot pixels.

In fact, without the tuning there are 1.3% of them, while after the tuning procedure there

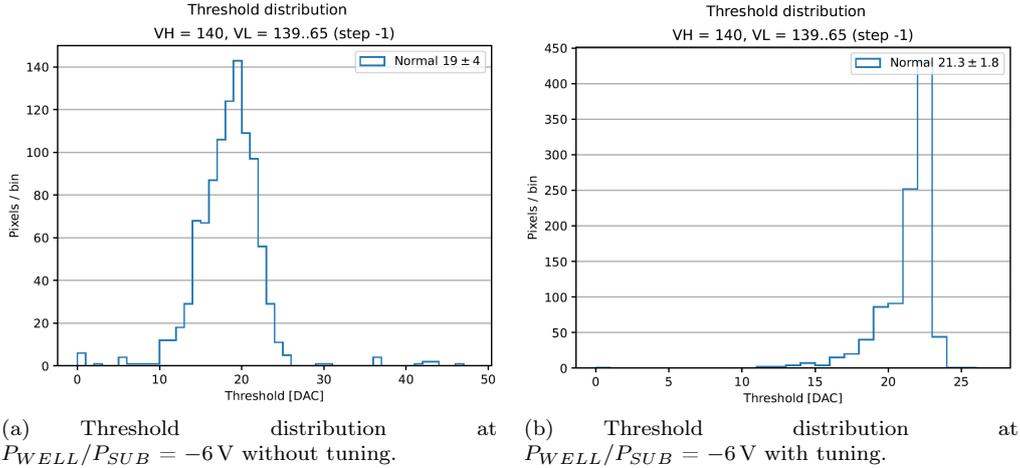


Figure 5.42: A comparison between the threshold distributions at -6 V bias voltage without tuning (a) and with tuning (b).

are none.

After these tests, we could conclude that with a bias voltage of -6 V and threshold tuning it was possible to operate the matrix with a global threshold of approximately $200/250 e^-$, much higher of the expected minimum threshold of $100 e^-$ (Table 5.1) indicated by the simulations done during the TJ-Monopix2 design, but still totally acceptable to keep a good detection efficiency even after irradiation.

These cross-talk studies were also crucial to indicate possible mitigation strategies for the next OBELIX chip. The origin of the this cross-talk coupling is still under investigation with detailed simulations of the matrix behaviour. At the same time in OBELIX we plan to increase the range of the threshold tuning DAC in order to have the possibility to selectively increase the pixel threshold that are in the tail of the threshold distribution and could otherwise become hot and disturb the readout.

It should be also noted the cross-talk charge injection happens during the readout and therefore does not affect the charge collected by pixel hit by real particles.

5.7 Test Beam results

The full characterization of the chip allowed to interpret data collected during the Test Beam campaign at Desy (June 2022). Several tests have been conducted to study the electrical characteristics and the hit detection efficiency of the unirradiated modules [38].

Experimental apparatus and DUTs

The measurements have been performed using an electron beam with energies in the range of 3-5 GeV, at DESY II testbeam facility at DESY, Hamburg [39]. The experimental apparatus consisted of a beam telescope, a Trigger Logic Unit to provide trigger and control signals employed during test beams, a scintillator trigger and a rotation-translation stage on which install the device under test. (Figure 5.43).

Three different modules have been tested with different sensor geometries, among which the chip W14R12 that we have studied in depth through laboratory measurements. All results

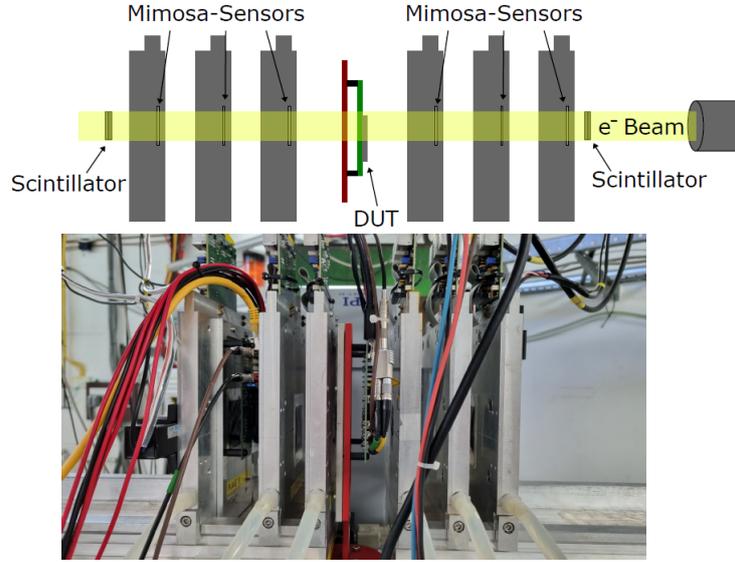


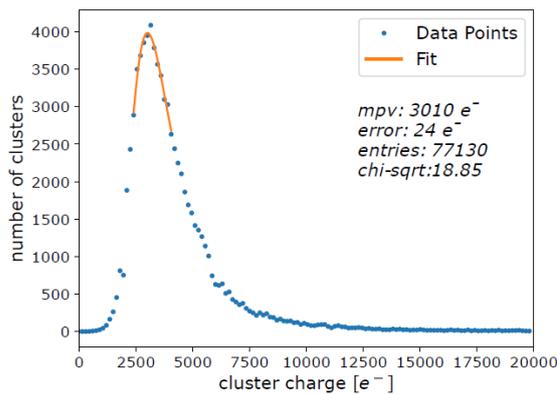
Figure 5.43: Test Beam experimental apparatus.

described in previous sections have been crucial to interpret data obtained during these tests. Preliminary settings were used (Table 5.2), featuring high threshold ($\approx 500 e^-$).

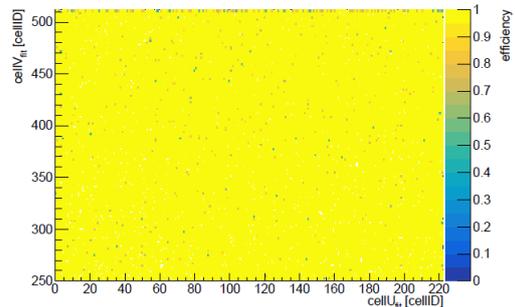
We will briefly describe the results obtained in the following [40].

Cluster charge distribution and efficiency measurements

The cluster distribution obtained from electron beam measurements is shown in Figure 5.44(a). This distribution was fitted by a Landau function and the signal of each pixel was corrected by a calibration factor, which we have discussed in depth in section 5.4.4. The MPV of the cluster charge was obtained from the fit, resulting in $(3010 \pm 24) e^-$, which is in line with the calculated deposited energy of about $\approx (3200 \pm 194) e^-$.



(a) Cluster charge distribution.



(b) $(99.54 \pm 0.04) \%$ hit detection efficiency of DC-coupled flavours at 3 V bias voltage.

The hit detection efficiency ϵ is evaluated by the ratio of the matched hit to the total tracks. For the DC-coupled flavors, for a threshold of $\approx 500\text{-}600 e^-$ (Table 5.3), the hit efficiency measured is $99.54 \pm 0.04 \%$, as shown in Figure 5.44(b). This result will be compared with the efficiency of the irradiated chip, to test the good efficiency performance even after irradiation.

In the next phase the sensors have been irradiated ranging from 10^{14} to $10^{15} \text{ n}_{\text{eq}}/\text{cm}^2$, and in a new test beam campaign in July 2023, the chip response have been examined under these conditions.

6. Conclusions

This thesis presents the characterization of the TJ-Monopix2 chip (W14R12), a small collection electrode DMAPS prototype in a modified TowerJazz Semiconductor imaging process.

Threshold and noise distributions have been studied for each frontend, using the same register setting and biasing voltages employed during the Test Beam campaign in Desy (June 2022). The registers' values were not optimized to take measurements at low threshold, therefore the results obtained in the first part of the work, cannot be compared with the specifications of the chip, but they have defined the working point of the matrix, crucial for interpreting the data collected during the Test Beam. We have measured a threshold of $\approx 500\text{-}600 e^-$ with a noise of about $25 e^-$ and a threshold dispersion of $\approx 20\text{-}30 e^-$ (for Normal and Cascode flavours).

We have characterized the Time Over Threshold (TOT) calibration function, which allows to translate the time information to the charge collected in the pixel, using the internal injection circuit. A first absolute calibration has been performed using a ^{55}Fe radioactive source with a known emission line of 5.9 KeV ($\approx 1616 e^-$).

We have also measured the response of the matrix to other radioactive sources like ^{241}Am and ^{109}Cd , to extend the study of the TOT spectrum above the limit imposed by the saturation of the internal injection circuit.

We have studied the operation of the matrix at low threshold, that is necessary to maintain good efficiency after irradiation, when radiation damage causes a reduction in the collected charge. Exploiting the threshold tuning circuit, which allows to adjust the threshold of the individual pixels, we have achieved a global threshold of $\approx 224 e^-$ with a threshold dispersion of $\approx 7 e^-$. These values could be compared with the chip specifications, which in a optimized operating point, indicate a threshold of $100 e^-$ and a threshold dispersion $\leq 10 e^-$. The charge released by a MIP in this thin sensor is approximately $(3010 \pm 24) e^-$. The possibility to reach lower threshold has been limited by the presence of a cross-talk issue, discovered during this investigation. After an accurate analysis, we have found the cause in one of the readout logic signals, the FREEZE. This signal is responsible for freezing the matrix during the readout cycle, preventing new hit from disturbing the readout of the previous one, before the cycle is completed. The cross-talk is induced by the leading and the trailing edge of the FREEZE, and if the height of this induced signal is higher than the threshold of the pixel, it causes spurious hits flooding the readout and invalidating the measurements.

The characterization of the TJ-Monopix2 chip has been useful to interpret data collected during the Test Beam (2022) with unirradiated sensor. The analysis conducted about the origin of the cross-talk issue, has been relevant in the design of the next chip, OBELIX, which will be the final prototype chosen for the VTX upgrade program.

Bibliography

- [1] E. K. et al., *The Belle II Physics Book.*, *Progress of Theoretical and Experimental Physics* **2019** (December 2019) .
- [2] BELLE-II collaboration, *Belle II Technical Design Report*, [1011.0352].
- [3] Y. Ohnishi et al., *Accelerator design at SuperKEKB*, *PTEP* **2013** (2013) 03A011.
- [4] SUPERB collaboration, *SuperB Technical Design Report*, [1306.5655].
- [5] BELLE collaboration, *Physics Achievements from the Belle Experiment*, *PTEP* **2012** (2012) 04D001 [1212.5342].
- [6] SUPERKEKB, BELLE II collaboration, *Recent progress of SuperKEKB project and future prospect*, *JACoW IPAC2023* (2023) TUOGB1.
- [7] N. Iida et al., *Beam injection issues at SuperKEKB*, *JACoW IPAC2023* (2023) MOPL120.
- [8] A. Natchii et al., *Beam background expectations for Belle II at SuperKEKB*, in *Snowmass 2021*, 3, 2022, [2203.05731].
- [9] BELLE-II collaboration, *Detectors for extreme luminosity: Belle II*, *Nucl. Instrum. Meth. A* **907** (2018) 46.
- [10] BELLE-II collaboration, *Snowmass Whitepaper: The Belle II Detector Upgrade Program*, in *Snowmass 2021*, 3, 2022, [2203.11349].
- [11] US BELLE II GROUP, BELLE II/SUPERKEKB E- POLARIZATION UPGRADE WORKING GROUP collaboration, *Snowmass 2021 White Paper on Upgrading SuperKEKB with a Polarized Electron Beam: Discovery Potential and Proposed Implementation*, in *Snowmass 2021*, 5, 2022, [2205.12847].
- [12] Z. J. Liptak, *Toward a new era of spin-polarized electron beams at SuperKEKB*, *JACoW IPAC2023* (2023) MOPL073.
- [13] A. Natchii et al., *Measured and projected beam backgrounds in the Belle II experiment at the SuperKEKB collider*, *Nucl. Instrum. Meth. A* **1055** (2023) 168550 [2302.01566].
- [14] ALICE collaboration, *Upgrade of the ALICE inner tracking system: Construction and commissioning*, *Phys. Scripta* **95** (2020) 084011.
- [15] A. Ishikawa et al., *New pixel detector concept DuTiP for Belle II upgrade and the ILC with an SOI technology*, *Nucl. Instrum. Meth. A* **978** (2020) 164404.

- [16] ALICE collaboration, *The ALPIDE pixel sensor chip for the upgrade of the ALICE Inner Tracking System*, *Nucl. Instrum. Meth. A* **845** (2017) 583.
- [17] H. Murayama et al., *Development of monolithic SOI pixel sensors capable of fine measurements of space and time*, *Nucl. Instrum. Meth. A* **978** (2020) 164417.
- [18] M. Garcia-Sciveres and N. Wermes, *A review of advances in pixel detectors for experiments with high rate and radiation*, *Rept. Prog. Phys.* **81** (2018) 066101 [1705.10150].
- [19] DEPFET, BELLE II PXD collaboration, *Commissioning of the Belle II Pixel Vertex Detector*, *PoS EPS-HEP2019* (2020) 122.
- [20] H. Kolanoski and N. Wermes, *1Introduction*, in *Particle Detectors: Fundamentals and Applications*, Oxford University Press, (06, 2020), 10.1093/oso/9780198858362.003.0001.
- [21] L. Greiner, E. C. Anderssen, H. G. Ritter, J. Silber, T. Stezelberger, X. Sun et al., *A MAPS based pixel vertex detector for the STAR experiment*, *PoS Vertex2012* (2013) 010.
- [22] ALICE collaboration, *Technical Design Report for the Upgrade of the ALICE Inner Tracking System*, *J. Phys. G* **41** (2014) 087002.
- [23] W. Snoeys et al., *A process modification for CMOS monolithic active pixel sensors for enhanced depletion, timing performance and radiation tolerance*, *Nucl. Instrum. Meth. A* **871** (2017) 90.
- [24] M. Battaglia, D. Bisello, D. Contarato, P. Denes, P. Giubilato, L. Glesener et al., *Monolithic Pixel Sensors in Deep-Submicron SOI Technology*, *JINST* **4** (2009) P04007 [0903.3205].
- [25] C. Bespin et al., *DMAPS Monopix developments in large and small electrode designs*, *Nucl. Instrum. Meth. A* **978** (2020) 164460 [2006.02297].
- [26] M. Barbero et al., *Radiation hard DMAPS pixel sensors in 150 nm CMOS technology for operation at LHC*, *JINST* **15** (2020) 05 [1911.01119].
- [27] D. Kim et al., *Front end optimization for the monolithic active pixel sensor of the ALICE Inner Tracking System upgrade*, *JINST* **11** (2016) C02042.
- [28] G. Aglieri Rinella et al., *Charge collection properties of TowerJazz 180nm CMOS Pixel Sensors in dependence of pixel geometries and bias parameters, studied using a dedicated test-vehicle: The Investigator chip*, *Nucl. Instrum. Meth. A* **988** (2021) 164859 [2009.10517].
- [29] K. Moustakas et al., *Development in a Novel CMOS Process for Depleted Monolithic Active Pixel Sensors*, in *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference*, p. 8533114, 2017, 10.1109/NSSMIC.2017.8533114.
- [30] I. Caicedo et al., *The Monopix chips: Depleted monolithic active pixel sensors with a column-drain read-out architecture for the ATLAS Inner Tracker upgrade*, *JINST* **14** (2019) C06006 [1902.03679].
- [31] M. Munker, M. Benoit, D. Dannheim, A. Fenigstein, T. Kugathasan, T. Leitner et al., *Simulations of CMOS pixel sensors with a small collection electrode, improved for a faster charge collection and increased radiation tolerance*, *JINST* **14** (2019) C05013 [1903.10190].

- [32] K. Moustakas et al., *CMOS Monolithic Pixel Sensors based on the Column-Drain Architecture for the HL-LHC Upgrade*, *Nucl. Instrum. Meth. A* **936** (2019) 604 [1809.03434].
- [33] K. Moustakas, *Design and Development of Depleted Monolithic Active Pixel Sensors with Small Collection Electrode for High-Radiation Applications*, Ph.D. thesis, U. Bonn (main), Bonn U., 2021. <https://cds.cern.ch/record/2782279>.
- [34] BELLE II VTX collaboration, *CMOS MAPS upgrade for the Belle II Vertex Detector*, *Nucl. Instrum. Meth. A* **1048** (2023) 168015.
- [35] F. Forti for Belle II, *The Belle II Detector Upgrades Conceptual Design Report*, <https://docs.belle2.org/record/3930?ln=en> (2023) .
- [36] BELLE-II collaboration, *Simulation of an all-layer monolithic pixel vertex detector for the Belle II upgrade*, *Nucl. Instrum. Meth. A* **1045** (2023) 167616.
- [37] K. Terada, S. Nakamura, T. Nakao, A. Kimura, O. Iwamoto, H. Harada et al., *Measurements of gamma-ray emission probabilities of ^{241}Am , ^{243}Am and ^{239}Np* , *Journal of Nuclear Science and Technology* **53** (2016) 1881.
- [38] C. Bespin et al., *Charge collection and efficiency measurements of the TJ-Monopix2 DMAPS in 180 nm CMOS technology*, *PoS Pixel2022* (2023) 080 [2301.13638].
- [39] R. Diener et al., *The DESY II Test Beam Facility*, *Nucl. Instrum. Meth. A* **922** (2019) 265 [1807.09328].
- [40] BELLE II VTX UPGRADE GROUP collaboration, *Upgrade of Belle II Vertex Detector with CMOS Pixel Technology*, in *24th international Workshop on Radiation Imaging Detectors*, 11, 2023, [2311.13360].