



UNIVERSITA' DEGLI STUDI DI PAVIA

FACOLTA' DI INGEGNERIA

**DIPARTIMENTO DI INGEGNERIA
INDUSTRIALE E DELL'INFORMAZIONE**

**MIXED SIGNAL READOUT CIRCUITS
FOR PIXEL DETECTOR SYSTEMS
BASED ON HIGH DENSITY
MICROELECTRONIC TECHNOLOGIES**

Tutor:

Prof. Lodovico Ratti

Coordinatore del Dottorato:

Prof. Franco Maloberti

Tesi di Dottorato
di Alessia Manazza

Anno Accademico 2011/2012

Contents

Introduction	1
1 Front-end for semiconductor detectors	3
1.1 Detectors for particle physics applications	4
1.1.1 Microstrip detectors	5
1.1.2 Hybrid pixel detectors	7
1.1.3 Monolithic pixel detectors	8
1.2 Architectures for analog readout	11
1.2.1 The charge sensitive amplifier	11
1.2.2 The shaper	12
1.2.3 Front-end readout	13
1.3 Readout architectures	14
1.3.1 Chips without data buffering	15
1.3.2 Chips with zero suppression and data buffering	15
1.3.3 Counting chips	21
1.4 Examples of pixel detector applications	22
1.4.1 The ATLAS pixel detector	23
1.4.2 The CMS pixel detector	28
2 Superpix0	33
2.1 The SuperB experiment	33
2.1.1 Design specifications for the Layer0 of SuperB SVT	34
2.2 The Superpix0 chip	35
2.2.1 The pixel sensor matrix	36
2.2.2 The analog front-end	37
2.2.3 Injection circuit for the chip calibration	63
2.2.4 Power distribution	64
2.2.5 Power consumption	66
2.2.6 Digital front-end	67

2.2.7	Readout architecture	68
2.2.8	Layout	71
2.3	Characterization results	73
2.3.1	Noise measurement with threshold scan techniques . . .	75
2.3.2	Inject threshold scan measurement	77
2.3.3	Measurements with radioactive sources	78
2.3.4	Beam test results	79
3	Superpix1	83
3.1	Vertical integration technologies	83
3.1.1	The Tezzaron/Globalfoundries technology	86
3.2	Characterization of the T/G prototypes	91
3.2.1	Devices under test	92
3.2.2	Characterization of the SDR1 analog front-end	96
3.2.3	Characterization of the SDR1 readout circuits	99
3.2.4	Characterization of the APSEL5T-TC analog front-end	104
3.3	The Superpix1 chip	111
3.3.1	Analog front-end	112
3.3.2	Injection circuit for the chip calibration	123
3.3.3	Threshold dispersion and detection efficiency	128
3.4	Layout	146
	Conclusions	150
	Bibliography	151

Introduction

Experiments at the future high luminosity colliders, like the SuperB Factory, will set severe requirements on each of the parts making up the silicon vertex tracker (SVT), including the detector and the readout electronics. In order to separate the very dense particle jets emerging from the interaction region, the first detector layer will be placed very close to the pipeline axis and will have to provide remarkably high spatial resolution, with a pitch in the order of $50\text{ }\mu\text{m}$ or smaller. Thin detectors and readout electronic chips, with overall thickness not exceeding a few hundred microns, will be required for the purpose of minimizing the amount of material in the sensitive region of the tracker, therefore reducing multiple scattering and improving momentum measurement accuracy. These specifications have a number of implications on the design of the readout electronics and on the technology choice. The use of low mass cooling systems, which is mandatory to comply with the material budget requirements, sets a limit to the maximum acceptable power dissipation in the front-end chip. A high granularity detector placed at a small distance from the interaction point results in an increased data rate, which may be dealt with by means of selective, or sparsified, readout architectures. Selective data readout requires that some amount of intelligence is included in the front-end electronics and that digital blocks are placed in the elementary cell, together with the analog front-end, and in the chip periphery. In order to cope with such severe functional density requirements, a vertical integration, 3D, processes have been taken into consideration. Vertical integration technologies have been recently proposed for the design of particle sensors for high energy physics (HEP) applications. 3D circuit manufacturing involves

the independent fabrication of two or more planar circuits on different wafers, which are subsequently bonded together after precise alignment and thinning, therefore providing higher integration density. Actually, 3D technologies allow the designer to comply with the increasing spatial resolution requirements set by HEP applications and provide the opportunity of developing more complex in-pixel circuits, improving the readout performance. Also, use of two layers makes it possible to improve the electrical isolation between the digital and analog sections of the front-end, therefore strongly reducing cross-talk issues in a mixed-signal circuit. This work discusses the design and characterization of a readout chip for high granularity hybrid pixel detectors in a planar 130 nm CMOS technology and presents a much improved version of the same circuit in a vertical integration CMOS process, both conceived for application to the SuperB vertex detector. The structure of the work is as follows.

The first chapter begins with a short description of the fundamental operating principles of microstrip detectors, hybrid pixel detectors and monolithic active pixel sensors (MAPS), which are the options under consideration for the Layer0, the innermost layer of the SuperB SVT. The second part of the chapter will be devoted to the description of the main features of some classical front-end (FE) schemes and, for the sake of example, of some readout chips for pixel detectors currently in operation at the Large Hadron Collider.

The second chapter will describe the design and the characterization of the Superpix0 readout chip, the first prototype of a readout chip for fine-pitch hybrid pixel sensors to be used in the SuperB Layer0. The Superpix0 chip contains 4096 $50\ \mu\text{m} \times 50\ \mu\text{m}$ cells arranged in a 32x128 matrix and organized in MacroPixels (MP).

The third chapter starts by introducing the reader to the 130 nm CMOS Tezzaron/Globalfoundries vertical integration technology, chosen for the design of a second front-end chip for hybrid pixels, called Superpix1. The use of a 3D technology makes it possible to increase the chip functional density and to implement a more complex in-pixel logic with respect to Superpix0. The first experimental results from the test of a 3D DNW MAPS prototype in the same 3D CMOS technology are also presented, demonstrating the functionality of the vertical integration process.

Chapter 1

Mixed-signal integrated circuits for semiconductor radiation detectors

A radiation detection system may be used to measure the amount of energy released by a charged particle or a photon while passing through the sensor volume, the position (in one or two dimensions) of a particle passing through the surface of the detector and/or the time of arrival of the particle. In particular, semiconductor microstrip and pixel detectors are used to measure the position of a particle hitting the detector surface. In this kind of detection systems, the signal from a capacitive detector, like the ones mentioned above, is first read-out by an analog processing channel, which in its optimum version includes a charge preamplifier and a shaper for signal-to-noise ratio maximization. A discriminator is generally used to compare the signal at the shaper output to a preset threshold voltage, therefore providing information about the presence of a significant event, called hit. The digital hit signal at the discriminator output must be further processed by circuitry in the pixel or at the chip periphery. Processing may just involve reading out the single hit/no hit bit of information or using it to perform more complex operations. Choice of the set of operations to be performed on the data collected in a detector before sending them out depends on the target application. Also, the design approach and the available technology may impact on the design of the readout chip. Evolution of CMOS technologies and, in general, of microelectronic processes, may help improve performance and include more functionalities in the readout circuits. Also, higher degree of radiation hardness of modern CMOS process makes them the ideal candidate for the readout of segmented semiconductor

detectors in present and next to come high luminosity colliders.

This chapter introduces the fundamental operating principles of microstrip detectors, hybrid pixel detectors and monolithic active pixel sensors (MAPS), on which the activity of the community of microelectronic designers in the field of radiation detection is mostly focused. The second part of the chapter will be devoted to the description of the main features of some classical front-end (FE) schemes and typical readout architectures. This will be followed, to provide some example of systems currently in operation, by the description of the silicon vertex tracker, based on pixel detectors, of two LHC (Large Hadron Collider) experiments, ATLAS and CMS.

1.1 Detectors for particle physics applications

The notion of pixel has been introduced in image processing to describe the smallest discernable element in a given process or device. A pixel detector is therefore a device able to detect an image and the size of the pixel defines the granularity of the image. Digital cameras are a typical example of pixel detectors, where photons of different energies are integrated in the sensing elements, during short exposure times and generate an image as an intensity distribution. The images or patterns considered in this work are not generated by visible light, but by relativistic charged particles or photons in the keV to MeV energy range. The charge generated by ionizing radiation is transformed into images through dedicated electronic circuits. The main characteristics of this kind of devices are high speed, good time resolution and the ability to select hit patterns. In general they are developed for the specific needs of a single particle physics experiment. Pixel detectors for high energy physics (HEP) applications have the purpose of studying short-lived particles, emerging from the collision of other particles. In modern particle accelerators, rates and energies are dramatically increased, with respect to previous facilities. Therefore, detectors have to detect particles emerging from every collision at a rate which can be of the order of a few tens of MHz. As already mentioned, in a HEP experiment, some particles emerge from the interaction point, also called collision vertex (CV). Some rare, but scientifically interesting, particle lives for few picoseconds and then decay into daughter particles. The particle physics requirements are satisfied by a detector with a high granularity able to detect multiple tracks with good space and time resolution. Moreover, its electronics should be capable of selecting the interesting patterns, which are very rare. To do this the readout electronics should be designed to temporarily store the hit patterns, belonging to an individual event which is possibly interesting on

the basis of its topological and dynamical variables. These variables, derived from the event itself, are digitized and then used in a combinatorial circuit whose output will eventually provide the events of interest. In a typical HEP application, the information is not uniformly distributed in all hit patterns, but concentrated in some rare patterns that have to be identified with appropriate readout architectures. In other applications, like X-ray radiography, the information is uniformly distributed on all events, whose integration will result in a meaningful image.

1.1.1 Microstrip detectors

Single-sided microstrip detectors [?] [1] are semiconductor detectors in which one electrode is segmented in thin parallel strips. Ion implantation and photolithographic techniques are used to selectively dope the surfaces of the semiconductor wafer of typically 300 μm thickness and to deposit the metallization patterns necessary to extract the signals. This kind of technique is derived from the standard processing used in microelectronics and therefore profits from the large investments and the high quality standard of the integrated circuit industry. Fig. 1.1 shows a schematic view of a silicon single sided microstrip detector in which each strip is a p-implant on a n-type substrate and acts as a pn-junction (diode). The interface region between the n-doped and the p-doped regions will be emptied of free charges through the following mechanism. The majority carriers in each region will diffuse through the junction and recombine with the opposite sign charge carriers. This will generate an electric field due to excess charge from the immobile doping atoms, which counterbalances the diffusion and established an equilibrium. This equilibrium, characterized by an absence of charges, which can move freely, extends to some thickness W (depletion zone). If diodes are reversely biased by applying a positive voltage on the n-side and connecting each p-implant to virtual ground through the input terminal of its readout amplifier, W varies according to the following equation,

$$W = \sqrt{2\varepsilon_0\varepsilon_{Si} \frac{V}{eN}}, \quad (1.1)$$

where ε_0 is the dielectric constant in vacuum, ε_{Si} is the silicon relative dielectric constant, e is the elementary charge, V is the reverse bias voltage, and N is the dopant density. Charges are built up on both sides of the junction and therefore the depletion zone can be seen as a capacitor of value C_A per unit area:

$$C_A = \frac{\varepsilon_0\varepsilon_{Si}}{W} = \sqrt{\frac{e\varepsilon_0\varepsilon_{Si}N}{2V}}. \quad (1.2)$$

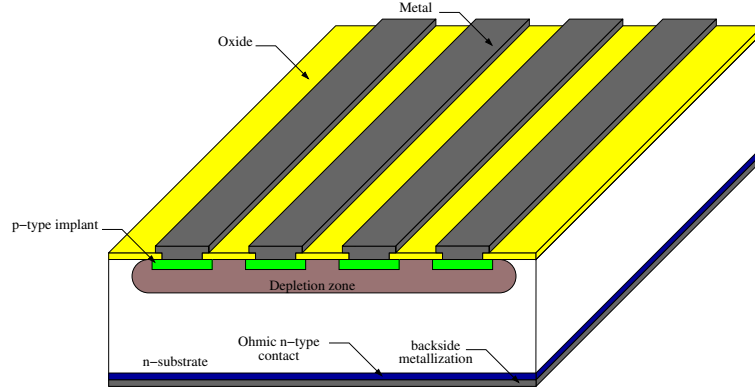


Figure 1.1: schematic representation of a single sided microstrip detector.

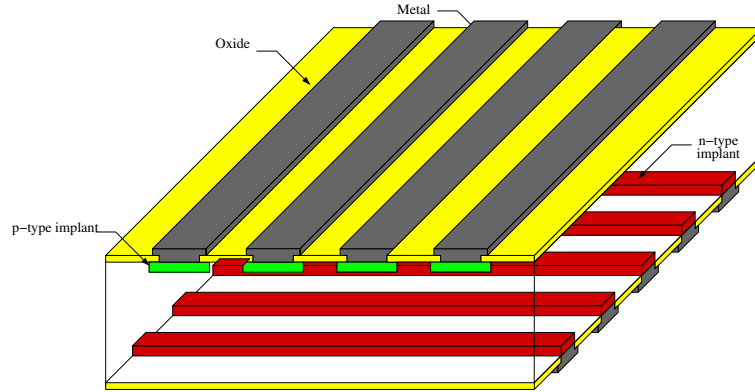


Figure 1.2: schematic representation of a double sided microstrip detector.

Increasing the reverse bias voltage increases the thickness of the depletion zone and reduces the capacitance of the sensing element and both these effects enhance the Signal-to-Noise Ratio (SNR). The best SNR is given by detectors with the depletion zone extending to the whole thickness of the silicon bulk (fully depleted detectors). Particles crossing the silicon detector, or photons absorbed in it, generate on average 1 electron-hole pair per 3.6 eV of energy deposited [2]. If the charged carriers are generated in the depletion zone, the active volume of the detector, they lead to a current signal. In the case of a relativistic particle, the energy lost through many collisions with the electrons of the crystal generates about 80 electron-hole pairs per micrometer of path. The charges drift under the action of the external electric field at a speed that

depends on the electric field, but saturates at a value of 10^7 cm/s for fields close to 10^4 V/cm. Therefore, in the case of a typical detector thickness of $300\text{ }\mu\text{m}$, the charge is collected in less than 10 ns. During the drift, the charges do not exactly follow the electric field lines, but diffuse as a consequence of the random thermal motion in the crystal lattice. Spread of the arrival position of the charge due to this effect can be described as a gaussian distribution with standard deviation

$$\sigma = \sqrt{2Dt}, \quad (1.3)$$

where D is the electron diffusion constant (typical value is $35\text{ cm}^2/\text{s}$) and t is the transit time of the carriers ($t \approx 10$ ns). As electrons and holes in a strip detector are swept by an electric field to opposite sides of the wafer, it is possible to use both types of charge carriers for position measurement by providing charge collection electrode on both sides of the wafer. This double sided read-out brings about the obvious advantage of providing twice the information for the same amount of scattering material. With crossed strips on the two detector faces, projective two dimensional measurement is obtained from one single detector.

1.1.2 Hybrid pixel detectors

The fabrication of a pixel sensor is very similar to the fabrication of a microstrip sensor. In the pixel case, the implants have a higher segmentation, because every microstrip diode would be further subdivided along its length. The charge collection mechanism is the same as the one that described in

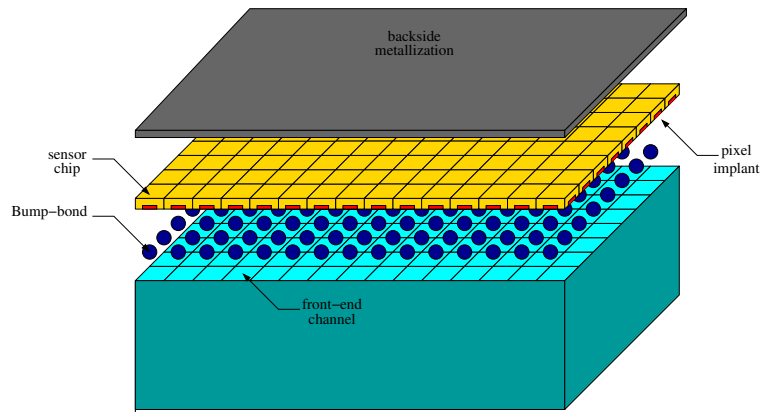


Figure 1.3: breakdown structure of a hybrid pixel detector.

the case of the single sided microstrip detectors. In a hybrid pixel detector, electronics and sensors are fabricated separately and then mated (Fig. 1.3). The detector consists of a matrix of diodes and a geometrically matching array of electronics circuits. The connection between the two chips is done by the so called “flip-chip” bonding technique. The chips are mounted face to face and tiny bumps of indium, gold, solder or conducting glue provide the electrical connection between the detector and its electronics. Depending on the application, the electronics has to provide several functions for each individual cell and may include: charge signal amplification, noise filtering and signal storage for later readout, circuits for pixel coordinate and time stamp storage and reset devices for the restoration of the pixel electronics after the readout. These functions have to be implemented in an area which is equal to the pixel size. An additional space is needed for readout control and signal busses. Some characteristics of the pixel detectors are related to the small dimension of the sensing elements. Each pixel covers, in fact, a very small area over a thin (about $300\text{ }\mu\text{m}$) layer of silicon. It therefore exhibits a very low capacitance which is dominated by the coupling to the neighboring pixels rather than to the backside plane. The direct interpixel coupling has to be kept to a minimum with proper sensor design to avoid crosstalk between pixels. The low capacitance is one of the key advantage of hybrid pixel detectors since it allows fast signal shaping with very low noise. An hybrid detector is compatible with work in a hostile environment existing close to the interaction region of a particle accelerator because, generally, it is radiation hard and can survive high integral particle flux. Freedom in the choice of the sensitive material is an advantage in application of hybrid pixel detectors also in other fields, like medical diagnostics.

1.1.3 Monolithic pixel detectors

The idea of a monolithic pixel, which means a device with both electronics and sensor in the same substrate and fabricated with the same technology, is based on a compromise between the sensor and the readout functionality. Since silicon is the material most commonly used in microelectronics, a monolithic detector is more robust and less expensive than a hybrid pixel and, at the same time, makes it possible to avoid expensive high density interconnection techniques and the related manipulations. They offer the possibility of very low input capacitance and hence very high signal-to-noise ratio. Commercial CMOS technologies use low resistivity silicon which is not suited for charge collection. However, an epitaxial layer of a few tens of micrometers can keep the charge by potential wells at the boundary, allowing them to reach an n-well

collection diode by thermal diffusion. The signal charge is very small (less than 1000 electrons) and its time development is slower (about 100 ns) than in detectors with highly resistive depleted bulk. Fig. 1.4 shows the cross-section of a monolithic active pixel sensor (MAPS), together with some transistor level detail. Only NMOS transistors are allowed in the active area in order not to degrade the charge collection properties of the device. Actually, the n-wells needed for the integration of PMOS transistors would steal charge from the collecting electrode. Therefore, pixel readout is performed using a standard three-transistor circuit (line select, source follower stage, reset).

Monolithic pixel detectors are being considered for application to HEP experiment, as they have the potential to comply with the specifications of the experiments at the future high luminosity colliders, like the International Linear Collider (ILC) and the SuperB Factory. The deep n-well MAPS (DNW MAPS), recently proposed for HEP applications, is based on the same working principle as standard MAPS, where minority carriers generated by charged particles in a p-type, lightly doped epitaxial layer diffuse and are collected by

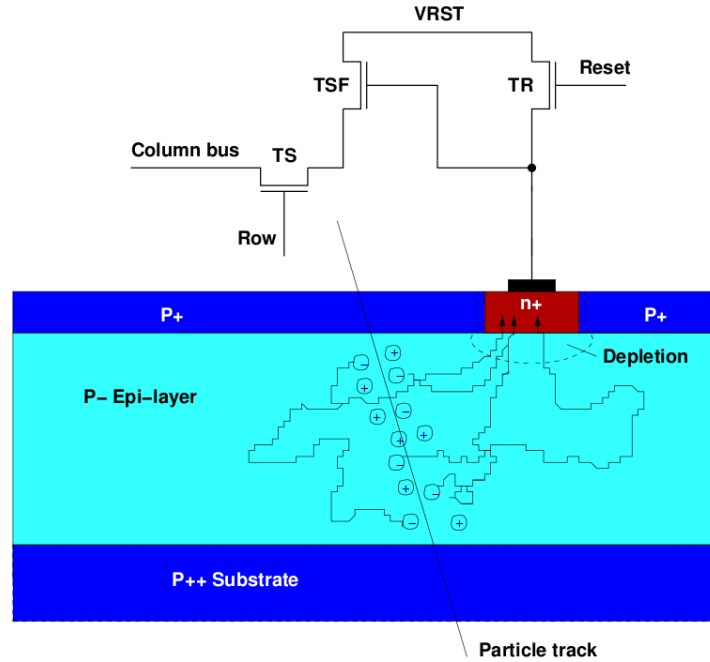


Figure 1.4: cross-section of a monolithic active pixel detector, with some transistor level detail of a typical 3-NMOS front-end channel.

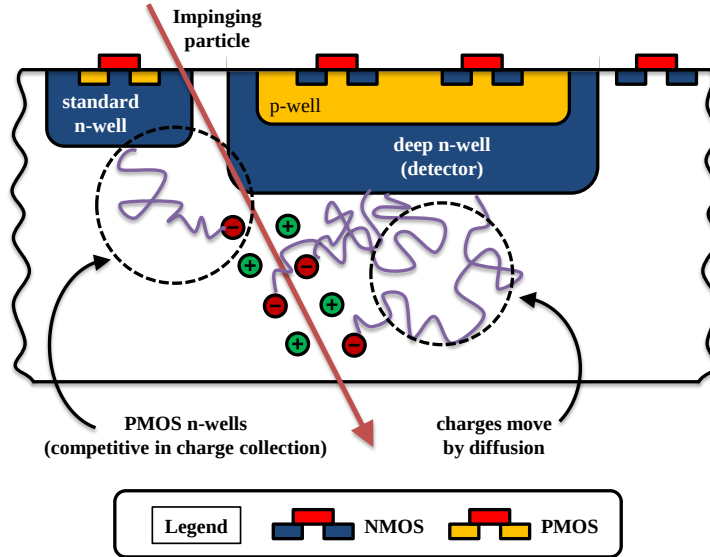


Figure 1.5: cross-section of a DNW MAPS detector.

n-type electrodes. The DNW MAPS sensor, whose structure is illustrated in Fig. 1.5, differs from standard MAPS in two main characteristics: an n-well with a deep junction is used to collect the charge released in the substrate and a classical readout chain for capacitive detectors is used to process the charge signal. The two design choices are closely related to each other. The deep n-well, which in modern, triple-well CMOS technologies is used to shield NMOS devices from substrate coupled noise in mixed signal circuits, may house N-type devices, therefore mitigating the constraints set by the readout circuits on the sensor area and geometry. Moreover, if a large area DNW sensor, as compared to standard MAPS, is laid out, PMOS devices may be used in the elementary cell at the expense of a certain amount of charge collection inefficiency depending on the ratio of the DNW area to that of all of the N-type wells, both deep and standard. Use of a large collecting electrode would impair the properties of classical three-transistor (3T) readout schemes (see Fig. 1.4), as the increased capacitance would unacceptably degrade the noise performance and the charge sensitivity at the same time. In such sensors, in fact, the foremost contribution to noise arises from the reset operation, so that the equivalent noise charge grows proportionally to the square root of the collecting junction capacitance. Processing the signal from the deep n-well sensor by means of a charge sensitive amplifier decouples the charge sensitivity from

the sensor capacitance and, therefore, from its area. Furthermore, the large scale of integration of deep submicron CMOS technologies in the 100 nm scale and the area saved by integrating NMOS devices inside the DNW sensor can be exploited to add both analog functions, such as signal shaping, and digital functions to the pixel level processing electronics [3].

1.2 Architectures for analog readout

1.2.1 The charge sensitive amplifier

The standard problem in the readout of a semiconductor detector is the low noise measurement of the charge signal, usually under severe constraints such as the requirement for high speed operation and low power consumption, limited space and high radiation levels. The charge sensitive amplifier (CSA) consists of an inverting amplifying circuit which, in the ideal case, delivers an output voltage inversely proportional to a feedback capacitor C_F and proportional to the input charge: In addition a high resistance is needed in the feedback loop, in order to bring the circuit into a stable operating condition. In Fig. 1.6, C_D represents the detector capacitance and C_{in} the capacitive load at the preamplifier input, usually dominated by the gate capacitance of the input transistor, and C_F is the feedback capacitor. A Q_{in} charge at the CSA input will result in an output voltage change of:

$$V_{out}(s) = \frac{Q_{in}}{s \left(C_F + \frac{C_D + C_{in} + C_F}{A} \right)}, \quad (1.4)$$

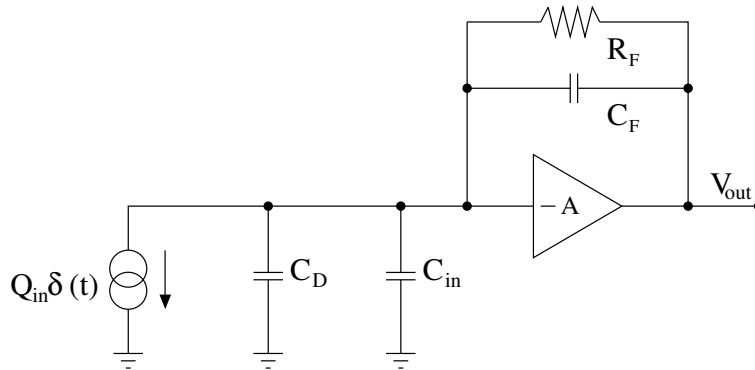


Figure 1.6: block diagram of a charge sensitive amplifier.

where $R_F \rightarrow \infty$, $V_{out}(s)$ is the voltage at the circuit output in the Laplace domain, A is the amplification of the stage, which is assumed to have infinite bandwidth. For large amplification ($A \rightarrow \infty$), V_{out} is given by

$$V_{out}(s) = \frac{Q_{in}}{sC_F} \quad (1.5)$$

A further important consideration in applying a CSA circuit to the readout of a detector concerns the noise properties of the amplifier. The contribution of the amplifier to the overall system noise is due to the noise generated by the electronic components of the circuit. In a properly designed amplifier, the noise contribution is dominated by the noise generated in the input transistor. The preamplifier is one of the most crucial parts in the readout channel. It must provide high gain with sufficient bandwidth. Its power consumption must be kept very low in most applications in order to limit the heat dissipated in the active area. Another design goal is a good immunity of the amplifier to fluctuations in the supply voltage, which could be generated by changes in chip activity and cross coupling from the digital to the analog part. A high power supply rejection ratio in the frequency range of interest is therefore desirable.

1.2.2 The shaper

The signal produced by the amplifier is further amplified and shaped with the aim to optimize the signal to noise ratio and to minimize the overlap between subsequent signals. Fig. 1.7 shows a simple shaper circuit scheme, represented by a RC-CR filter (an integration followed by a differentiation). For an input voltage step provided by the CSA with Q_{in}/C_F amplitude, the signal at the shaper output, if the time constant is the same for the integrating and the differentiating stage ($\tau = R_1C_1 = R_2C_2$), is given by

$$v_{out}(t) = \frac{Q_{in}}{C_F} \frac{t}{\tau} e^{-\frac{t}{\tau}}. \quad (1.6)$$

The peak value is proportional to the charge signal

$$V_{peak} = v_{out}(\tau) = \frac{Q_{in}}{C_F} e^{-1}. \quad (1.7)$$

The noise voltage is superimposed on the signal. A parameter for defining the quality of the analog readout channel is the signal-to-noise ratio (SNR), defined as the ratio of the peak value of the signal to the root mean square value of the noise voltage measured at the same point in the circuit. In order

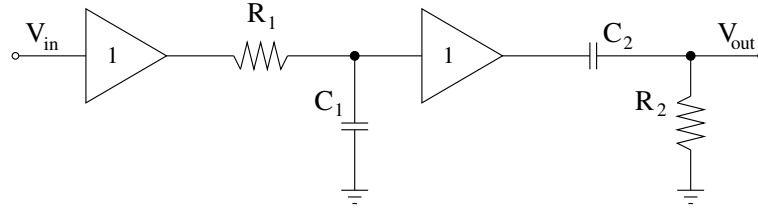


Figure 1.7: RC-CR shaping stage.

to find the noise at the output, each noise source in the circuit has to be traced to the output and the resulting voltages added in quadrature. This procedure will be treated in more details in chapter 2. More sophisticated continuous time filtering methods exist, like for example, Gaussian shape filtering, which can be approximated by several RC integration and differentiation steps in sequence.

1.2.3 Front-end readout

Once a hit has been detected, amplitude analog information at the shaper output can be discarded, if considered as the result of a non interesting event, or retrieved in one of four possible ways:

- **Pure binary readout:** analog information is discarded, just hit/no hit information obtained by comparing the analog signal to a preset threshold through a discriminator, is provided by the readout chip for each channel (Fig. 1.8(a));
- **Time over Threshold (ToT):** amplitude is converted to a time duration by comparing the shaper output to a preset threshold; numerical conversion can be obtained through direct digitalization methods (Fig. 1.8(b));
- **Peak & hold:** peak voltage at the shaper output is sampled and transferred as an analog piece of information, to the chip periphery (or to the data acquisition system) where it is converted to a number (Fig. 1.8(c));
- **In-pixel A/D conversion:** peak voltage at the shaper output is sampled, locally converted to a digital word and transferred to the chip periphery (Fig. 1.8(d)).

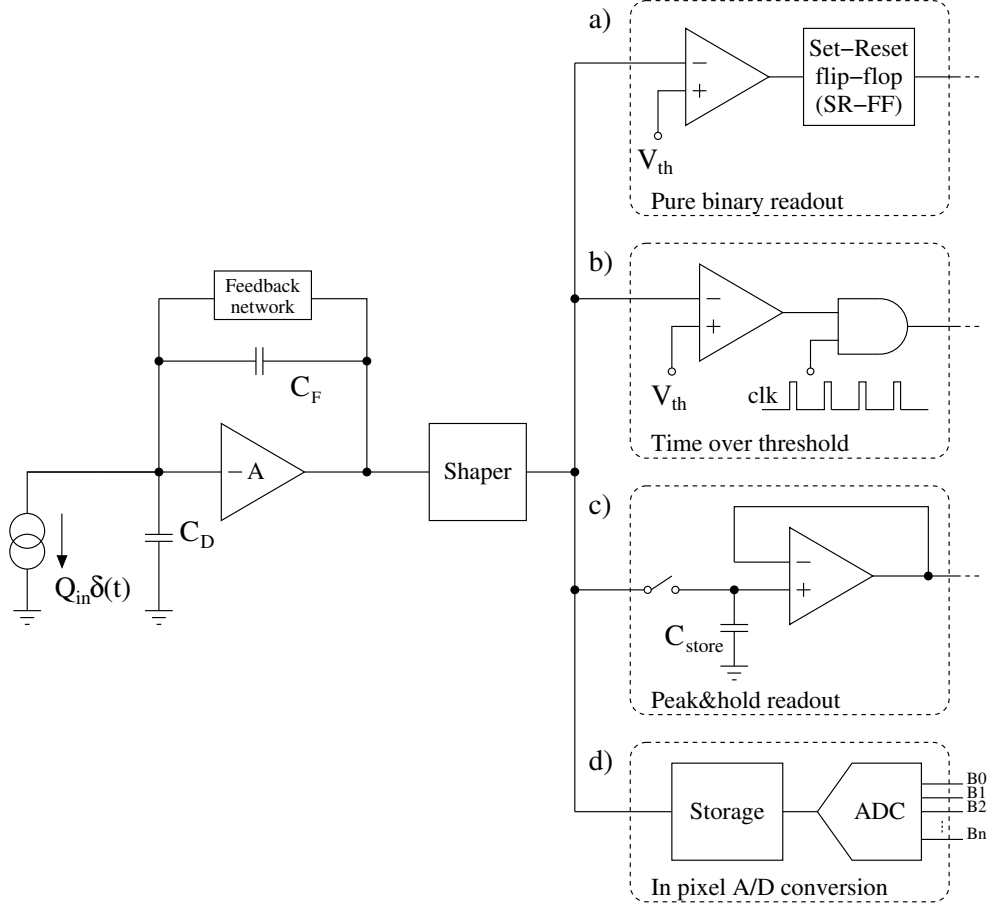


Figure 1.8: architectures for analog readout.

1.3 Readout architectures

A systematic classification of readout architectures is quite hard, since several different solutions have been proposed to face the challenges set by different experiments and applications. However, some high level design guidelines may be emphasized. The choice of a suited architecture mainly depends on the target application, on the available technology and on the acceptable hit losses, which may significantly vary for different readout concepts. Detailed simulation of the hit losses are therefore required before a choice can be made. An important decision to be made is whether the analog pulse height information of every hit is required, as some architectures are not suited for analog readout.

A first important distinction can be made between chips with or without data buffering. The latter case is used when the hit rate is relatively low, or in (still or slow) imaging applications, and every event can be read out. Sometimes the data rate is so high that some technique has to be implemented to discard not interesting events, in which case data buffering is needed. A variety of architectures have been proposed for the readout. In virtually all of them sparse readout, namely the skipping over empty pixels, is foreseen. Most of them are “column based structures”, where each column of pixels is treated separately and the result of the readout stored in an end-of-column buffer that can be read out asynchronously to the data capture. The advantage of the column structure with respect to an “x-y” structure is in the tolerance of malfunctions: an error in one pixel will usually affect only one column rather than the whole device. In this section, a few examples of readout architectures, used in different applications, will be presented and discussed.

1.3.1 Chips without data buffering

The first pixel readout chips were designed for the relatively low interaction rates of the LEP accelerator so that every event could be read out. The pixel detector readout chip featured an X-Y scanning scheme [4] represented in Fig. 1.9, implemented in the DELPHI experiment, to find the hit pixel in the matrix. In each pixel, the flip-flop (FF) is set by the discriminator signal when an event is detected. Horizontal stop lines are pulled high if a hit is present in the row. An asynchronous vertical scan is initiated by injecting a start-row-scan token into a scan unit chain. The token propagates through the scan units until a stop signal is encountered in the first row with hits. The stopped output signal from the scan unit is used to generate the row address and to select the active row for the secondary column scan. This scan stops at the first hit column so that the X-Y coordinate of the first hit is obtained. The hit flip-flop are successively reset so that the scan skips from hit pixel to hit pixel.

1.3.2 Chips with zero suppression and data buffering

A large family of pixel readout chips is used for tracking in high energy physics where accelerated particles collide with a fixed target or with other particles from a beam running in the opposite direction. Collisions occur at intervals of 25 ns in the case of the LHC collider at CERN and the bunch crossing clock (BC) can be used to synchronize data taking. Most events do not contain interesting information because the interesting physics processes are very rare.

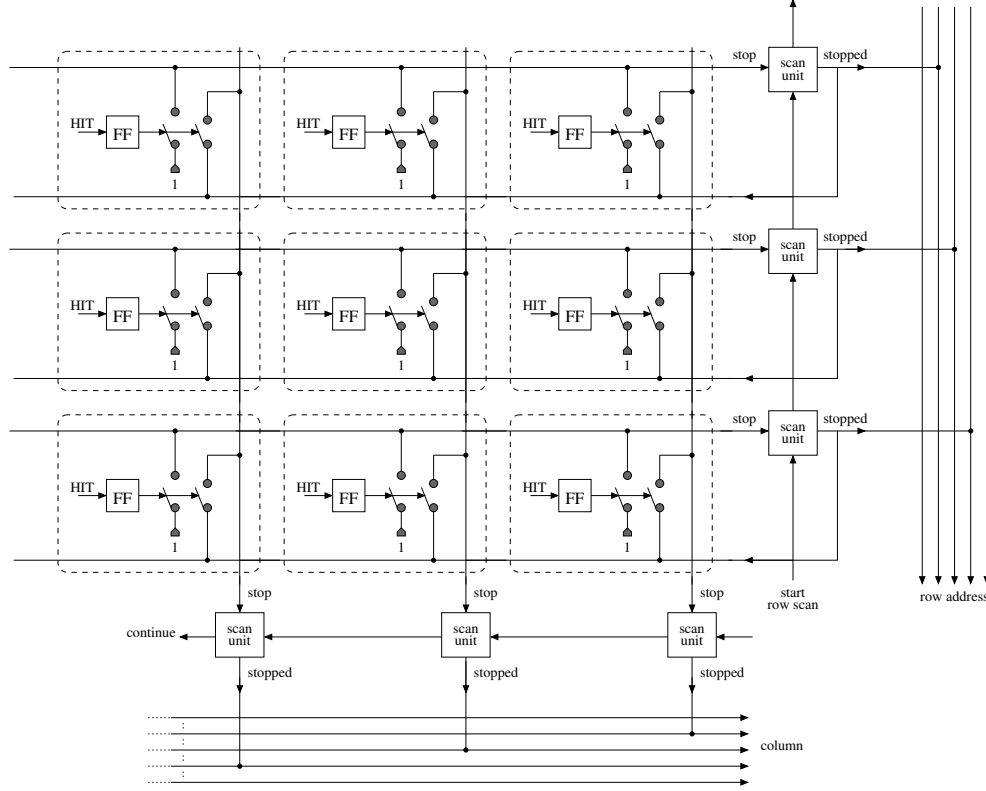


Figure 1.9: the readout architecture of the DELPHI pixel chip.

A selection of potentially interesting events is therefore made by other detectors of the experiment. This trigger decision requires several microseconds of processing time and the trigger signal arrives at the pixel chip with a delay of more than 100 bunch crossing clock cycles. The hit information of many events must therefore be stored on the pixel chip during this fixed latency interval. Only triggered hit information is sent to the data acquisition system of the experiment. All other hit information is discarded after the latency. Although the trigger rate is low (less than 1% of the events), several nearly consecutive readout requests may occur. The chips must therefore be capable of accepting new triggers before the data of a previous event have been completely sent out. Several solutions have been proposed and implemented, some of which are discussed in the following.

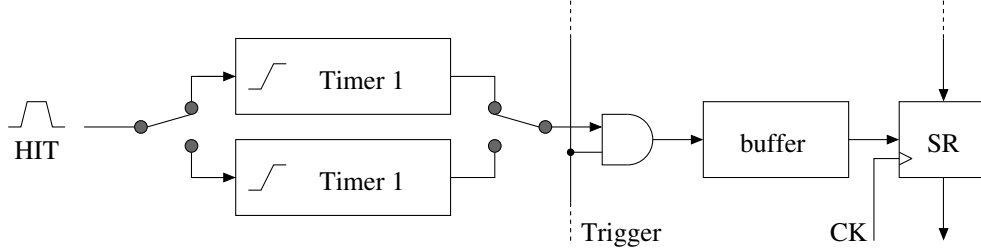


Figure 1.10: readout by using in-pixel timers.

In-pixel storage

One possible approach is to store the information in the pixel by starting a local timer with the same duration as the trigger latency (Fig. 1.10). A hit belongs to the bunch crossing of interest when the falling edge of the timer coincides with the trigger signal which is sent to all pixels. The readout of the valid hit is achieved with a shift register in the OMEGA chip family [5]. Two timers can be used and activated alternately to reduce the pixel dead time, during which no hit can be processed. Multiple hit flags can be buffered in the pixel to reduce dead time during the readout and to accommodate several closely spaced triggers.

Conveyor belt architecture

Another readout concept uses a vertically running shift register (6-bit wide with a column of 63 pixel high in the ATLAS FEA chip [6]) in every column to clock the row number of a pixel down the column as soon as a rising edge of the pixel discriminator occurs (Fig. 1.11). A shift register value X (ID in Fig. 1.11) arriving at the bottom indicate that the pixel in row X has been hit and that the hit has occurred X clock cycles in the past. The hit position is stored in one of several buffers at the bottom of the column. The “age” of the hit is reduced by the trigger latency and the result is written to a counter (incremented with the system clock) in the same buffer unit. The total time since the hit has occurred equals the latency when the counter reaches zero. Trigger coincidence is made at this moment. A hit would be lost in this architecture in the event that the shift register cell is full at the moment when a new row ID has to be written. Several tries to write the ID is then permitted (the number of trials is taken into account in additional late bits). Falling edges of the discriminator can be used to determine the width of the

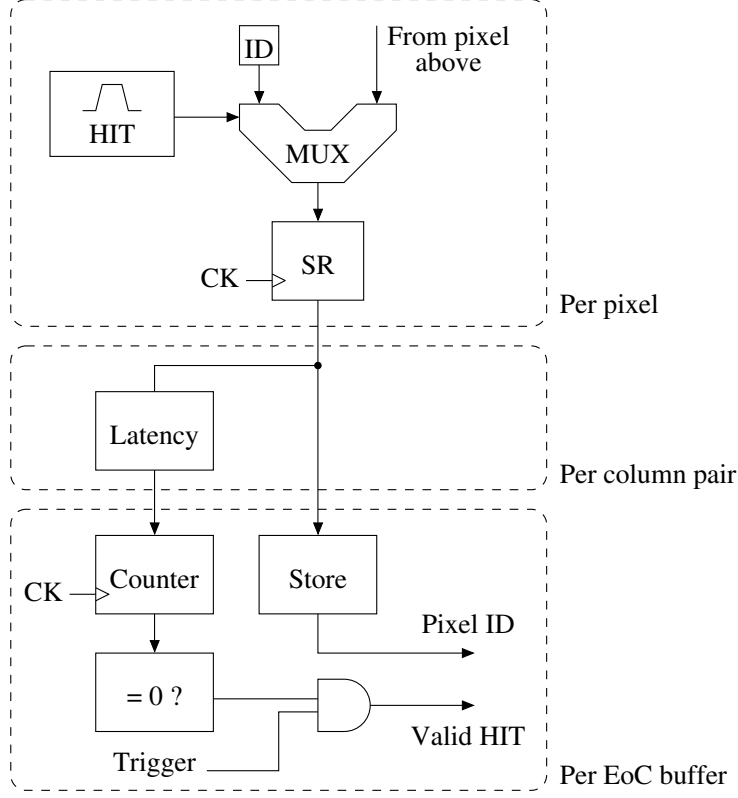


Figure 1.11: the conveyor belt architecture uses a digital shift register to transport the ID of a hit pixel (the X coordinate) to the bottom where it arrives exactly after X clock pulses. The trigger coincidence is performed after further clock pulses (corresponding to the pixel ID minus the latency) in a buffer in the chip periphery.

output pulse (ToT method).

Time stamp readout

This approach basically consists of recording the time at which a hit has occurred in digital form (time stamp). When the trigger signal selects a certain bunch crossing for readout, the time information of all accumulated hits is compared to the interesting time interval referred to by the trigger signal. Hits with the correct time stamp are read out, all older hits are rejected. The time of the falling edge of the discriminator output can also be stored in such

a way that the pulse width (ToT) can be determined digitally by calculating the difference of the two values (the falling edge time and the time stamp, corresponding to the rising edge time) The time stamp for a hit could be stored in the pixel and the trigger coincidence made there. This procedure would block the pixel for the entire latency time duration with a consequent, significant inefficiency. Time stamp values are therefore transferred to buffers at the bottom of the pixel columns as fast as possible in the architecture of the FEI chip used in the ATLAS experiment [7].

Column drain readout

In the column drain architecture, developed for the CMS experiment [8], all hits occurring within one clock cycle are sent to a buffer in the periphery. Fig. 1.12 shows the main features of the architecture. One or more hits produced by the pixel discriminators in a column pair are flagged to the end of column (EoC) by a FastOR signal. A single time stamp for all of these hits is stored in one of the available digital buffer locations. The hit pixels in a

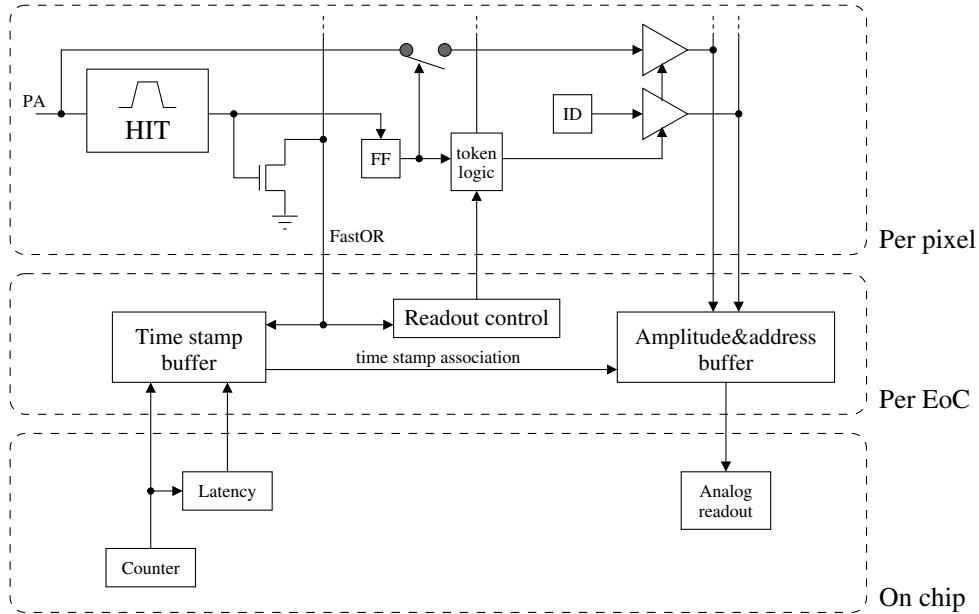


Figure 1.12: the column drain readout architecture transfers the amplitude and the address of hit buffers to the end of column (EoC) section, where the information is associated to a single digital buffer holding the time stamp of the event.

column pair are found by a fast scan mechanism passing a token from cell to cell through closed switches until a hit cell is found. Pixels with no hit are still sensitive during the scan. The analog information is stored on capacitors in the pixels and is sent out by the pixel which has been identified through the token scan. Then, all hit information is sent sequentially to a set of buffers (amplitude&address buffer in Fig. 1.12), each containing an analog cell for the amplitude information and digital cells for pixel address, in such a way to be associated with the corresponding time stamp (TS) of the digital buffer. Time stamp in the digital buffers are permanently compared with a second TS value offset by the latency. If the two values match and no trigger is present, the event is discarded.

Self-triggered readout architecture

Some pixel readout chips can be operated either in an externally triggered or a self-triggered mode. In the FPIX chip [9], four identical EoC readout controllers at the bottom of the columns communicate with the pixels in the column through 4x2 command lines. The controllers can issue one of four states: “look for data”, “idle”, “output”, “reset”. Readout operation can be divided into several steps.

- One of the EoC controllers is requested by a priority encoder to send a “look for data” command to the pixels; this controller will be responsible for processing the next event in the column.
- When a pixel is hit, the pixel is linked to the command set carrying the “look for data” pattern. The hit FastOR line is activated, informing the EoC that a pixel has been hit. The active EoC controller stores the time stamp and switches its state from “look for data” to “idle” and another free EoC controller is selected to issue a “look for data” command.
- All “idle” EoC controllers and the associated pixels wait until their readout is requested: in the triggered operation mode, a desired event TS is presented and compared to the TS stored in the EoC controller; readout is started if the values are the same, the event is discarded otherwise; in self-triggered mode, all events are read out.
- When a readout is requested, at least one controller issues an “output” command to its associated pixel, which pulls the read FastOR line low. The EoC bus controller starts a token scan; the token stops at the first pixel requiring it. The first pixel found outputs its address and the A/D converted amplitude onto the column bus and reset itself.

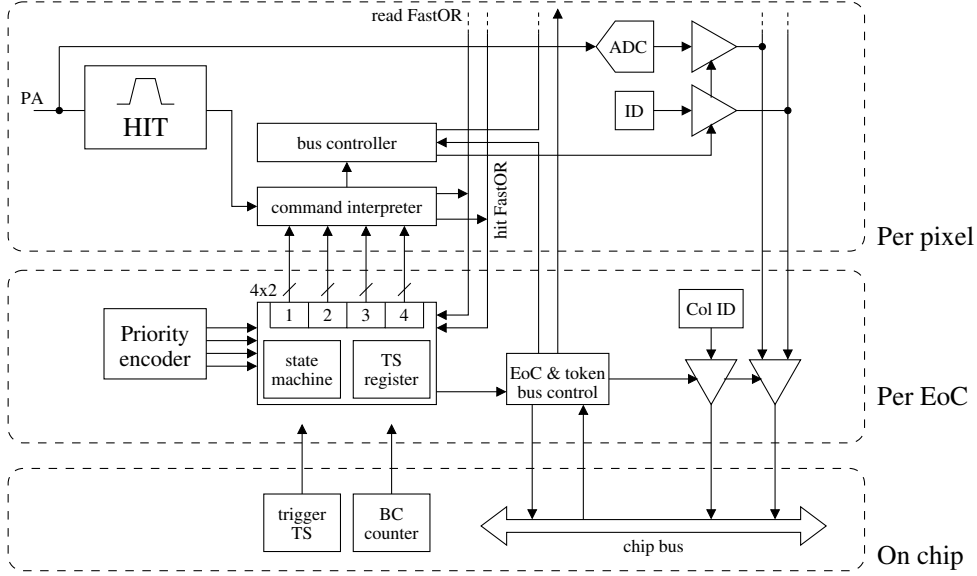


Figure 1.13: The self triggered architecture implemented in the FPIX chip family uses several readout controllers at the end of the column. All hits within a column occurring at the same bunch crossing (BC) are associated with one of the four available controllers where the time stamp (TS) is stored.

- All pixels have been found when the read FastOR goes high. Then, the EoC controller is ready to issue a new “look for data” and send a “reset” signal.

The self-triggered readout architecture is implemented in such a way that all hits can be read out immediately (for relatively low rates) for early trigger decisions.

1.3.3 Counting chips

For applications in biomedical imaging, synchrotron radiation experiments and autoradiography, the number of particles absorbed in every pixel during a given time interval must be determined. The hit signals are therefore counted in every pixel and read out after the measurement interval. Practical implementation requires very compact counters of 15 bits to cope with the hit rate in brightly illuminated pixels. Classical binary counters are space consuming and require a dedicated readout. A particularly simple design, represented in Fig. 1.14, is a linear shift register fed back with an EXOR gate from two or

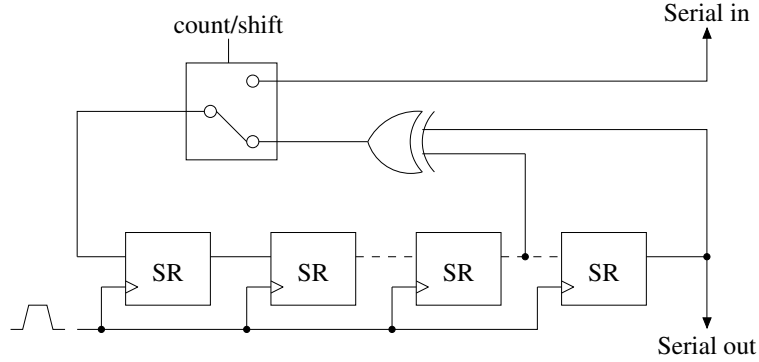


Figure 1.14: a counter implemented as a linear feedback shift register. The solution shown here includes a simple serial readout.

more taps (if taps are chosen correctly, $2^N - 1$ states can be achieved). The bit patterns in this case are pseudorandom numbers, not easily translated to the counted hit number (lookup tables may be needed). The shift counter has the additional advantage of a simple serial readout. Dynamic range limitations can be overcome by adding an overflow bit, which is set when the counter wraps around. External circuitry regularly scans the overflow bits and increments external counters. Virtually infinite dynamic range can be achieved if all the overflow bits are recorded before the counter wraps around for the second time.

1.4 Examples of pixel detector applications

A readout chip for semiconductor detectors includes both analog and digital blocks. It contains a section where a cell is periodically replicated based on the detector chip geometry, and a common (completely digital) section servicing all the cells. There is a trend to reduce the pitch of semiconductor detectors to improve resolution. This has an impact on the readout electronics, which has to be designed in such a way to fit into the interelectrode spacing. The analog front-end performs the task of amplifying and suitably shaping the charge signal in order to maximize the signal to noise ratio. Digital blocks may perform several tasks, like data selection (sparsification), zero suppression, hit counting, analog-to-digital conversion, time stamping, data storing, buffering and serialization (or parallelization). A readout chip has to provide digital information in a form that requires the least readout bandwidth and processing

time possible before being stored in a memory. Therefore, as many functions as possible are moved from the acquisition system to the chip itself. In order to provide a couple of paradigmatic examples of detectors currently in use, in this section the pixel detector systems used in the ATLAS and CMS experiments at the Large Hadron Collider (LHC).

1.4.1 The ATLAS pixel detector

The pixel detector system for the ATLAS experiment (A Toroidal LHC ApparatuS) [10] is a general purpose detector for the study of proton-proton collisions at the LHC. The pixel detector contains approximately 80 million channels and provides pattern recognition capability in order to meet the track reconstruction requirements of the ATLAS experiment at the full luminosity of the LHC of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. The performance requirements for the ATLAS Inner Detector (ID) were formulated in the Inner Detector Technical Design Report (TDR) [11][12]. The general performance requirements for the pixel system are: a resolution better than about $15 \mu\text{m}$, allowing vertex reconstruction of charged tracks, minimal material budget for all elements in the system, in order to reduce multiple scattering and secondary interactions, excellent efficiency for all pixel layers and radiation hardness of the pixel detectors elements to operate after a total dose of 500 kGy or about $10^{15} \text{ n}_{eq} \cdot \text{cm}^2$ (lifetime dose). These performance requirements lead to the following design choice: minimal radius of the innermost layer set at 5 cm and the smallest pixel size at $50 \mu\text{m} \times 400 \mu\text{m}$. The dose for the innermost layer is expected to reach 500 kGy after approximately five years of LHC operation. The active region of the pixel detector is shown in a schematic view in Fig. 1.15. The active part of

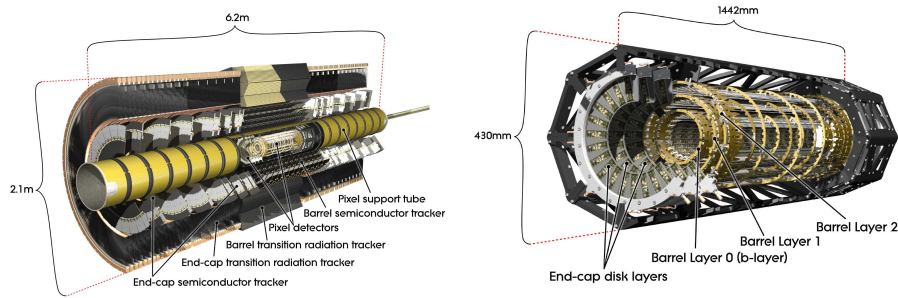


Figure 1.15: the ATLAS Inner Detector (left) and a schematic view of the active region of the pixel detector consisting of barrel and endcap layers (right).

the pixel system consists of three barrel layers (Layer 0, so-called b-layer, Layer 1 and Layer 2) and two identical endcap regions, each with three disk layers. The ATLAS pixel sensor is an array of detectors placed on a high resistivity n-type bulk close to the intrinsic charge concentration. The sensor is made by implanting high positive (p^+) and negative (n^+) dose regions on each side of a wafer. An asymmetric depletion region at the p^+ -n junction is operated in reverse bias and extends over the whole sensor bulk volume. Here, one is able to collect and detect charge carriers generated by ionizing particles passing through the active volume. The sensor design guarantees single pixel isolation, minimizes leakage current and makes the sensor testable as well as tolerant to radiation damage. The readout chip for the ATLAS pixel detector [13] [14], shown in Fig. 1.16, contains 2880 pixel cells of $50 \times 400 \mu\text{m}^2$ size arranged in an 18×160 matrix. Each pixel cell contains an analog block where the sensor charge signal is amplified and compared to a programmable threshold using a comparator. The digital readout part transfers the hit pixel address, a hit Leading Edge (LE) timestamp, and a Trailing Edge (TE) timestamp to the buffers at the chip periphery. In these buffers, a Time-over-Threshold (ToT) is calculated by subtracting the TE from the LE timestamp. These hit-buffers monitor the time of each stored hit by inspecting the LE time stamp. When a hit time becomes longer than the latency of the Level1 trigger (approximately $3.2 \mu\text{s}$) and no trigger signal is recorded, the hit information is deleted. Hits marked by trigger signals are selected for readout. Triggered hit data are then transmitted serially out of the chip in the same order as the trigger arrival. The FEI3 chip, the final version of the front-end for the ATLAS pixel detector, is implemented in a standard $0.25 \mu\text{m}$ CMOS technology. In order to obtain the required radiation tolerance, special layout rules have been used, i.e., all NMOS transistors have annular gates [15]. In the analog front-end, the charge-sensitive amplifier uses a single-ended, folded-cascode topology (Fig. 1.17), which is a common choice for low-voltage and high gain amplifiers. The amplifier is optimised for a nominal capacitive load of 400 fF and designed for the negative polarity signals expected from n-on-n-bulk detectors. The design of the charge amplifier was particularly influenced by requirements pertaining to sensor irradiation, which can produce leakage currents up to 100 nA . The preamplifier, which is operated at about $8 \mu\text{A}$ bias, has a 5 fF feedback capacitor with a current-source-continuous reset, and a 15 ns risetime. Since the input is DC-coupled, a compensation circuit is implemented that drains the leakage current and prevents it from influencing the continuous reset circuit. The implementation, shown in figure 1.17, uses two PMOS devices, one (M2) providing leakage current compensation and the other (M1) continuously re-

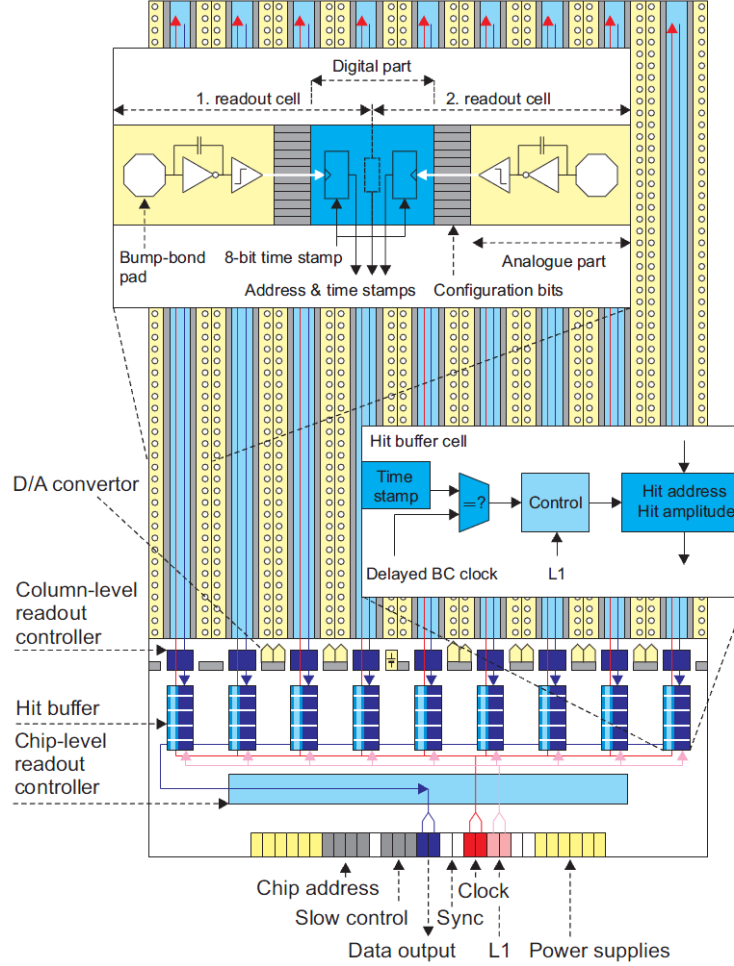


Figure 1.16: schematic plan of the front-end chip (FE-I3) with the main functional elements [13].

setting the feedback capacitor. An important property of this feedback circuit is that the discharge current provided by the reset device saturates for high-output-signal amplitudes. The return to baseline is, therefore, nearly linear and a discriminator pulse width proportional to the input charge is obtained. The width of the discriminator output, or Time-over-Threshold, can therefore be used to measure the signal amplitude. The duration of the ToT is measured by counting the cycles of the 40 MHz master chip clock. The feedback current

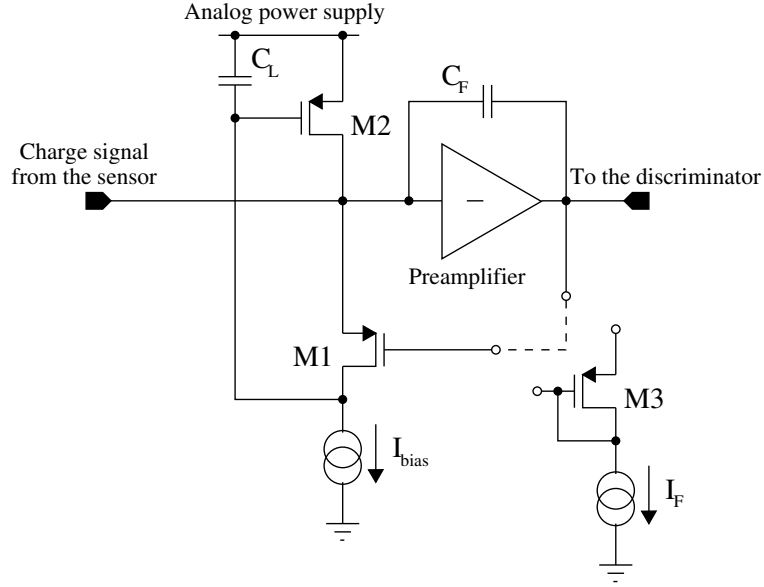


Figure 1.17: charge preamplifier of the ATLAS pixel detector readout channel, with emphasis on the feedback and leakage compensation circuit.

is 4 nA for a 1 μ s return to baseline in the case of a 20000 electron-equivalent input. The feedback circuit has an additional diode-connected transistor M3, which acts as a level shifter so that the DC-levels of the input and the output nodes are nearly equal. It also simplifies the DC-coupling between the amplifier and the discriminator, as described below. Signal discrimination is made by a two-stage circuit: a fully differential, low-gain amplifier, where the threshold control operates by modifying the input offset, and a DC-coupled, differential comparator. The first stage has a bias current of about 4 μ A, whereas the second uses a current of about 5 μ A. A local threshold generator is integrated in every pixel in order to make the threshold independent of the local digital supply voltage for each pixel and of the amplifier bias current I_{bias} . Seven-bits are used for each pixel to adjust the discriminator threshold. A complete block diagram of the analog part with several additional circuit blocks is shown in Fig. 1.18. Each pixel has several parameters that can be tuned through a 14 bit control register. These bits are:

- FDAC[0:2]: 3 bits to trim the feedback current for tuning the ToT response;

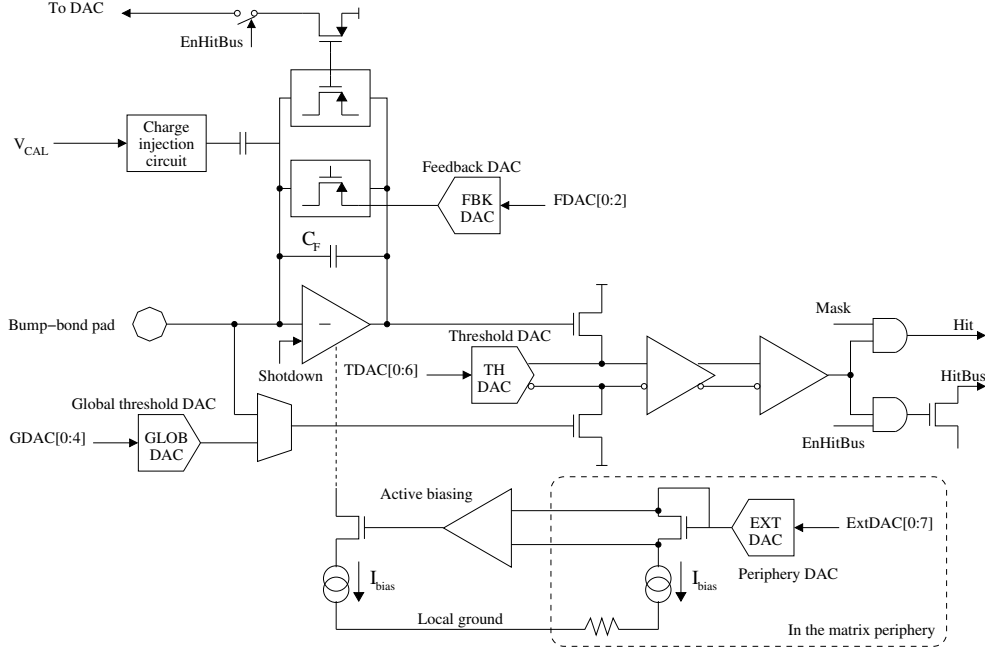


Figure 1.18: pixel cell block diagram.

- TDAC[0:6]: 7 bits to trim the threshold in each pixel;
- MASK: the digital output of the analog part can be switched off locally by setting this bit;
- EnHitBus: the digital outputs of all readout channels can be directly observed using a wired OR which is locally enabled with this bit. This bit also controls the summing of a current proportional to the feedback plus leakage current in the preamplifier, making it possible to monitor the feedback current and the leakage current from the sensor;
- Select: enables the pixel for test charge injection. The amplitude is generated from V_{CAL} (voltage proportional to the injected calibration charge);
- Shutdown: disables the charge amplifier so that no output is generated from the pixel.

The design requirements for the pixel front-end electronics come from operation at high radiation doses, from the time resolution of 25 ns to separate two

contiguous bunch crossings, from noise, from the minimum operation threshold and dispersion and from the overall power budget. The calibration relies on a 7 bit adjustment of individual pixel thresholds (tuning). The threshold dispersion is 800 electrons (e^-) and can be reduced to a 80 e^- within a calibration phase. The noise with the sensor attached is 160 e^- (for a pixel size of $50\text{ }\mu\text{m} \times 400\text{ }\mu\text{m}$) and the typical operating threshold is 4000 e^- , which results in hits with signals over 5500 e^- appearing in the correct 25 ns time bucket (described as in-time threshold). Neither the dispersion, nor the noise depend on the choice of threshold. The tuned thresholds have been observed to re-disperse with moderate radiation dose in prototypes, and periodic threshold re-tuning is needed. Measurements made on a few modules irradiated to 600 kGy show a negligible tuned threshold dispersion and a 20% increase in the noise, despite the very high induced sensor leakage current (the typical value is 60 nA at 7 Celsius degrees). For a configured chip, the typical digital current is 45 mA at 2 V and the analog current is 75 mA at 1.6 V for a total power of 220 mW.

1.4.2 The CMS pixel detector

The silicon pixel detector for CMS (Compact Muon Solenoid) has been designed to meet the requirements of position resolution, rate capability and radiation tolerance with a minimal amount of material set by the experiment. Its ability to provide three dimensional high precision space points plays an

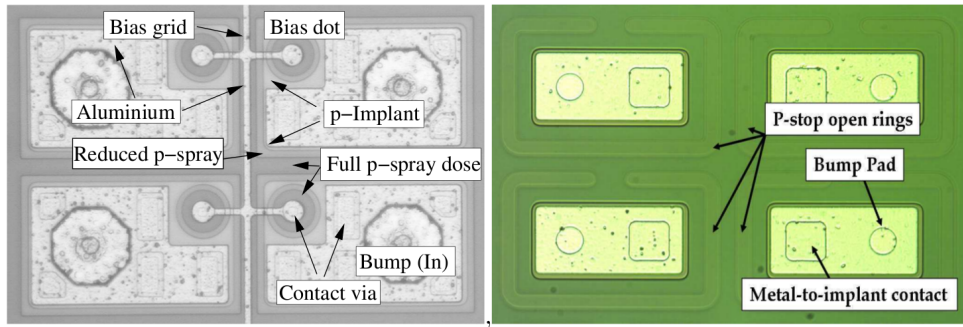


Figure 1.19: layout of the CMS pixel sensor for the forward (right) and barrel (left) detector. In order to isolate the pixels from each other, the barrel sensor uses p-spray isolation and the forward detectors use p-stop isolation. The pixel size is $100\text{ }\mu\text{m} \times 150\text{ }\mu\text{m}$ [16].

important role in the tracking system of the CMS detector. In the inner tracking detectors at the LHC, thousands of hits must be detected, time-stamped and stored every 25 ns, the same bunch crossing interval of the ATLAS experiment. Every cm^2 of a detector close to the beam pipe is expected to be hit by 10 million particles per second when the LHC runs at the full design luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Vertex detection of particles with sub-millimeter decay length requires precise track measurements close to the production point. The resolution required for CMS is on the order of $100 \mu\text{m}$. The extrapolation uncertainty is largely due to multiple scattering in the beam pipe and detector and it grows with the distance between the interaction point and the first measurement. More specific for the CMS pixel detector are the requirements of radiation hardness to at least $6 \times 10^{14} n_{eq}/\text{cm}^2$ and a readout architecture that handles a 40 MHz bunch crossing frequency with 20 simultaneous collisions. Only a small fraction of the bunch crossings will be read out, but the latency of the trigger is more than $3 \mu\text{s}$. Trigger rates up to 100 kHz are foreseen. The sensors for the forward and barrel detectors have a thickness of about $300 \mu\text{m}$ and adopt the n-in-n concept, where the pixels are formed by high dose n-implants introduced into a highly resistive n-substrate (Fig. 1.19) [16]. The junction is formed with a p-implant on the back-side. The front-end chip, whose size is coupled to the size of the detector module, serves the purpose of registering the signals produced by particles in the sensor, storing time, position, and the amount of collected charge of all channels during the trigger latency and, at last, sending out data for bunch crossings. The analog front end, shown in Fig. 1.19, includes amplifiers with push-pull stages. The input transistors operate in the weak inversion regime, where the transconductance depends only on the drain current I_D and g_m/I_D is maximal, ensuring the speed needed to separate the LHC bunch crossings. For the charge sensitive preamplifier, a solution with a rather large feedback capacitance ($C_F = 20 \text{ fF}$) and an additional gain stage was preferred over a single-stage design with small C_F . Passive feedback with a $\text{M}\Omega$ resistor was chosen in order to keep the shaping time bigger than the amplifier risetime and at the same time absorb the leakage current without causing an offset. The resistor is implemented as a weak p-transistor operating in its linear region. The preamplifier input is DC coupled to the sensor pixels and its feedback must absorb the expected sensor leakage current of 10 nA per pixel. The second stage (shaper) has the same push-pull design and is AC-coupled to the preamplifier. The AC coupling serves a dual purpose. It removes offsets caused by leakage current and it is part of the gain stage. The gain can be calculated as the ratio of coupling-capacitance to feedback-capacitance. The shaper output is connected

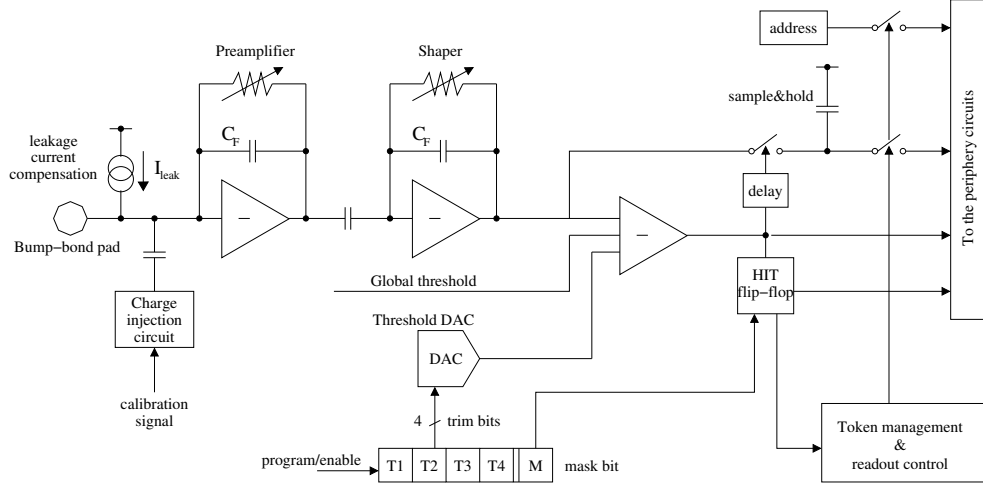


Figure 1.20: block diagram of a CMS pixel cell.

to a comparator and a sample-and-hold circuit, where the analog pulse-height information is stored for later readout. The comparator threshold is adjustable with an 8 bit DAC. Process related random variations of transistors lead to pixel-to-pixel threshold mismatch, but good threshold uniformity is essential to obtain a low global threshold without driving a large number of pixels into saturation. Additional 4 bit DACs in each pixel can compensate these variations and make fine-adjustments for individual pixel thresholds. The threshold dispersion in the final version of the chip for the CMS experiment (in a $0.25\ \mu\text{m}$ technology) is approximately 450 electrons, which can be reduced by trimming to 80 electrons. The input capacitance provided by the sensor chip is designed to be small enough to make the intrinsic noise of the front-end almost negligible. The readout circuit is organized in double columns of pixels that operate independently. The pixels communicate the detection of a hit over a wired OR signal. No clock or bunch crossing numbers need to be distributed over the pixel matrix. The periphery synchronizes the wired-OR with the LHC clock and latches the current bunch crossing number in a time-stamp buffer whenever a hit is found. The state of the pixels that have a hit at that time is frozen and their data is subsequently collected. A token passing from pixel to pixel controls the transfer. Even though empty pixels are skipped relatively fast, finding the next pixel with a hit can take longer than one bunch crossing. The data transmitted and stored in the data-buffers are the pixel address and the analog pulse-height. Marker bits in the data buffers keep track of the

association between time-stamps and hits.

Chapter 2

Superpix0, a front-end chip for hybrid pixel detectors in a planar CMOS technology

This chapter is focused on a prototype front-end chip for hybrid pixel detectors, named Superpix0, the first step towards the development of a detector to be used for the Layer0 upgrade. Section 2.1 describes the main features of the SuperB experiment and the characteristics of its Layer0, the part of the vertex detector closest to the particle interaction region, for which Superpix0 has been designed. Section 2.2 describes the device sensor matrix, the analog front-end, the in-pixel logic and the digital readout architecture. Section 2.3 will be concerned with the chip characterization results, obtained through tests in laboratory and on the SPS H6 beam line, at CERN.

2.1 The SuperB experiment

The SuperB Factory [17], a new asymmetric e^+e^- collider dedicated to heavy-flavour physics and expected to deliver unprecedented luminosities in excess of $10^{36} \text{ cm}^{-2}\text{s}^{-1}$, has been funded by the Italian Ministry of Education, University and Research in the framework of the 2011-2013 National Research Plan. Its reduced center of mass boost with respect to previous B-Factories (BaBar [18] and Belle [19]) asks for a factor two improvement on typical vertex resolutions to fully exploit the accelerator potential for new physics discoveries. In addition, the high luminosity, obtained with moderate beam currents, and large backgrounds expected at SuperB, determine stringent requirements in terms of granularity, time resolution and radiation hardness of all subdetectors

and, in particular, of the vertex detector, which is the closest to the interaction point, and its innermost layer, the Layer0.

2.1.1 Design specifications for the Layer0 of SuperB SVT

The design of the SuperB Silicon Vertex Tracker (SuperB SVT) follows the model of the BaBar SVT [20] but comprises both an extended coverage and an additional innermost layer, the Layer0, located at about 1.5 cm radius from the beam line. In the SuperB experiment, the silicon vertex detector will have the role of providing precise information on both the position and direction of charged particles emerging from the interaction point. The purpose of the SVT Layer0, which is the most critical part of the detector, is to measure the first hit of very high density tracks close to the production vertex. The Layer0 should offer a low material budget, to minimize multiple scattering, therefore meeting the requirements on vertex resolution. The detector must be also provided with a high speed readout to minimize the acquisition dead time. Intense R&D studies on various emerging technologies have been carried out to address further requirements such as a small pitch to guarantee a hit resolution at the level of 10 μm and to limit detector occupancy, the capability to withstand background hit rates up to a few tens of MHz/cm^2 , large signal-to-noise ratio and low power dissipation. Three different solutions are being considered for the Layer0: high resistivity short strip silicon detectors (or triplets), hybrid pixel sensors and CMOS monolithic active pixel sensors (MAPS). Standard high resistivity silicon detectors with short strips will be used for the Layer0 during the first period of operation, when the luminosity will be gradually increased to reach the design value. In fact, triplets offer a reasonably low material budget (about 0.2-0.3% X_0 for 200-300 μm silicon thickness) together with the required hit resolution. However, the detector occupancy becomes unaffordable at background rates larger than 5 MHz/cm^2 as expected at full luminosity, and a detector replacement is already scheduled after the first period of running. Hybrid pixel devices are a well established technology in HEP experiments. The fully depleted high-resistivity sensors and the readout integrated circuits are built on different substrates and then connected via high density bump bonding. Hybrid pixel sensors usually provide high signal-to-noise ratio, high radiation tolerance and 100% fill factor. Furthermore, this technology offers the possibility to implement advanced in-pixel functions such as low-noise amplification, zero suppression and threshold tuning without the problem of cross-talk between the readout logic and the sensor. The relatively large amount of material they are made of (the sensor and the readout chip, one on the top of the other) represents a disadvantage in terms of probability

of particle scattering, although a reduction of material budget may become possible with the latest technology advances [21]. Another possible choice for the Layer0 is represented by CMOS MAPS detectors, with a deep N-well acting as the collecting electrode and signal processing implemented at the pixel level [22] [23]. CMOS MAPS can meet the stringent material budget requirements set by SuperB Layer0, as their substrate can be thinned down to a few tens of microns with no significant signal loss. The most critical point of DNW MAPS is the low signal provided by the sensor and the relatively low degree of radiation hardness.

2.2 The Superpix0 chip

This section discusses the design of a front-end chip for hybrid pixel detectors in view of application to the SVT Layer0 of the SuperB experiment. The main feature of the Superpix0 prototype are the sensor size ($50 \times 50 \mu\text{m}^2$) and thickness ($200 \mu\text{m}$), as well as the custom front-end chip architecture providing a sparsified and data-driven readout. A prototype readout chip with 4096 cells arranged in a 32×128 matrix was submitted for fabrication in standard 130 nm CMOS technology by STMicroelectronics. The sensor layer was fabricated by Fondazione Bruno Kessler-IRST (FBK-IRST) and interconnected with the readout chip by the Fraunhofer Institute for Reliability and Microintegration

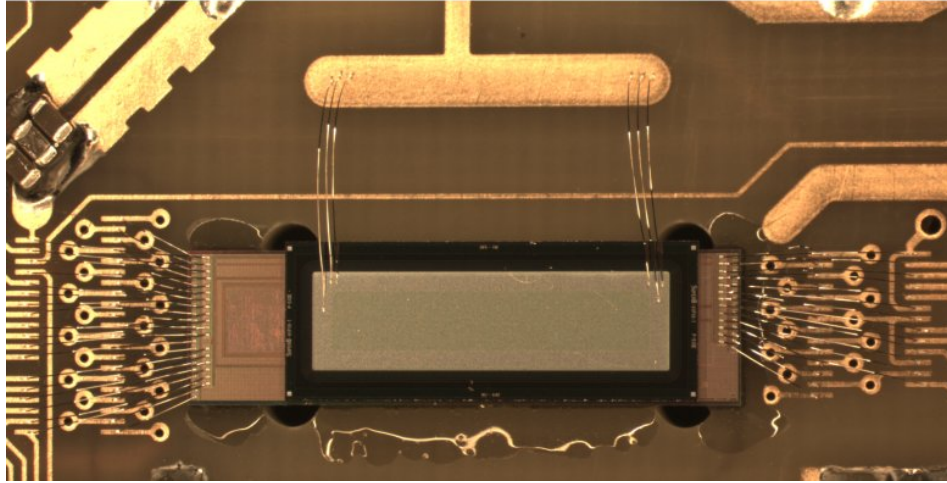


Figure 2.1: photograph of the Superpix0 readout chip bump-bonded to a pixel sensor matrix.

(IZM). Fig. 2.1 shows the detector, including the pixel sensor array and the readout chip, with the matrix of front-end channels, not visible because located below the sensor array, and the back-end digital circuits in the chip periphery, visible on the left in the picture. The chip is wire bonded on both the left and the right sides to a board, used for the laboratory tests, which makes it possible to feed all the input command signals to the chip. One side of the sensor matrix is kept at the front-end channel input potential through the chip to sensor interconnections, while the other side is biased at 40 V to achieve full depletion of the substrate.

2.2.1 The pixel sensor matrix

The pixel sensors, whose layout is shown in Fig. 2.2, are made from n-type, float zone, high-resistivity silicon wafers, with a thickness of 200 μm and a nominal resistivity larger than 10 $\text{k}\Omega\cdot\text{cm}$. Sensors are of the “n-on-n” type and were fabricated at FBK (Trento, Italy) with a double-sided technology [24]. N+ pixels are arranged in a matrix of 32x128 elements with a pitch of 50 μm in both X and Y directions, for a total active area size of 10.24 mm^2 . All around the pixels, a large n+ guard ring, extending up to the cut-line, has been designed. The electrical isolation between neighboring n+ pixels has been obtained by means of a uniform p-spray implantation. A large p+ diode is on the bias side: it has the same size as the active area and is surrounded by 6 floating rings. From electrical tests performed on wafers, before bump-bonding, by connecting the sensors from the bias side only with a probe on the diode and a probe

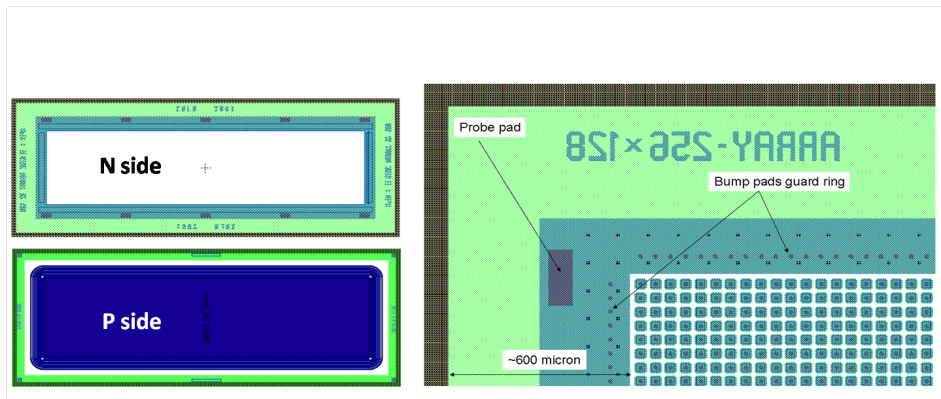


Figure 2.2: layout of the “n-on-n” pixel sensor matrix fabricated by FBK: p-spray isolation on n-side, p implant on the back side.

on the scribe line [25], the total leakage current is about 1 nA, the depletion voltage is about 10 V and the breakdown voltage in the order of 70 V, due to a relatively high p-spray dose. The pixel capacitance has also been estimated from measurements performed on a special test structure. The resulting values are in the order of 50 fF (i.e. close to the capacitance contribution expected from the bumps).

2.2.2 The analog front-end

The Superpix0 front-end chip has been designed taking into consideration the specifications set by the SVT Layer0. The pixel size should be as small as possible for good spatial resolution. A pixel pitch of $50 \times 50 \mu\text{m}^2$, while ensuring adequate spatial resolution, is about the largest pixel density available without using vertical integration technologies. Compared to a standard analog chain, the charge signal processor integrated into this chip has been compacted into a single block and does not include a standalone shaping stage. This was done to have more room in the pixels for connections between the front-end electronics and the peripheral readout logic. Indeed, since the number of connections between the detector array and the periphery increases with the

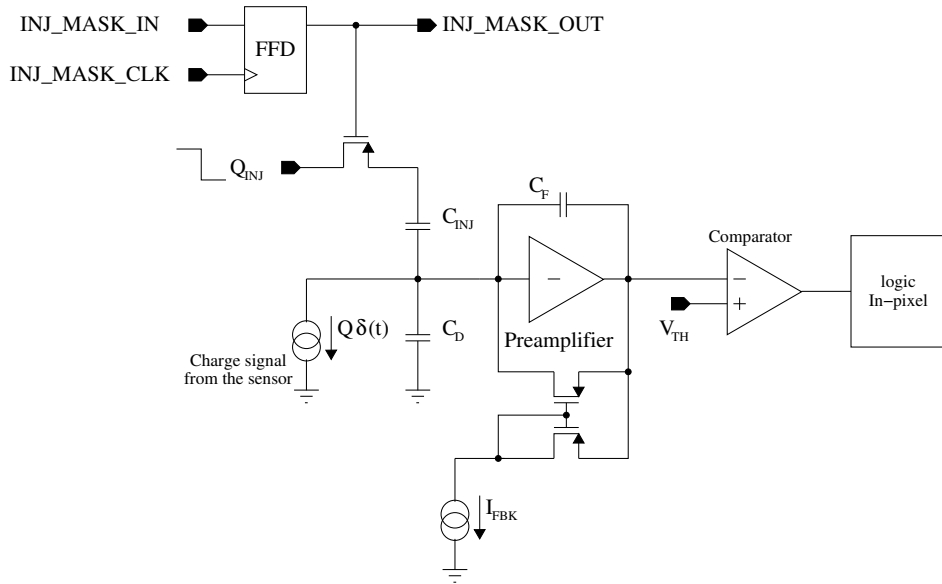


Figure 2.3: analog front-end electronics integrated in the pixel cell of the Superpix0 chip.

size of the matrix, the available area in the elementary cell for the layout of these connections poses a limit on the maximum size of the detector. By means of this design solution, the target pixel pitch and, consequently, the required spatial resolution was achieved. In each pixel the sensor signal is processed by an analog block (shown in Fig. 2.3) performing charge amplification and shaping, and compared to a chip-wide preset threshold by a discriminator. The in-pixel digital logic, which follows the comparator, stores the hit in an edge-triggered set reset flip-flop and notifies the periphery readout logic of the hit. The 20 fF preamplifier feedback capacitor (C_F) is discharged by a constant current which can be externally adjusted (I_{FBK}), giving an output pulse shape that is dependent upon the input charge. The sensor elements are DC coupled to the preamplifier input in order to avoid biasing and decoupling structures on the sensor. Important design parameters are the total (detector+bonding) parasitic capacitance C_D , that is about 100 fF, and the signal provided by the pixel sensor. When a charged particle passes through a hybrid pixel, typically 80 e/h pairs per micron are generated in the sensor. In a fully depleted detector, virtually all the electrons reach the preamplifier input. So, in the case of a wafer thickness of 200 μm , an input charge of about 16000 electrons is generated for a minimum ionizing particle, (a MIP, i.e., a particle with very high kinetic energy, which transfers the minimum possible energy to the sensitive medium). Assuming charge sharing, the signal charge per pixel may be as small as 4000 e. If a signal to noise ratio of 25 is required even in case of charge sharing, the equivalent noise charge (ENC) of the front-end electronics has to be smaller than 160 e-. The peaking time slightly increases with the collected charge and is in the order of 100 ns for 16000 electrons injected. The charge collected in the detector pixel reaches the preamplifier input via the bump-bond connection. Alternatively, a calibration charge can be injected at the preamplifier input through a 10 fF internal injection capacitance (C_{INJ}) so that threshold, noise and crosstalk measurements can be performed. The calibration voltage pulse is provided externally by a dedicated line (Q_{INJ}). Channel selection is performed by means of a control section implemented in each pixel. This control block, which is a cell of a shift register, enables the injection of the charge through the calibration capacitance. Each pixel features a digital mask, implemented in the readout logic, used to isolate single noisy channels.

The charge sensitive amplifier is the first part of the processing chain and is the most critical stage. Its purpose is to perform a charge to voltage conversion, while maximizing the signal to noise ratio. The detector signal $i_D(t)$, featuring a very short duration with respect to the front-end channel processing time,

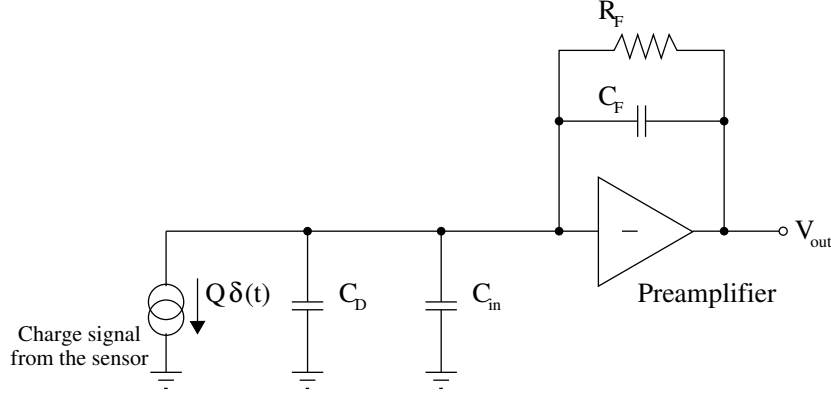


Figure 2.4: simplified scheme of the charge sensitive amplifier employed in the Superpix0 analog front-end.

can be safely approximated with a Dirac delta

$$i_D(t) = Q \cdot \delta(t) \quad (2.1)$$

with Laplace transform

$$I_D(s) = \mathcal{L}\{i_D(t), s\} = Q, \quad (2.2)$$

where Q is the charge generated in the detector by the impinging particle. If the simplified scheme of Fig. 2.4 is considered, where the current mirror in the preamplifier feedback network has been replaced with an equivalent resistor R_F , under the assumption that the open loop gain of the preamplifier tends to infinity, then

$$v_{out}(t) = \mathcal{L}^{-1}\{V_{out}(s), t\} = \mathcal{L}^{-1}\left\{\frac{R_F}{1 + sR_FC_F} Q, t\right\} = \frac{Q}{C_F} e^{-t/\tau_F}, \quad (2.3)$$

where $\mathcal{L}^{-1}$ represents the Laplace antitransform operation and $\tau_F = R_F C_F$. A more accurate analysis of the front-end channel response accounting for the non linear behavior of the feedback network, will be proposed later in this section. In the Superpix0 chip a charge sensitivity of 50 mV/fC was chosen (corresponding to a feedback capacitance of 20 fF) in order to comply with an input dynamic range of 64000 electrons (4 MIPs) and an output dynamic range of about 500 mV.

Stability study of the charge sensitive amplifier stability

In order to study the stability of the charge preamplifier, the open loop gain G_{loop} of the circuit can be calculated. If the output impedance of the amplifier

is negligible, it is possible to open the loop at the preamplifier output without the need for restoring the impedance at the same node (see Fig. 2.5). In the Laplace domain, at the input terminal of the amplifier

$$V'_-(s) = \frac{Z_{D,in}}{Z_{D,in} + Z_F} V'_{in}(s), \quad (2.4)$$

where

$$Z_{D,in} = \frac{1}{s(C_D + C_{in})}, \quad (2.5)$$

$$Z_F = \frac{R_F}{1 + sR_FC_F}, \quad (2.6)$$

and V'_{in} is the voltage source used for the G_{loop} calculation. Replacing the $Z_{D,in}$ and the Z_F in (2.4) leads to

$$\begin{aligned} V'_-(s) &= \frac{\frac{1}{s(C_D + C_{in})}}{\frac{1}{s(C_D + C_{in})} + \frac{R_F}{1 + sR_FC_F}} V'_{in}(s) \\ &= \frac{1 + sR_FC_F}{1 + sR_F(C_D + C_{in} + C_F)} V'_{in}(s). \end{aligned} \quad (2.7)$$

At the amplifier output, if a single pole transfer function is assumed for the open loop gain

$$V'_{out}(s) = -\frac{A_O}{1 + s\tau_A} V'_-(s). \quad (2.8)$$

Therefore the loop gain is given by

$$G_{loop}(s) = \frac{V'_{out}(s)}{V'_{in}(s)} = -\frac{A_O}{1 + s\tau_A} \frac{1 + sR_FC_F}{1 + sR_F(C_D + C_{in} + C_F)} \quad (2.9)$$

The relevant Bode plots are shown in Fig. 2.6. The loop gain (2.9) features two poles, $\omega_{p1} = 1/\tau_A$ and $\omega_{p2} = 1/[R_F(C_D + C_{in} + C_F)]$, and a zero, $\omega_z = 1/R_FC_F$. In order for the phase margin to exceed 45 degrees and meet the Bode stability criteria, the following condition on the gain-bandwidth product (GBP) of the amplifier has to be satisfied:

$$GBP = \frac{A_O}{\tau_A} < \frac{1}{R_F(C_D + C_{in} + C_F)}. \quad (2.10)$$

Indeed, the amplifier has to be designed taking into account the feedback network and the characteristics of the sensor. In this case, as already mentioned, the feedback capacitor value is 20 fF, the detector capacitance is

$C_D = 100$ fF, the input capacitance generally is in the order of 10 fF and R_F is about 10 M Ω . Then, the gain bandwidth product of the amplifier must be smaller than 120 dB. It is worth emphasizing that these results are valid if R_O is negligible, R_O being the open loop output impedance of the amplifier. The study of the system stability can be complicated by assuming that the amplifier has two poles $1/\tau_A$ and $1/\tau_B$, with $\tau_B < \tau_A$. In this case (2.9)

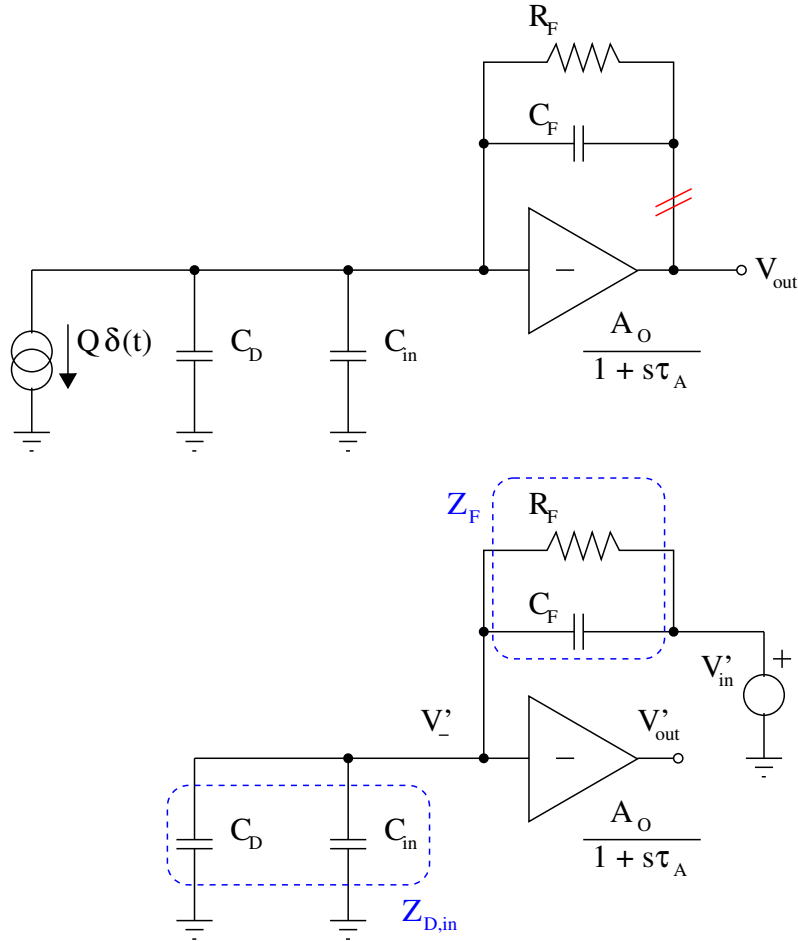


Figure 2.5: charge sensitive amplifier schemes used for the open loop gain calculation: closed loop scheme, with indication of the point chosen for loop opening (top) and open loop scheme (bottom).

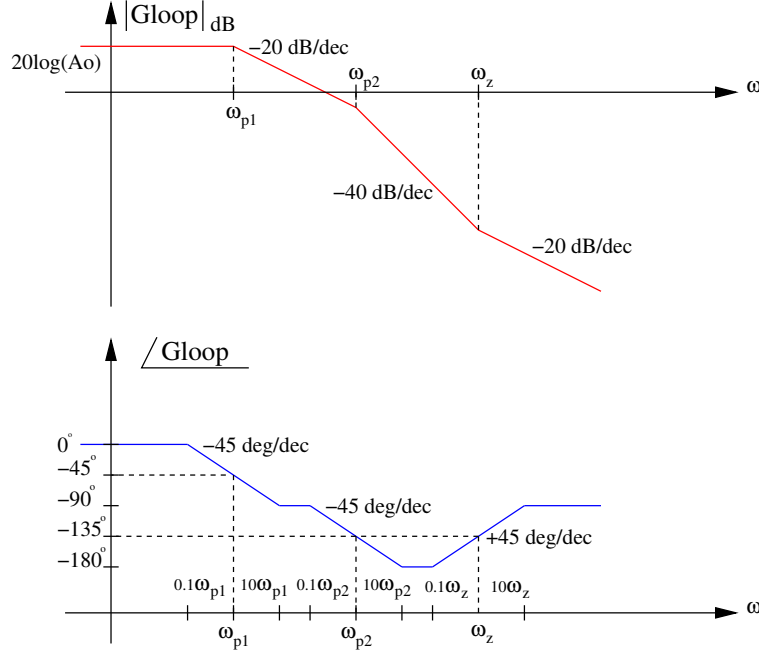


Figure 2.6: Bode plot of the magnitude and phase of the loop gain function G_{loop} as represented in (2.9).

becomes

$$G_{loop} = -\frac{A_O}{(1 + s\tau_A)(1 + s\tau_B)} \frac{1 + sR_FC_F}{1 + sR_F(C_D + C_{in} + C_F)}. \quad (2.11)$$

Under the condition that $\tau_A \gg \tau_B$ and R_F is very large, the phase at -135 degrees in the G_{loop} can be found at the angular frequency $\omega_s = 1/\tau_B$, where the condition $|G_{loop}| \leq 1$ has to be satisfied

$$\frac{C_F}{C_D + C_{in} + C_F} \frac{A_O\tau_B}{\tau_A\sqrt{2}} \leq 1 \quad (2.12)$$

Therefore, if the amplifier has two poles, a limit exists for the value of τ_B in order to obtain a phase margin larger than 45 degrees:

$$\tau_B \leq \frac{\tau_A\sqrt{2}}{A_O} \frac{C_D + C_{in} + C_F}{C_F}. \quad (2.13)$$

The case where the impedance R_O of the amplifier is no longer negligible is considered in the following. The stability study will be performed with

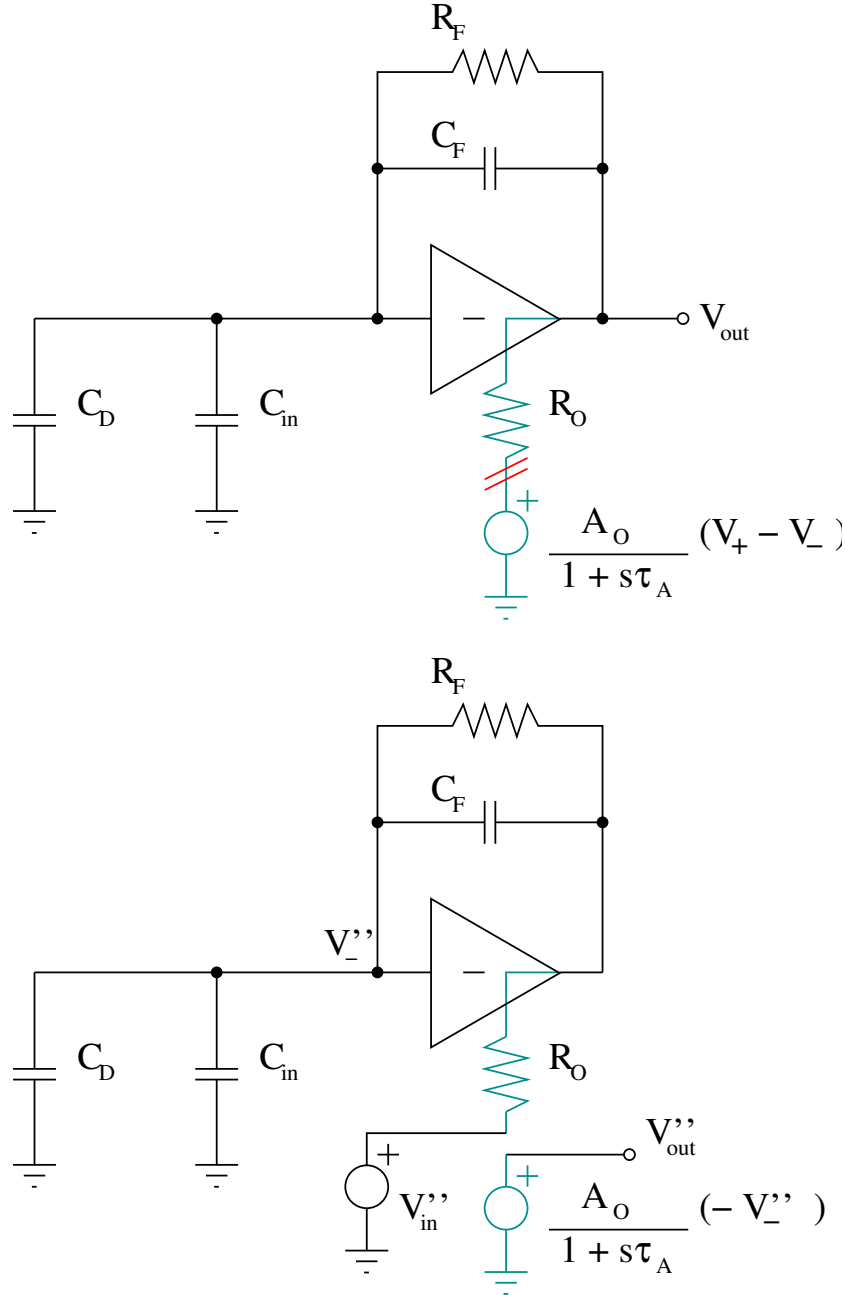


Figure 2.7: CSA scheme used for the open loop gain calculation in case where the output resistance of the amplifier cannot be neglected: closed loop scheme, with indication of the point chosen for loop opening (top) and open loop scheme (bottom).

reference to Fig. 2.7. The dominant pole approximation will be used for the open loop gain of the amplifier. If a R_F which tends to infinity is assumed, then the input terminal voltage in the Laplace domain is given by

$$V_-(s) = \frac{\frac{1}{s(C_D+C_{in})}}{\frac{1}{s(C_D+C_{in})} + \frac{1+sR_OC_F}{sC_F}} V_{in}''(s), \quad (2.14)$$

and the output voltage is

$$V_{out}''(s) = -\frac{C_F}{C_D + C_{in} + C_F} \frac{A_O}{1 + s\tau_A} \frac{1}{1 + sR_O \frac{C_F(C_D+C_{in})}{C_D+C_{in}+C_F}} V_{in}''(s). \quad (2.15)$$

Hence

$$G_{loop}'' = -\frac{C_F}{C_D + C_{in} + C_F} \frac{A_O}{1 + s\tau_A} \frac{1}{1 + sR_O \frac{C_F(C_D+C_{in})}{C_D+C_{in}+C_F}}. \quad (2.16)$$

Equation (2.16) shows that an amplifier with a single dominant pole and a feedback resistance which tends to infinity are not a sufficient condition for stability, at least theoretically. However, in general, the second pole is located at very high frequencies. This is easily obtained by suitable design of the output stage of the amplifier.

The Superpix0 charge sensitive amplifier

The charge sensitive amplifier, shown in Fig. 2.8, uses a single-ended folded cascode topology, which is a common choice for low-voltage, high gain amplifiers. The M1 and M3 devices implement the folded cascode configuration. Moreover, in order to increase the impedance seen at the M2 drain, a local feedback is used involving the M6 transistor (active folded cascode). M4, M5 and M9 implement the active preamplifier load, while the output stage is a source follower consisting of the M10 and M11 NMOS transistors. A MOS capacitor (M12) has also been used with the aim of limiting the bandwidth of the preamplifier and of reducing high frequency noise contributions. In the 130 nm ST Microelectronics process, two types of transistors are available: high speed (HS) transistors, featuring a threshold voltage of about 190 mV, and low leakage (LL) transistors with a threshold voltage of about 300 mV. Table 2.1 indicates the size of the transistors used in the preamplifier and the transistor type (HS or LL). As depicted in Fig. 2.8, the two voltage references, required in the preamplifier stage, are obtained by means of a simple voltage partition scheme, whose dimensions are shown in table 2.2. The first, including M18,

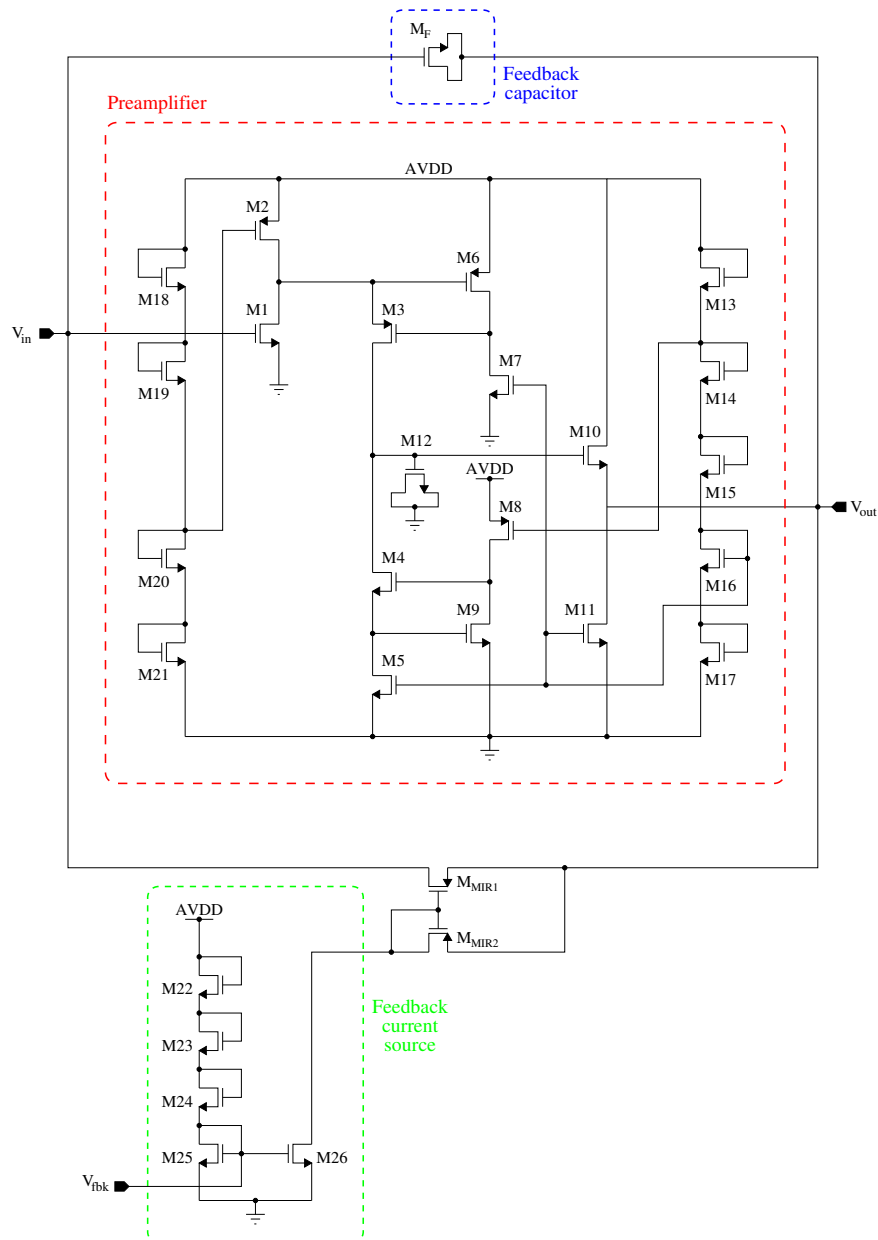


Figure 2.8: schematic circuit of the Superpix0 charge preamplifier.

M19, M20 and M21, sets the current in the input branch. A similar voltage reference, composed of M13, M14, M15, M16 and M17, and shared with the threshold discriminator, is used to set the current flowing into the cascode branch (220 nA).

The input device M1, whose dimensions were chosen, as will be explained later, as a compromise between the noise and the output baseline dispersion performance of the preamplifier, featuring an aspect ratio $W/L=18 \mu\text{m}/0.3 \mu\text{m}$ and a drain current of about $1 \mu\text{A}$, is biased in the weak inversion region. A non-minimum length has been chosen to avoid short channel effects. The size of the PMOS current source in the input branch (M2) has been chosen to have a smaller transconductance than the input transistor. The feedback capacitor C_F is the parasitic gate capacitance of the M_F transistor. Charge restoration in the preamplifier feedback network, which is discussed in details in the next section, is obtained through a current mirror stage (M_{MIR1} and M_{MIR2}), providing an almost linear discharge of the capacitor C_F . As already mentioned, the design charge sensitivity is around 50 mV/fC . In the design of the circuit, some measures have been taken to prevent the noise performance from degrading as a consequence of the elimination of the shaping stage from the readout chain. High frequency noise contribution has been reduced by purposely limiting the charge preamplifier bandwidth. The equivalent noise charge (ENC)

Table 2.1: dimension and type of the transistors used in the charge preamplifier.

Device	W/L [$\mu\text{m}/\mu\text{m}$]	Type
M1	18 / 0.3	Low Leakage
M2	0.8 / 4.6	Low Leakage
M3	0.18 / 0.36	High Speed
M4	0.18 / 0.45	Low Leakage
M5	0.5 / 7	Low Leakage
M6	0.4 / 0.4	High Speed
M7	0.25 / 3	Low Leakage
M8	0.58 / 0.4	Low Leakage
M9	0.18 / 0.4	High Speed
M10	0.8 / 0.8	High Speed
M11	0.4 / 3	Low Leakage
M12	0.8 / 4	Low Leakage

Table 2.2: MOSFET dimension and type included in voltage references and in the feedback current source.

Device	W/L [$\mu\text{m}/\mu\text{m}$]	Type
M13	0.2 / 0.6	Low Leakage
M14	5 / 0.15	Low Leakage
M15	0.8 / 2.4	Low Leakage
M16	1.2 / 2.4	Low Leakage
M17	0.8 / 2.4	Low Leakage
M18	0.15 / 2	Low Leakage
M19	4 / 0.2	Low Leakage
M20	0.2 / 7	Low Leakage
M21	0.4 / 3.5	Low Leakage
M22	0.15 / 3	Low Leakage
M23	0.15 / 3	Low Leakage
M24	0.15 / 3	Low Leakage
M25	6.23 / 0.15	High Speed
M26	0.2 / 1.5	High Speed

is an important parameter to evaluate the noise performance of the front-end circuits. The ENC can be defined as the charge that has to be injected at the input of the charge measuring system in order to have an output waveform with peak amplitude equal to the output rms noise. Based on this definition, the following equation holds,

$$ENC = \frac{\sqrt{\overline{v_n^2}}}{G_Q}, \quad (2.17)$$

where $\overline{v_n^2}$ is the mean square value of the noise at the preamplifier output. Fig. 2.9 shows, for a detector capacitance of 100 fF, an equivalent noise charge of 140 e- obtained from circuit simulations. The noise contribution arising from the leakage current can be neglected for the range considered in the simulations, $I_{LEAK} \leq 2$ pA, which corresponds to twenty times the anticipated leakage current for the pixel sensor. Particular care was put in the design of the front-end stage from the standpoint of threshold dispersion, mainly arising from dispersion in the threshold voltage of the preamplifier input device and of the PMOS and NMOS pairs in the discriminator. An overall input referred threshold dispersion of 350 e- rms was computed from Monte-Carlo simula-

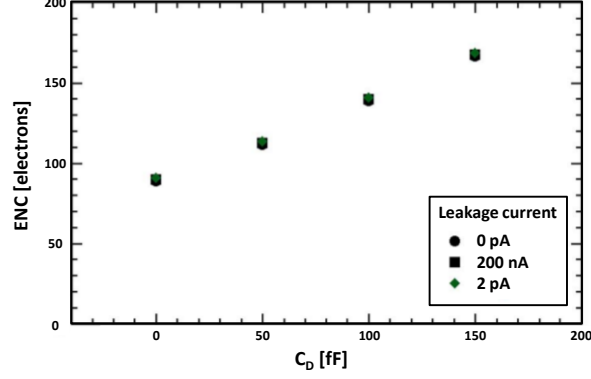


Figure 2.9: Simulated equivalent noise charge as a function of the detector capacitance C_D for different values of the leakage current.

tions. Since SuperPix0 is the first iteration step in the framework of an R&D activity aimed at the development of a readout chip for small pitch hybrid pixel sensors, in this design only the main functionalities have been integrated in the pixel cell. Threshold dispersion is a crucial characteristic to be considered in order to meet the specifications in terms of noise occupancy and efficiency. Therefore, circuits for in-pixel threshold fine-adjusting have been implemented in the new version of the chip, named Superpix1 and described in the last chapter of this work.

Small signal circuit analysis

The signal analysis of the preamplifier stage is carried out with reference to Fig. 2.10, which is the small signal equivalent circuit of the scheme in Fig.2.8. In the following analysis, g_{dsi} is the drain to source transconductance of the Mi transistor, g_{mi} the channel transconductance of the same device. The transfer function between the gate voltage of M10 (v'_{out}) and the input source (v_{in}) can be expressed as:

$$\frac{v'_{out}(s)}{v_{in}(s)} = \frac{g_{m1}}{g_{out1} + g_{out2} + s c'_{out}} \quad (2.18)$$

where g_{out1} and g_{out2} represent the transconductance seen, respectively, at the drain of M3 and M4, and

$$c'_{out} = c_{dd3} + c_{dd4} + c_{gg10} + c_{gg12} \quad (2.19)$$

is the capacitance loading the same node (from circuit simulation $c'_{out} = 40$ fF). Due to the local negative feedback, the gate voltage of the

M3 transistor results to be an amplified version of its source voltage (see Fig. 2.10(b))

$$v_{g3} = \frac{g_{m6}}{g_{ds6} + g_{ds7} + sC_3} v_{s3} = \frac{\frac{g_{m6}}{g_{ds6} + g_{ds7}}}{1 + s \frac{C_3}{g_{ds6} + g_{ds7}}} v_{s3} \approx \frac{g_{m6}}{g_{ds6} + g_{ds7}} v_{s3} = A_1 v_{s3} \quad (2.20)$$

where $c_3 = c_{dd6} + c_{dd7} + c_{gg3}$. Circuit simulations show that the time constant $\tau_3 = c_3/(g_{ds6} + g_{ds7})$ is very small ($\tau_3 = 12ns$) as compared to the characteristic times of the circuit and can therefore be neglected. The g_{out1} transconductance, featuring a value of 283 pS, has the following expression:

$$\frac{1}{g_{out1}} = \frac{1}{g_{ds1} + g_{ds2}} + \frac{1}{g_{ds3}} + \frac{g_{m3}}{(g_{ds1} + g_{ds2})g_{ds3}}(1 + A_1). \quad (2.21)$$

The same procedure can be used to calculate the transconductance g_{out2} seen at the drain of M4,

$$\frac{1}{g_{out2}} = \frac{1}{g_{ds4}} + \frac{1}{g_{ds5}} + \frac{g_{m4}}{g_{ds4}g_{ds5}}(1 + A_2), \quad (2.22)$$

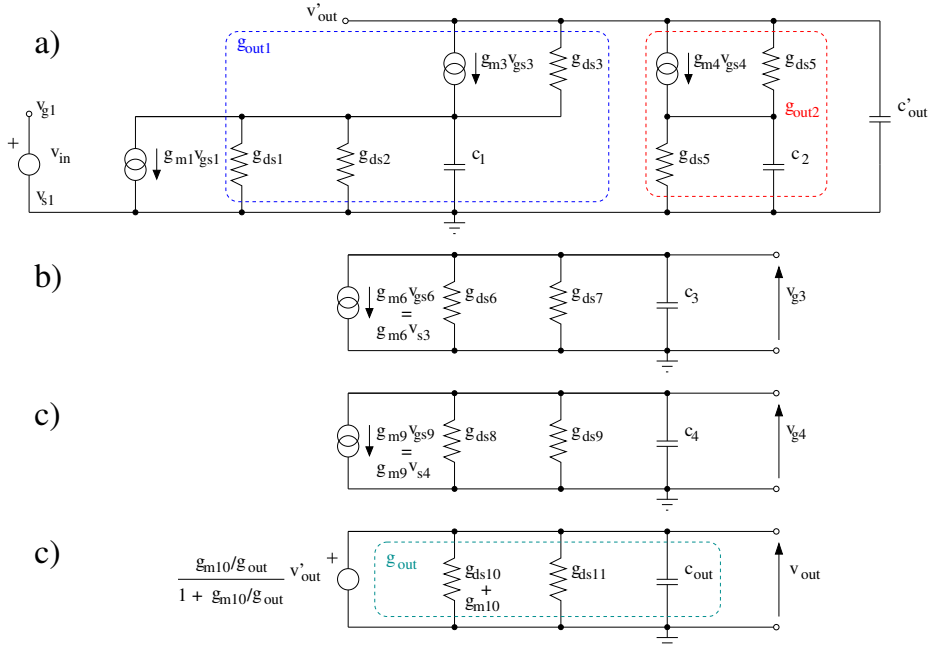


Figure 2.10: small signal circuit for the preamplifier stage.

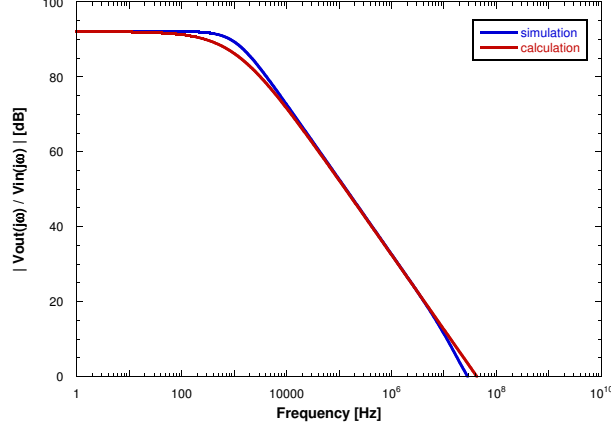


Figure 2.11: magnitude of the open loop gain of the charge preamplifier. Simulation results are compared to the calculated response.

where A_2 is a multiplicative factor, resulting from the presence of a second negative feedback loop, shown in Fig. 2.10(c), involving the M4 transistor

$$\begin{aligned} v_{g4} &= \frac{g_{m9}}{g_{ds8} + g_{ds9} + sC_4} v_{s4} = \frac{\frac{g_{m9}}{g_{ds8} + g_{ds9}}}{1 + s \frac{C_4}{g_{ds8} + g_{ds9}}} v_{s4} \\ &\approx \frac{g_{m9}}{g_{ds8} + g_{ds9}} v_{s4} = A_2 v_{s4}. \end{aligned} \quad (2.23)$$

In (2.23), $c_4 = c_{dd8} + c_{dd9} + c_{gg4}$. Again, simulation results show that $c_4/(g_{ds8} + g_{ds9}) = 6$ ps and, therefore, can safely be neglected. From simulation results $g_{out2} = 83$ nS $\gg g_{out1}$ is obtained. Then, (2.18) can be approximated to

$$\frac{v'_{out}(s)}{v_{in}(s)} \approx \frac{g_{m1}}{g_{out1} + sC'_{out}}. \quad (2.24)$$

Therefore, the transfer function between v_{in} and the v_{out} is

$$\begin{aligned} \frac{v_{out}(s)}{v_{in}(s)} &= \frac{g_{m1}}{g_{out1} + sC'_{out}} \cdot \frac{\frac{g_{m10}}{g_{out}}}{1 + \frac{g_{m10}}{g_{out}}} \\ &\approx \frac{g_{m1}}{g_{out1} + sC'_{out}} \end{aligned} \quad (2.25)$$

where

$$g_{out} = g_{m10} + g_{ds10} + g_{ds11} + sC_{out} \quad (2.26)$$

and

$$C_{out} = c_{ss10} + c_{dd11} + c_{ggDISC} \quad (2.27)$$

The c_{ggDISC} capacitor is the input capacitance of the discriminator stage that follows the preamplifier. The DC gain is

$$A_O = \frac{g_{m1}}{g_{out1}} \simeq 92.1 \text{ dB} \quad (2.28)$$

and the time constant is

$$\tau_A = \frac{c'_{out}}{g_{out1}} \simeq 151 \text{ } \mu s \quad (2.29)$$

Fig. 2.11 compares the open loop gain of the charge preamplifier obtained from simulations (blue line) with the one obtained from small signal analysis. It is worth noticing that the value of A_O is larger enough to ensure the condition $A_O C_F \gg C_D$. Therefore C_D does not affect significantly the charge sensitivity.

Feedback network analysis

The current mirror used for the charge restoration in the preamplifier feedback network provides an almost linear discharge of the feedback capacitor and a slight increase of the peaking time with the amplitude of the input charge pulse. This feature can be accounted for by studying the delta response of the circuit as it is modeled in Fig. 2.12, where the non-linear behavior of the feedback network is taken into consideration. In this scheme, g_{m1} is the transconductance of the preamplifier input device, while the capacitor c'_{out} and the resistor $r'_{out} = 1/g_{out1}$ are used to model the impedance in the high gain node of the amplifier, as it has been done in the previous section. Actually, c'_{out} is dominated by a 40 fF capacitor used to limit the amplifier bandwidth, as mentioned above. The unity gain block represents the second stage of the circuit, consisting of a source follower. The input signal is modeled, as previously done, by means of a current source generating a delta shaped current signal, with an area corresponding to the amount of charge Q collected by the sensor. The feedback capacitor C_F is continuously reset by means of a PMOS current mirror, which can be considered off until a signal is sent to the preamplifier input and the output voltage v_{out} starts increasing. The reset current i_{FBK1} can be expressed as follows:

$$i_{FBK1} = \alpha(v_{out} - v_{in}) I_{FBK2} \quad (2.30)$$

where I_{FBK2} is the reference source in the current mirror, amounting to a few nA and $\alpha(x)$ is a monotonically increasing function of x . In particular, α can be assumed to be 0 for $v_{out} - v_{in} \leq 0$ when the mirror is off. On the other hand, α reaches a constant value α_0 when $v_{out} - v_{in} \geq V_{SD,sat}$ ($V_{SD,sat}$ being

$$g_{m1}v_{in}(t) + \frac{1}{r'_{out}}v_{out}(t) + c'_{out}\frac{dv_{out}(t)}{dt} = 0, \quad (2.31)$$

$$\begin{aligned} \frac{d^2 v_{out}(t)}{dt} + \left(\frac{1}{r'_{out} c'_{out}} + \frac{g_{m1}}{c'_{out}} \right) \cdot \frac{dv_{out}(t)}{dt} + \\ + \frac{g_{m1}}{c'_{out} C_F} \alpha(v_{out} - v_{in}) I_{FBK2} - \frac{g_{m1} Q}{c'_{out} C_F} \delta(t) = 0. \end{aligned} \quad (2.33)$$

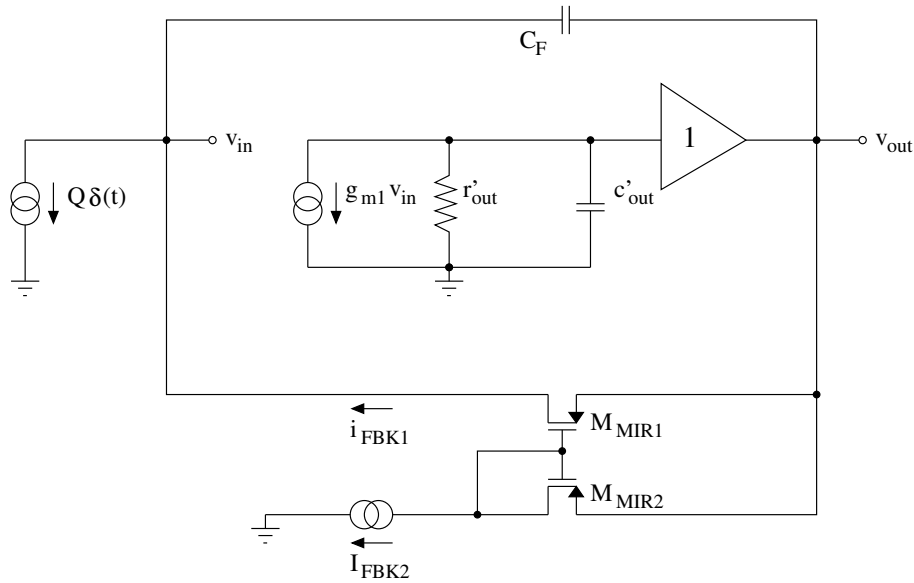
$$v_{out}(t) = H(t) \cdot \left[\frac{Q}{C_F} (1 - e^{-t/\tau}) - \frac{\alpha_0 I_{FBK2}}{C_F} t \right], \quad (2.34)$$


Figure 2.12: charge preamplifier model for studying the behavior of the feedback network as a response to a delta shaped current signal.

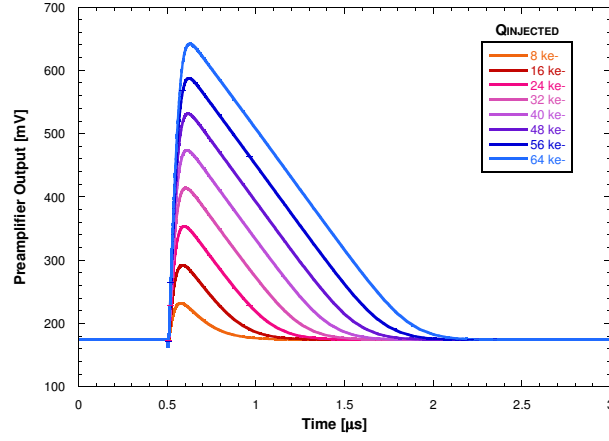


Figure 2.13: simulated response of the charge preamplifier to an input charge pulse with varying amplitude.

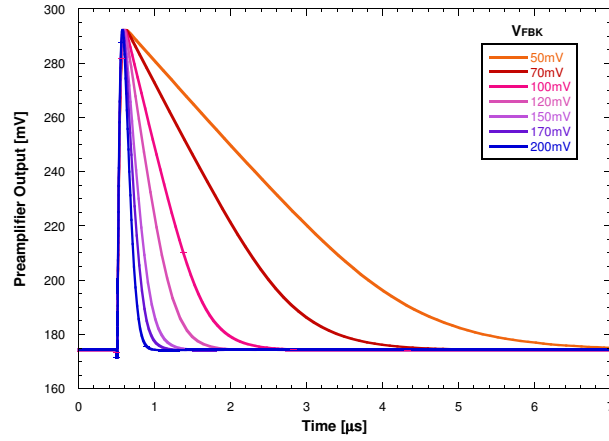


Figure 2.14: simulated response of the charge preamplifier to a 16000 input charge for different values of the current in the feedback current mirror, obtained by changing the control voltage V_{FBK} externally provided to the chip.

where $H(t)$ is the Heviside function and

$$\tau = r'_{out}c'_{out} + \frac{c'_{out}}{g_{m1}}. \quad (2.35)$$

In order to incorporate in (2.34) the non-linear behavior of α , $v_{out}(t)$ should be zero for $t > t_0$, t_0 being such that $v_{out}(t_0) = 0$:

$$\left[\frac{Q}{C_F}(1 - e^{-t_0/\tau}) - \frac{\alpha_0 I_{FBK2}}{C_F} t_0 \right] = 0. \quad (2.36)$$

This is needed to account for the fact that, when v_{out} returns to zero, the current mirror is switched off and the discharge phase must come to an end. From (2.34), the peaking time t_p can be found to change with the injected charge according to the following equation, showing that t_p is a monotonically increasing function of the input charge:

$$t_p(Q) = \tau \cdot \ln \left(\frac{Q}{\tau \alpha_0 I_{FBK2}} \right). \quad (2.37)$$

It is worth noticing that (2.34) predicts, as expected, a constant slope decrease of $v_{out}(t)$ after the peaking time. This behavior, however, is valid only for large value of the injected charge, while at smaller values of Q , the slope changes significantly, as a consequence of the approximation that was made to linearize (2.33). Actually, for small values of Q , when $v_{out} - v_{in}$ is too small for M_{MIR1} to work in the saturation region, $\alpha < \alpha_0$ and the discharge of C_F is slower than for higher Q values.

Noise performance analysis

In readout channels for capacitive detectors, the main noise sources are located in the charge preamplifier. Fig. 2.15 shows the same schematic diagram as in Fig. 2.4 with equivalent, series and parallel noise sources, e_n and i_n respectively, at the preamplifier input, and the parallel noise contribution, i_F , in the feedback network. Again, C_F and C_{in} are, respectively, the preamplifier feedback and input capacitance and C_D is the detector parasitic capacitance. Actually, e_n and i_n correspond to the noise sources in the input device, that is the input electrode of the charge preamplifier, all the other sources being negligible in a well designed circuit. In the circuit of Fig. 2.15 the charge resetting PMOS current mirror has been replaced again with an equivalent resistor R_F , with the purpose of simplifying circuit analysis in the small signal regime. A

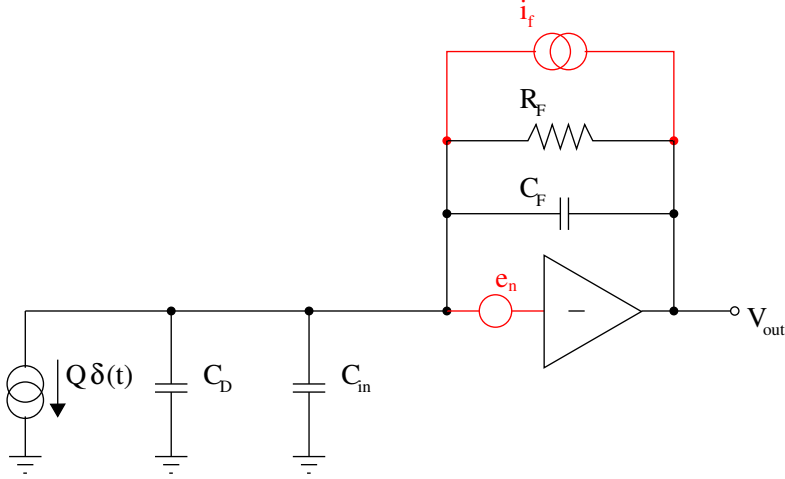


Figure 2.15: schematic circuit of the charge preamplifier with noise sources.

fairly general hypothesis for the power spectral density of the noise (assumed monolateral) is provided by the following equation,

$$\frac{de_n^2}{df} = S_w + \frac{S_f}{f}, \quad (2.38)$$

where a white and a flicker noise contributions, S_w and S_f/f respectively, are assumed for the series noise source. S_w is the gate referred channel thermal noise in the input transistor and S_f is the power coefficient of the flicker noise. For the open loop gain, a single pole transfer function will be assumed. The transfer function between the input current source modeling the detector signal and the preamplifier output is given by

$$\frac{V_{out}(s)}{Q} = \frac{\frac{A_O R_F}{1+A_O}}{\frac{R_F(C_D+C_{in}+C_F)\tau_A}{1+A_O} s^2 + \frac{\tau_A + R_F(C_D+C_{in}+C_F(1+A_O))}{1+A_O} s + 1}. \quad (2.39)$$

With the conditions that $A_O \gg 1$ and $A_O C_F \gg (C_D + C_{in})$

$$\begin{aligned} \frac{V_{out}(s)}{Q} &\approx \frac{A_O R_F}{R_F(C_D + C_{in} + C_F)\tau_A s^2 + (R_F C_F A_O) s + A_O} \\ &= \frac{R_F}{\frac{R_F(C_D+C_{in}+C_F)\tau_A}{A_O} s^2 + R_F C_F s + 1} \\ &= R_F \frac{1}{a s^2 + b s + 1}, \end{aligned} \quad (2.40)$$

where

$$a = \frac{R_F(C_D + C_{in} + C_F)\tau_A}{A_O}, \quad b = R_F C_F.$$

Series white noise contribution

The thermal noise in the channel of the input transistor leads to an equivalent white noise source at its gate with a spectral density, which can be expressed as

$$S_w = 4kT\Gamma \frac{1}{g_{m1}}, \quad (2.41)$$

where g_{m1} is the transconductance of the preamplifier input device, k is the Boltzmann constant, T is the absolute temperature, and Γ is channel thermal noise coefficient, whose value depends on the operating conditions of the transistor and varies between 1/2 in weak inversion and 2/3 in strong inversion. Under the hypothesis that $A_O \gg 1$ and $A_O C_F \gg C_D + C_{in}$ the transfer function from the series noise source to the circuit output is given by

$$\begin{aligned} H_{en}(s) = \frac{V_{out}(s)}{e_n} &\approx \frac{A_O R_F (C_D + C_{in} + C_F) s}{R_F (C_D + C_{in} + C_F) \tau_A s^2 + (R_F C_F A_O) s + A_O} = \\ &= \frac{R_F (C_D + C_{in} + C_F) s}{\frac{R_F (C_D + C_{in} + C_F) \tau_A}{A_O} s^2 + R_F C_F s + 1} \\ &= R_F (C_D + C_{in} + C_F) \frac{s}{a s^2 + b s + 1}. \end{aligned} \quad (2.42)$$

The mean square noise at the preamplifier output can be calculated as

$$\begin{aligned} \overline{v_{out}^2} &= \int_0^{+\infty} S_w |H_{en}(j\omega)|^2 d\omega \\ &= 4kT\Gamma \frac{R_F^2 (C_D + C_{in} + C_F)^2}{g_{m1}} \int_0^{+\infty} \frac{1}{2\pi} \frac{\omega^2}{a^2 \omega^4 + (b^2 - 2a)\omega^2 + 1} d\omega \\ &= 4kT\Gamma \frac{R_F^2 (C_D + C_{in} + C_F)^2}{g_{m1}} \frac{1}{2\pi} \left(\frac{\pi}{2ab} \right) \\ &= kT\Gamma \frac{A_O (C_D + C_{in} + C_F)}{\tau_A g_{m1} C_F}. \end{aligned} \quad (2.43)$$

The value extracted from (2.43) is $\sqrt{\overline{v_{out,en1}^2}} = 683 \mu\text{V}$, with the input capacitance $C_{in} = C_{gg1} = 30 \text{ fF}$, $g_{m1} = 11.8 \mu\text{S}$ and $\Gamma = 2/3$. The result obtained through (2.43) is compatible with simulations results providing a noise contribution from the M1 transistor of $\sqrt{\overline{v_{out,en1}^2}} = 694 \mu\text{V}$.

Series low frequency noise contribution

Low frequency contributions to the preamplifier noise performance are mainly provided by fluctuations of the $1/f$ kind in the drain current of the NMOS input device. The relevant monolateral noise spectral density can be expressed by means of the relationship

$$\frac{\overline{de_{n,1/f}^2}}{df} = \frac{S_f}{f} = \frac{K_f}{WLC_{ox}f}, \quad (2.44)$$

where S_f is a $1/f$ noise coefficient, K_f is a technology dependent parameter, W and L the preamplifier input device dimensions, and f is the frequency. The mean square noise at the preamplifier output due to low frequency noise is given by

$$\begin{aligned} \overline{v_{out}^2} &= \int_0^{+\infty} \frac{S_f}{f} |H_{en}(j\omega)|^2 df \\ &= \frac{K_f}{WLC_{ox}} R_F^2 (C_D + C_{in} + C_F)^2 \int_0^{+\infty} \frac{\omega}{a^2\omega^4 + (b^2 - 2a)\omega^2 + 1} d\omega \\ &\approx \frac{K_f}{WLC_{ox}} R_F^2 (C_D + C_{in} + C_F)^2 \frac{b^2}{b^4 - a^2} \ln\left(\frac{b^2}{a}\right) \\ &= \frac{\frac{K_f}{WLC_{ox}}}{\left(\frac{C_F}{C_D + C_{in} + C_F}\right)^2 - \left(\frac{\tau_A}{R_F C_F A_O}\right)^2} \ln\left(\frac{A_O R_F C_F^2}{\tau_A (C_D + C_{in} + C_F)}\right) \end{aligned} \quad (2.45)$$

The square root of the value extracted from (2.45) $\sqrt{\overline{v_{out, en1}^2}} = 47.9 \mu\text{V}$ was obtained, with $K_f = 10^{-20} \text{ V}^2\text{F}$, $C_{ox} = 18 \text{ fF}/\mu\text{m}^2$ [30] [31] and $R_F = 1/g_{ds1} = 110 \text{ nS}$. The value of the flicker noise contribution of M1 transistor, obtained by circuit simulations, is equal to $\sqrt{\overline{v_{out, flicker M1}^2}} = 35.8 \mu\text{V}$ and, actually, can be neglected with respect the series frequency noise contribution.

Feedback noise contribution

The feedback noise current source i_f models the white noise in the channel of the feedback transistor M_{MIR1} with a monolateral spectral density

$$S_w = 4kT\Gamma g_{m, MIR1}, \quad (2.46)$$

where $g_{m, MIR1}$ is the transconductance of the M_{MIR1} transistor.

$$i_f = \frac{\frac{V_{out}(s)}{A_O}}{\frac{1}{s(C_D + C_{in})}} + \frac{V_{out}(s) + \frac{V_{out}}{A_O}}{\frac{R_F}{1 + sR_F C_F}} \quad (2.47)$$

The transfer function between v_{out} and the noise signal i_f is:

$$\begin{aligned}
 H_{if}(s) = \frac{V_{out}(s)}{i_f} &\approx \frac{A_O R_F}{R_F(C_D + C_{in} + C_F)\tau_A s^2 + (R_F C_F A_O) s + A_O} = \\
 &= \frac{R_F}{\frac{R_F(C_D + C_{in} + C_F)\tau_A}{A_O} s^2 + R_F C_F s + 1} \\
 &= R_F \frac{1}{a s^2 + b s + 1}
 \end{aligned} \tag{2.48}$$

The rms noise obtained by integrating the square module of the tranfer function $H_{if}(s)$ multiplied by the spectral density of the noise S_w is

$$\begin{aligned}
 \overline{v_{out}^2} &= \int_0^{+\infty} S_w |H_{if}(j\omega)|^2 df \\
 &= \frac{8}{3} kT g_{m,MIR1} R_F^2 \int_0^{+\infty} \frac{1}{2\pi} \frac{1}{a^2 \omega^4 + (b^2 - 2a)\omega^2 + 1} d\omega \\
 &= \frac{8}{3} kT R_F \frac{1}{2\pi} \left(\frac{\pi}{2b}\right) \\
 &= \frac{2}{3} kT \frac{1}{C_F}
 \end{aligned} \tag{2.49}$$

The root mean square of the feedback noise contribution, calculated with the (2.49) is 371 μV , compatible with the value obtained from circuit simulation (357 μV).

Equivalent Noise Charge

The noise performance of the analog channel has been evaluated based on noise simulation results providing a total root mean square noise of 1.1 mV. This

Table 2.3: Main noise contributions in the Superpix0 analog front-end.

Device	Noise contribution [μV]
M1	694
M2	393
M _{MIR1}	357
M5	290
M6	220

result has been obtained with a detector capacitance C_D of 100 fF. Table 2.3 summarizes the main noise contributions in the circuit. The preamplifier input transistor (M1) provides the most significant contribution to the total noise. The resulting equivalent noise charge, calculated based on (2.17) is 140 electrons rms.

Threshold discriminator

The last block in the analog signal processing channel is the comparator. From this point on, the signal takes a digital nature, being low (zero volts), if the signal is below a globally preset threshold, or high (1.2 V) if it is above. The digital signal is fed to the in-pixel logic and then sent to the periphery. The threshold is set as low as possible in order to maximize the detection efficiency. On the other hand, the rate of noise hits must be kept at an acceptable level. Variation of the threshold of individual pixels, caused by transistor mismatch, voltage drops or preamplifier gain variations can lead to an increased noise hit rate or to a reduced sensitivity. In the Superpix0 front-end chip, special care has been used for the design of the discriminator, in particular for the device dimensions in order to minimize the mismatch between transistors. In the second prototype chip a local threshold tuning system is included in each pixel to compensate mismatch induced variations. The discriminator, shown in Fig. 2.16, consists of a differential pair with mirrored active load followed by a gain stage. A small positive feedback is applied to produce a regenerative action, with the advantage of inserting a small hysteresis and thus avoiding re-switching induced by noise. In the following, the transfer function of the discriminator stage is derived in when the inputs are balanced. In this case, if the presence of $M8$ is neglected, the transfer function $v_{int,disc}/v_{in}$ is given by

$$\frac{v_{int,disc}}{v_{in}} = - \frac{g_{m(1,2)}}{g_{ds(1,2)} + g_{ds(3,4)}} \frac{1}{1 + s \frac{C_{int,disc}}{g_{ds(1,2)} + g_{ds(3,4)}}} \quad (2.50)$$

where the body effect is neglected and the transistors $M1$ - $M2$ are considered identical, and so $M3$ and $M4$ are considered equals in pairs

$$\begin{aligned} g_{m(1,2)} &= g_{m1} = g_{m2}, \\ g_{ds(1,2)} &= g_{ds1} = g_{ds2}, \\ g_{ds(3,4)} &= g_{ds3} = g_{ds4}, \end{aligned} \quad (2.51)$$

Then, the transfer function v_{out}/v_{in} is

$$\frac{v_{out}}{v_{in}} = - \frac{g_{m5}}{g_{ds5} + g_{ds7}} \frac{v_{int,disc}}{v_{in}}$$

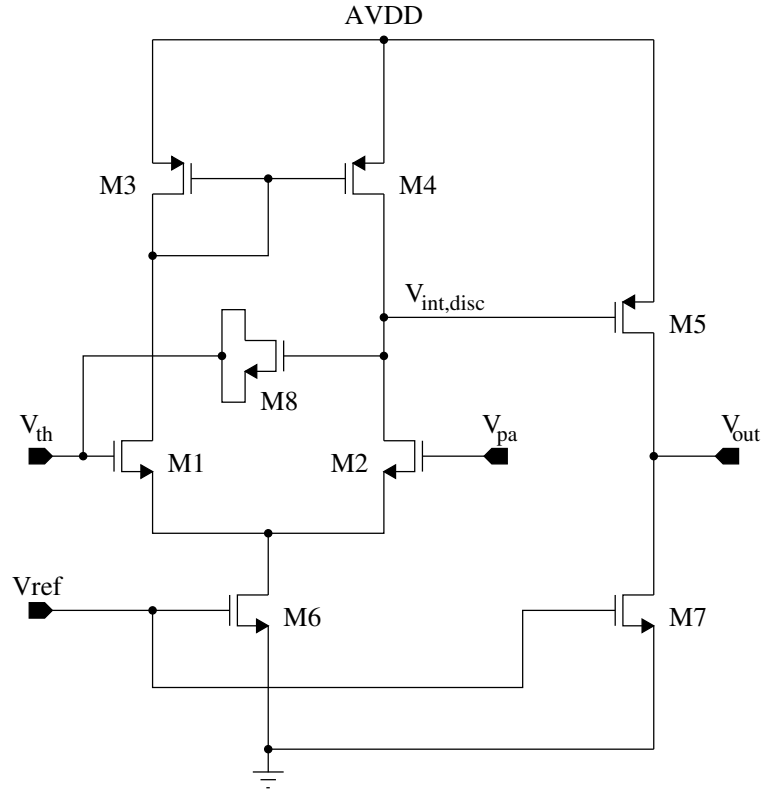


Figure 2.16: discriminator circuit.

$$\begin{aligned}
&= \frac{g_{m5} \ g_{m(1,2)}}{(g_{ds(1,2)} + g_{ds(3,4)})(g_{ds5} + g_{ds7})} \cdot \frac{1}{1 + s \frac{c_{int,disc}}{g_{ds(1,2)} + g_{ds(3,4)}}} \\
&= \frac{A_{O,disc}}{1 + s\tau_{disc}} \tag{2.52}
\end{aligned}$$

where

$$\begin{aligned} A_{O,disc} &= \frac{g_{m5} \ g_{m(1,2)}}{(g_{ds(1,2)} + g_{ds(3,4)})(g_{ds5} + g_{ds7})}, \\ \tau_{disc} &= \frac{C_{int,disc}}{g_{ds(1,2)} + g_{ds(3,4)}}. \end{aligned} \quad (2.53)$$

From circuit simulation $A_{O,disc} = 38.6$ dB and $f_{T,disc} = 1/(2\pi\tau_{disc}) = 720$ kHz have been obtained.

Threshold dispersion analysis and design criteria for the optimum detection efficiency

The Superpix0 chip has been designed for multichannel detection systems performing high performance parallel processing. In the case of binary front-end chains, channel-to-channel non uniformity has to be considered in setting the discriminator threshold in order to obtain the best trade-off between detection efficiency and noise occupancy. The overall effect of such a non uniformity is usually referred to as threshold dispersion, as it can be conveniently represented in terms of a statistical distribution of the discriminator threshold voltage. As far as CMOS processes are concerned, differences in the device threshold voltage V_{TH} and in the current gain factor β are the predominant mismatch sources in closely spaced, identical-by-design MOS transistors. However, under common operating conditions, and all the more so for devices operated close to or in weak inversion (as is the case in the circuits under analysis), effects from mismatch can be safely neglected. In the widely accepted model of V_{TH} mismatch [32], the variation in the MOSFET threshold voltage ΔV_{TH} has a normal distribution with zero mean and a variance $\sigma^2(\Delta V_{TH})$ inversely proportional to the device gate area,

$$\sigma^2(\Delta V_t) = \frac{A_{vth}^2}{WL}, \quad (2.54)$$

where A_{vth}^2 is a proportionality constant obtained from the characterization of a statistically significant number of device pairs. It is generally provided by the foundry together with the device models and depends on the transistor polarity. The main contributions to threshold dispersion come from device mismatch in the charge preamplifier and in the discriminator. The total equivalent threshold dispersion can be written as

$$\sigma(\Delta V_{t,eq}) = \sqrt{\sigma^2(\Delta V_{t,bl}) + \sigma^2(\Delta V_{t,disc})} \quad (2.55)$$

where $\sigma(\Delta V_{t,bl})$ is the contribution resulting from the baseline dispersion at the preamplifier output and $\sigma^2(\Delta V_{t,disc})$ is the equivalent preamplifier output baseline dispersion due to mismatch in the discriminator MOSFET pairs. In the shaperless processor of Superpix0, the input device of the charge preamplifier not only accounts for the main noise source, but also provides the main contribution to the threshold dispersion properties of the system. Contributions from the discriminator, which can be managed independently of the preamplifier design and can in principle be made negligible by acting on the device dimensions, will not be considered here. It can be easily demonstrated

that the effects of the random variation in the preamplifier input offset $\Delta V_{t,in}$, topologically equivalent to the noise source e_n , can be referred to the preamplifier input as

$$\sigma(Q_t) = C_F \sigma(\Delta V_{t,in}) = C_F \frac{A_{vth}}{\sqrt{WL}} \quad (2.56)$$

Therefore the input device dimensions affect not only the noise, but also the threshold dispersion properties of the charge preamplifier. Actually, in a multichannel, binary system, both noise and threshold dispersion characteristics have to be considered in order to determine the discriminator threshold which optimizes detection efficiency. Optimum efficiency is generally constrained by the maximum rate of noise induced transitions at the discriminator, due to the signal output, or maximum noise hit rate, at the preamplifier output randomly crossing the discriminator threshold. Such a noise hit rate limit is strongly dependent on the readout architecture and on the target readout efficiency. Under some very general assumptions, in a binary channel the noise hit rate f_n can be written as

$$f_n = f_{n0} e^{-\frac{Q_{t0}}{2ENC^2}} \quad (2.57)$$

where Q_{t0} is the mean value of the random variable Q_t and f_{n0} is the noise hit rate at zero threshold, i.e., at $Q_t = 0$. If we neglect the effect of threshold dispersion, the minimum input referred discriminator threshold may be set, for all the channels, based on the maximum noise hit rate the system can afford and on the ENC performance of the charge preamplifier:

$$Q_{t,min} = ENC \sqrt{2 \ln \left(\frac{f_n}{f_{n0}} \right)} = ENC \rho(f_{n,max}) \quad (2.58)$$

where $\rho(f_{n,max})$ is an extremely slowly increasing function of the ratio between the zero threshold noise hit rate and the noise hit rate. Due to threshold dispersion, a significant fraction of the channels could exceed the maximum admissible noise hit rate $f_{n,max}$ if the threshold is set according to (2.58). Therefore, $Q_{t,min}$ should be suitably changed to take into account threshold dispersion effects:

$$\begin{aligned} Q_{t,min} &= \rho ENC + \lambda \sigma(Q_t) \\ &= \rho ENC + \lambda C_F \frac{A_{vth}}{\sqrt{WL}}. \end{aligned} \quad (2.59)$$

If Q_t has a normal distribution with mean value Q_{t0} and standard deviation $\sigma(Q_t)$ and $\lambda = 0$, half the channels will find themselves exceeding the maximum noise hit frequency. In order to keep 98% of the channels above the

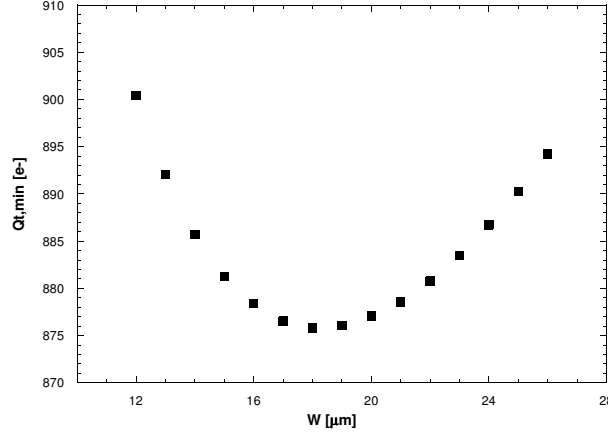


Figure 2.17: simulated $Q_{t,min}$ as a function channel width W of the preamplifier input transistor.

minimum tolerable threshold, $\lambda = 2$ should be chosen. Fig. 2.17 shows the behavior of (2.59) with $\rho = 4$, $\lambda = 2$ and channel length $L = 300$ nm. There is a minimum in correspondence of $W = 18 \mu\text{m}$, which is the chosen dimension.

2.2.3 Injection circuit for the chip calibration

In hybrid pixel detectors, precise analog measurements for every pixel are required to characterize the design before mounting the front-end chips onto the sensor array. It is therefore crucial to foresee internal circuitry for testing purposes. These features can also be used to monitor the chip operation also after connection to the sensor. In the Superpix0 front-end chip a very simple structure has been included in the elementary cell to perform individual channel characterization. The system makes it possible to simultaneously stimulate several pixels in a programmable pattern, so that a realistic signal activity can be generated on the chip. This feature is useful, for instance, to measure whether a significant activity in the digital readout degrades the analog performance by cross-talk. Simultaneous injections can also be used to characterize several pixels at the same time. The injection circuitry is designed very carefully in order not to increase the noise due to the extra components connected to the amplifier input. The test charge is generated by applying a voltage step to small injection capacitors, $C_{INJ} = 10$ fF, located into the pixel. This solution guarantees fairly identical input signals in all pixels, differing only because of the mismatch between the C_{INJ} capacitor. In the ST Microelectronics technology used to implement the Superpix0 hybrid pixel chip, metal-insulator-metal

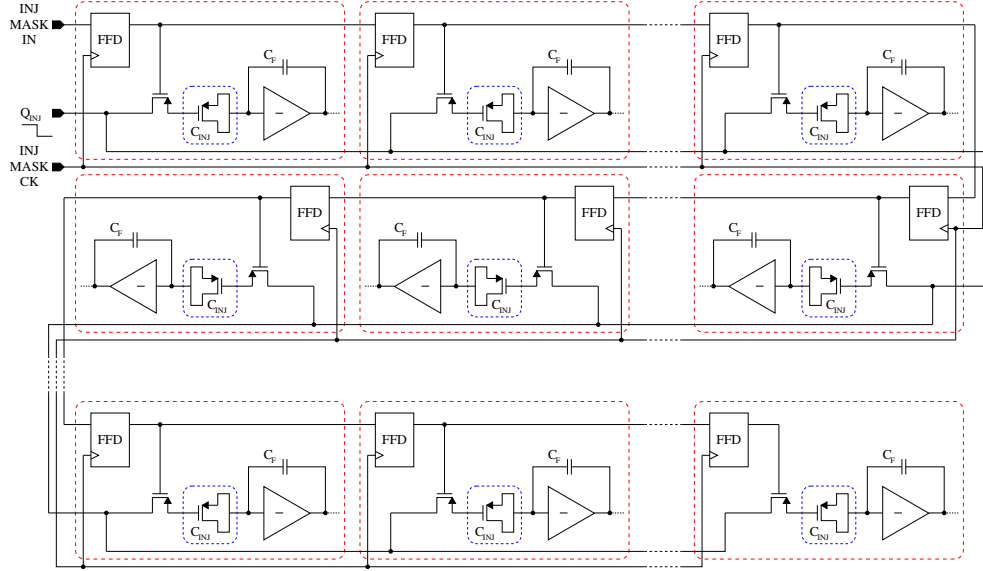


Figure 2.18: schematic circuit of the injection circuit for the chip calibration.

(MIM) capacitors could not be used because incompatible with the bump bonding process used for the connection to the sensor. Therefore, a MOSFET capacitance was used, with its substrate in a deep n-well, in order to isolate the device from the bulk of the pixel circuitry. Fig. 2.18 shows a scheme of the calibration system. The single channel selection is performed by means of a flip-flop implemented in each pixel and representing the cell of a shift register, running along the entire matrix and clocked by the INJ_MASK_CK . The input signal (INJ_MASK_IN , the injection enable signal) of each flip flop is the output signal from the previous pixel. The control signals of the shift register flow in one direction on the even-numbered rows of the matrix, while travel in the reverse direction on odd rows. The injection charge signal (Q_{INJ}) is common to all channels.

2.2.4 Power distribution

The front-end cell uses two power supplies. The analog supply (AVDD) is referenced to AGND, while the digital supply (DVDD) is referenced to DGND. Both supplies have a nominal operating value of 1.2 V. The dual power supply solution has been adopted to avoid voltage fluctuations on the supply lines, due to digital blocks, possibility degrading the performance of the charge pream-

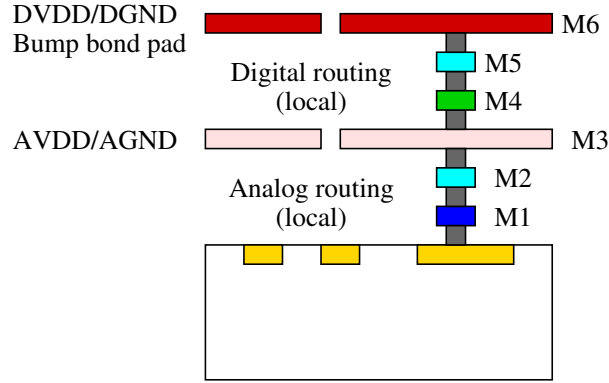


Figure 2.19: Cross sectional view of the metal layers used in the layout of the SuperPIX0 chip.

plifier. Therefore, the threshold discriminator and the voltage references are connected to the analog supply. The in-pixel digital logic is instead connected to the digital supply. The substrate of the transistors is connected to a separate net (SUB) and merged to the analog ground at the border of the matrix. The SuperPix0 chip has been fabricated in a six metal layer technology. Two levels of metal have been used to route the analog signals, two for the digital ones and two for distributing the analog and digital supplies. In particular, Fig. 2.19 shows a cross sectional view of the metal layers used to lay out the front-end processor. Metal1 and Metal2 are used to route the analog signals, Metal3 for distributing AVDD and AGND, Metal4 and Metal5 for digital routing. Metal6 is used to distribute the digital supply DVDD and the digital ground DGND. At the same time, Metal3 is used to shield the analog signals from the digital activity and Metal6 to shield the sensor bump bonding pad from the digital lines. Actually, pixel electronics may be particularly sensitive to cross-talk from the digital to the analog section because they are densely packed in the same substrate and located very close to each other. Very small disturbances can significantly degrade the overall performance. A transition in a digital line, corresponding to a voltage step of 1.2 V (in the 130 nm CMOS technology used for SuperPix0), injected through an extremely small parasitic coupling capacitance of only 1 fF, generates a cross-talk charge signal of 1.2 fC or 7500 electrons, which is of the same order of the signal injected in common operating conditions.

2.2.5 Power consumption

The maximum allowed power dissipation in the pixel chip depends on the application. The front-end chips can be cooled relatively easily in systems where there is only one sensor layer in which the particles are absorbed like, for example, X-ray detectors. In this case, the cooling equipment, if placed underneath the front-end chips, does not degrade the system performance. In pixel detectors for particle physics applications, on the contrary, a number of layers of sensors with front-end chips are used to measure several points of the track of particles emerging from the interaction point. The flight path of such particles is influenced by multiple scattering in the material, which therefore must be kept to a minimum. The cooling is simplified if less power must be taken out of the system. As the dissipation of the front-end chips is usually the dominant heat source (other sources are the module control chip, the interface chips, the power supply cables and the sensor itself), their power consumption must be as small as possible. For nominal bias conditions, the analog power consumption is about $2.5 \mu\text{W}$ per channel. The total power dissipation of the chip is about 1 W/cm^2 (for a 160 MHz readout clock) and is dominated by the digital power. The table 2.2.5 shows the current dissipated by single block in the analog front-end.

Analog FE block	Current consumption
Preamplifier	990 nA
Discriminator	$1.03 \mu\text{A}$
Reference voltages	115 nA
TOTAL	$2.135 \mu\text{A}$

Table 2.4: current consumption by single block in the analog front-end.

2.2.6 Digital front-end

The SuperPix0 digital readout architecture is an evolution of the one adopted for the APSEL4D chip [33] [34] and was originally designed to read out matrices of 320×256 pixels and sustain particle rates of 100 MHz/cm^2 . At the base of the readout technique there is a PIXEL_DATA bus, that is switched over the matrix columns in order to perform a sweep over the sensors grid. The whole bus is read in a parallel way in one clock cycle by the sparsifiers. Their role is to identify all the hits read out of the PIXEL_DATA bus with a proper coordinate label (and the indication about the hit time) and store them into an asymmetric FIFO: the barrel. The MacroPixel (MP) structure, a group of pixels with shared interconnections towards the readout logic, has been adopted, but with a different MP shape with respect to [34]. This technique allows a greater pixel density since fewer connections have to be routed over the sensor area. 2×8 pixel rectangles replace 4×4 pixel squares in order to minimize the matrix mean sweeping time (MST) in the presence of hit-clusters as expected in the data. The developed readout logic has been tailored to cope with wide sensor matrices (in particular to a 320×256 pixel matrix) but in this test chip, the sensor is only 128×32 pixels wide. In case of such large matrices, using a column scan strategy, the key point to reduce the average dead time is strictly bound to the mean sweeping time of the matrix. For this purpose, the decision was taken to involve more than one readout instance, dividing the matrix into sub-matrices. A final output stage retrieves data from all the readouts involved and compresses them into a single data stream. Several studies showed that the best way to exploit the parallelism of the structure and reduce the mean dead time is to have a vertical subdivision of the whole matrix and a vertical shape of the MPs. As already said, the Superpix0 matrix, 128×32 pixels wide, is made up of binary pixels, giving a 0/1 kind of information. When a sufficient amount of charge is released within a pixel, the hit information is stored in a digital MOS latch. The time granularity is provided by a dedicated clock, called BCO, which increments the time counter register. This clock determines the time window in which the hits are collected. The readout is responsible for the association of a hit MP to a determined time window. The MP is enabled when the acquisition starts, which means all its latches can be triggered. When a hit (particle or noise) activates at least one pixel of a MP, the MP is considered fired and the FAST_OR signal, which is the logic OR of all the pixel latches, goes high. When the current time window ends, with the arrival of a BCO rising edge, all the currently active and fired MPs get frozen. Each MP has a Latch enable line which, if not active, stops the collection of hits (the not-fired pixel can latch no more, even

if the threshold is crossed). In this way, the hit pattern of a MP is preserved till the readout phase; all the hits of the pattern refer to single and precise time stamps. The freezing logic and the time counter are implemented in the readout block and not into the MP itself, but they have been introduced here for a clearer explanation of the matrix features. The readout of the matrix takes place by columns; a column-wide `PIXEL_DATA[0:31]` bus is driven in turn by the columns of those MPs which contain at least one hit. The readout logic will perform a sweep activating only the columns of the MPs previously frozen. As shown in Fig. 2.20, the three-state outputs of the 16 latches of a MP are controlled by three enable signals: the `OUTPUT_ENABLE` and 2 bits of `COLUMN_ENABLE`. The `OUTPUT_ENABLE` is shared among all the MPs of a row while the `COLUMN_ENABLE` is shared among the pixels in a column. The cross combination of `COLUMN_ENABLE` and `OUTPUT_ENABLE` signals individuates the pixels that are meant to drive the `PIXEL_DATA` bus. To avoid conflicts on the bus, only 1 column of pixels at a time can be enabled, while the full bus width can be exploited activating all the `OUTPUT_ENABLE` lines. A dedicated hardware reset signal for the latches of a MP is not foreseen; each MP is provided with an auto-reset logic that activates after the `COLUMN_ENABLE` sequence “010” when `OUTPUT_ENABLE` is high. The 128×32 matrix is divided vertically into 2 independent 64×32 sub-matrices. Each one has its own and independent readout instance. This means that two identical sets of signals connect the readout and the two parts of the matrix.

2.2.7 Readout architecture

As already mentioned, the readout design was meant to cope with wide matrices, in particular the target was 320×256 pixels. This area is 20 times greater than the Superpix0 test chip (128×32). Once the target dimension was established, a readout strategy was studied which is expected to afford a 100 Hz/cm^2 rate keeping the readout efficiency higher than 98%. This efficiency refers to readout only, no internal pixel dead time is taken into account. It has been evaluated with systematic simulation of the device models, using a BCO signal with a $1 \mu\text{s}$ period and a 60MHz readout clock. The high data throughput generated must be granted by a wide-band external bus. This led to choose the implementation of several instances of readout working in parallel. Since the efficiency drops with the mean sweeping time, it should be kept as low as possible. One solution is to keep the regions to sweep as narrow as possible. That is why the whole matrix was divided vertically into 4 sub-matrices 80×256 pixel wide. All the parallel structures foreseen for the final application are excessive here but, in order to test their simultaneous func-

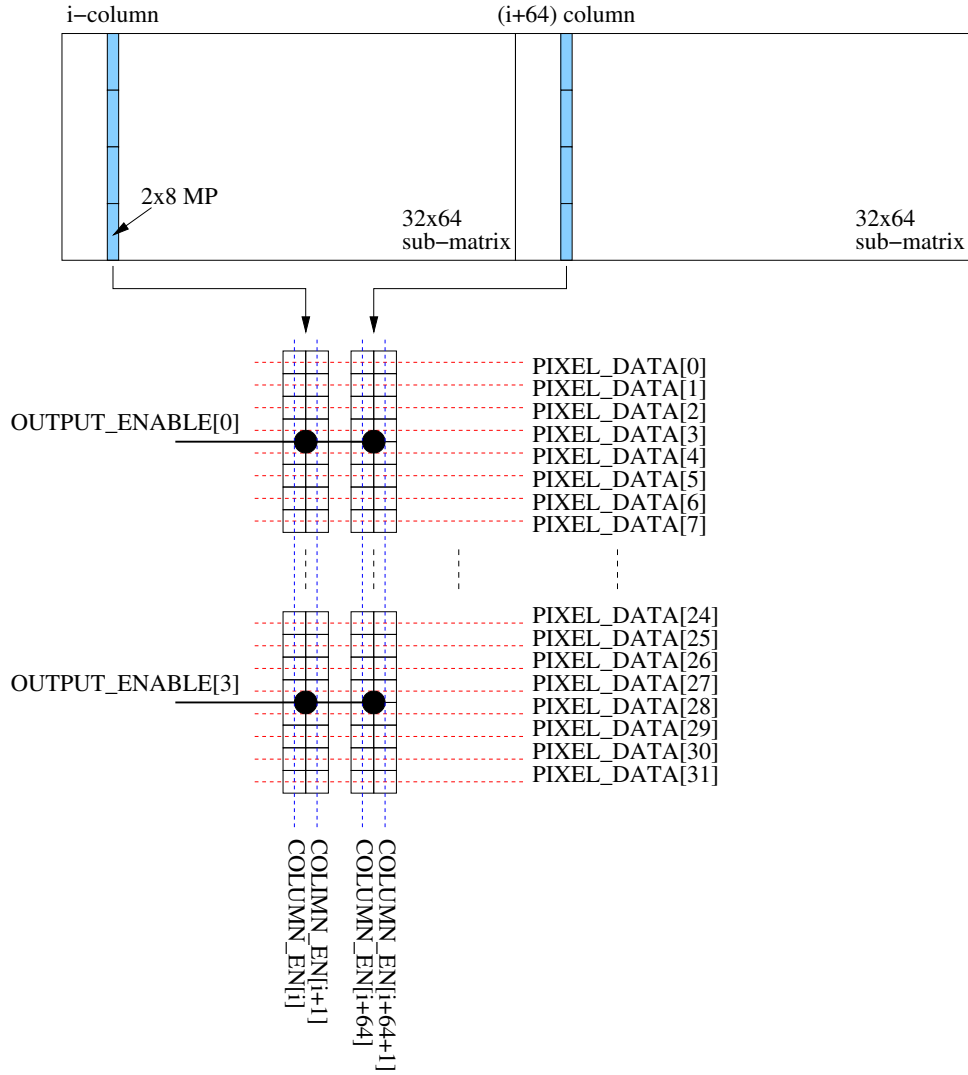


Figure 2.20: schematic representation of the Superpix0 matrix helpful to understand how the digital readout takes place. In the figure the main signals concerning the MP enabling are also highlighted: The PIXEL_DATA bus is driven by those pixels with COLUMN_ENABLE[i], COLUMN_ENABLE[i+1], COLUMN_ENABLE[i+64], COLUMN_ENABLE[i+64+1], OUTPUT_ENABLE[j] enable (where $i = 0 \div 63$ and $j = 0 \div 3$).

tionality, the readout logic is left a bit oversized with respect to the Superpix0 needs. A schematic of the periphery digital logic is shown in Fig 2.21. The pixel data are encoded by the sparsifier elements. The sparsifiers are the elements that read a portion of the PIXEL_DATA bus in order to label the hits with a (x,y,t) coordinate set. Then, they create a formatted list of all the hits found on the PIXEL_DATA bus and write it into a dedicated memory element called barrel. This component is a FIFO memory with multiple write ports (one for each word in the list) and a conventional single output port. A data concentrator controls the flux of data preserving the time-sorting of the hits and puts

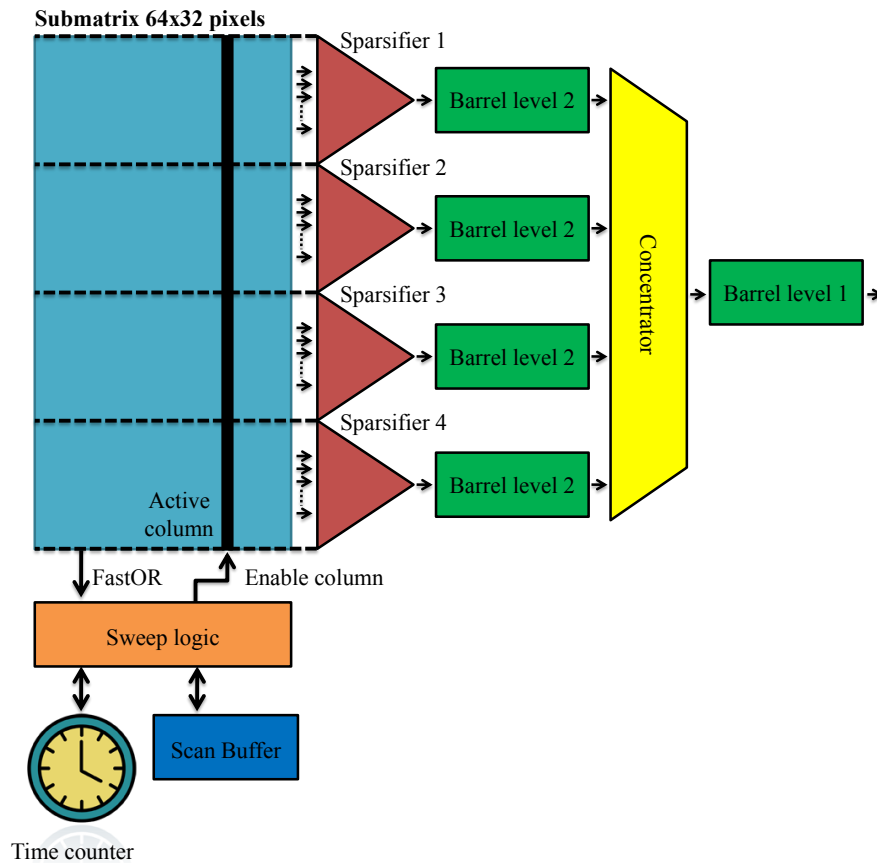


Figure 2.21: readout architecture.

information of coordinates and hit time of MP on the PIX_DATA [0:31] bus.

2.2.8 Layout

Fig. 2.22 shows the complete layout of Superpix0. The pixel matrix, which is a spatially periodic structure, is easily recognizable. Readout circuits, responsible for data management, are located on the left side. Although the matrix of 128×32 pixels is divided into two submatrices of 64×32 cells, the readout electronics of both submatrices has been placed on one side of the matrix. The pads, besides providing the control signals for proper management of the system, bring the supply voltages from both sides, both analog (AVDD and AGND) and digital (DVDD and DGND), to the matrix and the periphery (CORE VDD and CORE GND). Figs. 2.23 and 2.24 show the layout of a pixel, in which the charge sensitive amplifier, the discriminator and the digital front-end are implemented. Both the injection capacitor and the feedback capacitor have been implemented with an NMOS device. In order to isolate them from the rest of the circuit they have been included in a deep n-well. The internal substrate of the deep n-well is connected to the AVDD power supply. The bump bonding pad, clearly visible at the center of Fig. 2.24, with its exagonal shape, has been laid out with the metal 6 layer.

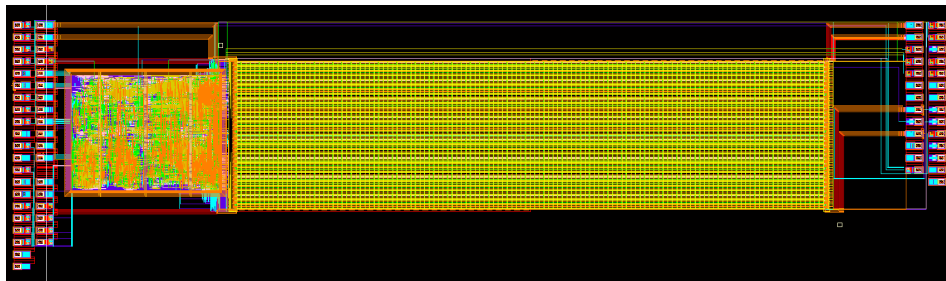


Figure 2.22: Superpix0 chip layout.

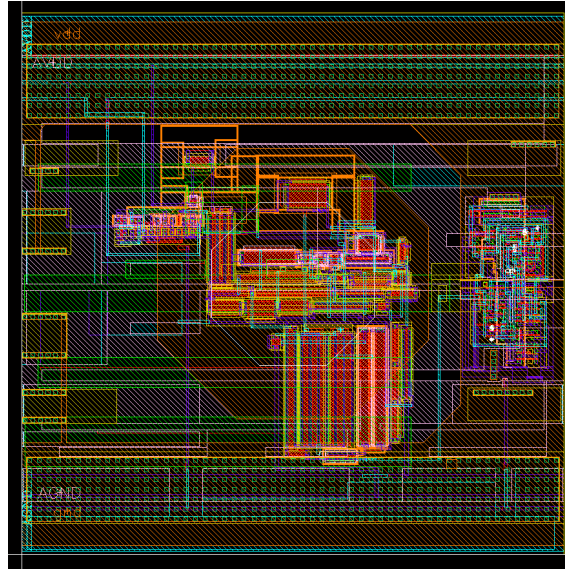


Figure 2.23: pixel layout.

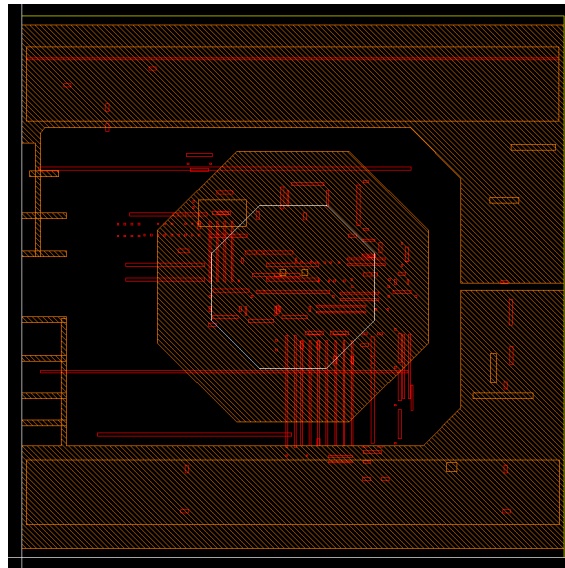


Figure 2.24: pixel layout with only metal 6 layer; the central structure is the bump-bond pad for the connection to the pixel sensor.

2.3 Characterization results

The Supersp0 chip, fabricated by STMicroelectronics in a 130 nm CMOS technology, is mounted on a carrier, designed to be connected to a printed circuit board, along with bias circuits, analog buffers and a programmable 12 bit DAC for threshold dispersion measurements. In order to verify the functionality of the front-end chip, preliminary tests have been performed without the sensor connected to the device. The digital signals needed to drive the digital readout of the chip are provided by a Pattern Generator (TPG) in a Tektronix TLA715 mainframe. A Logic Analyzer (TLA) plugged in the same mainframe acquires the digital output signals provided by the buffers on the test board.

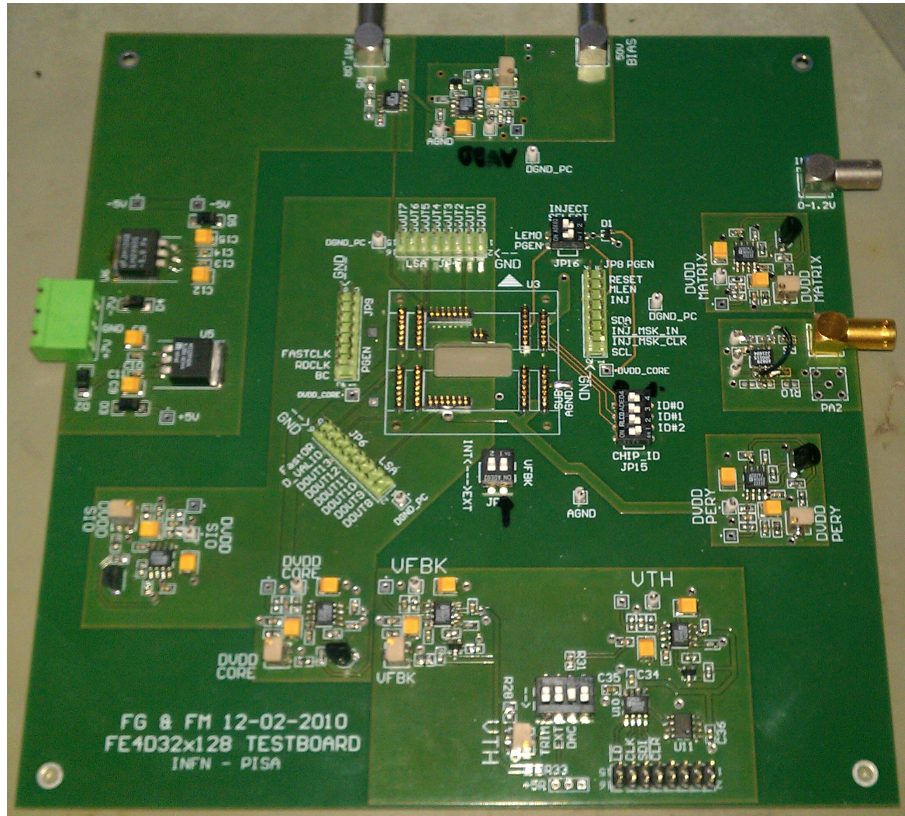


Figure 2.25: the test board used for the characterization of the Supersp0 chips.

In the following, the signals to be controlled by the pattern generator (input) or to be read by the logic analyzer (output) are listed:

- **RD_CLK**: readout clock that feeds the matrix sweep and the first stages of readout (sparsifiers, barrel2, concentrator); it serves also most of the service logic, like the I2C interface, the registers and all the slow control features (mask loading, test cycles...);
- **FAST_CLK**: signal used to transmit on a broad band external bus; it drives the final stages of readout (barrel1, final concentrator);
- **BC**: timing clock; it increments the time counter register and determines the time granularity of the acquisition;
- **RESET**: asynchronous reset, active high; when released (and clocks are running) an automatic matrix cycle starts to reset all the MPs. This operation takes 64 RD_CLK cycles;
- **MASTER LATCH ENABLE (MLEN)**: a global enable signal; if set to 0, all the MPs get frozen until it is re-activated; when set to 1, the MP can be hit and be frozen by the internal logic following the acquisition strategy;
- **GLOBAL_FAST_OR**: signal whose purpose is to export outside the chip the global OR of the FAST_OR of all the MPs; it is intended for debug purposes, not directly useful to a standard acquisition process;
- **SDA**: the I2C-like data line; it is a bidirectional pin which must be connected to a pulled-up line where multiple chips can be connected together to the master;
- **SCL**: timing signal (clock) of the I2C-like transaction; it must be at least 4 times slower than the RD_CLK.
- **Chip_addr[0:2]**: pins that need to be hard-wired to assign a hardware address to the chip. Address 111 is reserved as a broadcast address;
- **DATA_OUT[13:0]**: a data bus for the fast bus, running synchronous on the FAST_CLK; The hits on this bus are coded as shown in Fig. 2.26;
- **DATA_VALID**: bit that indicates if the DATA_OUT bus contains a valid data.

13	12	11	10	9	8	7	6	5	4	3	2	1	0
0	sparsifier		Y	X address						zone pattern			

13	12	11	10	9	8	7	6	5	4	3	2	1	0
1	000			sub-matr.		time stamp field							

Figure 2.26: data compression scheme relevant to the DATA_OUT bus.

The Slow Control Interface is based on an I2C-like environment with a fixed and predefined master-slave hierarchy. The I2C bus is used for read/write operations over a set of registers. All the instructions and the settings are passed to the chip by writing on these registers, such as the pixel mask settings.

2.3.1 Noise measurement with threshold scan techniques

The threshold voltage is set through a 12 bits digital-to-analog converter controlled with the pattern generator. The timing diagram used for the characterization is shown in Fig. 2.27, where the observation window corresponds to the time interval during which the latch is enabled to store an hit event (MLEN = 1). For each value of the threshold voltage, the measurements is repeated until the internal memory of the logic analyzer instrument is full. Fig. 2.28(a) shows an example of firing efficiency of a pixel as a function of the comparator threshold obtained without charge injection. The firing efficiency is zero at low and high threshold values since the latch is edge sensitive. The firing efficiency curve of each pixel is interpolated with the function

$$O(V_{th} - V_{bl}) = 1 - e^{-\tau_{oss} V_0 e^{-\frac{(V_{th} - V_{bl})^2}{2\sigma^2}}}, \quad (2.60)$$

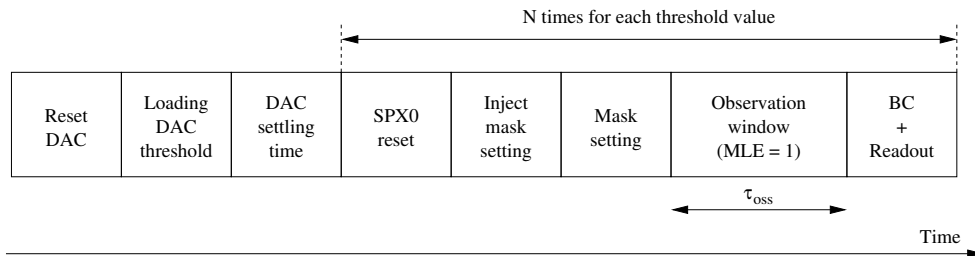


Figure 2.27: Timing diagram

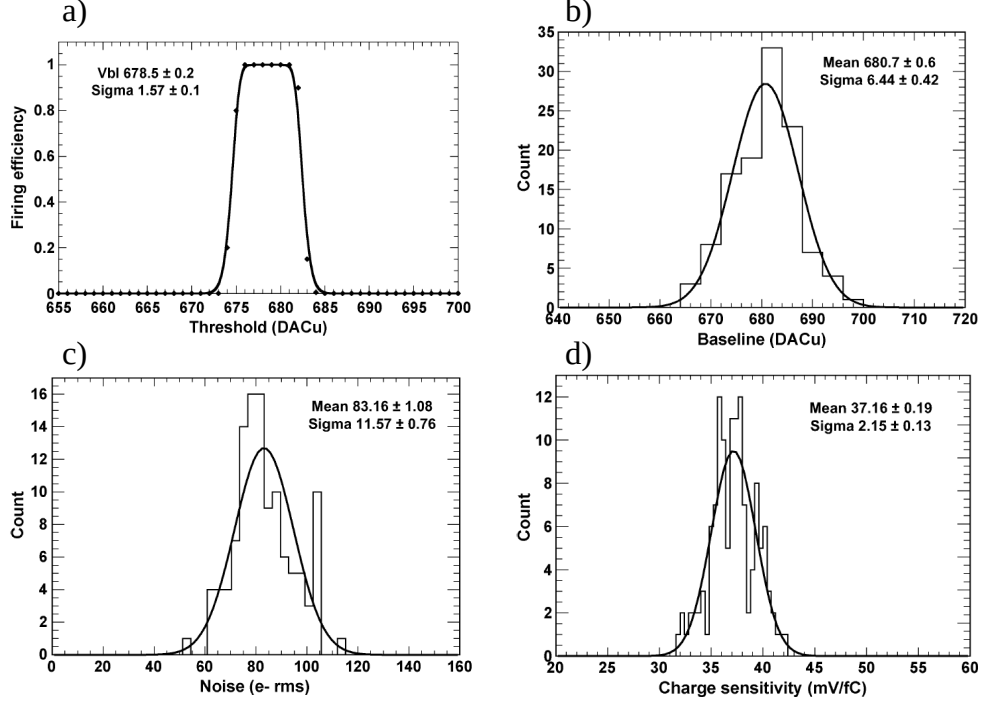


Figure 2.28: Superpix0 characterization (chip 7): a) firing efficiency of a sample pixel; b) dispersion of the DC voltage value at the preamplifier output; c) dispersion of the equivalent noise charge measured on 128 pixels; d) distribution of the charge sensitivity measured on the same pixels.

where V_{bl} is the DC value at the preamplifier output, V_{th} the threshold value, σ the root mean square noise at the output of the channel, τ_{oss} the observation time and ν_0 the noise hit rate at $V_{th} = V_{bl}$. First laboratory tests have revealed an issue in the readout architecture, causing the generation of spurious data on the output stream. The problem was fully understood and a special configuration was found to avoid it already in the current chip version, although at the expenses of a considerable reduction in the area of the matrix that can be used at the same time. In such a configuration, only 128 pixels at a time can be activated. Noise measurements and an evaluation of the threshold dispersion have been performed by measuring the hit rate as a function of the discriminator threshold. With a fit to the turn-on curve we obtain a pixel average equivalent noise charge (ENC) of about 80 e- (with no sensor connected to the front-end chip) and a threshold dispersion of about

Table 2.5: characterization results from a noise threshold scan performed on 3 different Superpix0 samples.

CHIP	Baseline [mV]	Threshold dispersion [mV]	Noise [μ V]
4	208.5	2.38	503
6	207.3	2.19	504
7	207.8	1.98	496

Table 2.6: characterization results from an inject threshold scan performed on 3 DUTs.

CHIP	Charge sensitivity [mV/fC]
4	37
6	38.6
7	37.2

360 e-, which motivated the design of a threshold tuning circuit at pixel level in the new prototype (Superpix1 see chapter 3). The histogram in Fig. 2.28(b) is obtained from the DC baseline values extracted from the fit, the threshold dispersion being the standard deviation of the distribution. Fig. 2.28(c) was obtained from the ENC values again extracted from the firing efficiency curve fit. In table 2.5 the measurement results relevant to 3 chips are reported. The data were obtained from the characterization of 3 macro-columns (384 pixels).

2.3.2 Inject threshold scan measurement

In order to test the operation of the front-end electronics without connection to the detector, measurements with the charge injection circuit have been performed. The absolute calibration of the gain of the chip matrix was performed by using the internal calibration circuit described in section 2.2.3, which enables the injection of charge from 0 to 12 fC in each preamplifier. Fig. 2.28(d) shows the distribution of the gain measured on several pixels of chip 7. The charge sensitivity is on average 37 mV/fC with a standard deviation of about 5%. The average value of the charge sensitivity measured in three Superpix0

chips is shown in Table 2.6.

2.3.3 Measurements with radioactive sources

Both beta (^{90}Sr) and gamma (^{241}Am) radioactive sources were used in order to test the sensor response and the interconnections between the pixel electronics and the sensor. The hit rate as seen from the sensor matrix when exposed to ^{90}Sr is shown in Fig. 2.29. The illumination of the matrix is not uniform due to the collimation of the source. The two blank columns were not scanned due to the aforementioned workaround in the readout scheme. All tested chips showed a very good quality of the interconnections at $50\text{ }\mu\text{m}$ pitch. Only four channels out of more than 8 thousands showed interconnection problems. Continuous data acquisition with pixel thresholds corresponding to less than $1/2$ of the signal released by a minimum ionizing particle revealed another undesired effect ascribable to inter-pixel inductions. However, a solution has been found to profit of the high SNR (200) of the sensor during the beam test, making it possible to reduce the threshold to $1/4$ of a MIP at the cost of a reduction in the acquisition seed.

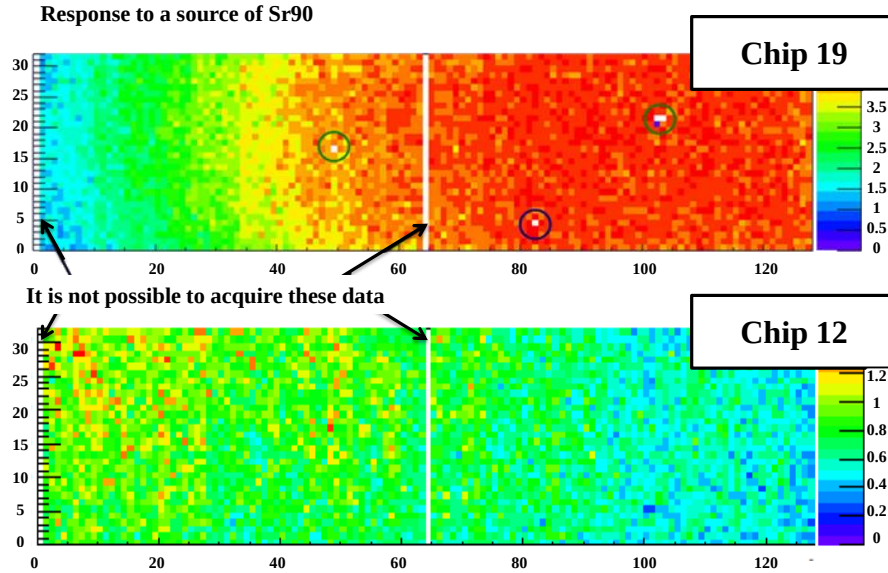


Figure 2.29: Hit rate (Hz) measured on two chips exposed to a ^{90}Sr source.

2.3.4 Beam test results

The beam test was carried out at CERN, at the SPS H6 beam line delivering 120 GeV pions in spills lasting 9.5 s and separated by about 40 s. In the region of the experimental setup the beam was characterized by widths of about 8 mm and 4 mm rms on the horizontal and vertical planes, respectively. As a reference, a six planes telescope of $2 \times 2 \text{ cm}^2$, composed by double-sided silicon strip detector with 25 μm strip pitch on the p-side and 50 μm pitch on the n-side were used [35]. The readout pitch was 50 μm on both sides. Three planes were placed before the devices under test (DUTs), and three after them, at distances of 3.5 cm from each other and either 25 cm or 35 cm from the DUTs, depending on their number: either one or two SuperPix0 chips could be accommodated in the beam-line. All detectors were placed on a custom motorized table with remote control (see Fig. 2.30). The reference telescope was used both to trigger events and determine the impact point of tracks at the DUT. One chip at a time was used to study the dependence of the efficiency on the angle of the impinging particles, whereas either one or two chips were put in the beam line when studying the dependence of the efficiency on the value of the discriminator threshold. For each event, tracks are reconstructed from the silicon telescope hits. Fig. 2.31 shows the functionality of the measurement setup, through a graphic display of fired pixels and the correlation plots between a layer of the telescope and the DUT. The reconstruction algorithm of the tracks relies on the fact that the telescope planes have high efficiency and low noise, and that most triggered events contain just one track with all its related hits, with nothing else but a very small number of noise hits. Adjacent fired strips are grouped in clusters, and the position of each cluster is calculated by weighting the strip positions with their measured charge. The clusters consist of one or two strips, in similar proportions. All possible straight lines connecting the space-points of the two outer detectors are considered, together with the closest space-points in the intermediate detectors. The efficiency of Superpix0 is studied both as a function of the angle of the track with respect to the normal to the detector (angle of incidence θ) and as a function of the threshold used in the pixel charge comparators. In order to vary the angle of incidence, the DUTs are rotated in the xz plane from 0 to 60 degrees in steps of 15 degrees. In runs at normal incidence, pixel charge thresholds are varied from 730 to 820 DACu (1DACu = 305 μV) counts, corresponding to a range from about 12.5% to 40.6% of the charge released by a minimum ionizing particle. The measured efficiency is shown on the left side of Fig 2.32 as a function of the angle of incidence of tracks for the three pixel sensors under test. Data has been collected with a threshold of 770 DACu

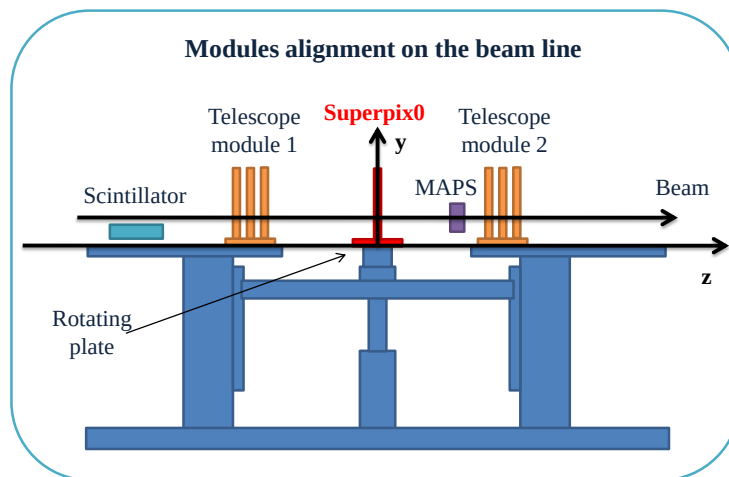
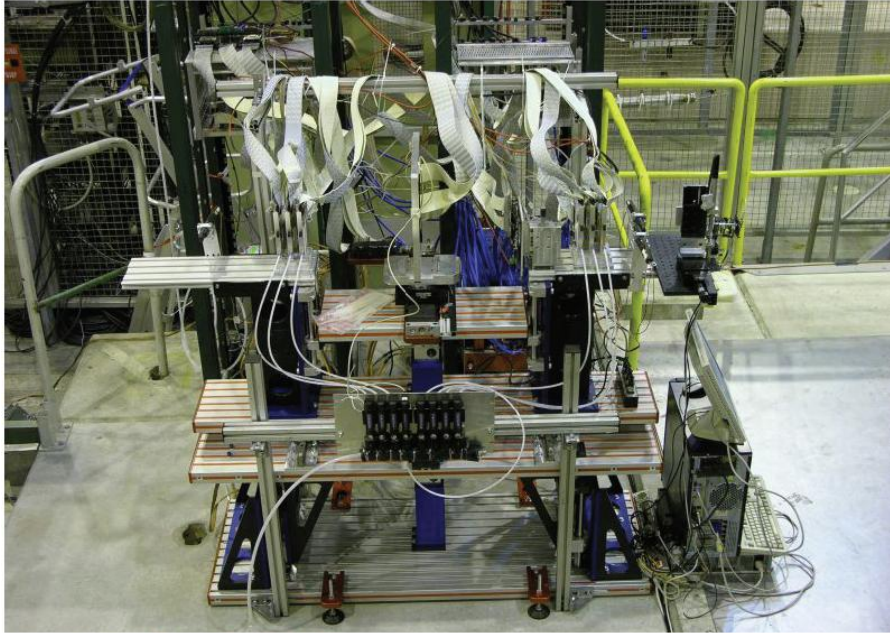


Figure 2.30: modules alignment on the beam line.

counts, corresponding to about 1/4 of the signal of a MIP. The chip efficiency at normal incidence and as a function of the value of the charge threshold is shown on the right side of Fig 2.32. Inefficiencies are uniformly distributed among pixels and no insensitive pixel was found. However the efficiency is found to depend on the distance of the track extrapolation to the center of the closest pixel in the y coordinate. Pixel sensors at the reference threshold, corresponding to 1/4 of a MIP, are close to full efficiency for normal-incidence tracks. More precisely, for this threshold and for smaller ones, an apparent maximum efficiency plateau at about 99.5% is reached. When the threshold is increased above the reference of 1/4 of a MIP, the data indicate a progressive and significant efficiency drop. The analysis also shows that the efficiency decreases considerably for non-zero angle of incidence. For all chips under test,

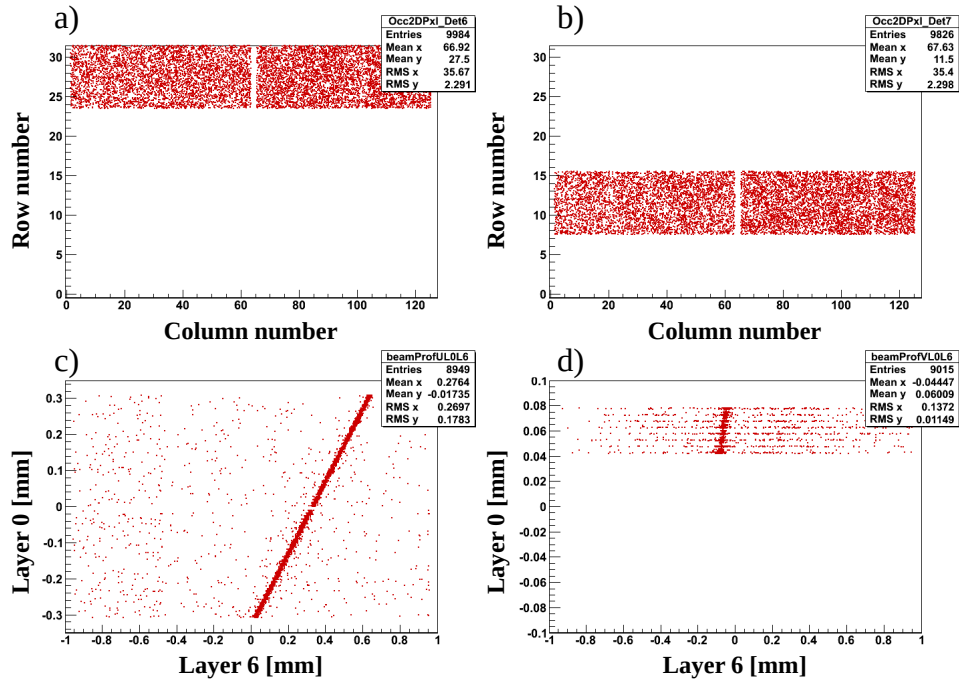


Figure 2.31: a) and b) shows the fired pixels during the 9.5 s observation window in which there is a spill and the Superpix0 active regions at a time; c) correlation between the layer 6 telescope strip columns and Superpix0 matrix pixel columns; d) correlation between telescope strip rows and Superpix0 matrix pixel rows.

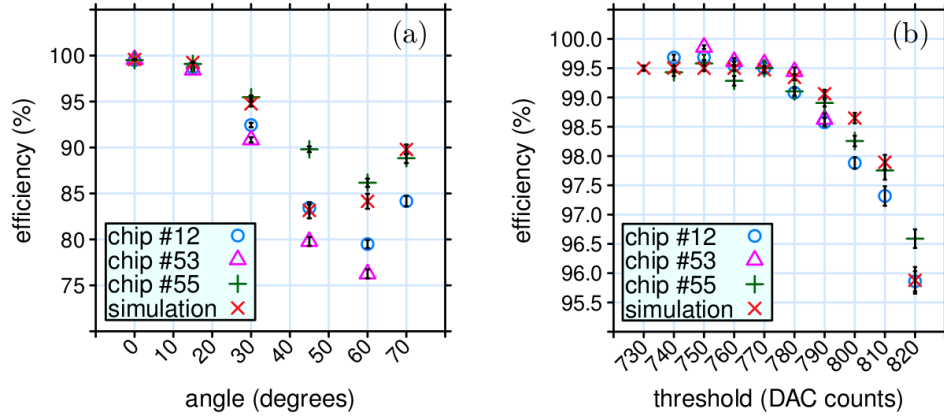


Figure 2.32: Superpix0 efficiency as a function of the angle (a) and of the threshold (b).

the efficiency falls from 99.5% to a minimum of about 80% for 60 degrees angle of incidence, and then moderately increases at 70 degrees. The efficiency drop with the track incidence angle and with the distance from the closest pixel center in the y coordinate is understood as caused by the fact that track ionization in the sensitive area of the detector is shared among a larger number of pixels: along the x coordinate because of the incidence angle, and along the y coordinate as the track hit moves from the pixel center to the pixel border. For increasing angles of incidence, the released ionization also increases because a larger amount of silicon is traversed. However, charge sharing among an increased number of pixels is such that the probability that all involved pixels remain under threshold increases, up to angles of incidence of about 60 degrees. The moderate efficiency increase between 60 and 70 degrees can be explained by the fact that the increase in the number of involved pixels prevails on the moderate reduction of average ionization per pixel.

Chapter 3

Superpix1, a front-end chip for hybrid pixel detector in 3D CMOS technology

This chapter is concerned with the design of a second front-end chip for hybrid pixel detection for applications to the SVT Layer0 of the SuperB experiment. The first prototype, thoroughly described in chapter 2, was produced with a 130 nm CMOS planar technology. For the design of its successor, called Superpix1, a 130 nm, vertically integrated, or 3D, CMOS technology has been used, whose main features are described in section 3.1. In section 3.2, the first experimental results from the test of a 3D DNW MAPS prototype in the same 3D CMOS technology will be presented. The complete characterization of the 3D chips have provided useful information for the design of the analog front-end of Superpix1, described in detail in section 3.3.

3.1 Vertical integration technologies

The development of Integrated Circuit (IC) technology is driven by the need to increase both performance and functionality while reducing power and costs. This goal has been achieved by means of two solutions. The first one involves scaling devices and associated interconnecting wires (process known as scaling down) through the implementation of new materials and processing innovations. The second one consists of new materials and processing innovations, and of introducing architecture enhancements to reconfigure routing, hierarchy, and placement of critical circuit building blocks. A promising solution is the 3D integration technology (also known as vertical integration), in which

multiple layers of active devices are stacked with vertical interconnections between the layers (see the example in Fig. 3.1) to form 3D integrated circuits. Three dimensional integrated circuits have the potential to dramatically enhance chip performance, functionality, and device packing density. They also provide advantages in system architecture and may facilitate the integration of heterogeneous materials, devices, and signals [36]. Multiple 2D-IC circuits can be fabricated in parallel and then assembled to form 3D-ICs. Such an approach enables the performance optimization of each layer and, depending on the application, its functional verification prior to stacking. This should lead to an acceptable yield and reasonable manufacturing cost. 3D-IC fabrication technology can be accomplished by the implementation of different processing sequences. The simplest way to distinguish among various methods is by determining whether the layer stacking was done using a “face-to-face” or “face-to-back” approach [37]. Depending on the position of the top of the second layer with respect to the top of the first layer after stacking, the process can be described as “face-to-face” if the two tops are facing each other, or “face-to-back” if they are not (as shown in Fig. 3.1). A ‘face-to-back’ structure typically has the largest interlayer via dimensions and the lowest via density, along with a more relaxed alignment tolerance. Conversely, the ‘face-to-face’

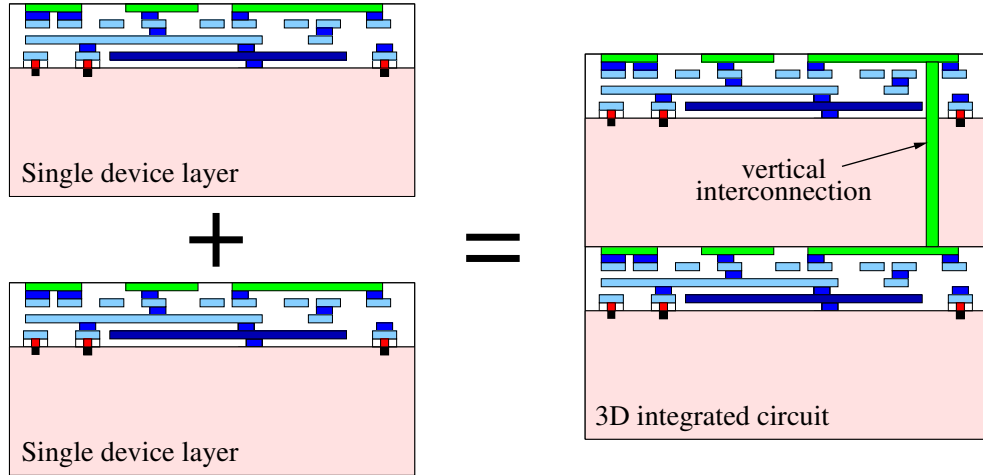


Figure 3.1: schematic representation of a three-dimensional integrated circuit (3D-IC), obtained through the interconnection of two single CMOS wafers, with their metallization levels and inter-device-layer connections (vertical interconnection). The 3D-IC in this example is based on a “face-to-back” process.

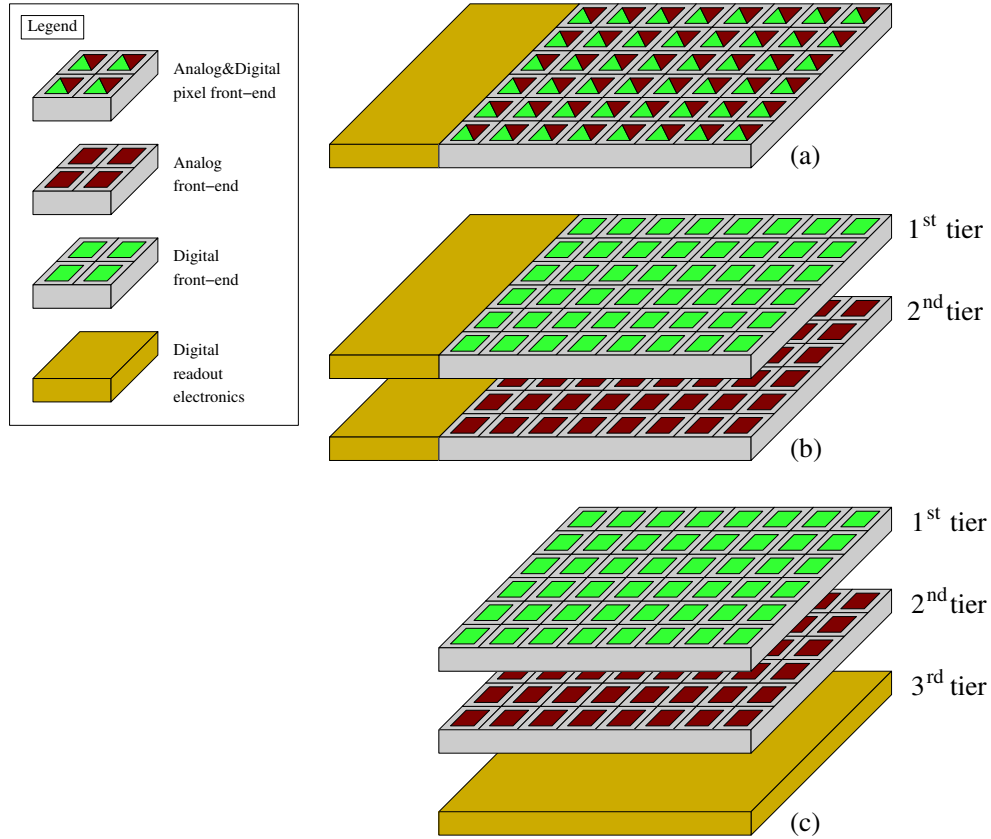


Figure 3.2: breakdown structure of a multiple layer 3D front-end for pixel detectors.

process is characterized by having the shortest distance between stacked device layers, the highest interconnection density, and extremely aggressive wafer-to-wafer alignment requirements. The choice of structure and fabrication method depends on the specific goal and application of the 3D-IC technology. Independent of the final 3D-IC structure, the assembly method always involves the integration of four key technology areas: thinning of the wafers, inter-device layer alignment, bonding, and inter-layer contact process. An additional challenge in achieving high density I/O signal through the layer stack arises from thermal mismatch between the bonded layers, affecting alignment tolerance. Also, thermal dissipation of high performance CMOS devices is already a concern in 2D ICs; for 3D circuits, heat spreading and self heating become critical issues. Actually, vertical integration technologies allow the designer to cope

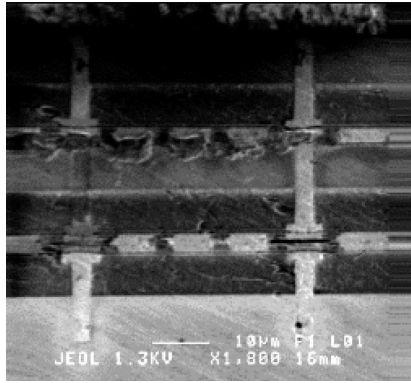
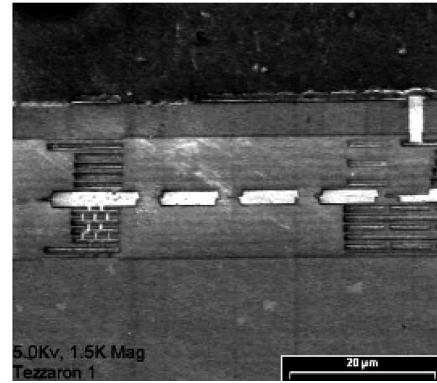
with the increasing spatial resolution requirements set by High Energy Physics (HEP) applications. Although the existing pixel chips use different geometries, readout philosophies and analog circuits, several building blocks and properties are common to most part of the designs. As illustrated in Fig. 3.2(a), the chip sensor can be divided into an active area, containing a repetitive matrix of identical pixels, and the chip periphery, from where the active part is controlled, data is buffered and global functions common to all pixels are located. With the use of a 3D vertically integrated technology, active devices can be stacked and the size of the chip footprint can be reduced. This choice adds a third dimension to the conventional two-dimensional device layout, improving the packing density, since circuit components can be stacked on top of each other. Fig. 3.2(b) illustrates the pursued strategy, which consists of dividing, in a double layer 3D process, the analog front-end from the digital front-end of each single pixel. The immediate advantage is a virtual doubling of the silicon area available, which can be used to implement additional features. A larger area is also available for the peripheral electronic readout. In Fig. 3.2(c), the case of a 3-layer process front-end chip is proposed. With this solution, peripheral area for the digital readout electronics is no longer needed and four-side buttable chips with no, or very small, dead area in the periphery can theoretically be fabricated.

3.1.1 The Tezzaron/Globalfoundries technology

The 130 nm CMOS process chosen for the hybrid pixel front-end chip design described in this chapter is provided by Chartered Semiconductor, now part of Globalfoundries. Vertical integration of the wafers is manufactured by Tezzaron Semiconductor. Wafers are face-to-face bonded by means of thermo-compression techniques. Inter-tier bond pads on each layer are laid out on the copper top metal layer and provide electrical connection between devices integrated in different tiers. Tezzaron has developed two types of vertical interconnect. The size, pitch, and parasitics of Tezzaron vertical interconnects are given in Table 1. The first-generation super-via process had the advantage of being applied to wafers after they were completely processed at a vendor fab. The disadvantage was in the required via size and the need for keep out areas in all layers. The second-generation super-contact process requires a new process module at the vendor fab. This module has proven relatively easy to add and does not introduce any new materials at the stage where the contact is added. Images of Tezzaron wafer stacks implementing the two types of interconnect are shown in Figs. 3.3 and 3.4. With the last generation contact, one of the two layers is thinned down to about 12 μm in order to expose Through Silicon

Table 3.1: main features of Tezzaron interconnections.

	First generation super-via	Second generation super-contact	Face-to-face
Size [μm]	4 x 4	1.2 x 1.2	1.7 x 1.7
Minimum pitch [μm]	6.08	< 4	2.4
Feed through capacitance [fF]	7	2-3	≈ 0
Series resistance [Ω]	< 0.25	< 0.35	≈ 0

**Figure 3.3:** three wafers, stacked and connected with Tezzaron first generation super-via interconnect process [39].**Figure 3.4:** two wafers, stacked and connected with Tezzaron second generation super-contact interconnect process [39].

Vias (TSVs), used to connect bonding pads to the buried circuits [38]. Note that the bottom wafer retains its complete thickness during the stacking process. After the stack is completed, the thick bottom wafer can be thinned and finished with standard wire bonding or flip-chip assembly for bump bonding connections. The Tezzaron process merges all wafers into a single stack. Vertical through-wafer connections are made directly through each substrate to the next wafer and its layer of transistors. The density of interconnect depends strongly on the accuracy of alignment. Tezzaron process can achieve alignment of less than a micrometer. Because of the greater alignment accuracy and higher degree of surface planarity, wafer-level stacking supports a low cost per connection and better interconnect density than flip-chip techniques. Millions of vertical connections can be made at only micrometers of spacing. The use of mixed substrates is limited only by the process temperatures. Also, high temperatures can cause misalignment due to non equal expansion. In wafer-level

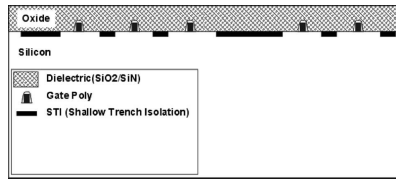


Figure 3.5: cross sectional view of a wafer, immediately after transistors have been created, but before contact metal formation.

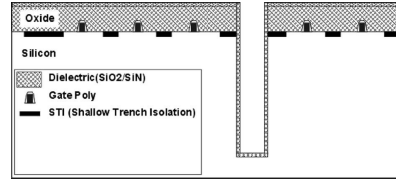


Figure 3.6: the vertical Super-Contact is etched through the oxide and into the silicon substrate. The walls are covered with SiO_2/SiN .

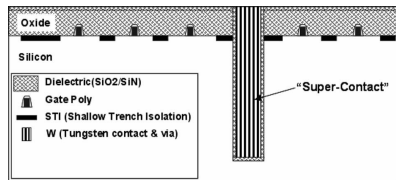


Figure 3.7: the Super-Contact is filled with tungsten and finished with chemical mechanical polishing (CMP).

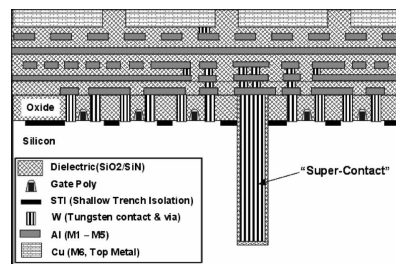


Figure 3.8: the wafer is finished with metal layer formation (aluminum and copper). The last layer must be copper.

stacking, all processing is done at the wafer level. Wafer handling equipment protects against static discharge, so designs do not need I/O buffering between the layers. Another advantage is that wafer stacking lends itself to standard lithography and processing techniques. Tezzaron has used copper metal bonding to produce fully functional wafer-level stacking. Metal bonding provides electrical interconnection, but bonding is done at 400 Celsius degrees, making alignment more difficult, especially with mixed substrates. Figs. 3.5 - 3.13, illustrate Tezzaron stacking method with Super-Contact interconnections and a 3-layer stacking process. Fig. 3.5 shows a cross section of a wafer with active elements, such as transistors, fabricated with the standard planar technology. At this stage, the interconnections between layers are not yet formed. Fig. 3.6 emphasizes the creation of a Super-Contact, whose walls must be insulated from the substrate with a layer of silicon dioxide or silicon nitride. The super contact is then filled with tungsten (Fig. 3.7) and the excess metal is removed with chemical mechanical polishing (CMP). After that, it is possible to continue the wafer production with the metal layer formation (aluminum and copper), up to the maximum level allowed by the technology (Fig. 3.8). The

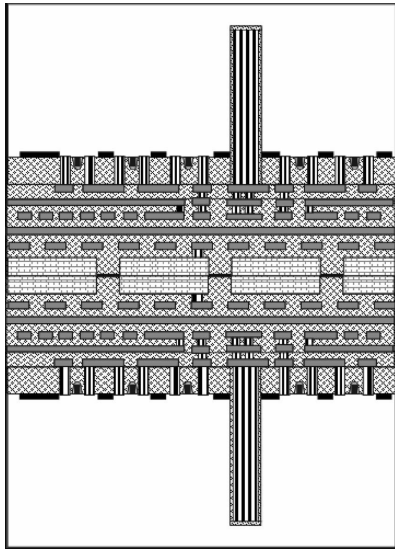


Figure 3.9: the wafers are aligned and bonded in a copper thermal diffusion process that takes place at approximately 400 Celsius degrees.

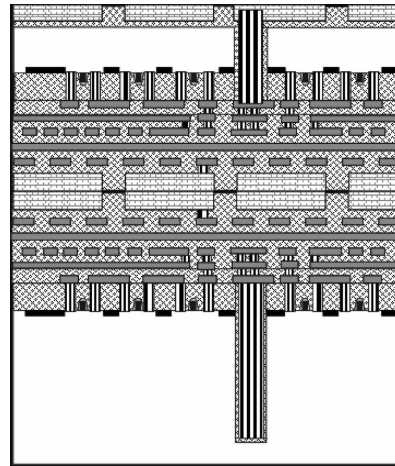


Figure 3.10: the top wafer is thinned and covered by an oxide, then a single damascene copper process creates bonding pads for subsequent stacking.

first and second wafers are then aligned and bonded with thermocompression techniques (Fig. 3.9). After bonding, the top wafer is thinned (Fig. 3.10) to

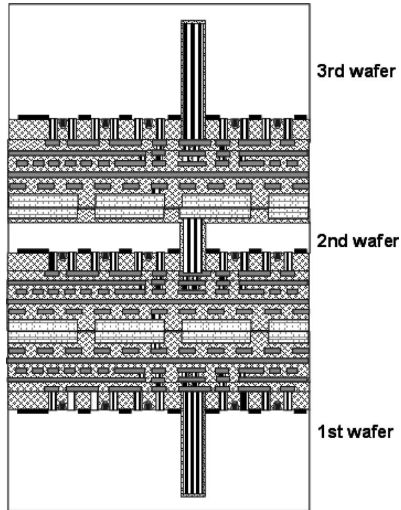


Figure 3.11: a third wafer has been added to the stack, with the same technique of Fig. 3.9.

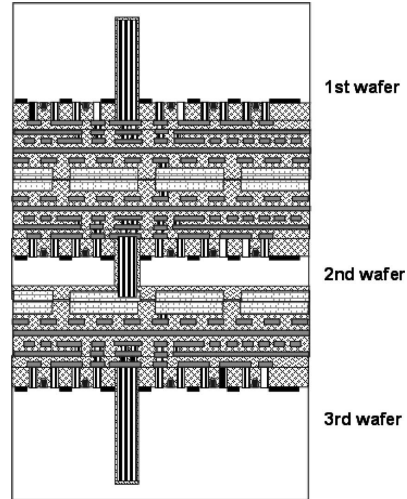


Figure 3.12: the stack is inverted; the first wafer is going to be processed.

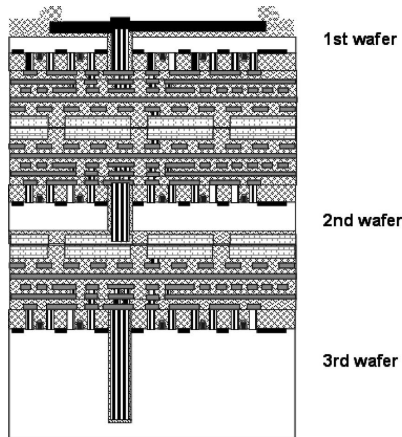


Figure 3.13: the first wafer undergoes the same thinning process used in Fig. 3.13 stopping at the tungsten super-contacts. Instead of a copper damascene process for bonding pads, an aluminum layer is deposited for wire bonding.

the bottom of the super-contacts. This leaves a wafer thickness of about $12\ \mu\text{m}$ (including metal layers and inter-metal oxides). Thinning is done with a combination of wafer grinding, CMP, and etching. The backside of the thinned wafer is covered by an oxide, then a single damascene copper process creates bonding pads for subsequent stacking. Fig. 3.11 shows the addition of a further layer using the same technique used to merge the second layer to the first one. The stack must then be reversed (Fig. 3.12), as the external connections must be to the first wafer. The substrate now located at the top of the stack is thinned. An aluminum layer is finally used to layout the wire bonding pads.

3.2 Characterization of the first prototypes in the T/G technologies

In the first multi-project wafer organized by the 3DIC Consortium, which collected several different designs [43] from a number of institutions and research groups, also the first prototype of a 3D deep n-well (DNW) MAPS was included. Deep n-well monolithic active pixel sensors (DNW MAPS) in planar CMOS technology were proposed a few years ago as a possible approach to the design of monolithic detectors with similar functionalities as in hybrid pixels, targeting applications to the SuperB and the ILC vertex detectors [44], [45]. A DNW MAPS takes advantage of the triple well option available in modern deep submicron CMOS technologies for the design of a relatively large collecting electrode, featuring a buried n-type layer contacted with standard n-wells on its contour and covering a substantial part of the elementary cell. The use of a large sensor enables the design of high performance, fully CMOS analog and digital blocks in the pixel element [44]. This can be done at the expense of a small charge collection inefficiency as long as the sensing electrode area significantly exceeds the area covered by PMOS transistors and by their n-wells. The use of a 3D technology may significantly improve the performance of a DNW MAPS in terms of charge collection properties, readout efficiency, functional density, point resolution and cross-talk between the digital and the analog section and between the digital section and the sensor. Among other benefits, use of two layers makes it possible to improve the electrical isolation between digital and analog sections of the front-end, including the sensor in the case of MAPS, therefore strongly reducing cross-talk issues in a mixed-signal circuit. In MAPS, n-well areas, hosting P-channel devices, compete with the sensor in the collection of charge. In 3D design, the collection efficiency may be improved, placing most of the competitive n-wells in a different tier from the DNW. An evaluation of the Tezzaron/GlobalFoundries 130 nm CMOS tech-

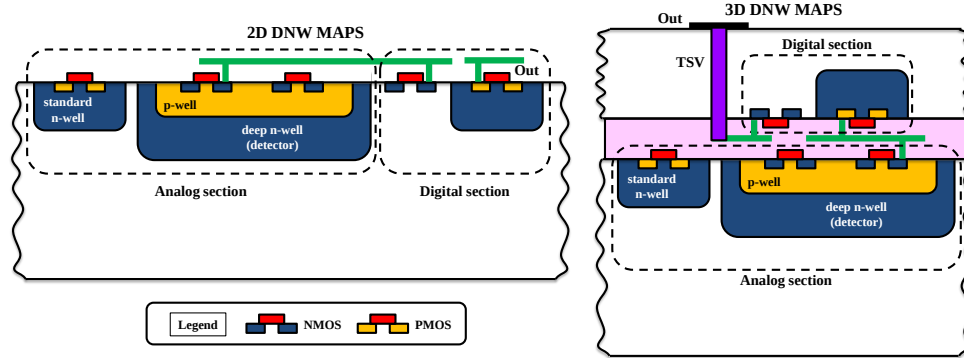


Figure 3.14: cross sectional view of a DNW CMOS MAPS: from a planar CMOS technology to a 3D process.

nology has been carried out through characterization of prototypes produced within the first 3D MPW run. The prototypes include DNW MAPS designed according to the International Linear Collider (ILC) vertex detector specifications [42]. Fig. 3.15 shows the photograph of the vertically integrated chip under test. The visible structures are bonding pads only, that are connected to the internal circuits by means of TSVs. Fig. 3.16 shows a view of the chip side. The thicker tier, containing the analog circuits and the DNW sensors, acts as a support for the 3D structure. The substrate of the digital section, included in the brown region of Fig. 3.16, is vertically interconnected to the first tier and thinned down to about $12\ \mu\text{m}$.

This section presents the first results from the test of 3D DNW MAPS prototypes, which was developed having in mind the requirements of the SVT at the International Linear Collider (ILC) and the SuperB experiments. The activity has been focused on the characterization of the analog front-end. Also, the functionality of the digital readout blocks has been verified.

3.2.1 Devices under test

The chip shown in Fig. 3.17 was designed in the Tezzaron/GlobalFoundries dual-tier 130 nm CMOS vertical integration process. It includes a number of different structures, with different pitch and conceived for different applications. A set of structures, identified as SDR1 (Sparsified Digital Readout 1) are based on the same readout architecture as the SDR0, designed in a 130 nm planar CMOS process for application to the ILC [40].

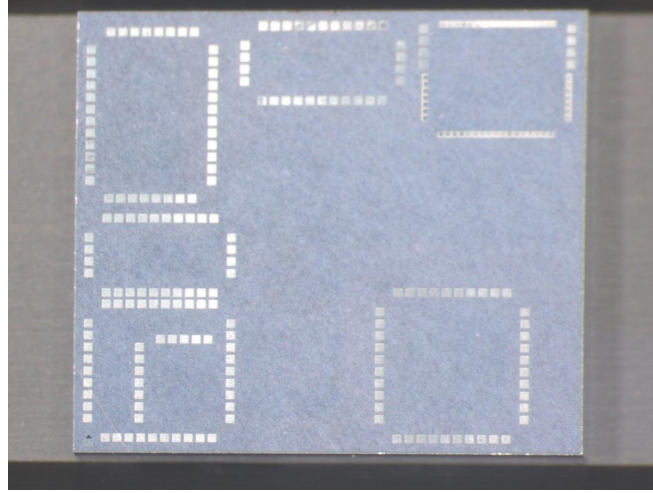


Figure 3.15: photograph of a vertically integrated chip. Pads are connected to internal circuits by means of TSVs.

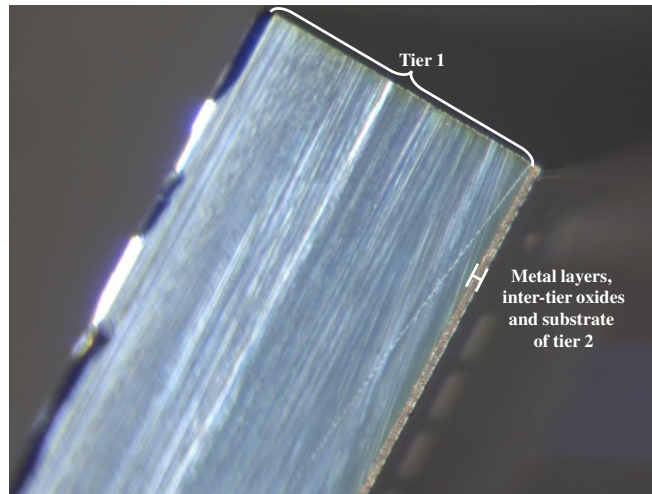


Figure 3.16: side view of the 3D chip of Fig. 3.15. The whole three dimensional integrated circuit is supported by the $675\ \mu\text{m}$ thick analog tier (Tier 1). The metal layers (five for each tier), the inter-metal oxides and the substrate of the second tier (digital layer) are visible through a color change. Wafer thinning techniques based on mechanical grinding and selective etching make it possible to reach about $12\ \mu\text{m}$ thickness for the digital tier.

The main features of the SDR1 structures are listed in the following:

- **single channels:** three different individual channels not connected to the DNW sensor but featuring a MIM (metal-insulator-metal) capacitor shunting the front-end input terminal (100 fF, 150 fF and 200 fF respectively), all with a 60 fF injection capacitance and output terminal available;
- **3 x 3 matrix:** output available for all of the pixels, the central one featuring a 60 fF injection capacitance;
- **8 x 8 matrix:** serial data readout based on a token passing scheme, selectable access to the output of the front-end in each cell;
- **16 x 16 matrix:** the serial data readout is again based on a token passing scheme.

In each SDR1 pixel, the bottom layer includes the deep n-well sensor, whose signal is processed by a charge sensitive amplifier (CSA) placed in the same tier. The CSA is followed by a threshold discriminator, which has been integrated part in the bottom and part in the top tier, as shown in Fig. 3.18. In the elementary cell, the top tier also includes a number of digital blocks taking care of double-hit detection with the relevant time stamping, data sparsification and pixel masking. Readout is based on a token passing architecture. Data sent-off by the pixels (X and Y coordinates and time stamps) during the readout phase are serialized by means of a multiplexer located in the top tier,

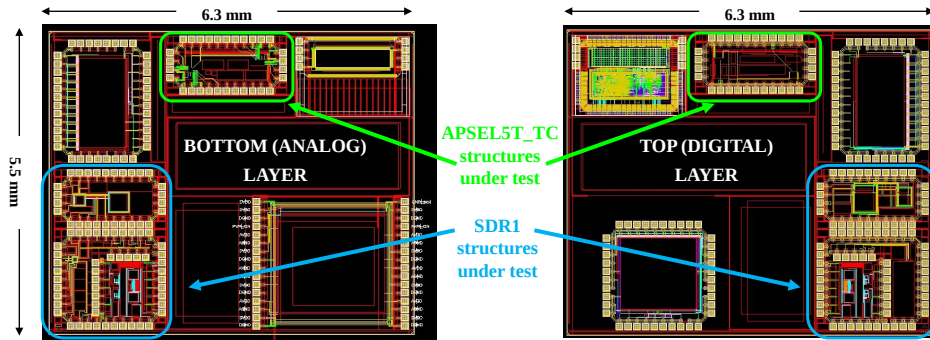


Figure 3.17: bottom (left) and top (right) tiers of the test structures, with different pitch, integrated in the sub-reticule E included in the first MPW of the 3DIC consortium.

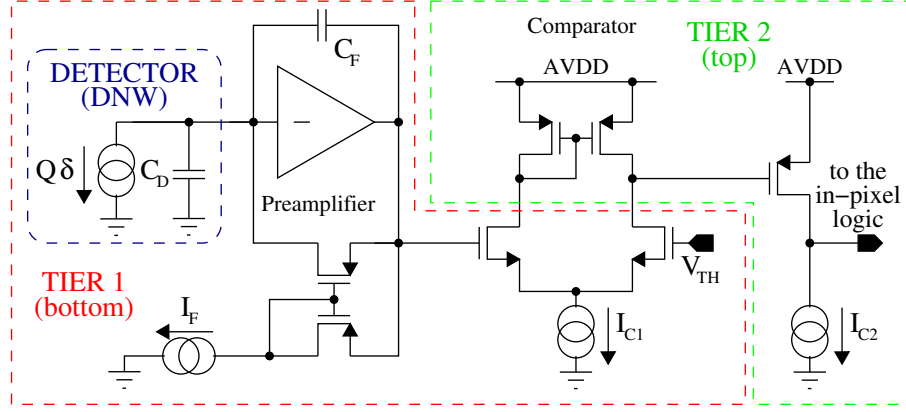


Figure 3.18: block diagram of the analog front-end of the SDR1 and APSEL5T-TC prototypes

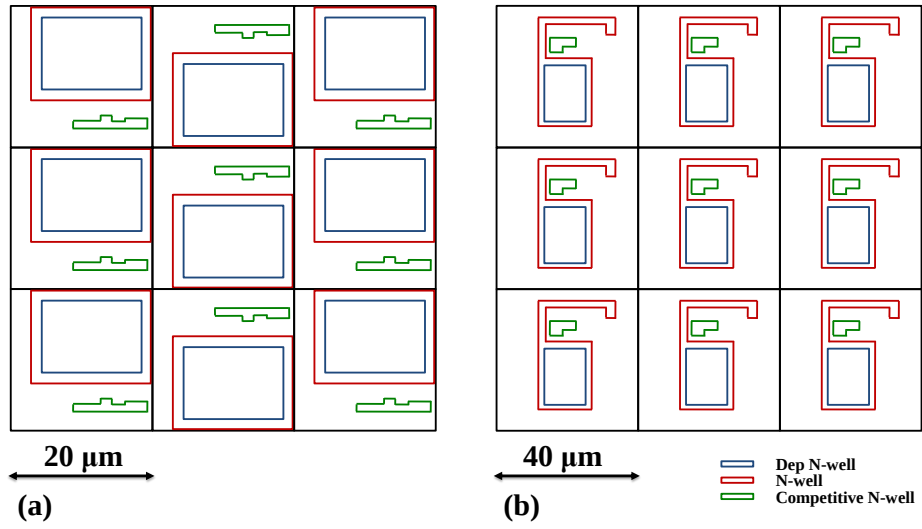


Figure 3.19: layout of 3×3 structures of the SDR1 (a) and the APSEL5T-TC (b) chips.

at the periphery of the 8×8 and 16×16 matrices [42].

Other 3D MAPS structures, of the APSEL type (Active Pixel Sensor ELectronics) [41] developed for application to the SuperB-Layer0 and featuring a $40 \mu\text{m}$ pixel pitch, are included in the chip of Fig. 3.17.

Their characteristics are listed in the following:

- **3 × 3 matrix, called M1:** all analog outputs accessible, injection capacitance ($C_{INJ} = 60$ fF) for the central pixel and preamplifier input transistor with standard layout;
- **3 × 3 matrix, called M2:** all analog outputs accessible, injection capacitance ($C_{INJ} = 60$ fF) for the central pixel and preamplifier input transistor with enclosed layout.

These two structures are called APSEL5T-TC. Also, in the case of the APSEL 3D structure the analog front-end electronics can be described with the diagram in Fig. 3.18. The layout of the collecting electrode, designed with DNW and n-well layers (in red in figure) is described in Fig. 3.19 for the SDR1 (left) and the APSEL (right) DUTs. The n-wells used for PMOS transistors are shown with a green contour.

3.2.2 Characterization of the SDR1 analog front-end

Measurements on 3×3 matrices and single channels proved that the analog section of vertically integrated chips is fully functional. Fig. 3.20 shows the test setup used for circuit characterization, emphasizing the path followed by signals and power lines to reach the circuits or to get out from them through the 3D integrated structure. The signal passes through several TSVs in order to reach the analog circuits placed in the bottom tier. A variable input charge is injected through a 60 fF MIM capacitor (C_{INJ}) in series with the preamplifier input. Figure 3.21 shows the preamplifier output of the central pixel in the 3×3 matrix (Pixel (2,2)). Fig. 3.22 shows the input-output characteristic of the charge preamplifier, which is the peak amplitude of the response as a function of the injected charge. The plot begins to saturate for a 2000 electrons injected charge, in fair agreement with circuit simulations. The linear interpolation of the data points for charge of 2000 electrons yields a charge sensitivity G_Q of about 800 mV/fC. Table 3.2 shows the charge sensitivity of the central pixel of the 3×3 matrix in different chips and the relevant equivalent noise charge. Actually, a significant charge sensitivity dispersion, with values ranging between 550 and 800 mV/fC, was found in the tested samples. This may be ascribed to a large spread in the value of the feedback capacitor C_F . In several test structures, the noise was measured by means of a digital scope. Given a set of N samples, acquired with a sampling period equal to T_s ,

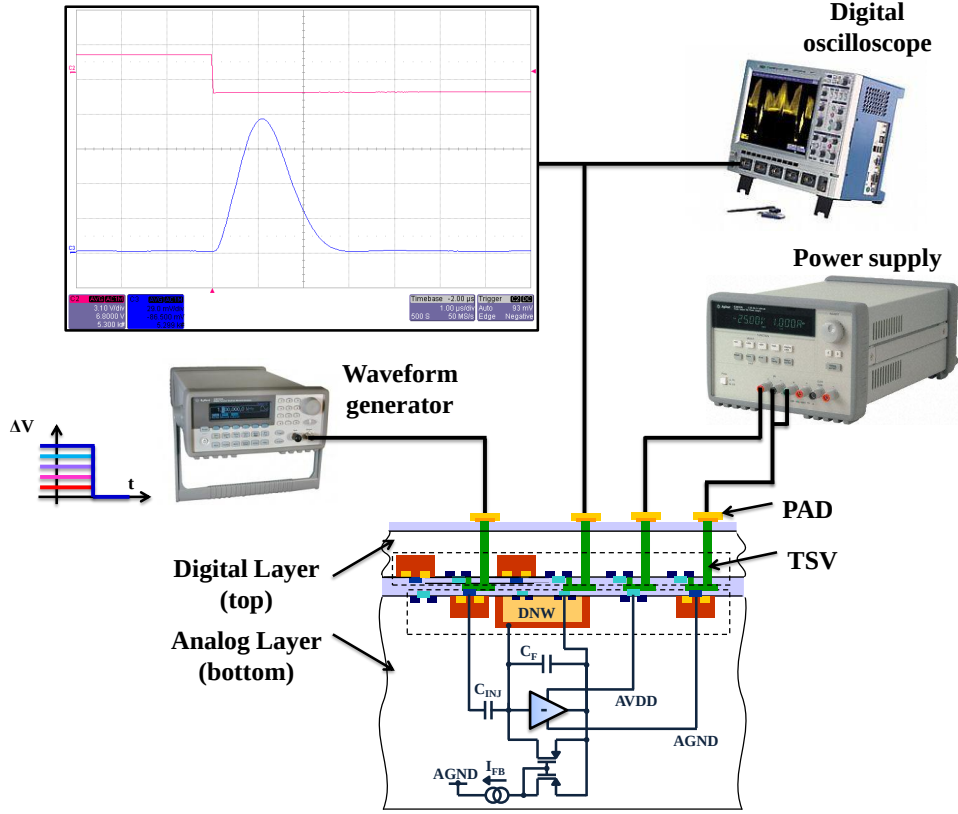


Figure 3.20: setup for the characterization of the analog section of vertically integrated chips.

the rms value $v_{n,rms}$ can be calculated as

$$v_{n,rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N v^2(iT_s) - \left[\frac{1}{N} \sum_{i=1}^N v(iT_s) \right]^2}, \quad (3.1)$$

where $v(t)$ is the signal at the preamplifier output. The noise was referred to the input of the preamplifier by dividing the rms output noise by the charge sensitivity. Fig. 3.23 shows the average ENC measured in three sets of single devices. The ENC is plotted as a function of C_D^* , which includes the detector emulating capacitor shunting the channel input (100, 150 and 200 fF) and the injection capacitance (60 fF). The ENC slope (i.e., the slope of the straight line interpolating the data points) is about 110 e-/pF. From the data of Fig. 3.23,

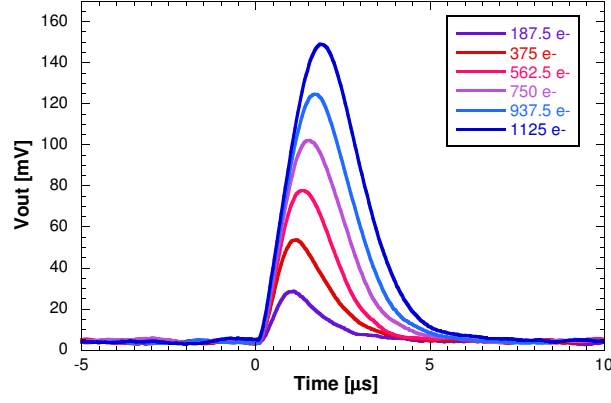


Figure 3.21: response to different injected charge values at the preamplifier output of the central pixel in a 3×3 matrix.

Table 3.2: measurement results for SDR1 3D DNW MAPS prototypes.

	Charge sensitivity [mV/fC]	ENC [$e^- rms$]
CHIP1	550	40
CHIP2	750	35
CHIP3	800	27
CHIP4	620	31

also the capacitance of the DNW collecting electrode can be extrapolated. For this purpose, the average ENC of DNW monolithic sensors (central pixels of 3×3 matrices) from the same chips as the above three sets of single devices was plotted following the trend set by the other data points. The estimated capacitance is in the order of 240 fF (about 300 fF of C_D , which, as already mentioned, also includes 60 fF from the injection capacitor), which, based again on the interpolating straight line of Fig. 3.23, yields an average ENC value of about 30 electrons.

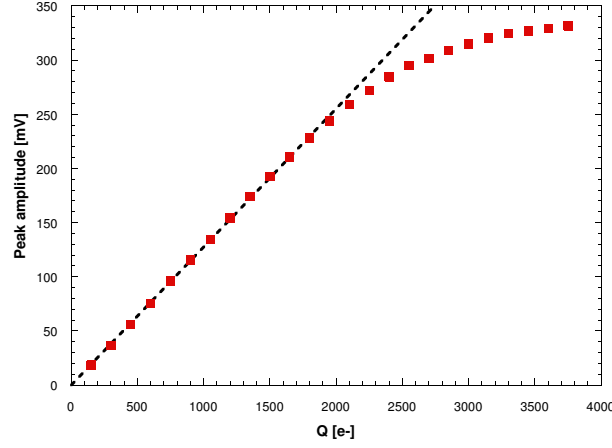


Figure 3.22: input-output transcharacteristic of the charge preamplifier, showing the peak amplitude of the response as a function of the injected charge.

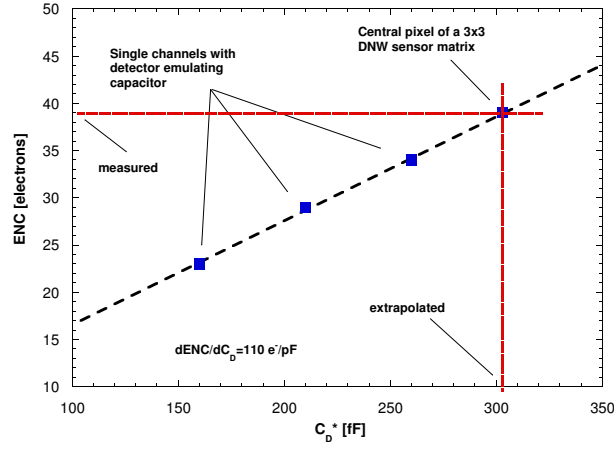


Figure 3.23: equivalent noise charge as a function of C_D^* measured in a DNW monolithic sensor and in three single channels devices with detector emulating capacitor at their input.

3.2.3 Characterization of the SDR1 readout circuits

Tests on SDR1 digital circuits show the full functionality of vertically integrated chips. As already said, use of a 3D process makes it possible to separate the analog signal section and the digital section by relocating them on two different layers. The charge sensitive preamplifier has been integrated on

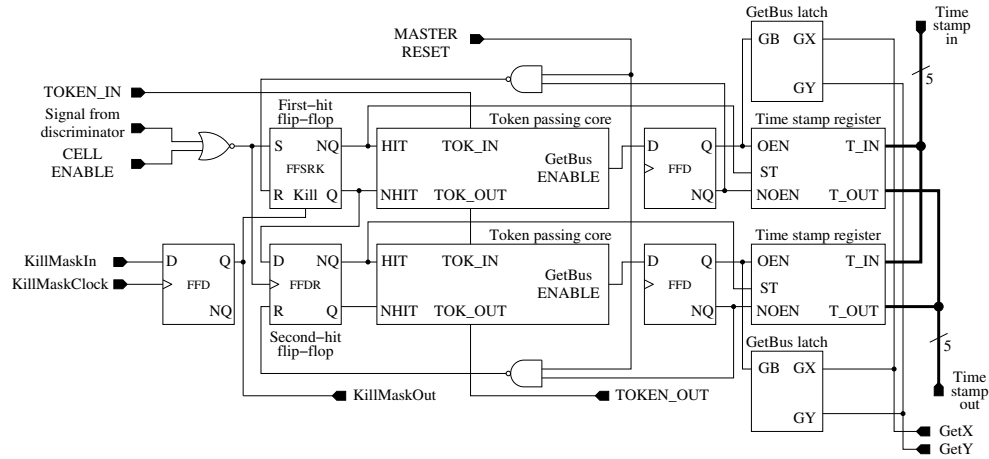


Figure 3.24: block diagram of the SDR1 digital front-end located in the second layer of the 3D chip.

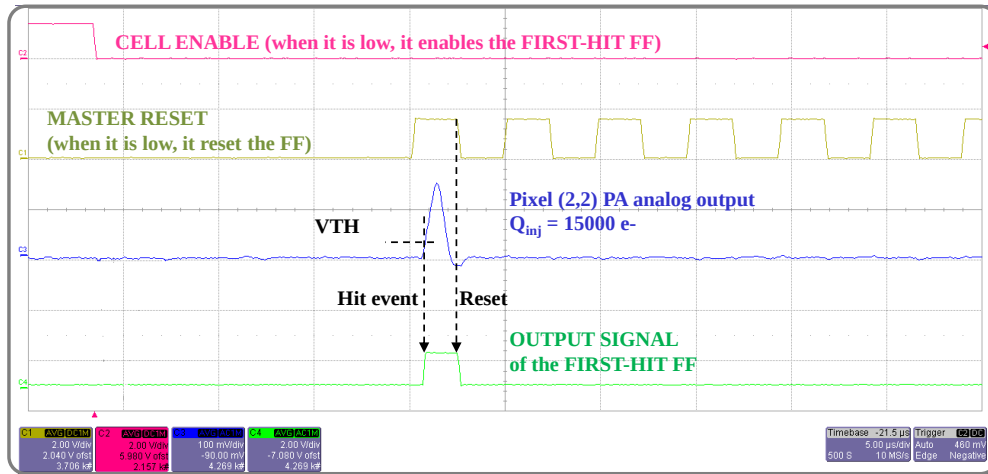


Figure 3.25: control (CELL ENABLE and MASTER RESET) and output (preamplifier output signal and output signal of the hit-FF) signals of a single pixel, displayed on a digital oscilloscope.

the first layer (thick layer), while the second layer (thin layer) contains part of the discriminator circuits and the digital front-end (see Fig. 3.24), featuring double hit detection capabilities. The proof of the interconnection between the two layers has been obtained by monitoring the output signal of the first flip-flop, showed in Fig. 3.25. When the CELL ENABLE signal is low and the

MASTER RESET signal is high, the flip-flop is enabled. In particular, if the preamplifier output exceeds the threshold V_{TH} , the digital output switches to a high logic state (hit event). If the MASTER RESET is activated, the pixel cell is resetted. As already mentioned, among the tested structures, 8×8 and 16×16 matrices are included, whose readout is performed by means of a token passing architecture. For the sake of clarity, it is worth describing here

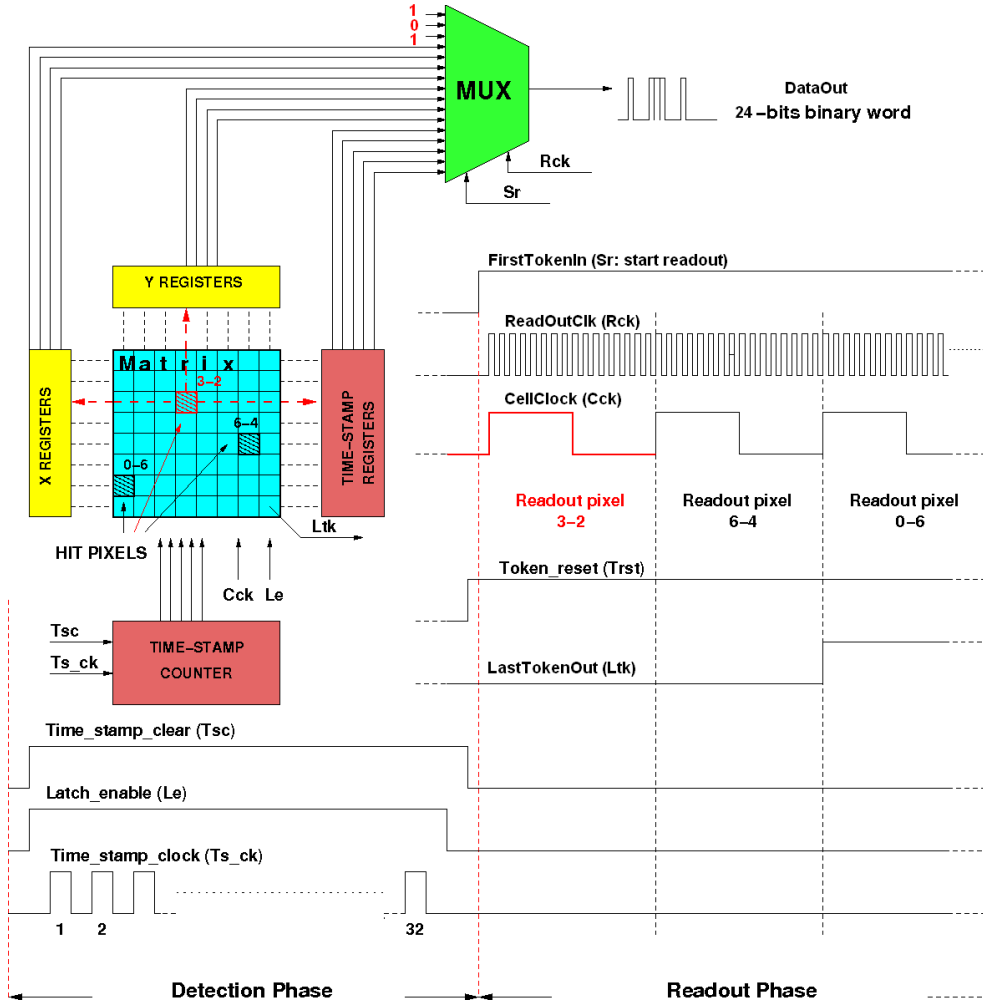


Figure 3.26: block diagram of SDR1 back-end circuits with the time evolution of the digital signals in the detection and the readout phases (in the example pixels store a single hit event).

in more detail the chip readout operation, with reference to Fig. 3.26. In each pixel cell, when a hit event is detected, a bit is set in the pixel and the content of a timestamp register, receiving a timestamp clock signal from a counter in the chip periphery, is frozen. This can happen a second and last time after a second threshold crossing, since each cell is able to detect two (and no more than two) events before being read out and reset. At the beginning of the readout phase, a token is launched across the pixel array and stops in each pixel which has detected at least one event. When receiving the token, the pixel sends off the time stamp (5 bits) and enables a couple of 8-bit registers, containing the row and column coordinates, at the periphery of the sensor matrix (the register size was chosen to be compatible with a larger, 256×256 matrix). The data are sent off chip by a serializer also located in the periphery. The one described here is a selective (sparsified) readout technique, since pixels with no event to report on are automatically discarded. In order to test the digital readout section, a Tektronix TLA715 instrument, performing both pattern generator and logic analyzer functions, was used to provide the circuit with the signal to launch the token (FirstTokenIn) and the clock signals (ReadOutClk, clocking the serializer and setting the bit period in the serial output stream, and CellClk, whose period corresponds to the time needed to readout a single event from a cell). The output signals (DataOut, the data from the hit pixels, and LastTokenOut, the signal indicating that the token has finished scanning the matrix) were read out by means of a logic analyzer. Figs. 3.27 and 3.28 show a set of signals from the characterization of an 8×8 matrix taken at a readout clock frequency (ReadOut signal) of 50 MHz. The interval T_{ro} between the start of the FirstTokenIn signal and the arrival of the LastTokenOut signal, $76.8 \mu s$, corresponds to the time needed to read out two events (each corresponding to a 24 bit word, i.e., 5 bits for the timestamp, 16 bits for the coordinates plus 3 spare bits, hard-wired to the pattern 101) for each of the 64 cells at a 40 ns bit period. In Fig. 3.28, the DataOut signal shows the data relevant to the first three cells of the array under test:

- 101000... (3 spare bits, followed by X=0, Y=0, timestamp=0),
- 101100... (3 spare bits, X=1, Y=0, timestamp=0),
- 101010... (3 spare bits, X=2, Y=0, timestamp=0).

The CellClk period is 24 times the ReadOutClk period ($T_{cellCK} = 24 \times T_{rdCK}$, corresponding to the time needed to read out a bit times the number of bits, 24, in each event). The DataOut and the LastTokenOut signals of the 8×8 matrix have been acquired, for different values of the threshold voltage of the

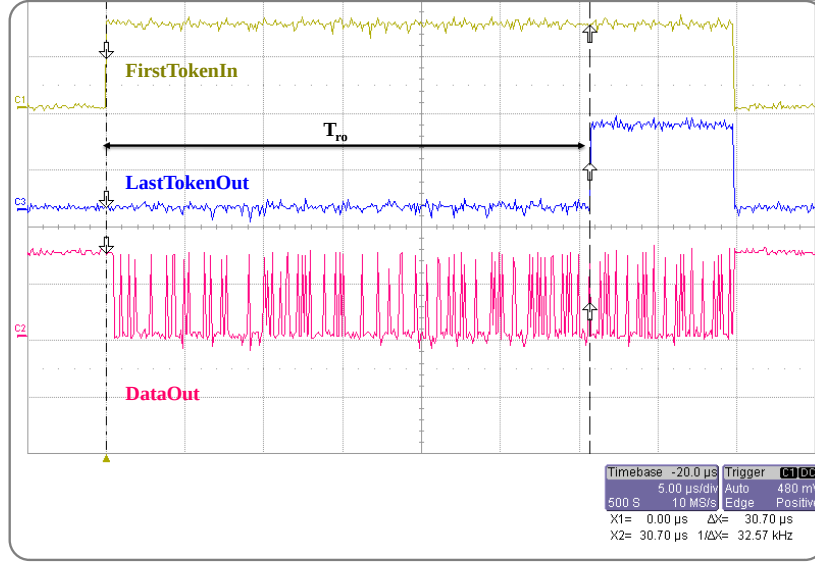


Figure 3.27: digital signals from an 8×8 3D DNW pixel array.

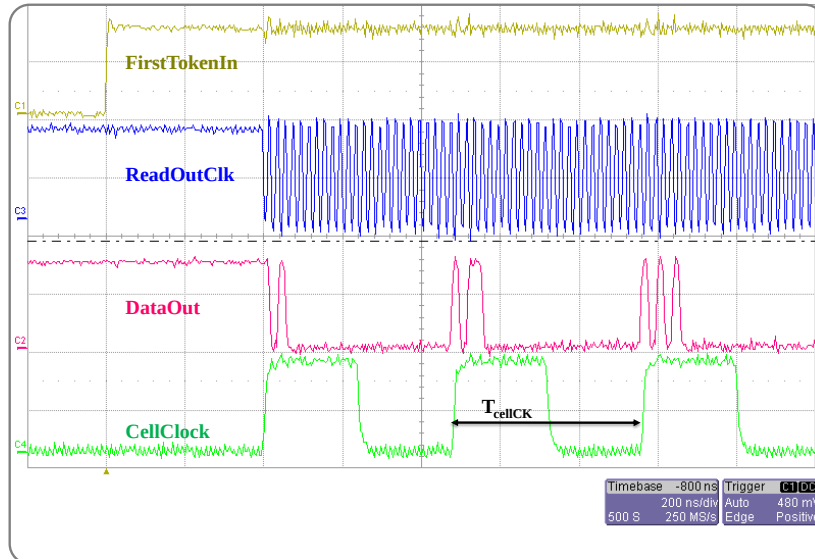


Figure 3.28: focus on the first data after readout start from Fig. 3.27.

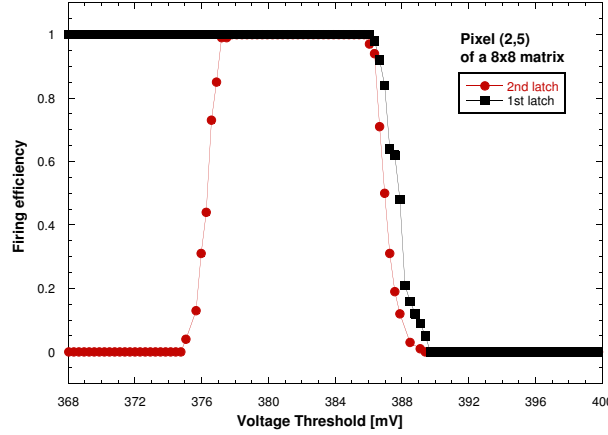


Figure 3.29: firing efficiency of the two flip-flops integrated in the digital front-end as a function of the threshold voltage V_{TH} .

comparator and a $100 \mu\text{s}$ observation window. With a suitable program written in the *root* environment, the occupancy of each pixel has been obtained. Fig. 3.29 shows, as an example, the firing efficiency of the two flip-flops in one of the matrix pixels related to a pixel in the pixel array. Black squares are referred to the occupancy curve of the first hit flip-flop, which is level sensitive, red circles to the occupancy the second-hit flip-flop, which is edge sensitive.

3.2.4 Characterization of the APSEL5T-TC analog front-end

The APSEL5T-TC prototype was designed for application to the SuperB SVT and features a $40 \mu\text{m}$ pixel pitch. As already illustrated, the APSEL5T-TC chip contains two 3×3 matrices, called M1 and M2. The main difference between the two matrices is the layout of the input device of the charge preamplifier, a standard multifinger layout, in the case of M1, enclosed and rad-hard, in the case of M2. Charge sensitivity measurements have been performed by means of charge injection techniques through an injection capacitance connected to the input of the central pixel of the matrices. An average gain of about 250 mV/fC and an ENC of about 40 electrons was found.

Measurements through a ^{55}Fe radioactive source have been also performed on the 3×3 matrices. ^{55}Fe X-rays release their entire energy in the detector substrate through photoelectric interaction. Photons from the ^{55}Fe 5.9 keV line generate about 1640 electron/hole pairs each. When photons convert into the junction depleted region, the released charge is virtually entirely collected, yielding the peak in the ^{55}Fe spectrum. The charge not completely collected

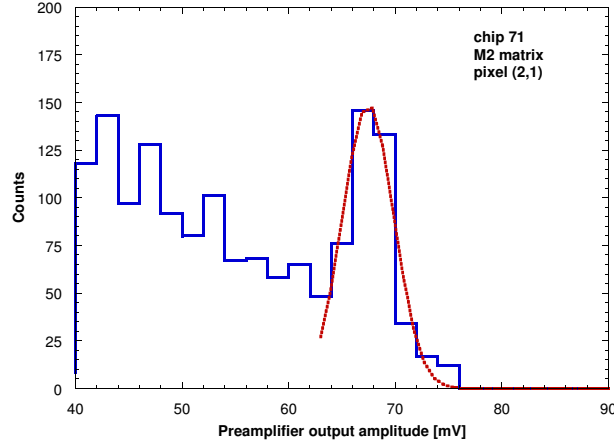


Figure 3.30: event count distribution of a sample pixel in the M2 matrix measured with a ^{55}Fe source.

by the sensor, released far from the junction, is responsible for the pedestal at lower amplitudes. Fig. 3.30 shows the event count distribution for a sample pixel in the M2 matrix. The results from charge sensitivity measurement obtained by means of the ^{55}Fe source are in good agreement with the values obtained with charge injection method. Response of the sensor to charge particle has been also measured on the 3×3 matrix with full analog output, using electrons from a ^{90}Sr source. The measurement setup used for the tests with ^{90}Sr is shown in Fig. 3.31. Electrons released by ^{90}Sr through beta decay have a broad continuous spectrum, with endpoint energy larger than 2 MeV. After detection of an electron, the pulse from a scintillator is used to generate a 500 ns gate signal. If, within this time interval, the signal in the central pixel of the 3×3 matrix exceeds a threshold voltage of $5\sigma_n$, σ_n being the rms noise at the preamplifier output, a trigger is issued and the waveforms at the output of the 9 channels are stored. For each event, the amplitudes of the signals are summed and then stored as a measurement of the charge collected by the cluster of 9 pixels. The signal of the 3×3 cluster is shown in Fig. 3.32. The most probable value (MPV) of the Landau distribution, used to fit the data, is about 1100 electrons and can be assumed to be reasonably close to a MIP signal. Few samples were subjected to a lapping operation, with an Engis Kent3 lapping machine, necessary to remove the roughness on the back of the die. This was done in order to characterize the DUTs with a laser source, a test which would not have been otherwise feasible because of scattering phenomena caused by the irregular surface chip backside. Indeed, in

the test, the DUTs were backside illuminated (i.e. illuminated from the substrate side) to minimize the reflections from the dense metal network located in the frontside region of the chip. The source used for the characterization is a double heterojunction InGaAs/GaAlAs/GaAs Fabry-Perot laser, operating at 1060 nm wavelength. The laser beam is connected to a single-mode operating

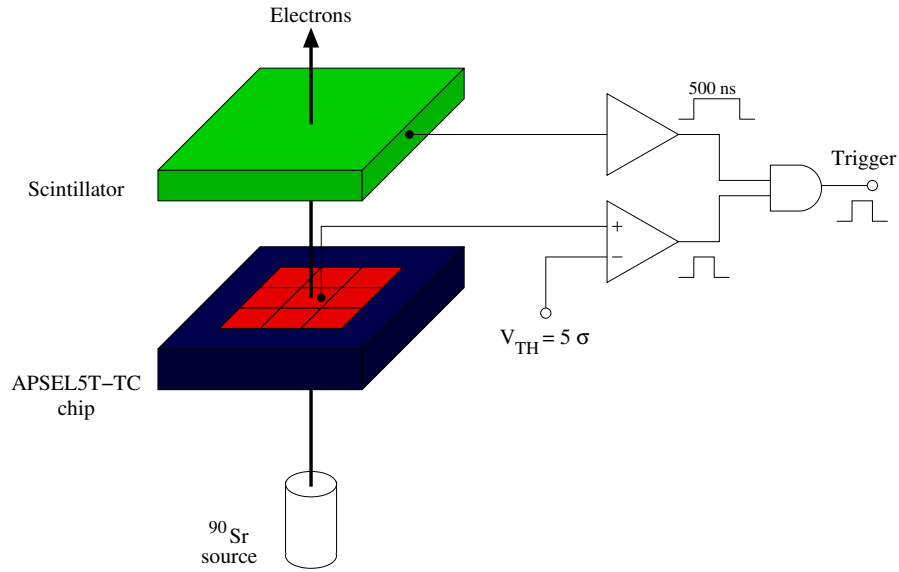


Figure 3.31: measurement setup for the characterization of the APSEL5T-TC chips with a ^{90}Sr source.

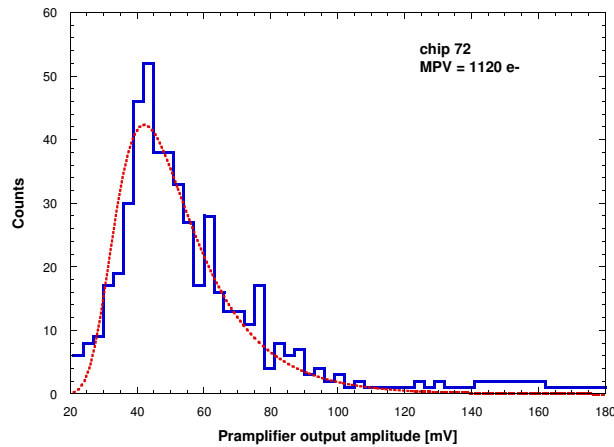


Figure 3.32: event count in a 3×3 matrix tested with a ^{90}Sr source.

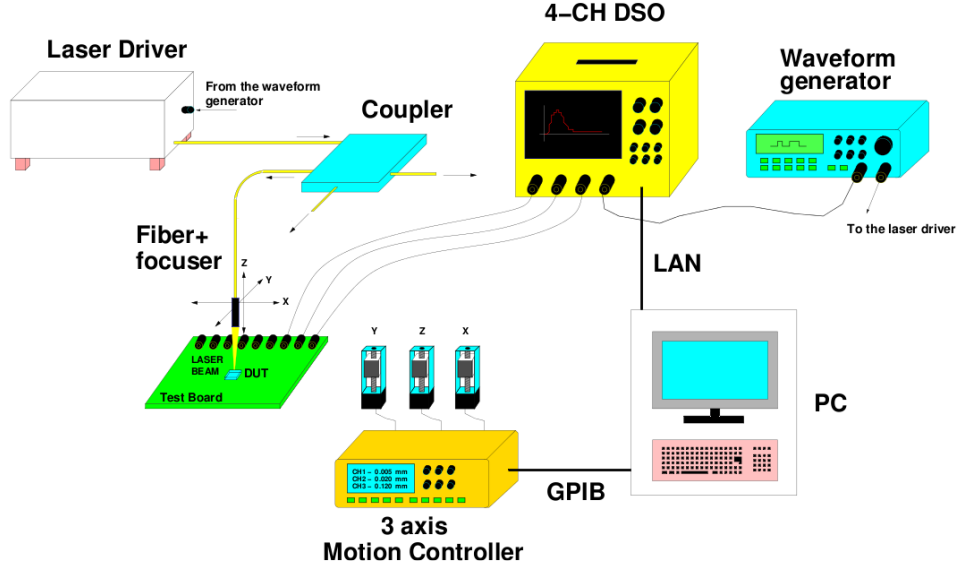


Figure 3.33: measurement setup used in the laser scan tests.

coupler ($\lambda = 1060$ nm), that has the purpose to reduce the power of the laser source. The intensity profile of the laser beam has a Gaussian shape, with a standard deviation (σ) of about $3 \mu\text{m}$. A second focuser, with a σ of about $1.3 \mu\text{m}$, has been used to characterize the sensor with a better spatial resolution. The 3 axes motion system, used in the test, can move the laser beam in the three directions with a resolution of approximately 20 nm. Tests have been performed on the available APSEL5T-TC structures in order to get information about the physical properties of the DNW-MAPS collecting elements in terms of charge collection and charge sharing among pixels. The detector performance was assessed by stimulating three adjacent pixels in the two matrices (M1 and M2), where the analog output is available for each matrix element. Fig. 3.33 shows the setup for laser tests, including the laser source with the relevant optical components (the fiber coupler and the focuser) the Newport Universal Motion Controller (3 axis) and a LeCroy Waverunner 64Xi digital storage oscilloscope (DSO). All instruments are controlled by means of an NI Labview virtual instrument (VI). An Agilent 33250A waveform generator is used to control the laser operation, also providing the trigger signal for the oscilloscope through the sync output. The need for an automatic setup arises from the relatively large number of measurements necessary to perform a full characterization of a MAPS matrix with an incremental step, along both the

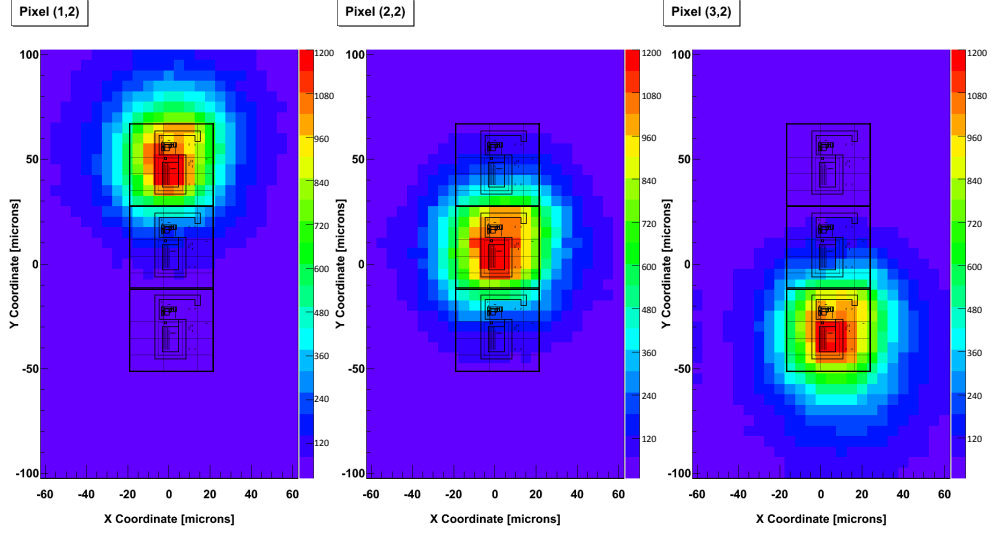


Figure 3.34: charge collected, in electrons, by three adjacent pixels, (2,1), (2,2) and (2,3), of the M1 matrix as a function of the laser position (focuser with $\sigma = 3 \mu\text{m}$).

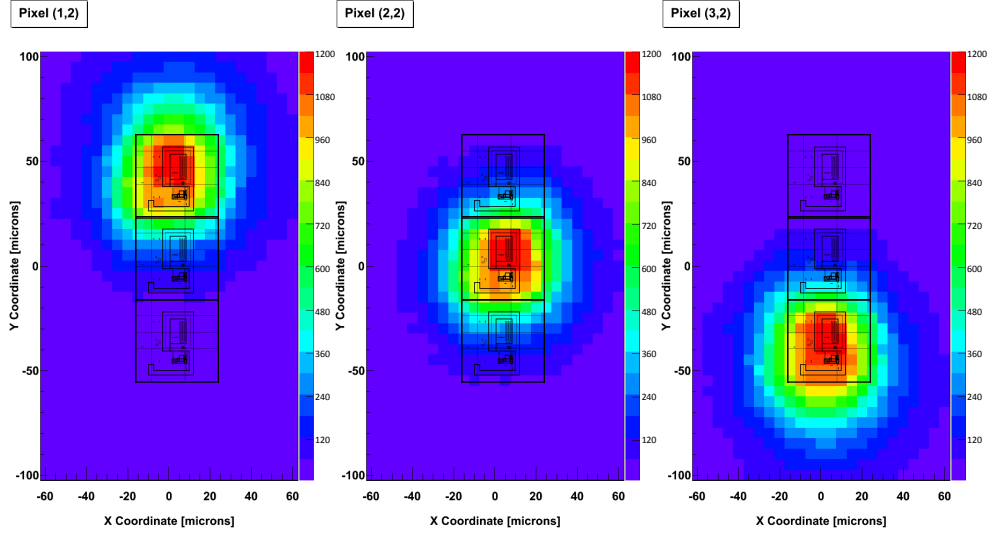


Figure 3.35: charge collected, in electrons, by three adjacent pixels, (2,1), (2,2) and (2,3), of the M2 matrix as a function of the laser position (focuser with $\sigma = 3 \mu\text{m}$).

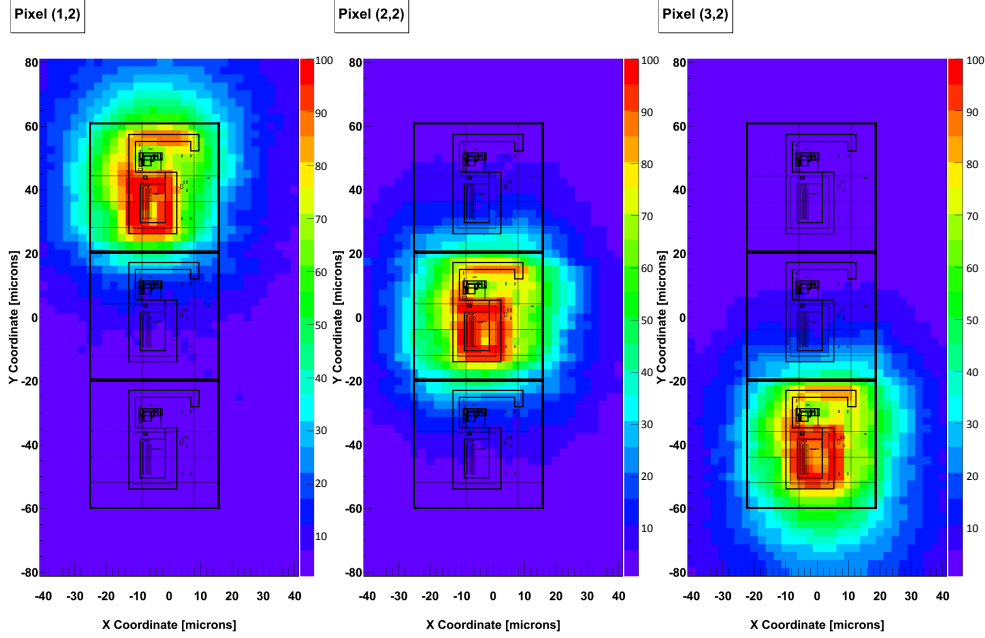


Figure 3.36: percentage of the peak collected charge by three adjacent pixels, (2,1), (2,2) and (3,2), of the M1 matrix as a function of the laser position (focuser with $\sigma = 1.3 \mu\text{m}$).

X and Y directions, not larger than $5 \mu\text{m}$. For each step, the virtual instrument acquires the waveforms from three oscilloscope channels and calculates the position of the next point. Moreover, for each channel, it provides the possibility to calculate the average waveform thus improving amplitude measurement accuracy. At the end of the acquisition, measurement results are summarized in a text file with indications about laser beam coordinates and the relevant signal amplitude at the preamplifier output. Figs. 3.34 and 3.35 show the charge collected by three adjacent pixels of both the M1 and the M2 matrices as a function of the laser position. The measurements were performed using the focuser with $\sigma = 3 \mu\text{m}$. The figure also shows the layout of the sensor, consisting of the n-well diffusions connected to the deep n-well in such a way to form a 6 shape. The collected charge was calculated as the ratio between the amplitude of the preamplifier output response to the laser pulse and the charge sensitivity measured by means of charge injection through an external pulser. The points corresponding to the maximum collected charge value can be observed to be located next to the regions covered by the col-

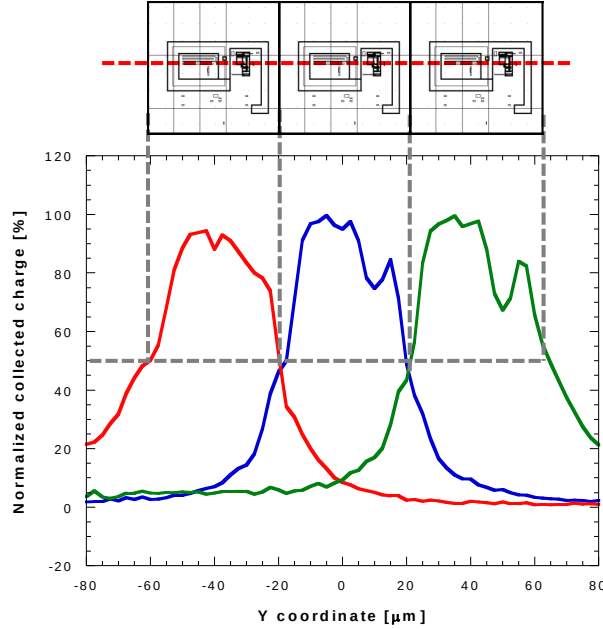


Figure 3.37: collected charge along a cross-section in a 3×3 matrix for charge collection distribution analysis. The values are expressed as the percentage of the peak collect charge in the central pixel.

lectin electrode diffusions. As anticipated, the sensor has been characterized also with a focuser having a smaller σ ($1.3 \mu\text{m}$) with the purpose of going a better understanding of the relative charge collection as a function of the laser spot position and of the effectiveness of competitive n-well diffusions. Fig. 3.36 shows the charge collected by three adjacent pixels as a function of the laser spot position, when a smaller laser spot is used. As expected, the maximum value of the collected charge is obtained in the region covered by DNW and n-well collectin electrodes, whereas a small charge loss can be detected close to the competitive diffusions and to the pixel corners. In Fig. 3.37 the effect of the charge stealing PMOS n-wells is emphasized by the clip in the curves, essentially in the blue and green ones. Note that the collected charge within the pixel area never gets smaller than 50% of the peak collected charge.

3.3 The Superpix1 chip

The main advantage provided by 3D technologies, with respect to planar processes, in high energy physics applications is the opportunity of developing a more complex in-pixel logic. Fig. 3.38 shows a comparison between the two prototype front-end chips for hybrid pixels developed for SuperB-Layer0 vertexing applications. The first device, Superpix0, with a $50\ \mu\text{m}$ pitch, includes a single stage, providing both signal amplification and shaping. Thanks to the use of a vertical integration process, Superpix1, the successor of Superpix0, implements, in the same pitch, a standard analog processing chain, with a preamplifier followed by a shaping block, together with some additional features, such as the possibility to manage charge signals with opposite polarities (electron and hole signals) and circuits for finely adjusting the comparator threshold and for channel calibration.

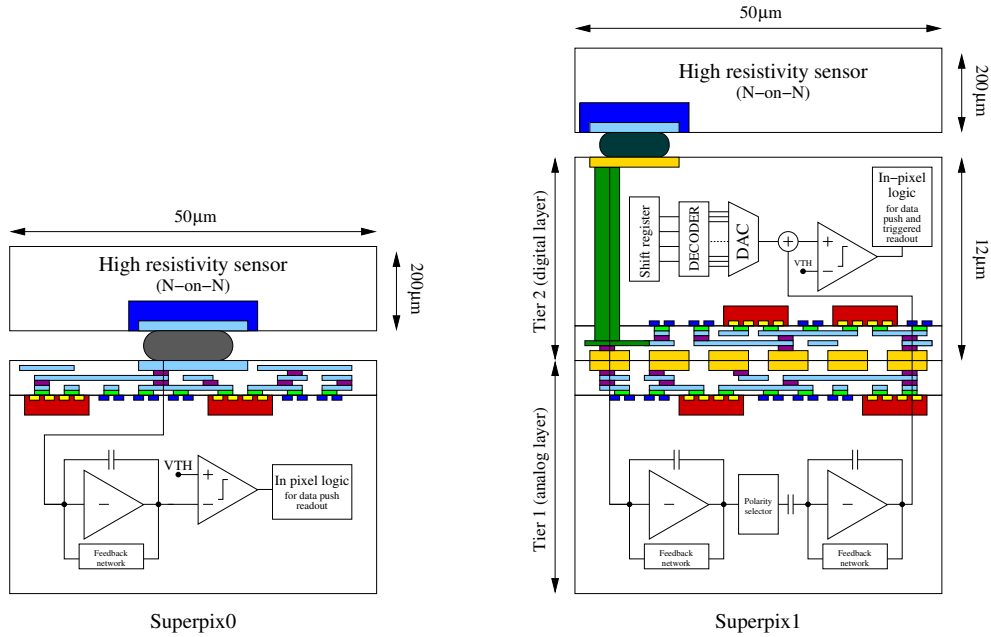


Figure 3.38: schematic representation of the elementary cell of two front-end chips for hybrid detectors developed for the SuperB Layer0. Superpix0 (left) has been integrated in a planar 130 nm CMOS technology provided by STMicroelectronics. Superpix1 has been designed in the 130 nm, 3D CMOS Tezzaron/Globalfoundries process.

3.3.1 Analog front-end

This section is concerned with the detailed description of the analog section of the Superpix1 front-end chip, whose simplified block diagram is shown in Fig. 3.39. The detector is represented with a current source delivering a Dirac delta shaped pulse $Q \delta(t)$, where Q is the charge released in the sensor by an impinging particle, and a capacitance C_D . Candidate sensors, for future interconnection with the front-end chip, are fabricated by FBK-IRST using an N-on-N process on high resistivity substrate with wafer thickness of $200 \mu\text{m}$ and a sensor capacitance of 150 fF . The first stage of the analog signal processor is a charge sensitive amplifier, where the input charge is integrated on the feedback capacitor C_F . In order to make the system compatible both with sensors collecting electrons and sensors collectin holes, the preamplifier is capable of handling signals with positive and negative polarity. The preamplifier output is fed to a polarity selector stage, whose function is to provide signal with

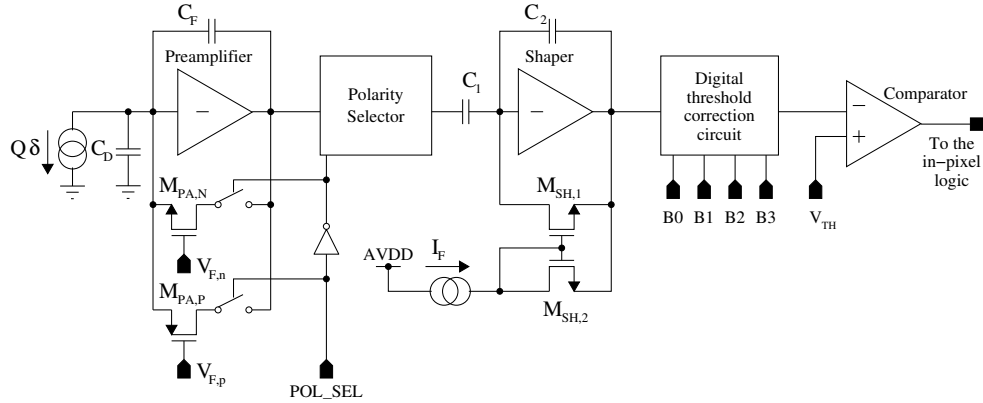


Figure 3.39: block diagram of the analog front-end circuit of Superpix1.

Table 3.3: dimensions and type of the transistor used in the front-end channel of the Superpix1 chip.

Device	W/L [$\mu\text{m}/\mu\text{m}$]	Type
$M_{PA,N}$	0.2 / 20	standard
$M_{PA,P}$	0.15 / 1	low voltage threshold
$M_{SH,1}$	0.2 / 4	low voltage threshold
$M_{SH,2}$	7 / 0.2	low voltage threshold

Table 3.4: capacitance values in the front-end channel of the Superpix1 chip.

Device	Capacitance [fF]
C_D	150
C_F	50
C_1	180
C_2	45

Table 3.5: Main features and simulation results for the Superpix0 and the Superpix1 prototypes.

	Superpix0 prototype	Superpix1 prototype
Charge sensitivity [mV/fC]	50	48
Peaking time [ns]	100	250
Input dynamic range [e^-]	64000	80000
ENC [$e^- rms$]	140	180
Threshold dispersion (bef./aft. correction) [$e^- rms$]	360	560/65
Pixel power dissipation [μW]	10	13.5

positive polarity to the shaper. The polarity selection signal (POL_SEL) is common to all the pixels and can be externally set through a pad. The shaper, which follows this block, has the function of optimizing the signal to noise ratio. In order to minimize the threshold dispersion, which was a weak point of the Superpix0 chip, a new feature has been added to the system. This is the threshold correction circuit, controlled by a 4-bit binary code, located be-

The charge sensitive amplifier

Figure 3.40: Superpix1 charge sensitive amplifier scheme.

Table 3.6: dimensions and type of the MOSFETs used in the Supersp1 charge preamplifier.

Device	W/L [$\mu\text{m}/\mu\text{m}$]	Type
M1	18 / 0.25	low voltage threshold
M2	3.8 / 2	low voltage threshold
M3	0.8 / 0.8	standard
M4	0.5 / 1	low voltage threshold
M5	0.4 / 5.5	low voltage threshold
M6	0.2 / 0.5	standard
M7	0.2 / 2.5	standard
M8	0.2 / 3	low voltage threshold
M9	0.4 / 2.5	low voltage threshold
M10	3 / 0.5	low voltage threshold
M11	0.4 / 5.5	low voltage threshold
M12	0.8 / 2.5	standard

in the Supersp0 chip). The role of the CSA, in this case, is to convert the charge pulse delivered by the detector to a voltage step, whose amplitude is proportional to the input charge. In the design of the charge preamplifier (see Figs. 3.40 and 3.41), the same scheme as in Supersp0, consisting of an NMOS in common source configuration (M1) with a folded cascode stage (the M3 PMOS) was used. The output stage is an NMOS source follower. The CSA scheme is showed in Fig. 3.40 and the relevant device dimensions are shown in table 3.6. The reference voltages ($V_{ref,PA,1}$, $V_{ref,PA,2}$ and $V_{ref,PA,3}$) are generated in the pixel cell. The charge preamplifier is completed with a feedback network including a capacitor C_F and a selectable charge restoration block. With reference to Fig. 3.41, if the POL_SEL signal is set to 1, the switch SW1 is open and SW2 is closed, therefore enabling the discharge of the feedback capacitor through the PMOS transistor, (case of hole collection in the detector). If the POL_SEL signal is set to 0, then the capacitor is reset through the NMOS transistor (case of electron collection in the sensor). Fig. 3.42 shows the simulated preamplifier output waveform, for different injected charge values and type (electron/hole, e-/h+), up to about 80000 e-/h+. The $V_{F,n}$ and $V_{F,p}$ voltage, controlling the feedback circuits, are internally produced, but are also connected to external pads, enabling extensive characterization of the test structures. The design values are 450 mV for $V_{F,n}$ and 105 mV for $V_{F,p}$. The peak value of the preamplifier output waveforms are plotted in Fig. 3.43 as

a function of the input charge. The charge sensitivity G_q has been defined as the slope of the straight line interpolating all the points of the characteristic and forced to pass through the axis origin. The value of charge sensitivity obtained from circuit simulation is about 20 mV/fC both for electron and hole collection. The linearity of the analog readout channel has been evaluated by means of the integral non-linearity (INL) parameter, defined as

$$INL = \frac{\Delta V_{CSA,max}}{V_{Q,max} - V_{Q,min}} \quad (3.2)$$

where $\Delta V_{CSA,max}$ is the maximum difference between the peak value obtained by simulation and the interpolating line showed in Fig. 3.43, $V_{Q,min}$ and $V_{Q,max}$ are the peak values of the pramplifier in correspondence to, respectively, the minimum and maximum injected charges (0 and 80000 e-/h+). The INL calculated the definition in (3.2) is 0.6%.

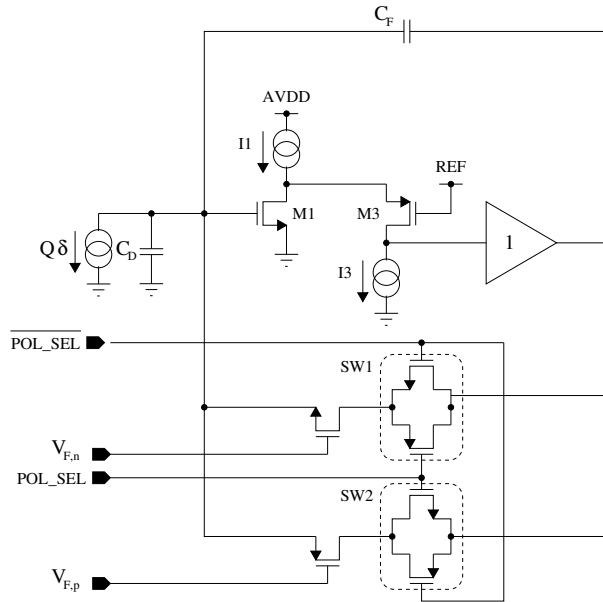


Figure 3.41: simplified schematic diagram of the CSA including the feedback network.

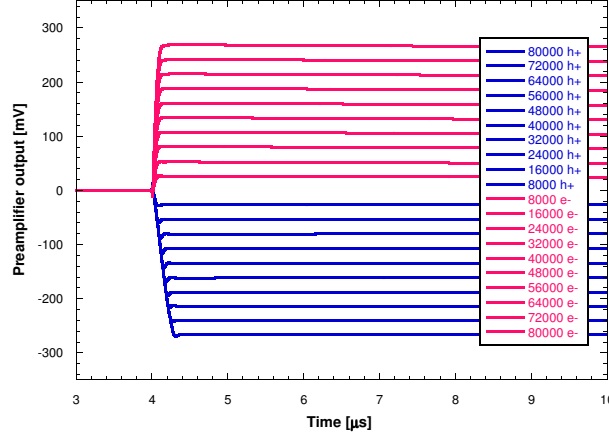


Figure 3.42: Preamplifier output waveforms simulated for different values of the injected charge.

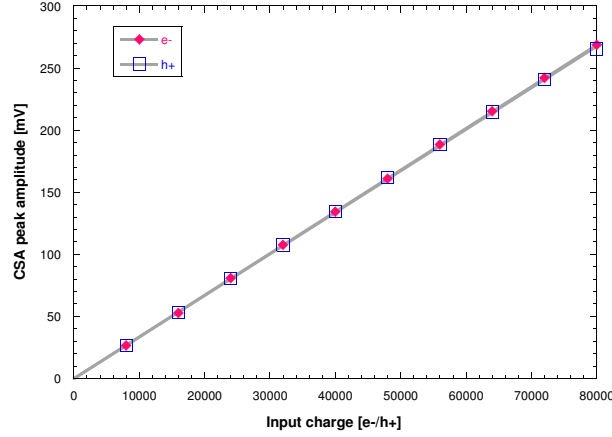


Figure 3.43: CSA input/output characteristic.

The polarity selector

The polarity selection circuit, shown in Fig. 3.44, is based on a differential NMOS pair (M3 and M4), with active load (M5 and M6). The output polarity can be chosen by means of two switches, located in the two branches of the differential pair. The purpose of this block is to provide a signal with positive polarity to the shaper, independent of the polarity of the signal at the preamplifier output. Small signal analysis of the circuit is carried out with reference

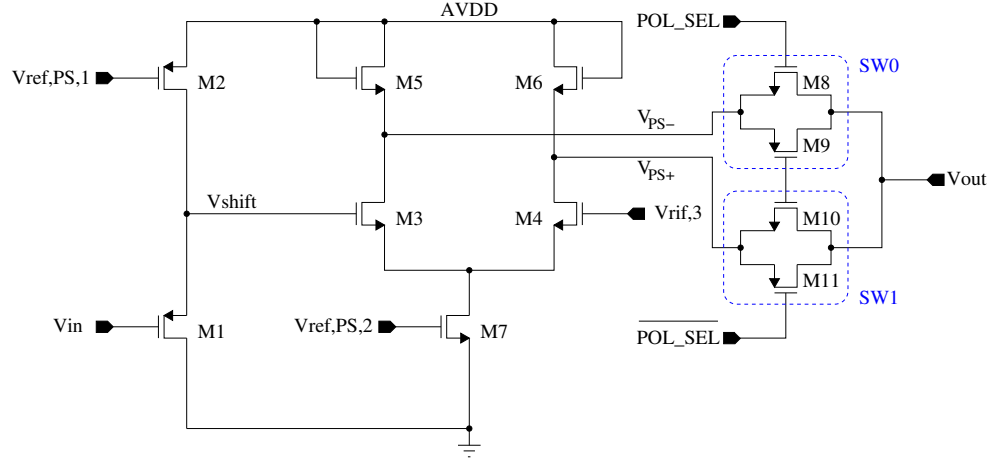


Figure 3.44: polarity selection circuit.

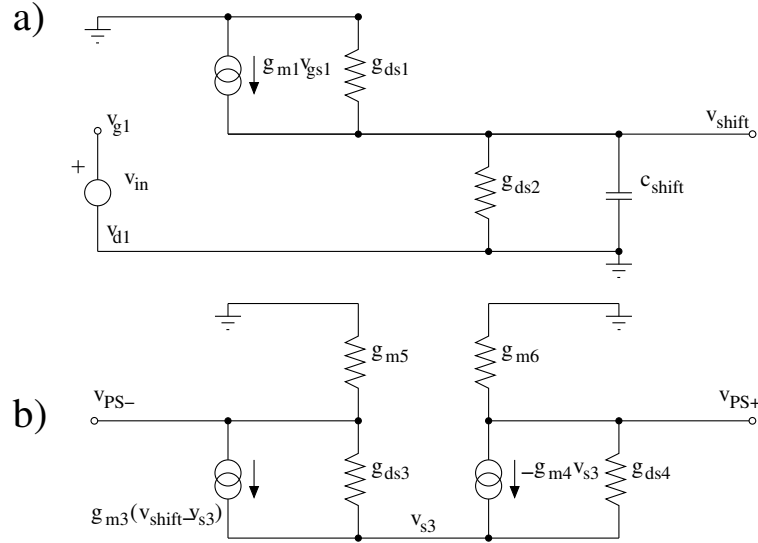


Figure 3.45: small signal equivalent circuit of the polarity selection block.

to Fig. 3.45, g_{mi} being the transconductance of the M_i transistor in Fig. 3.44 and g_{dsi} being its drain to source conductance. The $M1$ transistor is a source follower included in the circuit with the aim of shifting the baseline voltage at the preamplifier output. In particular the voltage at the gate of the $M3$ device,

v_{shift} (Fig. 3.45(a)), is given by

$$v_{shift} = \frac{g_{m1}}{g_{ds1} + g_{ds2} + sC_{shift}} \quad (3.3)$$

where

$$C_{shift} = C_{ss1} + C_{dd2} + C_{gg3}. \quad (3.4)$$

In the two previous equations, v_{shift} can also be seen as the voltage at the input of the NMOS differential pair. The M2 and M7 transistors act as current sources for biasing, respectively, the input branch and the differential pair. The SW0 (M8 and M9) and SW1 (M10 and M10) switches connect the output terminal v_{out} to the V_{PS+} or the V_{PS-} node, depending on the system configuration bit set through the POL_SEL signal. v_{PS-} and v_{PS+} (Fig. 3.45(b)) can be expressed as:

$$v_{PS-} = -\frac{g_{m3}}{g_{ds3} + g_{m5}} v_{shift} \quad (3.5)$$

$$v_{PS+} = +\frac{g_{m4}}{g_{ds4} + g_{m6}} v_{shift} \quad (3.6)$$

The gain of the polarity selector block in the two branches, expressed as the ratio between the v_{PS-} and v_{in} or v_{PS+} and v_{in} ,

$$\frac{v_{PS-}}{v_{in}} = -\frac{g_{m1}}{g_{ds1} + g_{ds2}} \cdot \frac{g_{m3}}{g_{ds3} + g_{m5}} \quad (3.7)$$

$$\frac{v_{PS+}}{v_{in}} = +\frac{g_{m1}}{g_{ds1} + g_{ds2}} \cdot \frac{g_{m4}}{g_{ds4} + g_{m6}} \quad (3.8)$$

is less than 1. A value of 0.8 was actually extracted from circuit simulations.

The shaping stage

The shaper stage is decoupled from the CSA by means of the capacitance C_1 (Fig. 3.46). An NMOS current mirror structure, implemented by $M_{SH,1}$ and $M_{SH,2}$, is used to continuously reset the shaper feedback capacitor. The design choice of a feedback network with constant current discharge of C_2 , proportional to the mirror reference current I_F (about 230 nA), is dictated by noise constraints. Since C_2 is discharged by a constant current, the recovery time increases linearly with the signal amplitude. The shaper circuit, shown in Fig. 3.46, is composed by a cascode input stage (M1 and M2) with active load (M3) and a source follower output stage (M5). The input branch current, 1.2 μ A, is set by the M3 transistor, biased by means of the $V_{ref,SH,1}$ reference voltage, which is shared with the M2 device and with the polarity selection

block ($V_{ref,SH,1} = V_{ref,PS,1}$). Also, the $V_{ref,SH,2}$ voltage reference, used to bias M4 in the output stage of the shaper, is internally generated and shared with the polarity selection circuit ($V_{ref,SH,2} = V_{ref,PS,2}$). In order to increase the output baseline value, i.e. the V_{out} DC voltage in Fig. 3.46, and, then, improve the output dynamic range of the shaper, M1 features a thick oxide transistor. The W/L aspect ratio and the type of the MOSFETs part of the shaper stage are shown in table 3.7. The small circuit analysis has been performed with reference to Fig. 3.47. The transfer function $V_{out}(s)/V_{in}(s)$ of the shaper stage

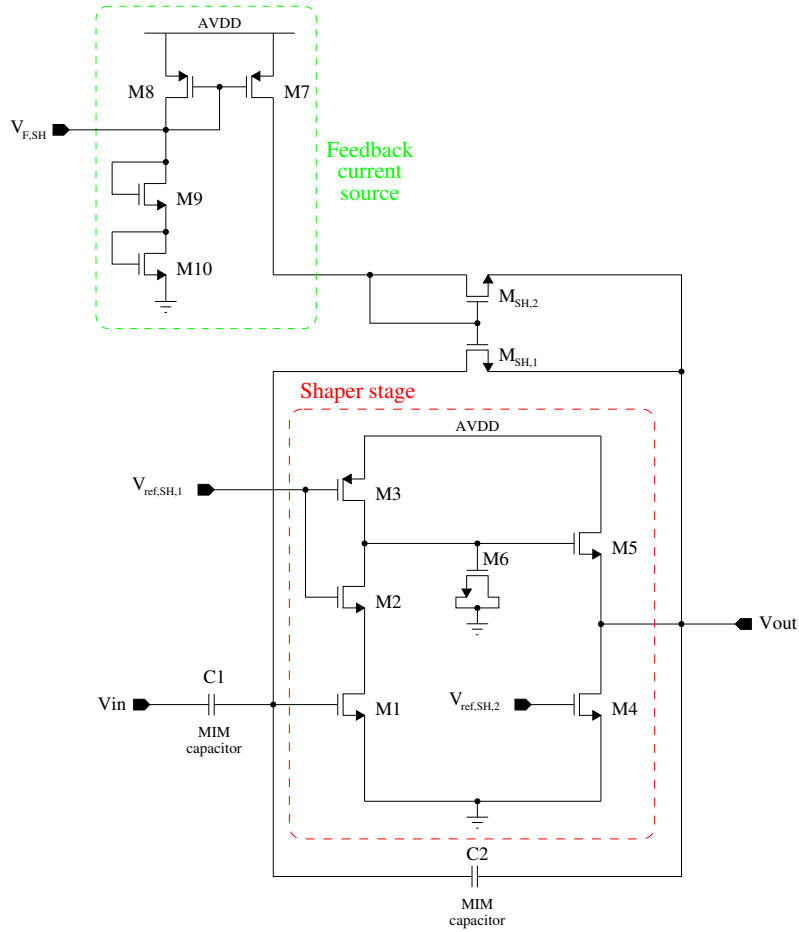


Figure 3.46: schematic diagram of the shaping stage.

can be expressed as

$$\frac{V_{out}(s)}{V_{in}(s)} = - \frac{g_{m1}}{\left(\frac{1}{\frac{1}{g_{ds1}} + \frac{1}{g_{ds2}} + \frac{g_{m2}}{g_{ds1}g_{ds2}}} \right) + g_{ds3} + s c_{gg6}} \cdot \left(\frac{g_{m5}}{g_{ds4} + g_{ds5} + g_{m5}} \right) \quad (3.9)$$

where $V_{in}(s)$ and $V_{out}(s)$ are the input and output signals in the Laplace domain, c_{gg6} is the gate capacitance of the M6 MOSFET, and, again, g_{mi} is the transconductance and g_{dsi} the output conductance of the Mi transistor in Fig. 3.46. The term $g_{m5}/(g_{ds4} + g_{ds5} + g_{m5})$ is approximately equal to 1. Then, the DC open loop gain of the shaper stage is given by

$$A_{O,SH} = - \frac{g_{m1}}{\left(\frac{1}{\frac{1}{g_{ds1}} + \frac{1}{g_{ds2}} + \frac{g_{m2}}{g_{ds1}g_{ds2}}} \right) + g_{ds3}} \quad (3.10)$$

$A_{O,SH}$ has been calculated and a value of about 45 dB has been obtained. The time constant of the open loop gain can be expressed as

$$\tau_{SH} = c_{gg6} \cdot \frac{A_{O,SH}}{g_{m1}}, \quad (3.11)$$

resulting in a cutoff frequency of about 80kHz. These results are in good agreement with the simulation data of Fig. 3.48, showing the modulus of the open

Table 3.7: transistor dimension and type in the Superpix1 shaper stage.

Device	W/L [$\mu\text{m}/\mu\text{m}$]	Type
M1	10 / 4	thick oxide
M2	6 / 1	low voltage threshold
M3	5 / 2	low voltage threshold
M4	3 / 0.85	low voltage threshold
M5	3 / 3	low voltage threshold
M6	3.2 / 2	standard
M7	2 / 1	standard
M8	2 / 1	standard
M9	0.2 / 7	low voltage threshold
M10	0.2 / 7	low voltage threshold

loop gain of the shaping stage. The value of charge sensitivity, obtained from circuit simulation is about 48 mV/fC both for electron and hole collection, evaluated after the threshold correction circuit. Indeed the DAC circuits introduce a gain of about 0.9 in the front-and processing chain. The linearity of the analog readout channel has been evaluated, again, by means of the integral non-linearity parameter, resulting in an INL less than 1%.

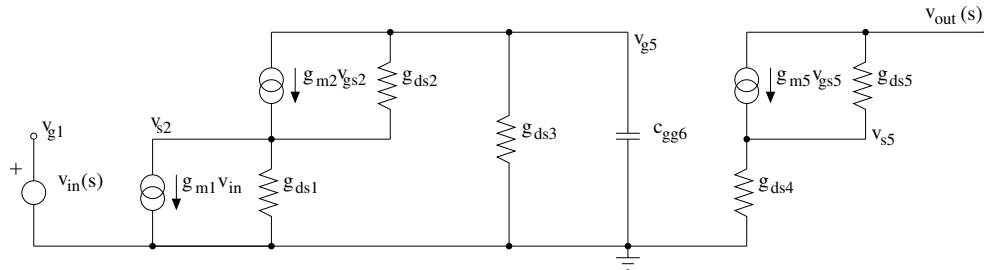


Figure 3.47: shaper small signal equivalent circuit.

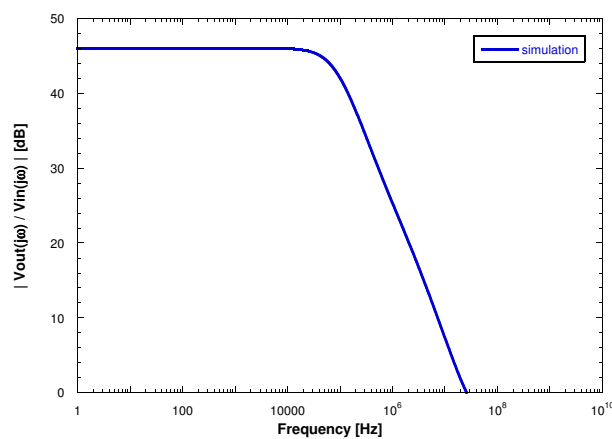


Figure 3.48: modulus of the open loop gain $|V_{out}/V_{in}|$ as a function of the frequency in the shaper stage as obtained from circuit simulations.

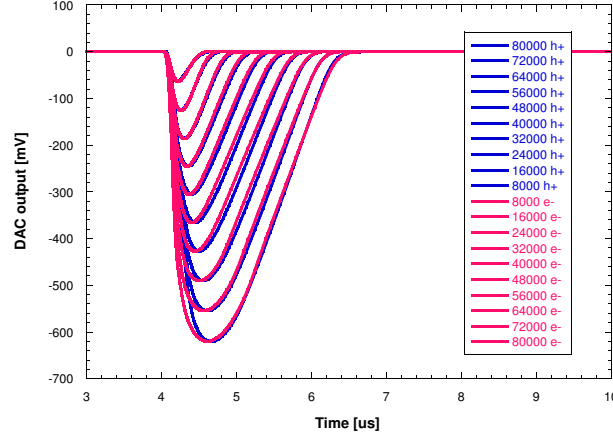


Figure 3.49: Superpix1 output waveforms, after the DAC stage (discussed in the following), simulated for different values of the injected charge.

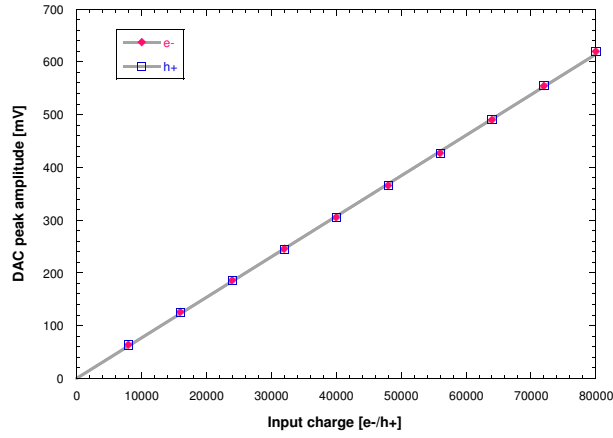


Figure 3.50: Superpix1 input/output characteristic after the DAC stage (discussed in the following).

3.3.2 Injection circuit for the chip calibration

In the Superpix1 chip also an injection circuit for front-end calibration has been included. The circuit was designed according to the scheme of a system for pixel-level calibration of readout electronics developed in the framework of the European large-scale X-ray Free Electron Laser (XFEL) collaboration [50].

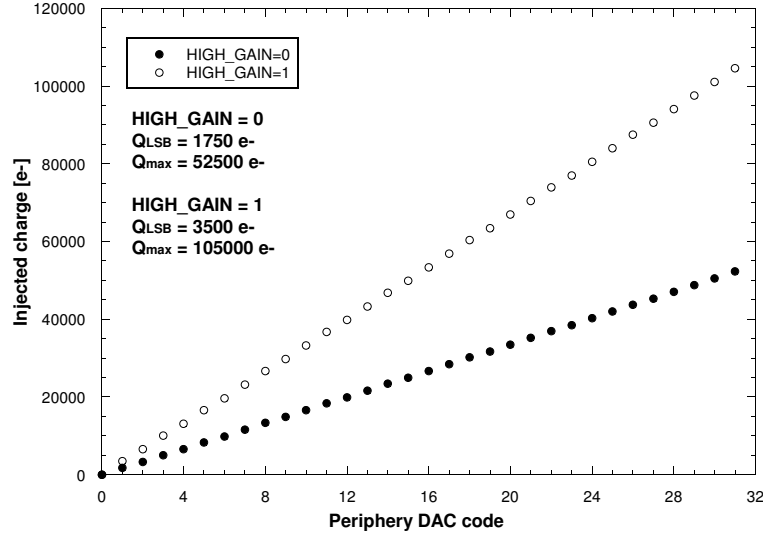


Figure 3.51: simulation of the charge injected by the pulser circuit as obtained from circuit simulations.

The injection circuit can be used for testing the functionality of the analog and digital section of the front-end chip with or without the sensor connected. Moreover, it provides a means for assessing the chip functionality also once the detector has been connected to the chip. The injection circuit was designed very carefully in order not to degrade the channel performance due to excess noise contributions at the preamplifier input. A charge sensitivity measurement can be performed by injecting a well known fixed signal into each individual readout channel and then analyzing the data at the digital readout output. The signal generated by the injection circuit is a voltage step, which is converted to charge by means of a capacitance located at the input of the front-end electronics. The injection circuit must be able to cover the full dynamic range of the signals delivered by the sensor, corresponding in the case of Superpix1, to about 80000 electrons. The injection circuit consists of two main parts: the pulser circuit, hosted in the pixel cell and connected to the input of the preamplifier through the injection capacitance, and a stage needed to select the amount of charge to be injected. The latter task is performed by a 5 bit current steering DAC, located in the periphery of the chip. In order to make the system more flexible and enable circuit testing with large injected charge without increasing the DAC complexity and area, a bit to select the DAC dynamic range has been added. If this feature is disabled, the injection circuit provides a high resolution in setting the amount of charge to be in-

jected, with a dynamic range of 80000 electrons (Fig. 3.51, plot with full block marks), while if the HIGH_GAIN signal is set to 1, the resolution is reduced, but a larger dynamic range is available for charge injection (Fig. 3.51, plot with open block marks). The digital-to-analog converter features an integral non linearity (INL) error smaller than 0.5 LSB, therefore fully complying with a common specification for DACs [48]. If the INL is smaller than 0.5 LSB then the differential non linearity (DNL) is always smaller than 1 LSB, which ensures that the converter is monotonic. The current provided by the DAC is mirrored into each pixel, where the pulser circuit generates the relevant signal at the preamplifier input. The pulser current consumption can be set to 0 by applying a zero binary word to the DAC periphery. The circuit is required to simultaneously stimulate several pixels in a freely programmable spatial pattern. Fig. 3.52 actually shows that the injection signal into each pixel (INJ) can be independently enabled or disabled by means of the INJ_EN signal.

The pulser circuit in the pixel cell

The pulser circuit, shown in Fig. 3.52, injects a charge into the input of the preamplifier by applying a voltage step to an injection capacitor $C_1 = 35$ fF. The current from the 5 bit current steering DAC is mirrored into each pixel by means of a 10:1 mirror. The minimum current step in the cell is $I_{LSB} = 34$ nA. By means of the inject signal (INJ), the mirrored current is switched from the left branch of the differential stage to the right one to produce a negative voltage step applied to the injection capacitance C_1 . In order to fit the requirements on resolution and dynamic range of the injected charge and on

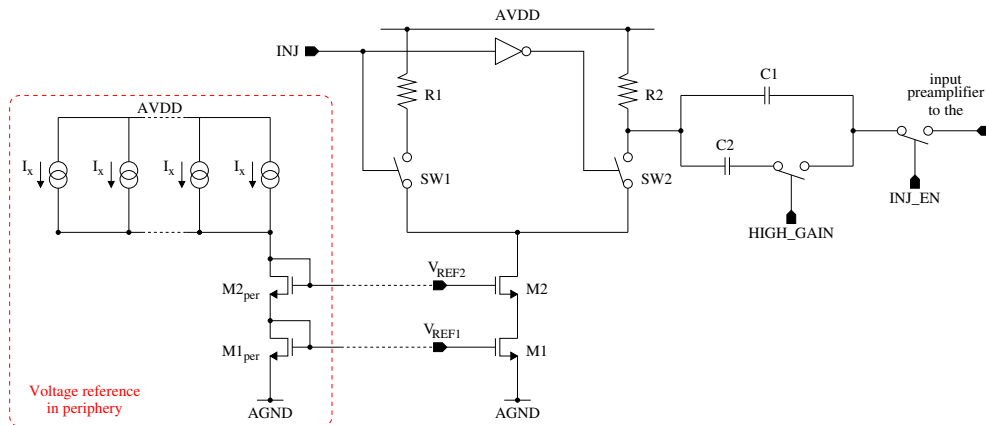


Figure 3.52: schematic diagram of the pulser circuit.

power consumption, a value of 226.5 k Ω has been chosen for resistors R_1 and R_2 . Charge injection into each pixel cell can be disabled by opening a switch in series with the injection capacitance. The status of the switch is established by a local control bit (INJ_EN), stored in a register placed in the pixel.

The D/A converter in the periphery

The current steering D/A converter, shown in Fig. 3.53, can switch each of the 31 unity current generators toward the output node or to a dummy load under the control of the 5 bit digital input. The current cell consists of a cascoded current source ($M_{i,1}$, $M_{i,2}$) and current-steering switches ($M_{i,3}$, $M_{i,4}$). The current generated by the current reference cell is mirrored into each unary current cell by means of a 1:1 cascode current mirror scheme. The current provided by the reference is 340 nA. The current-steering switches ($M_{i,3}$, $M_{i,4}$) send the current of the cascode current mirror to the output ($M_{out,1}$, $M_{out,2}$) when the signal EN_CELL*i* is at a high level and routes the current to the dummy load when EN_CELL*i* is low. For a current-steering DAC, the INL is mainly determined by the matching features of the current sources. Inaccuracies on the generated currents can be caused, again, by random and systematic variations of device parameters. Both effects have been taken into account in the DAC design. One source of systematic non linearity is the finite output impedance of the current source: when the input code varies between zero and

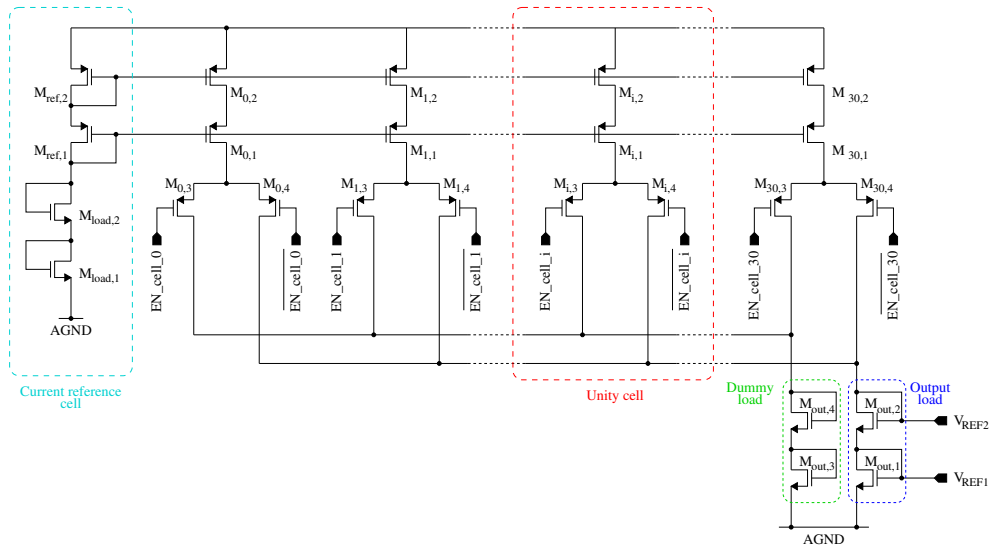


Figure 3.53: schematic diagram of the current steering DAC.

the full scale, an increasing number of current sources are connected in parallel, resulting in a decreasing impedance at the output node. This effect on

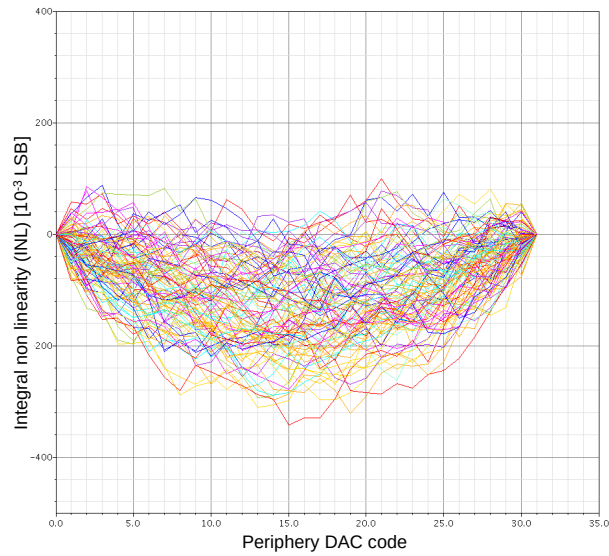


Figure 3.54: integral non linearity of the injection circuit.

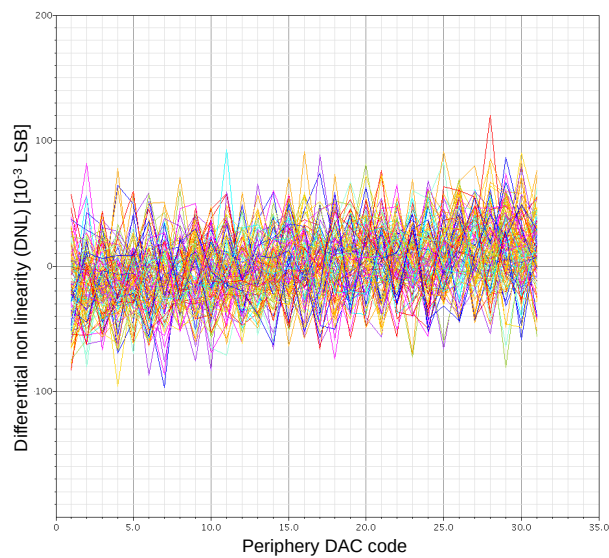


Figure 3.55: differential non linearity of the injection circuit.

the output current has been reduced by means of a cascoded current source in the unary current cell, which increases the output impedance. Moreover, the random error of the current source, which is mainly determined by device mismatch, has been considered in determining the dimension of the critical unit current cell transistor ($M_{i,1}$, $M_{i,2}$). Monte Carlo simulations, whose results are showed in Fig. 3.54 and Fig. 3.55, have been performed on the injection circuit. The maximum value of the integral non linearity parameter is 34% of the least significant bit (LSB) and the differential non linearity parameter is 12% of the LSB, as expected from design specification.

3.3.3 Threshold dispersion and detection efficiency

As already mentioned, a threshold correction circuit has been included in the Superpix1 front-end circuit, based on a digital to analog converter (DAC), to minimize the overall threshold dispersion. In the following, a set of criteria for optimum design of DACs for threshold correction in multichannel circuits is introduced and discussed. Fig. 3.56 shows the simplified block diagram of a charge measuring system (like Superpix1), including a charge preamplifier with feedback capacitor C_F , its charge restoring network and a filtering stage with transfer function $T(st_p)$, where t_p is a characteristic time of the filter (typically the peaking time of the delta response of the system), and V_{bl} the DC output voltage. At the end of a typical analog processing chain, a discriminator compares the signal at the shaper output with a preset threshold V_{th} . The voltage threshold is chosen in such a way to maximize the rate of detected true events without having the system flooded with false, noise-induced hits. The detection efficiency of the system is defined as the ratio between the true

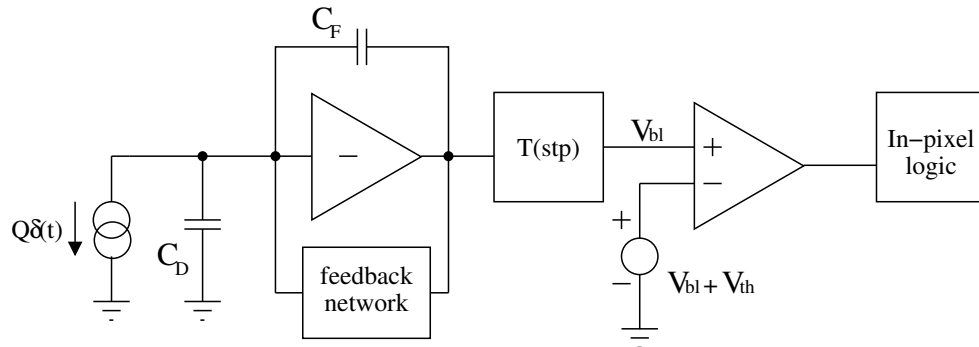


Figure 3.56: simplified block diagram of a typical front-end analog processor for capacitive detectors.

hits detected and sent off chip and the true hits. Efficiency optimization is generally constrained by the chip readout architecture, i.e., by the speed of the readout electronics in collecting data from the detector matrix and sending them off chip in the case of data-push systems, or by the pipeline depth in data-pull (triggered) architectures. Actually, channels with a low threshold are susceptible to noise hits, while channels with a high threshold may miss some true hit. Typically, in a multichannel binary readout system, a maximum noise hit rate $f_{n,max}$, i.e. a maximum rate of noise-induced transitions at the discriminator output, exists beyond which the system is not capable to guarantee a given, target readout efficiency (defined as the ratio between the number of hits actually sent off chip and the number of detected, real or noise-induced, hits). Under some very general assumptions, in a binary channel, the noise hit

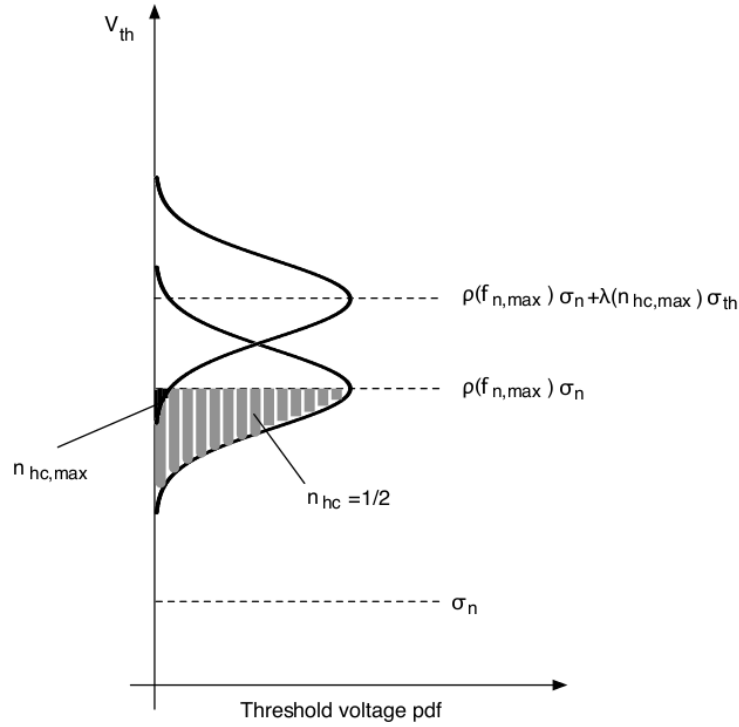


Figure 3.57: threshold setting in a multichannel chip. The fraction of hot channels is represented by the gray plus black shaded areas if the threshold is set without accounting for threshold dispersion effects, by the black shaded area alone when such effects are taken into consideration.

rate f_n , whose analysis represents a specific aspect of the more general level crossing problem, can be written as [46]:

$$f_n = f_{n0} \cdot e^{-\frac{V_{th}^2}{2\sigma_n^2}} \quad (3.12)$$

where f_{n0} is the noise hit rate at zero threshold (i.e., when the threshold voltage equals the mean value of the voltage at the other discriminator input, corresponding to the output of the analog processor) and σ_n is the root mean square value of the noise at the analog processor output. As already said, in a multichannel binary readout circuit, random and systematic variations of process (doping) and geometrical (device dimensions, thickness of the various involved layers) parameters may be responsible for introducing non uniformities in the parallel path followed by the signals. As a result, two channels like that in Fig. 3.56, nominally identical to each other and featuring a common threshold at the inverting input of the discriminator, may provide different responses to the same charge pulse at the preamplifier input. Typically, the main contributions to such non uniformities come from the shaper (generally AC coupled to the preamplifier) and from the discriminator. The overall effect is usually simply referred to as threshold dispersion, as it can be equivalently represented in terms of a statistical distribution of the discriminator threshold voltage [46]. If the effect of threshold dispersion is neglected, a minimum threshold value $V_{th,min}$ can be found for all the channels to satisfy the requirements on the maximum noise hit rate,

$$V_{th,min} = \sigma_n \cdot \sqrt{2 \cdot \ln \left(\frac{f_{n0}}{f_{n,max}} \right)} = \rho(f_{n,max}) \cdot \sigma_n \quad (3.13)$$

where $\rho(f_{n,max})$ is a slowly decreasing function of $f_{n,max}$. Actually, if the threshold is set according to (3.13), due to threshold dispersion, the noise hit rate could exceed $f_{n,max}$ in a significant fraction of the channels (also called hot channels). In order to limit the fraction of hot channels, threshold dispersion has to be taken into account and $V_{th,min}$ has to be moved towards higher values according to the following equation,

$$V_{th,min} = \rho(f_{n,max}) \cdot \sigma_n + \lambda(n_{hc,max}) \cdot \sigma_{th} \quad (3.14)$$

where the threshold dispersion is represented by means of the standard deviation σ_{th} of the threshold distribution in the multichannel chip. In (3.14), λ is a decreasing function of $n_{hc,max}$, the maximum acceptable fraction of hot

channels. If V_{th} is normally distributed, for the fraction of hot channels not to be larger than $n_{hc,max}$, λ should be chosen such that

$$\lambda(n_{hc,max}) = \sqrt{2} \cdot \text{Erfc}^{-1}(2n_{hc,max}) \quad (3.15)$$

where $\text{Erfc}(x)$ is the complementary error function. While, on the one hand, increasing the threshold voltage allows the system to comply with the noise hit rate constraints, on the other hand it reduces the detection efficiency, in particular for those channels which are located towards the higher end tail of the distribution. Therefore, in order to make the detection efficiency as high and uniform as possible, measures have to be taken to minimize threshold dispersion.

Optimum DAC design for threshold correction

Excessive threshold dispersion in a multichannel chip for capacitive detectors is generally managed by means of a D/A converter trimming the threshold in each individual channel. Fig. 3.58 shows the final part of a typical readout channel, with a block diagram of the discriminator and a DAC for threshold correction acting on the non inverting input of the discriminator. The correction could be equivalently applied to the other input. It is quite reasonable to expect that the larger the resolution (number of bits) of the DAC, the finer the threshold adjustment and, eventually, the narrower the dispersion among the channels. On the other hand, a wide output range might be necessary to cope with channels featuring a largely offset threshold, resulting in a degradation of the DAC trimming performance. In order to represent threshold dispersion in a multichannel chip, the threshold voltage V_{th} is generally treated as a random variable with Gaussian distribution,

$$p(V_{th}) = \frac{1}{\sigma_{th}\sqrt{2\pi}} \cdot e^{-\frac{(V_{th}-\langle V_{th} \rangle)^2}{2\sigma_{th}^2}} \quad (3.16)$$

where $\langle V_{th} \rangle$ is the average value of the threshold calculated over the set of channels. For the sake of simplifying the calculations, in the following, $\langle V_{th} \rangle = 0$ will be assumed with no effect on the validity and generality of the results. In order to correct the threshold in each individual pixel, a digital to analog converter can be included in the elementary cell in such a way that the voltage generated at the DAC output V_{DAC} is added to the baseline voltage V_{bl} at the analog channel output. Threshold correction is studied under the following hypothesis:

- the output range of the n bit DAC is a (not necessarily integer) multiple of the threshold dispersion σ_{th} by a factor θ ; the output range is therefore subdivided in 2^n equal intervals, numbered from 0 to $2^n - 1$;
- in each cell, the threshold correction is obtained by programming the DAC so as to shift the actual threshold V_{th} as close as possible to $\langle V_{th} \rangle$, under the constraint that, due to the very nature of the correction technique, involving a D/A converter, the shift can be applied only in discrete steps, depending on the DAC resolution.

Application of the above procedure is graphically represented in Fig. 3.59. Correction of the threshold through a D/A converter has the same effect as folding the probability density function (PDF) $p(V_{th})$ onto the average value of the distribution. The threshold PDF is subdivided into $(2^n + 2)$ intervals; of them, 2^n are $\theta\sigma_{th}/2^n$ long, while the remaining two, going to $+\infty$ and to $-\infty$ respectively, include the probability density function portions not covered by the DAC range $\theta\sigma_{th}$. When correcting V_{th} , the same shift is applied to all the channels with a threshold voltage lying in a given interval. In Fig. 3.59, the black strips at the left and right ends of the graph lay outside the DAC output range $\theta\sigma_{th}$ (which is centered on the distribution average value). As a consequence, the corresponding portion of the PDF cannot be turned right onto the central, obliquely striped area but just pushed as close as possible to

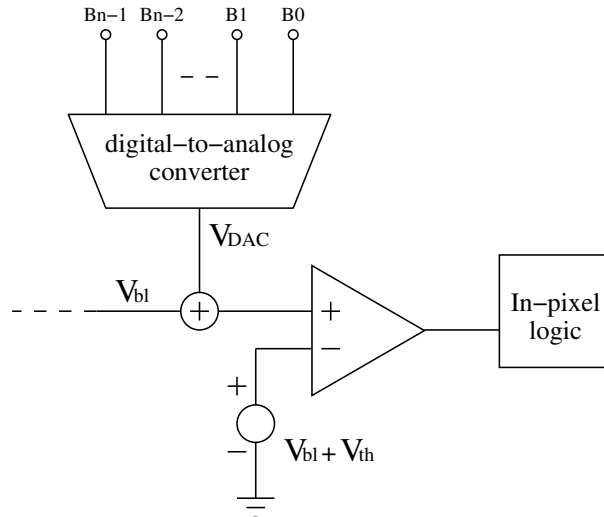


Figure 3.58: block diagram of the discriminator with the DAC for threshold correction.

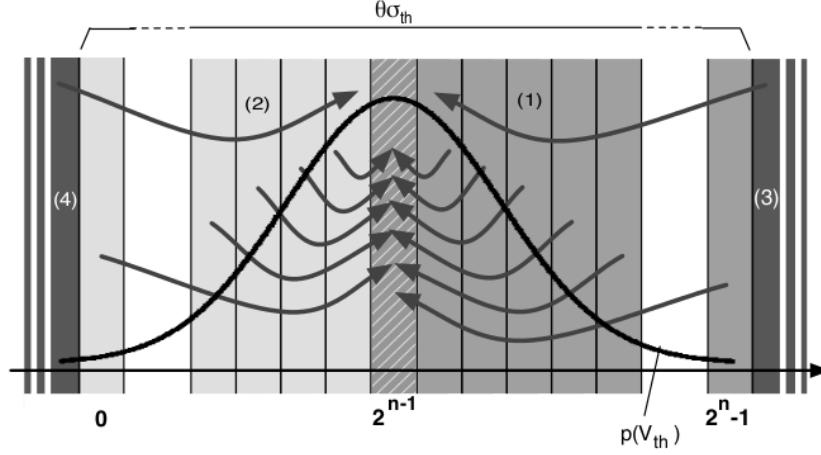


Figure 3.59: graphical representation of the threshold correction as the folding of the threshold probability density function onto the average value.

it. The best that can be done in this case is apply the maximum voltage shift made available by the DAC. All the other portions of the PDF, laying against the dark and light gray strips and included in the DAC output range, can be folded directly onto the central one. This operation results in a new PDF for the corrected threshold $p_{n,\theta}(V_{th})$, depending on the parameter n (the DAC resolution) and θ (the ratio between the DAC output range and the threshold dispersion before correction). The analytical expression of the PDF is provided by

$$\begin{aligned}
 p_{n,\theta}(V_{th}) = & \underbrace{\frac{1}{\sigma_{th}\sqrt{2\pi}} \cdot H\left(\frac{\theta\sigma_{th}}{2^{n+1}} - |V_{th}|\right) \cdot \sum_{i=0}^{2^{n-1}} \exp\left[-\frac{\left(V_{th} + i\frac{\theta\sigma_{th}}{2^n}\right)^2}{2\sigma_{th}^2}\right]}_{(1)} \\
 + & \underbrace{\frac{1}{\sigma_{th}\sqrt{2\pi}} \cdot H\left(\frac{\theta\sigma_{th}}{2^{n+1}} - |V_{th}|\right) \cdot \sum_{i=1}^{2^{n-1}-1} \exp\left[-\frac{\left(V_{th} - i\frac{\theta\sigma_{th}}{2^n}\right)^2}{2\sigma_{th}^2}\right]}_{(2)}
 \end{aligned}$$

$$\begin{aligned}
& + \underbrace{\frac{1}{\sigma_{th}\sqrt{2\pi}} \cdot H\left(V_{th} - \frac{\theta\sigma_{th}}{2^{n+1}}\right) \cdot \exp\left[-\frac{\left(V_{th} + \frac{\theta\sigma_{th}}{2^n}\right)^2}{2\sigma_{th}^2}\right]}_{(3)} \\
& + \underbrace{\frac{1}{\sigma_{th}\sqrt{2\pi}} \cdot H\left(V_{th} + \frac{\theta\sigma_{th}}{2^{n+1}}\right) \cdot \exp\left[-\frac{\left(V_{th} - (2^{n-1} - 1)\frac{\theta\sigma_{th}}{2^n}\right)^2}{2\sigma_{th}^2}\right]}_{(4)},
\end{aligned} \tag{3.17}$$

where the equation terms are linked to the various sections of Fig. 3.59, marked with shades of gray and numbered from 1 to 4 (numbers 3 and 4 referring to the PDF portions laying outside the DAC output range). In (3.17) $H(x)$ is the Heaviside function. The standard deviation $\sigma_{th,c}$ of the threshold distribution after correction can be calculated as

$$\begin{aligned}
\sigma_{th,c} &= \sqrt{\int_{-\infty}^{+\infty} p_{n,\theta}(V_{th}) V_{th}^2 dV_{th}} \\
&= \sigma_{th} \sqrt{\int_{-\infty}^{+\infty} \sigma_{th} \cdot p_{n,\theta}(\sigma_{th}u) u^2 du}.
\end{aligned} \tag{3.18}$$

Therefore,

$$\frac{\sigma_{th,c}}{\sigma_{th}} = \sqrt{\int_{-\infty}^{+\infty} \sigma_{th} \cdot p_{n,\theta}(\sigma_{th}u) u^2 du}, \tag{3.19}$$

where, taking into account that, for $a \neq 0$, $H(ax) = H(x)$, it can be demonstrated that $\sigma_{th} \cdot p_{n,\theta}(\sigma_{th}u)$ and, as a consequence $\sigma_{th,c}/\sigma_{th}$ is independent of σ_{th} and is a function of n and θ only. Fig. 3.60 shows the $\sigma_{th,c}/\sigma_{th}$ ratio, (in the following also referred to as correction factor) as a function of the parameter θ for different values of the DAC resolution n . As already suggested, for a given resolution, small values of the DAC range leave a significant fraction of the thresholds out of the correction span, therefore limiting the effectiveness of the process. Suboptimal results are obtained also when too large a range is set, as the consequently large width of the correction step is unsuitable for fine threshold adjustment. Actually, an optimum value of the DAC output range can be found, depending on the resolution. Fig. 3.61 shows θ_{opt} , i.e. the optimum DAC range divided by the threshold dispersion σ_{th} as a function of the number of bits of the DAC. The points can be interpolated by the following linear equation,

$$\theta_{opt}(n) = a + b \cdot n, \tag{3.20}$$

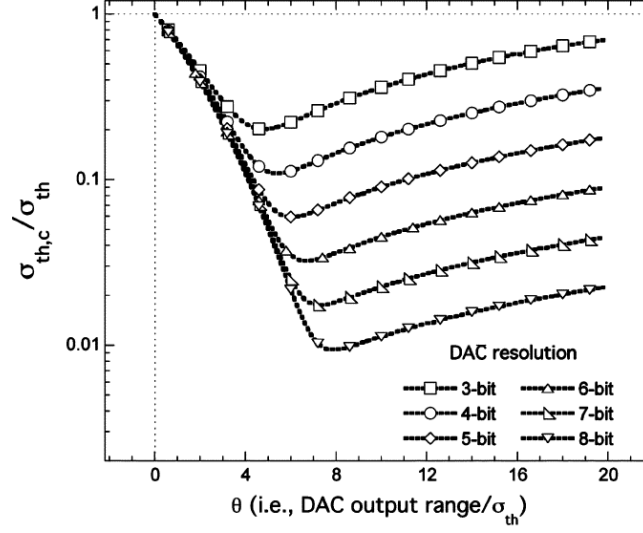


Figure 3.60: correction factor ($\sigma_{th,c}/\sigma_{th}$), as a function of the parameter θ for different values of the DAC resolution.

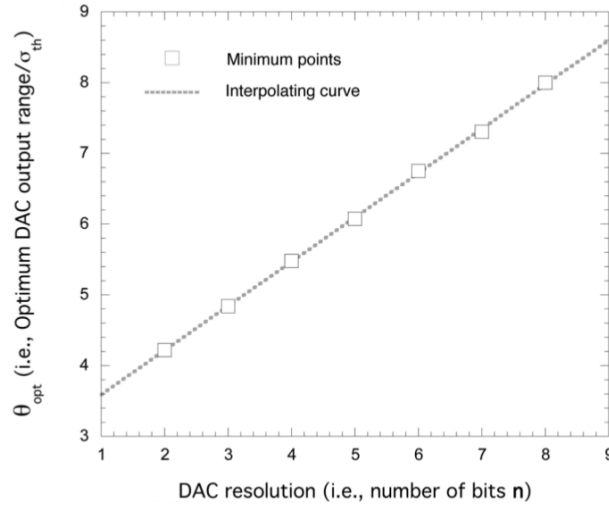


Figure 3.61: optimum DAC range divided by the threshold dispersion σ_{th} , as a function of the number of bits of the DAC. The interpolating function (3.20) is shown together with the minimum points (open squares) obtained from Fig. 3.60.

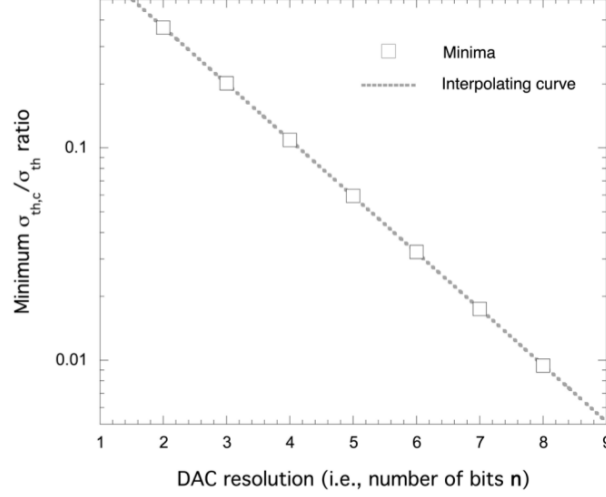


Figure 3.62: minimum $\sigma_{th,c}/\sigma_{th}$ ratio as a function of the correction DAC resolution. The interpolating function (3.21) is shown together with the minima (open squares) obtained from Fig. 3.60.

where $a \simeq 2.96$ and $b \simeq 0.63$. Fig. 3.62 shows the minimum $\sigma_{th,c}/\sigma_{th}$ ratio as a function of the correction DAC resolution. The minimum correction factor was found to follow an exponential law, given by the following interpolating function,

$$\min_{\theta} \left\{ \frac{\sigma_{th,c}}{\sigma_{th}} \right\} (n) = c \cdot e^{-d \cdot n} \quad (3.21)$$

where $c \simeq 1.26$ and $d \simeq 0.61$. It is worth noting that a 4 bit ($n = 4$) correction DAC can theoretically achieve a reduction in threshold dispersion by a factor of about 10. From (3.21), once the required correction factor cf has been specified, the minimum theoretical resolution n_{min} of the correction DAC can be derived as

$$n_{min} = \left\lceil d \cdot \ln \left(\frac{c}{cf} \right) \right\rceil, \quad (3.22)$$

where $\lceil x \rceil$ is the minimum integer larger than x . Fig. 3.63 shows the threshold correction factor as a function of the DAC resolution for different values of the DAC output range. For a given value of the DAC range, the correction factor as a function of the DAC resolution is found to reach an asymptotic value, which can be shown to be given by

$$\lim_{n \rightarrow +\infty} \frac{\sigma_{th,c}}{\sigma_{th}} = \left[\sqrt{\frac{2}{\pi}} \int_0^{+\infty} e^{-\frac{(u+\theta/2)^2}{2}} u^2 du \right]^{\frac{1}{2}}$$

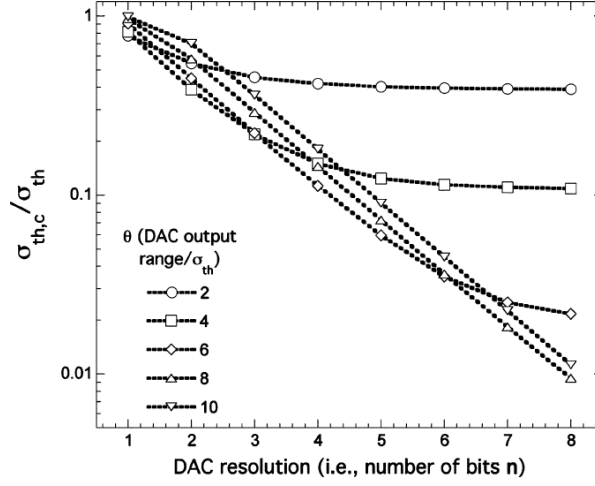


Figure 3.63: threshold correction factor as a function of the DAC resolution for different values of the DAC output range/ σ_{th} ratio.

$$= \frac{1}{2} \left[\frac{-4e^{-\theta^2/8} + \sqrt{2\pi}(4 + \theta^2) \text{Erfc}\left(\frac{\theta}{2\sqrt{2}}\right)}{\sqrt{2\pi}} \right]^{\frac{1}{2}}. \quad (3.23)$$

The locus of the asymptotic values of the correction factor, which is a function of θ , is shown in Fig. 3.64. For DAC output ranges exceeding about $6\sigma_{th}$, the distribution tails outside the DAC range can be neglected in the calculation of the post-correction threshold voltage distribution. In this case, $p_{n,\theta}(V_{th})$ can be approximately expressed as

$$p_{n,\theta}(V_{th}) \simeq \frac{2^n}{\theta\sigma_{th}} \int_{-\infty}^{+\infty} p(V_{th}) dV_{th} = \frac{2^n}{\theta\sigma_{th}}, \quad (3.24)$$

which leads to the following expression for the threshold correction factor

$$\frac{\sigma_{th,c}}{\sigma_{th}} = \frac{1}{\sigma_{th}} \left(\int_{-\frac{\theta\sigma_{th}}{2^{n+1}}}^{\frac{\theta\sigma_{th}}{2^{n+1}}} \frac{2^n}{\theta\sigma_{th}} V_{th}^2 dV_{th} \right)^{\frac{1}{2}} = \frac{\theta}{2^n\sqrt{12}}. \quad (3.25)$$

The approximated expression shown in (3.24) for the post-correction distribution is obtained by moving all the thresholds of the original distribution in a bin $\theta\sigma_{th}/2^n$ wide. Note that, from the previous equation

$$\sigma_{th,c}(\theta) = \frac{\theta\sigma_{th}}{2^n\sqrt{12}} = \frac{DACrange}{2^n\sqrt{12}} = \frac{LSB(\theta)}{\sqrt{12}} \quad (3.26)$$

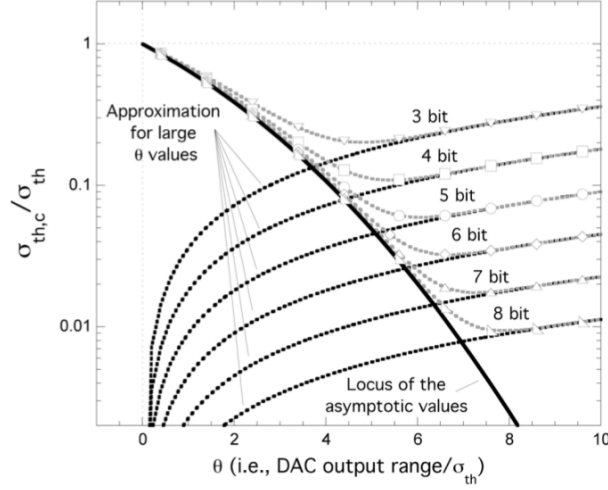


Figure 3.64: approximated correction factor curves as obtained for large θ values together with the curves obtained from the complete expression of $p_{n,\theta}(V_{th})$ given by (3.17) (same curves as in Fig. 3.60). The locus of the asymptotic values of the correction factor for large DAC resolution is also shown.

where $LSB(\theta)$ is the DAC least significant bit. Equation (3.26) indicates that, for large enough values of θ , the corrected threshold dispersion equals the quantization error of the system. The behavior of the correction factor as obtained from (3.25) for large θ values is compared in Fig. 3.64 with the curves of Fig. 3.60. As expected, at large values of the coefficient, the approximated curves are virtually indistinguishable from those obtained from the complete expression of $p_{n,\theta}(V_{th})$ given by (3.17).

Monte Carlo Model

The results presented in the previous section, in particular the correction factor curves, were obtained by means of the Mathematica software by Wolfram Research. Calculation of the correction factor involves computing $\sigma_{th,c}/\sigma_{th}$ as represented in (3.19), which includes the distribution of the corrected threshold voltage $p_{n,\theta}(V_{th})$ increases exponentially with the DAC resolution n , obviously impacting on the computing time. In order both to validate the results obtained in the previous section and to implement a faster and more versatile tool for DAC design, a model of the system, based on a Monte Carlo (MC) algorithm, has been developed in the LabVIEW environment. The block diagram of Fig. 3.65 describes the operation and features of the program. The

first block generates a random m -element vector (with the number of channels of the system) of threshold voltages. The vector elements are distributed with a Gaussian probability density function with standard deviation σ_{th} provided by the program user. The threshold correction is obtained by subtracting from each threshold voltage element of the random vector the value of the same element after quantization with a suitable stair-step function. The amount of correction that can be made to the threshold is limited by the DAC output range. This is accounted for by the limiter block following the quantizer, featuring a negative saturation level, $-q(2^n - 1)/2$ and a positive one, $+q(2^n + 1)/2$. The developed Monte Carlo model also offers the possibility to include DAC non-idealities in the analysis, in particular, as shown in Fig. 3.65, offset and gain errors, differential non linearity (DNL) and integral non linearity (INL). Fig. 3.66 shows the correction factor as a function of the correction DAC output range for different values of the DAC resolution obtained from the described Monte Carlo model. The results are perfectly consistent with the values obtained through computation with the analytical model discussed in the previous section, also displayed in the figure. As an example of the Monte Carlo tool capabilities, Fig. 12 shows the correction factor as a function of the DAC output range as obtained from MC simulations in the case of a 6-bit converter. The ideal curve is compared with four other plots accounting for the effects of differential non-linearity on the DAC correction capabilities. DNL is forced into the DAC response by generating a set of 2^n uniformly distributed, pseudorandom numbers in the range

$$\left[-\frac{DNL_{max} \cdot LSB}{2}, +\frac{DNL_{max} \cdot LSB}{2} \right] \quad (3.27)$$

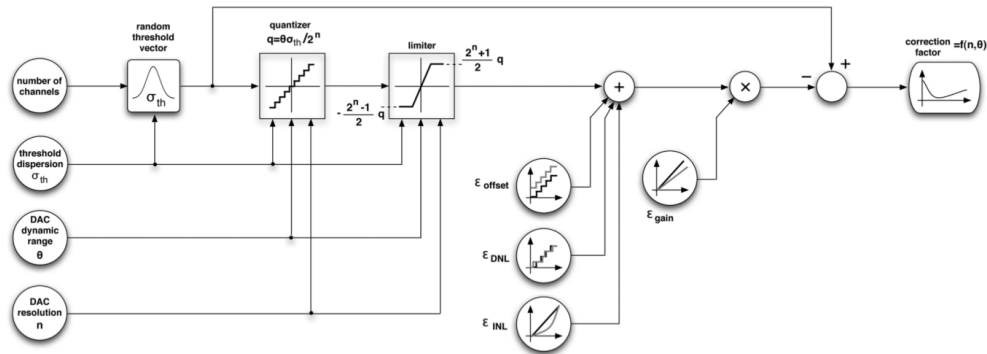


Figure 3.65: block diagram of the Monte Carlo tool for threshold correction modeling and simulation.

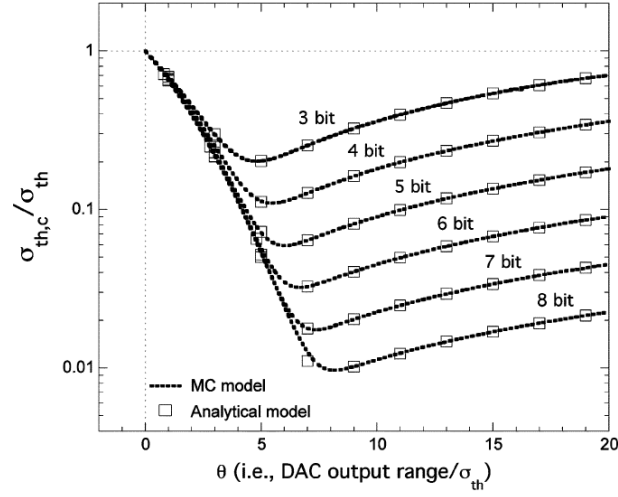


Figure 3.66: correction factor as a function of the correction DAC output range for different values of the DAC resolution n obtained from the Monte Carlo (MC) model. The curves are compared with the values obtained with the analytical model.

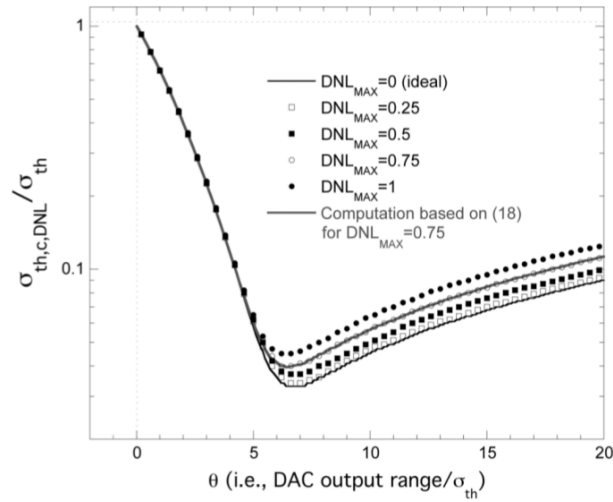


Figure 3.67: effect of differential non-linearity on the correction factor in the case of a 6 bit DAC.

to be added to the 2^n levels of the converter input-output characteristic. As a result, $|DNL| \leq DNL_{max}$. While the correction factor is barely affected by

DNL at small values, around the minimum and for larger values of the DAC output range, the effect becomes significant. This result can be explained by assuming that, in the case of a DAC affected by a differential non-linearity, the corrected threshold dispersion $\sigma_{th,c,DNL}$ accounting for the DNL can be obtained as the quadratic sum of the ideal corrected threshold dispersion $\sigma_{th,c}$ and the standard deviation of the DNL-induced error in the DAC output level which can be calculated to be $(DNL_{max} \text{ LSB})/\sqrt{12}$,

$$\begin{aligned}\sigma_{th,c,DNL} &= \sqrt{\sigma_{th,c}^2 + \left(\frac{DNL_{max} \text{ LSB}}{\sqrt{12}}\right)^2} \\ &= \sqrt{\sigma_{th,c}^2 + \left(\frac{DNL_{max}^2 \theta^2 \sigma_{th}^2}{12 \cdot 2^{2n}}\right)}\end{aligned}\quad (3.28)$$

From the previous equation, the expression of the correction factor accounting for a DNL error in the DAC can be obtained,

$$\frac{\sigma_{th,c,DNL}}{\sigma_{th}} = \sqrt{\left(\frac{\sigma_{th,c}}{\sigma_{th}}\right)^2 + \left(\frac{DNL_{max} \theta}{\sqrt{12} \cdot 2^n}\right)^2}\quad (3.29)$$

In Fig. 3.67, as an example, (3.29) has been computed in the case of $DNL_{max} = 0.75$. The resulting curve is in good agreement with the outcomes of the Monte Carlo simulations.

The Superpix1 threshold correction circuit

As already said, in order to limit the effects of the threshold dispersion σ_{th} , a local fine tuning system was designed in the Superpix1 prototype. In the Superpix1 front-end circuit, a source follower, biased by means of the I_{BIAS} current source, is placed between the shaper stage and the discriminator (see Fig. 3.68). The DC voltage at the source of the PMOS depends on the current flowing through the device. In particular, a variable current I_{DAC} , set by a current steering DAC, is added to the bias current. The DAC circuit, integrated in each pixel cell, contains 15 unity current sources selected through a thermometric code. A decoder is used to convert the 4 bit binary word, stored in a shift register, to the thermometric code used by the DAC. The design choice of using 4 bit for the threshold dispersion correction is a compromise between the DAC area and the achievable correction factor $\sigma_{th,c}/\sigma_{th}$. Both DAC and decoder circuits are laid out in the analog layer of the pixel. In the Superpix1 chip, a corrected threshold dispersion of $65 \text{ e}^- \text{ rms}$ can be achieved starting

from a pre-correction value of $560 e^-$ rms. In order to achieve the needed correction factor, the DAC has been designed in such a way to feature an output range of about $5.5\sigma_{th}$, which can be also adjusted. The PMOS M_P of Fig. 3.68 biased through the current source I_{BIAS} , acts as a level shifter between the shaper output and the discriminator input. The voltage shift V_{DAC} can be changed by varying the current in the I_{DAC} current source. In particular:

$$V_{disc} - V_{sh} = V_{DAC} = V_{DAC0} + v_{DAC} \quad (3.30)$$

where, V_{sh} is the voltage at the shaper output, V_{disc} is the voltage at the discriminator input, V_{DAC0} is the value of V_{DAC} for $I_{DAC} = 0$, while v_{DAC} depends on I_{DAC} . Fig. 3.68 shows the threshold distribution in the case of a set of 9000 channels before and after correction with the 4 bit DAC. The starting threshold distribution has been obtained through a circuit Monte Carlo simulation, accounting for random parameter variations in the microelectronic process and for the subsequent channel-to-channel mismatch in the DC operating point at the discriminator input. In the case considered in Fig. 3.69, M_P was operated at $I_b = 350$ nA, while an elementary step of 10 nA was used in the correction DAC, resulting in a correction factor of 0.117. Fig. 3.70 shows

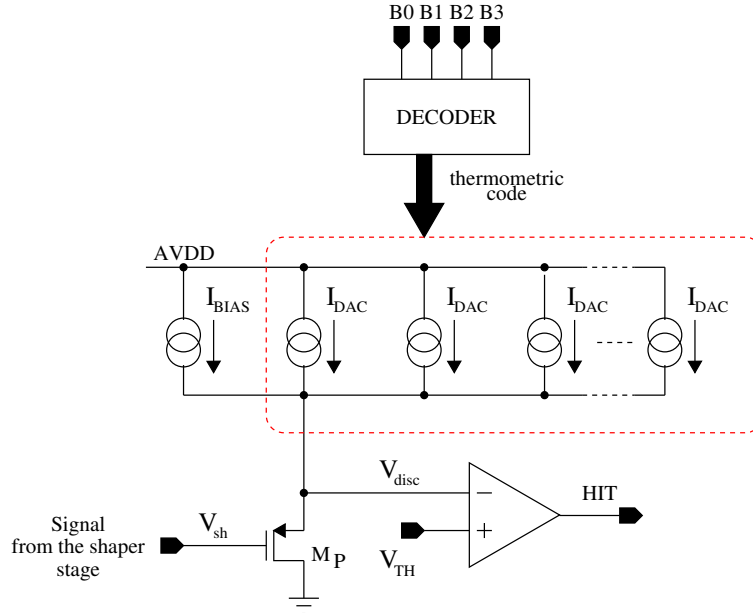


Figure 3.68: circuit schematic describing the operation of the threshold correction DAC.

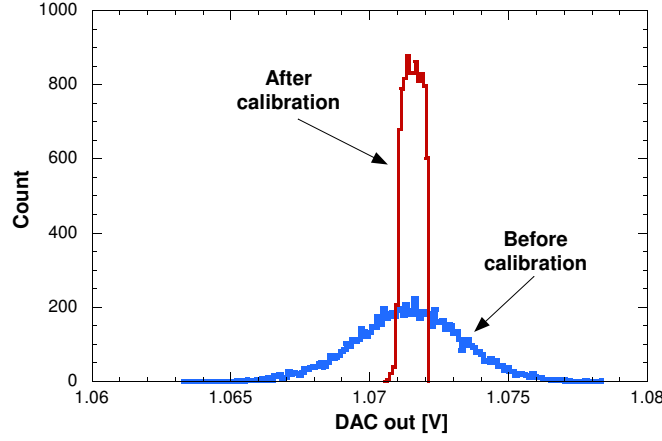


Figure 3.69: threshold distribution of the DC voltage at the inverting discriminator input before (blue) and after (red) correction. The graph is obtained with Monte Carlo simulations of 9000 analog channels.

the correction factor $\sigma_{th,c}/\sigma_{th}$ as a function of the parameter θ as obtained through circuit simulations (open square markers). The simulation data are compared to the ideal threshold correction curve resulting from 3.19 for $n = 4$. They can be observed to be in fair agreement, especially around the minimum values of the plot. At larger output ranges of the DAC, a slight discrepancy (about 10% in the worst case) can be detected between circuit simulations and theoretical curve. In order to understand the results of Fig. 3.70, the operation of the correction circuit of Fig. 3.68 is worth a more detailed treatment. Even in the ideal case where the DAC controlling the I_{DAC} current is not affected by any non-linearity, the change in V_{DAC} , i.e. v_{DAC} , corresponding to a change in I_{DAC} , is intrinsically non linear with the programming code of the converter due to the non linear $I_D - V_{GS}$ characteristic of the MOSFET M_P . In the simulations shown in Fig. 3.70 M_P , which features $W/L = 10 \mu\text{m}/4 \mu\text{m}$, is made to work at currents ranging from 150 nA to 3 μA , corresponding to a normalized drain current $I_{D,norm} = I_D \cdot L/W$ ranging from 60 nA to 1.2 μA . If compared to the characteristic normalized drain current I_z^* , which is by definition located at the center of the moderate region of operation, separating the weak from the strong inversion region, the considered interval of I_b values forces the device to operate in moderate inversion. Actually, although experimental data are not available for the 130 nm CMOS technology considered here, an I_z^* of a few hundreds of nA can be extracted from simulations. This is also in fair agreement with the value of I_z^* obtained from experimental

characterization of PMOS devices belonging to a different 130 nm process [51]. In the moderate region of operation, no simple representation of the $I_D - V_{GS}$ relationship can be used, as it would be the case instead in the weak or in the strong inversion approximations. Still, some general considerations can be made by referring to Fig. 3.71, representing the small signal behavior of the threshold correction circuit at each incremental step of the I_{DAC} current. The circuit also accounts for the body effect in the M_P transistor through the bulk transconductance $g_{mb,j-1}$. In the following, the DAC will be assumed to be ideal. Therefore, if

$$I_{DAC,j} = j \cdot I_{cell} \quad (3.31)$$

where I_{cell} is the current of the unity cell of the DAC, $j = 0, \dots, 2^n - 1$ and n is the DAC resolution, then

$$\Delta I_{DAC,j} = I_{DAC,j} - I_{DAC,j-1} = I_{cell} \quad (3.32)$$

is independent of j . From the small signal model of Fig. 3.71,

$$\Delta v_{DAC,j} = \frac{I_{cell}}{g_{m,j-1} + g_{mb,j-1}}, \quad (3.33)$$

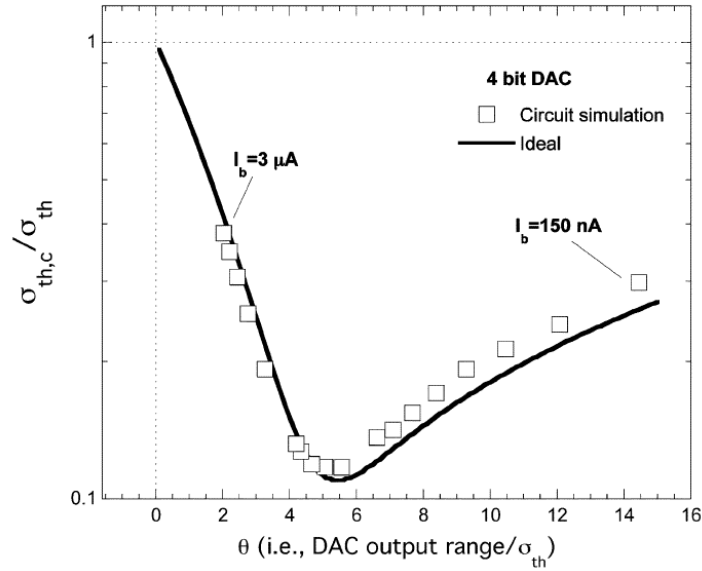


Figure 3.70: simulated threshold correction factor compared with the performance of an ideal DAC and of the Superpix1 current steering DAC with its intrinsic non-linearity.

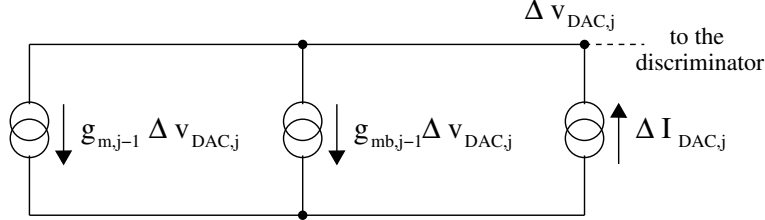


Figure 3.71: small equivalent circuit of the threshold correction DAC at each incremental step of the I_{DAC} current.

where $\Delta v_{DAC,j}$ is the change in $v_{DAC,j}$ due to j -th step in the DAC current and $g_{m,j-1}$ and $g_{mb,j-1}$ are the channel and the bulk transconductance respectively as they result from the $(j-1)$ -th incremental in I_{DAC} . If $v_{DAC}(j)$ is the voltage change at the output of the correction circuit after the j -th step of the DAC current, then

$$v_{DAC}(j) = \sum_{i=1}^j \Delta v_{DAC,i} = I_{cell} \cdot \sum_{i=1}^j (g_{m,j-1} + g_{mb,j-1})^{-1} \quad (3.34)$$

Note that, since both the channel and the bulk transconductances increase with the drain current, then $\Delta v_{DAC,j} < \Delta v_{DAC,j+1}$ from which it can be concluded that $v_{DAC,j}$ is not linear with I_{DAC} . A non linearity error $\epsilon(j)$ can be defined as

$$\epsilon(j) = \frac{v_{DAC}(j) - v_{DAC,id}(j)}{LSB_{id}}, \quad (3.35)$$

where

$$LSB_{id} = \frac{\sum_{i=1}^{2^n-1} \Delta v_{DAC,i}}{2^n - 1} \quad (3.36)$$

and

$$v_{DAC,id}(j) = j \cdot LSB_{id}. \quad (3.37)$$

Obviously, $\epsilon(j) = 0$ would indicate that the correction circuit is linear with I_{DAC} . Fig. 3.72 shows the non linearity error as obtained from circuit simulations of the 4-bit DAC at different values of the bias current. The non-linearity, which can be seen to increase with decreasing I_b , can actually provide an explanation for the slight discrepancy observed in Fig. 3.70 between the simulation points and the ideal curve. If (3.34) is used in the definition of $\epsilon(j)$, then

$$\epsilon(j) = (2^n - 1) \cdot \frac{\sum_{i=1}^j (g_{m,j-1} + g_{mb,j-1})^{-1}}{\sum_{i=1}^{2^n-1} (g_{m,j-1} + g_{mb,j})^{-1}} - j. \quad (3.38)$$

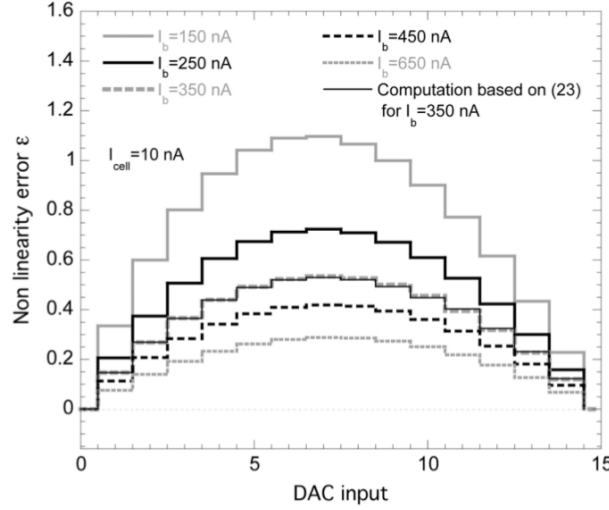


Figure 3.72: non linearity error as resulting from circuit simulations.

Again in Fig. 3.72, the non-linearity error as obtained from simulations by straightforward application of (3.35) is compared with the error computed by means of (3.38) with $g_{m,j-1}$ and $g_{mb,j-1}$ provided by circuit simulations. The two staircase curves are in very good agreement, demonstrating that the small signal model of Fig. 3.71 is able to predict the nonlinear behavior of the correction circuit.

3.4 Layout

Fig. 3.73 and Fig. 3.74 show the Superpix1 pixel cell layout. The front-end circuits are divided in the two layers, available in the Tezzaron/Globalfoundries process. The charge sensitive preamplifier, the polarity selector and the shaper stage are located in the bottom, thin layer, of the pixel cell. In order to contact the bump bond pad, for the connection to the sensor, three supercontact have been laid out near the preamplifier input. The bump bond pad has the same dimension as the one designed for Superpix0, therefore making the Superpix1 chip compatible with the same sensors used in the first 2D prototype. The threshold correction circuits, consisting of the digital to analog converter and the binary to thermometric decoder are emphasized with blue box in Fig. 3.73. A shift register is also included for the storage of the configuration bit: 4 bit for the DAC (B0, B1, B2, B3), 1 bit for enabling the charge injection (INJ),

1 bit for disabling the pixel digital front-end. The threshold discriminator is located on the top layer of the pixel cell and is directly connected to the in-pixel logic through several inter-tier pads, showed in more detail in Fig. 3.75. The digital layer layout, shown in Fig. 3.74, contains the in-pixel logic and the injection circuit. A large portion of the pixel cell is dedicated to the $226.5\text{ k}\Omega$ resistance used in the pulser, integrated with a poly layer, which provides the highest integration density among the options offered by the Globalfoundries process. As already said, the inter-tier bond pads provide both mechanical and electrical connection between the face-to-face bonded wafers. Each single connection between the two layers is performed by means of four inter-tier bond pads. This design choice was made in order to achieve some degrees of tolerance to a possible misalignment of the two layers. For this reason, the five inter-layer connections, relevant to analog power lines (AVDD and AGND) necessary for the pulser circuit, the discriminator output signal (HIT), the injection enable bit (INJ) and the mask bit (MASK), are located as far as possible from each other. Particular care has been taken in the design of the inter-tier connections, as they are very sensitive to capacitive coupling and, in addition, subtract area from the metal layers used for power distribution, therefore increasing their resistance.

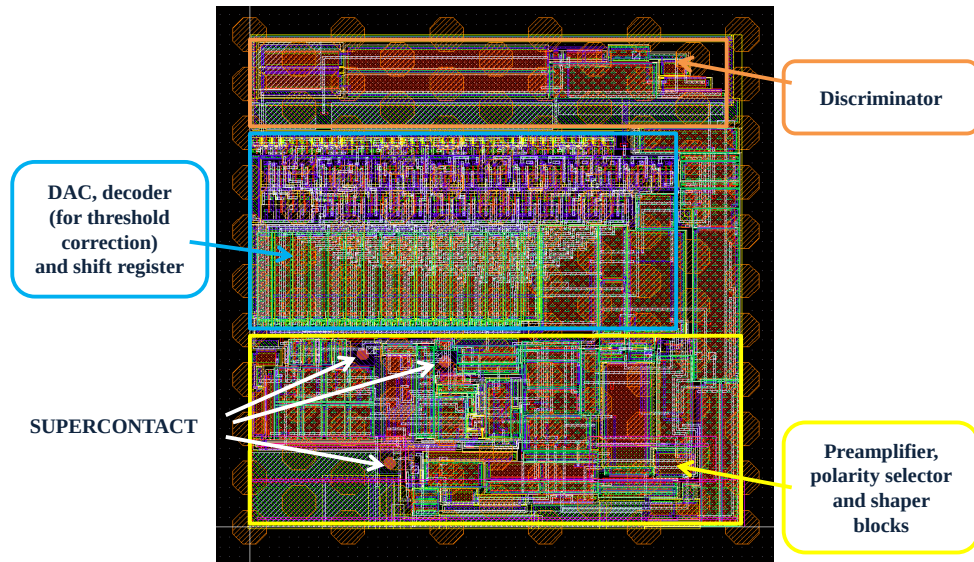


Figure 3.73: Superpix1 analog bottom layer layout.

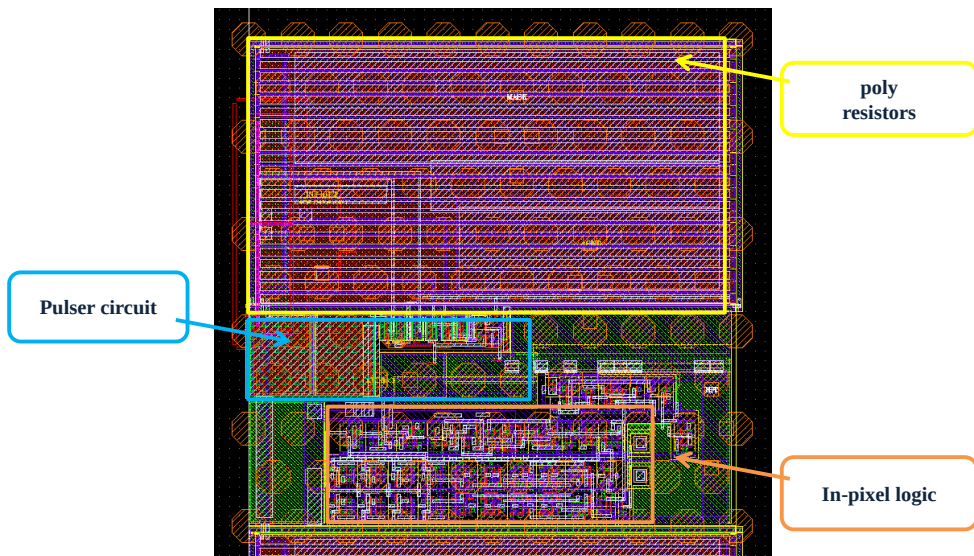


Figure 3.74: Superpix1 digital top layer layout.

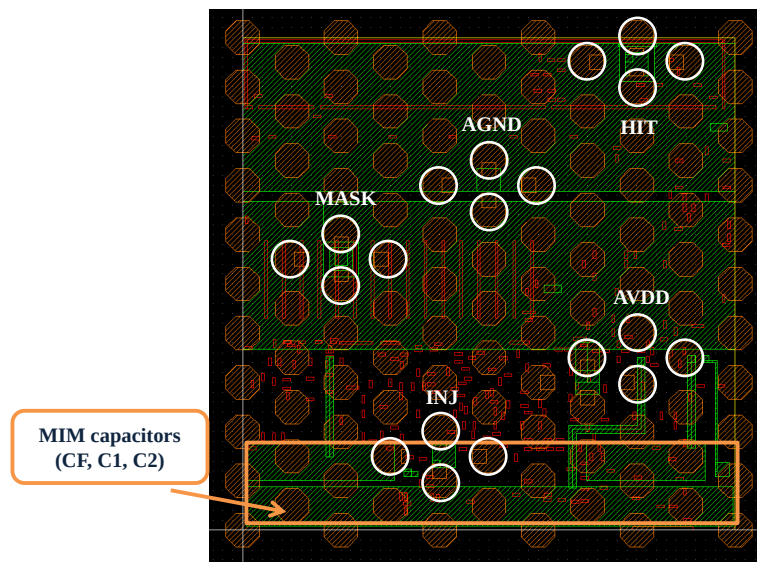


Figure 3.75: analog layer layout with emphasis on the inter-tier connections.

Conclusions

In this thesis work, the design and the experimental characterization of a front-end chip for hybrid pixels, fabricated in a planar, 130 nm CMOS process, for applications to the SVT Layer0 of the SuperB Factory has been presented. This front-end chip, also called Superpix0, was characterized in terms of charge sensitivity, noise and threshold dispersion through laboratory tests. Characterization of chips interconnected to high resistivity sensors has been successfully performed with radioactive sources. The Superpix0 hybrid sensor has been also characterized from the standpoint of detection efficiency through tests on a particle accelerator beam. A second prototype, called Superpix1 and conceived, like the first one, for application to the SVT Layer0, has been designed in a vertical integration, 130 nm CMOS technology, whose properties have been exploited to increase the functional density of the mixed-signal front-end circuits. The Superpix1 chip includes new features with respect to Superpix0, such as a threshold correction stage and a charge injection circuit. First experimental results from the test of a DNW MAPS prototype in the same 3D technology have also been presented demonstrating the functionality of the vertical integration process. Fabrication of the Superpix1 prototype is expected for the first semester of 2013. Thorough characterization of the chip will follow, including tests on a particle beam for detection efficiency measurements.

Bibliography

- [1] E. H. M. Heijne, “Semiconductor micropattern pixel detectors: a review of the beginnings,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 465, pp. 1-26, 2001
- [2] R. Muller, T. Kamins, “*Device Electronics for Integrated Circuits, 3rd Edition*,” edited by Wiley, 2003
- [3] L. Ratti et al., “Monolithic pixel detectors in a 0.13 μm CMOS technology with sensor level continuous time charge amplification and shaping,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 568, iss. 1, Pages 159-166, Nov. 2006.
- [4] P. Delpierre et al., “The DELPHI Very Forward Tracker for LEP200,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 367, pp. 198-201, 1995.
- [5] H. Beker et al., “A hybrid silicon pixel telescope tested in a heavy-ion experiment,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 332, pp. 188-201, 1993
- [6] L. Blanquart et al., “Pixel readout electronics for LHC and biomedical applications,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 439, pp. 403-412, 2000
- [7] P. Fischer, “Pixel electronics for the ATLAS experiment,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 465, pp. 153-158, 2001
- [8] R. Baur, “Readout architecture of the CMS pixel detector,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 465, pp. 159-165, 2001

- [9] D.C. Christian et al., “Development of a pixel readout chip for BTeV,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 435, pp. 144-152, 1999
- [10] The ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider,” 2008 JINST 3 S08003, doi:10.1088/1748-0221/3/08/S08003
- [11] “Inner Detector: Technical Design Report 1,” The ATLAS Collaboration, Available online at <http://cdsweb.cern.ch/record/331063>
- [12] “Inner Detector: Technical Design Report 2,” The ATLAS Collaboration, Available online at <http://cdsweb.cern.ch/record/331064>
- [13] I. Peric et al., “The FEI3 readout chip for the ATLAS pixel detector,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol 565, no. 1, pp. 178-187, 2006.
- [14] J. Grosse-Knetter, “The ATLAS pixel detector,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 568
- [15] G. Anelli, F. Faccio, S. Florian, P. Jarron, “Noise characterization of a 0.25 μm CMOS technology for the LHC experiments,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol 457, no. 1, pp. 361-368, 2001.
- [16] A. Dominiguez, “The CMS pixel detector,” vol 581, pp. 343-346, 2007.
- [17] “The SuperB Conceptual Design Report”, INFN/AE-07/02, SLAC-R-856, LAL 07-15. Available online at <http://www.pi.infn.it/SuperB>
- [18] Babar “The Babar collaboration” Available online at <http://www.slac.stanford.edu/BFROOT>
- [19] “The Belle collaboration” Available online at <http://belle.kek.jp>
- [20] B. Aubert on behalf of the BABAR Collaboration, “The BABAR detector,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 479, no. 1, pp. 1-116, Feb. 2002.
- [21] A. Huffman, “Fabrication, Assembly and Evaluation of Cu-Cu Bump-Bond Arrays for Ultra-fine Pitch Hybridization and 3D Assembly,” presented at *Pixel 2008 International Workshop*, September 22-26, 2008, Fermilab, Batavia (IL), USA

- [22] G. Rizzo et al., "Recent Development on Triple Well 130nm CMOS MAPS of Deep N-Well MAPS with In-Pixel Signal Processing and Data Sparsification Capability," *2007 IEEE Nuclear Science Symposium Conference Record*, pp. 927-930, Oct. 26 2007-Nov. 3 2007.
- [23] A. Gabrielli et al., "Proposal of a data sparsification unit for a mixed-mode MAPS detector," *2007 IEEE Nuclear Science Symposium Conference Record*, vol. 2, pp. 1471-1473, Oct. 26 2007-Nov. 3 2007.
- [24] N. Zorzi, et al., "Fabrication and characterization of n-on-n silicon pixel detectors compatible with the Medipix2 readout chip," *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 546, no. 1, pp. 46-50, 2005.
- [25] F. Huegging, et al., "Design and test of pixel sensors for operation in severe radiation environments", *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 439, pp. 529-535, 2000.
- [26] L. Ratti et al., "Design of Time Invariant Analog Front-End Circuits for Deep N-Well CMOS MAPS," *IEEE Transactions on Nuclear Science*, vol. 56, no. 4, pp. 2360-2373, Aug. 2009.
- [27] H.L. Hughes, J.M Benedetto, "Radiation effects and hardening of MOS technology: devices and circuits," *IEEE Transactions on Nuclear Science*, vol. 50, no. 3, pp. 500-521, 2003.
- [28] N. S. Saks, S. Nelson, M. G. Ancona, J. A. Modolo, A. John, "Generation of Interface States by Ionizing Radiation in Very Thin MOS Oxides," *IEEE Transactions on Nuclear Science*, vol. 33, no. 6, pp. 1185-1190, 1986.
- [29] G. Anelli et al., "Radiation tolerant VLSI circuits in standard deep sub-micron CMOS technologies for the LHC experiments: practical design aspects," *IEEE Transactions on Nuclear Science*, vol. 46, no. 6, pp. 1690-1696, 1999.
- [30] L. Ratti, M. Manghisoni, V. Re, G. Traversi, "Design Optimization of Charge Preamplifiers With CMOS Processes in the 100 nm Gate Length Regime," *IEEE Transactions on Nuclear Science*, vol. 56, no. 1, pp. 235-242, 2009.
- [31] V. Re, L. Gaioni, M. Manghisoni, L. Ratti, G. Traversi, "Mechanisms of Noise Degradation in Low Power 65 nm CMOS Transistors Exposed to Ionizing Radiation," *IEEE Transactions on Nuclear Science*, vol. 57, no. 6, pp. 3071-3077, 2010.

- [32] Y. Tsividis, "Operation and modeling of the MOS transistor," edited by McGraw-Hill, Boston, 1999.
- [33] G. Rizzo et al., "Development of deep N-well MAPS in a 130 nm CMOS technology and beam test results on a 4k-pixel matrix with digital sparsified readout" *2008 IEEE Nuclear Science Symposium Conference Record*, pp. 3242-3247, 19-25 Oct. 2008
- [34] A. Gabrielli et al., "On-chip fast data sparsification for a monolithic 4096-pixel device", *IEEE Transactions on Nuclear Science*, vol. 56, iss. 3, pp. 1159-1162, 2009.
- [35] M. Villa et al, "Beam-test results of 4k pixel CMOS MAPS and high resistivity triplet detectors equipped with digital sparsified readout in the Slim5 low mass silicon demonstrator", *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 617, pp. 596-596, 2010.
- [36] P. Garrou, C. Bower, P. Ramm, "*Handbook of 3D Integration: Technology and Application of 3D Integrated Circuit*," edited by WILEY-VCH, Weinheim, 2008.
- [37] Jian-Qiang Lu, "3-D Hyperintegration and Packaging Technologies for Micro-Nano Systems," *2009 IEEE Nuclear Science Symposium Conference Record*, vol. 97, no. 1, pp. 18-30, Jan. 2009
- [38] B. Patti, "3D Bonding At Tezzaron", presented at the *Pixel 2008 International Workshop*, Fermilab, Batavia, September 23-26 2008.
- [39] A. W. Topol, D. C. La Tulipe, L. Shi, D. J. Frank, K. Bernstein, S. E. Steen, A. Kumar, G. U. Singco, A. M. Young, K. W. Guarini, M. Jeong, "Three-Dimensional Integrated Circuits," *IBM Journal of Research and Development*, vol. 50, no. 4-5, July 2006.
- [40] G. Traversi, A. Bulgheroni, M. Caccia, M. Jastrzab, M. Manghisoni, E. Pozzati, L. Ratti, V. Re, "Design and Performance of a DNW CMOS Active Pixel Sensor for the ILC Vertex Detector," *IEEE Transactions on Nuclear Science*, vol. 56, no. 5, pp. 3002-3009, Oct. 2009
- [41] G. Rizzo et al., "Triple Well CMOS Active Pixel Sensor with In-Pixel Full Signal Analog," *2005 IEEE Nuclear Science Symposium Conference Record*, vol. 3, pp. 1485-1489, 23-29 Oct. 2005

- [42] L. Ratti et al., “Vertically integrated deep N-well CMOS MAPS with sparsification and time stamping capabilities for thin charged particle trackers,” *Nuclear Instrumentation and Methods in Physics Research Section A*, vol. 624, no. 2, pp. 379-386, Dec. 2011.
- [43] G. Traversi, “Recent results and plans of the 3D IC consortium,” *Proceedings of Science*, VERTEX, vlo. 32, 2010
- [44] L. Ratti, “Continuous Time-Charge Amplification and Shaping in CMOS Monolithic Sensors for Particle Tracking,” *IEEE Transactions on Nuclear Science*, vol. 53, no. 6, pp. 3918-3928, Dec. 2006.
- [45] G. Traversi, M. Manghisoni, L. Ratti, V. Re, V. Speziali, “CMOS MAPS with pixel level sparsification and time stamping capabilities for applications at the ILC,” *Nuclear Instruments and Methods in Physics Research Section A*, vol 581, iss. 12, pp 291-294, Oct. 2007.
- [46] S. O. Rice, “Mathematical analysis of random noise,” *Bell Syst. Tech. J.* vol. 24, pp. 46156, 1945.
- [47] J. F. Colombeau, “Generalized Functions and Infinitesimal,” arXiv:math/0610264v1 [math.FA].
- [48] F. Maloberti, “*Data Converters*”, edited by Springer, Boston, 2001
- [49] Y. P. Tsividis, “*Operation and Modeling of the MOS Transistor*,” 2nd ed. New York: McGraw-Hill, 1999.
- [50] M. Manghisoni, E. Quartieri, L. Ratti, G. Traversi, ‘ ‘High accuracy injection circuit for pixel-level calibration of readout electronics,’ ‘ *2010 IEEE Nuclear Science Symposium Conference Record*, pp. 1312-1318, Oct. 30 - Nov. 6 2010.
- [51] M. Manghisoni, L. Ratti, V. Re, V. Speziali, and G. Traversi, ”Resolution limits in 130 nm and 90 nm CMOS technologies for analog front-end application,” *IEEE Transactions on Nuclear Science*, vol. 54, no. 3, pp. 531537, Jun. 2007.