Summary

Characterization of monolithic CMOS pixel sensors for charged particle detectors and for high intensity dosimetry

Chapter 1

Introduction

Pixel detectors, members of the semiconductor detector family, have significantly been used since () at the first accelerator experiments for energy and position measurement. Because of their dimension (today $\sim$ 30 μm or even better) and their spatial resolution ($\sim$ 5-10 μm), with the availability of technology in 1980s they proved to be perfectly suitable for vertex detector in the inner layer of the detector.
Technological development has been costant from then on and today almost every high energy physics (HEP) experiment employs a pixels detector; hybrid pixel currently constitute the state-of-art for large scale pixel detector but experiments began to look at the more innovative monolitic active pixels (MAPS) as perspective for their future upgrades, as BelleII, or they already have installed them, as ALICE.

Requirement imposed by accelerator are stringent and they will be even more with the increase of luminosity/intensity, in terms of radiation hardness, efficiency and occupancy, time resolution, material budget and power consumption.

Qual è invece la richiesta per la dosimetria?

While CCDs pioneered the use of silicon pixels for precision tracking,

Chapter 2

Pixel detectors

2.1 Signal formation

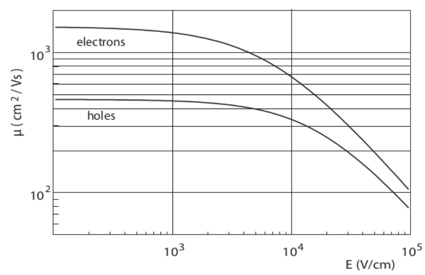
When a charge particle passes through a pixel and loses energy by ionization a part of that energy is used to generate electron-hole pairs (another part is used for other processes, as the lattice excitation) which are then separated by the electric field and collected at their respectively electrodes (p for holes and n for electrons)[1]; by the drift of these charges, a signal i_e is generated on the electrode e as stated by the Shockley–Ramo's theorem:

$$i_e(t) = -q\, v(t)\, E_{WF,e} \tag{2.1}$$

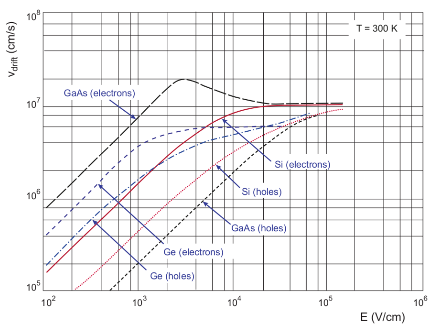
where $v(t)$ is the instantaneous velocity of the charge q and E_{WF} is the weighting field, that is the field obtained biasing the electrode e with 1V and all the others with 0V. The drift velocity of the charge depends on the electric field and on the mobility of the particle:

$$v = \mu(E)\, E \tag{2.2}$$

where $\mu(E)$ is a function of the electric field and is linear with E only for small E: at higher values the probability of interactions with optical phonons increases and the mobility drops and this leads to an independence of the velocity from the electric field (fig. 2.1b).



(a) Typical values for electrons and holes mobility in silicon at room temperature are $\mu_n \sim 1450\ cm^2/Vs$, $\mu_h = 500$



(b) Drift velocity at room temperature in different semiconductors

The average energy needed to create a pair at 300 K in silicon is $w_i = 3.65$ eV, that is more than the mean ionization energy because of the interactions with phonon, since for a minimum ionizing particle (MIP) the most probable value (MPV) of charge released in the semiconductor is 0.28 keV/μ, hence the number of e/h pairs is:

$$\langle \frac{dE}{dx} \rangle \frac{1}{w_i} \sim 80\, e/h \sim \frac{1.28\, 10^{-2} fC}{\mu m} \tag{2.3}$$

[1]Even if in principle both the electrode can be used to read a signal, for pixel detectors, where the number of channel and the complexity of readout are high, only one is actually used. In strip and pad detectors, instead, is more common a dual-side readout

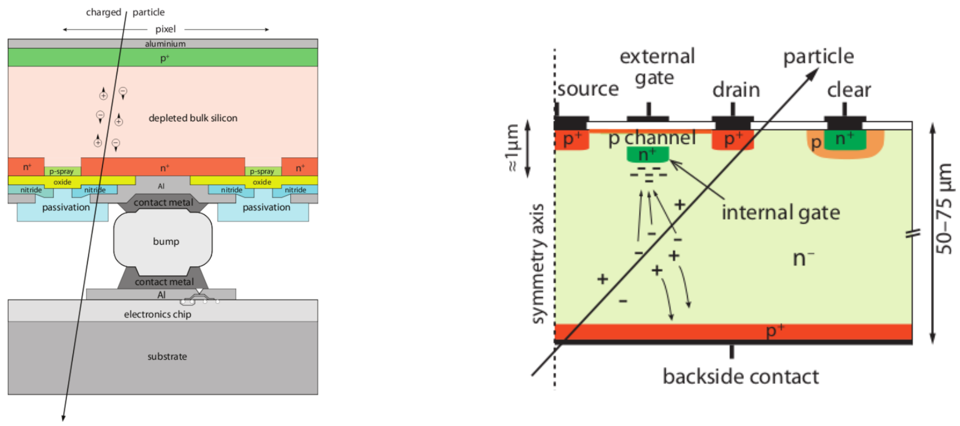


Figure 2.2: Concept cross-section of hybrid pixel (a) and of a DEPFET (b)

CON UN'INCERTEZZA CHE È RADICE DI N; ED EVENTUALEMTNE SI AGGIUNGE IL FATTORE DI FANO NEL CASO DI ASSORBIMENTO TOTALE. iL FATTORE DI FANO È 0.115 NEL SILICIO. ecc It is fundamental that pairs e/h are produced in the depleted region of the semiconductor where the probability of recombination with charge carriers is low to avoid loss of signals. Pixel detectors are then commonly reverse biased: a positive bias is given to the n electrode and a negative to the p to grow the depletion zone in the epitaxial layer below the electrode. The width of the depletion region is related with the external bias V_{ext}, the resistivity ρ and also with the dopant:

$$d_n \sim 0.55\sqrt{\frac{\rho}{\Omega cm}\frac{V_{ext}}{V}}\mu m \tag{2.4}$$

$$d_p \sim 0.32\sqrt{\frac{\rho}{\Omega cm}\frac{V_{ext}}{V}}\mu m \tag{2.5}$$

For that reason high resistivity wafers ($100\ \Omega cm - k\Omega cm$) are typically preferred because they allow bigger depletion zone with smaller voltage bias.

2.2 CCDs

ens of ms due to the need to transfer the charge signals pixel by pixel through a single output circuit For photon imaging the need of high assorbtion efficiency

2.3 Hybrid pixels

METTI IN EVIDENZAZ CHE PUOI FARE UN READOUT CON TECNOLOGIA CMOS Hybrid pixels are made of two parts (fig. 2.2a), the sensor and the electronics: for each pixel these two parts are welded together through microconnection (bump bond).
They provide a practical system where readout and sensor can be optimized separately, although the testing is less easy-to-do since the sensor and the R/O must be connected together before.
In addition, the particular and sophisticated procedure to bond sensor and ASIC (application specific integrated circuit) makes them difficult to produce, delicate, especially when exposed to high levels of radiation, and also expensive.
A critical parameter for accelerator experiments is the material budget, which represents the main limit factor for momentum measurement resolution in a magnetic field; since hybrid pixels are thicker ($\sim$ hundreds of μm) than monolithic ones (even less than 100 μm), using the latter the material budget can be down by a third: typical value for hybrid pixels is 1.5 % X_0 per layer, while for monolithic 0.5 % X_0.
Among other disadvantages of hybrid pixels there is the bigger power consumption that implies, by the way, a bigger cooling system leading in turn to an increase in material too.

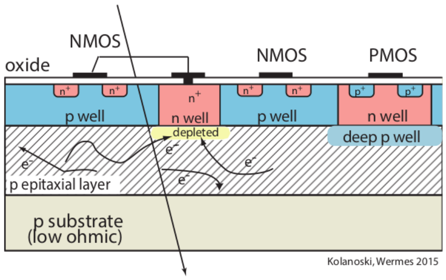


Figure 2.3: Concept cross-section of CMOS MPAS pixel

DEPFET are the first attempt towards the integration of the front end (FE) on the sensor bulk: they are typically mounted on a hybrid structure but they also integrate the first amplification stage.
Each pixel implements a MOSFET (metal-oxide-semiconductor field-effect transistor) transistor (a p-channel in fig. 2.2b): an hole current flows from source to drain which is controlled by the external gate and the internal gate together. The internal gate is made by a deep $n+$ implant towards which electrons drift after being created in the depletion region (to know how the signal is created in a pixel detector look at appendix A); the accumulation of electrons in the region underneath the n implant changes the gate potential and controls the transistor current.
DEPFET typically have a good S/N ratio: this is principally due the amplification on-pixel and the large depletion region. But, since they need to be connected with ASIC the limiting factor still is the material budget.

2.4 CMOS MAPS and DMPAS

With respect to CCDs, the radiation tolerance could be greatly increased by sensing the signal charge within its own pixel, instead of transporting it over thousands of pixels. The readout speed could also be dramatically increased by in-pixel amplitude discrimination, followed by sparse readout of only the hit pixels Monolithic active pixels accommodate on the same wafer both the sensor and the front end electronics, with the second one implanted on top within a depth of about 1 µm below the surface.
MAPS have been first proposed and realized in the 1990s and their usage has been enabled by the development of the electronic sector which guarantees the decrease in CMOS transistors dimension at least every two years, as stated by the Moore's law[2].
As a matter of fact the dimension of components, their organization on the pixel area and logic density are important issues for the design and for the layout; typically different decisions are taken for different purposes.

Monolithic active pixel can be distinguished between two main categories: MAPS and depleted MAPS (DMAPS).
MAPS (figure a 2.3) have typically an epitaxial layer in range 1-20 μm and because they are not depleted, the charge is mainly collected by diffusion rather than by drift. This makes the path of charges created in the bulk longer than usual, therefore they are slow (of order of 100 ns) and the collection could be partial expecially after the irradiation of the detector (look at A for radiation damages), when the trapping probability become higher.
In figure 2.3 is shown as example of CMOS MAPS: the sensor in the scheme implements an n well as collection diode; to avoid the others n wells (which contain PMOS transistor) of the electronic circuit would compete in charge collection and to shield the CMOS circuit from the substrate, additionally underlying deep p well are needed. DMAPS are instead MAPS depleted with d typically in $\sim$ 25-150 μm (eq. 2.1) which extends from the diode to the deep p-well, and sometimes also to the backside (in this case if one wants to collect the signal also on this electrode, additional process must be done).

[2]Moore's law states that logic density doubles every two years.

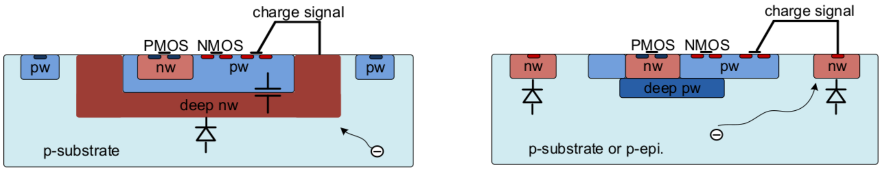


Figure 2.4: Concept cross-section with large and small fill factor

2.4.1 DMAPS: large and small fill factor

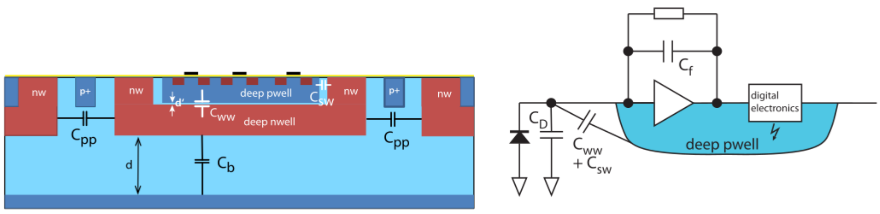
There are two different sensor-design approaches (figure 2.4) to DMAPS:

- large fill factor: a large collection electrode that is a large deep n-well and that host the embedded electronics
- small fill factor: a small n-well is used as charge collection node

To implement a uniform and stronger electric field, DMAPS often uses large electrode design that requires multiple wells (typically four including deep n and p wells); this layout adds on to the standard terms of the total capacity of the sensor a new term (fig. 2.5), that contributes to the total amplifier input capacity. In addition to the capacity between pixels (C_{pp}) and between the pixel and the backside (C_b), a non-negligible contribution comes from the capacities between wells (C_{SW} and C_{WW}) needed to shield the embedded electronics. These capacities affect the thermal and 1/f noise of the charge amplifier and the τ_{CSA} too:

$$ENC^2_{thermal} \propto \frac{4}{3}\frac{kT}{g_m}\frac{C_D^2}{\tau_{sh}} \qquad (2.6) \qquad\qquad \tau_{CSA} \propto \frac{1}{g_m}\frac{C_D}{C_f} \qquad (2.7)$$

where g_m is the transconductance, τ_{sh} is the shaping time.
Among the disadvantages coming from this large input capacity could be the coupling between the sensor and the electronics resulting in cross talk: noise induced by a signal on neighbouring electrodes; indeed, since digital switching in the FE electronics do a lot of oscillations, this problem is especially connected with the intra wells capacities. So, larger charge collection electrode



(a) (b)

Figure 2.5: C_{pp}, C_b, C_{WW}, C_{SW}

sensors provide a uniform electric field in the bulk that results in short drift path and so in good collection properties, especially after irradiation, when trapping probability can become an issue. The drawback of a large fill-factor is the large capacity (~100 fF): this contributes to the noise and to a speed penalty and to a larger possibility of cross talk.

The small fill-factor variant, instead, benefits from a small capacity (5-20 fF), but suffers from a not uniform electric field and from all the issue related to that. Ho già detto prima parlando dei MAPS, devo ripetere qui?
As we'll see these two different types of sensor require different amplifier: the large elcrtrode one is coupled with the charge sensitive amplifier, while the small one with voltage amplifier (sec 2.5.1).

2.4.2 A modified sensor

A process modification developed by CERN in collaboration with the foundries has become the standard solution to combine the characteristics of a small fill factor sensor (small input amplifier

	small fill factor	large fill factor
small sensor C	$\surd$ (< 5 fF)	$\times$ ($\sim$ 100-200 fF)
low noise	$\surd$	$\times$
low cross talk	$\surd$	$\times$
velocity perfomances	$\surd$	$\times$ ($\sim$ 100 ns)
short drift paths	$\times$	$\surd$
radiation hard	$\times$	$\surd$

Table 2.1: Small and large fill factor DMAPS characteristics

capacity) and of large fill factor sensor (uniform electric field) is the one carried out for ALICE upgrade about ten years [1].
A compromise between the two sensors could also be making smaller pixels, but this solution requires reducing the electronic circuit area, so a completely new pixel layout should be though. The modification consists in inserting a low dose implant under the electrode and one its advantage lies in its versatility: both standard and modified sensor are often produced for testing in fact.

Before the process modification the depletion region extends below the diode towards the substrate, and it doesn't extend laterally so much even if a high bias is applied to the sensor (fig. 2.6). After, two distinct pn junctions are built: one between the deep p well and the n^- layer, and the other between the n^- and the p^- epitaxial layer, extending to the all area of the sensor. Since deep p well and the p-substrate are separated by the depletion region, the two p electrodes can be biased separately[3] and this is beneficial to enhance the vertical electric field component. The doping concentration is a trimmer parameter: it must be high enough to be greater than the epitaxial layer to prevent the punchthrough between p-well and the substrate, but it must also be lower enough to allow the depletion without reaching too high bias.

2.5 Analog front end

After the creation of a signal on the electrode, the signal enters the front end circuit (fig.2.7), ready to be molded and transmitted out of chip. Low noise amplification, fast hit discrimination and an efficient, high-speed readout architecture, consuming as low power as possible must be provided by the readout integrated electronics (ROIC).
Let's take a look to the main steps of the analog front end chain: the preamplifier (that actually often is the only amplification stage) with a reset to the baseline mechanism and a leakage current compensation, a shaper (a band-pass filter) and finally a discriminator. The whole chain must be optimized and tuned to improve the S/N ratio: it is very important both not to have a large noise before the amplification stage in order to not multiply that noise, and chose a reasonable threshold of the discriminator to cut noise-hits much as possible.

2.5.1 Preamplifier

Even if circuits on the silicon crystal are only constructed by CMOS, a preamplifier can be modeled as an operational amplifier (OpAmp) where the gain is determined by the input and feedback impedance (first step in figure 2.7):

$$G = \frac{v_{out}}{v_{in}} = \frac{Z_f}{Z_{in}} \tag{2.8}$$

Depending on whether a capacity or a resistance is used as feedback, respectively a charge or a voltage amplifier is used: if the voltage input signal is large enough and have a sharp rise time, the voltage sensitive preamplifier is preferred. Consequently, this flavor doesn't suit to large fill factor MAPS whose signal is already enough high: $v_{in} = Q/C_D \approx$ 3fC/100 pF = 0.03 mV, but it's fine for the small fill factor ones: $v_{in} = Q/C_D \approx$ 3fC/3 pF = 1 mV.

In the case of a resistor feedback, if the signal duration time is longer than the discharge time ($\tau = R_S C_D$) of the detector the system works as current amplifier, as the signal is immediately

[3] This is true in general, but it can be denied if other doping characteristics are implemented, and we'll see that this is the case of TJ-Monopix1

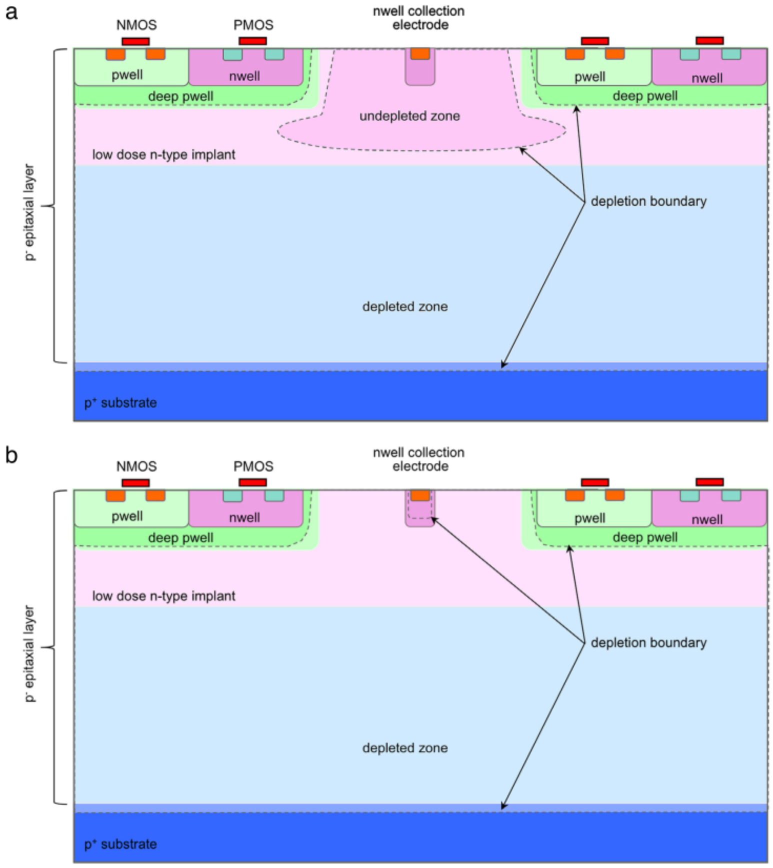


Figure 2.6: A modified process for ALICE tracker detector: a low dose n implant is used to create a planar junction. In (a) the depletion is partial, while in (b) the pixel is fully depleted.

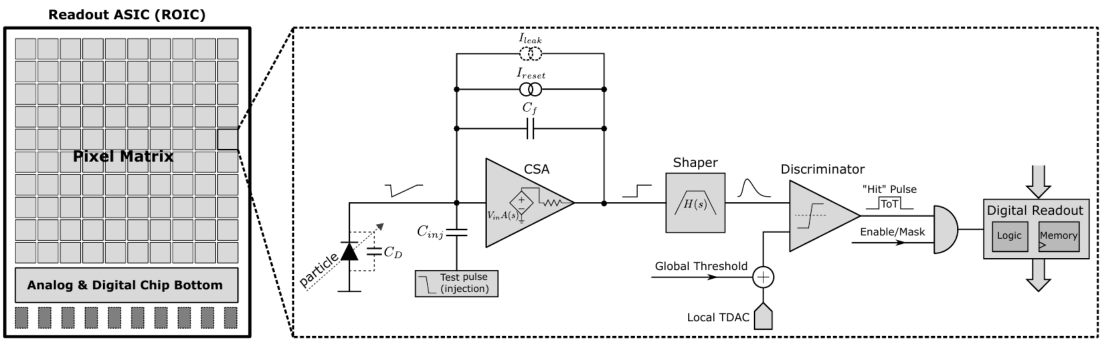


Figure 2.7: Readout FE scheme: in this example the preamplifier is a charge sensitive one (CSA) but changing the capacitive feedback into a resistive one, this can be converted in a voltage or current amplifier.

trasmit to the amplifier; in the complementary case (signal duration longer than the discharge time) the system integrates the current on the C_D and operates as a voltage amplifier.

2.6 Readout logic

Readout logic includes the part of the circuit which takes the FE output signal, processes it and then transmit it out of pixel and/or out of chip; depending on the situation of usage different readout characteristics must be provided.
To store the analogical information (i.e. charge collected, evolution of signal in time, ...) big buffers and a large bandwidth are needed; the problem that doesn't occur, or better occur only with really high rate, if one wants record only digital data (if one pixel is hit 1 is recorded, and if not 0 is recorded).

A common compromise often made is to save the time over threshold (ToT) of the pulse in clock cycle counts; this needs of relatively coarse requirement as ToT could be trimmer to be a dozen bits but, being correlated and hopefully being linear with the deposited charge by the impinging particle in the detector, it provides a sufficient information. The ToT digitalization usually takes advantage of the distribution of a clock (namely BCID, bunch crossing identification) on the pixels' matrix. The required timing precision is at least around 25 ns, that corresponds to the period of bunch collisions at LHC; for such reason a reasonable BCID-clock frequency for pixels detector is 40 MHz.
Leading and trailing edges' timestamp of the pulse are saved on pixel within a RAM until they have been read, and then the ToT is obtained from their difference.

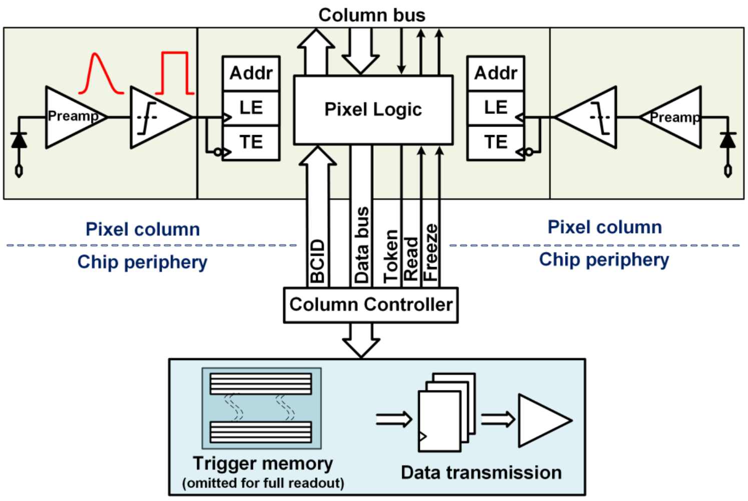


Figure 2.8: Column drain R/O scheme where ToT is saved

Moreover, the readout architecture can be full, if every hit is read, or triggered, if a trigger system decides if the hit must be store or not. On one hand the triggered-readout needs buffers and storage memories, on the other the full readout, because there is no need to store hit data on chip, needs an high enough bandwidth.
A triggered readout is fundamental in accelerator experiments where the quantity of data to store is too large to be handled, and some selections have to be applied by the trigger: to give an order of growth, at LHC more than 100 TBit/s of data are produced, but the storage limit is about 100 MBit/s [2] (pag. 797).
Typically the trigger signal is processed in a few μs, so the pixel gets it only after a hundred clock cycles from the hit arrival time: the buffer depht must than handle the higher trigger latency.

After having taken out the data from the pixel, it has to be transmitted to the end of column (EoC) where a serializer deliver it out of chip, typically to an FPGA.
There are several ways of transmitting data from pixel to the end of column: one of the most famous is the column-drain read out, developed for CMS and ATLAS experiments [3]. All the pixels in a double-column share a data bus and only one pixel at a time, according to a priority chain, can be read. The reading order circuit is implemented by shift register (SR): when a hit

arrives, the corresponding data, which can be made of timestamp and ToT, is temporarily stored on a RAM until the SH does not allow the access to memory by data bus.
Even if many readout architectures are based the column-drain one, it doesn't suit for large size matrices. The problem is that increasing the pixels on a column would also raise the number of pixels in the priority chain and that would result in a slowdown of the readout.

If there isn't any storage memory, the double-column behaves as a single server queue and the probability for a pixel of waiting a time T greater than t, with an input hit rate on the column μ and an output bandwidth B_W is [4]:

$$P(T > t) = \frac{\mu}{B_W} e^{-(B_W - \mu)t} \tag{2.9}$$

To avoid hit loss (let's neglect the contribution to the inefficiency of the dead time τ due to the AFE), for example imposing $P(T > t) \sim 0.001$, one obtains $(B_W - \mu)\, t_t \sim 6$, where t_t is the time needed to transfer the hit; since t_t is small, one must have $B_W \gg \mu$, that means a high bandwidth [4].

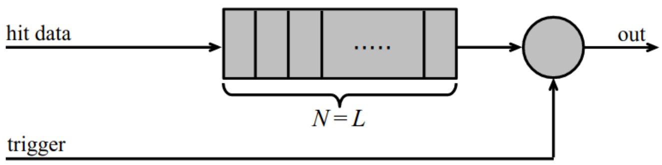


Figure 2.9: Block diagram of a pipeline buffer: N is the dimension of memory buffer and L is the trigger latency expressed in BCID cycles

Actually the previous one is an approximation since each pixel sees a different bandwidth depending on the position on the queue: the first one sees a full bandwidth, but the next sees a smaller one because occasionally it can be blocked by the previous pixel. Then the bandwidth seen by the pixel i is $B_i = B - \sum_j \mu_j$, where μ_j is the hit rate of the jth pixel.
The efficiency requirement on the bandwidth and the hit rate becomes: $B_{W,i} > \mu_i$, where the index i means the constraint is for a single pixel; if all the N pixels on a column have the same rate $\mu = N\mu_i$, the condition reduces to $B_W > \mu$. The bandwidth must be chosen such that the mean time between hits of the last pixel in the readout chain is bigger than that.
In order to reduce the bandwidth a readout with zero suppression on pixel is typically employed; this means that only information from channels where the signal exceeds the discriminator threshold are stored. QUalcosa sulla zero suprpession? La metto qui questa affermazione?

If instead there is a local storage until a trigger signal arrives, the input rate to column bus μ' is reduced compared to the hit rate μ as: $\mu' = \mu \times r \times t$, where r is the trigger rate and t is the bunch crossing period. In this situation there is a more relaxed constraint on the bandwidth, but the limiting factor is the buffer depth: the amount of memory designed depends both on the expected rate μ and on the trigger latency t as $\propto \mu \times t$, that means that the higher the trigger latency and the lower the hit rate to cope with.

In order to have an efficient usage of memory on pixels' area it's convenient grouping pixels into regions with shared storage. Let's compare two different situations: in the first one a buffer is located on each pixel area, while in the second one a core of four pixels share a common buffer (this architecture is commonly called FE-I4).

Consider a 50 kHz single pixel hits rate and a trigger latency of 5 μs, the probability of losing hits is:

$$P(N > 1|\nu) = 1 - P(N = 0|\nu) - P(N = 1|\nu) = 1 - e^{-\nu}(1 + \nu) \approx 2.6\% \tag{2.10}$$

where I have assumed a Poissonian distribution with mean $\nu = 0.25$ to describe the counts N.
To get an efficiency ϵ greater than 99.9 % a 3 hit depth buffer is needed:

$$P(N > 3|\nu) = 1 - \sum_{i=0}^{3} P(N = i|\nu) < 0.1\% \tag{2.11}$$

Considering the second situation: if the average single pixel rate is still 50 kHz, grouping four pixels the mean number of hits per trigger latency is $\nu = 0.25 \times 4 = 1$. To get an efficiency of 99.9% (eq. 2.11) a buffer depth of 5 hits in the four-pixels region, instead of 3 per pixels, is needed.

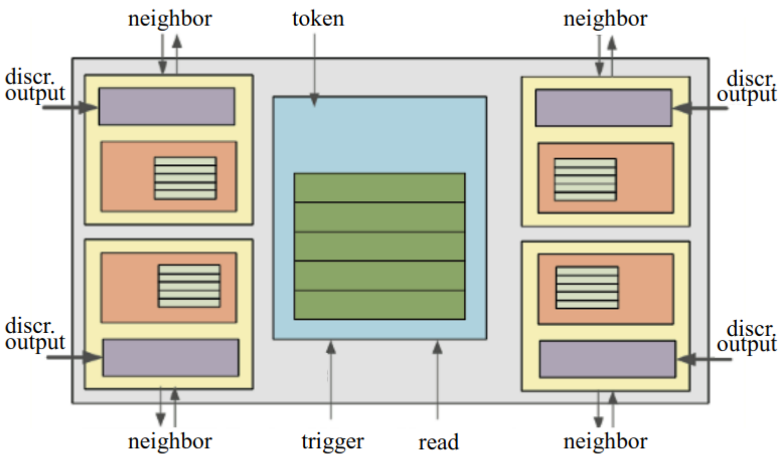


Figure 2.10: Block diagram of the FE-I4 R/O. Read and memory management section is highlighted in light blue; latency counters and trigger management secrion are highlighted in green; hit processing blocks are highlighted in purple; ToT counters and ToT management units are highlighted in orange

Chapter 3

Use of pixel detectors

There always was a tight relation between the development of cameras and pixel detectors since 1969, when the idea of CCDs, thanks to whom Boyle and Smith were awarded the Nobel Prize in Physics in 2009, revolutionized photography allowing light to be captured electronically instead of on film. Even though the CMOS technology was already known when CCDs spread, the costs of productions were too high to allow the diffusion of these sensors for which needed to wait untill 1990s. From that period on, the fast diffusion of CMOS was mainly due to the less cost than CCD, and the less power required for supply.

The principal use cases of pixel detectors are particle tracking and imaging: in the former case individual charged particles have to be identified, in the latter instead an image is obtained by the usually un-triggered accumulation of the impinging radiation. Also the demands on detectors performance depends on their usage, in particular tracking requires high spatial resolution, fast readout and radiation hardness.

3.1 Tracking in HEP

Historically, the first pixel detector employed in particle physics was a CCD: it was installed in the spectrometer at the CERN's Super Proton Synchrotron (SPS) by the ACCMOR Collaboration (Amsterdam, CERN, Cracow, Munich, Oxford, RAL) at mid 1980s, with the pourpose of studing the recently-discovered charm particles. The second famous usage of CCDs took place at SLAC in the Large Detector (SLD) during the two years 1996-98. From that period on particle tracking in experiments have been transformed radically: it was mandatory for HEP experiments to build a inner vertex detector. In 1991, the more demanding environments led to the development of hybrid pixel detectors: a dedicated collaboration, RD19, was established at CERN with the specific goal to define a semiconductor micropattern detector with an incorporated signal processing at a microscopic level. In those years a wide set of prototypes of hybrid pixel has been manufactured; among the greatest productions a mention goes to the huge ATLAS and CMS vertex detectors. From the middle of 2013 a second collaboration, RD 53, has been established with the new goal to find a pixel detector suitable for phase II future upgrades of those experiments. Even if the collaboration is specifically focused on design of hybrid pixel readout chips (aiming to 65 nm tecnique so that the electronics fits within the pixel area), also other options have been taken in account and many test have been done on MAPS for example. Requirements imposed by HL-LHC will become tigher in time: for example, a dose and radiation of 5 Mrad and 1016NIEL are exepcted after 5 years of operation. Time resolution, material budget and power consumption are also issues for the upgrade: a time resolution better than 25 ns for a bunch crossing frequency of 40 MHz, a material budget lower than 2% and a power consuption lower than $500\,\mathrm{mW/cm^2}$ are required.

Amidst the solutions proposed 3D silicon detector, invented by Sherwood Parker in 1995, and MAPS are the most promising. In 3D sensors the electrode is a narrow column of n-type implanted vertically across the bulk instead of being implanted on the wafer's surface. The charge produced by the impinging particle is then drifted transversally within the pixel, and, as the mean path between two electrode can be soufficent low, the trap probability is not an issue. 3D pixels have been already proved in ATLAS tracker qualcosa? tipo anno e rif a caso. Even if 3D detector are adequately radiation hard, MAPS architecture looked very promising from the beginning: they overcome both the CCDs long reading time and the hybrid problems (I have already explained in section ?? the benefits of MAPS). Experiments such as ALICE at LHC and STAR at RHIC have already introduced the CMOS MAPS technology in their detectors. ALICE Tracking System (ITS2), upgraded during the LHC long shut down in 2019-20, was the first large-area ($\sim$10 m^2 covered by 2.5 Gpixels) silicon vertex detector based on CMOS MAPS.

3.1.1 Hybrid pixels at LHC: ATLAS, CMS and LHC-b

ATLAS

ATLAS is one of two general-purpose detectors at the LHC and has the largest volume detector ever constructed for a particle collider (46 m long and 25 m in diameter). The inner detector consists of three different systems all immersed in a magnetic field parallel to the beam axis. The main components of the Inner Detector are: the pixel, the micro-strips and transition radiation trackers. 92 million pixels are divided in 4 barrel layers and 3 disks in each end-cap region, covering a total area of 1.9 m^2 and having a 15 kW of power consumption.

As stated by the ATLAS collaboration the pixel detector is exposed by an extreme particle flux: "By the end of Run 3[1], the number of particles that will have hit the innermost pixel layers will be comparable to the number it would receive if it were placed only a few kilometres from the Sun during a solar flare". And the particle density will increase even more with HL-LHC. The most ambitious goal is employ a MAPS-based detector for the inner-layer barrels, and for this reason the RD53 collaboration is performing many test on MAPS prototypes, as Monopix of which I will talk about in section ??.

Up to now this possibility will be eventualy implemented during the second phase of the HL-LHC era, as at the start of high-luminosity operation the selected option is the hybrid one. The sensor, which is not selected yet, will be bonded with ITkPix,the first full-scale 65 nm hybrid pixel-readout chip and is developed by the RD53 collaboration. For the sensor a valueable option is using 3D pixels, which have already proved themselves in ATLAS, for the insertable B layer (IBL).qualcosa sui 3d usati a ATLAS. These pixel tracking systems will increase the number of pixels of a factor about 7, passing from 92 milioni to 6 miliardi.

CMS

LHCb

LHCb is a dedicated heavy-flavour physics experiment by exploiting pp interactions at 14 TeV at LHC. It was the last experiment to upgrade the vertex detector Vertex Locator (VELO) replacing the silicon-strip with pixels in May 2022. As the instantaneous luminosity in Run3 is increased by a factor $\lesssim$10, much of the readout electronics and of the trigger system have been developed in order to cope with the large interaction rate. To place the detector as close as possible to the beampipe and reach a better track reconstruction resolution, the VELO can be moved: during the injection of LHC protons it is parket at 3 cm from the beams and only when the stability is reach it is brought at $\sim$5 mm. Radiation hardness as well as readout speed are then a priority for the detectors: that's why the collaboration opted for a hybrid system. The Velopix is made bonding sensors, each measuring 55 $\times$ 55 micrometers, 200 µm-thick to a 200 µm-thick ASIC specially developed for LHCb and coming from the Medipix family (sec. ??). It is capable of handling up to 900 million hits per second per chip, while withstanding the intense radiation environment. The new VeloPix readout chips have a readout speeds of up to 20 Gb/s each, resulting in 3 Tb/s torrent of data Since the detector is operated under vacuum near the beam pipe, the heat removal is particularly difficult and evaporative CO2 microchannel cooling are used.

3.1.2 A DEPFET example: Belle-II

3.1.3 CMOS MAPS: ALICE and STAR

ALICE

ALICE (A Large Ion Collider Experiment) is a detector dedicated to heavy-ion physics and to the study of the condensed phase of the chromodynamics at the LHC. The tracking detector consists of the Inner Tracking System (ITS), the gaseous Time Projection Chamber (TPC) and the Transition

[1] Run 3 start in June 2022

Radiation Detector (TRD) and those are embedded in a magnetic field of 0.5 T. The ITS is made by six layers of detectors, two for each type, from the interaction point outwards: Silicon Pixel Detector (SPD), Silicon Drift Detector (SDD) and Silicon Strip Detector (SSD). Contrary to the others LHC experiments, ALICE tracker in placed in a quite different environments: the expected dose is smaller by two order of magnitude and the rate of interactions is few MHz instead of 40 MHz, but the number of particles comes out of each interaction is higher (the SPS is invested by a density of particles of $\sim$100 1/cm^2). The reconstruction of very complicated events whit a large number of particle is a challenge, hence to segment and to minimize the amount of material, which may cause secondary interaction complicating futher the event topology, is considered a viable strategy.

Upgrade con Monopix1 Thanks to the reduction of the material budget, ITS2, which uses the ALPIDE chip developed by ALICE collaboration, obtained an amazing improvement both in the position measurement and in the momentum resolution, improving the efficiency pf track reconstruction for particle with very low transverse momentum (by a factor 6 at pT $\sim$ 0.1 GeV/c). Further advancements in CMOS MAPS technology are being aggressively pursued for the ALICE ITS3 vertex detector upgrades (foreseen around 2026-27), with the goals of further reducing the sensor thickness and improving the readout speed of the devices, while keeping power consumption at a minimum. **STAR**

MIMOSA-28 devices for the first MAPS-based vertex detector: a 356 Mpixel two-layer barrel system for the STAR experiment at Brookhaven's Relativistic Heavy Ion Collide

3.2 Application in imaging

The counting mode represents the principal imaging tecnique used by detectors for that pourpose: l'obiettivo è tipicamente o single particle count, For photons imaging other materials with higher atomic charge than silicon could be prefered, as a high photon absorption efficiency is needed.

Two noteworthy example of microchips originally meant for detectors in particle physics at the LHC, and later employed in other fields are Medipix and Timepix. They are read-out chips developed by the Medipix Collaborations since early 1990s. Different chips with different imaging process have been produced in the last two decades, up to 2016 when Medipix4 and Timepix4 appeared. The applications in scientific imaging vary from atrophysics and medical imaging to more exotic domains as studies of protein dynamics, art authentication and dosimetry.

3.2.1 Medipix

The aim of the Medipix4 collaboration is designing pixel read-out chips that for the first time are fully prepared for TSV processing and may be tiled on all four sides. Two new chips are foreseen:Medipix4, which will target spectroscopic X-ray imaging at rates compatible with medical CT scans, and Timepix4, which will provide particle identification and tracking with higher spatial and timing precision.

Utilizzi in medicina: Radiography and computed tomography (CT) use X-ray photons to study the human body. The Medipix chips that implement on-pixel single photon counting provide many advantages for use in these fields. The technology has been applied in X-ray CT, in prototype systems for digital mammography, in CT imagers for mammography and for beta- and gamma-autoradiography of biological samples. , the images are no longer black and white

-First 3D colour X-ray of a human using CERN technology

3.2.2 Timepix

Timepix is being exploited for radiation monitoring in NASA's orion rocket and at the International Space Station.

Timepix Detector for Imaging in Ion Beam Radiotherapy

NASA, CERN Timepix Technology Advances Miniaturized Radiation Detection

As an example, the chips are now being used in the ATLAS experiment to provide independent, real-time information about the radiation environment in the experimental cavern - in principle the same task that Timepix does at the International Space Station. A direct 'spin back' to high-energy physics is VELOpix, based on Timepix3. Benefiting from developments made in the framework of the Medipix3 consortium which were not originally intended for high-energy physics, VELOpix will serve as the read-out chip in the new vertex detector of tje LHCb experiment, which is planned for installation in 2018.

Timepix and Timepix3 Optical Cameras

3.2.3 Applicability to FLASH radiotherapy

Chapter 4

TJ-Monopix1

TJ-Monopix1 is a small electrode DMAPS with fast R/O capability, fabricated by TowerJazz foundry in 180 nm CMOS imaging process. It is part, together with prototypes from other series such as TJ-MALTA, of the ongoing R&D efforts aimed at developing DMAPS in commercial CMOS processes, that could cope with the requirements at accelerator experiments. Both TJ-Monopix and TJ-MALTA series [5], produced with the same technology by TowerJazz (the timeline of the foundry products is shown in figure 4.1), are small electrode demonstrators and principally differ in the readout design: while Monopix implements a column-drain R/O, an asynchronous R/O without any distribution of BCID has been used by TJ-Malta in order to reduce power consumption.

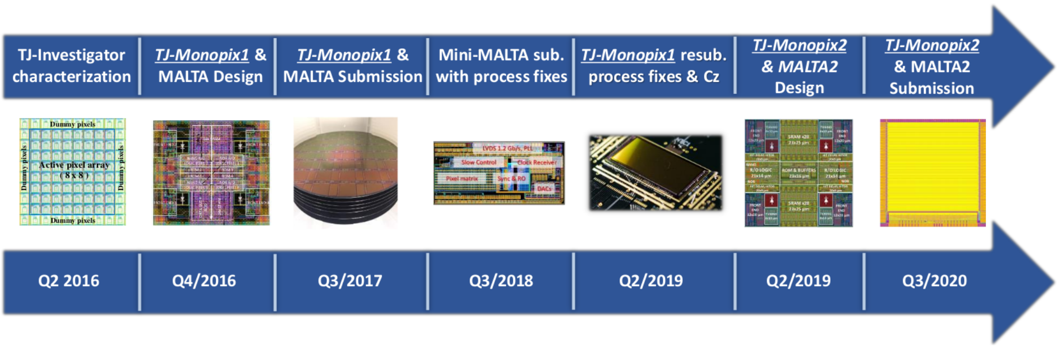


Figure 4.1: Timeline in TowerJazz productions in 180 nm CMOS imaging process

Another Monopix series, but in 150 nm CMOS technology, has been produced by LFoundry [6]. The main differences between the LF-Monopix1 and the TJ-Monopix1 (summarized in table 4.2), lay in the sensor rather than in the readout architecture, as both chips implements a fast column drain R/O with ToT capability [7][8]. Concerning the sensors, either are based on a p-type substrate, but with slightly different resistivities; in addition LFoundry pixels are larger, thicker and have a large fill factor (the very deep n-well covers $\sim$55% of the pixel area). The primary consequence is that LF-Monopix1 pixels have a higher capacity resulting in higher consumption and noise. As I discussed in section 2.4.1, the fact that LF-Monopix has a large fill factor electrode is expected to improve its radiation hardness. Indeed, a comparison of the perfomance of the two chips showed that TJ-Monopix suffers a comparatively larger degradation of efficiency after irradiation, due to the low electric field in the pixel corner; on the other hand, a drawback of the large fill factor in LF-Monopix is a significant cross-talk.

The TJ-Monopix1 chip contains, apart from the pixels matrix, all the required support blocks used for configuration and testing:

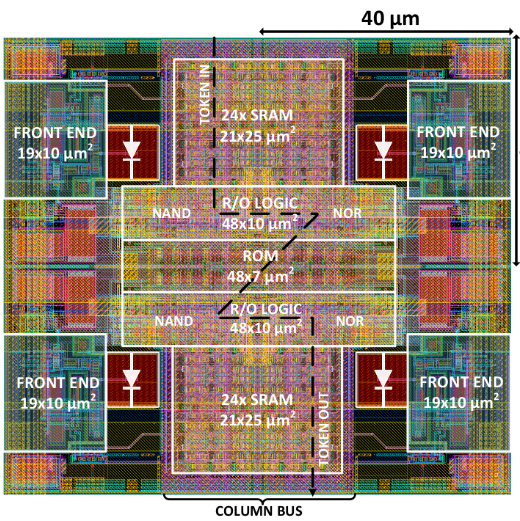
- the whole matrix contains 224 $\times$ 448 pixels, yielding a total active area approximately equal to 145 mm^2 over a total area of $1\times2\text{cm}^2$;
- at the chip periphery are placed some 7-bit Digital to Analog Converter (DAC), used to generate the analog bias voltage and current levels and to confiugure the FE;

	LF-Monopix1	TJ-Monopix1
Resistivity	>2 kΩcm	>1 kΩcm
Pixel size	50 × 250µm^2	36 × 40 µm^2
Depth	100-750 µm	25 µm
Capacity	∼ 400 fF	∼ 3 fF
Preamplifier	charge	voltage
Threshold trimming	on pixel (4-bit DAC)	global threshold
ToT	8 bits	6 bits
Consumption	∼ 300mW/cm^2	∼ 120mW/cm^2
Threshold	1500 e^-	∼ 270 e^-
ENC	100 e^-	∼ 30 e^-

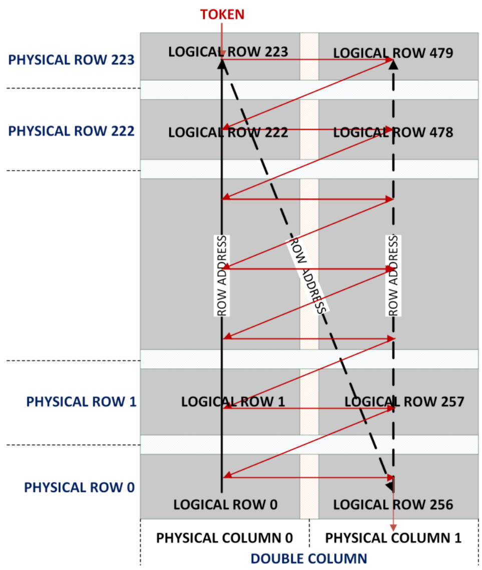
Table 4.1: Main characteristics of Monopix1 produced by TowerJazz and LFoundry [7][8]

- at the EoC is placed a serializer to transferred datas immediately, indeed no trigger memory is implemented in this prototypes;
- the matrix power pads are distributed at the sides
- four pixels which have analog output and which can be monitored with an oscilloscope, and therefore used for testing

Pixels are grouped in 2×2 cores (fig. 4.2a): this layout allows to separate the analog and the digital electronics area in order to reduce the possibile interference between the two parts. In addition it semplifies the routing of data as pixels on double column share the same column-bus to EoC. Therefore pixels can be addressed through the physical column/row or through the logical column/row, as shown in fig. 4.2b: in figure is also highlighted the token propagaion path, whose I will discuss later.



(a) Layout of a core containing 4 pixels. The analog FE and the digital part are separated in order to reduce cross-talk be



(b)

4.1 The sensor

As already anticipated, TJ-Monopix1 has a p-type epitaxial layer and a n doped small collection electrode (2 µm in diameter); to avoid the n-wells housing the PMOS transistors competing for the charge collection, a deep p-well substrate, common to all the pixel FE area, is used. TJ-Monopix1 adopts the modification described in section 2.4.2 that allows to achieve a planar depletion region

Parameter	Value
Matrix size	$1 \times 2\mathrm{cm}^2$
Pixel size	$36 \times 40\ \mu\mathrm{m}^2$
Depth	$25\,\mu\mathrm{m}$
Electrode size	$2\,\mu\mathrm{m}$
BCID	40 MHz
ToT-bit	6
Power consumption	$\sim 120\ \mathrm{mW/cm}^2$

Table 4.2

near the electrode applying a relatively small reverse bias voltage. This modification improves the efficiency of the detector, especially after irradiation, however a simulation of the electric field in the sensor, made with the software TCAD (Technology Computer Aided Design), shows that a nonuniform field is still produced in the lateral regions of the pixel compromising the efficiency at the corner. Two variations to the process have been proposed in order to further enhance the transversal component of electric field at the pixel borders: on a sample of chip, which includes the one in Pisa, a portion of low dose implant has been removed, creating a step discontinuity in the deep p-well corner (fig. 4.3); the second solution proposed[**MOUSTAKAS THESYS**, **PAG 58**] consists in adding an extra deep p-well near the pixel edge. A side effect of the alteration in the low dose implant is that the separation between the deep p-well and the p-substrate becomes weak to the point that they cannot be biased separately to prevent the punchthrough.

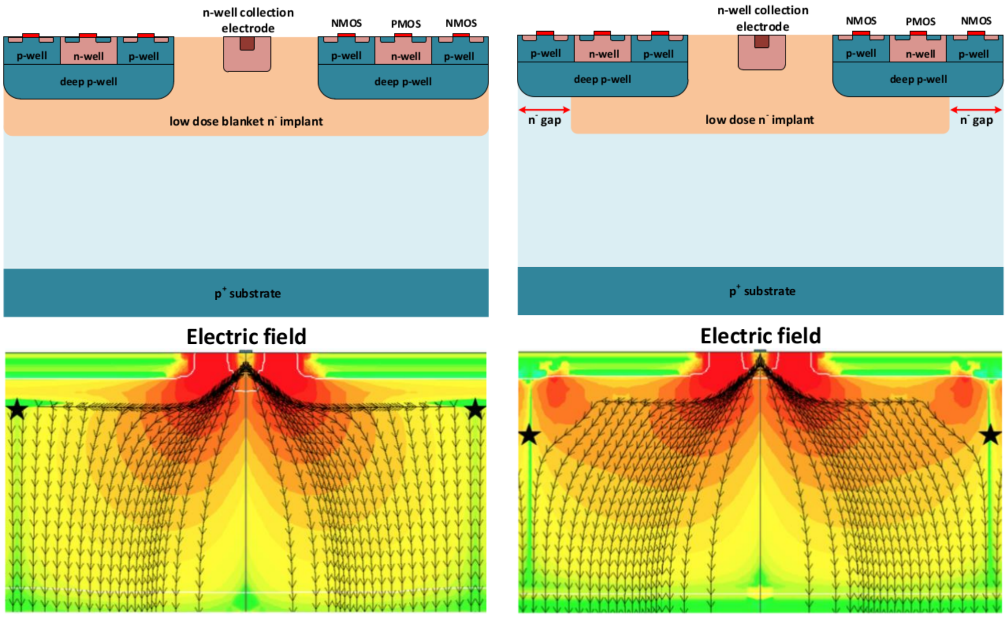


Figure 4.3: (a) The cross-section of a monolithic pixel in the TJ-Monopix with modified process; additionally in (b) a gap in the low dose implant is created to improve the collection of charge due to a bigger lateral component of the electric field. this point in figure is indicated by a star . transversal component of the electric field drops at the pixel corner

Moreover, to investigate the charge collection properties, pixels within the matrix are split between bottom top half and bottom half and feature a variation in the coverage of the deep p-well: the electronics area can be fully covered or not. In particular the pixels belonging to rows from 0 to 111 are fully covered (FDPW) and pixels belonging to rows from 112 to 223 have a reduced p-well (RDPW), resulting in a enhancement of the lateral component of the electric field.

4.2 Front end

The matrix is split in four sections, each one corresponding to a different flavor of the FE. The four variation have been implemented in order to test the data-bus readout circuits and the input reset modes.

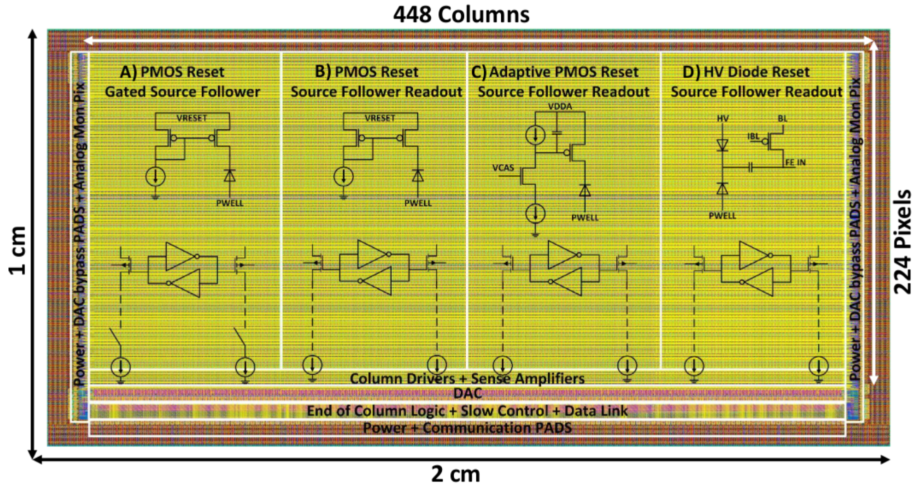


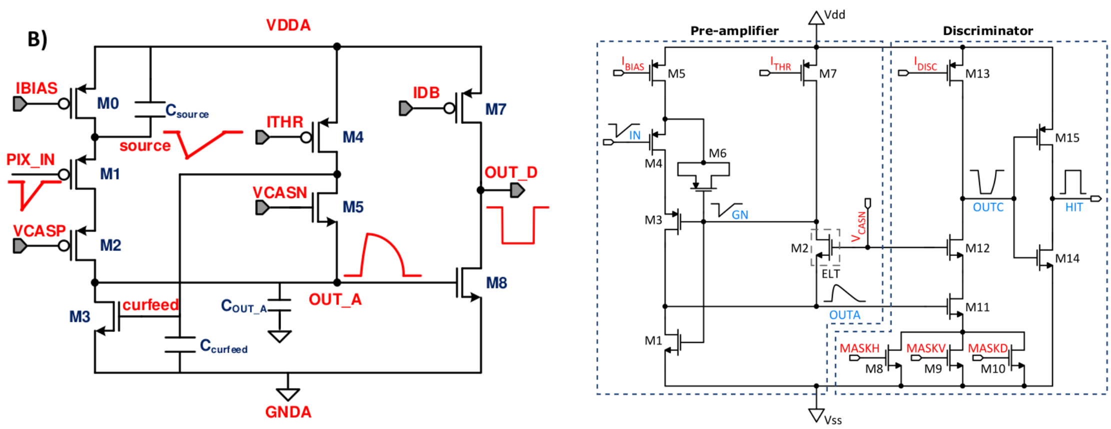
Figure 4.4

All the flavors implement a source-follower double-column bus readout: the standard variation is the flavor B, that features a PMOS input reset (refered as "PMOS reset"). Flavor A is identical to flavor B except for the realization of the source follower (it is a gated one) that aim to reduce the power consumption.cosa significa? C instead implements a novel leakage compensation circuit. Moreover the collection electrode in flavors A, B, C is DC-coupled to the front-end input, while in D is AC-coupled, providing to applu a high bias voltage; for this reason flavor D il called "HV flavor".

R resistenza di reset deve essere abbastanza grande in modo da far si che il ritorno allo zero è abbastanza lento (non devi "interferire" con la tot slope e non devi più corto del tempo del preamplificatore, sennò hai perdita di segnale). Baseline reset: all'input solitamente hai un PMOSS o un diodo; R reset; Voltage amplifier

4.2.1 ALPIDE-like

ALPIDE chips, developed by the ALICE collaboration, implemented a standard FE to the point that many CMOS MAPS detectors used a similar FE and are called "ALIPDE-like". Concidering that both TJ-Monopix1 and ARCADIA-MD1 have an ALPIDE-like FE, I am going to explain the broad principles of the early FE stage. The general idea is of the amplification to transfer the



(a) ALPIDE-like

(b)

charge from a bigger capacity[9], C_{source}, to a smaller one, C_{out}: the input transistor M1 with current source IBIAS acts as a source follower and this forces the source of M1 to be equal to the gate input $\Delta V_{PIX\_IN} = Q_{IN}/C_{IN}$.

$$Q_{source} = C_{source}\Delta V_{PIX\_IN} \tag{4.1}$$

The current in M2 and the charge accumulates on C_{out} is fixed by the one on C_{source}:

$$\Delta V_{OUT\_A} = \frac{Q_{source}}{C_{OUT\_A}} = \frac{C_{source}\Delta V_{PIX\_IN}}{C_{OUT\_A}} = \frac{C_{Source}}{C_{OUT\_A}}\frac{Q_{IN}}{C_{IN}} \tag{4.2}$$

A second branch (M4, M5) is used to generate a low frequency feedback, where VCASN and ITHR set the baseline value of the signal on C_{OUT_A} and the velocity to goes down to the baseline.
IL RUOLO DI CURVFEED NON L'HO CAPITO.
Finally IDB defines the charge threshold with which the signal OUT_A must be compared: depending on if the signal is higher than the threshold or not, the OUT_D is high or low respectively.

The actual circuit implemented in TJ-Monopix1 is shown in figure 4.5b: the principal difference lays in the addition of disableing pixels' readout. This possibility is uttermost important in order to reduce the hit rate and to avoid saturating the bandwidth due to the noisy pixels, which typically are those with manufacturing defects. In the circuit transistors M8, M9 and M10 have the function of disabling registers with coordinates MASKH, MASKV and MASKD (respectivelly vertical, orizontal and diagonal) from readout: if all three transistors-signals are low, the pixel's discriminator is disabled. Compared with a configurable masking register which would allow disableing pixels individually, to use a triple redundancy reduces the sensistivity to SEU[1] but also gives amount of intentionally masked ("ghost") pixels. This approach is suitable only for extremely small number N of pixel has to be masked: if two coordinate projection scheme had been implemented, the number of ghost pixels would have scale with N^2, if instead three coordinates are used, the N's exponential is lower than 2 (fig. 4.6)

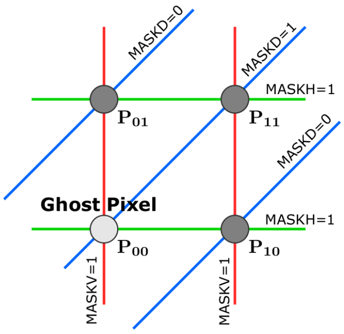


Figure 4.6

4.3 Readout logic

The simplest readout is "rolling shutter", in which peripheral logic along the chip edge addresses rows in turn, and analogue signals are transmitted by column lines to peripheral logic at the bottom of the imaging area. TJ-Monopix1 has a triggerless, fast and with ToT capability R/O which is based on a column-drain architecture. On the pixel are located two Random Access Memory (RAM) cells to store the 6-bit LE and 6-bit TE of the pulse, and a Read-Only Memory (ROM) containing the 9-bit pixel address. Excluded these memories, TJ-Monopix1 hasn't any other buffer: if a hit arrives while the pixel is already storing a previous one, the new data get lost.

[1] Single Event Upset, in sostanza è quando un bit ti cambia valore (da 0 a 1 o viceversa) perché una particella deposita carica nell'elettronica che fa da memoria registro/RAM/.... Questo tipo di elettronica ha bisogno di un sacco di carica prima che il bit si "flippi" (cambi valore), infatti tipicamente per avere un SEU non basta una MIP che attraversa esattemente quel pezzo di chip in cui è implementata la memoria, ma un adrone che faccia interazione nucleare producendo più carica di quanto farebbe una MIP. Questo metodo pur essendo più comodo richieda less amount of area ha però come drawback che il registro può essere soggetto a SEU problema non trascurabile in acceleratori come HL-LHC adronici

Parameter	Meaning	
IBIAS		
IDB	sets the discriminator threshold	yes
ITHR	sets the velocity of the return to the baseline	yes
ICASN	sets the baseline of the signal	yes
VRESET	sets the gain of the preamplifier	yes
IRESET	sets the gain of the preamplifier	no

Table 4.3: FE parameters which must be setted through the DAQ. "Function" means that higher parameter implies higher value

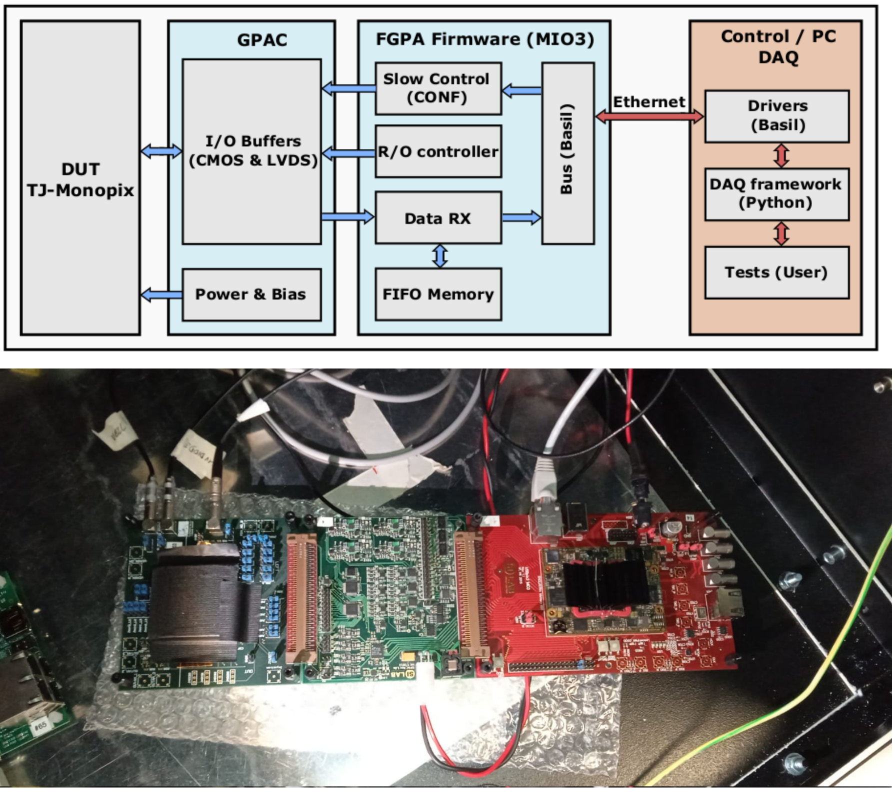
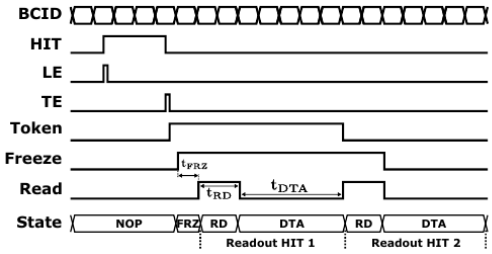


Figure 4.7: Main caption

After being read, the data packet is sent to the EoC periphery of the matrix, where a serializer transfers it off-chip to an FPGA (4.7). There a FIFO is used to temporarily stored the data, which is transmitted to a computer through an ethernet cable in a later time.

The access to the pixels' memory and the transmission of the data to the EoC, following a priority chain, is managed by control signals and is based on a Finite State Machine (FSM) composed by four state: no-operation (NOP), freeze (FRZ), read (RD) and data transfer (DTA). The readout sequence (**??**) starts with the TE of a pulse: the pixel immediately tries to grab the column-bus turning up a hit flag signal called *token*. The token is used to control the priority chain and propagates across the column indicating what pixel that must be read. To start the readout and avoid that the arrival of new hits disrupt the priority logic, a *freeze* signal is activated, and then a *read* signal controls the readout and the access to memory. During the freeze, the state of the token for all pixels on the matrix remains settled: this does not forbid new hits on other pixels from being recorded, but forbids pixels hit from turning on the token until the freeze is ended. The freeze stays on until the token covers the whole priority chain and gets the EoC: during that time new token cannot be turned on, and all hits arrived during a freeze will turn on their token at the end of the previous freeze. Since the start of the token is used to assign a timestamp to the hit, the token time has a direct impact on the time resolution measurement; this could be a problem coping with high hits rate.



(b) Readout sequence timing diagram. In this example two hits are being processed.

Figure 4.8: Readout timing diagram: in this example two hits are being processed

The analog FE circuit and the pixel control logic are connected by an edge detector which is used to determine the LE and the TE of the hit pulse(fig. 4.9): when the TE is stored in the first latch the edge detector is disabled and, if the FREEZE signal is not set yet, the readout starts. At this point the HIT flag is set in a second latch and a token signal is produced and depending on the value of Token in the pixel can be read or must wait until the Token in is off. In figure an OR is used to manage the token propagation, but since a native OR logic port cannot be implemented with CMOS logic, a sum of a NOR and of an inverter is actually used; this construct significantly increases the propagation delay (the timing dispersion along a column of 0.1-0.2 ns) of the token and to speed up the circuit optimized solution are often implemented. When the pixel become the next to be read in the queue, and at the rising edge of the READ signal, the state of the pixel is stored in a D-latch and the pixel is allowed to use the data bus; the TE and the HIT flag latches are reset and a READINT signal that enable access of the RAM and ROM cells is produced.

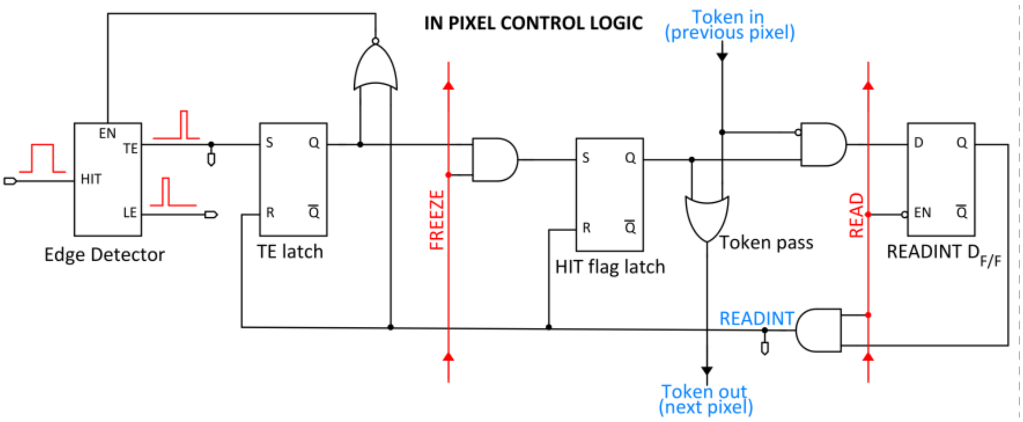


Figure 4.9

The final data must provide all the hits' information: the pixel address, the ToT and the timestamp. All those parts are assigned and appended at different time during the R/O chain:

- **Pixel address:** while the double column address (6-bit) is appended by the EoC circuit, the row address (8-bits for each flavor) and the physical column in the doublet (1-bit) are assigned by the in-pixel logic
- **ToT:** is obtained offline from the difference of 6-bits TE and 6-bits LE, stored by the edge detector in-pixel; since a 40 MHz BCID is distributed across the matrix, the ToT value is range 0-64 clock cycle which corresponds to 0-1.6 µs
- **Timestamp:** The timestamp of the hit correspond to the time when the pixel set up the token; it is assigned by the FPGA, that uses the LE, TE and a 640 MHz clock to derive it. For all those hits which arrived while the matrix is frozen, the timestamp is no more correlated with the time of arrival of the particle

Parameter	Value [DAC]	Value [µs]
START_FREEZE	64	1.6
STOP_FREEZE	100	2.5
START_READ	66	1.65
STOP_READ	68	1.7

Table 4.4: Default configuation of the R/O parameters

When the bits are joined up together the complete hit data packet is 27-bit.

4.3.1 Dead time measurements

The hit loss is due to analog and digital pile up: the first one occurs when a new hit arrives during the pre-amplifier response, the second instead, which is the more relevant contribution, while the information of the previous hit has not yet been transferred to the periphery. As only one hit at a time can be stored on the pixel's RAM, until the data have completed the path to get out, the pixel is paralyzed and the dead time τ almost corresponds with the time needed to trasmit the data-packets off-chip. Since the exportation of data from pixel to the EoC occurs via a 21-bits data bus, only one clock cycle is need to transfer the data to the end of column and the dead time bottleneck is given by the bandwidth of the serializer at the EoC. In our setup it operates at 40 MHz, thus to transmit a data packet (27-bit) at least 675 ns are needed. For what we have said so far, the R/O is completely sequential and therefore is expected a linear dependence of the reading time on the number of pixels to read:

$$\tau = 25\,\text{ns} \times (\alpha\, N + \beta) \tag{4.3}$$

where α and β are parameters dependent on the readout chain setting.

To measure and test the linearity of the reading time with the number of pixels firing, I have used the injection mode available on the chip. Indeed, the injection mode allows fixing not only the amplitude of the pulse, which corresponds to the charge in DAC units, but also the period and the width. I have injected a fix number of pulses (100) and looked for the rate when the efficiency decreases. Moreover to test that there is no dependece of the digital readout time from the charge of the pulse, I have try to change the amplitude of the pulse injected, but the parameters found were consistent with the default configuration ones.

Un esempio se leggo un singolo pixel: LA SLOPE CON CUI L'EFFICIENZA SCENDE È ABBSTANZA UNIFORME? perchè satura a 50?

While the single pixel reading time and the dead time do not depend on the position on the pixel matrix and are equal to 106 (46+60) clock counts within 1 clock count, on the other hand the τ depends on the pixel position on the matrix when more than one pixel are firing. In particular the priority chain goes from row 224 to row 0, and from col 0 to 112, that means the last pixels to be read is the one on le bottom right corner of the matrix.

In figure 4.11 is reported the reading time versus the number of pixels injected; the R/O parameters that control the reading time and their default values are reported on table **??**.

The factor α, referring to eq. 4.3 is proportional to the difference (STOP_FREEZE - START_READ), while the offset β lies between 5 and 15 clock counts. Since through the injection a random hit rate on the matrix can't be simulated, as the coordinates of the pixels to inject must be specified, for convenience I used the pixels on the same column/row. No difference in the α and β coefficients has been observed between the two case.

A tutte le hit di una iniezione che arrivano contemporaneamente viene assegnato lo stesso timestamp; Risoluzione temporale??

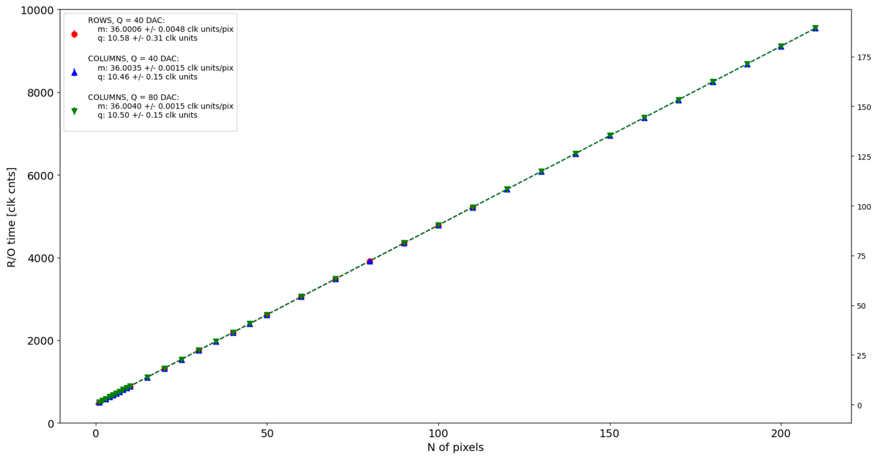


Figure 4.10

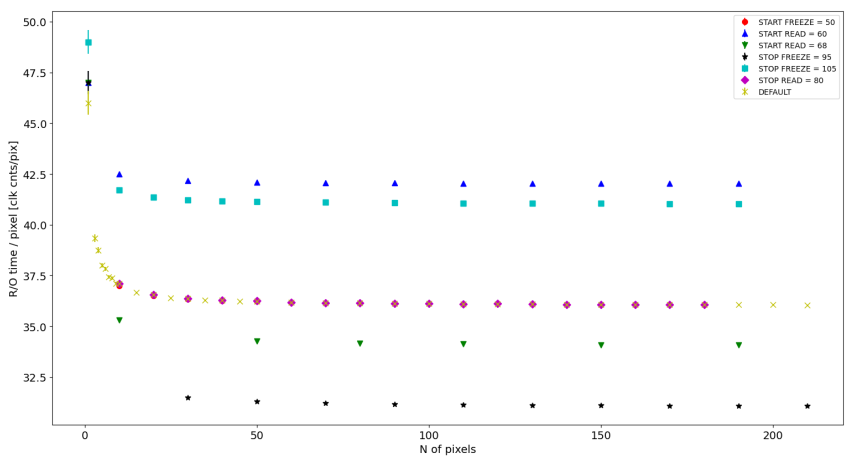


Figure 4.11

Chapter 5

Arcadia-MD1

[10] [11]

Breve introduzione analoga a Monopix1 in cui descrivo brevemente la ”timeline” da SEED Matisse a Md1 e Md2

5.1 The sensor

ARCADIA-MD1 is an LFoundry chip which implements the technology 110 nm CMOSS node with six metal layer **??**. The standard p-type substrate was replaced with an n-type floating zone material, that is a tecnique to produce purified silicon crystal. (pag 299 K.W.).

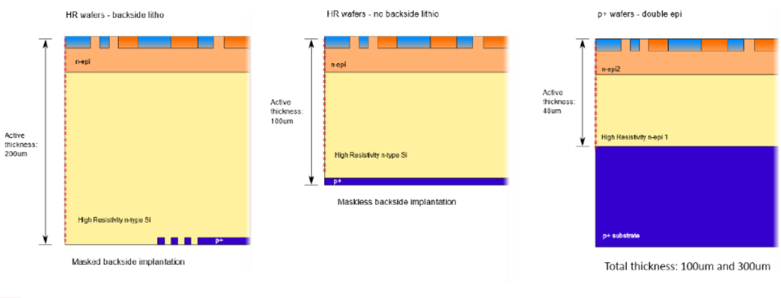


Figure 5.1

Wafer thinning and backside litography were necessary to introduce a junction at the bottom surface, used to bias the substrate to full depletion while maintaining a low voltage at the front side. C’è un deep pwell per - priority chainseparare l’elettronica dal sensore; per controllare il punchthought è stato aggiunto un n doped epitaxial layer having a resistivity lower than the substrate.

RILEGGI SUL KOLANOSKY COS’È IL PUNCHTHROUGHT, FLOAT ZONE MATERIAL, COME VENGONO FATTI I MAPS COME FAI LE GIUNZIONI

It is part of the cathegory of DMAPS Small electrode to enhance the signal to noise ratio.

It is operated in full depletion with fast charge collection by drift.

Prima SEED si occupa di studiare le prestazioni: oncept study with small-scale test structure (SEED), dopo arcadia: technology demonstration with large area sensors Small scale demo SEED(sensor with embedded electronic developement) Quanto spazio dato all’elettronica sopra il pwell e quanto al diodo. ..

5.2 Readout logic and data structure

5.2.1 Matrix division and data-packets

The matrix is divided into an internal physical and logical hierarchy: The 512 columns are divided in 16 section: each section has different voltage-bias + serializzatori. Each section is devided in cores () in modo che in ogni doppia colonna ci siano 1Pacchetto dei dati 6 cores. ricordati dei serializzaatori: sono 16 ma possono essere ridotti ad uno in modalità spazio

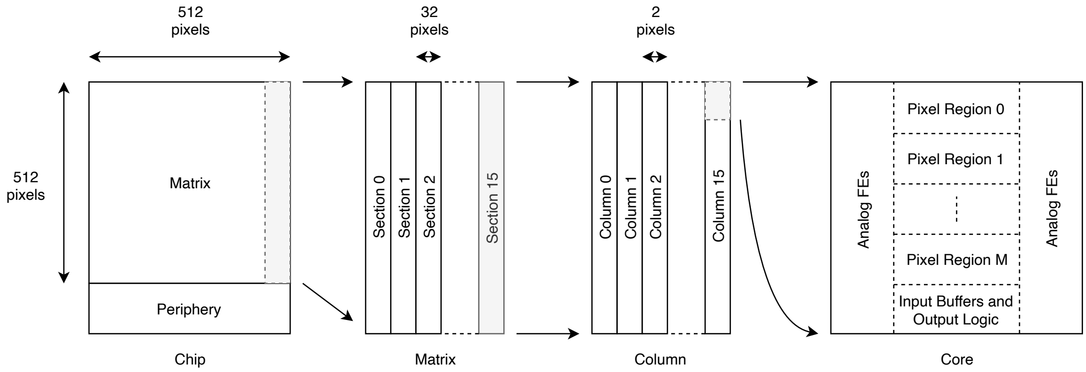


Figure 5.2

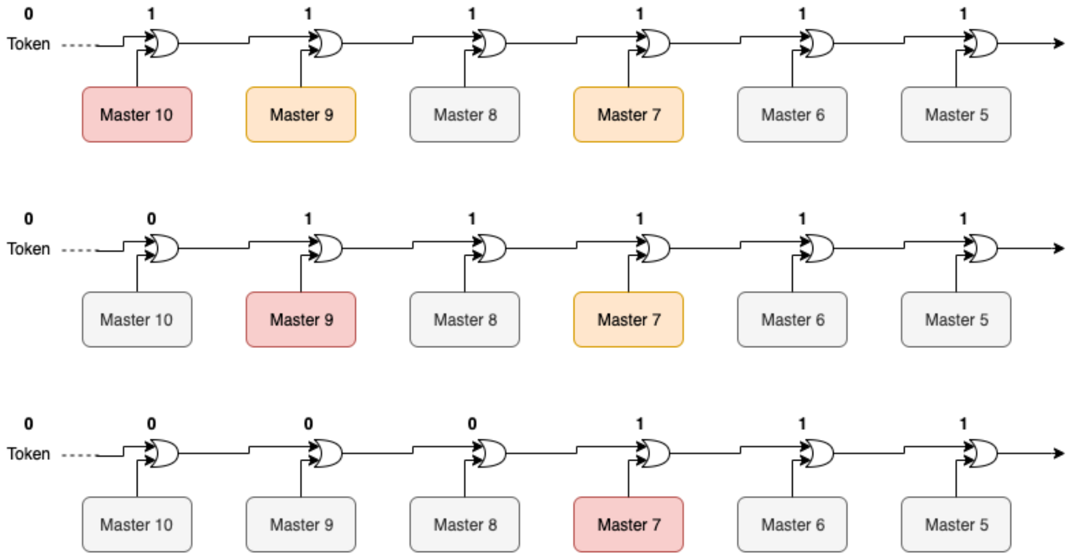


Figure 5.3

Questa divisione si rispecchia in come sono fatti i dati: scrivi da quanti bit un dato è fatto e le varie cordinate che ci si trovano dentro; devi dire che c'è un pixel hot e spieghi dopo a cosa serve, e devi accennare al timestamp

"A core is simply the smallest stepped and repeated instance of digital circuitry. A relatively large core allows one to take full advantage of digital sybthesis tools to implement complex functionality in the pixel matrix, sharing resources among many pixels as needed.". pagina 28 della review.

TABELLA: con la gerarchia del chip Matrix (512x512 pixels) Section (512x32 pixels) Column (512x2) Core (32x2) Region (4x2)

Nel chip trovi diverse padframe: cosa c'è nelle padframe e End of section.

"DC-balance avoids low frequencies by guaranteeing at least one transition every n bits; for example 8b10b encoding n =5"

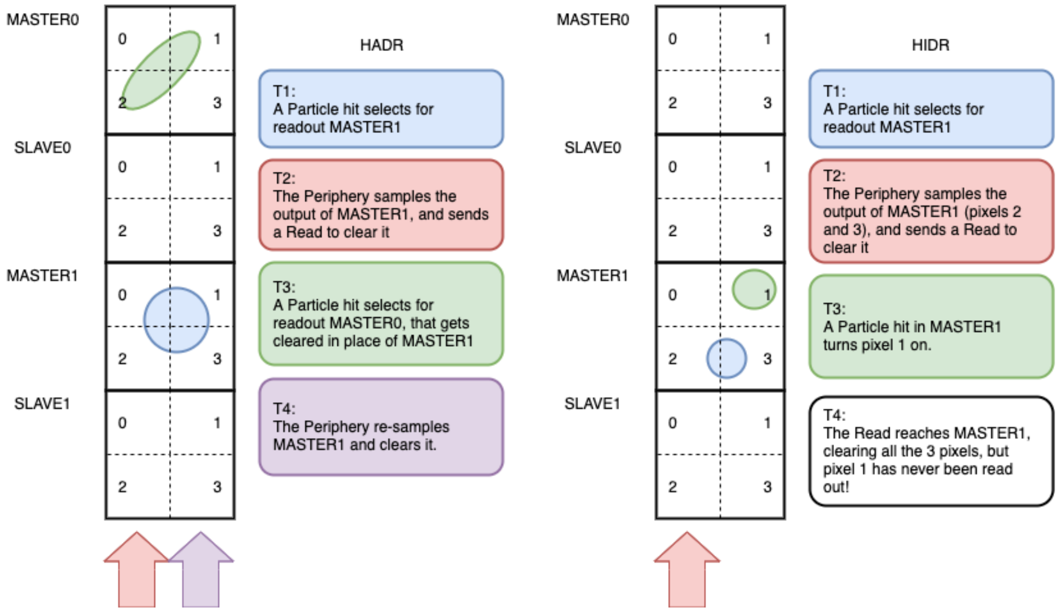


Figure 5.4

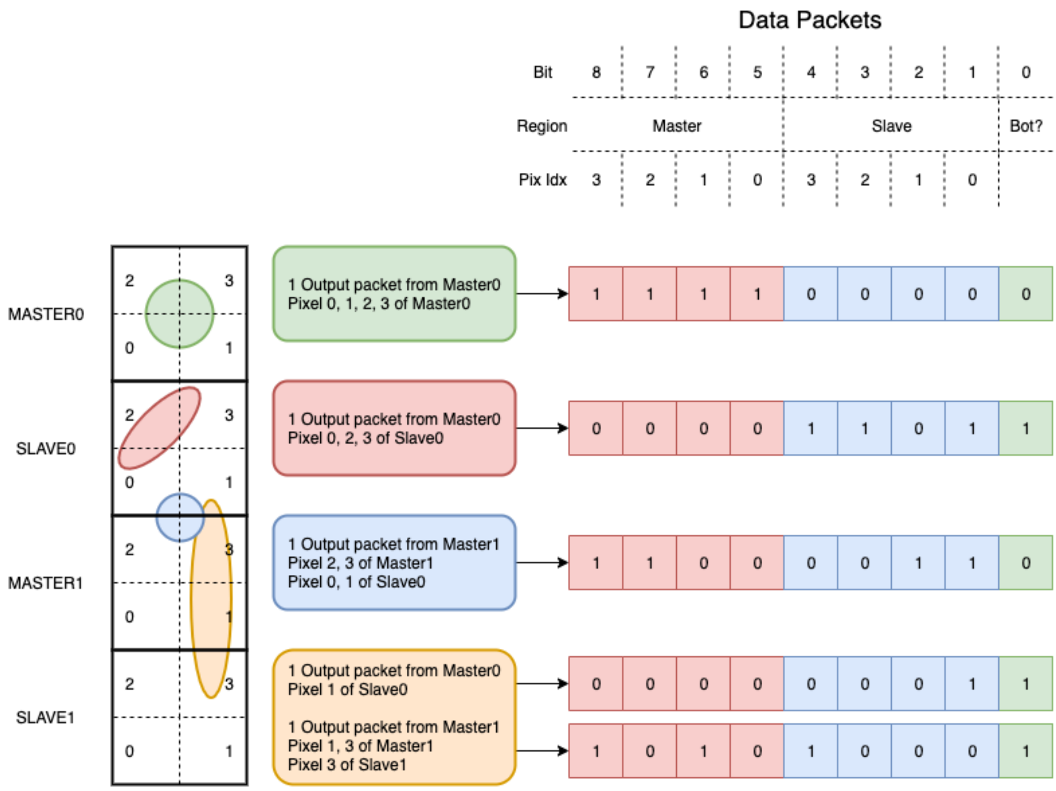
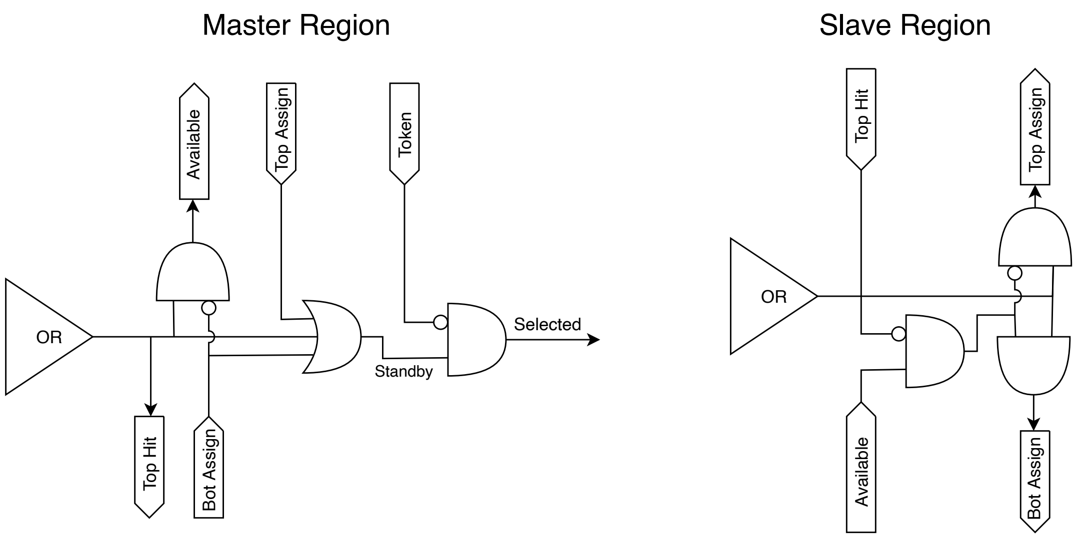


Figure 5.5



Master Region
Available
Top Assign
Token
OR
Selected
Standby
Top Hit
Bot Assign
Slave Region
Top Hit
Top Assign
OR
Available
Bot Assign

Figure 5.6

Chapter 6

Threshold and noise characterization

6.1 Threshold and noise: figure of merit for pixel detectors

The signal to threshold ratio is the figure of merit for pixel detectors.

la soglia deve essere abb alta da tagliare il rumore ma abb bassa da non perdere efficienza. Invece di prendere il rapporto segnale rumore prendi il rapporto segnale soglia. Perchè? la soglia è collegato al rumore, nel senso che: supponiamo di volere un occupancy di 10-4 allora sceglierò la soglia in base a questo. (plot su quaderno) Da questo conto trovo la minima soglia mettibile
In realtà quello che faccio è mettere una soglia un po' più grande perchè il rate di rumore dipende da molti fattori quali la temperatura, l anneling ecc, e non voglio che cambiando leggermente uno di questi parametri vedo alzarsi molto il rate di rumore. In realtà non è solo il rumore sensibile a diversi fattori, ma anche la soglia: ad esempio la cosa classica è la variabilità della soglia da pixel a pixel.
In questo modo rumore e soglia diventano parenti.
Review pag 26.

The noise requirement can be expressed as:

Questo implica tra le altre cose che voglio poter assegnare delle soglie diverse a diversi pixel: Drawback è dare spazio per registri e quantaltro.
Questo lascia però ancora aperto il problema temporale delle variazioni del rumore: problema per cui diventano necessarie le misure dei sensori dopo l'irraggiamento.

Per arcadia i registri (c'è un DAC) per la soglia (VCASN) si trovano in periferia. Non fare trimming sulla soglia è uno dei problemi che si sono sempre incontrati: a casusa dei mismatch dei transistor le soglie efficaci pixel per pixel cambiano tanto. La larghezza della s curve è il noise se assumi che il noise è gaussiano

Il trimming della soglia avviene con dei DAC: la dispersione della soglia dopo al tuning e dovuta al dac è:

$$\sigma_{THR,tuned} = \frac{\sigma_{THR}}{2^{nbit}} \tag{6.1}$$

dove il numero di bit cambia varia tra 3-7 tipicamente. Monopix è 7 Arcadia 6

Each ROIC is different in this respect, but in general the minimum stable threshold was around 2500 electrons (e) in 1st generation ROICs, whereas it will be around 500 e for the 3rd generation. This reduction has been deliberate: required by decreasing input signal values. Large pixels (2 104 um2), thick sensors (maggiore di200 um), and moderate sensor radiation damage for 1st generation detectors translated into expected signals of order 10 ke, while small pixels (0.25 104 um2), thinner sensors (100 um), and heavier sensor radiation damage will lead to signals as low as 2 ke atthe HL-LHC
The ENC can be directly calculated by the Cumulative Distribution Function (CDF) (scurve) obtained from the discriminator "hit" pulse response to multiple charge injections

6.2 TJ-Monopix1 characterization

6.3 ARCADIA-MD1 characterization

Chapter 7

Test beam measurements

Epitaxial layer thickness: più grande è e più carica viene depositata da una MIP, però devi fare attenzione alla forma della zona svuotata perchè può portare ad un aumento della charge sharing tra pixel vicini. Se il diodo è molto piccolo rischi che l'efficienza di collection è diminutia perchè l'intensità del campo elettrico è più bassa intorno al diodo, e hai più charge sharing.

Possibilità di integrare carica sul pixel: due elettroni consecutivi su un pixel ogni quanto arrivano? Fai il conto del tempo medio

Vogliamo sfruttare l'analog pile up, per fare questo dobbiamo fare attenzione a non finire nel digital pile up Devi avere che il tot dell'elettrone (cioè MIP) è maggiore del deltat medio; in questo caso potresti riuscire ad integrare carica.

Appendix A

Pixels detector: a brief overview

A.1 Radiation damages

Radiation hardness is a fundamental requirement for pixels detector especially in HEP since they are almost always installed near the interaction point where there is a high energy level of radiation. At LHC the ϕ_{eq} per year in the innermost pixel detector is $10^{14} n_{eq}/cm^2$; this number reduces by an order passing to the outer tracker layer [2] pag 341 Wermes. Here the high fluence of particles can cause a damage both in the substrate of the detector and in the superficial electronics.

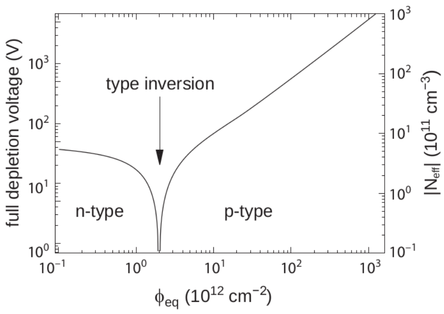
The first one has a principal non ionizing nature, due to a non ionizing energy loss (NIEL), but it is related with the dislocation of the lattice caused by the collision with nuclei; by this fact the NIEL hypothesis states that the substrate damage is normalized to the damage caused by 1 MeV neutrons. Differently, surface damages are principally due to ionizing energy loss.

DUE PAROLE IN PIÙ SUL SURFACE DAMAGE A charge accumulation in oxide (S_iO_2) can cause the generation of parasitic current with an obvious increase of the 1/f noise. Surface damages are mostly less relevant than the previous one, since with the development of microelectronics and with the miniaturization of components (in electronic industry 6-7 nm transistors are already used, while for MAPS the dimensions of components is around 180 nm) the quantity of oxide in circuit is reduced.

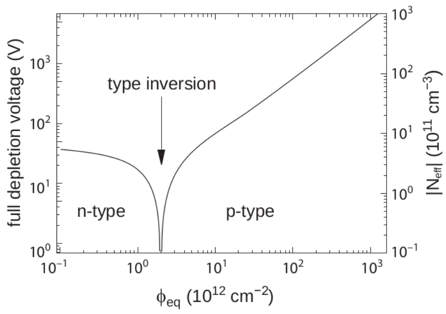
Let's spend instead two more other words on the more-relevant substrate damages: the general result of high radiation level is the creation of new energy levels within the silicon band gap and depending on their energy-location their effect can be different, as described in the Shockely-Read-Hall (SRH) statistical model. The three main consequence of radiation damages are the changing of the effect doping concentration, the leakage current and the increasing of trapping probability.

Changing of the effective doping concentration: is associated with the creation/removal of donors and acceptors center which trap respectively electrons/holes from the conduction band and cause a change in effective space charge density. Even an inversion (p-type becomes n-type[1]) can happen: indeed it is quite common at not too high fluences ($\phi_{eq} 10^{12-13} n_{eq} cm^{-2}$). A changing in the doping concentration requires an adjustment of the biasing of the sensor during its lifetime (eq.2.1) and sometimes can be difficult keeping to fully deplete the bulk.

[1] L'INVERSIONE OPPOSTA NON CE L'HAI PERCHÈ?



(a) 1a



(b) 1b

Leakage current: is associated with the generation-recombination centers. It has a strong dependence with the temperature ($I_{leak} \propto T^2$), whose solution is therefore to operate at lower temperature.

Increase of trapping probability: since the trapping probability is constant in the depleted region, the collected charge decreases exponentially with the drift path. The exponential coefficient, that is the mean trapping path, decreases after irradiation and typical values are 125-250 μm and must be compared with the thickness of the depleted region which () corresponds to the mean drift path.

Different choices for substrate resistivity, for junctions type and for detector design are typically made to fight radiation issues. Some material with high oxygen concentration (as crystal produced using Czochralki (Cz) or float-zone (Fz) process (CONTROLLA LA DIFFERENZA TRA I DUE)) for example, show a compensation effect for radiation damage; another example is the usage of n+ -in-p/n sensors (even if p+ -in-n sensors are easier and cheaper to obtain) to get advantage of inversion/to have not the inversion (since they are already p-type). After inversion the n+p boundary, coming from n+ in-n, but to keep using the sensor the depletion zone still must be placed near the diode.

Appendix B

FLASH radiotherapy

La radioterapia si usa nel 60 per cento dei pazienti, sia come cura che come trattamento palliativo. Si associa spesso ad altre cure e si può fare prima/durante/dopo un intervento.

Si può fare in modi diversi: da dentro (brachytherapy) oppure da fuori (quella standard). Un requisito importante è la delinazione del target (non vuoi rischiare di beccare i tessuti sani), per cui prima tipicamente si fanno esami di imaginig del tumore. TIpicamente anche gli acceleratori stessi per la terapia sono provvisti di radiografia.

Un problema dei fotoni ad esempio è che il loro rilascio di dose è lineare, per cui danneggi anche i tessuti sani. Il problema dei protoni invece è che hanno un picco troppo strtto per cui non puoi coprire grosse zone e sorpattuto se sbagli rischi davvero di danneggiare moooolto i tessuti sani.

B.1 Cell survival curves

Curva di efficacia del trattamento in funzione della dose:

$$\frac{S(D)}{S(0)} = e^{-F(D)} \tag{B.1}$$

dove F(D)

$$F(D) = \alpha D + \beta D^2 \tag{B.2}$$

dove α e β rappresentano due tipi di danno diversi: coefficients, experimentally determined, characterizing the radiation response of cells. In particularly, alpha represents the rate of cell killing by single ionizing events, while beta indicates the maximal rate of cell killing by double hits observed when the repair mechanisms do not activate during the radiation exposure. Si ottiene una curva di sopravvivenza dove si vede la possibilità delle cellule di autoripararsi. A basse dosi infatti le cellule possono ripararsi.

Per introdurre l'effetto FLASH instroduco prima la therapeutic window.

TCP è la tumor control Probability che indica la probabilità delle cellule del tumore di essere uccise dopo una certa dose (con in riferimento a dose in acqua)
Se una media di $\mu(D)$ di cellule di tumore are killed con una dose D, la probabilità che n cellule sopravvivono è data da $P(n|\mu)$ poisson:

$$P(n|\mu) = \frac{\mu(D)^n e^{-\mu(D)}}{n!} \tag{B.3}$$

$$TPC(D) = P(n=0|\mu(D)) = e^{-\mu(D)} \tag{B.4}$$

D'altra parte hai una probabilità di fare danno su normal tissue NTCP Normal Tissue Complication Probability, che rappresenta il problema principale e che limita la massima radiazione erogabile Una scelta bilanciata si applica guardando a questi due fattori; si usa il therapeutic index definito come TCP/NTCP.

La cosa ottimale è ampliare la finestra del therapeutic ratio.

CONV-RT 0.01-5 Gy/min. A typical RT regime today consists of daily franctions of 1.5 to 3 Gy given over several weeks.
Nell Intra operative radiation therapy (IORT), where they reach values respectively about 20 and 100 times greater than those of conventional radiation therapy.
FLASH vuole ultrahigh mean dose-rate (maggione di 40 Gy/s) in modo da ridurere anche il trattamento a meno di un secondo.

B.2 FLASH effect

Ci sono due effetti che affect the flsh effect and la sua applicabilità: Dose rate effect e oxygen

Cellule che esibiscono hypoxia (cioè cellule che non hanno ossigeno sono radioresistenti); al contrario normoxia e physoxia non lo sono. la presenza di ossigeno rende la curva steeper indicando che lo stesso danno si raggiunge a livelli di dose più bassi rispetto al caso senza ossigeno.
FIGURA con una curva a confronto con e senza ossigeno.
Typically, the OER is in the order of 2.5–3.5 for most cellular systems
Quindi si vogliono sfruttare questi effetti per diminuire la tossicità sui tessuti sani

Bibliography

[1] W. Snoeys et al. "A process modification for CMOS monolithic active pixel sensors for enhanced depletion, timing performance and radiation tolerance". In: (2017). DOI: https://doi.org/10.1016/j.nima.2017.07.046.

[2] H. Kolanoski and N.Wermes. *Particle Detectors: Fundamentals and Applications*. OXFORD University Press, 2020. ISBN: 9780198520115.

[3] E. Mandelli. "Digital Column Readout Architecture for the ATLAS Pixel 0.25 um Front End IC". In: (2002).

[4] M. Garcia-Sciveres and N. Wermes. "A review of advances in pixel detectors for experiments with high rate and radiation". In: (2018). DOI: https://doi.org/10.1088/1361-6633/aab064.

[5] M. Dyndal et al. "Mini-MALTA: Radiation hard pixel designs for small-electrode monolithic CMOS sensors for the High Luminosity LHC". In: (2019). DOI: https://doi.org/10.1088/1748-0221/15/02/p02005.

[6] M. Barbero. "Radiation hard DMAPS pixel sensors in 150 nm CMOS technology for operation at LHC". In: (2020). DOI: https://doi.org/10.1088/1748-0221/15/05/p05013.

[7] K. Moustakas et al. "CMOS Monolithic Pixel Sensors based on the Column-Drain Architecture for the HL-LHC Upgrade". In: (2018). DOI: https://doi.org/10.1016/j.nima.2018.09.100.

[8] I. Caicedo et al. "The Monopix chips: depleted monolithic active pixel sensors with a column-drain read-out architecture for the ATLAS Inner Tracker upgrade". In: (2019). DOI: https://doi.org/10.1088/1748-0221/14/06/C06006.

[9] D. Kim et al. "Front end optimization for the monolithic active pixel sensor of the ALICE Inner Tracking System upgrade". In: *JINST* (2016). DOI: doi:10.1088/1748-0221/11/02/C02042.

[10] L. Pancheri et al. "A 110 nm CMOS process for fully-depleted pixel sensors". In: (2019). DOI: https://doi.org/10.1088/1748-0221/14/06/c06016.

[11] L. Pancheri et al. "Fully Depleted MAPS in 110-nm CMOS Process With 100–300-um Active Substrate". In: (2020). DOI: 10.1109/TED.2020.2985639.