UNIVERSITA' DEGLI STUDI DI PAVIA

DIPARTIMENTO DI ELETTRONICA

DOTTORATO DI RICERCA IN MICROELETTRONICA

CMOS PIXEL SENSORS AND MIXED-SIGNAL READOUT ELECTRONICS IN A 3D INTEGRATION TECHNOLOGY

Tutor:
Prof. Valerio Re

Coordinatore del Dottorato:
prof. Rinaldo Castello

Tesi di Dottorato
di Luigi Gaioni

Anno Accademico 2008/2009

Ringraziamenti

Desidero ringraziare il prof. Valerio Re, mio tutor nel presente lavoro di ricerca, già mio relatore di tesi di laurea, per i suoi preziosi insegnamenti, per tutte le opportunità che mi ha offerto, per i consigli e per tutto il tempo a me dedicato nel corso di questi utlimi anni.
Ringrazio con sincero affetto la prof. Valeria Speziali per essermi sempre stata vicina, per tutti i suggerimenti e tutta l'attenzione che ha avuto nei miei confronti.
Sarò sempre grato a Massimo Manghisoni, Enrico Pozzati, Lodovico Ratti e Gianluca Traversi per i costanti consigli, la pazienza e la disponibilità: di loro sono debitore di molte delle conoscenze acquisite durante questi anni.
Doveroso, infine, ringraziare Fabio, Marco, Claudio, Alessia, Stefano, Emanuele e Valentina per la pazienza con la quale hanno sopportato Bach.

Contents

Introduction

The continued reduction of microelectronic device size over the past few decades has led to pioneering improvements in performance of electronic products. This miniaturization has been associated with unrelenting improvements in process technologies. Beside the evident benefits introduced by the scaling process, such as high level of integration density, there are also some other aspects which have to be considered. Smaller feature sizes imply lower supply voltages, sharply increasing mask costs, and larger gate leakage currents. Moreover, small feature size technologies also have to cope with increasing interconnect density and delays. Now, perhaps, the semiconductor industry is approaching a domain where physics makes it difficult to achieve smaller gate technologies. This scenario has led to consideration of alternative arrangements of electronics based on vertical interconnection of layers of circuitry. These technologies, usually referred to as “3D”, rely on layering tiers of active circuitry interconnected to each other. A vertically interconnected wafer, in addition to having increased overall circuit density, reduces the overall length of the device interconnections, increasing the speed by reducing resistance, inductance and parasitic capacitance. Power consumption is also decreased due to the reduced wire length and smaller capacitance. In addition, the layers may be fabricated in different technologies, each optimized for a particular application. A variety of integration technologies for 3D integrated circuits (IC) have been demonstrated. Silicon-on-Insulator (SOI) technology is particularly appealing in these processes.

This thesis work discusses the 3D integration of CMOS monolithic active pixel sensors (MAPS). MAPS sensors represent nowadays a promising alternative with respect to other mature technologies in different sectors. Initially conceived for imaging application, in competition to CCD devices, CMOS MAPS with simple readout schemes have gained a significant share of the consumer electronic, market where they are used in such commercial devices as videocameras or cameras for digital still photography. Vertical integration process capabilities make it possible to fabricate compact 3D MAPS detectors in which the different sections included in a traditional pixel (e.g. the sensing element, the analog front-end, digital processing blocks and so on) can be split into different layers, each optimized for the particular function accomplished by the layer itself, with significant benefits in terms of pixel size and functionality. Moreover, CMOS MAPS have been proposed in the last years as suitable candidates for charged particle trackers at the next generation colliders like International Linear Collider (ILC) and Super B-Factory as alternative tracking devices with respect to hybrid pixel detectors. MAPS devices, indeed, may comply with the severe constraints set by the future experiments at these colliders, which require highly granular and low mass detectors. CMOS MAPS are adequate in terms of material budget, since their sensing element shares the same substrate with the readout electronics; furthermore, substrate thickness can be reduced to a few tens of microns with no significant signal loss. The proposed 3D MAPS, object of this thesis works, involves the use of a deep n-well/p-substrate junction, provided by triple-well CMOS technologies, as collecting element. In this way, the sensor can be extended to cover a large area of the pixel cell. This solution, inherited from the so called deep n-well MAPS (DNW MAPS) fabricated in planar (2D) technologies, allow designers to realize more complex readout circuits, taking advantage of fully CMOS architectures: the effects of charge collection from PMOS n-wells, which acts as competitive electrodes against the main collecting electrode, may be significantly limited. Vertical integration processes, by stacking two or more layers one on the top of the other, make it possible to reduce interaction effects between different sections of the pixel cell and may also improve collection efficiency since PMOS wells can be placed in a different layer with respect to the sensor.

The first chapter introduces the 3D integration process, the benefits related to this technology and the main approaches involved in 3D IC fabrication. At the end of the chapter, a brief overview of the Tezzaron Semiconductor and MIT Lincoln Laboratory processes will be provided: these processes, indeed, are those employed in the design and fabrication of the devices subject of this thesis work. The second chapter provides the static, signal and noise characterization of devices fabricated in a 180 nm SOI technology suitable for 3D integration.

The third chapter discusses the features of a new kind of DNW MAPS, called SDR1 (Sparsified Data Readout), which exploits the capabilities of vertical integration processing in view of the design of high granularity detectors.

SDR1 inherits and extends the functional capabilities of DNW MAPS fabricated in planar CMOS technology and is expected to show better performance with respect to 2D versions.

Chapter 1

Three dimensional integrated circuits

Over the last years, CMOS scaling has made it possible to comply with the demand of the computer and information technology industry, that is, very large scale integrated (VLSI) circuits with increasing functionality and performance at minimum cost and power dissipation. In planar (2D) technologies, scaling is reducing gate delays but also increasing interconnect ones: interconnect density turns out to be increased in the design of microelectronic circuits realized with small feature sizes technologies. Increasing interconnect density affects the power consumption in high-performance 2D chips, where a significant fraction of the total chip power consumption can be due to the wiring network used for clock distribution, which is usually realized using long global wires. Another aspect which has to be taken into account is represented by the impact of the interconnect scaling on the traditional computer-aided-design (CAD) methodologies and tools, which are causing the design cycles to increase, thus increasing the time-to-market and the cost per chip function.
Furthermore, increasing drive for the integration of disparate signals (digitals, analog, RF) and technologies (SOI, SiGe HBTs, GaAs, and so on) is introducing various system-on-a-chip (SoC) design concepts, for which existing planar (2D) technologies may not be suitable. In order to overcome critical issues related to 2D designs, dozens of companies and organizations are working on 3D integrated circuits in the United States, Europe and Asia: companies such as IBM, Intel, Philips, STM, Tezzaron Semiconductor and Samsung, and organizations such as MIT Lincoln Labs and Fraunhofer Institute IZM-Munich are very active in 3D. Each of them has its own approach to working on the key technologies needed for 3D. A review of the options for the 3D integration

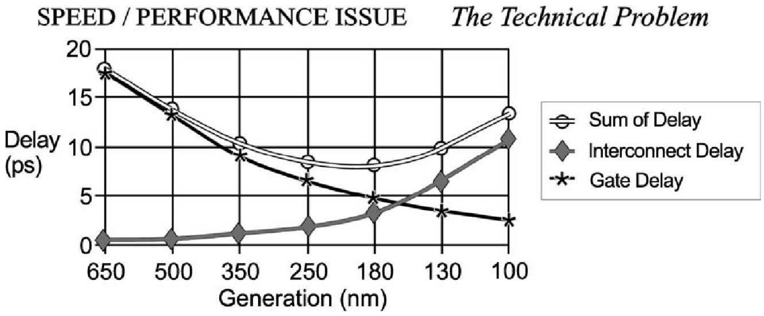


Figure 1.1: Gate and interconnect delays as a function of gate technology.

is presented in the next sections; also, a brief description of the methods employed by Tezzaron and MIT Lincoln Labs, whose technology processes have been involved in this thesis work, will be presented at the end of this chapter.

1.1 Motivation for 3D ICs

Because of advances in IC technology, such as lithography, etching and reduction in defect density, which have led to fabrication of very small feature size devices, 2D chip size is continually increasing [1]. This is due to the increase in chip complexity required to satisfy the ever-growing demand for functionality and higher performance, which in turn requires more and more transistors to be closely packed and connected. Although smaller feature sizes have improved device performance [2], the miniaturization process has led to less positive consequences in terms of performance of interconnect wires. Indeed, smaller wire cross sections, smaller wire pitch, and longer lines running along the chips have increased the resistance and the capacitance of these lines resulting in a significant increase in signal propagation (RC) delay. In particular, as interconnect scaling continues, RC delay represents the dominant factor determining the performance of advanced ICs [3]. At 250 nm technology node, copper with low-k dielectric was introduced in order to limit the effect of increasing interconnect delay [4]. Fig. 1.1 shows the rapid increase in delay time caused by the interconnect loading. It is possible to notice how, beyond the 180 nm regime, interconnect delays affect the benefits introduced by small feature size devices and by the use of new materials in ICs fabrication. Nonetheless, more scaled technologies employing low-k dielectric has been realized, even if these processes turn out to be quite expensive.

Three-dimensional integrated circuits, which contain multiple layers of active devices, extensively utilize the vertical dimension to connect components and are expected to address interconnect delay related problems and to enable integration of heterogeneous technologies [5],[6]. It can be shown that 3D architectures can reduce the overall global wire-length, while increasing the number of local wires [7]. Moreover, the decrease in the number of long interconnects could directly translate to an increase in device density, provided that the devices are efficiently packed, placed, and wired. Another interesting feature is the fact that the 3D architecture lends itself to the realization of a SoC.

SoC is a broad concept that refers to the integration of nearly all aspects of a system design on a single chip. These chips are often mixed-signals and/or mixed-technology designs, including such diverse combinations as analog, RF, low-power logic and embedded DRAM.

There is a plethora of 3D integration techniques in literature, but the most appealing and competitive schemes to date are those involving either low-temperature silicon epitaxy or wafer bonding. A brief review of the technology options for 3D fabrication will be presented in the following sections.

1.2 Benefits of 3D ICs

In the 3D design, an entire (2D) chip is divided into a number of blocks, and each block is placed on a separate layer of silicon that are stacked on top of each other, as schematically shown in Fig. 1.2. The 3D architecture offers extra flexibility in system design, placement, and routing, reducing the earlier discussed negative impact of deep-submicrometer interconnects on VLSI design. Other benefits of 3D ICs include improved packing density, noise immunity, improved total power due to reduced wire length and lower capacitance, and the ability to implement added functionality. Furthermore, the 3D chip design technology may be exploited to build SoC by placing circuits with different

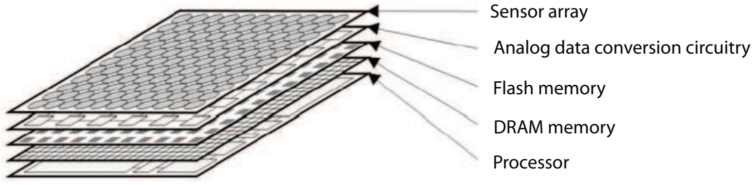


Figure 1.2: A possible design for a 3D Soc.

voltage and performance requirements in different layers.

1.2.1 Power and performance

The effects associated with long interconnect paths in 2D chips are mitigated by the use of short wires in 3D designs. These shorter wires will decrease the average load capacitance and resistance and decrease the number of repeaters which are needed to regenerate signal on long wires. A significant portion of the total power consumption is indeed due to interconnect wires with their supporting repeaters: 3D ICs, reducing the average interconnect length with respect to 2D counterparts, will improve wire efficiency (about 15%) and reduce total active power by more than 10% [8]. Moreover, dynamic power consumption is in the charging and discharging of the interconnect capacitance. The RC delay of signals ultimately limits the maximum speed of a circuit; this delay increases as the square of the wire's length l. The experimental derived forumla is:

$$t_d = 0.35\ r\ c\ l^2. \tag{1.1}$$

Since resistance (r) and capacitance (c) per unit length are increasing with the scaling process, by means of 3D technology is it possible to reduce propagation delay by reducing the wire length l.

Also, regenerative repeaters used in CMOS circuits to restore signal on long wires, may introduce noise through the substrate: reducing the number of the repeaters makes it possible to curb this effect. Moreover, the shorter interconnects in 3D ICs, with the consequent reduction of load capacitance and the lower wire-to-wire capacitance, will reduce the noise due to simultaneous switching events and the noise coupling between signal lines.

1.2.2 Density

By adding a third dimension to the conventional two-dimensional device layout, the transistor packing density is improved, thereby allowing a reduced chip footprint. As an example, Fig. 1.3 compares layout designs of 2D and 3D inverters: it is possible to notice the area gain for the 3D design. When the total layout area (the sum of the device area and the metal routing area) is compared for 2D and 3D standard cells with different inverter designs, a 30% areal gain for the 3D cells can be achieved [9]. The ability to stack circuit elements, thus shrinking the footprint and potentially reducing the volume and/or the weight of a chip, is particularly appealing for wireless or portable electronics, where silicon real estate is at a premium. Reduced chip volume and weight are also motivated by military applications.

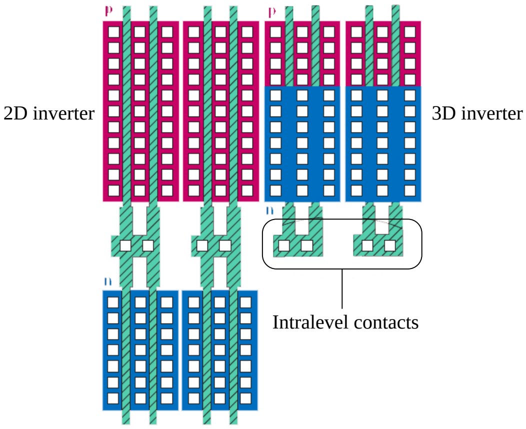


Figure 1.3: 2D and 3D inverter layout.

1.2.3 Functionality

Many of the general techniques for building 3D ICs will facilitate the integration of heterogeneous materials, devices, and signals and enable flexibility in device structure, system design, and routing, making it possible to integrate an entire system onto a single piece of silicon and to implement added functionality in a single 3D chip.

1.3 MAPS sensors: from 2D to 3D

Silicon devices have been used since the 1960s for the detection of radiation [10]. In the early 1990s monolithic pixel sensors have been proposed as a viable alternative to CCD's in visible imaging. These sensors are made in a standard VLSI technology, usually CMOS: in the literature, they are normally referred to as CMOS sensors.
The cross-section view of Fig. 1.4 shows the basic principles underlying CMOS sensors. In most modern CMOS process, n- and p-wells are fabricated on top of a thin p-doped epitaxial layer, with resistivity of the order of 1-10 Ω cm. The epitaxial layer thickness ranges between a few and up to about 20 μm and it is lightly doped with respect to the underlying p-substrate, whose main function is for mechanical support. A pn junction exists between the n-well and the p-epilayer and can be used as the detecting element. Because of the difference

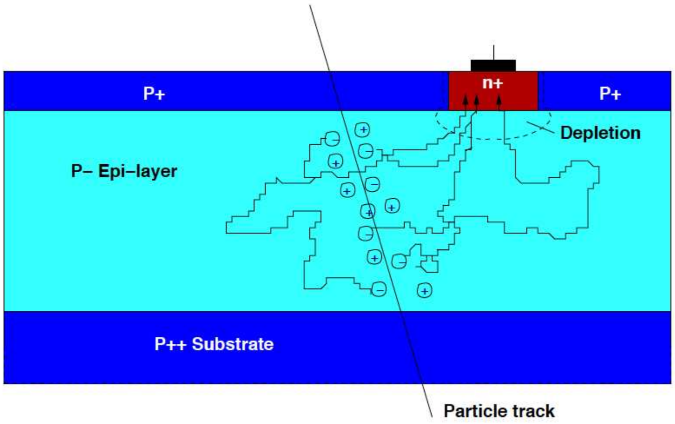


Figure 1.4: Simiplified cross section of CMOS pixel with epitaxial layer .

in doping between the epitaxial layer and the p-well and the p-substrate, a potential difference of a few times kT/q is created. The epitaxial layer acts as a shallow potential well for the electrons, which are the minority carriers. Electrons created by the radiation diffuse in the epitaxial layer till they are close enough to the nwell/p-epi diode, where they experience an electric field. They are then collected by the diode.

The first developments were based on the so-called Passive Pixel Sensors (see Fig 1.5 (a)). Only one selection transistor is integrated in the pixel together with the diode. The charge generated by the silicon/radiation interaction is integrated in the diode. The readout is performed by closing the selection switch and dumping the charge to a charge preamplifier, common to all the pixels in one column. This solution has the minimum amount of in-pixel electronics and thus has a very high fill factor, defined as the ratio between the detecting area and the total area of the pixel detector. It has however disadvantages in terms of speed and noise. The development of monolithic active pixel sensors (MAPS) was mainly pushed by the requirements of low power and low weight for space applications. In a MAPS pixel, an amplifier integrated in each cell directly buffers the charge signal. In the minimum configuration of a MAPS, three transistors (3-T) are integrated in the pixel, as shown in Fig. 1.5 (b). The transistor MRST is used to reset the pixel by dumping the integrated charge to the positive power supply line. The transistor MSEL is activated to select the readout of the pixel and MIN is the a source follower. The current source is common to all the pixels in one column. The diode is reverse biased by

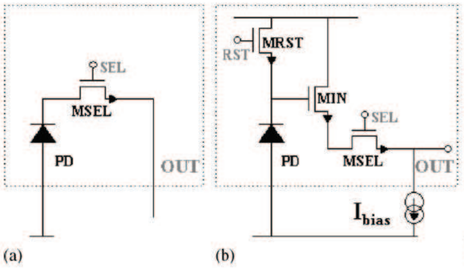


Figure 1.5: (a) Schematic of a Passive Pixel Sensor (PPS). (b) Baseline, 3-MOS Active Pixel Sensor (APS).

connecting to V_{DD} through the reset switch; the charge-to-voltage conversion gain G_{in} at the input is given by $1/C_{in}$, where C_{in} is the total capacitance seen in the input node, and mainly determined by the diode capacitance and the gate-to-source capacitance of the input transistor. Typical values for the sensitivity G_{out} at the output of the source follower are in the range of 10-50 μV/e$^-$.
Today most CMOS imagers have a MAPS structure for these main reasons:

- low cost, since they are fabricated in a standard VLSI technology;
- low power, since the circuitry in each pixel is active only during the readout and, contrary to CCD's, there is no clock signal driving large capacitances: the total power dissipation is usually in the range of 100mW for a few millions pixel device even with integrated analogue-to-digital conversion [11];
- random access, since each pixel can be addressed directly for readout;
- increased functionalities, taking advantage of the full capabilities of the CMOS technology: the control logic, the analogue-to-digital converter or other signal processing blocks can be integrated in the same substrate as the sensor array.

Because of these features, CMOS sensors are a candidate technology for demanding applications, which are typically found in space science. They also look attractive for particle tracking applications because of the following features:

- spatial resolution: a small pitch is achievable, and hence good spatial resolution even with a binary readout. Taking advantage of possible analogue readout and natural charge spread between neighbouring pixels, for very demanding application the spatial resolution can possibly be pushed down to less than 1 μm;

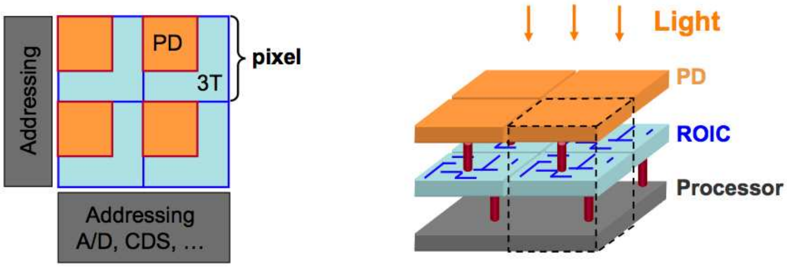
- low mass detectors: CMOS MAPS are adequate in terms of material budget, since their sensing element shares the same substrate with the readout electronics; furthermore, substrate thickness can be reduced to a few tens of microns with no significant signal loss;

- radiation tolerance, taking advantage of the reduced radiation sensitivity offered by nowadays submicron VLSI processes. [12]

Anyway, the use of large area electrodes is strongly discouraged in 3T-MAPS design: indeed, increased capacitance would unacceptably degrade the noise figure and the charge sensitivity at the same time. In such sensor, in fact, charge sensitivity is inversely proportional to the parasitic capacitance of the collecting electrode and the equivalent noise charge (ENC), originating mainly from the reset operation, is proportional to the same capacitance. Moreover, in standard CMOS MAPS, use of PMOS devices in the design of the front-end electronics is avoided as the n-well they are integrated in might subtract charge to the collecting electrode leading to potentially serious efficiency loss. A possible 3D translation of a monolithic active pixel sensor is shown in Fig. 1.6. The figure displays the layout of a conventional MAPS device and its 3D counterpart, where the different sections of the 2D pixel cell have been placed in different layers. In the 2D version, pixel electronics and detector share the same area, whereas in the 3D case an ideal 100% fill factor is achievable. By placing analog and digital sections on a separate layer from the sensor, both N-channel and P-channel devices can be used in the circuit design. Moreover, the separation of the analog and digital layers provides a large amount of functionality in each pixel cell. Last, in the 3D version, the processing for each layer can be optimized by layer function.

1.4 3D IC fabrication technology

3D ICs can be fabricated by means of four basic methods: chip stacking, transistor stacking, die-on-wafer stacking and wafer-level stacking. These four methods are described in this section. The most useful approaches for many applications, such as HEP (*High Energy Physics*) and imaging application, are



Conventional MAPS

3D MAPS

Figure 1.6: Comparison of MAPS and 3D pixel layout. In the 3D version, the photodetector (PD), the readout electronics (ROIC, more advanced with respect to the simple three-transistors, 3T, architecture employed in the 2D version) and the digital processing blocks are split into three different layers.

the last two, that is die-on-wafer and wafer-level stacking. These approaches share four common key technologies:

- bonding between layers using oxide to oxide fusion, copper tin eutectic bonding, or polymer bonding;
- wafer thinning using a combination of grinding, lapping, etching and CMP (*Chemical Mechanical Polishing*);
- through silicon vias (TSV) using different processes (some of which need to include hole passivation);
- high precision alignment of parts before bonding.

1.4.1 Chip stacking

This method stacks fully processed and tested stand-alone components to produce a System-in-Package (SiP). SiP products are fully functional systems or sub-systems in an IC package format. SiP may contain one or more IC chips (wirebonded or flip chip) plus other components that are traditionally found on a system board such as surface mount discrete passive components, connectors, EMI shields and so on; a schematic view of a SiP is shown in Fig. 1.7. The only significant benefit offered by chip stacking is the reduction in system size. Indeed, connecting wires may be slightly shorter, but the achievable integration density is not different if compared with a planar system. In the same

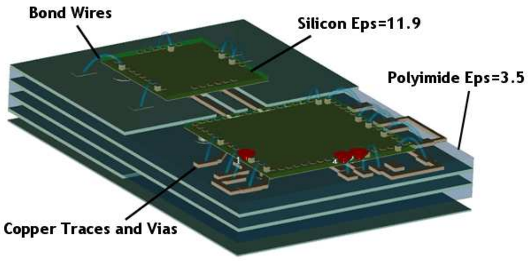


Figure 1.7: Schematic view of a SiP.

way as standard 2D system, signals traveling from one layer to another must be pushed off-chip and then brought on-chip. Existing market uses for SiP include RF and wireless devices (such as power amplifiers, GPS modules, cellular, Bluetooth solutions), digital baseband solutions for the wireless markets and controllers for hard drives in the storage market. SiP technology can also be used to enhance single component packages that require improved circuit performance and reduced board real estate.

1.4.2 Transistor stacking

This fabrication method makes it possible to create multiple levels of transistors on a single substrate. The success of this methodology is limited by thermal budget issues: copper or aluminum already laid down would be damaged by the temperatures required to build a new layer of transistors. Moreover, these temperatures could cause migration of transistor implants on other layers. Promising research on transistor stacking technologies has been performed by the Stanford University. Stanford's studies concern laser annealing and nickel nucleation. Laser annealing can be used for electrical activation of dopants without excessively heating material deeper within the work piece, making it possible to fabricate transistors on the upper levels of a general 3D structure without affecting the reliability of devices below [13]. Unfortunately, high defects density in these structures represents a problem. The nickel nucleation method allows to realize high quality multilayer structures at lower temperatures: nevertheless nickel ions can be kept in the structure, causing system failures. Matrix Semiconductor produces a highly successful variation on stacked transistors in its one-time programmable (OTP) memories [14]. This application turns out to be restricted only to special cases, and it can not be adopted in a large scale: indeed, this process doesn't provide the speed

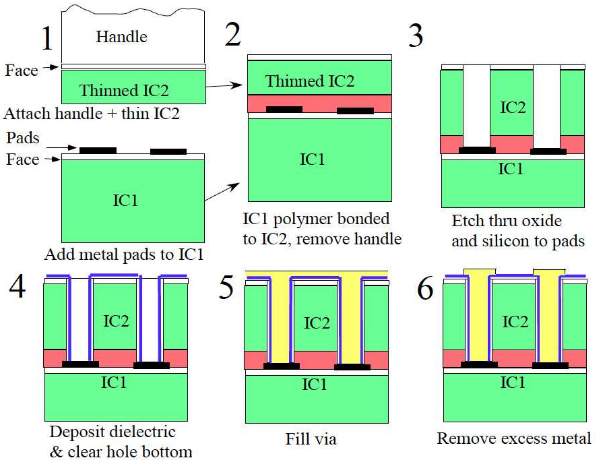


Figure 1.8: Steps showing polymer bonding and via formation for die-to-wafer process.

or actual transistors required by most other devices.

1.4.3 Die-on-Wafer stacking

In this method, system components are built on two semiconductor wafers. One wafer is diced: the singulated known good dies (KGDs) are aligned and bonded onto die sites of the second wafer. Dies can be attached to the host wafer by means of organic glues, oxide bonding or metal bonding. The resulting structure is further processed for thinning and formation of interconnects. In the die-on-wafer stacking approach interconnects can be made on the edge of the die, on the bonded faces themselves, or through-die. Depending on the type of interconnect, this method may realize a higher level of integration with respect to the chip-stacking one [15], with better cost per connection and higher interconnect density. The quality of die-on-wafer stacking is determined by the dies placement accuracy and therefore by the capabilities of the pick-and-place equipment that positions each die on its wafer. Presently, placement accuracy is about 10 µm, which limits the achievable density interconnect. Another issue which has to be taken into account is that the equipment cannot handle naked circuitry and so it does not protect adequately against static discharge. For this reason, the processed dies include protection structures which represent a cost in terms of power, speed, and die size. With die-on-wafer approach, the parts may be bonded face to back (that is circuit side to

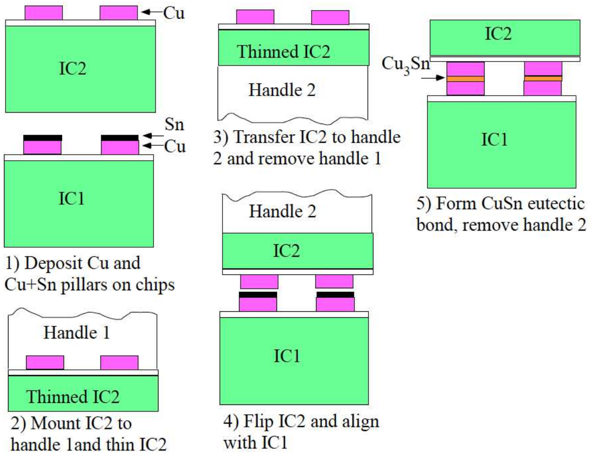

Cu
IC2
Sn
Cu
IC1
1) Deposit Cu and Cu+Sn pillars on chips
Handle 1
Thinned IC2
2) Mount IC2 to handle 1and thin IC2
Thinned IC2
Handle 2
3) Transfer IC2 to handle 2 and remove handle 1
Handle 2
IC2
IC1
4) Flip IC2 and align with IC1
IC2
Cu_3Sn
IC1
5) Form CuSn eutectic bond, remove handle 2

Figure 1.9: Face to face bond using Cu_3Sn.

substrate side) or face to face. Fig 1.8 shows a possible process sequence for device bonding and vias formation. Inter-device vias require to leave space in the design; if vias pass through a CMOS layer, they must be insulated from the substrate.

It is possible to realize face to face bonds by means of Cu_3Sn eutectic bond for both the electrical and mechanical connections between devices. The process for the Cu/Sn bond is shown in Fig. 1.9: a double handle transfer is needed as shown in steps 2 and 3.

1.4.4 Wafer-level stacking

This last method bonds entire wafers into a stack. Wafers are frequently bonded together using an SiO_2 bond. For good bonding, the wafers must be very flat and the surfaces must be extremely clean. As in die-on-wafer stacking, alignment accuracy, which is better in wafer-level stacking, determines the interconnect density. Technology processes provided by Tezzaron Semiconductor can achieve alignment of less than a micrometer. Wafer-level stacking supports a lower cost per connection and better interconnects density than die-on-wafer approach, because of the greater alignment accuracy and higher degree of surface planarity. As in the case of die-on-wafer stacking, process temperatures limit the use of mixed substrates: high temperatures, indeed, cause misalignment due to unequal expansions between mixed-wafer pairs. In wafer-level approach, all processing is done at the wafer level: wafer handling equipment protects against static discharge, so I/O buffering between layer is not required. Moreover, standard lithography and processing techniques, with few added process steps, can be used in structures fabrication with the wafer-level stacking. Bonding between wafer, as in the die-on-wafer method, can be made by means of organic gluing, oxide bonding and metal bonding. Each bonding technique has its benefits and drawbacks. Organic glue bonding reduces the possibility of particle contamination, but it does not provide a good metal interconnect. Oxide bonding makes it possible to improve alignment accuracy because of the room temperature prebonding process, but again it does not have an intrinsic metal connection. Metal bonding provides the interconnect, but alignment turns out to be more difficult because of the higher temperatures needed (about 400°C). Wafer-level stacking techniques are further differentiated by the method used to create TSVs: either via-first or via-last. In the via-last approach, interconnect is created after wafers are bonded, whereas via-first processes build the TSVs on each wafer prior to the bonding process, which is generally more efficient and cost-effective [16]. SOI wafers have several advantages for 3D wafer stacking. Each wafer can be easily thinned to the buried oxide layer (BOX) since the buried layer is used as an etch stop for the silicon etch to produce a uniformly thin active layer.

1.5 Tezzaron's 3D solution

Tezzaron 3D IC fabrication process approach, involved in the fabrication of the prototype chip subject of this thesis work, is wafer-level, via-first, with metal-to-metal thermal bonding. This kind of bonding, realized by means of copper, provides both mechanical and electrical connectivity in one step. Copper bonding has the main benefit, with respect to other materials, of being already used as standard part of normal CMOS processing. Moreover, it provides excellent electrical and thermal dissipation characteristics and it is easily planarized using existing chemical mechanical polishing (CMP) technologies. Tezzaron has developed two kind of vertical interconnection between the layers, whose main characteristic are summarized in Table 1.1. The first generation of Tezzaron's interconnect, namely SuperVia, has the advantage of not requiring any foundry process changes, as the vertical interconnect is entirely post processed. However, each SuperVia requires an open field area with no transistors and no interconnect. The SuperVias diameter has also to be quite large in order to keep a conservative aspect ratio, constraining the density of the interconnect. On the other hand, the second generation of Tezzaron's interconnect, namely SuperContact, is built during wafer fabrication, so it requires the foundry to

	Super-Via^{TM}	***Super-Contact***^{TM}
Size [µm x µm x µm]	4.0 x 4.0	1.2 X 1.2
Material	Copper	Tungsten
Minimum pitch [µm]	6.08	< 4
Feedthrough capacitance [fF]	7	2 - 3
Series resistance [Ω]	< 0.25	<0.6

Table 1.1: Tezzaron's interconnect properties. [17].

perform a unique process step. Pushing the vertical interconnect fabrication into the foundry significantly reduces the process complexity and equipment requirements for the subsequent stacking operation. The following sequence, Figs. 1.10-1.18, illustrates Tezzarons stacking method with the SuperContact interconnect.

Fig. 1.10 shows the cross-section of one wafer, immediately after transistors have been created, but before contact metal. In Fig. 1.11 the vertical SuperContact is etched through the oxide and into the silicon substrate approximately 6 μm; the walls are lined with silicon oxide or silicon nitride. The SuperContact is filled with tungsten and finished with CMP polishing, as shown in Fig. 1.12. This operation completes the unique processing requirements at the wafer level. The wafer is finished with its normal processing, as shown in Fig. 1.13, which can include a combination of aluminum and copper wiring layers. It is worth noticing that the last layer must be copper. All the wafers included in the 3D stack are processed by means of the discussed process. Then, the first and the second wafer are aligned and bonded in a copper thermal diffusion process that takes places in a vacuum at approximately 375° and 40 psi. The step is shown in Fig. 1.14. Several minutes are required to form the bond: typical cycle time within the bonder is 20 minutes. After bonding, the top wafer is thinned to the bottom of the SuperContact, as shown in Fig. 1.15. This leaves a substrate thickness of about 6 µm. 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. In Fig. 1.16 a third wafer has been added to the stack, using the same technique by which the second wafer was added. Then, the stack is inverted, as shown in Fig. 1.17: indeed final processing will be applied to the backside of the first wafer. The first wafer undergoes the same thinning process used before, stopping on the tungsten SuperContact, as shown in Fig. 1.18. Instead of a copper damascene process for bonding pads, an aluminum layer is deposited for normal wire bonding.

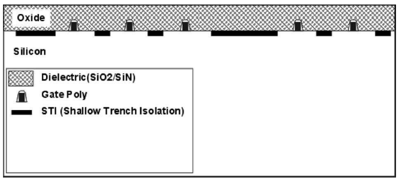


Figure 1.10: Cross-sectional view of the wafer.

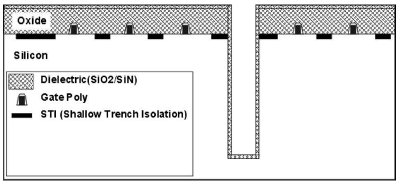


Figure 1.11: SuperContact creation.

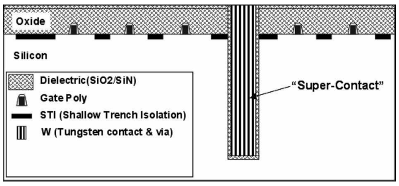


Figure 1.12: SuperContact is filled with tungsten.

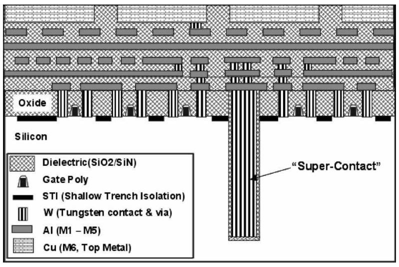


Figure 1.13: The first wafer is finished with its normal processing.

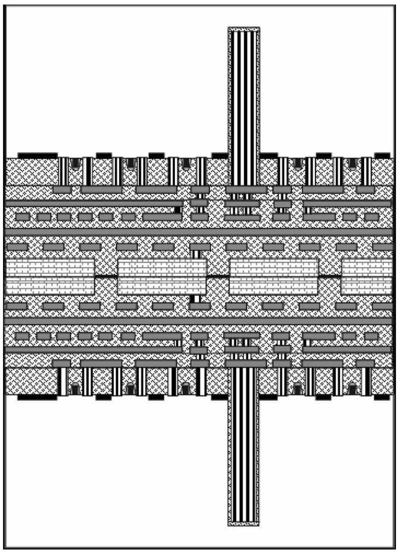


Figure 1.14: Interconnection of the two wafers

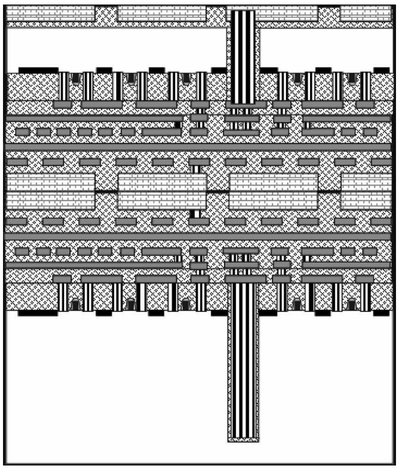


Figure 1.15: Fabrication of the bonding pads for subsequent stacking

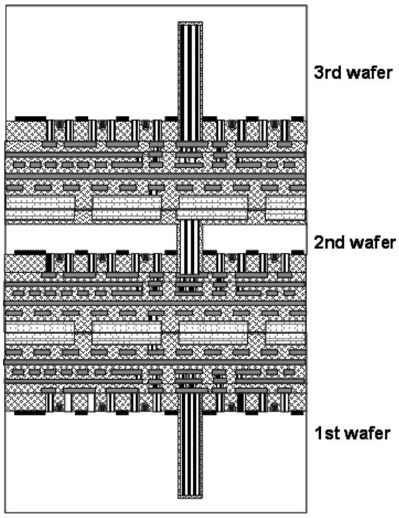


Figure 1.16: A third wafer is added to the stack.

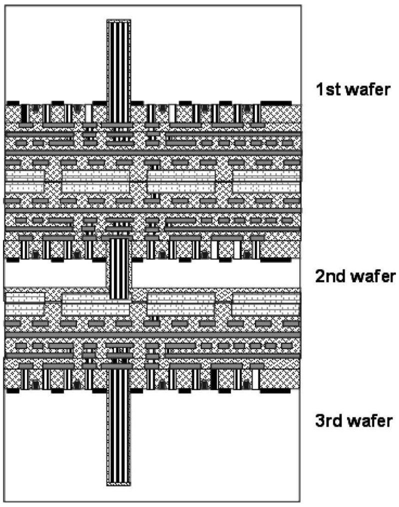


Figure 1.17: The stack is inverted.

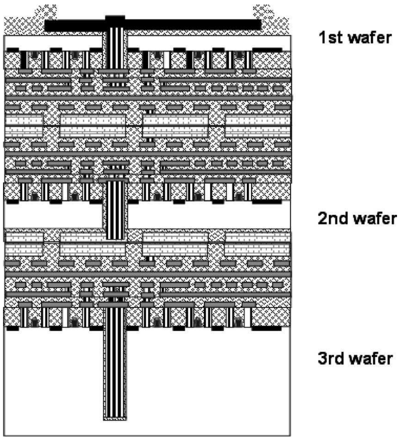


Figure 1.18: Aluminum layer is deposited for wire bonding.

Tezzaron has stacked a variety of wafer types, including SOI and bulk wafers, using both the original Super-Via and the new enhanced second generation process flows. Experiments have shown that the 3D-enabled wafers can be thinned to as little as a few microns. They can be stacked with sub-micron alignment. The bonded wafers have a *Cu-Cu* bond strength that is greater than required and actually stronger than the Cu-SiO_2 interface. Also, the interconnected wafers can be handled in a normal fashion without special handlers or precautions. The transistors on the stacked, bonded, and thinned wafers were shown to have no discernable performance differences from their original 2D form. Tezzaron Semiconductor has chosen the Chartered Semiconductor 0.13 µm process as the candidate process in the fabrication of their 3D devices. Chartered's 0.13 µm process offerings includes logic, high performance, low voltage and low power. Chartered's logic process is built upon a modular architecture that allows modules of embedded memory, analog/mixed signal and RF CMOS to be easily and efficiently incorporated. With up to eight metal layers available, Chartered's 0.13 µm solutions offer a variety of transistor options with multiple threshold voltages, including 1.0, 1.2 and 1.5V core, and 2.5 and 3.3V I/O, making them well-suited for a large variety of applications. Chartered's 0.13 µm process offers deep N-well (DNW) structure which can be incorporated for noise isolation from adjacent switching circuits or for other special functions, as in the case of DNW-MAPS discussed in the third chapter. The variety of transistors (see Fig. 1.19) can be used simultaneously in a design and are designed for compatibility with other industry processes,

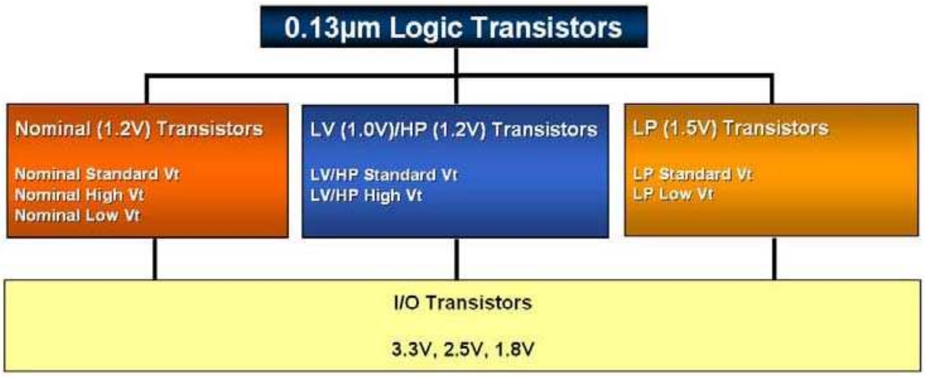


Figure 1.19: Chartered 0.13 µm process transistors.

allowing migration to other fabs. Chartered has an extensive silicon proven library for the 0.13 µm process which includes standard cells, memory compilers and I/O cells. The 0.13 µm process is currently available for 8-inch wafer production.

1.6 MITLL's 3D solution

Another 3D integration scheme, developed recently at MITs Lincoln Laboratory, can interconnect three separate layers of circuitry on fully fabricated SOI-based wafers. The method involves inverting and aligning one wafer over another using an infrared registration technique, and then forming a low-temperature oxide bond. Handle silicon is removed from the top-facing side of the inverted wafer in a stripping process, utilizing the buried oxide layer of the SOI wafer as an etch stop. Concentric 3D vias are then etched through both the top tier and the oxide-bonded layer all the way to the bottom-tier circuitry, and tungsten interconnect is deposited in a damascene process. The steps are repeated to add a third chip and to expose the underside of the top layer to serve as a bond pad in the final package. This process flow is schematically shown in Fig. 1.20. Proof-of-concept applications of Lincoln Labs three-tier integration approach include a ring oscillator test circuit, a 1024 x 1024 pixel imager with some one million 3D vias across the array, a photodiode laser radar and the bonding of a SOI CMOS circuit layer to an InP handle wafer, to enable higher-density and longer-wavelength focal plane detectors. Other

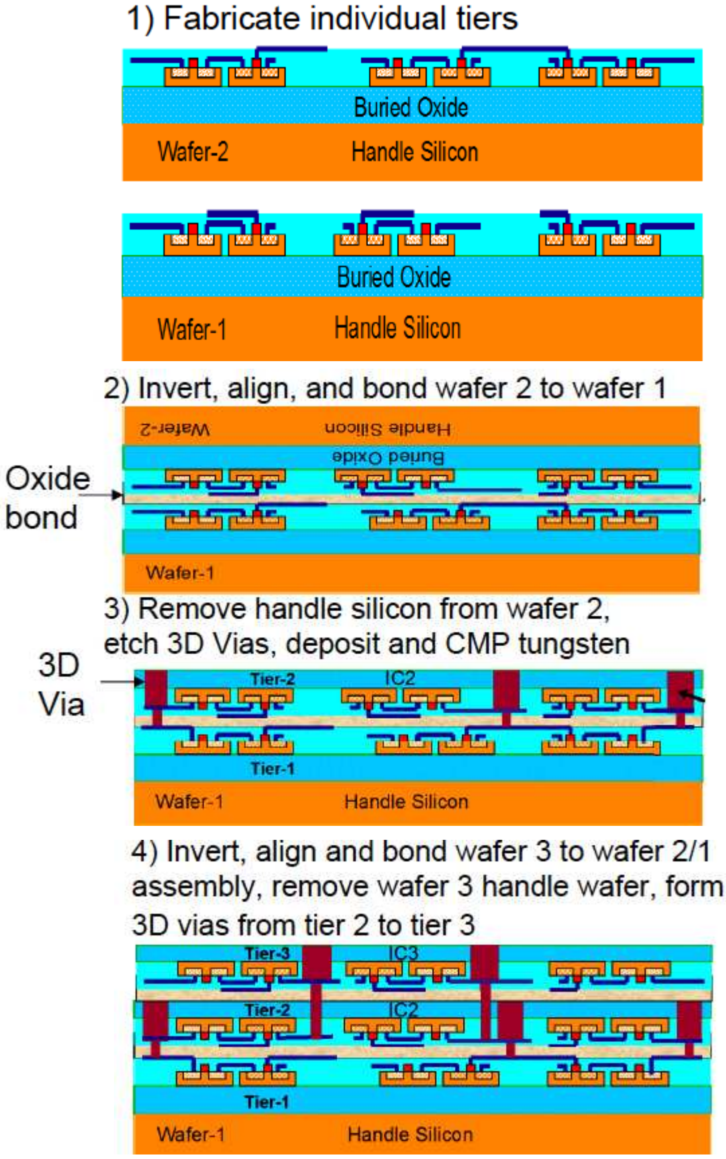

1) Fabricate individual tiers
Buried Oxide
Wafer-2
Handle Silicon
Buried Oxide
Wafer-1
Handle Silicon
2) Invert, align, and bond wafer 2 to wafer 1
Wafer-2
Handle Silicon
Buried Oxide
Oxide bond
Wafer-1
3) Remove handle silicon from wafer 2, etch 3D Vias, deposit and CMP tungsten
3D Via
Tier-2
IC2
Tier-1
Wafer-1
Handle Silicon
4) Invert, align and bond wafer 3 to wafer 2/1 assembly, remove wafer 3 handle wafer, form 3D vias from tier 2 to tier 3
Tier-3
IC3
Tier-2
IC2
Tier-1
Wafer-1
Handle Silicon

Figure 1.20: 3D SOI process flow.

3D concepts being explored by Lincoln Lab in conjunction with some industry and university partners include 3D FPGAs, ASICs, flash memory, and nano-radio and RF tags.

Chapter 2

Characterization of a 180 nm CMOS SOI Technology

Silicon-on-insulator (SOI) is emerging as a strong technology candidate for low-power, high-performance applications [18]. Integrated circuits fabricated on SOI substrates have been of increasing interest as the starting material has improved in quality, leading to highly promising circuit results. Moreover, as earlier discussed, SOI technology is particularly well-suited for 3D design. Since the substrate is not electrically connected to the SOI transistors, it can be removed through etching: the buried oxide layer (BOX) provides an adequate etch stop. In MIT-LL's 3D SOI process, based on their 0.18 μm technology, signals from adjacent levels in the stack can be interconnected through compact 2.5 μm wide interdie vias [19].

SOI technology can be combined with vertical integration techniques in order to realize 3D multilayer structures including sensors and mixed-signal readout electronics with high functional density [20]. Furthermore, CMOS SOI technologies have become very attractive for the design of advanced sensor and readout electronics for the future generation of high energy physics (HEP) experiments [21], [22]. SOI detectors wafers are formed by bonding together a top wafer with low resistivity and a bottom wafer with high resistivity, by means of a silicon oxide bond. After bonding, the top wafer is thinned to just a few microns using one of several different techniques. The low resistivity layer may host the readout electronics, whereas in the high resistivity one it is possible to integrate fully-depleted detector elements.

In view of employing this technology in the fabrication of 3D sensors for HEP and imaging applications, an accurate characterization of SOI devices is mandatory. In this chapter the SOI MOSFET will be introduced, followed by the study of static, signal and noise performance of fully depleted silicon-on-insulator (FD-SOI) devices, provided by MIT Lincoln Laboratory.

2.1 SOI MOSFETs

With silicon on insulator technology, MOSFETs are formed in a thin top silicon layer separated from the silicon substrate by an insulating layer. This structural feature provides SOI devices with several advantages for high-speed, low-power operation. Fig. 2.1 shows the SOI structure and its equivalent electrical representation. Depending on silicon thickness and body doping value, the channel region can be fully depleted (FD) or partially depleted (PD). The FD-SOI, in principle, are not affected by kink-effect or by other effects related to the floating body of the device. The key feature of a FD-SOI MOSFET is represented by the fact that the depletion region reaches all the way to the bottom of the silicon film. As a result, the body region is fully depleted, as the name of the device indicates. Usually, for deep submicron technologies, optimization of FD devices requires an ultra-thin silicon layer while only a thin silicon layer (about 100 nm) is adopted in the case of PD devices. Fig. 2.2 (a) shows the cross-sectional structure of an FD-SOI nMOSFET; (b) is an energy band diagram along the line A-A′ in (a), which runs through the source, body, and drain near the bottom of the body region; (c) is an energy band diagram showing how the energy bands change along the line B-B′ in (a), which runs from the gate oxide film down into the body region near the

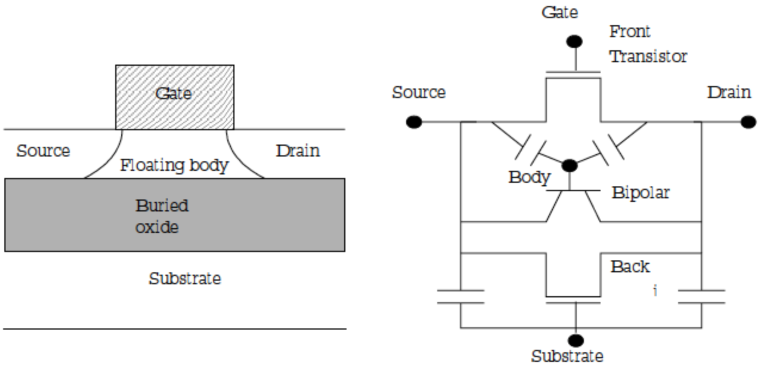


Figure 2.1: SOI transistors.

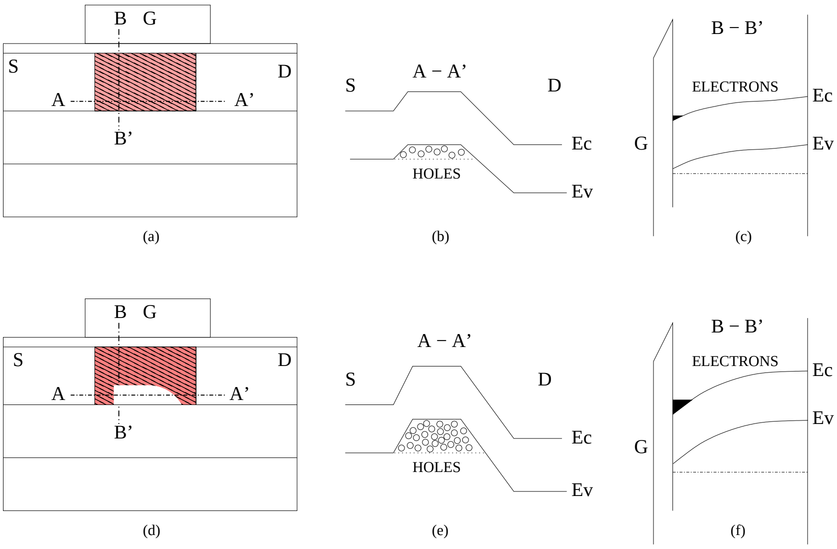
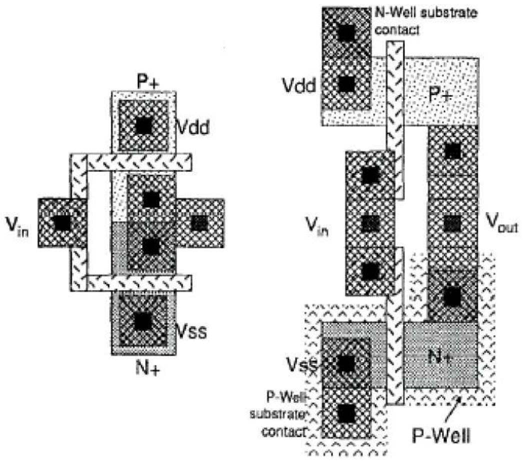


Figure 2.2: FD- and PD-SOI MOSFETs. (a) Cross section of an FD-SOI device, (b) the energy band diagram along the line A-A', and (c) the energy band diagram along the line B-B'. (d-f) show the corresponding figures for a PD-SOI device.

source end. The corresponding figures for a PD-SOI nMOSFET are shown in Fig. 2.2 (d-f). In a FD-SOI device, the entire body region is depleted in both the "on" and "off" states, as shown in Fig. 2.2 (a). In contrast to an FD-SOI device, a PD-SOI device has an undepleted neutral region at the bottom of the body region, as shown in (d). This difference results in a different potential distribution inside the body region. In an FD-SOI device, the entire body region has a potential gradient in the depth direction, as shown in (c), and the gate field extends right into the BOX. In a PD-SOI device, the influence of the gate field stops inside the body region, as shown in (f), and there is a neutral region with no potential gradient at the bottom of the body region. Accordingly, the potential difference between the top and the bottom of the body region is larger in PD device and the potential barrier to holes between the source and the body near the bottom of the body region is higher. This difference in the barrier height for holes leads to a difference in the number of holes that accumulate in the body region. Holes are generated by impact ionization near the drain. During the operation of an nMOSFET, when channel electrons pass through the high-electrical-field region near the drain, they



a) SOI inverter b) bulk inverter

Figure 2.3: SOI and bulk CMOS inverters.

gain energy from the field and jump to higher energy levels. The high-energy electrons collide with valence electrons and generate more electrons and holes. The electrons flow into the drain, and the holes flow toward the source via the bottom of the body region. When this happens, more holes accumulate at the bottom of the body region in a PD-SOI than in an FD-SOI device because PD-SOI device has a higher potential barrier. This leads to a large difference between the floating-body effects of the two types of devices, such as the kink in the drain current-voltage characteristics and the stability of the dynamic characteristics.

Compared to a classical bulk transistor, SOI can reduce the capacitance at the source and drain junctions significantly by eliminating the depletion regions extending into the substrate. This results in a reduction in the RC delay due to parasitic capacitance, and hence a higher speed performance of the SOI CMOS devices compared to bulk CMOS. The independent body bias of SOI MOSFETs makes them faster in a stacked-gate structure: in the stacked gates made with bulk MOSFET the negative body bias increases the threshold voltage and lowers the operating speed. In contrast, the body bias of stacked SOI MOSFETs is positive because it takes a value between the source and drain voltage. This yields a lower threshold voltage for stacked transistors, thereby enhancing the operating speed. Moreover, SOI devices are laterally isolated from each other by an insulator film, and vertically isolated from the substrate by the BOX, which makes the isolation ideal. As a result, SOI devices can be packed closer together than bulk ones. In addition, the n+ and p+ diffusion regions at the output of a CMOS inverter can be connected directly to each other, as shown in Fig. 2.3, wich makes the area of the device smaller than that of a bulk one.

2.2 Investigated devices

The static and noise analysis is carried out for both N-channel and P-channel FD devices with various gate geometries. The BOX thickness of the investigated devices is 400 nm and the thickness of the active silicon film above the BOX is 40 nm. The maximum allowed supply voltage V_{DD} is 1.5 V. For all devices a body contact pad, which was grounded during the measurements, is available. Gate dimensions of investigated devices (gate width W and gate length L) are shown in table 2.1. The same geometries are available for both N-channel and P-channel devices. The MOSFETs are mounted in a 40 pins DIL package; all device contacts have no protecting diodes. MOSFET featuring a W/L=8/0.2 were provided with a contact between body and source.

	W [µm]	L [µm]	*Note*
NMOS	100	0.5	
		0.2	
	60	0.5	
	8	0.2	
	8	0.2	S
PMOS	100	0.5	
		0.2	
	60	0.5	
	8	0.2	
	8	0.2	S

Table 2.1: Gate geometries of the N-channel and P-channel devices. Devices marked with S are of body-tied-to-source (BTS) type.

2.3 Static measurements

For each device the following measurements have been performed to study the static behavior:

- I_D versus V_{GS} with V_{DS} as a parameter and V_{BS}=0
- I_D versus V_{DS} with V_{GS} as a parameter and V_{BS}=0

The value of the transconductance g_m, defined as:

$$\frac{\partial I_D}{\partial V_{GS}} \tag{2.1}$$

was extracted from I_D-V_{GS} curves. Measurements of static parameters were carried out by means of an Agilent E5270B Precision Measurement Mainframe with E5281B SMU Modules.
Fig. 2.4 shows the I_D-V_{GS} characteristic for NMOS and PMOS devices featuring a gate width of 100 μm and gate lengths of 0.5 and 0.2 μm, for different values of the V_{DS} voltage, whereas Fig. 2.5 shows the same characteristics in a logarithmic scale. From these plots it is possible to notice a lateral shift in the curves, which points out that some devices are significantly affected by the drain induced barrier lowering: this effect will be discussed in section 2.5.
Fig. 2.6 shows the I_D-V_{DS} characteristic for different values of the gate-to-source voltage. It is possible to notice that, for lower V_{GS}, NMOS devices are more significantly affected by the kink effect with respect to the PMOS devices.
Figs. 2.7 and 2.8 show the gate transconductance as a function of the gate-to-source voltage and of the drain current respectively, for different values of the drain-to-source voltage.

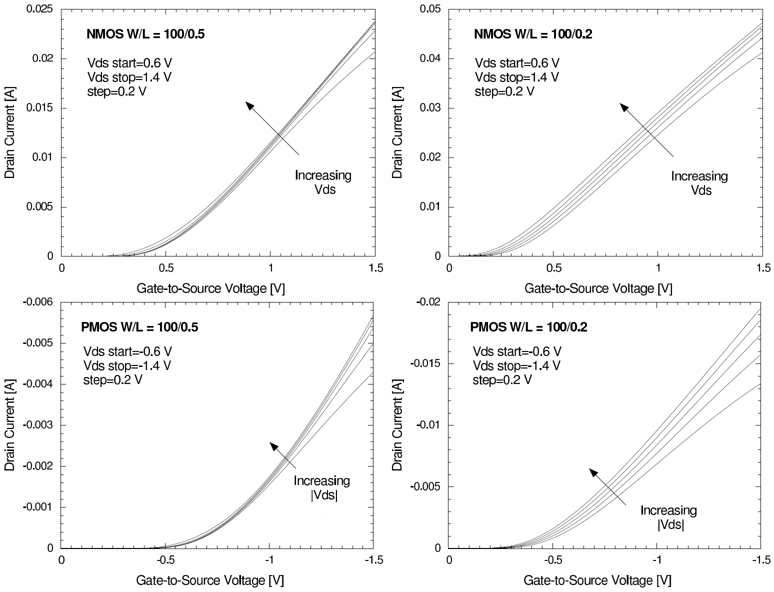


Figure 2.4: Drain current I_D versus V_{GS} at different V_{DS}, for both N-channel and P-channel devices.

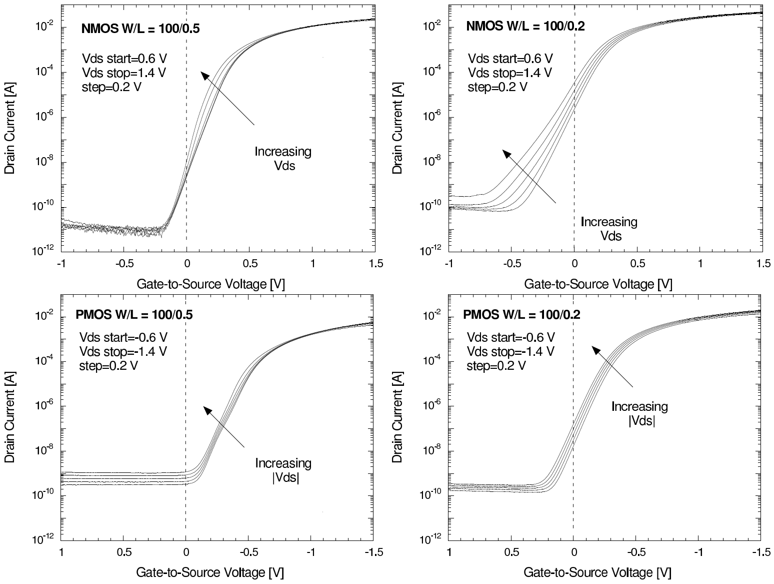


Figure 2.5: Drain current I_D (in logarithmic scale) versus V_{GS} at different V_{DS}, for both N-channel and P-channel devices.

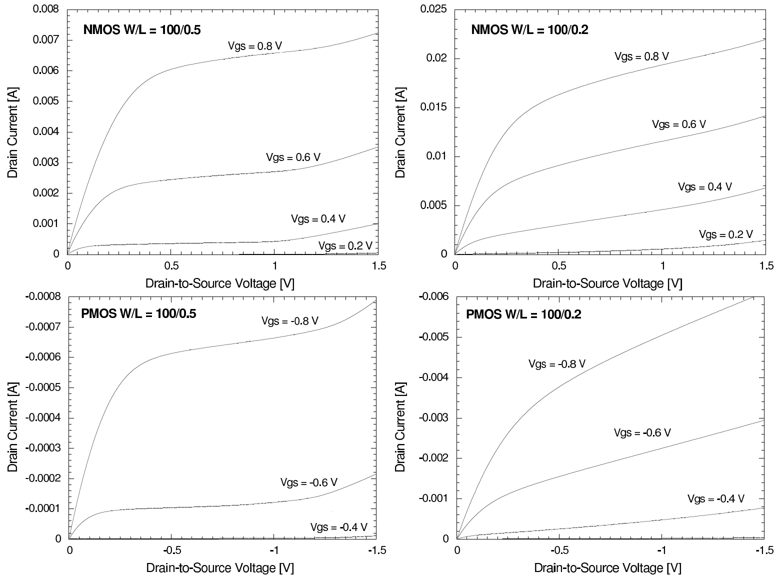


Figure 2.6: Drain current I_D versus V_{DS} at different V_{GS}, for both N-channel and P-channel devices.

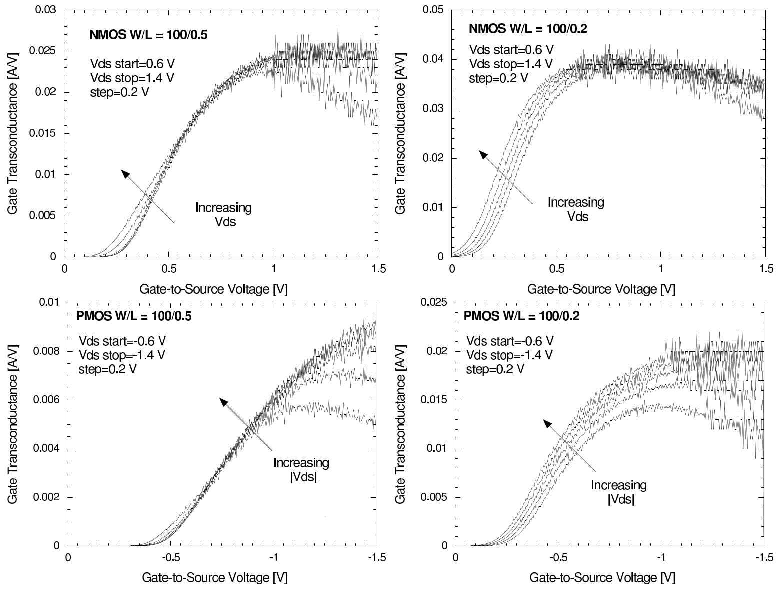


Figure 2.7: Gate transconductance g_m versus V_{GS} at different V_{DS}, for both N-channel and P-channel devices.

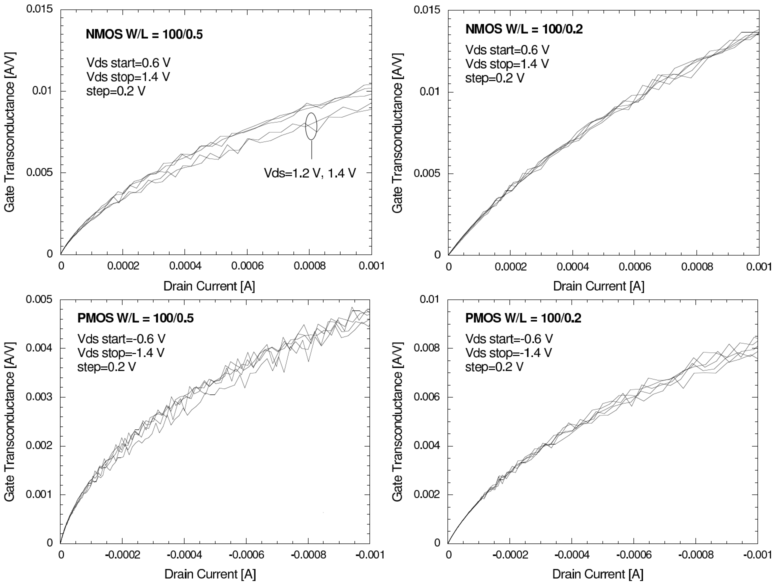
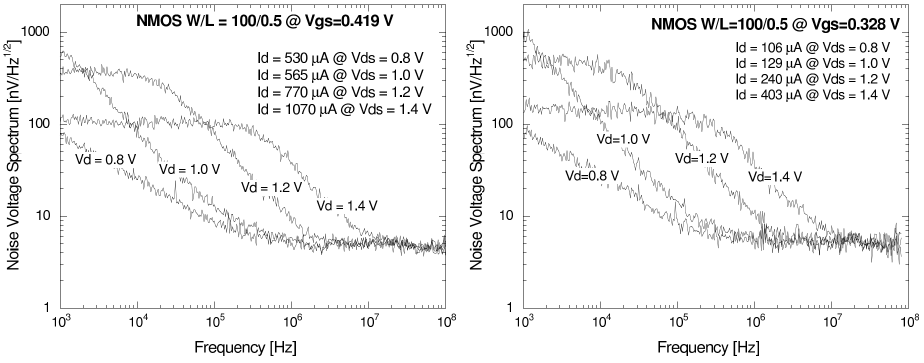


Figure 2.8: Gate transconductance g_m versus I_D at different V_{DS}, for both N-channel and P-channel devices.

2.4 Noise measurements

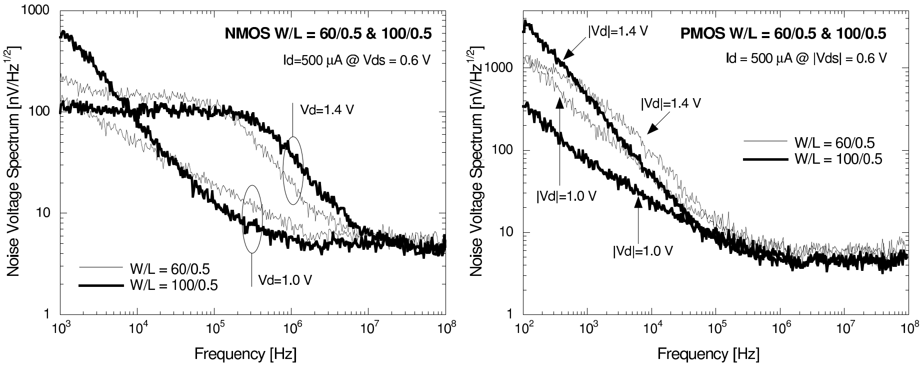
The spectral density of the noise in the channel current of the examined devices was studied by measuring the equivalent noise voltage spectrum referred to the gate. These measurements were carried out with a Network/Spectrum Analyzer HP4195A and a purposely developed interface circuit which allows for noise measurements in the frequency range 100 Hz-100 MHz [23]. Noise measurements have been performed in the low current density region which is of major concern for low-power applications. In particular, the devices were characterized at V_{DS}=0.6, 0.8, 1.0, 1.2, 1.4 V and at I_D=0.050, 0.10, 0.25, 0.50, 1.00 mA; body terminal was kept to ground for all noise measurements. A large set of noise measurements is focused on device behavior under different drain-to-source voltages in order to evaluate the effect of this bias condition on the low-frequency noise.
A subset of the results of the performed measurements, with some of relevant comparison between spectra, is presented in this section.

Fig. 2.9 show the noise voltage spectra of an NMOS device featuring a W/L=100/0.5 for different drain-to-source voltages and for different drain currents. It is possible to notice that the noise spectrum exhibits a lorentzian-like contribution as V_{DS} increases. The same behavior is shown in Fig. 2.10 (a), which displays the noise voltage spectra of two NMOS featuring a gate length of 0.5 μm and with different gate width. The same plot is shown in Fig. 2.10 (b) for PMOS devices.



(a) (b)

Figure 2.9: Noise voltage spectra of an NMOS with W/L=100/0.5 at different drain-to-source voltage. I_D=500 μA (a), I_D=100 μA (b) @ V_{DS}=0.6 V.



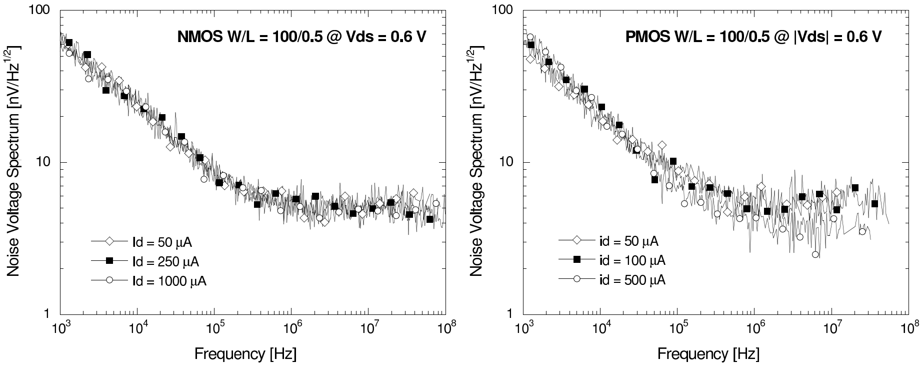
(a) (b)

Figure 2.10: Noise voltage spectra of two NMOS (a) and two PMOS (b) with L=0.5 μm and variable W at different drain-to-source voltage (I_D=500 μA @ V_{DS}=0.6 V).

Fig. 2.11 displays the noise voltage spectra for devices of both polarity for different drain currents. It is possible to notice, in the case of NMOS devices, that high frequency noise contribution is almost independent of the drain current. This may be due to parasitic resistances associated with the device, as discussed in section 2.6.

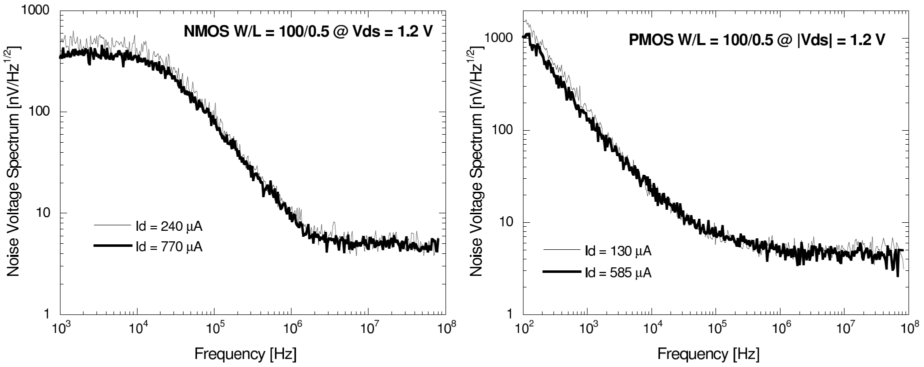
Fig. 2.12 shows the noise voltage spectra of NMOS and PMOS devices with W/L=100/0.5 for a drain-to-source voltage (absolute value) of 1.2 V and for two different values of the drain current, whereas plots in Fig. 2.13 display the noise voltage spectra of MOSFETs with a gate width of 100 μm varying the gate length. It is possible to notice that excess white noise, already highlighted in Fig. 2.11 (a), increases when the gate length is shrinked.

As a last comparison, Fig. 2.14 shows the noise voltage spectra of devices featuring a W/L=8/0.2 with and without body tied to source, for a drain current of 100 μA. Body-tied-to-source devices exhibit a smaller low frequency noise contribution.



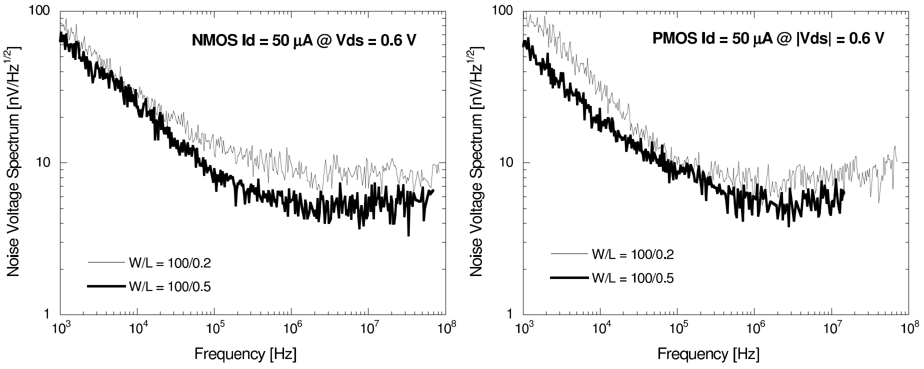
(a) (b)

Figure 2.11: Noise voltage spectra of an NMOS (a) and a PMOS (b) with W/L=100/0.5 at $|V_{DS}|$=0.6 V for different drain current.



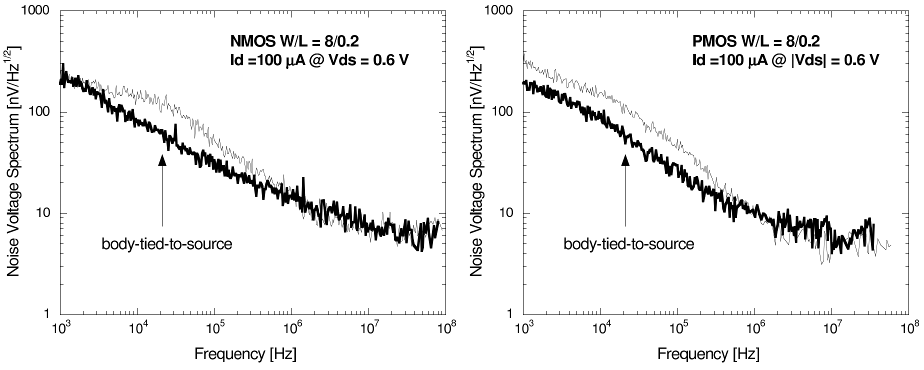
(a) (b)

Figure 2.12: Noise voltage spectra of an NMOS (a) and a PMOS (b) with W/L=100/0.5 at $|V_{DS}|$=1.2 V for two different bias conditions.



(a) (b)

Figure 2.13: Noise voltage spectra of two NMOS (a) and PMOS (b) with W=100 µm and different L (I_D=50 µA @ V_{DS}=0.6 V).



(a) (b)

Figure 2.14: Noise voltage spectra of two NMOS (a) and PMOS (b) with W/L=8/0.2 with and without body tied to source, for a drain current of 100 µA.

2.5 Analysis of static measurement results

Two interesting features related to SOI devices behavior have been highlighted by the static measurements: kink effect and drain induced barrier lowering (DIBL). It is well-know that SOI nMOSFETs exhibit floating body effect [24], which may result in typical kinks in the drain current I_D versus drain-to-source voltage V_{DS} characteristics. The kink effect is due to impact ionization arising from channel current carriers accelerated in the high-field depletion region, where they create electron-hole pairs.

The generated holes may accumulate in the device body, increasing its potential and decreasing the threshold voltage, as discussed in section 2.1. As a result, I_D shows a steep increase when V_{DS} exceeds a value needed for a significant impact ionization to occur. As shown in Fig. 2.15, the FD-SOI NMOS transistors clearly exhibit a kink effect in the I_D-V_{DS} characteristic, which points out that they do not operate in an ideal full depletion mode. In order to reduce the kink effect, a method is to provide an effective body contact for the device, but this will increase the area of the circuit losing the feature of high device density and small parasitic capacitance. For the investigated devices, the presence of a single body contact does not appear to be effective in suppressing these effects. This can be explained by the high resistance of the thin body region [25]. As expected, kink effect is less evident in pMOSFETs, since holes are less effective than electrons in giving impact ionization.

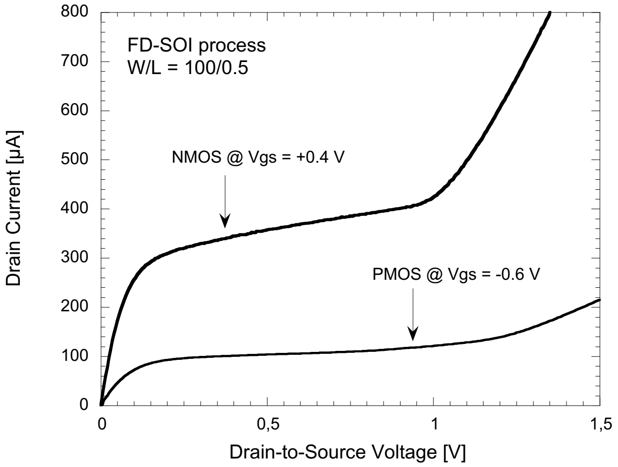


Figure 2.15: Drain current as a function of the drain-to-source voltage for FD-SOI MOSFETs with W/L = 100/0.5.

Kink effects worsen the differential drain conductance of the device as shown in Fig. 2.16 and affect the performance of analog circuits. For an amplifier, the gain at low frequency is substantially degraded with the kink effect.

Drain conductance can be affected also by the drain-induced barrier lowering (DIBL), especially for short-channel devices operating at low levels of inversion. DIBL effects is represented by a decrease in the threshold voltage, V_T, while increasing the V_{DS}. As V_T decreases, the overdrive voltage $V_{OV} = V_{GS} - V_T$ increases, increasing the drain current. This effect can be explained by either the barrier-lowering or charge-sharing concept [26]. In the barrier-lowering concept, decreasing the channel length places the drain and the source closer together resulting in a deeper depletion region under the channel. This depletion region is further deepened with increasing V_{DS} as the depletion region around the drain increases. This effect results in less substrate control, lower depletion capacitance, and increased silicon surface potential. Increasing V_{DS} then lowers the potential barrier, attracts more carriers to the channel and, correspondingly, increases the drain current. In the charge-sharing concept, channel operation is influenced by the source and the drain, in addition to the gate and the substrate. This is especially true for short-channel devices where the drain is close to the channel and acts as a secondary gate generating field lines terminating on the channel. Increasing V_{DS} then enhances the channel beyond its regular gate and substrate control and increases the drain current. Both the barrier-lowering and charge-sharing

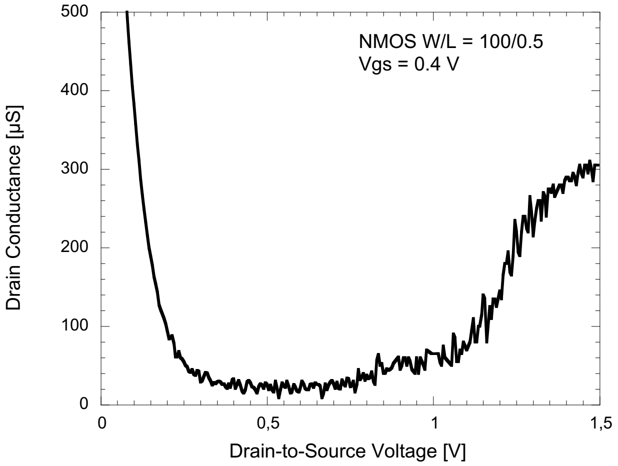


Figure 2.16: Drain conductance as a function of the drain-to-source voltage for a FD-SOI nMOSFETs with W/L = 100/0.5.

concepts describe increasing channel enhancement and drain current with increasing V_{DS} for a fixed value of V_{GS}.
The DIBL effect becomes obvious when looking at the transfer curves of a MOS transistor biased at different V_{DS}, in a semi-logarithmic plot as in Fig. 2.5 pointing out that investigated NMOS devices with small gate length are significantly affected by DIBL.
Without DIBL the curves would coincide in the subthreshold regime, otherwise a shift between curves can be observed in that operating region. With reference to Fig. 2.17, the DIBL effect can be measured by the lateral shift of the transfer curves in the subthreshold regime divided by the drain voltage difference of the two curves:

$$DIBL = \frac{\Delta V_T}{\Delta V_{DS}} \tag{2.2}$$

For the considered device, a NMOS with W/L=100/0.2, DIBL turns out to

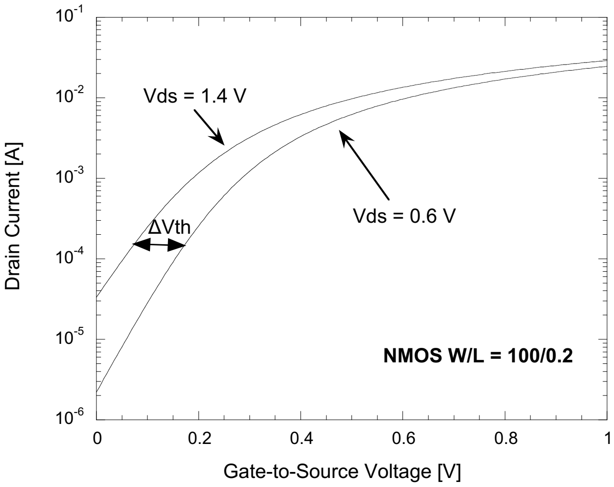


Figure 2.17: Drain current as a function of the drain-to-source voltage for FD-SOI MOSFETs with W/L = 100/0.5.

be about -125 mv/V. DIBL effects significantly increase the drain conductance g_{ds}: the change in drain current ΔI_D, resulting from a change in V_T, ΔV_T, due to DIBL is given by:

$$\Delta I_D = g_m(-\Delta V_T); \tag{2.3}$$

V_T acts as a negative, small-signal, gate to source voltage since its increase lowers the drain current. Dividing Equation (2.3) by ΔV_{DS} that causes ΔV_T

and ΔI_D gives:

$$g_{ds} = \frac{\Delta I_D}{\Delta V_{DS}} = \frac{g_m(-\Delta V_T)}{\Delta V_{DS}} \tag{2.4}$$

where $\Delta V_T / \Delta V_{DS}$ is the DIBL term which expresses the change in the threshold voltage with respect to V_{DS}. Since this term is negative, g_{ds} due to DIBL is positive, consistent with the increase in drain current caused by increasing V_{DS}.

2.6 Analysis of noise measurement results

In a MOSFET, the noise in the channel current can be expressed by an equivalent noise voltage source referred to the gate of the device. A rather general expression for the power spectral density of the voltage noise is:

$$S_V^2(f) = S_W^2 + S_{1/f}^2(f) + S_L^2(f). \tag{2.5}$$

2.6.1 White noise

In (2.5), the first term is given by channel thermal noise, and by other thermal noise sources associated to device parasitic resistors, such as the source resistance $R_{SS'}$, the gate resistance $R_{GG'}$ and the body resistance $R_{BB'}$. Therefore, the expression for the white noise spectral density can be approximated in the following way:

$$S_W^2 = 4k_BT\left[\frac{\Gamma}{g_m} + R_{SS'} + R_{GG'} + R_{BB'}\frac{g_{mb}^2}{g_m^2}\right] \tag{2.6}$$

where Γ is a coefficient whose value depends on the inversion region where the device is operated, g_m is the gate transconductance and g_{mb} is the bulk transconductance. Equation (2.6) can be also written as follows:

$$S_W^2 = 4k_BT\frac{\Gamma}{g_m} + 4k_BTR_{par} \tag{2.7}$$

where the first term in (2.7) is given by channel thermal noise while R_{par} includes all contributions from parasitic resistors. In the low current density operating region the contribution from parasitic resistors usually has a minor impact. Therefore, according to equations (2.6) and (2.7), the white component of the noise voltage spectrum decreases with the increase of the drain current since the transconductance correspondingly increases. Fig. 2.18 shows the measured values of white noise voltage spectrum as a function of the

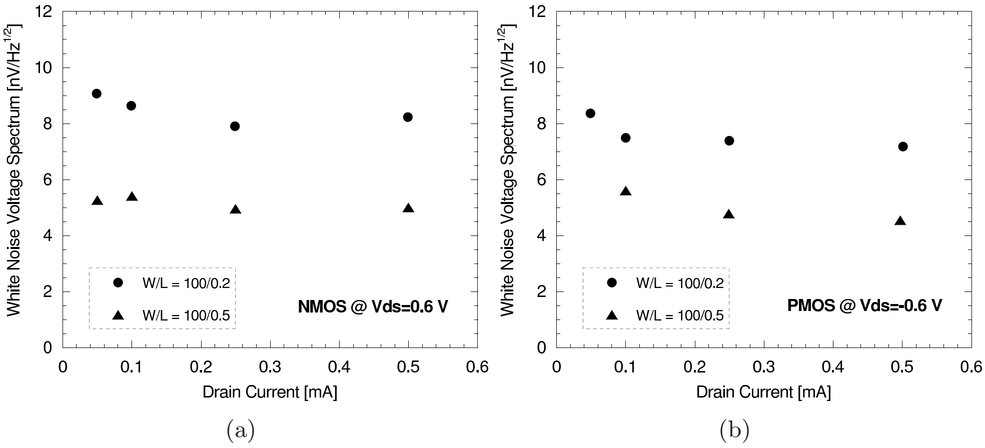


Figure 2.18: White noise voltage spectrum as a function of the drain current I_D for N-channel (a) and P-channel (b) devices with W=100 µm ($|V_{DS}|$=0.6 V).

drain current for N-channel and P-channel devices. From these plots a sizable contribution from R_{par} is noticed. In particular, parasitic body resistor $R_{BB'}$ generates drain current fluctuations through the body transconductance g_{mb}. Since the ratio g_{mb}/g_m is independent of the drain current [26], as shown in Fig. 2.19, this contribution is not expected to change at different I_D and results to be the dominant term in white noise in the I_D range investigated in the performed measurements, as shown in Fig. 2.11 (a).

This excess noise term increases as the gate length decreases, as shown in Fig. 2.13 (a), because $R_{BB'}$ also increases. For channel thermal noise to be dominant with respect to the $R_{BB'}$ contribution, the device has to be operated at very low current density. At smaller drain current, indeed, g_m decreases and channel thermal noise gives a larger contribution to the total white noise voltage spectrum. Fig. 2.20 shows the noise voltage spectra of an FD-SOI NMOS measured at I_D = 50 µA and I_D = 10 µA, together with the channel thermal noise values calculated according to the first term in (2.6) and to g_m values measured at the same drain current.

According to equation (2.7), plotting the measured values of white noise voltage spectrum as a function of $1/g_m$ it is possible to extract the experimental values of Γ and R_{par} from the resulting straight line as shown in Fig. 2.21. In the same plots theoretical values of S_W^2 obtained with Γ=2/3 and R_{par}=0 are also shown. The experimental values extracted for these parameters are

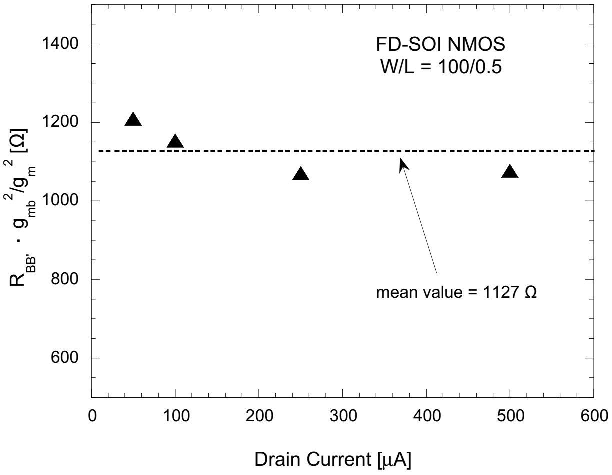
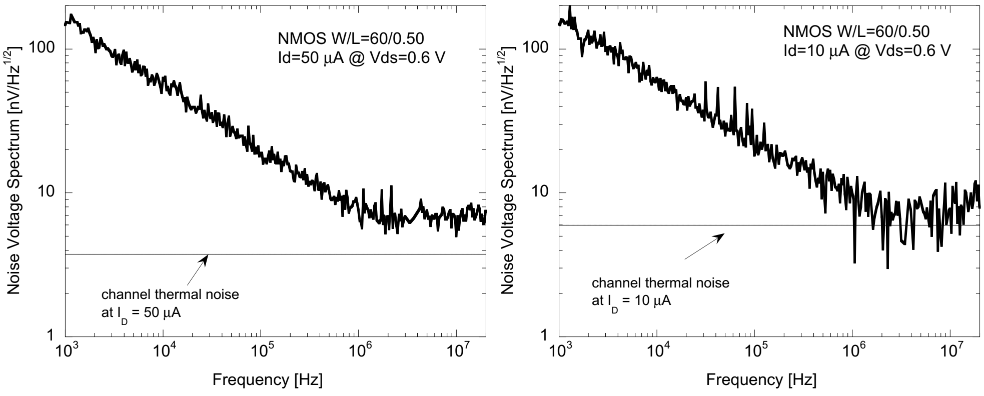


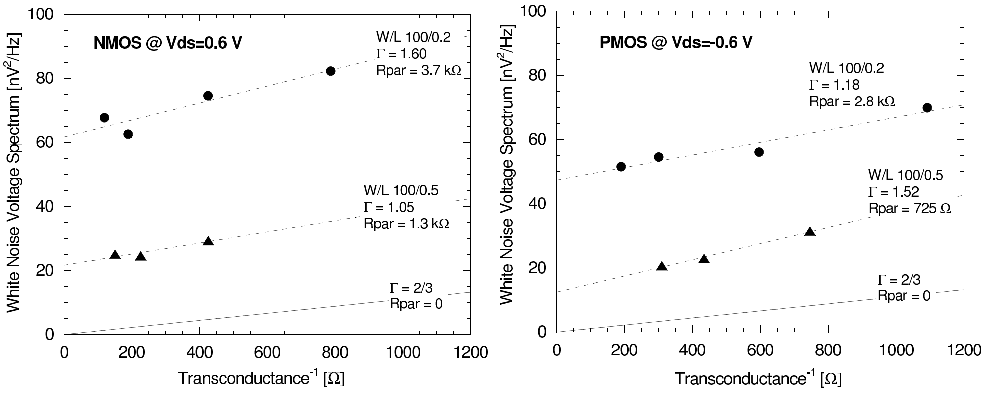
Figure 2.19: $R_{BB'}g_{mb}^2/g_m^2$ as a function of the drain current, for a NMOS with W/L = 100/0.5



(a) (b)

Figure 2.20: Noise voltage spectra of an FD-SOI NMOSFET with W/L = 60/0.5 at I_D = 50 µA (a) and I_D = 10 µA (b), and channel thermal noise value calculated at the same drain current.

summarized in Table 2.2. From these data, it is possible to notice how the parasitic resistance $R_{BB'}$ increases as the gate length decreases, accordingly with the behavior highlighted in Fig. 2.13.



(a) (b)

Figure 2.21: White noise voltage spectrum as a function of $1/g_m$ for N-channel (a) and P-channel (b) devices with W=100 µm (V_{DS}=0.6 V).

Polarity	L [µm]	Γ	R_{par} [Ω]
NMOS	0.2	1.60±0.40	3700±200
	0.5	1.05±0.40	1300±100
PMOS	0.2	1.18±0.20	2800±100
	0.5	1.52±0.10	725±70

Table 2.2: White noise parameters Γ and R_{par}.

2.6.2 1/f noise

The second term in (2.5) describes the 1/f noise in the channel current. In weak inversion and in saturation, the following relationship can be used to model this noise contribution:

$$S^2_{1/f}(f) = \frac{K_F}{C_{OX}WL}\frac{1}{f^{\alpha_F}} \tag{2.8}$$

where K_F is an intrinsic process parameter for 1/f noise, C_{OX} is the gate oxide capacitance per unit area whose value is about 8.8 fF/µm^2 for 4 nm of corresponding electrical gate oxide thickness. The exponent α_F determines the slope of this low frequency noise term. Values of coefficients K_F and α_F have been extracted for both N-channel and P-channel devices at $|V_{DS}|$=0.6 V and at different drain current values. Coefficient α_F does not exhibit any clear dependence on the drain current while a slight variation with the channel length is observed. Due to the small set of investigated devices it is not possible to establish a clear relationship between α_F and L.

As shown in Fig. 2.22 coefficient K_F is variable with channel length while

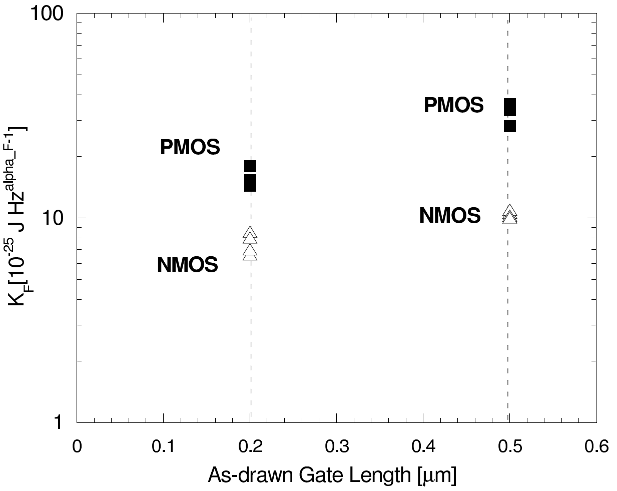


Figure 2.22: 1/f noise coefficient K_F as a function of the gate length L for N-channel and P-channel devices with W=100 µm ($|V_{DS}|$=0.6 V).

it does not exhibit any clear dependence on the drain current. Typical values of 1/f noise coefficients obtained for devices of both polarities are reported in Table 2.3. K_F values have been extracted by using the mean value of α_F shown in the table. For comparison purposes, the table reports data relevant to devices in a 0.18 µm bulk CMOS process from different foundries [27]. According to Table 2.3, K_F and α_F values for SOI NMOSFETs are very close to those featured by bulk NMOS devices. For SOI PMOSFETs, K_F coefficient is instead larger than for bulk counterparts, highlighting the worse 1/f noise properties of the examined process for this device polarity. In bulk CMOS, the fact that PMOSFETs feature a smaller low-frequency noise than equally sized NMOSFETs was generally related to buried channel conduction [28], In SOI CMOS, the active silicon film is so thin that the conduction takes place close to the surface and to oxide traps also in P-type devices. This might be the reason why the magnitude of their 1/f noise is higher than the bulk counterparts.

Foundry	Polarity	α_F	K_F [10^{-25}JHz$^{\alpha_F-1}$]
MITLL	NMOS	0.9	10
	PMOS	1.0	15
TSMC	NMOS	0.9	10
	PMOS	1.0	5
STM	NMOS	0.9	15
	PMOS	1.1	8

Table 2.3: 1/f noise coefficients K_F and α_F from the investigated technology and from devices belonging to bulk CMOS processes in the 0.18 µm node.

2.6.3 Lorentzian-like noise

At low frequency, in devices biased with V_{DS}>0.8 V for the investigated frequency range, the noise voltage spectrum is observed to deviate from the 1/f$^{\alpha_F}$ behavior, showing a much higher slope and sometimes exhibiting a plateau. This can be due to the existence of kink-related lorentzian-like noise overshoot superimposed on 1/f noise. The noise overshoot spectrum can be expressed as:

$$S_L^2(f) = \frac{K_L(V_{DS})}{1 + [f/f_L(V_{DS})]^2} \tag{2.9}$$

where f_L is the corner frequency and K_L is the noise level of the plateau. It is important to note that the corner frequency is a function of V_{DS}, which shift toward higher frequencies as V_{DS} increases, while the noise level plateau decreases as V_{DS} increases. Values of these coefficients as a function of V_{DS} are shown in Fig. 2.23 . The same values are summarized in Table 2.4.

W [µm]	L [µm]	V_{DS} [V]	K_L [nV^2]	f_L [kHz]
60	0.5	1.2	1.3·10^5	5.6
		1.4	2.0·10^4	130
100	0.5	1.2	1.4·10^5	21
		1.4	1.1·10^4	390

Table 2.4: Lorentzian-like noise coefficient K_L and f_L for N-channel devices.

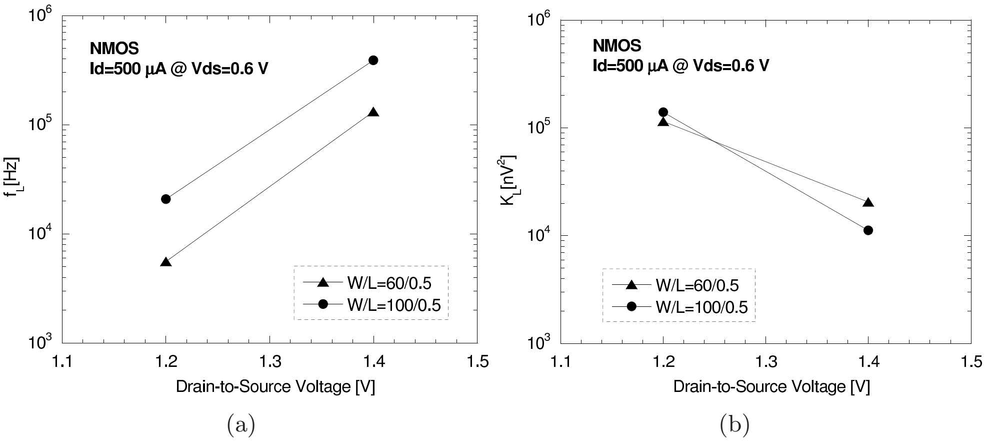


Figure 2.23: Lorentzian-like noise coefficients f_L (a) and K_L (b) as a function of the drain-to-source voltage for N-channel devices.

2.7 Conclusions

This chapter presented the static, signal and noise characterization carried out on devices from a 180 nm FD-SOI CMOS process provided by MIT Lincoln Laboratory. The results point out that SOI MOSFETs noise performance is very close to that featured by bulk devices, especially for low drain currents that is the operating region of major concern for low power applications. A prototype 3D chip designed by Fermilab group involving the investigated SOI technology is the so called VIP chip [20], a three-layer, 20 μm-pitch pixel proposed for ILC applications. The tested chip exhibited some problems, such as trapped charge between layers during the fabrication, causing shift in transistor voltage thresholds. Moreover, the tests carried out on the VIP chip also point out that precision circuits such as current mirrors are hard to be designed in SOI due to possible trapped charge effects and local heating. To avoid difficulties in using a non commercial foundry such as MIT Lincoln Laboratory, our interest has shifted toward looking for commercial vendors for 3D, as Tezzaron Semiconductor, which fabricates wafers in a well established, high yield process.

Chapter 3

The sparsified data readout (SDR1) chip

As discussed in the introduction, monolithic Active Pixel Sensors (MAPS) designed in a standard VLSI CMOS technology have recently been proposed as a compact pixel detector for the detection of charged particle in vertex/tracking applications. MAPS are already extensively used in visible light applications: with respect to other competing imaging technologies, MAPS sensors have several potential advantages in terms of low cost, low power, lower noise at higher speed, random access of pixels which allows windowing of regions of interest, ability to integrate several functions on the same chip.
This chapter discusses the properties of MAPS detectors involving a deep n-well (DNW) structure as the collecting element, and describes the main design features of the SDR1 (Sparsified Digital Readout) chip, which represents the first generation of vertically integrated MAPS with advanced readout architecture for high data rate. In particular, SDR1 is a general purpose prototype of a 3D device, being suitable for diverse applications such as imaging and particle tracking. However, its digital readout architecture was especially designed for vertexing applications to the International Linear Collider (ILC) facility: thus, in order to better understand the operating principles of the chip, the first section will be devoted to discuss the ILC specifications. The second section will introduce the general features of DNW monolithic active pixel sensors. The subsequent sections review the design criteria of the SDR1 chip and presents and discusses the operation and the expected performance of the sensor and the readout electronics. Finally, a set of results of analog and digital simulations, relevant to the analog front-end, the digital front-end and the digital back-end, will be presented.

3.1 Specifications for ILC

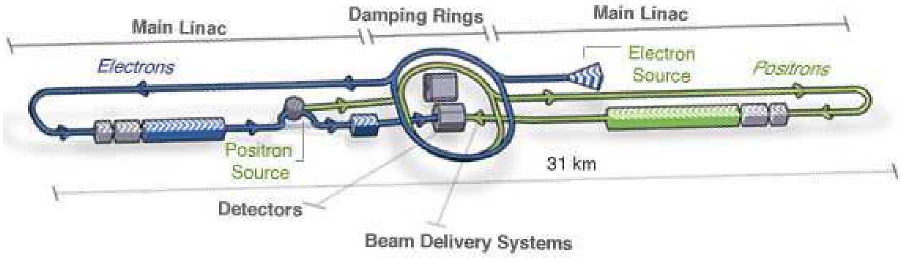


Figure 3.1: Schematic representation of the International Linear Collider

The International Linear Collider is one of the particle accelerators that is being proposed for the investigation of fundamental physics laws [29].

The ILC will accelerate elementary particles (electrons and positrons) along a straight path in beams focussed to a few nm in height and a few hundreds of nm in width and it is planned to have a collision energy of 500 GeV. In order to reconstruct particles trajectories, vertex detector will consists of five concentric cylinders [30] enclosing the beam interaction point, as shown in Fig. 3.2. The radius and length for each of the layers are shown in table 3.1 and result in a total vertex detector area of about 170000 mm^2.

For the design of the detector and the readout electronics, it is necessary to

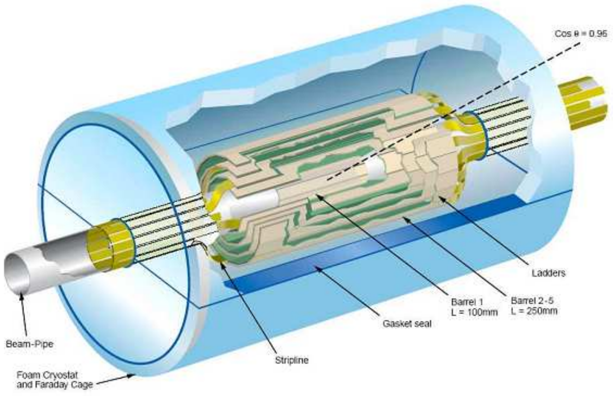


Figure 3.2: Schematic view of the vertex detector at the International Linear Collider.

Layer	Radius [mm]	Half-length [mm]
1	16	50
2	26	75
3	37	75
4	48	75
5	60	75

Table 3.1: Radius and half-length for the five layers of one of the design of the ILC vertex detector.

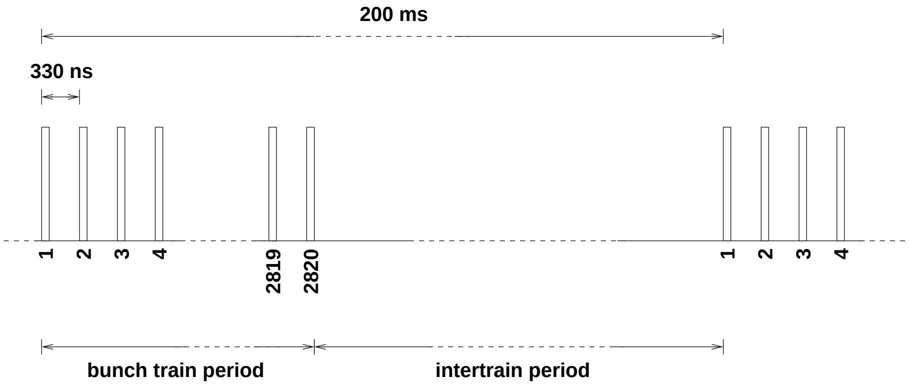


Figure 3.3: Schematic representation of the ILC beam structure.

take into account the specifications for the ILC vertex detector. In particular, the beam structure of ILC will feature 2820 bunch crossing per train, as shown in Fig. 3.3. The bunch train period lasts 1 ms, with an interbunch period of 330 ns and a repetition rate of 5 Hz.

Physics simulations show that for a linear e^+e^- collider operated at 500 GeV, a maximum hit occupancy of 0.03 particles/crossing/mm^2 can be considered as a reasonable assumption for the innermost layer of the detector [45]. If charge spreading in the sensor bulk and oblique trajectories with respect to the detector plane are accounted for, 3 pixels (considering a pixel size greater than 10μm x 10μm) can be expected to fire for every particle hitting the detector. Therefore

$$\begin{aligned} Hit\ rate &= 0.03\ particles/bco/mm^2\ \times\ 3\ hits/particle\ \times\ 2820\ bco/train = \\ &\simeq 250\ hits/train/mm^2. \end{aligned}$$

Detector pitch [μm]	**Resolution [μm]**	**Detection efficiency**
5	1.44	99.998%
10	2.89	99.968%
15	4.33	99.843%
20	5.77	99.518%
25	7.21	98.867%
30	8.66	97.756%

Table 3.2: Spatial resolution and detection efficiency for different sensor pitch values.

If a purely binary readout approach is adopted, the position resolution is (pixel pitch)/($\sqrt{12}$). This means that a position resolution better than 5 μm requires a square pixel with a pitch smaller than $5\sqrt{12}$ μm $\simeq$ 17.3 μm [46]. If a 15 μm pitch (both in X and Y directions) is assumed, then the occupancy O_c for a single cell is

$$O_c = 250\ hits/train/mm^2\ \times\ 15\ \mu m\ \times\ 15\ \mu m \simeq 0.056\ hits/train.$$

Let us assume a Poisson distribution for the number of times an elementary cell is hit during a bunch train period. The probability $P(n)$ of an elementary cell being hit exactly n times in a bunch train is then given by

$$P(n) = \frac{O_c^n}{n!} \cdot \exp(-O_c). \tag{3.1}$$

Therefore, the probability of an elementary cell being hit at least twice during a bunch train is

$$\begin{aligned} P(n \geq 2) &= \sum_{i=2}^{+\infty} P(i) = \sum_{i=2}^{+\infty} \frac{O_c^i}{i!} \cdot \exp(-O_c) = \exp(-O_c) \cdot \sum_{i=2}^{+\infty} \frac{O_c^i}{i!} = \\ &= \exp(-O_c) \cdot [\exp(O_c) - 1 - O_c] \qquad (3.2) \\ &= 1 - \exp(-O_c) - O_c \cdot \exp(-O_c). \qquad (3.3) \end{aligned}$$

In the above considered case (0.03 particle/bco/mm^2 hit rate, 15 μm pitch sensor, yielding $O_c = 0.056$ hits/train), the probability of a cell being hit at least twice during a bunch train is about 0.0016, which means that a pipeline with a depth of one (that is a pipeline with single-hit storing capability) would

Hit rate $[\frac{particles}{bco \cdot mm^2}]$	**Detector pitch [µm]**					
	5	**10**	**15**	**20**	**25**	**30**
0.03	99.99%	99.97%	99.84%	99.52%	98.87%	97.76%
0.05	99.99%	99.91%	99.59%	98.76%	97.14%	94.50%
0.06	99.99%	99.87%	99.39%	98.20%	95.91%	92.26%
0.09	99.98%	99.72%	98.69%	96.20%	91.70%	84.93%
0.12	99.97%	99.52%	97.76%	93.68%	86.66%	76.75%
0.15	99.95%	99.26%	96.62%	90.75%	81.13%	68.36%

Table 3.3: Detection efficiency for different sensor pitch and hit rate values achievable by means of a pipeline with a depth of one.

Hit rate $[\frac{particles}{bco \cdot mm^2}]$	**Detector pitch [µm]**					
	5	**10**	**15**	**20**	**25**	**30**
0.03	100%	100%	99.99%	99.98%	99.94%	99.83%
0.05	100%	100%	99.99%	99.93%	99.75%	99.31%
0.06	100%	100%	99.98%	99.88%	99.58%	98.87%
0.09	100%	99.99%	99.93%	99.62%	98.74%	96.76%
0.12	100%	99.98%	99.83%	99.17%	97.33%	93.49%
0.15	99.99%	99.97%	99.69%	98.50%	95.35%	89.18%

Table 3.4: Detection efficiency for different sensor pitch and hit rate values achievable by means of a pipeline with a depth of two.

be sufficient to record 99.84% of the events without any ambiguity. Detection efficiency will of course change with the elementary cell pitch, as shown in table 3.2. The changes in detection efficiency due to changes both in the hit rate and in the detector pitch are shown in table 3.3. Exploiting the same approach, for a pipeline with a depth of two, the probability of an elementary cell being hit at least three times during a bunch train is given by

$$P(n \geq 3) = 1 - \exp(-O_c) - O_c \cdot \exp(-O_c) - \frac{O_c^2}{2} \cdot \exp(-O_c). \qquad (3.4)$$

Changes in detection efficiency due to changes both in the hit rate and in the detector pitch are summarized in table 3.4.

3.2 Deep n-well MAPS sensors

An innovative solution of MAPS, the so called deep n-well MAPS sensor, was proposed a few years ago in order to deal with the large amount of data produced in the readout of large matrices of pixels [31]. In standard CMOS MAPS, as discussed in section 1.3, use of PMOS devices in the design of the front-end electronics is avoided as the n-wells they are integrated in act as competitive electrodes with respect to the main collecting diode. The lack of complementary devices represents a significant limitation to the design of stages with satisfactory properties, thereby restricting the set of available readout solutions. This issue is addressed by the solution represented by the DNW MAPS. The DNW MAPS is based on the same sensing principle as standard MAPS,

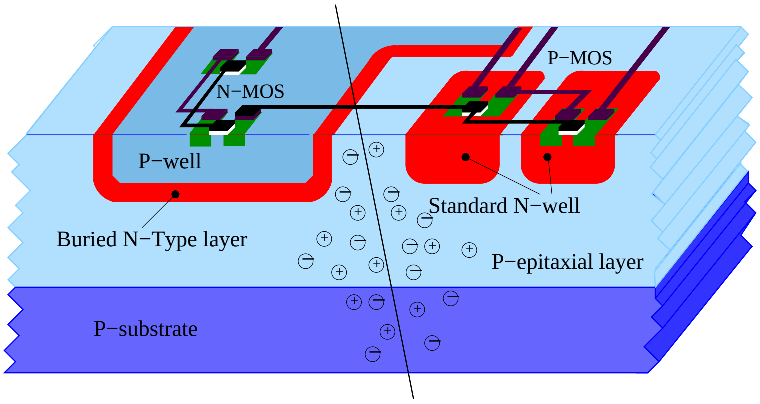


Figure 3.4: Simplified structure of a DNW MAPS. N-channel devices belonging to the analog section may be integrated inside the sensor, whereas the other transistors cover the remaining area of the pixel, with P-channel devices integrated in standard n-wells.

where minority carriers generated in a p-type, lightly doped epitaxial layer by the interaction between silicon and radiation, diffuse and are then collected by n-type electrodes. In a DNW MAPS sensor, whose structure is shown in Fig. 3.4 , an n-well with a deep junction acts as the collecting element for the charge released in the substrate. The deep n-well, which in modern, triple-well CMOS processes is used to shield NMOS devices from substrate coupled noise in mixed signal circuits, may host n-channel devices, thus relaxing the constraints set by the readout circuits on the sensor area and geometry. Moreover, designers may realize more complex readout circuits, taking advantage of fully CMOS architectures, laying out large area DNW sensors. In this way, the effects of charge collection from standard PMOS n-wells, which acts as competitive electrodes against the main collecting electrode, may be significantly limited.

As discussed earlier, the use of large area electrodes is strongly discouraged in 3T-MAPS design: indeed, increased capacitance would unacceptably degrade the noise figure and the charge sensitivity at the same time. In DNW MAPS, a readout scheme employing a charge sensitive amplifier decouples the charge sensitivity from the sensor capacitance and, therefore, from its area. The gain is set by the feedback capacitance C_F of the front-end amplifier as shown by the schematic in Fig. 3.5 which displays the typical chain for the readout of capacitive detectors.

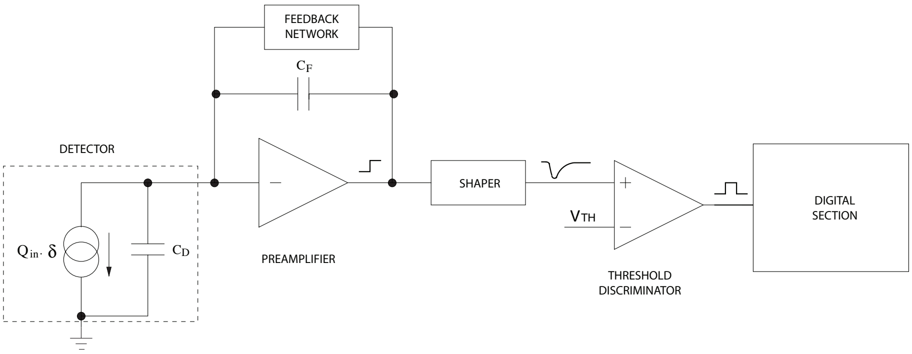


Figure 3.5: Typical chain for capacitive detectors, with binary readout.

The readout chain included in each cell belonging to the pixel matrix consists of a charge preamplifier (PA), a shaping stage, a threshold discriminator followed by the the digital blocks needed for the readout of the matrix. The collected charge Q_{IN} is integrated in the preamplifier feedback capacitance C_F, resulting in a voltage step at the preamplifier output. This signal is then further amplified and filtered by the shaper, which improves the performance of the analog processor in terms of signal-to-noise ratio. The shaper output signal is then further processed by the threshold discriminator and by the digital blocks of the pixel cell, which may include sparsification blocks for the readout of hit pixel only (sparsified readout).

In table 3.5, a comparison between the main features of standard MAPS and DNW-MAPS is given.

	Standard MAPS	**DNW-MAPS**
Technology process	CMOS Standard	CMOS Standard
In-pixel basic front-end	3T	PA
Transistor polarity	Only NMOS	NMOS and PMOS
Transistors per cell	∼10	>100
Foremost noise source	Reset transistor	PA input device
Minimum pixel pitch	∼10 µm	∼20 µm (130 nm CMOS)
Charge collecting element	N-well/Sub	Deep N-Well/Sub
$\frac{\textit{Collecting element area}}{\textit{Total pixel area}}$	≪0.1	>0.1
Collecting element parasitic capacitance	<10 fF	∼100 fF
Readout architecture	Sequential	Sparsified

Table 3.5: Comparison between Standard MAPS and DNW-MAPS.

The so called APSEL series is the first generation of DNW MAPS with on-pixel data sparsification and time stamping, successfully tested in a beam for the first time in September 2008 [32]. Another prototype chip, namely SDR0 [33], is a DNW MAPS featuring a different readout architecture with respect to the APSEL. In particular, this readout architecture has been used also in the SDR1 chip, described in the next section.

3.3 The SDR1 chip

The features of the Tezzaron/Chartered process have been exploited in the design of the SDR1 chip, which inherits the readout architecture as the SDR0 monolithic sensor, presented in [33], whose characteristics will be often taken into account in this work for comparison purpose. SDR1 features two vertically integrated layers each fabricated in a 130 nm CMOS process provided by Chartered Semiconductor, containing the analog and the digital front-end respectively. SDR1 operation is based on the ILC beam structure. It features two different processing phases: a detection phase (corresponding to the bunch train period) and a readout phase (corresponding to the intertrain period). SDR1 consists of a 240x256 MAPS matrix with a pixel pitch of 20 µm. The

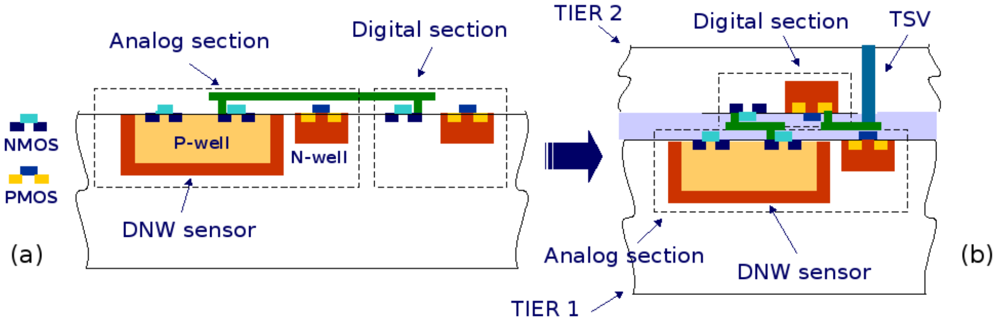


Figure 3.6: Cross-sectional view of a DNW CMOS MAPS from planar CMOS technology (a) to a 3D process (b).

analog tier includes the DNW sensor, a charge sensitive amplifier and the NMOS pair from the threshold discriminator, while the digital tier hosts the digital front-end, formed by two latches for hit storage, sparsification logic, two time stamp registers and kill mask blocks, the digital back-end, with X and Y registers, time stamp line drivers and a serializer and a PMOS pair from the discriminator. Logic blocks are able to keep information about two hits during each single bunch train with the relevant time stamps, thus providing a high detection efficiency. The evolution from the DNW MAPS in a planar CMOS technology to its vertically integrated version is sketched out in Fig. 3.6, showing a cross-sectional view of a 2D MAPS and of its 3D translation.
In the SDR1, the deep n-well sensor and the analog front-end are integrated on a different layer from the digital front-end, thus providing several advantages:

- P-channel devices belonging to the digital blocks are integrated in a different substrate from the sensor, therefore significantly reducing the area covered by standard n-wells (and thus reducing the parasitic charge collection effects), and improving the charge collection efficiency;

- although SDR1 features a 20% smaller pitch with respect to SDR0 monolithic sensor (20 µm for SDR1, 25 µm for SDR0), the available area for the device integration is larger, because of the 3D stack (625 µm^2 in the SDR0 MAPS, 800 µm^2 in the SDR1 chip); this has been exploited to add functionality to the pixel cell (in particular, to implement double-hit stroring capability), therefore providing an increased functional density of the sensor; the larger availability of area has also made it possible to lay out large area transistors in some blocks of the readout chain in order to reduce threshold dispersion;

- from the previous point it should be apparent that a better trade-off between integrated functionality and detector pitch can be achieved, yielding a smaller point resolution;

- stacking the analog processor layer and the digital front-end one on the top of the others, makes it possible to reduce interaction effects between different sections of the pixel cell, such as charge injection into the sensor through parasitic capacitive coupling.

3.4 Readout architecture

As earlier discussed, CMOS pixel sensors are generally used in visible imaging applications, and only in recent years they have been proposed for applications to charged particle tracking. With reference to a matrix of CMOS pixels, in the first case almost all the pixels are involved in photon detection, whereas in the second one few pixel interact with a charged particle passing through the matrix. In imaging applications, in order to collect information from all the pixel in the matrix, a sequential readout turns out to be adequate: this approach has been also employed in tracking applications.

The SDR1 architecture has the potential to fit into already available architectures performing sparsified readout at the pixel cell level, therefore, offering a substantial degree of flexibility in dealing with large flows of data from the detector to the chip periphery. This sparsified approach were proposed in the front-end of the pixel sensor for the BTeV experiment [34], [35]. In contrast to the sequential readout, this approach allows extremely fast non-sequential readout of hit pixels only.

Figure 3.7 shows the block diagram of a 16×16 MAPS matrix which will be used for a first, high level description of the detector operation and of the digital sparsification scheme. The operation of such a sensor has been tailored on the structure of the ILC beam and features two different processing phases:

- a detection (or acquisition) phase, corresponding to the bunch train period,

- a readout phase, corresponding to the intertrain period.

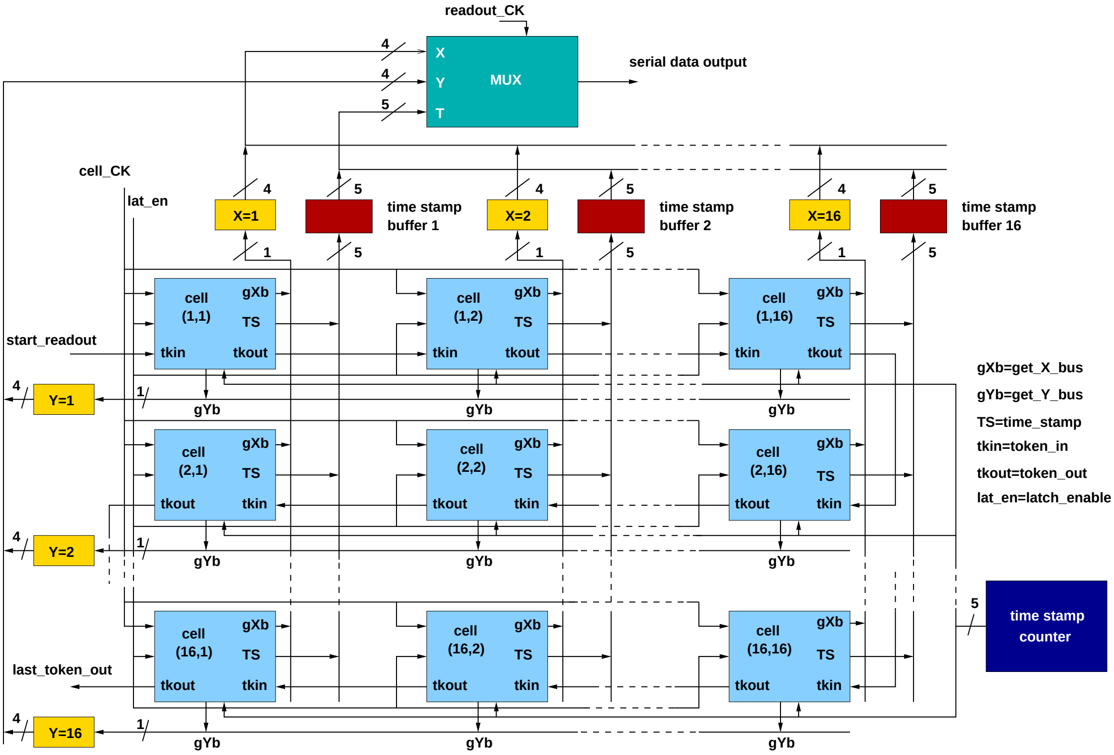


Figure 3.7: Digital readout architecture of the DNW-MAPS sensor with sparsified readout and time stamping capability.

3.4.1 Detection phase

During the detection (or acquisition) phase, the output of a 5-bit Gray code time-stamp counter, located at the periphery of the sensor matrix, is broadcast to all the matrix cells and written in their individual time stamp registers. When a cell is hit, the content of its time stamp register gets frozen, therefore preserving timing information of the hit. The 5-bit time stamp allows the bunch train interval to be split into 32 time slices, providing useful complementary information to other detectors for track reconstruction. When the beam train is over, the detector is made insensitive to any possibly noise-induced, fake hit by pulling down a latch-enable signal. Given the ILC bunch train period, about 930 µs, each time slot amounts to about 29 µs, corresponding to a time-stamp clock frequency of about 34.4 kHz.

3.4.2 Readout phase

At the beginning of the intertrain period, a token is launched through the detector starting from the cell in the first row and first column of the matrix. This is done by setting the `start_readout` signal. The token scans the matrix in a row by row fashion and stops in the first hit cell it finds along its path. Then, at the next rising edge of the cell clock (`cell_CK`), the pixel gets hold of the column and row buses, pointing to the `X` and `Y` registers at the periphery of the sensor matrix, and sends off the time stamp register content. Almost immediately, the token is released and scans ahead, searching for the next hit pixel, which will be readout during the next cell clock period. `X`, `Y` and time stamp data are serialized and sent off the chip by means of a multiplexer within a cell clock period. The readout clock (`readout_CK`) frequency is an integer multiple of the cell clock and is synchronous with it. In order to understand the time scale of the detection phase, let us consider a megapixel sensor with 20 μm pitch, yielding an overall detector area of $(20 \times 10^{-3})^2 \cdot 10^6$ mm^2 = 400 mm^2. Based on the hit rate of 250 $hits/train/mm^2$ derived in section 3.1, the number of hits per megapixel sensor would be about 100000 per train. Assuming a readout clock of 50 MHz and a 30 bit word for each hit pixel (10 bits for the Y coordinate, 10 bit for the X coordinate, 5 time-stamp bits and 5 bits for other pieces of information such as, for instance, chip number), the time required for reading out an entire chip would be $30 \times 100000 \times 20$ ns=60 ms, well below

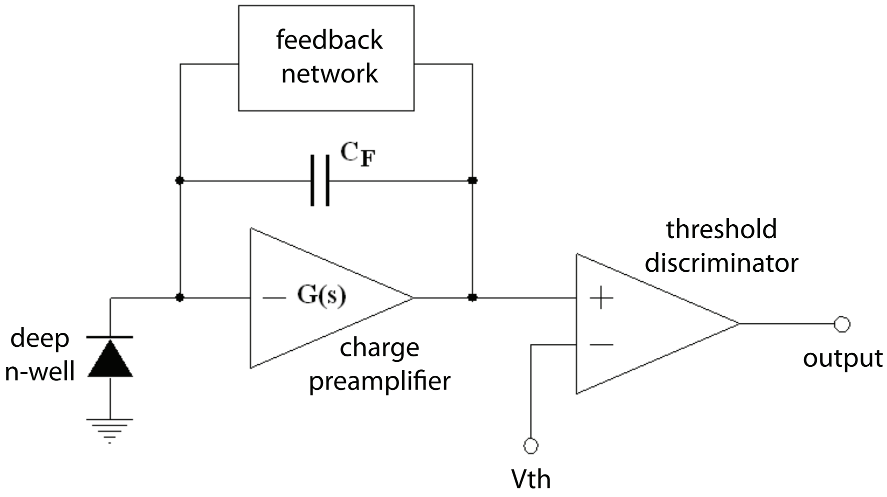


Figure 3.8: The analog front-end

the 199 ms intertrain period featured by the ILC beam structure.

3.5 Analog front-end

In the SDR1 design, the pixel-level front-end processor, whose block diagram is shown in Fig. 3.8, includes a charge sensitive amplifier and a threshold discriminator. This is a modified, shaper-less version of typical channel for capacitive sensors discussed in section 3.2: this reduction in the analog front-end complexity makes it possible to comply with the point resolution constraints set for the ILC vertex detector by reducing the pixel size and pitch. Moreover, the threshold discriminator has been only partially integrated in the bottom tier has pointed out in Fig. 3.9, which shows the block diagram with some transistor-level details of the analog front-end electronics. Large area PMOS devices belonging to the threshold discriminator were laid out on the digital tier of the chip, thus remarkably reducing the competitive n-well areas. Charge restoration in the preamplifier feedback network is obtained through a PMOS current mirror stage, providing a linear discharge of the metal-oxide-metal capacitor C_F (about 1 fF). The following subsection will provide a detailed description of the analog blocks contained in the SDR1 pixel cell analog front-end.

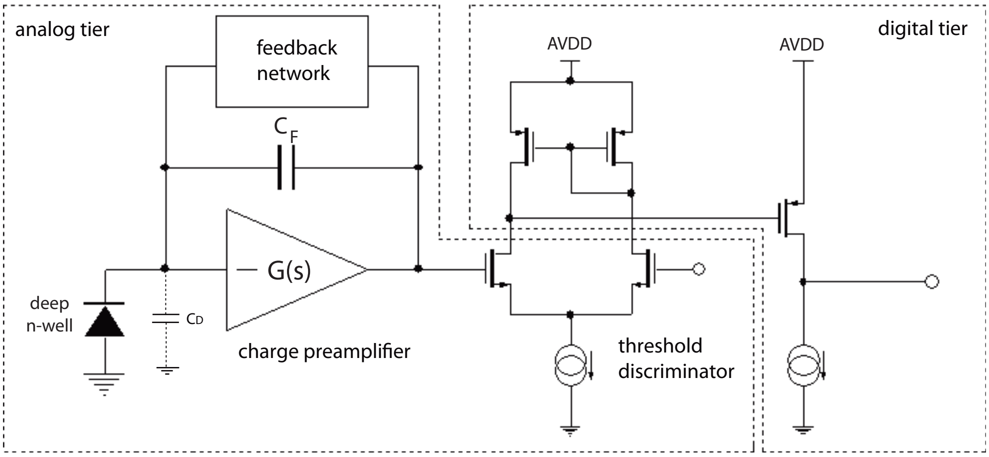


Figure 3.9: Block diagram, with some transistor-level details, of the analog front-end.

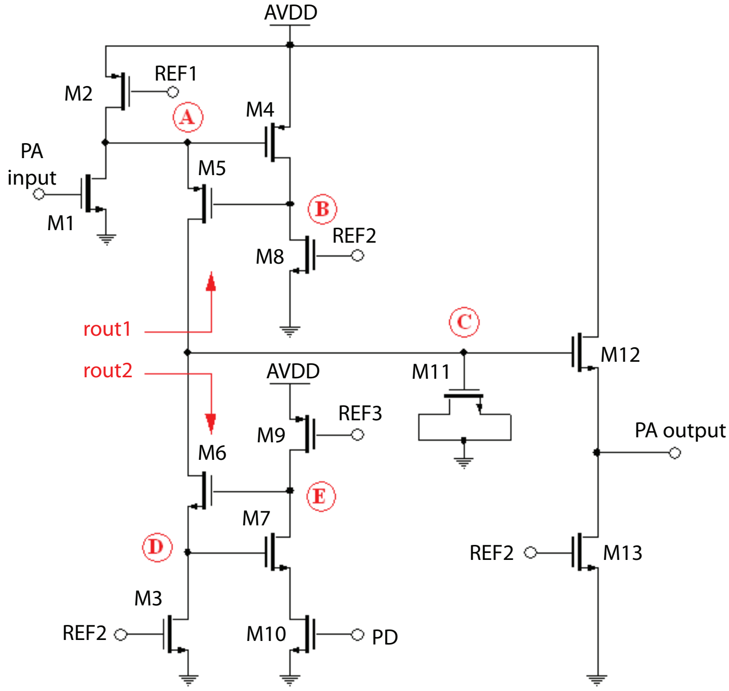


Figure 3.10: Schematic circuit of the charge preamplifier.

3.5.1 Charge preamplifier

Fig. 3.10 illustrates the schematic circuit of the charge preamplifier. The input NMOS device M1, which features a W/L=20/0.18, and M5 implement a folded cascode configuration. The choice of M1 dimensions was dictated by criteria for optimum detection efficiency, as discussed later. In order to increase the impedance seen at the gate of M12, two local feedback networks involving respectively M4-M5 and M6-M7 are used. A MOS capacitor, M11, has also been used with the aim of limiting the bandwidth of the preamplifier and to reduce high frequency noise contributions. The output stage consists of M13, operating in a source follower configuration. All the blocks implemented in the analog front-end involve structures controlled by the PD (Power Down) signal. This signal controls some transistors, operating as switches, which make it possible to switch the analog front-end blocks off during the readout phase, thus minimizing the power consumption of the cell and complying with the severe power constraints set by ILC vertex detector. Actually, this feature turns

Device	Gate W/L	Type
M1	20 μm / 180 nm	standard
M2	800 nm / 2 μm	standard
M3	800 nm / 400 nm	low Vth
M4	400 nm / 150 nm	low Vth
M5	200 nm / 400 nm	low Vth
M6	500 nm / 400 nm	low Vth
M7	500 nm / 150 nm	low Vth
M8	200 nm / 300 nm	low Vth
M9	150 nm / 800 nm	low Vth
M10	300 nm / 150 nm	standard
M11	5 μm / 1.02 μm	standard
M12	400 nm / 180 nm	low Vth
M13	200 nm / 150 nm	low Vth

Table 3.6: Gate dimension and device type (standard or low threshold voltage) of the transistors belonging to the charge preamplifier.

out to be essential in order to operate the chip at low temperature without any cooling systems: these systems are indeed forbidden in the beam interaction point proximity, because of the material budget constraints. Gate geometry of the transistors belonging to the charge preamplifier are summarized in table 3.6, which also indicates the device threshold voltage (low threshold voltage device, or standard threshold voltage device, which are available options in the Chartered process).

Small signal analysis

The study of the parameters such as the open loop gain of the preamplifier and the position of the poles of the circuit is performed through the equivalent small signal circuit shown in Fig. 3.11. All transistors (Mi) are replaced by a voltage-controlled current source, controlled by the voltage between gate and source through the device transconductance g_m, and by the r_{ds} resistor. Since the MOS devices controlled by the PD signal operate as switches, in this analysis they will be neglected, assuming a very small drain to source resistance. Moreover, let us assume a unity gain for the device M12 operating as the source follower stage. Under these assumptions, the simplified transfer function of the circuit can be obtained from the circuit shown in Fig. 3.12.

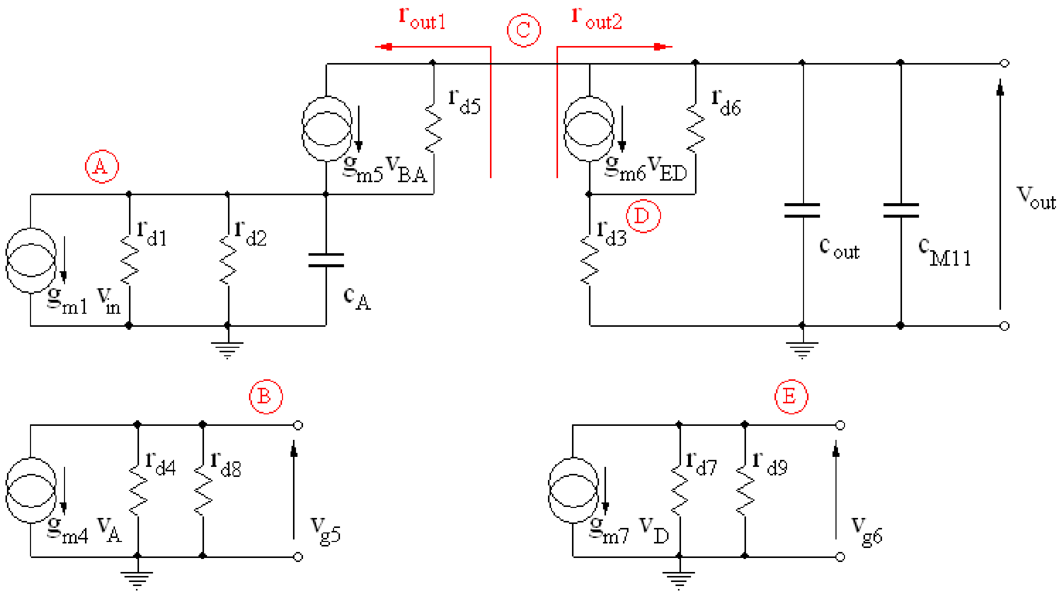


Figure 3.11: Small signal circuit for the preamplifier stage.

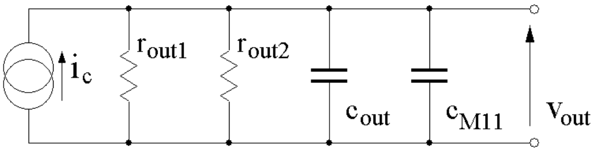


Figure 3.12: Simplified small signal circuit for the preamplifier stage.

The output voltage can be calculated by obtaining the current signal i_c and multiplying by the output impedance:

$$v_{out} = Z_{out}\ i_c, \tag{3.5}$$

where

$$Z_{out} = r_{out1} \parallel r_{out2} \parallel \frac{1}{sC_{out}} \parallel \frac{1}{sC_{M11}}. \tag{3.6}$$

The term Z_{out} includes r_{out1} e r_{out2}, which are defined as follows:

$$r_{out1} = r_d + r_{d5} + g_{m5}r_{d5}r_d(1 + A_1), \tag{3.7}$$

$$r_{out2} = r_{d3} + r_{d6} + g_{m6}r_{d6}r_{d3}(1 + A_2), \tag{3.8}$$

where

$$r_d = r_{d1} \parallel r_{d2}, \tag{3.9}$$

and A1, A2 are the loop gains of the local feedback provided by the pairs M4-M5 and M6-M7; in particular:

$$A_1 = g_{m4}(r_{d4} \parallel r_{d8}). \tag{3.10}$$

$$A_2 = g_{m7}(r_{d7} \parallel r_{d9}). \tag{3.11}$$

The i_c current can be expressed as follows:

$$i_c \simeq -g_{m1}v_{in}. \tag{3.12}$$

The value of r_{out1} is approximately 670 MΩ, while that of r_{out2} is greater than 10 GΩ: therefore, it is possible to simplify the calculation of the transfer function since $r_{out1} \ll r_{out2}$. The expression of the transfer function with a single pole approximation is:

$$\frac{v_{out}}{v_{in}}(s) = -g_{m1}\frac{r_{out1}}{1 + sr_{out1}(C_{out} + C_{M11})}, \tag{3.13}$$

where

$$C_{out} = C_{d5} + C_{d6}, \tag{3.14}$$

In the equation 3.14 C_{d5} and C_{d6} represent the sum of the drain parasitic capacitance of M5 and M6.

Actually, the transfer function has a second pole due to the capacitance at the node C. Taking into account this capacitance, the transfer function of the charge preamplifier can be written as follows:

$$\frac{v_{out}}{v_{in}}(s) = -g_{m1}r_{out1} \cdot \frac{1}{1 + sr_{out1}(C_{out} + C_{M11})} \cdot \frac{1}{1 + sr_pC_p}. \tag{3.15}$$

where r_p is the small signal resistance seen at the node C, and C_p represents the capacitance at the same node (its value is about 50 fF). DC gain and frequency of the two poles f_{P1} and f_{P2} are given by:

$$G_0 = -g_{m1}r_{out1}, \tag{3.16}$$

$$f_{P1} = \frac{1}{2\pi r_{out1}C_{out}}, \tag{3.17}$$

$$f_{P2} = \frac{1}{2\pi r_pC_p}. \tag{3.18}$$

Fig. 3.13 shows the frequency response of the charge preamplifier. The red line has been obtained by simulating the circuit with MOS capacitor M11, while the blue one was obtained by simulating the circuit without this capacitance. The DC gain, obtained both through simulations and calculations, is 89.8 dB and the frequency of the first pole, considering the bandwidth limiting capacitor M11, is 4.1 kHz.

Noise performance analysis

In the preamplifier shown in Fig. 3.10 the main noise source comes from the series noise of the input device. In order to perform an analysis of the preamplifier noise performance, reference will be made to Fig. 3.14, showing a simplified version of the feedback network, where the charge resetting PMOS current mirror has been replaced with an equivalent resistor R_F with the purpose of simplifying circuit analysis in the small signal regime. In Fig. 3.14, e_s is the input referred series noise of the M1 MOSFET, C_i is the preamplifier input capacitance and C_D is the detector capacitance. For the forward stage of the charge preamplifier, a single pole transfer function can be reasonably assumed:

$$G(s) = \frac{G_0}{1 + \frac{s}{\omega_0}} = \frac{G_0}{1 + s\tau}, \tag{3.19}$$

where G_0 and τ are the DC gain and the time constant of the forward stage respectively. The transfer function $T(s)$ between e_s and the preamplifier output can be written as follows:

$$T(s) = \frac{\frac{C_T+C_F}{C_F}}{1 + s\frac{\tau}{G_0}\frac{C_T+C_F}{C_F}}, \tag{3.20}$$

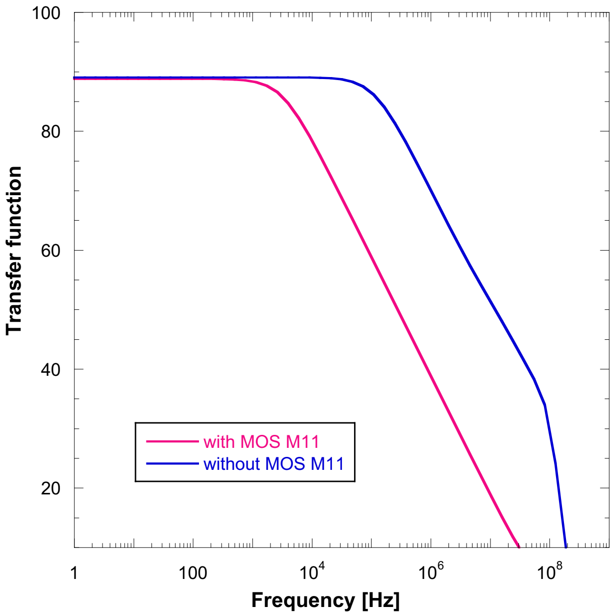


Figure 3.13: Simulated frequency response of the amplifier stage. The red line refers to the circuit with the bandwidth limiting capacitor M11.

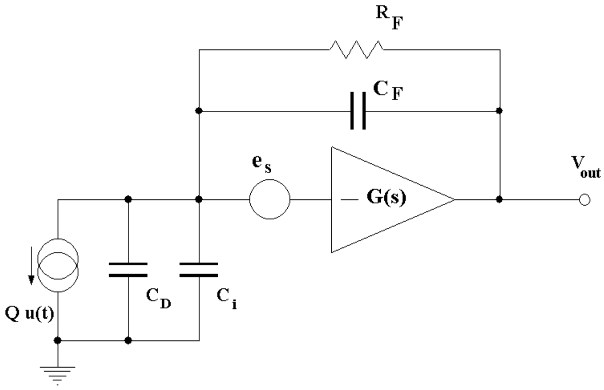


Figure 3.14: Block diagram of the charge preamplifier with noise sources.

assuming that R_F approaches infinity
The mean square value of the noise at the preamplifier output, $\overline{v^2_{N,e_s}}$, due to the input referred series noise is:

$$\begin{aligned}\overline{v^2_{N,e_s}} &= \int_0^{+\infty} S_w \cdot |T(j\omega)|^2 df = \\ &= \int_0^{+\infty} S_w \cdot \frac{(\frac{C_T+C_F}{C_F})^2}{1+(\omega\frac{\tau}{G_0})^2(\frac{C_T+C_F}{C_F})^2} df \\ &= \frac{1}{4} S_w \frac{C_T+C_F}{C_F} GBP, \qquad (3.21)\end{aligned}$$

where

- C_T represents the sum of the preamplifier input capacitance and the detector capacitance:
$$C_T = C_i + C_D; \qquad (3.22)$$
- S_w is the white noise contribution in the input referred voltage noise source:
$$S_w = \frac{4kT\Gamma}{g_{m1}}; \qquad (3.23)$$
- $GBP = G_0/\tau$ is the gain-bandwidth product of the preamplifier forward stage, which can be approximated by means of:
$$GBP = \frac{G_0}{\tau} = (g_{m1}r_{out1})(\frac{1}{r_{out1}C_{out}}) = \frac{g_{m1}}{C_{out}}. \qquad (3.24)$$

The equivalent noise charge, ENC, is defined as follows:

$$ENC = \frac{\sqrt{\overline{v^2_{N,e_s}}}}{v_{out,max}/Q}. \qquad (3.25)$$

where Q is the injected charge and $v_{out,max}$ is the peak value of the preamplifier response.

The transfer function between the delta-shaped current source modeling the detector signal and the preamplifier output is given by

$$\frac{V_{out}}{Q} = \frac{G_0 R_F}{s^2(\tau \cdot C_T R_F) + s[\tau + R_F(C_T + G_0 C_F)] + G_0 + 1} =$$

$$\begin{aligned}
&\simeq \frac{GBP}{C_T}\frac{1}{s^2+s\cdot GBP\cdot\frac{C_F}{C_T}+GBP\cdot\frac{1}{C_TR_F}} = \\
&= \frac{GBP}{C_T}\frac{1}{\left(s+\frac{1}{\tau_1}\right)\left(s+\frac{1}{\tau_2}\right)}, \qquad (3.26)
\end{aligned}$$

where V_{out} is the Laplace transform of the voltage signal at the shaper output v_{out}, C_T is the sum of the capacitances shunting the charge preamplifier input, $GBP = \frac{G_0}{\tau}$ is the gain-bandwidth product of the preamplifier forward stage, $\tau_1 \simeq R_FC_F$ and $\tau_2 \simeq \frac{C_TC_FR_F}{GBP\cdot C_F^2R_F - C_T}$. Using simple fraction expansion for (3.26) yields

$$\frac{V_{out}}{Q} \simeq \frac{1}{C_F}\left(\frac{1}{s+\frac{1}{\tau_1}} - \frac{1}{s+\frac{1}{\tau_2}}\right), \qquad (3.27)$$

where $GBP\cdot C_F^2\cdot R_F \gg 2C_T$ is assumed. The response of the charge preamplifier to a delta-shaped signal with area Q can be easily calculated from (3.27)

$$v_{out}(t) = \frac{Q}{C_F}\left[exp\left(-\frac{t}{\tau_1}\right) - exp\left(-\frac{t}{\tau_2}\right)\right]. \qquad (3.28)$$

The signal v_{out} features a peaking time

$$t_p \simeq \frac{C_T}{GBP\cdot C_F}ln\left(GBP\cdot R_F\frac{C_F^2}{C_T}\right). \qquad (3.29)$$

The peak value is

$$v_{out}(t_p) \simeq \frac{Q}{C_F}\left[\left(\frac{C_T}{GBP\cdot C_F^2\cdot R_F}\right)^{\frac{C_T}{GBP\cdot C_F^2\cdot R_F}} - \frac{C_T}{GBP\cdot C_F^2\cdot R_F}\right]. \qquad (3.30)$$

Note that,

$$\lim_{R_F\to+\infty} v_{out}(t_p) = \lim_{GBP\to+\infty} v_{out}(t_p) = \frac{Q}{C_F}, \qquad (3.31)$$

Substituting equation (3.21) in (3.25) yields:

$$ENC_{e_s,w} = \sqrt{S_wC_F(C_T+C_F)\frac{GBP}{4}}. \qquad (3.32)$$

This expression will be taken into account in the design criteria of the charge preamplifier.

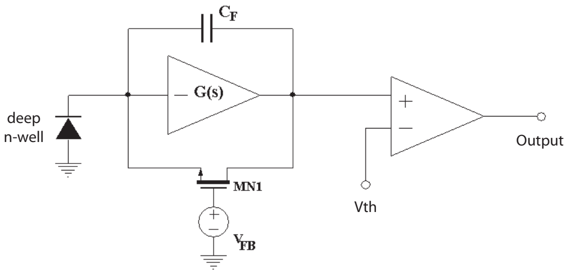


Figure 3.15: Charge restoration obtained by means of an NMOS biased by an external voltage reference V_{FB}.

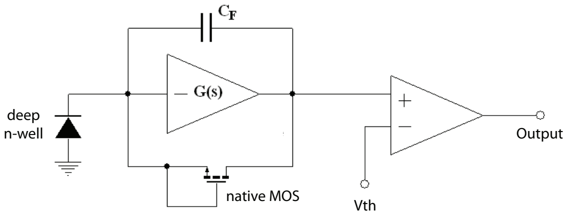


Figure 3.16: Charge restoration with a native NMOS (zero vth).

3.5.2 Reset operation

In order to work properly within the system, the preamplifier requires an appropriate network for charge restoration. The circuits implementing this solution can be chosen according to various criteria. Three possible solutions were studied. The first solution requires the use of a single NMOS device biased with a external voltage V_{FB} (see Fig. 3.15), the second one does not provide an external control and uses a native NMOS transistor with a zero threshold voltage (Fig. 3.16); finally, the third one uses a current mirror formed by PMOS devices (Fig. 3.17). The first solution, already used in other CMOS MAPS applications [36], is little expensive in terms of area, but is very sensitive to process variations and then it has been not exploited. The second solution does not provide a control on the reset operation velocity. The chosen solution provides better performance, but requires a larger area with respect to the previous ones.

The charge reset operation in the preamplifier of Fig. 3.17 can be analyzed by taking into account the large signal model of the MOS transistor in deep

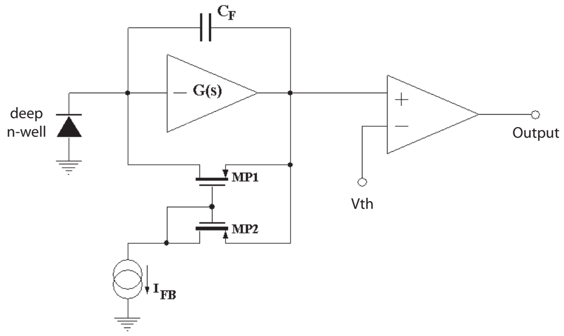


Figure 3.17: Charge restoration network with a current mirror stage.

subthreshold regime, where the current is due mainly to the diffusion current between drain and source. In this operating region, the drain current in a PMOS device can be approximated to the following expression [37]:

$$I_D \simeq I_{D0}\frac{W}{L} \cdot exp\left[\frac{q(V_{SG} - |V_{THp}|)}{nk_BT}\right]\left[1 - exp\left(\frac{-qV_{SD}}{k_BT}\right)\right], \qquad (3.33)$$

where,

$$I_{D0} = 6\mu_{p0}C_{ox}\left(\frac{k_BT}{q}\right)^2. \qquad (3.34)$$

In (3.33), W and L are the transistor channel width and length respectively, q is the elementary charge, V_{SG} is the source-to-gate voltage, V_{THp} is the threshold voltage (which is negative in a p-channel MOSFET), n is the subthreshold slope coefficient, k_B is the Boltzmann's constant, T is the absolute temperature, V_{SD} is the source-to-drain voltage, μ_{p0} is the zero-bulk-bias hole mobility and C_{ox} is the oxide capacitance per unit area. Since, in transistor M1, V_{SG} is kept constant by the current mirror action, also the first exponential term in (3.33) is constant. Therefore, the expression of I_D can be rewritten as follows

$$I_D = I'_{D0}\left[1 - exp\left(\frac{-qV_{SD}}{k_BT}\right)\right], \qquad (3.35)$$

where the constant term is

$$I'_{D0} = I_{D0}\frac{W}{L} \cdot exp\left[\frac{q(V_{SG} - |V_{THp}|)}{nk_BT}\right]. \qquad (3.36)$$

Note that the current mirror in the charge preamplifier feedback network is not active unless the stage has been thrown off its quiescent state by a charge signal. Upon arrival of a charge signal Q (which will be supposed to feature a Dirac delta-like shape) at the preamplifier input, the behavior of the circuit is governed by the following equation,

$$C_F \frac{d\Delta V_{out}}{dt} = -I'_{D0}\left[1 - exp\left(\frac{-q\Delta V_{out}}{k_B T}\right)\right], \tag{3.37}$$

where ΔV_{out} is the signal at the preamplifier output. Equation (3.37) can be solved by means of variable separation methods. If $\Delta V_{out} = 0$ for $t < 0$ and $\Delta V_{out} = Q/C_F$ for $t = 0$, then the following equation holds:

$$\int_{\frac{Q}{C_F}}^{\Delta V_{out}} \frac{du}{1 - exp\left(\frac{-qu}{k_B T}\right)} = -\frac{I'_{D0}}{C_F} t. \tag{3.38}$$

Equation (3.38) yields the following solution,

$$\Delta V_{out} = \frac{k_B T}{q} ln \left\{\left[exp\left(\frac{Q}{C_F}\frac{q}{k_B T}\right) - 1\right] exp\left(-\frac{q}{k_B T}\frac{I'_{D0}}{C_F} \cdot t\right) + 1\right\} \cdot 1(t), \tag{3.39}$$

where $1(t)$ is the step function. It is worth noticing that, for $t \ll \frac{C_F}{I'_{D0}} \frac{k_B T}{q}$

$$exp\left(-\frac{q}{k_B T}\frac{I'_{D0}}{C_F} \cdot t\right) \simeq 1 - \frac{q}{k_B T}\frac{I'_{D0}}{C_F} \cdot t,$$

and, therefore,

$$\begin{aligned} \Delta V_{out} &\simeq \frac{k_B T}{q} ln \left[\left(1 - \frac{q}{k_B T}\frac{I'_{D0}}{C_F} \cdot t\right) \cdot exp\left(\frac{Q}{C_F}\frac{q}{k_B T}\right) - \frac{q}{k_B T}\frac{I'_{D0}}{C_F} \cdot t\right] = \\ &\simeq \frac{Q}{C_F} + \frac{k_B T}{q} ln\left(1 - \frac{q}{k_B T}\frac{I'_{D0}}{C_F} \cdot t\right) \simeq \frac{Q}{C_F} - \frac{I'_{D0}}{C_F} \cdot t. \end{aligned} \tag{3.40}$$

Note that I'_{D0}, to a good extent, should be considered proportional to I_{FB}, which sets the discharge current when the mirror is active.

From (3.39), the reset time Δt_{res}, defined as the time needed in order for the preamplifier output to return to one hundredth of its peak value, can be calculated to be

$$\Delta t_{res} = \frac{k_B T C_F}{qNI_{FB}} \cdot ln\left[\frac{exp\left(\frac{Q}{C_F}\frac{q}{k_B T}\right) - 1}{exp\left(\frac{Q}{100 \cdot C_F}\frac{q}{k_B T}\right) - 1}\right]. \tag{3.41}$$

It is worth noticing that in this shaperless version of the analog processor, the charge preamplifier encompasses the functionalities usually accomplished by the shaper; indeed, noise has been decreased by purposely limiting the preamplifier bandwidth, and the peaking time of the signal at the preamplifier output can be adjusted by means of the current biasing the preamplifier feedback network.

3.5.3 Threshold discriminator

The threshold discriminator circuit provides two inputs: the first one is the output signal from the preamplifier, while the second one is an adjustable threshold voltage. This circuit generates an high output logic level if the signal coming from the preamplifier exceeds a set level. If this does not occur, the output will remain at a low logic level. The threshold discriminator is based on a differential pair followed by a common source PMOS gain stage as shown in

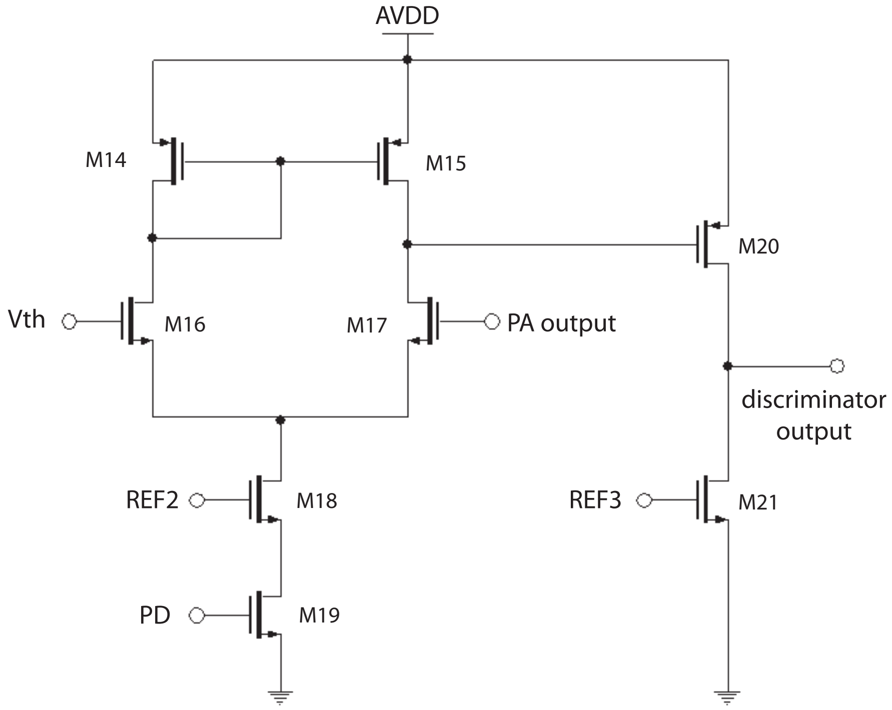


Figure 3.18: Schematic circuit of the threshold discriminator.

Devices	Gate W/L	Threshold voltage
M14 and M15	3 µm / 3 µm	standard
M16 and M17	16.3 µm / 1 µm	standard
M18	1.7 µm / 1 µm	low Vth
M19	300 nm / 150 nm	standard
M20	1 µm / 130 nm	low Vth
M21	1 µm / 130 nm	standard

Table 3.7: Dimensions of the devices belonging to the threshold discriminator.

Fig. 3.18. M16 and M17 transistors make up the input differential pair and M14 and M15 provide the PMOS active load. The area of these devices is relatively large with respect to the other front-end transistors (see table 3.7) in order to reduce the threshold dispersion as discussed in the next subsection. The common-mode current is provided by M18 and can be switched-off by means of the PD signal on the gate of M19 for power saving during the readout phase. It is worth noticing that all the devices were laid out on the analog layer, which also hosts the DNW sensor, except for the PMOS active load (M14, M15) and the output gain stage (M20, M21). This section of the comparator was integrated on the digital layer, as schematically shown in Fig. 3.9. In this way, it was possible to reduce the area of competitive n-wells in the sensor layer which, as discussed earlier, act as parasitic collecting electrodes with respect to the main DNW electrode. The 3D process makes this separation possible, interconnecting the two layers by means of three 3D contacts: two of them interconnect the NMOS differential pair with the active load, whereas the third provides the analog supply voltage to the comparator devices integrated in the digital layer.

3.5.4 Analysis of the discriminator threshold dispersion

High granularity detection systems are required to provide high performance parallel processing. Differences in the parallel paths followed by the signals may jeopardize the system response uniformity. In the case of a binary front-end like the one described in the previous section, channel-to-channel threshold nonuniformity (or dispersion) might make overall optimization of the system detection efficiency and minimization of noise occupancy quite a complicate matter. The main contributions to threshold dispersion come from device mismatch in the charge preamplifier and in the discriminator. A simplified circuit

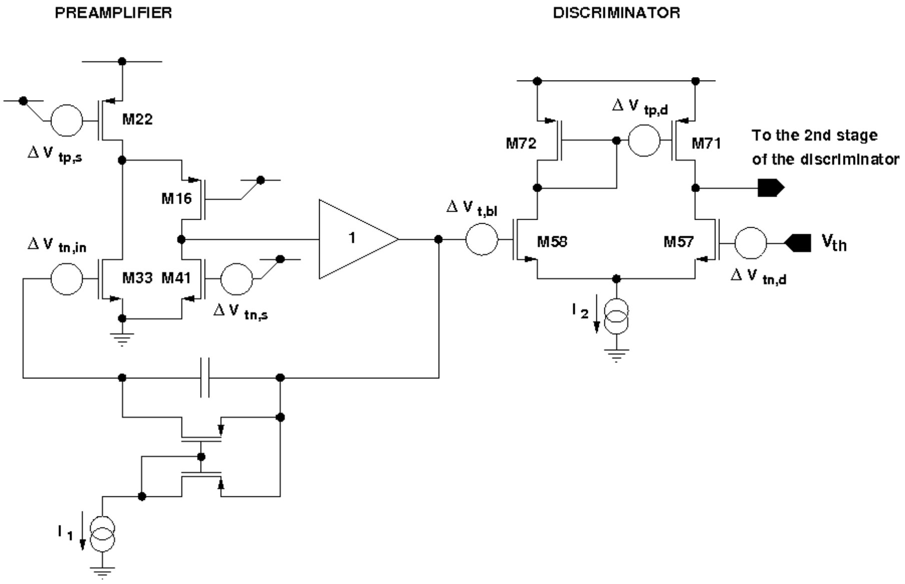


Figure 3.19: Simplified circuit schematic of the charge preamplifier and of the discriminator with voltage sources for threshold dispersion modeling and calculation.

schematic, emphasizing the main dispersion sources, is shown in figure 3.19. Differences in the threshold voltage V_t and in the current gain factor β are the predominant mismatch sources in closely spaced, identical-by-design MOS transistors [26]. Under common operating conditions, the contribution from V_t variations is indeed considerably more significant than the one from β mismatch [38]. Moreover, it becomes largely dominant in devices operated close to or in weak inversion [39]. Therefore, only the effect of V_t mismatch will be considered in the study of the channel-to-channel threshold dispersion. In the widely accepted model of V_t mismatch [40], the threshold voltage variation ΔV_t has a normal distribution with zero mean and a variance $\sigma^2(\Delta V_t)$ inversely proportional to the device gate area:

$$\sigma^2(\Delta V_t) = \frac{A_{vt}^2}{W \cdot L}. \tag{3.42}$$

In (3.42), A_{vt} is a constant obtained from the characterization of a statistically significant number of device pairs and generally provided by the foundry together with the device models. In figure 3.19, threshold mismatches are modeled by means of the voltage sources $\Delta V_{tp,s}$, $\Delta V_{tn,in}$, $\Delta V_{tn,s}$, $\Delta V_{tn,d}$ and $\Delta V_{tp,d}$. Spread of the threshold voltage of transistors M33, M22 and M41 in the charge preamplifier may change the preamplifier output level in one pixel relative to the level in another pixel. Assuming small mismatch values, small signal analysis of the circuit can be performed to calculate the resulting contribution to the baseline dispersion:

$$\sigma(\Delta V_{t,bl}) = \sqrt{\sigma^2(\Delta V_{tn,in}) + \frac{g_{mp,s}^2}{g_{mn,in}^2}\sigma^2(\Delta V_{tp,s}) + \frac{g_{mn,s}^2}{g_{mn,in}^2}\sigma^2(\Delta V_{tn,s})}. \quad (3.43)$$

In (3.43), $\sigma(\Delta x)$ is the standard deviation of Δx and $g_{mp,s}$, $g_{mn,in}$ and $g_{mn,s}$ are the channel transconductances of transistors M22, M33 and M41 respectively. The effect of mismatch in the (M58,M57) and (M72,M71) discriminator transistor pairs on the threshold dispersion can be modeled as a change $\Delta V_{t,d}$ in the externally set voltage V_{th} or, equivalently, as a variation -$\Delta V_{t,d}$ of the preamplifier output baseline. To calculate the resulting threshold dispersion, linear operation will be assumed for the discriminator. For this purpose it is worth observing that, although the discriminator almost always works in an unbalanced condition, transistor mismatch effects become detectable when the circuit is in its linear operating region, during 0-to-1 (or 1-to-0) transitions. Therefore, by performing small signal analysis of the circuit, the equivalent preamplifier output baseline dispersion due to mismatch in the discriminator MOSFET pairs is obtained as

$$\sigma(\Delta V_{t,d}) = \sqrt{\sigma^2(\Delta V_{tn,d}) + \frac{g_{mp2}^2}{g_{mn2}^2}\sigma^2(\Delta V_{tp,d})}, \quad (3.44)$$

where g_{mp2} and g_{mn2} are the channel transconductances of transistors M71 and M57 respectively. The total equivalent threshold dispersion can then be written as

$$\sigma(\Delta V_{teq}) = \sqrt{\sigma^2(\Delta V_{t,bl}) + \sigma^2(\Delta V_{t,d})}. \quad (3.45)$$

Threshold dispersion can be referred to the preamplifier input, and directly compared to the input charge signal, by taking into account the charge sensitivity G_Q:

$$\sigma(\Delta Q_t) = \frac{\sigma(\Delta V_{teq})}{G_Q}, \quad (3.46)$$

where ΔQ_t is the variation in the equivalent input threshold charge $Q_t = \frac{V_{teq}}{G_Q}$ due to device mismatch in the discriminator and in the preamplifier. Equation (3.46) shows that the dispersion in Q_t can be reduced by increasing the charge sensitivity.

3.6 Design criteria for analog processor

Optimum design criteria for readout chains processing the signals from capacitive detectors have been extensively discussed in the literature [41]. The criterion employed in the design of SDR1 chip relies upon the maximization of detection efficiency in multichannel systems with binary readout under noise hit rate constraints. Noise and threshold dispersion, both depending on the preamplifier input device dimensions, cannot be optimized separately. Nevertheless, the discriminator threshold can be minimized to maximize the detection efficiency, therefore providing a different design guideline for the charge preamplifier.

In the analog processor of the SDR1, 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 voltage, represented in Fig. 6 by the voltage source $\Delta V_{th,in}$ (topologically equivalent to the noise source e_n shown in Fig. 3.14), can be expressed in terms of an input charge as:

$$\sigma(\Delta Q_t) = C_F\ \sigma(\Delta V_{th,in}) = C_F\ \frac{A_{vt}}{\sqrt{WL}}, \tag{3.47}$$

where ΔQ_t represents the input referred dispersion in the preamplifier output baseline and W, L are the gate width and length of the input device. 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. If we neglect the effects of threshold dispersion, the minimum input referred discriminator threshold $Q_{t,min}$ may be set, for all the channels, based on the maximum noise hit rate $f_{n,max}$ (that is, the maximum rate of noise induced transitions at the discriminator output) the system can afford and on the ENC performance of the charge preamplifier:

$$\begin{aligned} Q_{t,min} &= ENC \cdot \sqrt{2 \cdot ln\left(\frac{f_{n0}}{f_{n,max}}\right)} = \\ &= \rho(f_{n,max}) \cdot ENC, \end{aligned} \tag{3.48}$$

where $\rho(f_n) = \sqrt{2 \cdot ln\left(\frac{f_{n0}}{f_n}\right)}$ is an extremely slowly increasing function of the ratio between the zero threshold noise hit rate f_{n0} and the noise hit rate, varying from about 3 to about 4.8 for $\frac{f_{n0}}{f_n}$ ranging from 100 to 100000. In order to take into account threshold dispersion effect, it is possible to modify (3.48) in the following way:

$$Q_{t,min} = \rho(f_{n,max}) \cdot ENC + \lambda \cdot \sigma(\Delta Q_t). \tag{3.49}$$

Let us assume that ΔQ_t features a normal distribution. If λ=0, then half the channels will find themself exceeding the maximum noise hit frequency. In order to keep 98% of the channels above the minimum tolerable threshold, λ=2 should be chosen [42].
In the design of the SDR1 chip the value of the current biasing the input branch of the charge preamplifier has been chosen to comply with the constraints set by the ILC vertex detector specifications (which require power dissipation less than 10 mW/cm^2 with power cycling operations). Moreover, the gate length of the input device has been set just over the minimum channel length allowed by the technology to minimize possible excess noise associated to short channel effects. Therefore, for the purpose of emphasizing the dependence of $Q_{t,min}$ on the gate width of the preamplifier input device, (3.49) can be rewritten as:

$$Q_{t,min}(W) = \rho(f_{n,max}) \cdot ENC(W) + \lambda \frac{C_F A_{vt}}{\sqrt{WL}} \tag{3.50}$$

Driving criterion used for the design of the preamplifier input device relies upon minimization of (3.50) [43]. In particular, simulations of $Q_{t,min}$ were performed varying the channel width (assuming $\rho = 4$ and $\lambda = 3$, a "safety value" larger than the one needed to keep 98% of the channels above the minimum tolerable threshold): simulation results are shown in 3.20. It is possible to notice that minimum value for $Q_{t,min}$ is achieved for W = 16 µm. Nonetheless, a gate width of 20 µm has been adopted in order to relax threshold dispersion constraints. Moreover, the difference in $Q_{t,min}$ between W = 16 µm and W = 20 µm is less than one electron.

3.7 Digital front-end

In the DNW-MAPS for the ILC vertex detector, besides the analog circuits described in the previous section, the elementary cell also includes two 5-bit time stamp registers and a set of logic blocks implementing data sparsification, namely the hit latch (FFSR), the token passing core and the get bus D latch. A D-type flip-flop (namely FFDR) is used in order to detect the second hit on the pixel. Other flip-flops are employed to control the readout operation. As

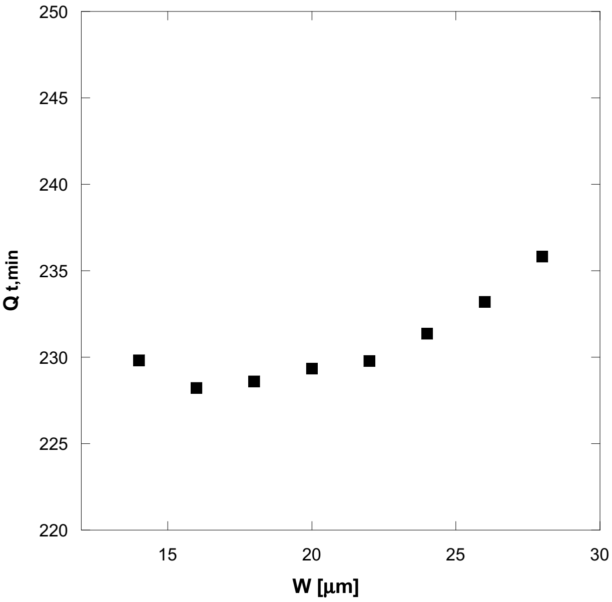


Figure 3.20: $Q_{t,min}$ as a function of the gate width W of the preamplifier input device.

already discussed in section 3.4.1, the 5 time stamp bits are fed to the registers by a Gray counter located at the chip periphery and allow the bunch train interval to be subdivided into 32 time slots. During the bunch train period, when a pixel is hit for the first time, its discriminator fires and sets the FFSR. The hit latch output is used to freeze the content of the time stamp register (through the, `WE` (`write_enable`) input in the time stamp register), therefore providing the arrival time of the hit with about 30 µs resolution in the case of the ILC beam structure. If a second hit occurs, the output of the flip-flop FFDR is set and the content of the second time stamp register gets frozen. At the end of the bunch train period, when the readout phase begins, the `LATCH_ENABLE` (which controls the buffer BUFR) signal is switched off (therefore preventing the latch from accidentally firing in those pixels which were not hit during the detection period) and the token is launched through the MAPS matrix. Each hit pixel, after receiving the token (`tkin` (`token_in`) signal), gets hold of the column and row buses (`gXb` (`get_X_bus`) and `gYb` (`get_Y_bus`) signals are pulled down) at the next cell clock rising edge, and releases position and time stamp information (by acting on the `OE` (`output_enable`), input of the time stamp register) within a cell clock period. Within a very short time

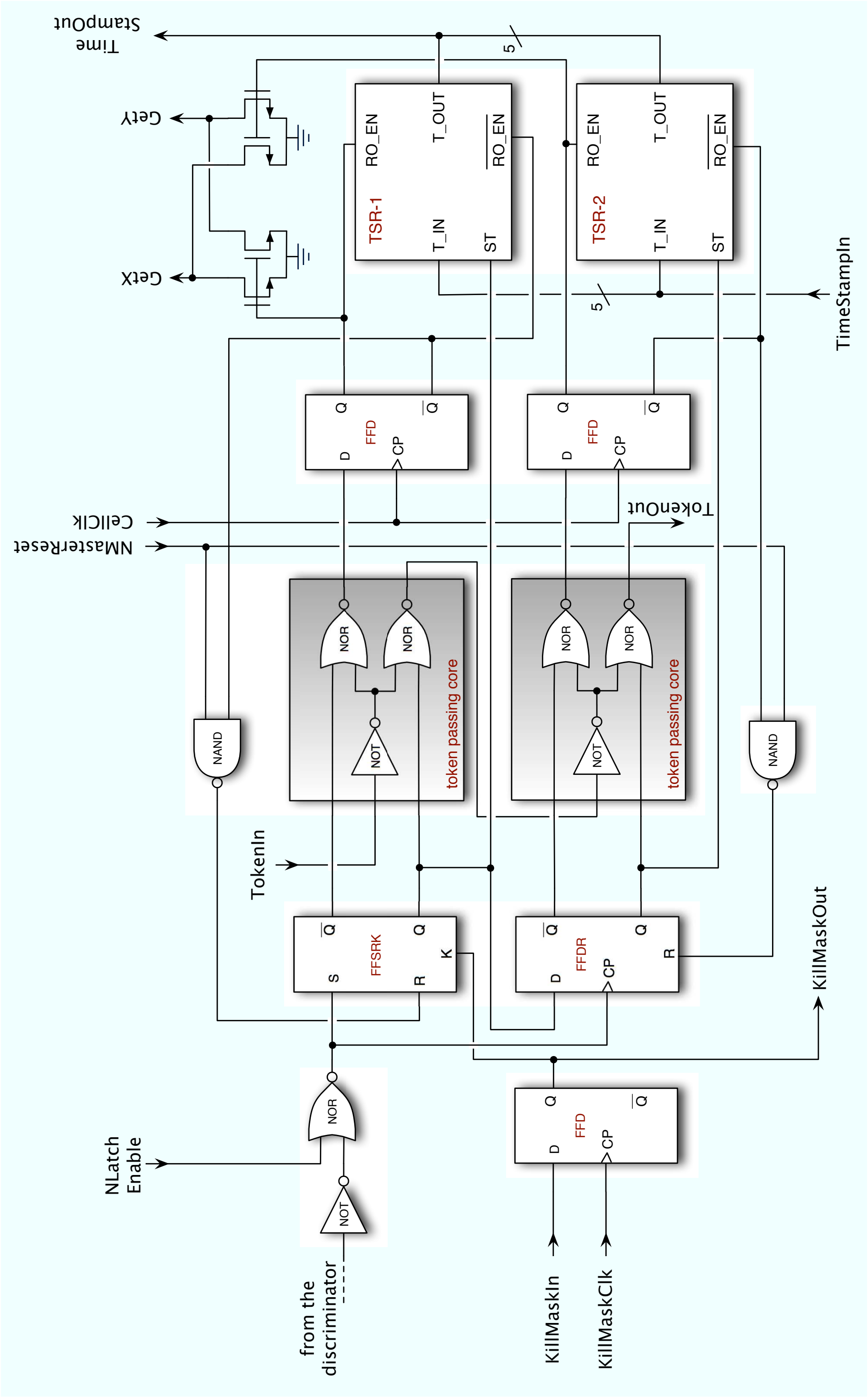


Figure 3.21: Block diagram of the in-pixel sparsification logic.

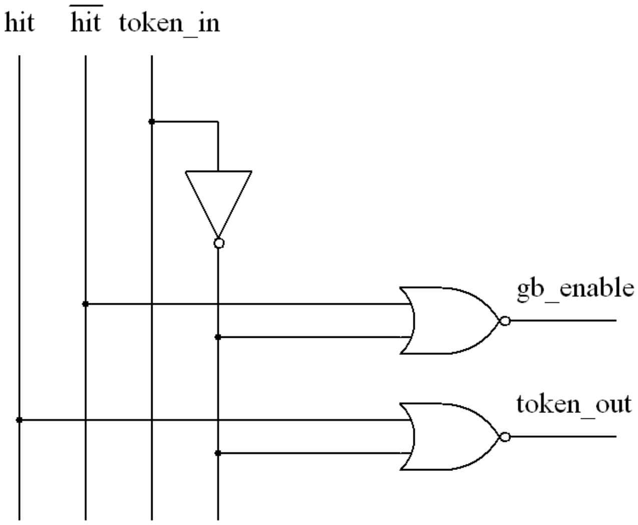


Figure 3.22: Schematic circuit of the token passing core stage.

interval, the hit latch is reset (this is possible if the MASTER_RESET signal is high, which is the common operating condition in both detection and readout phases); the token scans the second time stamp register and if there is not a second detected hit, it is released and sent out (tkout (token_out) signal) to the next hit pixel or to the matrix output (last_token_out signal, see figure 3.7). Note that, the token scans two time stamp registers per cells; two cell clock periods are needed to perform the readout of a pixel which has detected two hits. Details about the most important blocks and in particular the token passing core and the bit-storage unit employed in the time stamp register will be given in the following sections.

3.7.1 Token passing core

The token passing core is illustrated in figure 3.22. It is present in two equal areas (one for each of the two hit storage channel) within each cell; it is aimed at controlling propagation of the token signal during the readout phase. The token passing core is a digital combinatorial network that handles signals token_in, hit and its complement. The token passing core provides two output signals, the gb_enable and token_out signals. The first output, that is getb_en, is needed to enable row and column bus, thus making it possible to send the cell address to the output serializer. This happens at the first available rising edge of the `CELL_CLK`, only if a hit has been detected (`hit` = 1), and the cell got the *token* (`token_in` = 1). In this case the `gb_enable` signal is high, whereas `token_out` signal is low. At the end of the `CELL_CLK` period the hit latch is reset and the `token_out` signal gets high. The behavior of the second token passing core is similar at that of the first one.

3.7.2 Time stamp register

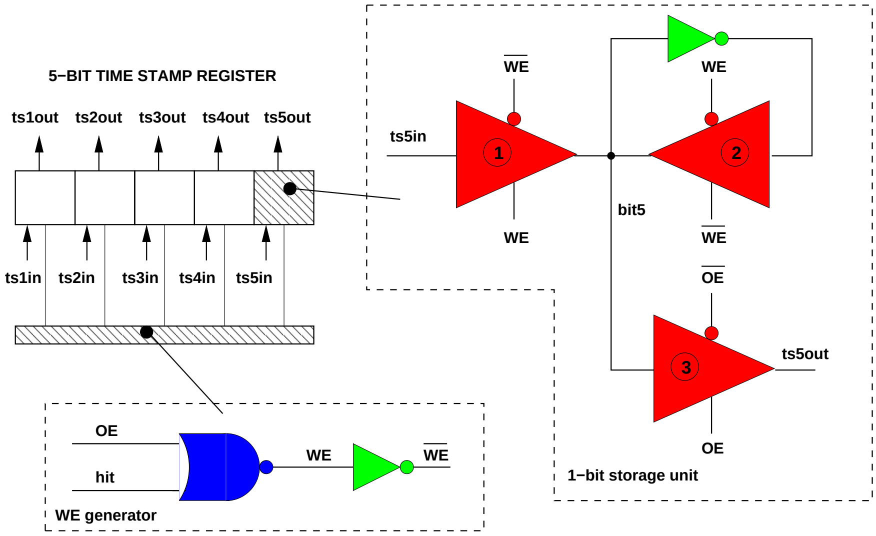


Figure 3.23: Schematic circuit of the time stamp register.

The 5-bit time stamp register, shown in figure 3.23, counts 55 minimum size transistors and consists of 5 identical 1-bit storage units and a couple of logic ports used to generate the `WE` signal. Each one of the storage units includes a latch whose schematic representation is shown in fig. 3.23, with the WE generator circuit.

When the `WE` signal is high, the time stamp register operates in the input configuration and is allowed to change its content. This condition persists until the hit takes place (`WE` gets low) and the connection with the time stamp counter gets interrupted by the input tri-state (1), which moves to the high impedance configuration. At the same time, the tri-state (2) and the inverter, which form a regenerative loop, store the content. The transfer of the information bit stored in the time stamp register takes place as soon as `OE` moves high and the output tri-state (3) is enabled.

3.7.3 Kill Mask

The 3D technology used in the production of the SDR1 chip enabled the integration of many additional functions with respect to the SDR0 chip: a new functional block involved in SDR1 is the so called *kill mask*.
The kill mask block consists in a shift register which involves all the cells of the matrix: the kill mask can be loaded in the matrix in order to disable noisy pixels. These pixels may produce spurious hits at the discriminator output, and turn out to be useless for the charge detection. The operation is controlled by means of the `KillMaskClock`, as shown in Fig. 3.22. The output of the FFD flip-flop involved in the kill mask block controls the FFRS flip-flop: if the `KILL` signal gets low, the cell is disabled. It is worth noticing that the preliminary setting of the system will necessarily involve the the kill mask loading.

3.8 Digital back-end

At the periphery of the matrix integrated in the SDR1 chip a further digital section is used in order to collect information from the pixels and to serialize the data. This section, in particular, includes a multiplexer, X and Y coordinate registers and a Gray counter used to provide the current time stamp value to time stamp registers in all the cells during the detection phase. Peripheral time stamp buffers are also included in this section. The pixels in each row or column in the pixel matrix share a specific coordinate register, respectively Y-register or X-register, in which the position is hardcoded with a number of bits equal to the smallest integer larger than

$$N_{bit} = log_2 N, \tag{3.51}$$

where N is the number of columns (X-register) or rows (Y-register). In the case of the SDR1 chip the matrix features240 rows and 256 columns so that 8 bits for both the coordinate registers are required. As far as the Gray counter is concerned, as earlier discussed a time stamp LSB of about 30 μs, corresponding to a 5 bit resolution, is regarded as sufficient for applications at the ILC vertex detector. Use of a Gray code minimizes time stamp inaccuracies and cross-talk effects between digital and analog section of the chip.
X, Y and time stamp data are serialized and sent off the chip by means of a 24 input (8 bits fot the X and the Y coordinates, 5 bits for the time stamp and 3 sync bits) multiplexer within a cell clock period. The readout clock (`Rck`) frequency is an integer multiple of the cell clock and is synchronous with it. Pull down transisotrs and tri-state buffers in the time stamp registers have

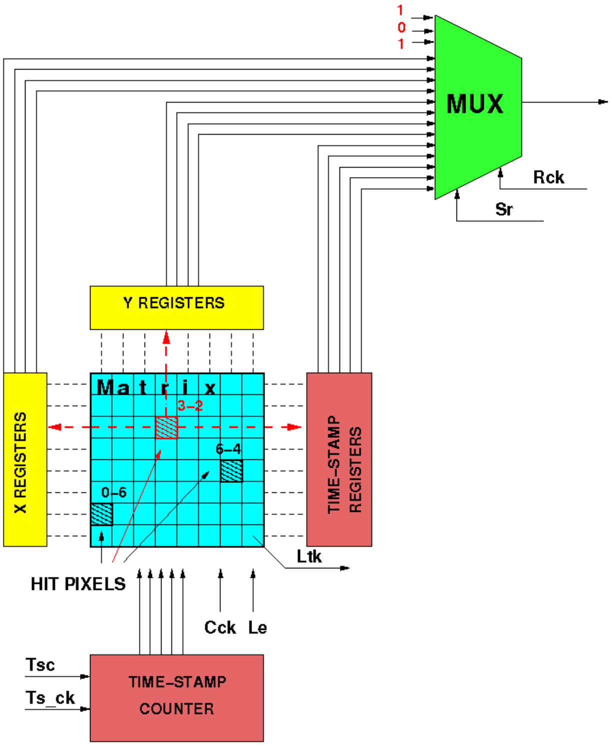


Figure 3.24: Block diagram of the SDR1 back-end electronics

been designed to enable readout clock operation at 100 MHz.

3.9 Geometrical features of the DNW sensor

As discussed earlier, the operation of the SDR1 sensor is based on a deep n-well structure acting as the collecting electrode. Fig. 3.25 displays a schematic representation of the elementary cell integrated in the SDR1 chip, where the distribution of the deep and standard n-wells is emphasized. The integration of large area PMOS devices in the digital top tier of the chip made it possible to reduce the area covered by the standard n-wells.

The detector capacitance includes two contributions: the first one is provided by the deep n-well/p-type epitaxial layer junction, whereas the second one comes from the deep n-well/internal p-well junction. In order to estimate the capacitance value, specific capacitances for the junctions were extracted from

the circuit simulations. The resulting values are summarized in table 3.8. In the same table, perimeter and area for the junctions composing the DNW sensor are also displayed.
The resulting total detector capacitance is therefore:

$$C_D = C_{DNW/P-sub} + C_{DNW/P-well} \simeq 47\ fF\ +\ 63\ fF \simeq 110\ fF \quad (3.52)$$

3.10 Cell layout

This section introduces the layout of the elementary pixel cell integrated in the SDR1 DNW MAP matrix. The elementary cell features two vertically integrated layers specifically processed by Chartered Semiconductor in a 130 nm CMOS technology, and vertical integrated by Tezzaron Semiconductor. Bonding between the two layer, whose layouts are shown in Fig. 3.26 , are obtained thanks to a grid of octagonal bond pads laid out on the top metal layer available for the Chartered process, with a grid pitch of 4 μm. Besides the

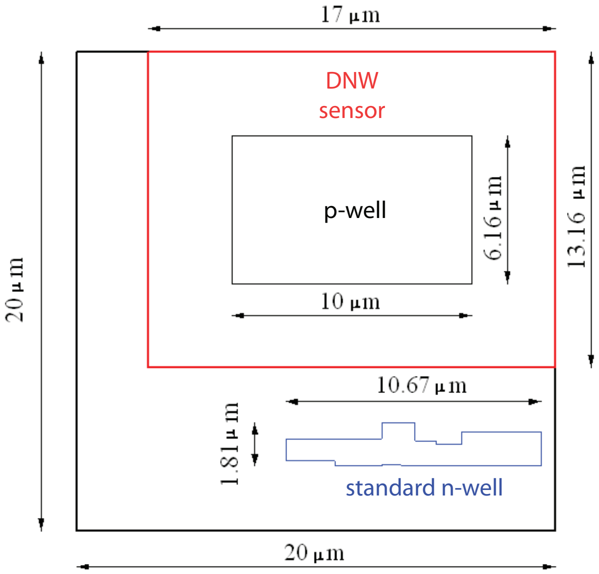


Figure 3.25: Deep and standard n-wells in the elementary cell in the SDR1 chip.

	DNW/P-sub	**DNW/P-well**
Capacitance per unit length [fF/µm]	0.664	0.231
Capacitance per unit area [fF/µm^2]	0.103	0.644
Perimeter length [µm]	60.32	32.32
Area [µm^2]	223.72	61.6
Resulting capacitance [fF]	47.14	63.09

Table 3.8: Specific capacitances and geometrical parameters relevant to the DNW sensor.

mechanical bonding, these pads are used to provide the electrical connection between the top and the bottom tiers. In the figure inter-tier connections are highlighted: two of them are used to connect the PMOS load of the discriminator integrated in the top digital layer with the NMOS pair placed in the analog one. Third connection provides the power supply to the analog portion integrated in the digital layer. The pixel size is 20 µm x 20 µm.

3.11 Circuit simulation results

In this section circuit simulation results relevant to SDR1 test structures will be shown. The simulations concern both the analog front-end and the digital section, with its front-end and back-end electronics.

3.11.1 Analog front-end

This section mainly discusses the noise contributions arising from the devices belonging to the charge sensitive preamplifier, evaluating their impact on the total output noise. Noise performance of the analog section will be evaluated in terms of equivalent noise charge (ENC), charge sensitivity, response linearity and threshold dispersion.

Noise contributions Circuit simulations point out that the sum of the noise contributions from devices belonging to the charge preamplifier, represented in terms of preamplifier output root mean square, $\sqrt{\overline{v_N^2}}$, is equal to 4.51 mV rms. Table 3.9 shows the main contributions arising from single devices, in decreas-

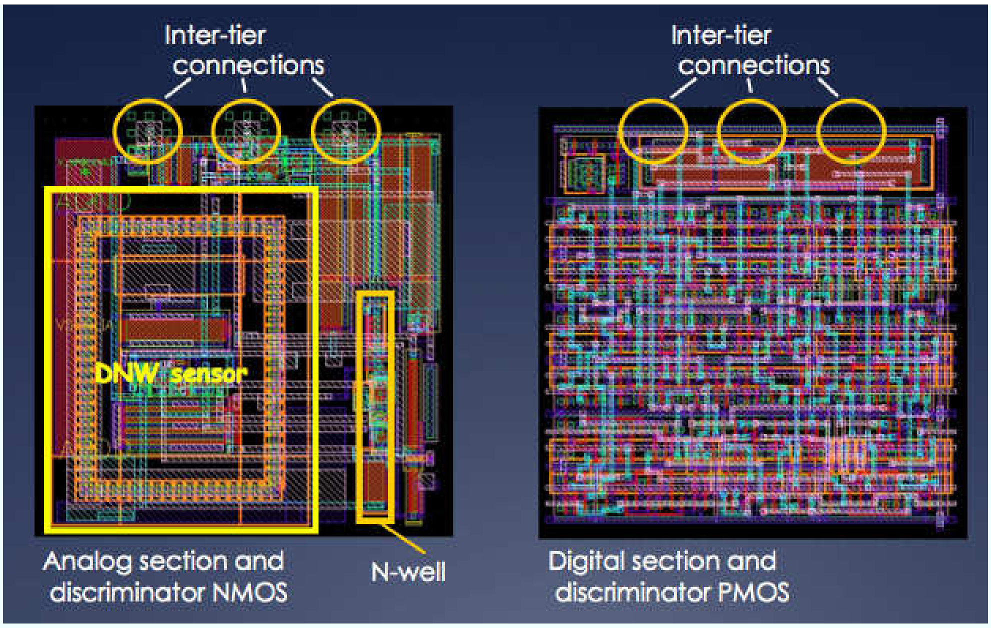

Inter-tier connections
Inter-tier connections
DNW sensor
Analog section and discriminator NMOS
N-well
Digital section and discriminator PMOS

Figure 3.26: Cell layout: analog and digital layers.

ing order. It is possible to notice that the main noise contribution comes from the input device of the preamplifier (M1 device in Fig. 3.10), as expected in a well-designed processor. Moreover, table 3.9 also points out that noise contributions from P-channel devices M2 in the input branch circuit and MP1 in the feedback current mirror stage (shown in Fig. 3.17), are not negligible. Actually, MP1 consists in the series of three PMOSFETs, namely MP1a, MP1b, MP1c, each featuring a W/L = 0.15/0.2. Bias voltages in the preamplifier are provided by internal voltage references shown in Fig. 3.27).

3.11.2 Equivalent noise charge

In charge measuring systems, the noise properties of the front-end processor are expressed by means of the equivalent noise charge (ENC). ENC has been defined by equation 3.25 as the charge that has to be injected at the input of the charge measuring system in order to obtain a unit signal-to-noise ratio at its output. In order to evaluate the ENC from the preamplifier, circuit simulations of the schematic shown in Fig. 3.28 have been performed. Charge injection at the preamplifier input is obtained by means of the current source

Device	**Noise contribution**	**Type**
M1	5.90 $(\mathrm{mV})^2$	channel thermal noise
M1	2.82 $(\mathrm{mV})^2$	1/f noise
M2	2.13 $(\mathrm{mV})^2$	channel thermal noise
MP1a	1.77 $(\mathrm{mV})^2$	channel thermal noise
MP1b	1.46 $(\mathrm{mV})^2$	channel thermal noise
MP1c	1.30 $(\mathrm{mV})^2$	channel thermal noise
M3	1.12 $(\mathrm{mV})^2$	channel thermal noise
M25	0.67 $(\mathrm{mV})^2$	channel thermal noise
M29	0.55 $(\mathrm{mV})^2$	channel thermal noise
M27	0.53 $(\mathrm{mV})^2$	channel thermal noise
M8	0.37 $(\mathrm{mV})^2$	channel thermal noise
M2	0.36 $(\mathrm{mV})^2$	1/f noise
M4	0.34 $(\mathrm{mV})^2$	channel thermal noise
M27	0.16 $(\mathrm{mV})^2$	1/f noise
M28	0.12 $(\mathrm{mV})^2$	channel thermal noise

Table 3.9: Noise contributions from devices belonging to the charge preamplifier. See Fig. 3.10 and 3.27 for references.

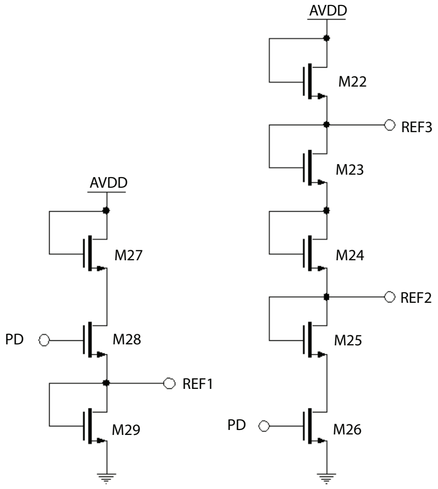


Figure 3.27: Schematic circuits of the voltage references in the preamplifier stage.

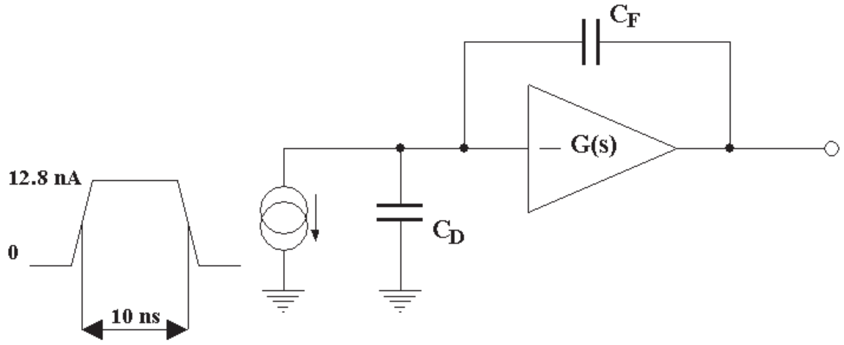


Figure 3.28: Schematic for the preamplifier ENC calculation.

providing a current pulse whose area corresponds to the injected charge.

In the performed simulation, the injected charge, 800 electrons, is the most probable value of the minimum ionizing particle response of the DNW MAPS sensors, resulting from the tests on previous prototypes, such as the APSEL series and the SDR0 chip, with radioactive sources [32], [44].

In Fig. 3.28 the detector is represented by means of its capacitance, C_D, at the preamplifier input; C_F is the preamplifier feedback capacitance.

Equivalent noise charge is obtained by means of:

$$ENC : \sqrt{\overline{v_N^2}} = Q : V_{max}. \tag{3.53}$$

where $\sqrt{\overline{v_N^2}}$ is the preamplifier output noise voltage and V_{max} is the preamplifier peak amplitude in response to an injected charge Q.

Fig. 3.29 shows the preamplifier output in response to an 800 electrons pulse: the peak amplitude is about 102 mV. Considering an output noise level of 4.51 mV rms, an equivalent noise charge of 35 electrons can be computed. This simulation has been carried out for a detector capacitance of 200 fF, which overestimates the capacitance of the deep n-well sensor, and for a feedback capacitance of 1 fF. This capacitance has been obtained by means of the fringing capacitance between two metal plates laid out in the feedback network of the charge preamplifier.

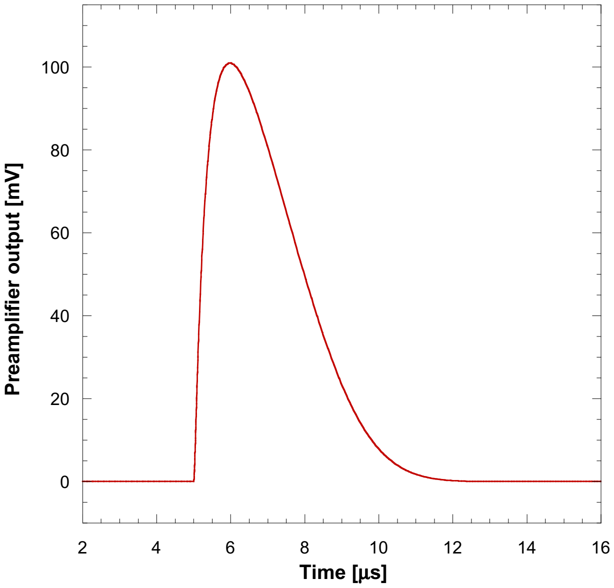


Figure 3.29: Simulated response of the charge preamplifier to an 800 e$^-$ input pulse (C_D = 200 fF, C_F = 1 fF).

3.11.3 Charge sensitivity and linearity

The charge sensitivity G_Q is defined as the slope of the fitting straight line of Fig. 3.30, showing the preamplifier output peak value V_{max} as a function of the injected charge Q. G_Q can be expressed by the following:

$$G_Q = \frac{\Delta V_{max}}{\Delta Q}. \tag{3.54}$$

From the previous equation, a charge sensitivity of 800 mV/fC can be computed.

An integral non-linearity (NLI) of about 2% has been obtained from circuit simulations over an input dynamic range of 2000 electrons, considering the following parameter:

$$NLI = \frac{\Delta V_{max}}{V_{max,2} - V_{max,1}} \cdot 100\%, \tag{3.55}$$

where ΔV_{max} represents the maximum gap between simulated values and the straight line passing through V_1 and V_2. It is worth noticing that linearity in binary readout channels does not play a fundamental role. Anyway, a good linearity over a large charge range can make the characterization of test structures easier.

3.11.4 Threshold dispersion

Threshold dispersion contributions related to the analog processor come from both the charge preamplifier and the threshold discriminator. The overall threshold dispersion can be calculated by means of Monte Carlo simulations where no charge is injected into the analog processor. The discriminator output is low when the level at the preamplifier output is lower than the set threshold; on the other hand, discriminator output gets high when the preamplifier output breaks that level. For each threshold value, 300 Monte Carlo simulations have been performed. These simulations kept count of the times in which the discriminator output was high. This value, normalized with respect to the number of simulations, is displayed in Fig. 3.31 as a function of the threshold voltage V_{th}.
At a given time the probability $F(V_{th})$ that a comparator with a threshold V_{th} is in high state is related to the probability density function $f(v)$:

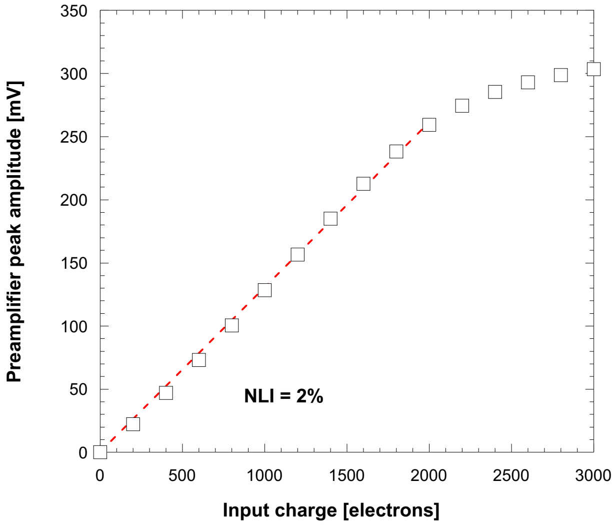


Figure 3.30: Peak amplitude of the preamplifier response as a function of the input charge. An integral non-linearity of 2% has been obtained in circuit simulations over an input dynamic range of 2000 electrons.

$$F(V_{th}) = \int_{V_{th}}^{\infty} f(v - V_{bl})dv, \tag{3.56}$$

where V_{bl} is equal to the DC preamplifier output level. In the hypothesis that $f(v)$ is Gaussian, with standard deviation $\sigma(\Delta V_t)$, equation (3.56) can be written as:

$$\begin{aligned} F(V_{th}) &= \frac{1}{\sigma(\Delta V_t)\sqrt{2\pi}} \int_{V_t}^{\infty} exp\left(\frac{-(v - V_{bl})^2}{2\sigma^2(\Delta V_t)}\right) dv = & (3.57) \\ &= \frac{1}{2} + \frac{1}{2} \cdot erfc\left(\frac{V_{th} - V_{bl}}{\sqrt{2}\sigma(\Delta V_t)}\right), & (3.58) \end{aligned}$$

where $erfc(x)$ is defined as follows:

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} exp(-t^2)dt. \tag{3.59}$$

It is possible to calculate the standard deviation $\sigma(\Delta V_t)$ related to threshold dispersion by interpolating the curve in Fig. 3.31 with the equation (3.58).

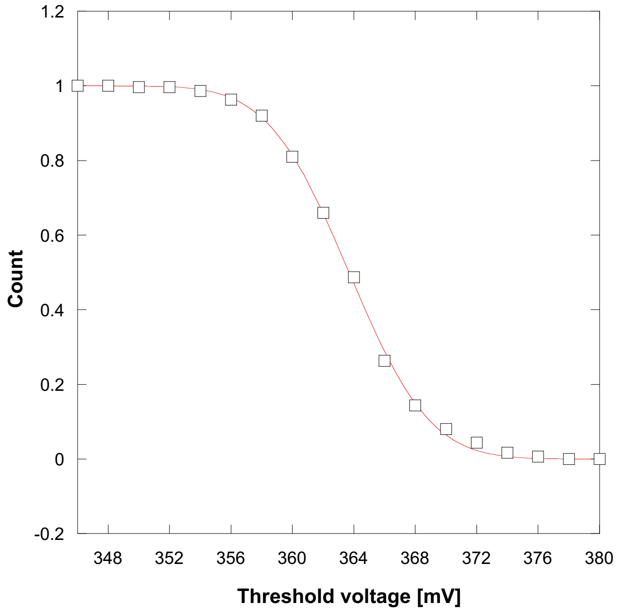


Figure 3.31: Normalized counts as a function of the comparator threshold.

Device	Dimensions	Simulated $\sigma(\Delta V_t)$
M1	20 $\mu m/180\ nm$	3.03
M2	800 $nm/2\ \mu m$	0.62
M29	150 $nm/2.2\ \mu m$	0.89

Table 3.10: Main contributions to the overall threshold dispersion related to the charge preamplifier.

The computed value is $\sigma(\Delta V_t) = 4.64$ mV,rms. An overall input referred threshold dispersion of 36 electrons is obtained by dividing $\sigma(\Delta V_t)$ for the charge sensitivity.
The computed value for the standard deviation, $\sigma(\Delta V_{t,bl})$, related to the dispersion at the preamplifier output baseline turns out to be 3.63 mV,rms. Table 3.10 summarizes the main threshold dispersion contributions coming from the charge preamplifier, with the relevant devices. The difference between $\sigma(\Delta V_t)$ and $\sigma(\Delta V_{t,bl})$ is due to the contributions from the devices in the discriminator.

3.12 Digital sections

This section introduces some of the circuit simulations related to the digital blocks, including the digital front-end and the digital back-end. In order to better understand these simulation results, Fig. 3.32 shows the pattern of the digital output words.
Output data are transmitted off the chip by means of a serializer controlled by a 5-bit binary counter, as shown in Fig. 3.33. Serializer input includes three hard-coded sync bits, row and column coordinates and time stamp data of the

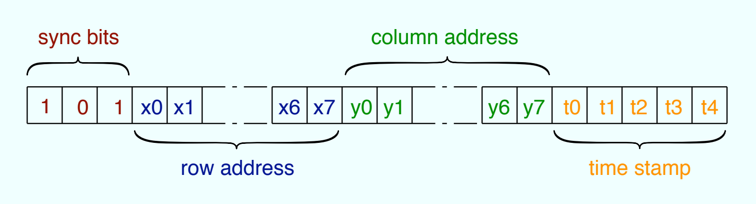


Figure 3.32: Serializer output words format.

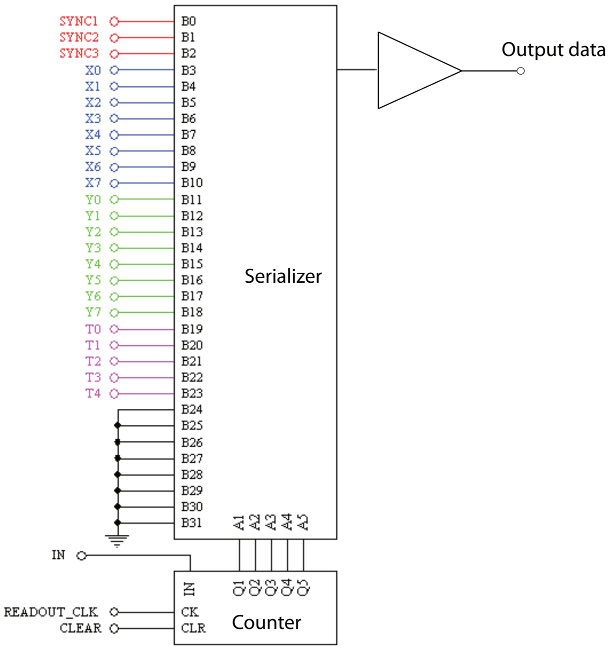


Figure 3.33: Serializer integrated in the 240X256 matrix periphery.

hit cells. Bits are serialized within a readout clock period: achievable bit rate is in the order of 100 Mbit/s. The cell clock period, since output data are 24 bit long words, is 24 times the readout clock period, and the two signals are synchronous.

In order to verify the correctness and eliminate possible bugs related to digital operations, circuit simulations on a 4x4 matrix, whose schematic representation is shown in Fig. 3.34, have been performed. In order to make simulations more realistic, bus lines from the cells to the matrix periphery were loaded with 3pF capacitors. This value has been obtained from geometrical considerations on the layout of the digital cell. The interconnections between cells in Fig. 3.34 represents the path for the `token` and `kill mask` signals.

Digital section simulations were performed by providing the `HIT` signal (threshold discriminator high-state) to some randomly selected cells. In particular, cell (1,0) in Fig. 3.34 is supposed to be hit twice during the detection phase, whereas cell (1,1) is supposed to be hit once. The behavior of the cell (3,0) is evaluated in order to verify the kill mask operations. Although the simulated matrix is undoubtedly smaller than the integrated matrix in the SDR1

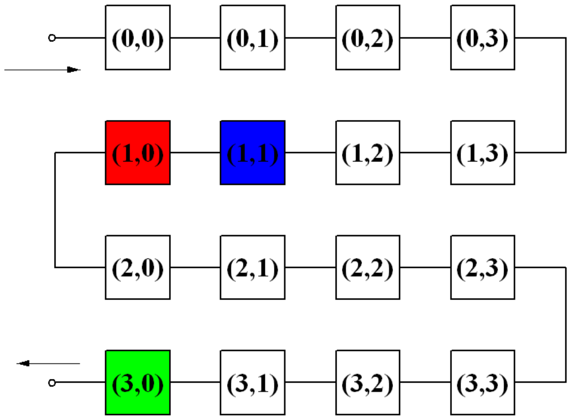


Figure 3.34: Schematic representation of a 4x4 matrix. During the detection phase, the red colored pixel is hit twice, the blue colored and the green colored pixels once; green one has been disabled by the kill mask.

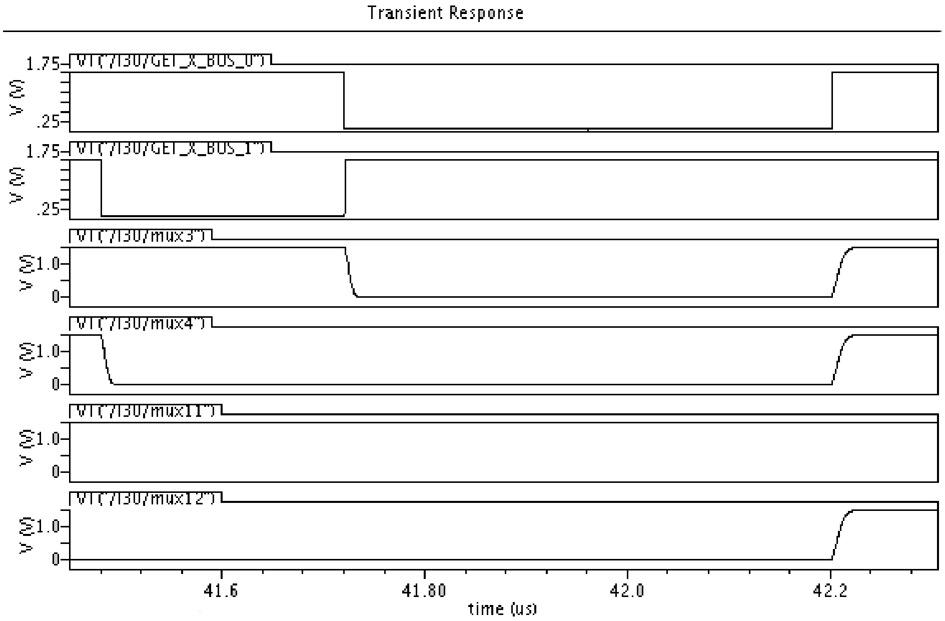


Figure 3.35: Coordinates signals related to pixels in which an `HIT` has been simulated.

chip, these simulations involve the same periphery electronics integrated in the SDR1.

X and Y coordinates

These simulations are suitable to check the correctness of X and Y coordinates writing at the serializer input. SImulation result is shown in Fig. 3.35. Accordingly to the token signal path, the bus related to the coordinate X=0 will be enabled first (GET_X_BUS_0 gets low); then, at the next CELL_CLK rising edge, GET_X_BUS_1$getslow, transmittingthecoordinateX = 1totheserializerinput.$ When GET_X_BUS_1 signal gets low, serializer input signals assume the following values:

- X1 = 1 (mux3)
- X2 = 0 (mux4)
- Y1 = 1 (mux11)
- Y2 = 0 (mux12)

which correspond to the binary code of the cell (1,1) address. Due to the fact that cell (1,0) is hit twice, during next two CELL_CLK periods GET_X_BUS_0 signal is in the low logic state. In this case, serializer input signals are:

- X1 = 0 (mux3)
- X2 = 0 (mux4)
- Y1 = 1 (mux11)
- Y2 = 0 (mux12)

corresponding to the binary code of the cell (1,0) address.
Fig. 3.36 shows the behavior of bits relevant to the row coordinates with the READOUT_CLK signal. It is worth noticing that serializer input signals (mux3 and mux4) are affected by a certain delay during the transitions. Because of this delay, due to parasitic effects on the bus lines, three sync bits are written at the beginning of the output words: in this way, mux_i signal can take advantage of three READOUT_CLK periods to reach its asymptotic value. Simulation results point out that this period is fully sufficient.

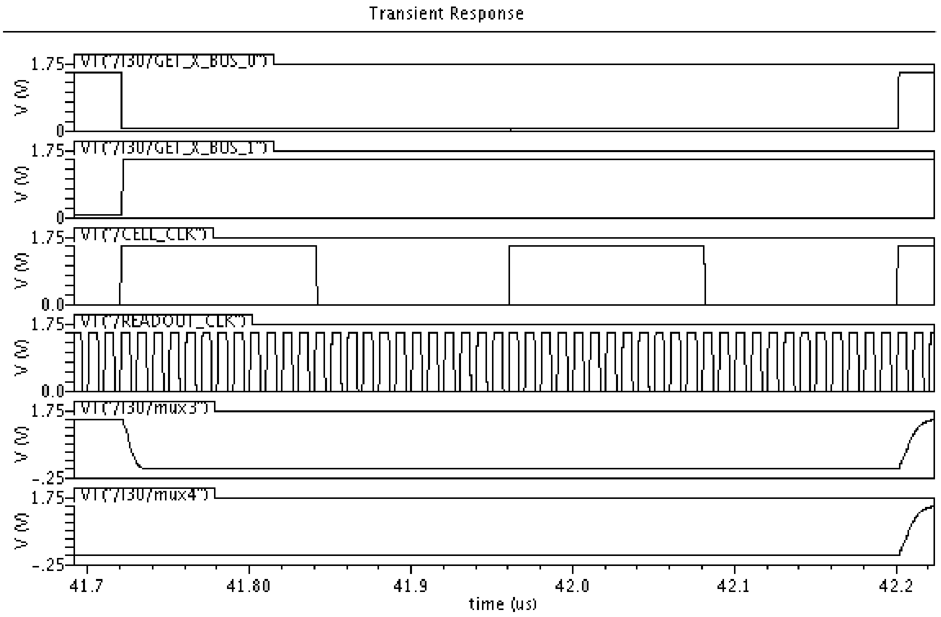


Figure 3.36: `GET_X_BUS_0` enables coordinate X=0 bus while X=1 bus gets released

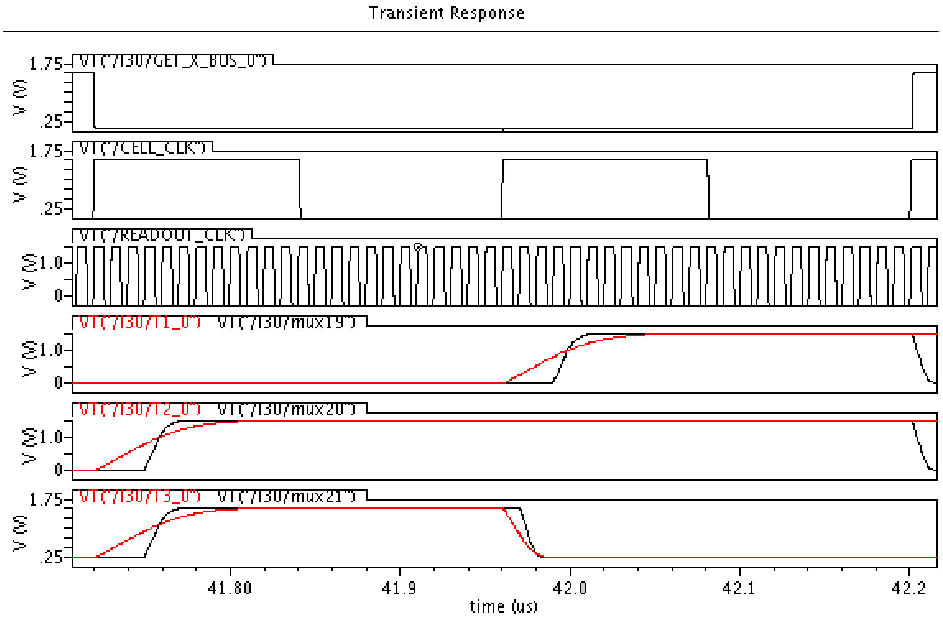


Figure 3.37: Read out of the data relevant to the time stamp register of the cell (1,0), in which two hits are simulated.

Time stamp register

Fig. 3.37 shows the transient response of the signals related to the time stamp register. It is possible to notice that these signals reach slowly their asymptotic value: this is due, besides the parasitic effects related to the bus lines, to the reduced size of the buffers belonging to the time stamp register, smaller than those integrated in the coordinate registers, because of the constraints on the pixel cell size. In order to cope with these effects, time stamp register data are read out after the coordinate data: in this way time stamp register bits can exploit 19 READOUT_CLK periods to reach its asymptotic value (3 sync bits, 8 bits for X coordinate and 8 bits for Y coordinate). Fig. 3.37 displays the behavior of three bits (red curves) relevant to the time stamp register of the cell (1,0), with the related signal at the serializer input (mux19, mux20 and mux21).

With reference to Fig. 3.37 it is also possible to observe the behavior of the cell (1,0), hit twice: fist, the GET_X_BUS_0 signal enables the bus 0 for two CELL_CLK periods. Then, it is possible to notice the transmission of the content of the two time stamp registers. Since only one cell is read, the signals related to the coordinates remain rather stable throughout this period.

Kill mask

The kill mask behavior has been simulated by setting the first bit of the mask and keeping at low logic level the remaining bits: this configuration makes the last cell of the matrix, with coordinates (3,0), insensitive to ionizing events. In order to verify the behavior of the kill mask blocks, let us consider the results relevant to simulations on cells (1, 0), (1, 1) and (3, 0). With reference to Fig. 3.38, 3.39 and 3.40 the first of the three shown signals is the HIT signal, the second one is the flip-flop FFSRK output, taking care of the first hit detection, and the third one is the flip-flop FFDR output, taking care of the second hit detection (reference to Fig. 3.21). Pixel cell with coordinates (1,1), the blue colored in Fig. 3.34, is hit once: from Fig. 3.38 it is possible to notice that the HIT signal set the FFRSK output. Pixel cell with coordinates (1,0), the red colored in Fig. 3.34, is hit twice: Fig. 3.39 points out that the HIT signal set both the FFRSK and FFDR output. Pixel cell with coordinate (3,0), the green colored, also is hit once, but, since it has been disabled by the kill mask, it can not detect any ionizing event concerning the cell.

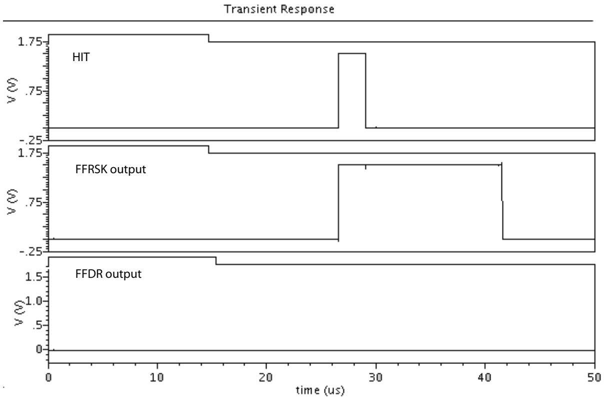


Figure 3.38: `HIT` signal and flip-flop FFRS and FFDR output signals of the cell (1,1).

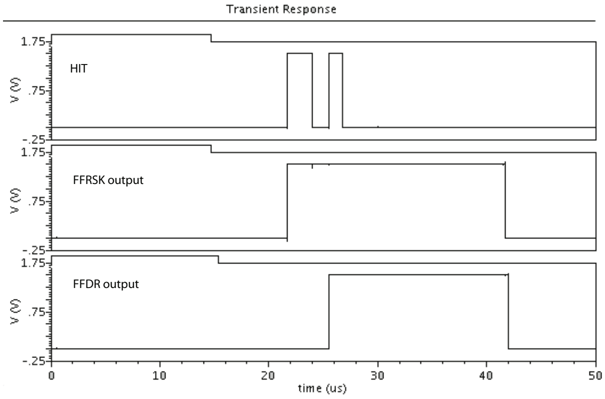


Figure 3.39: `HIT` signal and flip-flop FFRS and FFDR output signals of the cell (1,0).

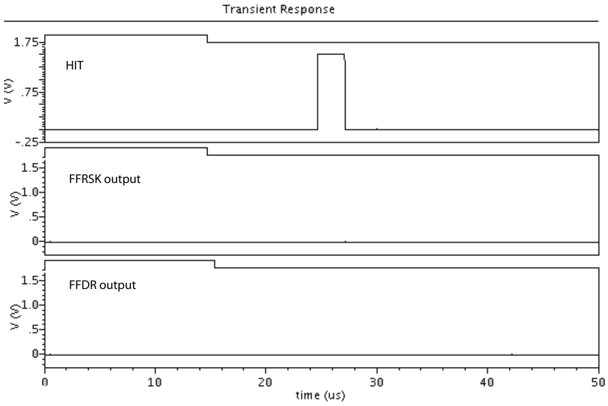


Figure 3.40: `HIT` signal and flip-flop FFRS and FFDR output signals of the cell (3,0).

3.13 Detection efficiency

Use of 3D processes is expected to provide significant benefits to the DNW MAPS sensor performance in terms of detection efficiency. The overall detection efficiency ϵ_T, of the DNW monolithic sensor integrated in the chips of the SDR series, under the hypothesis of factorizability, can be written as

$$\epsilon_T = \epsilon_{DNW} \cdot \epsilon_{AFE} \cdot \epsilon_{DFE} \cdot \epsilon_{ROA}. \tag{3.60}$$

In particular:

ϵ_{DNW} represents the contribution related to the sensor and is given by the ratio between the number of particles producing a signal which exceeds the discriminator threshold and the number of particles hitting the detector; it depends on the geometry of the DNW sensor and of the standard N-wells, on the threshold level and on technology parameters, such as the substrate resistivity which determines the amount of the collected charge;

ϵ_{AFE} represents the contribution provided by the analog front end and the threshold discriminator; it is basically related to the analog front-end dead time, that is the time during which the front-end output is over threshold and the cell is blind, and depends on the threshold level and on the current biasing the mirror stage in the preamplifier feedback network;

ϵ_{DFE} represents the contribution to the overall efficiency provided by the digital front-end and is given by the ratio between the number of hit stored in the pixel and the total number of hits on the cell during the detection phase; it depends on the storing capability of the cell, the hit occupancy (i.e., the number of particles hitting the detector per bunch crossover (BCO) per mm^2), the cluster multiplicity and the detector pitch;

ϵ_{ROA} represents the contribution to ϵ_T provided by the readout architecture and is given by the ratio between the number of read out hits and the number of stored hits; it mainly depends on the readout clock frequency, the number of pixels per chip and the hit occupancy.

Use of vertical integration technologies is expected to provide a significant impact mainly on ϵ_{DNW} and ϵ_{DFE}. In the following sections, such an impact will be evaluated, through Monte Carlo simulations and analytical calculations. Moreover, a comparison between detection efficiency performance in the SDR1 and in the SDR0 chips will be sketched out. It is worth recalling here that SDR0 chip is the 2D precursor of the SDR1 DNW MAPS.

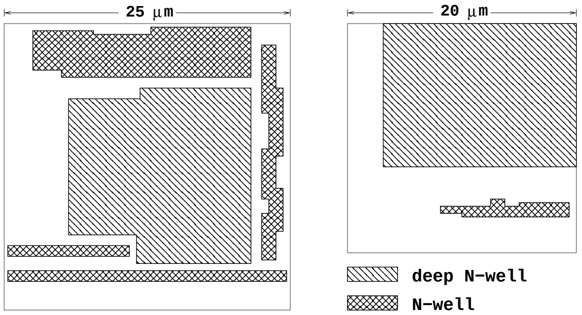


Figure 3.41: Layout of the DNW sensor and of the N-wells for PMOS transistors in (left) the SDR0 and (right) the SDR1 (bottom tier) MAPS detectors.

3.13.1 Sensor detection efficiency

As earlier discussed, in the design of DNW MAPS, 3D technologies can be used to move a large part of the PMOS devices and their N-wells to a different layer from the collecting electrode. Fig. 3.41 shows the layout of the DNW sensor and of the N-wells for PMOS transistors in the SDR0 and SDR1 (bottom tier) MAPS detectors. In the SDR0 pixel, the DNW sensor covers just 35% of the cell area, whereas it covers more than 50% in the case of SDR1. Moreover, a non negligible fraction of the cell area is covered by standard N-wells in the SDR0 pixel whereas they take a small portion (amounting to about one tenth of the total area covered by DNW and N-well diffusions) in the case of the SDR1 detector. In order to fully appreciate the beneficial effects of 3D design on the sensor collection efficiency, two series of Monte Carlo simulations have been performed. These simulations are based on a random walk algorithm developed to model the carrier motion in the undepleted substrate of monolithic pixel detectors [44]. For each series, a set of 10000 particles randomly hitting the central pixel of a 3×3 matrix have been simulated. Sensor layouts employed in these simulations are shown in Fig. 3.41. The resulting sensor detection efficiency, ϵ_{DNW}, is displayed in Fig. 3.42 as a function of the discriminator threshold. Sensor detection efficiency is still well over 99% at a discriminator threshold of 300 electrons in the SDR1 MAPS. Fig. 3.43 shows the average cluster size as a function of the discriminator threshold. It is worth noticing that SDR0 cluster size is always smaller than in the SDR1 case. This may be due to the larger area covered by standard N-wells in the SDR0 chip, which in turn reduces the available charge for the main DNW electrode, therefore reducing, at each threshold level, the average number of pixels over threshold.

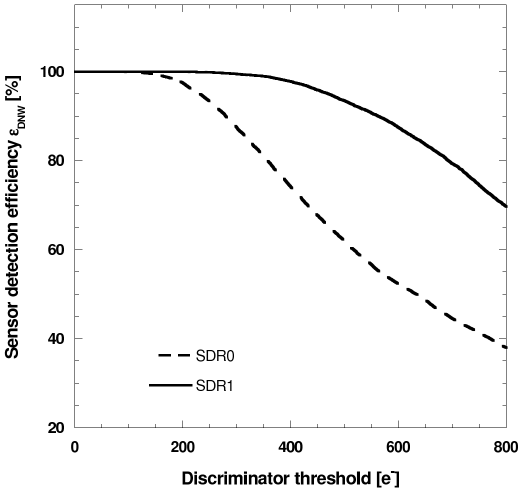


Figure 3.42: Detection efficiency contributed by the sensor as a function of the discriminator threshold in the case of the SDR0 and of the SDR1 chip.

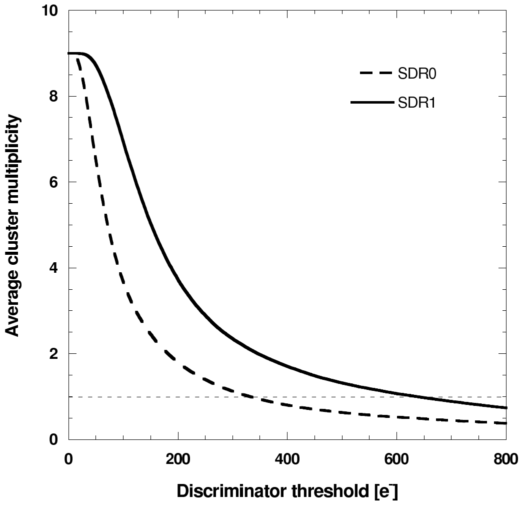


Figure 3.43: Average cluster multiplicity as a function of the discriminator threshold in the case of the SDR0 and of the SDR1 chip.

3.13.2 Digital front-end detection efficiency

Detection efficiency of the overall system may be increased by the capability for double-hit storing implemented in the digital front-end of the SDR1 chip. The benefits related to 3D technologies may be evaluated by directly calculating the detection efficiency contributed by the digital front-end. For this purpose, let h_o be the hit occupancy (i.e., the average number of particles hitting the detector per BCO per area unit), c_m the cluster multiplicity, n_b the number of bunches in a bunch train and p the detector pitch. The occupancy O_c for a single cell (i.e., the average number of hits per cell) is given by

$$O_c = h_o c_m n_b p^2. \tag{3.61}$$

It is worth recalling here that SDR1 features a different pixel pitch with respect to the SDR0 chip. Also the cluster multiplicity, which depends on the threshold, is not the same in the two sensors, as shown in Fig. 3.43. Let us assume a Poisson distribution for the number of times an elementary cell is hit during a bunch train period. The probability $P(n)$ of an elementary cell being hit exactly n times in a bunch train is then given by

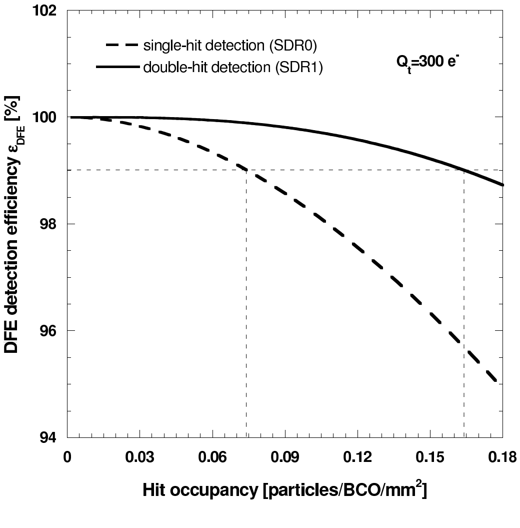


Figure 3.44: Detection efficiency contributed by the digital front-end as a function of the hit occupancy in the case of the SDR0 (single-hit detection) and of the SDR1 (double-hit detection) chip.

$$P(n) = \frac{O_c^n}{n!} \cdot \exp(-O_c). \tag{3.62}$$

If a pixel cell has the capability for storing m hits in a bunch train, the relevant digital front-end detection efficiency corresponds to the probability of the the pixel being hit no more than m times,

$$\epsilon_{DFE} = P(n \leq m) = \sum_{i=0}^{m} P(i). \tag{3.63}$$

Therefore the detection efficiency for the SDR0 MAPS, $\epsilon_{DFE,SDR0}$, corresponding to the probability of an elementary cell being hit no more than once, is given by

$$\epsilon_{DFE,SDR0} = \exp(-O_{c0}) + O_{c0} \cdot \exp(-O_{c0}), \tag{3.64}$$

where O_{c0} is the occupancy for the SDR0 pixel. On the other hand, the digital front-end detection efficiency for the SDR1 MAPS, $\epsilon_{DFE,SDR1}$, corresponding to the probability of an elementary cell being hit no more than twice, is given by

$$\begin{aligned}\epsilon_{DFE,SDR1} &= \exp(-O_{c1}) + O_{c1} \cdot \exp(-O_{c1}) + \\ &+ \frac{O_{c1}^2}{2} \cdot \exp(-O_{c1}),\end{aligned} \tag{3.65}$$

where O_{c1} is the occupancy for the SDR1 pixel. Fig. 3.44 shows the detection efficiency contributed by the digital front-end as a function of the hit occupancy in the case of the SDR0 (single-hit detection) and of the SDR1 (double-hit detection) chip at a discriminator threshold Q_t of 300 electrons. The curves take into account the cluster size as obtained from the data of Fig. 3.43. It is possible to notice that the detection efficiency in the SDR1 digital front-end is larger than 99% at h_o=0.15 particles/BCO/mm^2, a factor of five safety margin compared to the foreseen value for the hit occupancy in the innermost layer of the ILC vertex detector. [45].

3.14 Structures integrated in the SDR1 chip

In late 2008 a large number of international laboratories and universities with interest in high energy physics formed a consortium for the development of

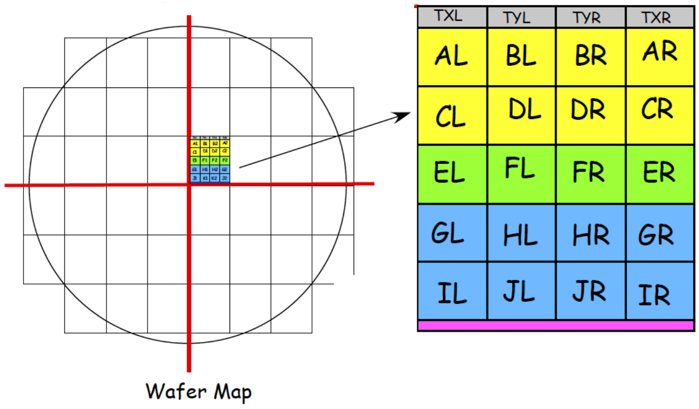


Figure 3.45: Frame for the first multi-project wafer run.

3D integrated circuits: the 3D IC Consortium [47]. The first joint effort was to organize a 3D multi-project run utilizing wafers from the Chartered 0.13 µm CMOS process and 3D assembly by Tezzaron.
The SDR1 design has been submitted in September 2009 in that multi-project wafer run and the delivery of the chips is foreseen in January 2010. The frame for the first multi-project run is divided into 12 sub-reticules among consortium members, as shown in Fig. 3.45. Fig. 3.46 displays the multi-project wafer full frame. SDR1 test structures are included in the E and F sub-reticules.
The list of the test structures designed in the SDR1 prototype includes:

- 3D DNW MAPS matrix, 240x256, 20 µm pitch pixels, intertrain sparsified readout architecture;

- smaller 3D DNW MAPS matrices, 16x16, 8x8, 3x3 20 µm pitch pixels and single pixel structures for easier laboratory testing of the sparsified architecture and of the analog front-end.

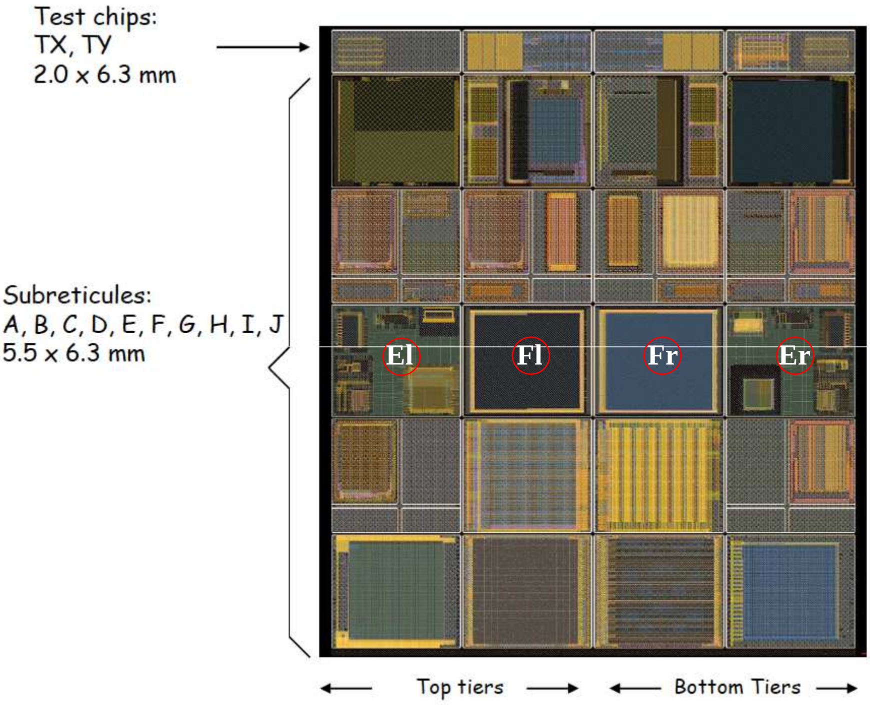


Figure 3.46: Multi-project wafer full frame: the sub-reticules E and F are highlighted.

Conclusions

In this thesis work, the design of the first generation of deep n-well CMOS monolithic sensors in a 130 nm vertical integration technology has been discussed. The use of a double-layer process in the development of the proposed DNW MAPS, the SDR1 chip, may address the main issues related to planar 130 nm CMOS technology. In particular, the 3D approach can cope with the low charge collection efficiency (due to the non negligible area covered by charge-stealing n-wells) and with the low detection efficiency (due to limitations in hit-storing capability) exhibited by the 2D detectors. Moreover, beside the higher functional density provided by the vertical integration processes, better point resolution and reduction of cross-talk phenomena between digital blocks and the analog section can be achieved by means of the 3D technology.
The design of the analog section of the SDR1 pixel, which includes a charge preamplifier and a threshold discriminator, was carried out paying particular attention to its noise and threshold dispersion properties. Significant improvements in terms of detection efficiency have been confirmed by Monte Carlo simulations and analytical calculations.
As far as the digital section is concerned, the main design guideline was the increase in digital functional density, translated in the capability for double-hit detection and double 5-bit time stamping.
The SDR1 prototype chip includes several different test structures, among which a 240X256 DNW MAPS matrix. Full experimental characterization of the submitted structures is foreseen in January 2010.

Bibliography

[1] *The International Technology Roadmap for Semiconductors (IRTS)*, 1999.

[2] C. Hu, “MOSFET scaling in the next decade and beyond”, *Semicond. Int.*, pp. 105-114, 1994.

[3] M. T. Bohr, “Interconnect scaling - The real limiter to high performance ULSI”, *IEDM Tech. Dig.*, pp. 241-244, 1995.

[4] D. Edelstein et al. , “Full copper wiring in a sub-0.25 μm CMOS ULSI technology”, *IEDM Tech. Dig.*, pp. 773-776, 1997.

[5] K. Banerjee, S.J. Souri, P. Kapur and K.C. Saraswat, “3-D ICs: A novel chip design for improving deep-submicrometer interconnect performance and systems-oc-chip integration”, *Proceedings of the IEEE*, vol. 89, no. 5, pp. 602-633, 2001.

[6] Y. Akasaka, “Three-Dimensional IC Trends.”, *Proceedings of the IEEE*, vol. 74, no. 12, pp. 1703-1714, December 2006.

[7] A. Rahman, A. Fan, R. Reif, and J.E. Chung, “Wire-Length Distribution of Three-Dimensional Integrated Circuits”, *Proceedings of the Int.Interconnect Technology Conference*, pp. 233-235, 1999.

[8] S. F. Al-sarawi, D. Abbot, P. D. Franzon “A review of 3-D packaging technology”, *IEEE Trans. Components, Pkg. & Manuf. Technolog.*, B 21, 1, 1998.

[9] M. Ieong, et al. “Three Dimensional CMOS Devices and Integrated Circuits”, *Proceedings of the IEEE Custom Integrated Circuits Conference*, 2003.

[10] E. R. Fossum, “CMOS Image Sensors: Electronic Camera-On-A-Chip”, *IEEE Trans. Elec. Dev.*, vol. 44, no. 10, pp. 1689-1698, Oct. 1997.

[11] J. Nakamura, B. Pain, T. Nomoto, T. Nakamura, E.R. Fossum , "On-focal-plane signal processing for current-mode active pixel sensors", *IEEE Trans. Elec. Dev.*, vol. 44, no. 10, pp. 1747-1758, Oct. 1997.

[12] V. Re, L. Gaioni, M. Manghisoni, L. Ratti, V. Speziali, G. Traversi, "CMOS technologies in the 100 nm range for rad-hard front-end electronics in future collider experiments", *Nucl. Instrum. Methods*, vol. A596, pp. 107-112, Oct. 2008.

[13] B. Rajendran et al. "CMOS transistor processing compatible with monolithic 3-D Integration", *Proceedings of VLSI Interconnection Conf. (VMIC)*, pp. 76-82, 2005.

[14] V. Dunton, T. Chen, M.Konevecki, U. Raghuram, S. Sivaram. "Zias: Vertical wires in 3-D memory devices", *Proceedings of VLSI Interconnection Conf. (VMIC)*, pp. 480-485, 2005.

[15] B. Markunas "3D architectures for semiconductor integration and packaging", *Presented at the RTI Int. Technology Venture Forum*, Burlingame, CA, 2004.

[16] J. Baliga, "Three-dimensional ICs solve the interconnect paradox", *Semiconductor Int.*, available online: http://www.reed-electronics.com/semiconductor/article/CA604503, 2005.

[17] B. Patti, "3D Bonding At Tezzaron", presented at *Pixel 2008 Intenational Workshop*, Fermilab, Batavia, September 23-26 2008.

[18] Colinge JP. "Silicon on insulator technology: materials to VLSI" *3rd ed. Kluwer Academic Publishers; 2004.*

[19] J. A. Burns, "3-D circuit integration technology for multiproject fabrication", *MIT Lincoln Laboratory, Lexington, MA*, 2001.

[20] R. Yarema, "Development of 3D Integrated Circuits for HEP", *12th LHC Electronics Workshop*, Sept. 25-29, 2006, Valencia, Spain.

[21] Y. Arai et al., "Monolithic Pixel Detector in a 0.15 μm SOI Technology", *IEEE NSS 2006 Conf. Rec.*, vol. 3, pp. 1440-1444, 2006.

[22] A. Bulgheroni et al., "Monolithic active pixel detector realized in silicon on insulator technology", *Nucl. Instrum. Methods*, vol. A535, pp. 398-405, 2004.

[23] M. Manghisoni, L. Ratti, V.Re, V. Speziali, “Instrumentation for noise measurements on CMOS transistors for fast detector preamplifiers”, *IEEE Trans. Nucl. Sci.*, vol. 49, no. 3, pp. 1281-1286, June 2002.

[24] M. Chan et al. , “Modeling the floating-body effects of fully depleted, partially depleted, and body grounded SOI MOSFETs”, *Solid State Elctron.*, vol. 48, pp. 969-978, 2004.

[25] J. Kuo, “*Low- Voltage SOI CMOS VLSI Devices and Circuits*”, New York, John Wiley, Sept 2001.

[26] Y. Tsividis “*Operation and Modeling of the MOS transisotr*”, McGraw-Hill, Boston, 1999.

[27] V. Re, M. Manghisoni, L. Ratti, V. Speziali, G. Traversi, “Survey of noise performances and scaling effects in Deep Submicron CMOS devices from different foundries”, *IEEE Trans. Nucl. Sci.*, vol. 52, no. 6, pp. 2733-2740, Dec. 2005.

[28] E. Simoen and C. Claeys, “On the flicker noise in submicron silicon MOSFETs”,, *Solid-State Electronics* 43, pp. 865-882, 1999.

[29] J. Bagger, T. Behnke, P. Burrows, J. Choi, E. Clements, J-P. Delahaye et al., “The International Linear Collider: Gateway to the Quantum Universe”, `http://www.linearcollider.org/gateway/`.

[30] The LDC outline document,
`http://www.ilcldc.org/documents/dod/outline.pdf`.

[31] L. Ratti, M. Manghisoni, V. Re, V. Speziali, G. Traversi, S. Bettarini et al., “Monolithic pixel detectors in a 0.13 μm CMOS technology with sensor level continuous time charge amplification and shaping”, *Nucl. Instrum. Methods*, vol. A568, pp. 159- 166, 2006.

[32] G. Rizzo et al, “Development of deep N-well MAPS in a 130nm CMOS technology and beam test results on a 4k-pixel matrix with digital sparsified readout”, *IEEE Nuclear Science Symposium Conference Record*, October 1925 2008, Dresden, Germany.

[33] 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”, *Nucl. Instrum. Methods*, vol. A581, pp. 291-294, 2007.

[34] D.C. Christian, J.A. Appel, G. Chiodini, J. Hoff, S. Kwan, A. Mekkaoui e al., “FPIX2, the BTeV Pixel Readout Chip”, *Nucl. Instrum. Methods*, vol. A549, pp. 165-170, 2005.

[35] V. Re, M. Manghisoni, L. Ratti, J. Hoff, A. Mekkaoui, R. Yarema, “FSSR2, a Self-Triggered Low Noise Readout Chip for Silicon Strip Detectors”, *IEEE Trans Nucl. Sci.*, vol. 53, no. 4, pp. 2470-2476, August 2006.

[36] G. Rizzo, G. Batignani, S. Bettarini, F. Bosi, G. Calderini, R. Cenci, “Recent Development on Triple Well 130 nm CMOS MAPS with In-Pixel Signal Processing and Data Sparsification Capability”, *2007 IEEE Nuclear Science Symposium Conference Record*, pp. 927-930, 2007.

[37] R. Muller, T. Kamins, ”*Dispositivi Elettronici nei Circuiti Integrati*”, Bollati Boringhieri, 1993.

[38] P. Kinget, M. Steyaert, “Impact of transistor mismatch on the speed-accuracy-power tradeoff of analog CMOS circuits”, *Proc. of the IEEE Custom Integrated Circuits Conference (CICC)*, 5-8 May 1996, pp. 333-336.

[39] K. Laker, W. Sansen, *Design of Analog Integrated Circutis and Systems*, McGraw-Hill, New York, 1994.

[40] K.R. Lakshmikumar, R.A. Hadaway, M.A. Copeland, “Characterization and modeling of mismatch in MOS transistors for precision analog design”, *IEEE J. Solid-State Circuits*, vol. 21, no. 6, pp. 1057-1066, Dec. 1986.

[41] G. De Geronimo, P. OConnor, “MOSFET Optimization in Deep Submicron Technology for Charge Amplifier”, *IEEE Trans. Nucl. Sci.*, vol. 52, no. 6, pp. 3223-3232, December 2005.

[42] L. Ratti, M. Manghisoni, V. Re, G. Traversi: “Design of time invariant analog front-end circuits for deep n-well CMOS MAPS”, *IEEE Trans. Nucl. Sci.*, vol. 56, no. 4, pp. 2360-2373, August 2009.

[43] L. Ratti, C. Andreoli, M. Manghisoni, E. Pozzati, V. Re, V. Speziali, G. Traversi, “Design of time invariant analog front-end circuits for deep n-well CMOS MAPS”, *IEEE Trans. Nucl. Sci.*, vol. 56, no. 4, pp. 2360-2373, August 2009.

[44] E. Pozzati et al., “MAPS in 130 nm and 90 nm triple well CMOS technologies for HEP applications”, *Proceedings of the Topical Workshop on Electronics and Particle Physics*, Prague, Czech Republic, 3-7 September 2007, pp. 492-497.

[45] J. Brau et al., “Monolithic CMOS pixel detectors for ILC vertex detection”, presented at the *2005 International Linear Collider Physics and Detector Workshop and second ILC Accelerator Workshop*, Snowmass, Colorado, August 14-27 2005.

[46] H. Spieler, *Semiconductor Detector Systems*, Oxford University Press, New York, 2005.

[47] The 3D IC Consortium, http://3dic.fnal.gov