**ALMA MATER STUDIORUM**
**UNIVERSITA' DI BOLOGNA**

ARCES – Advanced Research Center on Electronic Systems
for Information and Communication Technologies E. De Castro

# TECHNIQUES AND SOLUTIONS FOR 3D INTERCONNECTIONS

FEDERICO NATALI

**TUTOR**
Professor
**ELEONORA FRANCHI**

**COORDINATOR**
Professor
**CLAUDIO FIEGNA**

**PHD. THESIS**

**January, 2006 – December, 2008**

PhD Program in Information Technology
Cycle XXI – ING-INF/01

## KEY WORDS

CMOS, 3D Interconnections, Through Silicon Via Technologies,
Contact-less Communications, Latch-up Phenomena

**TO MY FAMILY**

**CONTENTS**

# CONTENTS

**CONTENTS**

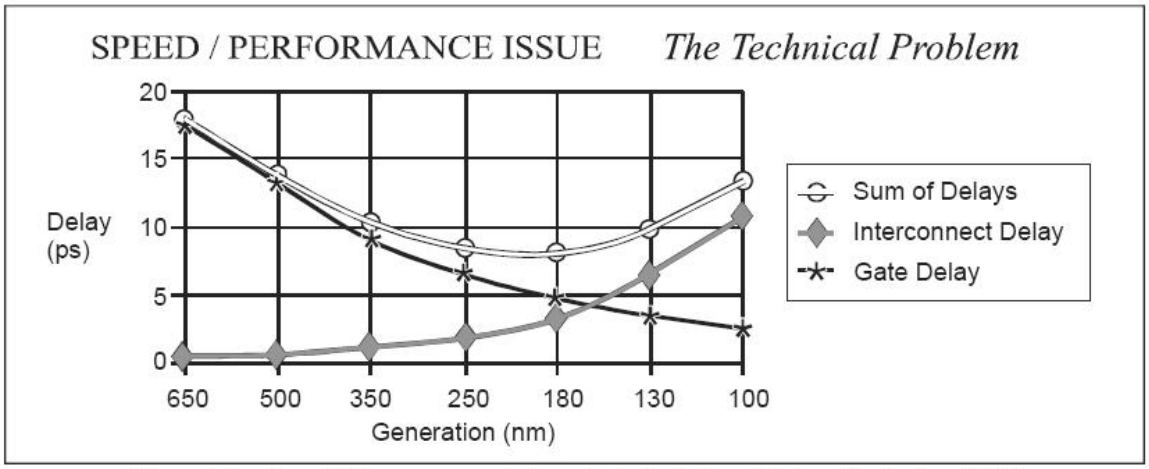
# ACRONYMS

| | |
|---|---|
| BOF | Bump on Flex |
| WOB | Wire on Bump |
| H3D | Interconnection attenuation |
| AGC | Automatic Gain Control |
| VGA | Variable Gain Amplifiers |
| BGWR | Band Gap Voltage Reference |
| HWR | Half wave rectifiers |
| | |
| TSV | Through Silicon Via |
| DRIE | Deep Reactive Ion Etching |
| CVD | Chimical Vapor deposition |
| | |
| SEE | Single Event Effect |
| SEU | Single Event Upset |
| SEB | Single Event Burnaut (in bipolar power transistors) |
| SEL | Single Event Latchup |
| EHP | Electron-Hole pairs |
| LET | Linear Energy Transfer |
| | |
| VDD2 | Secondary (vertical) supply grid |

# INTRODUCTION

## DRIVERS FOR 3D SYSTEM INTEGRATION

Transistors continue to improve in performance at smaller scales, but the wiring which connects them peaked in efficiency somewhere between 130nm and 90nm (Fig. 1). In fact, the performance improvement gained in transistor scaling is insignificant compared to the negative effects of interconnect scaling.




(Figure taken from "Heterogeneous Integration" *Tech Trend Notes*, September 2003)

Fig. 1 Technology node: ITRS Road Map

## INTRODUCTION

The only workable way to achieve real improvement in semiconductor systems is by reducing the average length of the interconnecting wire. Since several years packaging is driven by System in Package (SiP) solutions to meet the requirements of improved performance, miniaturization and cost reduction.

This leads to a number of technologies where 3D system integration is one of the main potential drivers [1]. In general, the introduction of 3D integration technologies is driven by:

• *Form factor*: Reduction of system volume, weight and footprint

• *Performance*: Improvement of integration density and reduction of interconnect length leading to improved transmission speed and reduced power consumption

• *High volume low cost production*: Reduction of processing costs for, e.g., mixed technologies

• *New applications*: e.g. ultra compact camera and detector systems and small wireless sensor nodes In competition to Systems on Chip (SoC) solutions, the 3D wafer level system integration enables the combination of different optimized production technologies.

3D integration is a possible solution to overcome the “wiring crisis” caused by signal propagation delay, both, at board and at chip level; because it allows minimal interconnection lengths and the elimination of speed-limiting intra- and inter-chip interconnects. The introduction of very advanced microelectronic systems, as e.g. 3D image processors, will be mainly driven by the enhancement of performance. The potential for low cost fabrication will be a further key aspect for future applications of 3D integration as well. Today, the fabrication of Systems on Chip (SoC) is based on embedding multiple technologies by monolithic integration. But there are serious disadvantages: the chip partition with the highest complexity drives the process technology which leads to a “cost explosion” of the overall system. In contrast to this, suitable 3D integration technologies enable the combination of different optimized base technologies, e.g. MEMS, CMOS, etc., with the potential of low cost fabrication

through high yield and high miniaturization degree. The characteristic parameters of different 3D Technologies as connection resistance or capacitance, minimum pitch and integration density, total power consumption, often are not the same so that it results difficult to compare the different interconnect approaches. Nevertheless, it is possible to determine terms of comparison considering the methods of assembly (die or wafer scale), packaging requirements, maximum number of tiers (tier refer to the chip in a stack, as opposed to the layers in a chip), pitch of vertical interconnects and achievable integration density, and amount of routing resources consumed on the chip.

# STRUCTURE OF THE THESIS

This thesis is subdivided into two main parts focusing on different aspects of 3D Interconnection Technologies. The first part describes systems developed for contact-less approaches. Appropriate circuits have been realized transmitting both digital signals and analog signals.

The research activities carried out during the last part of PhD course focuses on the analysis of Latch-up phenomena induced by heavy Ions in memories. It has been developed an equivalent model for circuital simulations in order to study Single Event Latch-up effects on CMOS technology. Then the proposed mitigation Latch-up scheme for memory it has been implemented in SRAM for 130nm technology node. The resulting system has been designed to part of a project for 3D integration using Through Vias Technology.

# INTRODUCTION

## CHAPTER 1: OVERVIEW OF 3D INTERCONNETION TECHNOLOGIES

The chapter presents a state of art in 3D technologies. Designers should know the general characteristics of 3D technologies in order to choice the most suitable for the specific application. It is necessary to set apart the approaches that realize 3D interconnections only on periphery (as the common Wire Bonding or variations of micro bump) from the technologies that realize a real 3D integration with very high integration densities. Among those it is possible to distinguish considering power consumption, packaging constrains, impact on fabrication process and cost. Moreover, it is also necessary to take in account the 3D limitations due to design resources, thermal management, and yield.

## CHAPTER 2: CAPACITIVE 3D INTERCONNECTIONS FOR HIGH SPEED APPLICATONS

The work focuses on 3D contact-less interconnections based on 3D capacitive coupling. Two different architectures have been developed and implemented on silicon prototypes; they are a Synchronous circuit and an Asynchronous circuit. Both the systems use the same type of assembly (i.e. face-to-face) and have been designed for high speed applications. The interface design aims to minimize the interconnection pitch in order to achieve high integration density and throughput. Moreover, developing very sensitive receivers and using appropriate mixes of high threshold and low threshold devices allows to drastically reducing the power consumption.

The research has been developed in collaboration with STMicroellectronics, as part of a project co-funding by the European Community (HIGH-TREE, IST2001-38931). The package assembly and support have been realized from Fraunhofer-IZM.

# CHAPTER 3: 3D CAPACITIVE TRANSMISSION OF ANALOG SIGNALS

Nowadays the SiP architectures mostly use digital implementations, but there are significant cases which require the transmission of analog signals. Among the several applications, two types are should be mentioned: i) low cost applications that do not have particular specification on resolution but need high rate of parallelism, i.e. high interconnection densities. Systems designed for transmission of high bandwidth analog signals where analog to digital conversion is not practicable or not cost-effective. The proposed solution aims to implement the 3D transmission of analog signals on the range of 100 KHz to 20MHz; thus it is suitable for the former case. Nevertheless, as it will be explained, the proposed architecture can implement communication at high frequencies at the cost of the redesign of a single block. The 3D communication is based on capacitive coupling and uses the same packaging assembly process developed for the systems described in the previous chapter. It will be explained the problem related to capacitive communication and then the proposed architectural solution. Further, the system specifications and related constrains are analyzed. They determine the project parameters in designing the several components. Finally, the input/output interfaces for testing are described, and the results on prototypes are shown.

The second part of the dissertation describes the research activity performed in collaboration with the CMP of Grenoble. The Multi-Circuits Projects laboratory is focused on developing system architectures for high performances systems. Specifically, applications involving Through Silicon Vias Technology is a matter of interest. The process technologies are furnished by Tezzaron which achieves high interconnection capabilities thanks to new techniques with pitches less than 4um. The project aims to integrate several SRAM memories on different layers connected with

TSV. It allows to realize a memory which presents twice or triple (and so on) the size (Kbits) of memories with almost the same footprint. It is of great interest both for general purpose applications and specific applications like military and avionics. Nevertheless, the TSV reliability has already been proven, thus we wanted to add an original contribute to the memories 3D stacking. The project also concerns the implementation and validation of a low-cost design for mitigation of Single Event Latch-up (SEL) induced by heavy ions memories. It is worth to notice that avoiding SEL is really important in avionic and space environments, but, since the devices are scaling down, it will be useful also common applications since the neutron induced up-set at ground level can result in similar threats. The modified SRAM will be integrated within two tiers connected with TSV technologies. Then the devices will be subjected to sources (i.e. Cyclotrons, tandem Van de Graff generator, 252Cf) providing heavy ions. The experiments aim to:

- Validate the realized Latch-up mitigation scheme for memories
- Analyze the effects of Latch-up on 3D integrated memories and compare the results with equivalent planar designs.

## CHAPTER 4: THROUHG SILICON VIA (TSV) TECHNOLOGY

This section presents an overview on TSV technologies. Different processes are analyzed and some interesting applications are presented. Finally the Tezzaron TSV technology is briefly introduced. The process steps and the achieved interconnection capabilities suggest using this specific technology to implements Logic-to-Memory integration or Memory-to-Memory integration (our case) instead of MEMS applications.

## CHAPTER 5: LATCH-UP PHENOMENA

The Single Event Effect (SEL) is introduced. The physical mechanism is not described in detail since it is not of interest for the current work, but the attention is focused on the processes that lead to Latch-up conditions. In order to study the effects on the CMOS circuits it has been necessary to extract an equivalent circuit starting from the physical model of the PNPN device. Moreover, since in literature there is not any standard approach to perform circuital simulations with latch-up events, it is very important determine the characteristic parameters and focalize on the critical operating points that characterize SEL. Typical I-V curve has been simulated for the given technology to extract the Triggering and Holding tension values. Then it has been possible analyze the relations between gains of the parasitic bipolar transistors, equivalent resistances (i.e. N-well distribute resistance, P-substrate distribute resistance, P and N tap distribute resistance) and Latch-up conditions.

## CHAPTER 6: SINGLE-EVENT LATCH-UP (SEL) MITIGATION SCHEME

This chapter describes the realized Latch up mitigation scheme. The basic principle is explained pointing out the advantages and back draws of the possible implementations. Then the system specifications are summarized and the different components are illustrated. Finally we present the simulations results.

The second part focuses on the description of the SRAM compiler and the generated memory. Layout design constrains the implementation of the proposed logic within the given memory. The solutions adopted to guarantee the flexibility of the system are presented within the modification made to the original memory.

# INTRODUCTION

# OVERVIEW OF 3D INTERCONNECTION TECHNOLOGIES

Performance of deep sub-micrometer very large scale integrated (VLSI) circuits is being increasingly dominated by the interconnects due to decreasing wire pitch and increasing die size. Additionally, heterogeneous integration of different technologies in one single chip is becoming increasingly desirable, for which planar (two-dimensional) ICs may not be suitable. An increasing number of integrated solutions involve the stacking of chips to reduce system size, to alleviate the interconnect related problems and to combine different technologies. Although in many applications is mandatory moving from planar design to three-dimensional chip design strategies, there are several benefits and drawbacks that should be taken in account. ICs designers must know the pros and cons of 3D technologies so that they can decide whether or not systems would work better as 3D IC and which approach is most suitable. Even if miniaturization is the primary motivation of the desire of vertical exploit, there are a growing number of reasons which justifies the cost through performance gains: realize very high bandwidth

communication systems, achieve efficient heat management and solve problems related to power consumptions.

The main strategies described so far are:

- Wire Bonded connections
- Micro-bump connections
- Contact-less communication
- Through Silicon Via Technology

## WIRE BOUNDED CONNECTIONS

The most common approach is wire bonded in which wires connect the individual die in a stack [1]. The connections between the chips go through the board or chip carrier and back to the other chip in the stack, or it bonds from chip-to-chip in the stack (Fig. 1.1). Typically, mechanical stresses require many metal layers to prevent tearing of the pad during bounding, and avoid pressure to destroy devices underneath the pad. Wire bound is already in wide-spread use in cell phones and continues to grown in low-power applications. Since the assembly is quite easy and this approach does not require many adjunctive steps respect to the standard process, the costs needed to implement this technology are not high. However, this approach is limited by the resolution of wire bonders and becomes increasingly difficult as the number of I/O in the chip stack increases. Moreover, due to large pitch and low integration density this technology is not suitable for those applications which require a large bandwidth. Unlike other 3D technologies, wire bonds are possible only on the chip's periphery, which severely limits the interconnect density and there are few benefits in the way of reduction of parasitic effects.

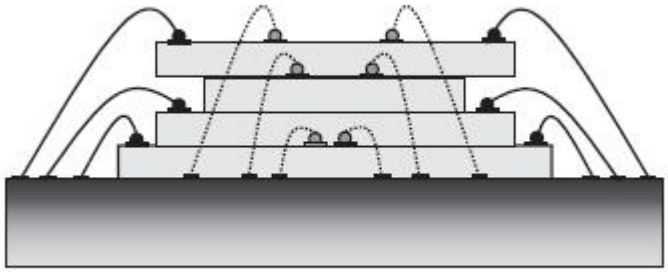


Fig. 1.1 Wire bonding technology

## MICRO-BUMP CONNECTIONS

Micro-bump technology involves the use of solder or gold bumps on the surface of the die to make connections. The mechanical stresses of assembly are much lower than with wire bonding. The bumps typically have a pitch of 50 to 500 um; anyway recently studies have proved that it is possible to have fairly smaller pitches. Back to face 3D package technology [2] involves embedding previously fabricated die into a set of carrier wafers with a fixed size, enabling engineers to assemble them into a tight cube. A layer of micro-bumps bond each die-carrier tier to an epoxy routing tier that brings signals to the edges of the cube. The tiers are fixed into a single stack, and then metallization is added to the sides to connect the routing tiers. The 3D package approach offers a much greater vertical interconnect density than the wire-bonded approach, but it does not significantly reduce parasitic capacitances because a micro-bump bonded cube must still route signals to the periphery before sending them back to the destination inside the cube. The number of tiers is not limited by the assembly process but rather by the heat inside the micro-bump cube.

Face to Face assembly technology offers the ability to shorten the wires between tiers and improves performance by reducing parasitic effects, this simultaneously decreases overall power consumption. In [3] it has been determined that, with proper placement of

blocks in the 3D architecture, it is possible to reduce the use of high-power dynamic logic circuits, repeaters, pipelined stages and long routing path. This approach is limited to two tiers and taking connections out of the chip stack requires the use of this technology in conjunction with a wire-bonded or through-via approach.

Two new 3D chip stacking technologies based on bumps have been recently proposed [4]:

- Wire-on-bump (WOB)
- Bump on-flex (BOF)

The WOB and BOF technologies are for low cost 3D stacking of memory chips by vertical side interconnection with metal wires and flex-circuits, respectively. These approaches aim to realize a high-performance vertical interconnection, such as through-via, while ensuring low cost in stacking process. In both those approaches, the process can be done on the wafer level to make multiple sets of back-to-back bounded chip.

WOB technology is for connecting stacked chips vertically with conductive wires, as shown in Fig. 1.2. Solder paste is first applied around peripheral solder balls/bumps of stacked chips. Then, wires are placed and bonded onto the solder balls/bumps during the solder re-flow. The solder paste should re-flow at a temperature lower than the solder balls/bumps on chips, so that the configuration of stacked chips is maintained without any collapse of the solder balls/bumps during the re-flow. For the wires in WOB interconnection, various low cost wires such as copper and aluminium with proper plating can be applied to obtain better wetting of solder paste onto wires and better thermo-mechanical reliability at the interconnection.

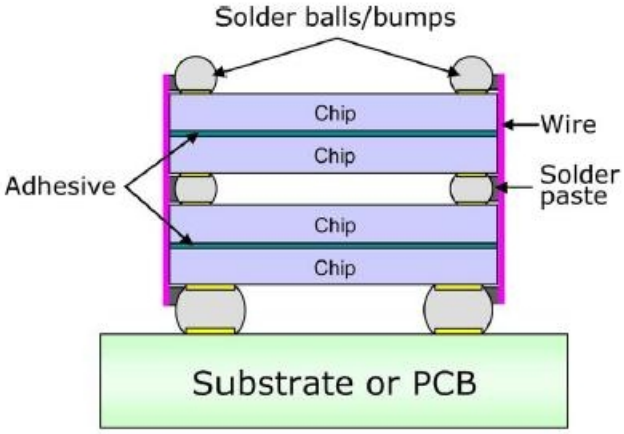



Fig. 1.2 Schematic cross section of 3D chip stacking by WOB

BOF technology, on the other hand, uses a flex circuit to vertically connect the chips (Fig. 1.3). This kind of connection includes metal patterns such as lines or pads on flexible organic materials. For increased component integration together with stacking, thin film and discrete components can be either embedded into or surface mounted onto the flex. Solders within the flex circuit are bonded to peripheral solder balls/bumps of stacked chips, leading to a BOF vertical interconnection. For this process, solder materials on the flex should be also selected to re-flow at a temperature lower than the solder balls/bumps on chips. The advantage is that stacked chips can be vertically connected by a single interconnection process using a flex circuit, while WOB chip stacking needs multiple interconnection processes with a multitude of individual wires. These technologies achieve better performance than wire-bonded technology since they allow denser communication lines integration. WOB and BOF interconnection also reduce the parasitic effects. In facts, they have been proved to have 64 mill ohm and 65 mill ohm respectively compared with 135 mill ohm presented by wire bonding case with similar size. Moreover, these technologies allow the possibility to stack multitudes of chips, either thin or thick, with a relatively low cost. However, WOB and BOF technologies can be difficultly compared to through-via technologies.

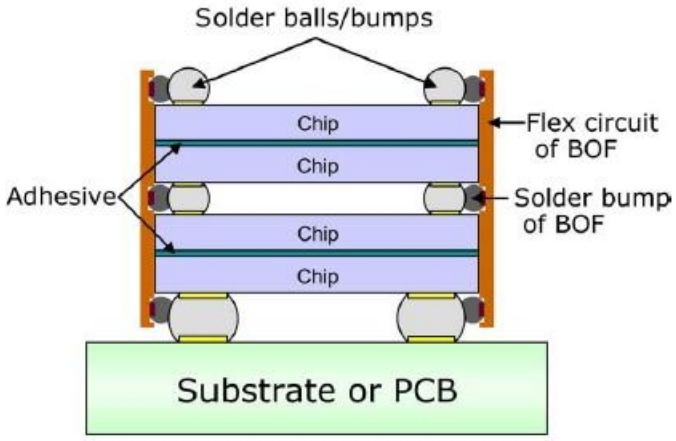



Fig. 1.3 Schematic cross section of 3D chip stacking by BOF

Since the inter-chip interconnection is made by vertical connections, it is necessary to bring the signals to the periphery. It means that the total electrical resistance and inductance are comparable with wire-bonded ones. All above, it has not proved the possibility to achieve a bandwidth as large as the bandwidth required by the high-performance multiprocessor systems.

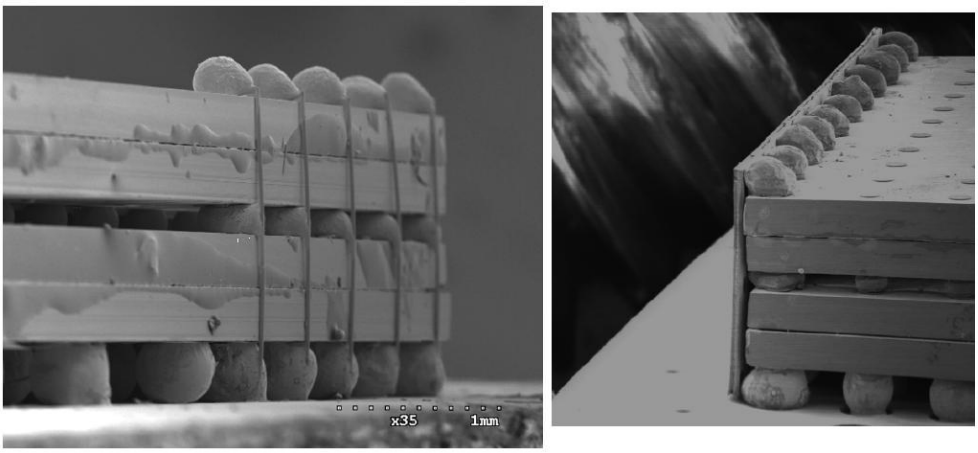



Fig. 1.4 SEM micrograph of WOB and BOF hip stacking

## CONTACT-LESS 3D COMMUNICATIONS

Contact-less or AC-coupled interconnection involves the use of capacitance or inductive coupling to communicate between tiers. This approach eliminates the processing steps for creating inter-tier DC connectivity and eliminates the need to route signals to periphery, allowing the reduced wire lengths. It requires only a minimal amount of processing for chip thinning; moreover the lack of specialized processing steps makes it much cheaper than micro-bump and through silicon via approaches. The main strategies described so far are: interconnections based on inductive coupling and on capacitive coupling.

Inductive coupling is more favorable for situations where the separation of the coupling elements, which is determined by the chip thickness, approaches the lateral dimensions of the coupling elements (Fig. 1.5). These 3D systems that use inductive coupling for tier-to-tier communications and wire bonding to provide DC power and external interfacing are inexpensive and relatively easy to construct.

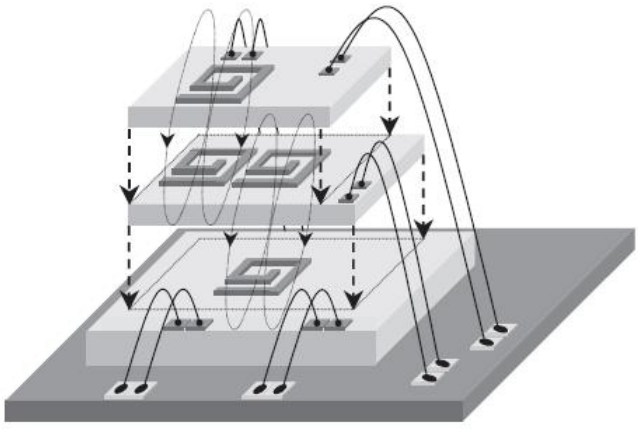


Fig. 1.5 Contact less technology based on inductive coupling

In [5] it has been developed a transceiver for inductive-coupling, by using a pulse-shaping circuit. It can be applied to more than two chips stacked in whether face-to-face, face-up/down, or even back-to-back, which provides high feasibility in assembly and higher scalability in system integration. One of the biggest challenges in the inductive-coupling transceiver is the energy reduction: pulse shaping reduces transmitter's energy dissipation and device scaling reduces receiver's energy dissipation. The transmitter chip should be thinned down to 10um thickness, as a result, communication distance between the transmitter and the receiver is 15um, including a glue layer of 5um. The data transceiver communicates at 1Gb/s/ch by a metal inductor with a diameter of 29um. 1024 data transceiver are arranged with a pitch of 30um, which yields aggregated data rate of 1 Tb/s (90nm CMOC technology). The keywords to minimize energy dissipation are: Pulse shaping where the pulse width is minimized for the transmitter's energy reduction. Robust Timing Design against PVT variation to maintain BER even with the narrower pulse.

Device scaling where the receiver's energy dissipation can be effectively reduced. The Test-Chip's performance is summarized in Table I [5].

**Table I**

**Performance of the transceiver for inductive coupling**

| **PROCESS** | **90nm CMOS; VDD = 1V** | **180nm CMOS; VDD = 1.8V** |
|---|---|---|
| **Energy Dissipation in TX/RX, $E_{TOT}$** | **0.14pJ/b** | **0.33pJ/b** |
| **Energy Dissipation in TX, $E_{TX}$** | **0.11pJ/b** | **0.13pJ/b** |
| **Energy Dissipation in RX, $E_{RX}$** | **0.03pJ/b** | **0.20pJ/b** |
| **Data Rate** | **1Gb/s** | |
| **Bit Error Rate** | **< 10e-12** | |
| **Clock Rate** | **1GHz** | |
| **Channel Area** | **30um x 30um** | |
| **Inter-Chips Distance** | **15um** | |

The inductive approach is current controlled and the transmission power can be tuned in order to deal with assembly issue, moreover it presents advantage in terms of packaging. On the other hand, the capacitive approach optimizes interconnection performance and area if a high-precision face-to-face assembly is available. In the contact-less communication based on capacitive coupling, dies are stacked face-to-face leaving the I/O pads uncovered, so that they can be further connected using wire-bonding (Fig. 1.6). Electrodes realized in the upper metal layer of each chip provide capacitive coupling between TX and RX circuits. For mechanical stability and inter-chip isolation, a 1um layer of acrylate-based adhesive is sandwiched between the vertically interconnected dies, with an alignment accuracy of ±1um.

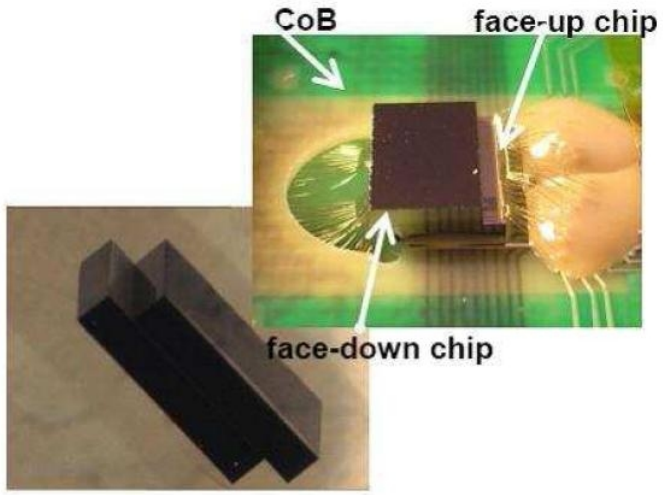



Fig. 1.6 Packaging and bonding of chips used for 3D integration based on capacitive coupling

The interface design aims to minimize the interconnection area and power consumption by implementing a very sensitive receiver. The interconnection attenuation depends of the ratio between the inter-electrode coupling capacity (C3d) and the

parasitic coupling to ground capacity (Cp). In order to reduce the interconnections pitch, the C3d is scaled down and the Cp doesn't scale with the same ratio, thus a receiver that can manage a significant attenuation is mandatory.

Depending by the applications two different topologies of circuits are implemented: Synchronous Circuit and Asynchronous circuit. The first communication approach exploits synchronism in order to reduce size and guarantee communication reliability. The second scheme significantly improves mono- and bi-directional transmission capabilities and it does not need any synchronization. The development of those communication systems has been one of the themes of the presented dissertation and they will be explained in the next chapter.

## THROUGH SILICON VIA TECHNOLOGY

Through-via interconnection has the potential to offer the greatest interconnect density but also the greatest cost. Assembly occurs at the wafer level, placing a second wafer face down on the first wafer (face-to-face) and subsequent wafers face down (face-to-back) as the number of tiers grows. The manufacturing process then etches holes through the upper wafer into the lower wafer and fills them to provide connectivity. Before placement of the next chip, the backside of the previously etched chip is thinned by polishing. The top tier has tungsten Vias that protrude along with cuts for bond pads to provide power, ground and I/O interconnectivity. The assembly process in Through-Via approach does not limit the number of possible tiers: rather, heat inside the stack is the limiting factor. Also, in this approach, the dies are not known to be good before assembly, so it is possible to attach a good die to a faulty one, making it necessary to reject the entire assembly. In such a situation, yield drops quickly with the addition of more tiers. Bulk technologies have demonstrated through-via interconnection by first coating the hole with an insulator, with achieved pitches of 50um. Silicon-on-Insulator technologies (SOI) avoid the need for passivating the hole

by polishing the substrate away completely, down to the buried oxide. SOI technologies have achieved the smallest inter-tier pitches yet, on the order of 5um (Fig. 1.7). Area of interest for implementation is the integration of MEMS and CMOS. Via technology provides new possibilities for such integration although yield factors for particular designs need to be carefully considered when integration method (wafer level or die level) is chosen. For many sensing devices such as flow- and pressure sensors, via technology simplifies packaging and connection to the chip.

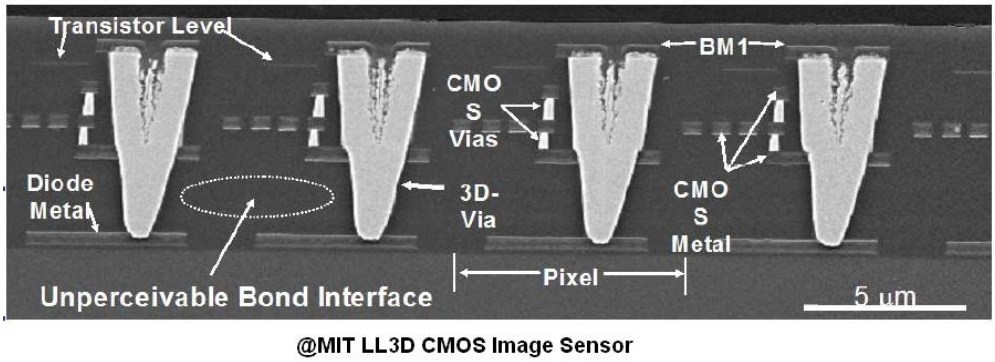



Fig. 1.7 Trough Silicon Via technology cross sections

The connection to the sensing chip can be made at the chip side facing away from the media while having the sensing part of the die in direct contact with the measured media. Using via technology in micro-fluidic devices enables creation of electrodes in the channels, and at the same time it allows the integration of more additional MEMS features such as fluidic filters, high aspect ratio structured channels and silicon to glass bonding.

The TSV technology, due to the extraordinary density of integration capability, is of great interest in every application which requires very high bandwidth. Routing and flip chip mounting is accommodated on the back side of the die beneficial for example in imaging applications. Via technology is applicable when using the System in Package

(SiP) approach, integrating a number of integrated circuits in a single module. The SiP can perform many of the functions of an electronic device, such as a mobile phone. This feature is particularly valuable in space constrained environments like mobile phones as it reduces the form factor and complexity of the PCB and overall design. An example of using TSV technology for memory integration will be described in the last part of the dissertation, since it has been a theme of the presented study. It has been developed as part of a larger project performed in cooperation with Circuits Multi-Projects INPG, Chartered Semiconductor Manufacturing LTD and Tezzaron Semiconductor. This work concerns the implementation, validation and evaluation results of a recent low-cost design for SEL mitigation scheme for memories. Moreover, the mentioned design is implemented in a 130nm process node destined for 3D integration aiming at validating various issues concerning implementation of 3D systems.

# LIMITATIONS OF 3D PACKAGING TECHNOLOGY AND SOLUTIONS

The 3D technology offers advantages for all types of electronic assemblies: computer, military, automotive and telecommunication applications. However, there are trade-offs which need to be taken into account when using 3D technology in system design [6].

## THERMAL MANAGEMENT

As the demand increases to build high performance systems, trends in electronic package design have moved toward larger chips, higher number of I/O ports, increased circuit density and improved reliability. Greater circuit density means increased power

density (W/cm2), so thermal management should be considered carefully. It has been addressed in two levels. The first is at the system design level by trying to evenly distribute the thermal energy across the 3D device surface. The second is at the package level. This is achieved through the combination of several techniques: by using low thermal resistance substrates such as diamond or chemical vapor deposition (CVD) diamond, by using forced air or liquid coolant to reduce the 3D device temperature. Finally, by using a thermally conductive adhesive and implementing thermal Vias between the stacked elements to extract the heat from inside the stack to its surface.

## DESIGN COMPLEXITY

Increased functional integration has lead to larger chip sizes, which has required materials development for increased wafer size and equipment development for handling larger wafer. A large number of systems have been implemented using 2D form and have demonstrated that such complexity can be managed. However, a few number of system and devices have been implemented using 3D technology, proving nevertheless that such devices or system are manageable, although complex.

## COST

With the emergence of any new technology, there is an expected high cost involved in using it. This is also the case with the 3D technology, due to the lack of infrastructure and the reluctance of manufacturers to change to new technologies for reasons associated with risk factor. Moreover, such cost is a function of the device complexity and the requirements.

## DESIGN SOFTWARE

Most manufacturers use their own design tool kits, which give the designers the ability to implement their design in accordance with the vendor's manufacturing requirements, while allowing the designers to focus on the design without getting involved with manufacturing and interfacing details. However, most of these design kits are not fully integrated or implemented in software that is easily accessible. Hence, for some of the manufacturers there is a need to port their design rules into available software or alternatively the costumer may buy their own software. In the first case there is a time and risk involved, while in the second the cost involved is the main issue. It's also worth to be noticed that new tools that consider thermally aware physical design implementation are crucial to the success or 3D integration. This requires strong linkage between architecture level analysis tools and 3D physical planning tools. It's needed both to optimize the 3D system implementation (3D place and route algorithms) and to adequately consider the physical impact of vertical Vias (area, latency, thermal impact).

## YIELD MANAGEMENT

This is another issue that has been little discussed. If the field of a single die at wafer level is 90%, then yield of a die from two stacked wafers would be 81%, from three wafers 73% and so on. This would create a clear financial disincentive for 3D technology. It is well known that yield as a quadratic dependency on die size and a linear dependency on chip count at a given die size. 3D design may incur some die loss due to vertical vias, and may be gain some yield due to density. Fortunately, there are at least two ways to alleviate this issue. One approach would be to use small die that have high yield. Another would be to use a die on wafer 3D integration approach, rather than wafer on wafer. If the separated die were even partially tested and sorted, the yield would be that of undiced wafer die.

# CAPACITIVE 3D INTERCONNECTIONS FOR HIGH SPEED APPLICATIONS

As reviewed in the previous chapter, several solutions for vertical interconnection have been proposed, among which it is worth to notice wireless inter-chip connections. The absence of Ohmic contacts allows the omission of ESD protections, reducing parasitic effects and improving overall performance. That technology is also valuable for heterogeneous process integration where different devices (i.e. digital components, analog components...) can be placed in different dies to avoid interferences. The main strategies described so far are: interconnections based on inductive coupling [9] [7] and on capacitive coupling [13] [4]. In the former approach communication is current controlled and the transmission power can be tuned in order to deal with assembly issue. It presents an advantage in terms of packaging, in fact chip are stacked and aligned face up, avoiding more critical manipulations. On the other hand, the capacitive approach optimizes interconnection performance and area if a high-precision face-to-face assembly is available.

The presented study focuses on contact-less communication based on capacitive coupling, the principle of the presented 3D technology is sketched in Fig 2.1. Dies are stacked face-to-face leaving the I/O pads uncovered, so that they can be further connected using wire-bonding. Electrodes realized in the upper metal layer of each chip provide capacitive coupling between TX and RX circuits.

## CHAPTER 2: CAPACITIVE 3D INTERCONNECTIONS FOR HIGH SPEED APPLICATIONS

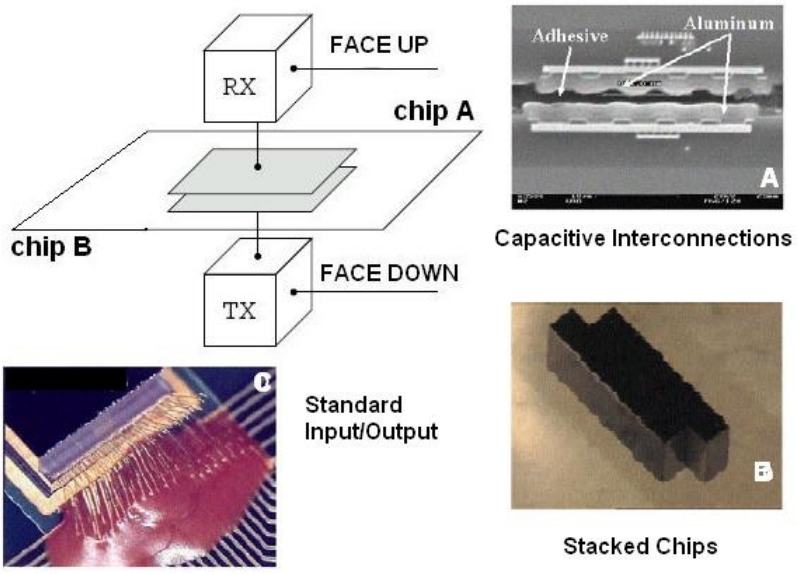



Fig. 2.1 3D interconnection scheme based on capacitive coupling. a) Cross section of the capacitive interconnections, b) Assembly packaging of the stacked chips, and c) bonding

The dissertation describes two different topologies of circuits:

- 3D Synchronous communication system
- 3D Asynchronous communication system

The first communication approach exploits synchronism in order to reduce size and guarantee communication reliability. The measurements demonstrate that a high bandwidth density (Maximum bit rate/Electrode area) is achieved [1] along with an energy consumption reduction by 38% with respect to previously works [14]. On the other hand, the Asynchronous scheme is valuable for applications that do not require a synchronization system and it significantly improves mono- and bi-directional transmission capabilities.

# SYNCHRONOUS COMMUNICATION CIRCUITS

Fig. 2.2 and Fig. 2.3 show TX and RX circuits respectively. The design has been carried out with the aim of minimizing the interconnection area by providing a very sensitive receiver circuit and a communication scheme that maximizes the transmission energy. The presented system is suitable in order to deal with the trade-off between receiver performance and electrodes size. Furthermore a dynamic topology provides high speed functionality, while an appropriate mix of high threshold (VTH) transistors (depicted with thicker gate) and low threshold (VTL) transistors enable to manage leakage current and preserve the correct functionality in low speed operating condition.

## TRANSMITTER

The transmitter is based on a 2-stage dynamic circuit: when the clock CKt is low, the output node (Qt) is pre-charged to VDD. When CKt rises, the transmitter input stage, formed by P1, P2, P3, N1, N2 MOS, samples the Dt value on CKt rising edge: the input data is evaluated and Qt voltage falls according to the sampled data while the state is latched at the dynamic node int_TX. The critical path consisting of the discharge of the output node has been optimized by an appropriate transistor sizing. That sequence is formed by P1-P2-P3 plus N2-N3: transistor P5 is over-sized in order to provide a large current so as to guarantee a quick switch to pre-charge phase, whereas a minimum size is enough for P4 and N3. If the sampled data is ’1’, Qt does not fall during the evaluation phase and its value is dynamically kept; N4 is a VTH device in order to prevent undesired voltage drop on the electrode in high impedance state because of the pull-up leakage due to P5 (VTL) is orders of magnitude larger than the pull-down leakage due to N4. Moreover, P4 and N2 are in feedback configuration which preserves the correct behavior of int_TX. When its voltage is high, an active feedback through P4 (on) makes leakage negligible. On the contrary when Dt changes to a logic ’0’ still in

the evaluation phase, int_TX is driven to high impedance state and N4 is switched off as the pre-charged value is to be kept on Qt node.

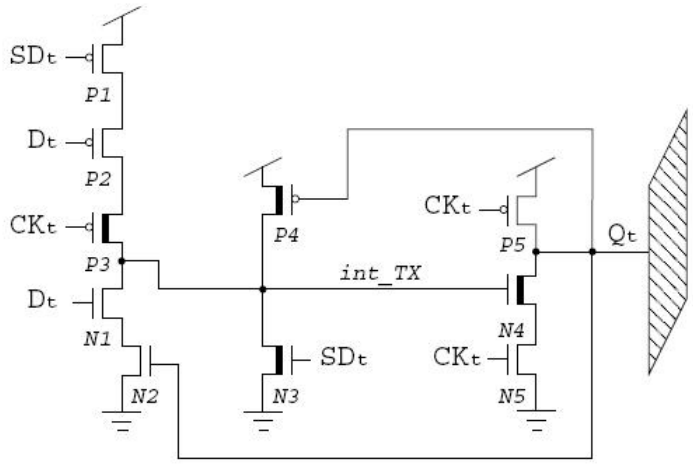



Fig. 2.2 3D Synchronous Transmitter (TX)

It is crucial to avoid that leakage currents partially charge int_TX and lead to an incorrect output value. According to it the pull down network formed by N1, N2 is implemented with low threshold devices, thus it has a much larger leakage than P3 and P4. Moreover P1 and N4 enable a synchronous preset of transmitter, controlled by the active-high signal SDt, by forcing the first stage of the circuit to drive low int_TX.

## RECEIVER

The receiver is implemented with a pre-charge and evaluation scheme that matches the transmitter topology. When the clock (CKr) is low, the input node (Dr connected to the 3D electrode) is pre-charged to VDD; during this phase, the first stage

of the receiver (P6, N6 and N7) is disabled while the subsequent stages keep the value received during the previous clock cycle. In particular the stage formed by P8, P9, N9, N10, and N11 is a latch scheme which keeps the QNr low value in low impedance mode while the high value in high impedance mode by the leakage current. The evaluation of received data takes place when CKr is high: during this phase Dr is kept in high impedance state, so that the electrode voltage can change according to the signal coming from transmitter. If the transmitted value is ’1’, then the transmitter does not generate any transition along the vertical channel thus Dr is left high; this high level is guaranteed despite leakage on Dr since pull-down leakage (N6) is lower than the pull-up leakage (P6). If the transmitted vale is ’0’, a falling edge is generated by transmitter and is propagated to Dr with an attenuation that depends on the channel characteristics; then its value is amplified by the subsequent stage and finally latched by the output stage. N6 provides a feedback that drives low the electrode voltage if a low level is correctly received (Dr falls over the logic threshold of receiver). This circuit does not need a dedicated reset signal to set its state; if the transmitter is in reset mode, than the receiver output QNr keeps a logic ’0’ after one CKr cycle.

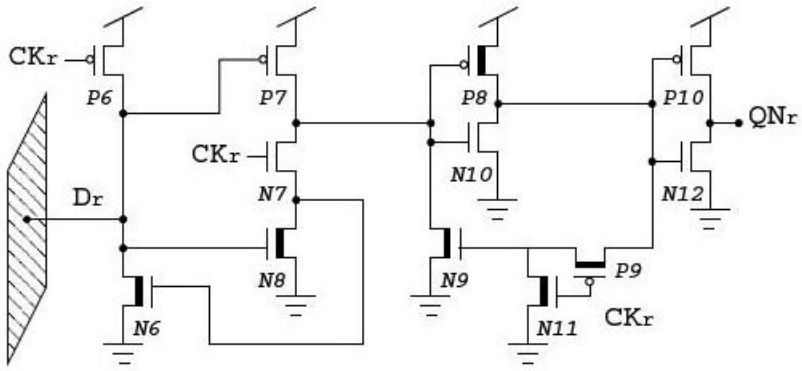



Fig. 2.3 3D Synchronous Receiver (RX)

## INTER-CHIP SYNCHRONIZATION

In order to guarantee functionality and performance, TX and RX need to be synchronized. A clock signal is transmitted through a capacitive interconnection by using asynchronous circuits. For synchronization purposes the clock is propagated from the receiver chip to the transmitter one. This approach guarantees the required skew between CK_RX and CK-TX. During the initialization phase, output node Qt and input node Dr are both pre-charged to VDD. Afterwards the evaluation phase of the transmitter is delayed in respect to the evaluation phase of the receiver. . When the transmission edge is generated the receiver is already out of the preset phase and ready to sample it [1]. In order to propagating the clock also in opposite direction, a delay block has been added. Thanks to this clock tree implementation, it is possible to under (or over) compensate the skew according as propagation from the receiver or from the transmitter respectively.

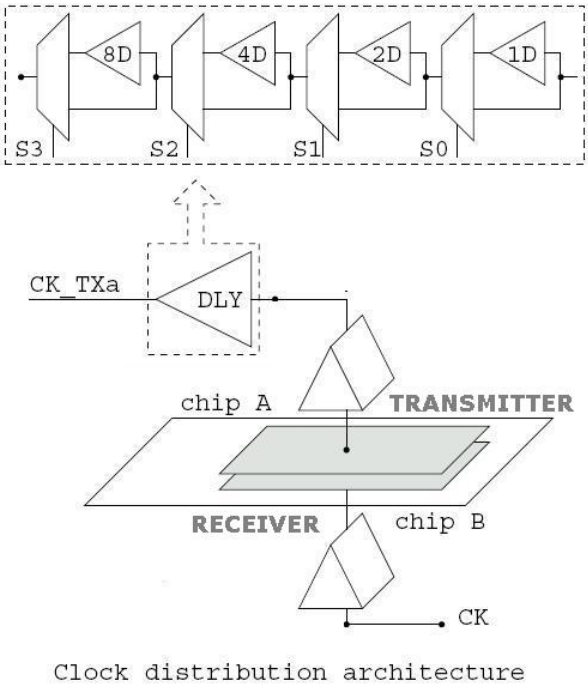



Fig. 2.4 3D clock architecture with programmable delay

Furthermore the clock distribution architecture includes a latency modulator circuit, as shown in Fig. 2.4, which allows to improve the performance of the system. In fact programmable delay can be obtained, for either chip, by using 4 digital signals (S0, S1, S2, and S3). More precisely, each configuration adds a different number (from 0 to 15) of a basic delay unit (50ps in the presented silicon implementation) to the corresponding clock path. A setting procedure tunes the latency of various clock paths in order to determine the best inter-chip clock skew; this enables a further optimization of performance.

## MEASUREMENT RESULTS

TX and RX circuits were included in a test chip fabricated in 0.13um, 1.2V CMOS technology. The test chip and the corresponding 3D package are sketched in Fig. 2.5 left and Fig. 2.5 right respectively. The prototype contains TX and RX circuits, arranged in arrays and connected to electrodes of different sizes, the synchronization logic, and a test interface that enables low speed input-output together with high speed vertical data transmission. The communication circuits are placed across an axis of symmetry in order to allow assembly of the system at die level. Mechanical stability and inter-chip isolation are provided by a 1um thick acrylate adhesive layer sandwiched between the chips. Chips are stacked face-to-face with 1um alignment accuracy. The package assembly has been performed by Fraunhofer-IZM. Further improvements in assembly techniques will result in enhancement of system performances. Reducing misalignment and inter chip distance will increase the capacitive coupling of the transmitting channel. It means the possibly faster data transmission or else it allows to have the same speed performance but with reduced electrode sizes; thus with higher density of connections and less power consumption.

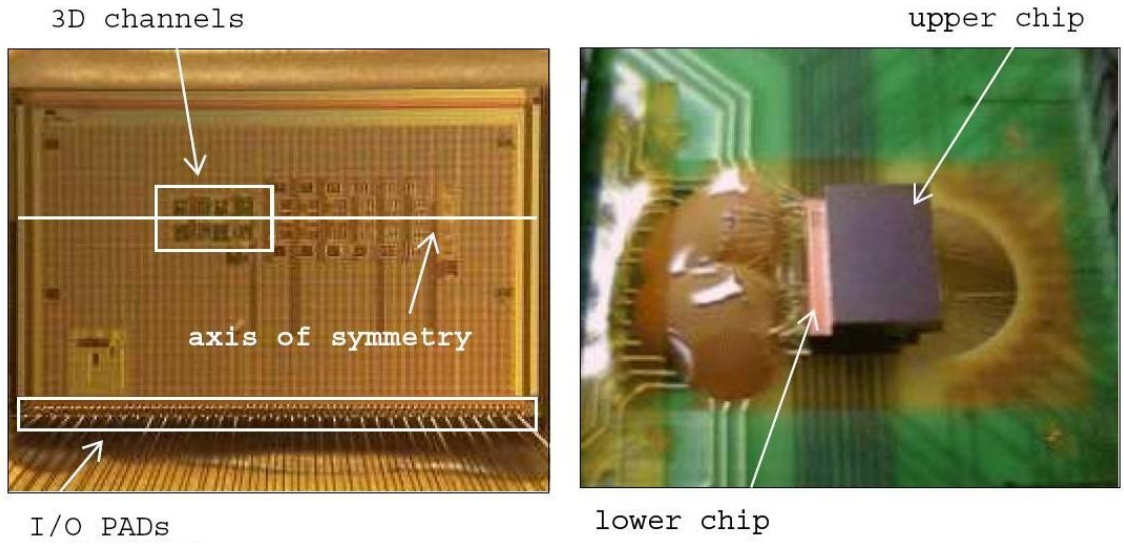



Fig. 2.5 0.13 um test chip

.

This silicon implementation demonstrates a throughput per pin of 900Mbps with electrodes area down to 8x8um2. Bandwidth measurements have been performed for three electrode sizes. Results of maximum bit-rate and propagation delay are shown in Table I. With the present implementation, a maximum bandwidth of 975 Mbps has been obtained with 15x15um2 electrodes. While, the 25x25um2 interconnections are affected by large parasitic effects, consequently the bit-rate is significantly reduced. About 8x8um2 channels, simulations show the best performance as summarized by the data propagation time, but noise margins are reduced by the low inter-chip coupling.

**Table I**

| SIZE | Max. Rate | T_data |
|---|---|---|
| 25x25um2 | 795Mbps | 304ps |
| 15x15um2 | 975Mbps | 256ps |
| 8x8um2 | 900Mbps | 240ps |

This effect explains the maximum rate of 900 Mbps. Anyway, test on silicon realization demonstrates a wide communication bandwidth, with full functionality down to 8x8um2 electrodes, which proves a great sensitivity of the receiver circuit. The measurement waveforms are show in Fig.2.6.

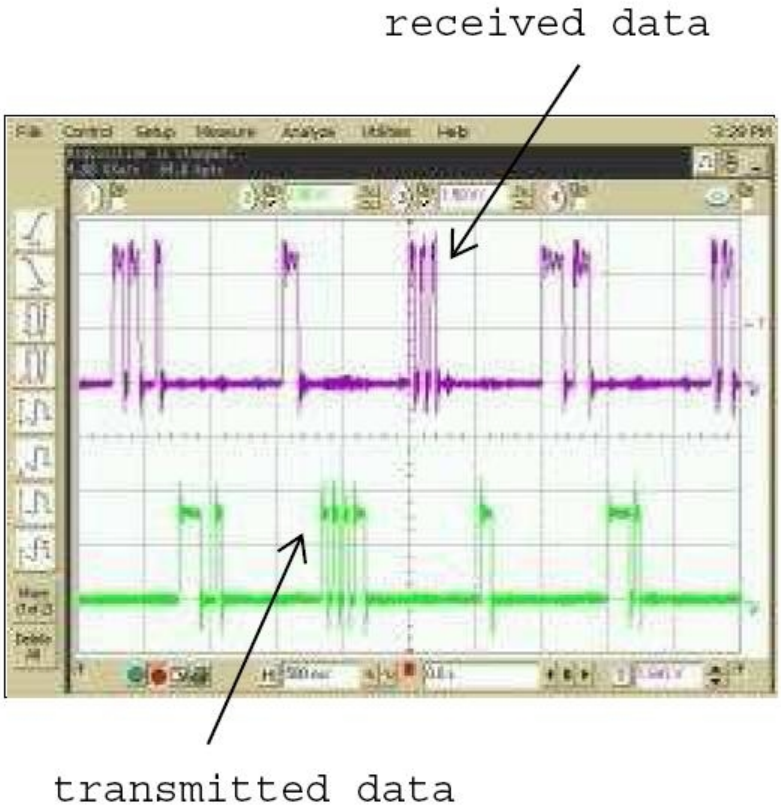



Fig. 2.6 Data transmission waveforms (measured waveforms)

The power consumption is estimated by using SPICE simulation (with extracted of capacitive parasitic) together with an optimization of 3D parameters (inter-electrode coupling) obtained by measurement results of silicon implementation. Table II summarizes power consumption characteristics. The dynamic power contribution is comparable to the one related to standard on-chip buffers; the static power is an aggregate of leakage and penetration current. The dynamic contribution increases with

electrode size (since it mainly depends on electrode load). The static term has opposite trend: increasing coupling we reduce voltage attenuation and the result is a reduced penetration current in receiver stage. Finally, this last contribution could be decreased changing clock duty-cycle in order to minimize the pre-set period. In the low-frequency range the static contribution represents a large percentage of the overall consumption, but this is a direct consequence of the sensitivity target.

**Table II**

| SIZE | Dinamic Power | Static Power |
|---|---|---|
| 25x25um2 | 0.13uW/MHz | 15uW |
| 15x15um2 | 0.11uW/MHz | 17uW |
| 8x8um2 | 0.10uW/MHz | 23uW |

It is worth to note that the energy consumption related to the transmission of logic ’0’ is much larger than the transmission of logic ’1’ (no commutation with respect to pre charged state); the average power consumption is 45uW/pin@900MHz (50fJ/bit).

# ASYNCHRONOUS COMMUNICATION CIRCUIT

The synchronous communication interface described above achieves multi-Gbps aggregate bandwidth and low power consumption. Nevertheless, it presents some disadvantages; in fact communication is constrained by pre-charge and evaluation scheme and demand of synchronization. Thus, it has been developed a scheme based on asynchronous communication. This circuit improves performance in terms of energy consumption and data throughput using a mix of multi threshold process features, high threshold (VTH) and low threshold (VTL) transistors to manage leakage current and preserve the correct functionality in low speed operating condition. Furthermore it

enables either mono- or bi-directional transmission capabilities providing more flexibility in system design. Circuits for asynchronous interconnections are sketched in Figure 2.7. The transmitter is a buffer that drives the electrode.

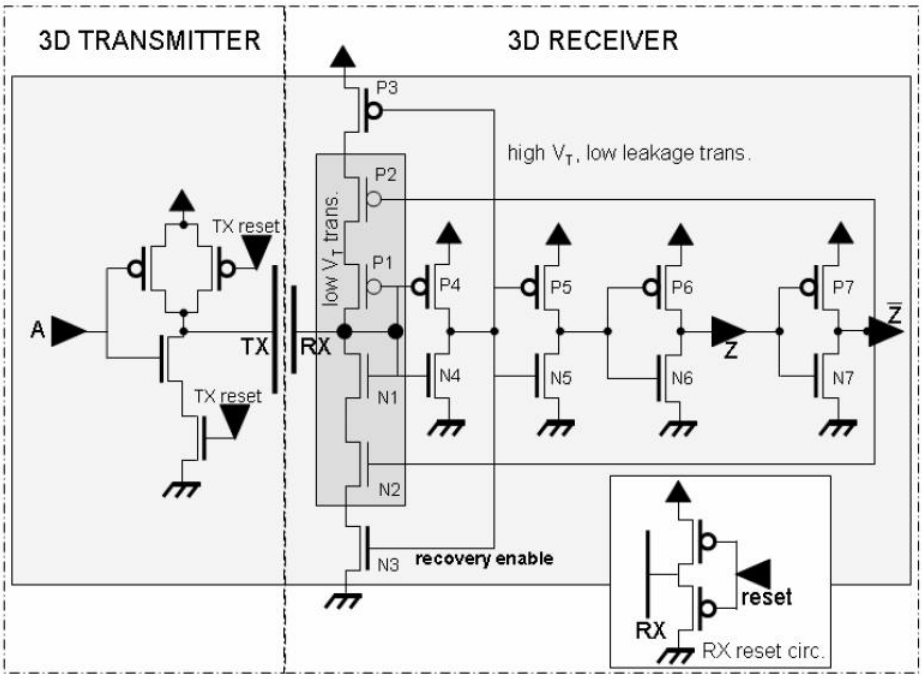



Fig. 2.7 Mono-directional AC interconnection for asynchronous communication

The receiver implements a double feed-back topology that provides voltage recovery as well as high impedance for the AC input. Within the receiver, the inverter N4-P4 amplifies the input signal and generates the Recovery Enable signal. This determines a positive feedback on the electrode (transistors N3, P3) that recovers the correct voltage level for the related logic state. After an appropriate delay, the received signal reaches *Z* and the feedback on the receiver electrode is switched off (N2, P2), so the electrode is finally left in high impedance state. It has been included the diode connected transistors P1 and N1 in order to reduce the voltage swing of the electrode and keep it close to the logic threshold of the amplification stage (N4-P4) while limiting

the direct leakage current. For a further reduction of static consumption, P1 and N1 are low-VT (VTL) transistors while N4 and P4 are high-VT (VTH) devices, so the leakage of the amplification stage is reduced by obtaining a positive cut-off (VTH-VTL) for N4 and P4. The use of devices with different threshold guarantees the receiver stability despite leakage on the high impedance input node: when RX is high (low) P3 is on and P2 is off (N3 is on and N2 is off), so the leakage of the pull-up (pull-down) branch is due to the low-VT device P2 (N2); in the same condition, N3 is off and N2 is on (P3 is off and P2 is on), so the leakage of the pull-down (pull-up) network is due to the high-VT device and is smaller than the one related to the complementary branch. This mechanism prevents wrong commutations even at low frequency. Furthermore, receiver electrode is smaller than transmitter electrode; this results in enhanced inter-die coupling and lower sensitivity to alignment while limiting the size of the receiver electrode and its parasitic coupling to ground.

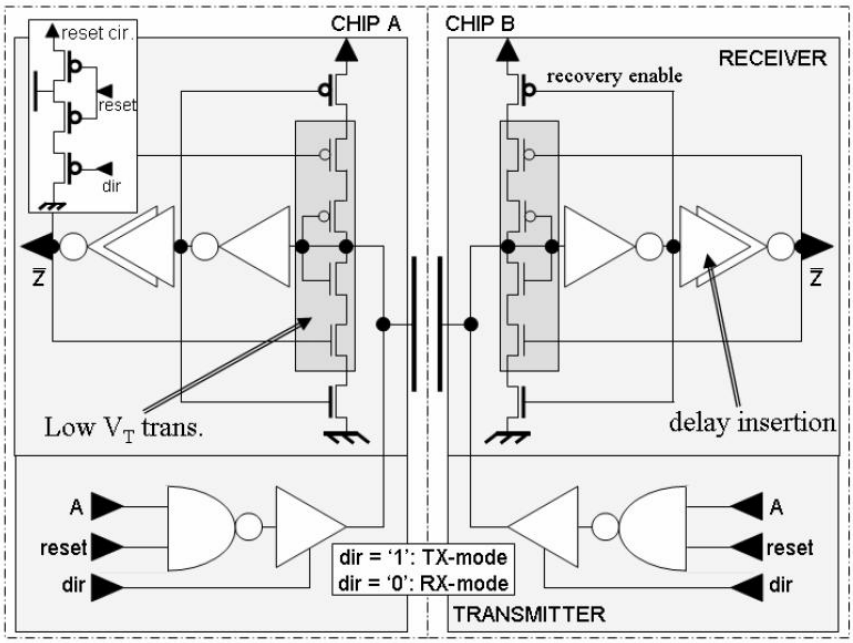



Fig. 2.8 Bi-directional AC interconnection

The interconnection described so far is suitable for a bi-directional realization as well; this option is sketched in Fig. 2.8. For this solution to work properly, the transmission buffer is replaced by a tri-state one that does not drive the electrode in reception-mode. The receiver sub-circuit drives the electrode during voltage recovery only, so it does not require any modification. When the transmission-mode is selected, the circuit dedicated to RX initialization is switched-off in order to prevent additional power consumption.

## EXPERIMENTAL RESULTS

The proposed communication circuits were included in a test chip fabricated in 0.13µm, 1.2V 6metal double threshold transistors CMOS process. Each 3D prototype includes 216 AC links with an overall consumption of 210mW@1.2GHz, including test structures and PLL for the generation of high-speed clock. Design summary is reported in Table III.

**Table III**

| | |
|---|---|
| DESIGN SUMMARY | **0.13 um CMOS by STMicroelectronics, 1.2V** |
| | **108 Mono-directional links in 3x3 Arrays** |
| | **108 Bi-directional links in 3x3 Arrays** |
| | **PLL for clock generation** |
| | **Test-structures for random pattern test** |
| | **Test-structures for clock transmission test** |

The silicon prototype contains TX and RX circuits (3D channels) arranged in arrays and connected to electrodes of different sizes. The communication circuits are placed across an axis of symmetry in order to allow assembly of the system at die level. Stacked chips are mounted on a Printed Circuit Board (PCB) by using a standard chip on board approach: face up chip is wire bonded to the top side of the board while face down chip is wire bonded to the opposite side of board through a PCB slot. The package assembly technology is the same previously described and it has been developed by Fraunhofer-IZM.

The test chip includes a PLL and dedicated test structures as frequency divisors to perform data and clock transmission through electrodes of different sizes with different operating frequency. In addition a dedicated test circuit enables high speed propagation of pseudo-random generated pattern in order to measure the Bit error rate during vertical contact-less communication. The integrated characterization structures as well as the measured waveforms are sketched in Fig. 2.9. For the mono-directional implementation, the transmission of clock signals up to 1.7GHz was verified with electrodes down to 6x6μm2 (TX electrode is 8x8μm2) and correct functionality after more than 1013 cycles; larger electrodes enable clock transmission up to 2.46GHz. The speed and power depend on the trade-off between parasitic coupling to ground and inter-die coupling: the fastest implementation uses 15x15μm2 RX electrodes and leads to a 380ps propagation delay with a power consumption of 0.15μW/MHz plus 1.6μW static. Bi-directional circuits show a slightly reduced performance because of the increased circuit complexity. The extra-parasitic effects, due to the implementation of a tri-state buffer on the RX node, result in a reduced sensitivity of the receiver. By increasing the electrode sizes, it is possible to improve the capacitive 3D coupling. Anyway, it does not substantially improve the ratio between transmission coupling and parasitic coupling to ground, since also the last increases. The fact explains why the bi-directional scheme cannot reach the performance of the mono-directional approach; in this case the peak frequency for clock transmission is 1.8GHz with a 20x20μm2 electrode area.

However, the possibility to transmit and receive signals on the same transmission channel can be an advantage and guarantee more flexibility design in several applications.

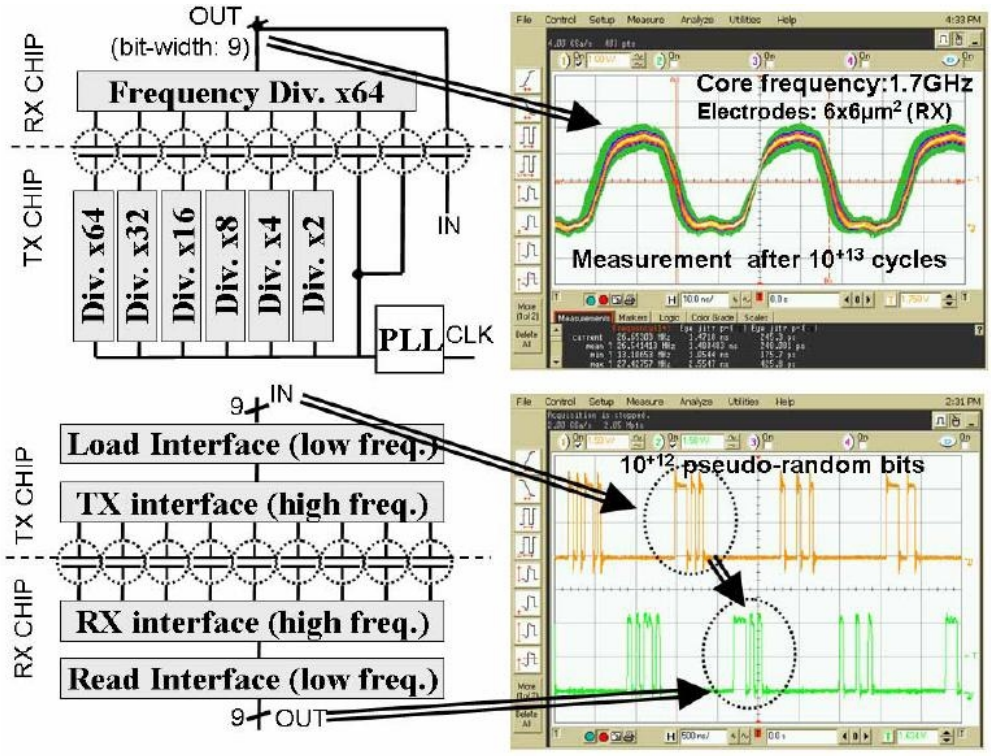



Fig. 2.9 Test structures and measurement waveforms

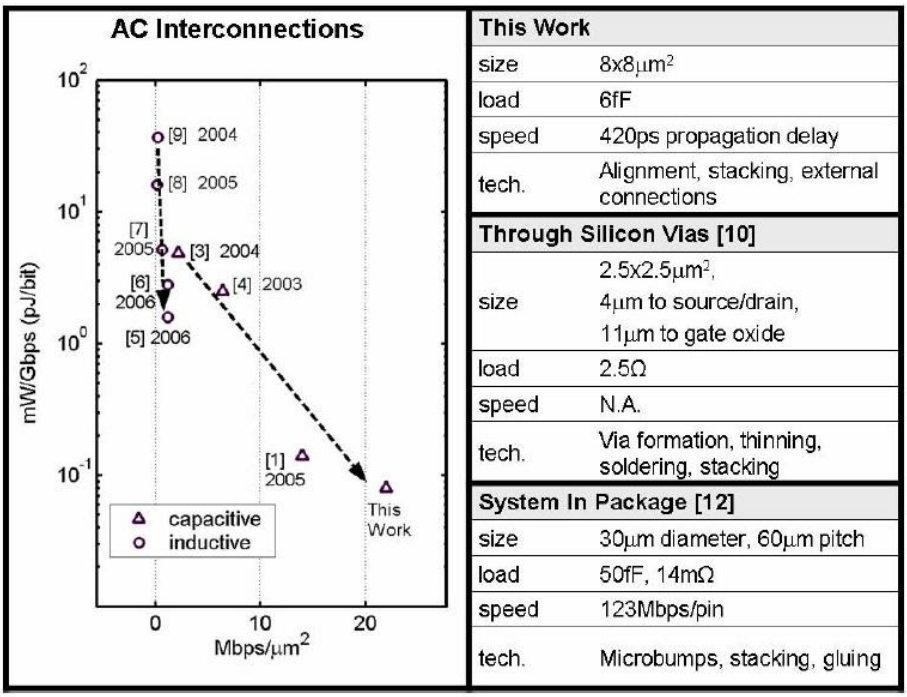
A propagation delay of 560ps was estimated for these structures by a test on more than 1012 pseudo-random bits with no error reported (BER <10−12) and the power consumption is 0.22μW/MHz plus 6μW static. Since asynchronous interconnections are really equivalent to standard buffers, their communication speed is fully summarized by the propagation delays in Table IV.

**Table IV**

| MONO-DIRECTIONAL | | | BI-DIRECTIONAL | | |
|---|---|---|---|---|---|
| **SIZE** | **DELAY** | **CLOCK TX** | **SIZE** | **DELAY** | **CLOCKTX** |
| **RX: 6um TX: 8um** | 420ps | 1.7GHz | **15um** | 615ps | 1.5GHz |
| **TX: 8um RX: 10um** | 395ps | 2.3GHz | **20um** | 560ps | 1.8GHz |
| **TX: 10um RX: 15um** | 385ps | 2.35GHz | **25um** | 610ps | 1.65GHz |
| **TX: 15um RX: 20um** | 380ps | 2.46GHz | **29um** | 640ps | 1.32GHz |
| **POWER** | **0.15uW/MHz + 1.6uW** | | **POWER** | **0.22uW/MHz + 6uW** | |

# 3D SOLUTIONS COMPARISON

Figure 2.10 highlights the comparison between several 3D solutions; it points out the performance trends of capacitive and inductive 3D interconnect technologies and reviews the advancements achieved in both bandwidth and power. The inductive approach is current-driven and enables an effective face-up assembly but requires large power consumption (1.7mW/Gbps [5]). The proposed capacitive solutions are voltage-driven and this results in lower power consumption. Moreover, face-to-face assembly and fine alignment lead to smaller interconnections and this advances the throughput per area. Although if the system architectures are very different, it has also been made an effort to compare the contact-less technology with Through Silicon Vias (TSV) and System-in-Package (SiP) approaches.




| This Work | |
|---|---|
| size | 8x8μm² |
| load | 6fF |
| speed | 420ps propagation delay |
| tech. | Alignment, stacking, external connections |
| **Through Silicon Vias [10]** | |
| size | 2.5x2.5μm², 4μm to source/drain, 11μm to gate oxide |
| load | 2.5Ω |
| speed | N.A. |
| tech. | Via formation, thinning, soldering, stacking |
| **System In Package [12]** | |
| size | 30μm diameter, 60μm pitch |
| load | 50fF, 14mΩ |
| speed | 123Mbps/pin |
| tech. | Microbumps, stacking, gluing |

Fig. 2.10 Comparison among 3D technologies

Reported TSV are 2.5μm per side but require a minimum distance to gate oxide (and thus to functional circuits) of 11μm [10]. The AC electrodes have a minimum area of 8x8μm2, but they do not affect the silicon substrate so that adjacent circuits can be as close as allowed by the CMOS design rules; for this reason the pitch enabled is, in our judgement, equivalent to TSV. SiP solutions allow die connection using well known technologies, as micro-bumps, but suffer from a large interconnection pitch (50-60μm [11] [12]) and from a limited throughput (123Mbps/pin [12]). From a technological point of view, our approach does not require any addition to Back-End-Of-the-Lines, so the packaging effort is only slightly larger than usual SiP assembly procedures.

# CHAPTER 2: CAPACITIVE 3D INTERCONNECTIONS FOR HIGH SPEED APPLICATIONS

# 3D CAPACITIVE TRANSMISSION OF ANALOG SIGNALS

Wireless interconnections based either on capacitive or inductive coupling have already been proven for transmitting digital data. So far great efforts have been done in order to improve the system performances. Miura et. all developed a transceiver for inductive coupling [1], which can be applied to more than two chips stacked in whether face-to-face, face-up/down, or even back-to-back. Despite the mentioned implementation is valuable from design standpoint, since provides a fairly feasibility in assembly and high scalability in system integration, it requires to increase the transmission energy in order to deal with assembly issues. On the other hand, interface design for connections based on 3D capacitive coupling presented in [2], [3] aimed to minimize the interconnection area; thus optimize the integration density and the power consumption reduction. So far, the works presented in literature describe solutions focused on obtaining high-performance in digital data systems.

Nevertheless, even though a large part of SiP is digital based, some significant applications require the transmission of analog signals.

The aim of the work explained in the following sections is to prove the possibility to implement a chip-to-chip transmission of analog signals. Moreover, it is demonstrated that the same technology presented in [3] can be used in analog applications as well. The proposed solution is suitable for low-cost applications which also require high communication density. Although the project has been carried out to operate in low frequencies domain, it worth to notice that the architecture can find applications for the transmission of wide band analog signals where analog to digital conversion is impractical or very costly. In this case only the Variable Gain Amplifier implemented in the receiver would have to be redesigned for high bandwidth [4], while every other system component can still operate at low frequencies.

# PROBLEM DEFINITION AND SOLUTION

In Digital Transmission Systems, the voltage or current pulse induces a positive or negative pulse voltage on the Receiver. The main matter is guarantee high sensitivity and noise immunity on the receiver node. It determines the lowest signal variation that it is recognized as logic swing, thus it fixes the current pulse power or the electrodes size. The different architectures can implement several topologies of signal recovery in order to carry the voltage swing to the correct logic level 1 or 0. Obviously it is not possible when it operates in analog domain, where it is needed to recover the signal to the original variable amplitude. The error made in such process determines the resolution of the system, thus the performances. In contact-less capacitive approach the signal attenuation is due to the presence of parasitic capacitances on the RX node. The parasitic effects cause signals attenuation and result in a reduction of communication efficiency. Carry back the signal amplitude to his original value will be easily accomplished if the attenuation is known or at least it could be accurately estimated.

## CHAPTER 3: 3D CAPACITIVE TRANSMISSION OF ANALOG SIGNALS

But in 3D chips stacking the coupling capacitance of the transmitting channel is affected by large uncertainty. In order to better understand the problem it is necessary to examine the several components to the overall capacitive connections. The interconnection attenuation is expressed by the formula:

$$H_{3D} = \frac{C_{cc}}{(C_{cc} + Crx)}$$

Where $C_{CC}$ is the inter-electrode coupling and $C_{RX}$ is the parasitic coupling to ground of the receiver electrode, as sketched in Fig 3.1.

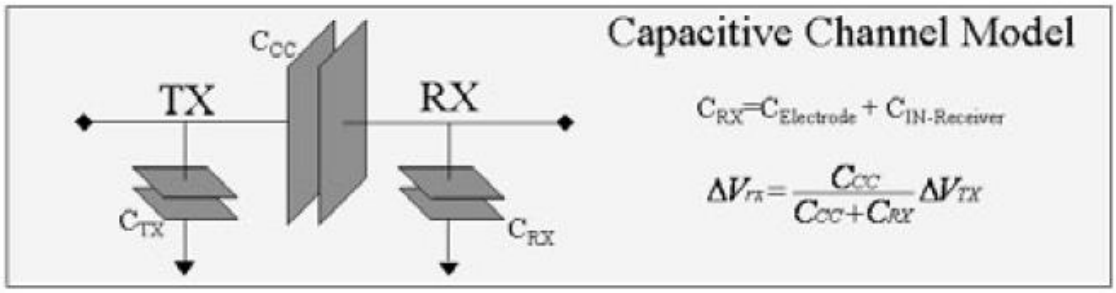



Fig 3.1 Capacitive Channel Model of the 3D communication system

The capacitance $C_{RX}$ on the RX node consists in two components:

- $C_{Electrode}$: electrode coupling capacitance to ground
- $C_{IN\text{-}Receiver}$: receiver input load

$C_{Electrode}$ is basically formed by the capacitor existing between the top metal rectangle (i.e. electrode) and the underlying metal lines, also considering contributes of

via connections. Thus, it mostly depends by the size of the electrode itself and it cam be extracted from the layout design. Table I reports the capacitance values for different metal sizes.

**Table I**

**Capacitance values corresponding to different electrodes sizes**

| SIZE | CAPACITANCE |
|---|---|
| **5x5um2** | 2.83fF |
| **10x10um2** | 4.00fF |
| **15x15um2** | 5.23fF |
| **20x20um2** | 6.59fF |
| **25x25um2** | 8.18fF |
| **30x30um2** | 10.07fF |

$C_{IN\text{-}Receiver}$ is related to the load capacitance on RX node, thus it depends by the solution adopted to realize the receiver circuit. In the previously works and in the proposed one, it is generally represented by the input capacitance of inverter circuit or NMOS pair. Anyway the capacitance contribute can be analytically calculated during the first phase of the project using the technology electrical parameters specification.

Thus, the main problem is due to a large uncertainty on the inter-coupling capacitance $C_{CC}$. This value varies with the thickness of the inter-chip isolation layer and with misalignment of the electrodes. The packaging assembly is realized by Fraunhofer-IZM. In the die-to-die option, it is possible to stack the devices face-to-face with a inter chips distance down to less than 1um. The developed methodology also guarantees an alignment accuracy of 1um. Nevertheless, it is not a standard process which can be completely controlled thus the thickness of the acrylate based adhesive, which provides inter-electrode dielectric and mechanical stability, can vary on the range of hundreds nanometres [from 750nm to 1000nm] and it is not homogeneous all over the surfaces. It means that different prototypes present different inter-electrode coupling

or also the attenuation varies within the prototype depending by the position. Moreover, misalignment is more critical when implementing really small electrodes (i.e. 8x8um$^2$) in order to reduce interconnection pitch and achieve higher connection density.

The inter-coupling capacitance is expressed by the formula:

$$C = \epsilon_0 \epsilon_r \frac{A}{D}$$

Where the constant value $\varepsilon_0$ is known as the electric constant or the permittivity of free space, and has the value $\varepsilon_0 \approx 8.854187 \times 10^{-12}$ F/m. $\varepsilon_r$ is the relative permittivity equals to 2.8 for the acrylate adhesive. *A* is the area of the conductors, thus electrode size, and *D* is the inter chips distance. Taking in account the errors due to misalignment (X1) and variable adhesive thickness (X2), it can be re-write as:

$$C = \epsilon \frac{(A+X_1)}{(D+X_2)}$$

Table II reports values of the coupling capacitance in the best case (i.e. no process errors) and worse case (i.e. both misalignment error and inter-chip distance equal to 1um) for different electrode sizes

**Table II**

**Comparison between 3D capacitances in Nominal case and Worse case**

| ELECTRODE SIZE | Nominal Capacitance | Worse Case Capacitance | RATIO |
|---|---|---|---|
| 5x5um$^2$ | 0.826fF | 0.347fF | 0.420 |
| 10x10um$^2$ | 3.306fF | 2.008fF | 0.607 |
| 15x15um$^2$ | 7.438fF | 4.859fF | 0.653 |
| 20x20um$^2$ | 13.223fF | 8.950fF | 0.676 |
| 25x25um$^2$ | 20.660fF | 14.279fF | 0.691 |
| 30x30um$^2$ | 29.751fF | 20.850fF | 0.701 |

It shows that inevitable errors in packaging assembly result in a reduction up to 40% of the inter chip capacitive coupling with respect to the nominal estimated value. As consequence, when the interconnection attenuation, for a specific receiver scheme, with defined capacitive channel size, in given technology process, is estimated for example H3D = $C_{CC\text{-}Nominal}$ / ($C_{CC\text{-}Nominal}$ + $C_{RX}$) = ½, due to the packaging assembly errors, the value of H3D in different samples can vary from 1/2 in the best case to 1/3 in the worst case. It should be added that the value of parasitic capacitance $C_{RX}$ can be calculated but not exactly predicted, resulting in further uncertainty. Those reasons explain the necessity to realize a system able to compensate a signal reduction, when it is required to transmit analog signals. The attenuation depends by package assembly, architecture, design but it is reasonable to consider a value in the range of 2 to 4.

Nevertheless, it is still possible to rely on a reasonable degree of matching of the inter-chip capacitances across a significant area. It suggests the idea to implement a scheme which can afford to automatically compensate the signal attenuation and locally correct the unpredicted variations. It results in the automatic gain control (AGC) system sketched in Fig. 3.2.

## CHAPTER 3: 3D CAPACITIVE TRANSMISSION OF ANALOG SIGNALS

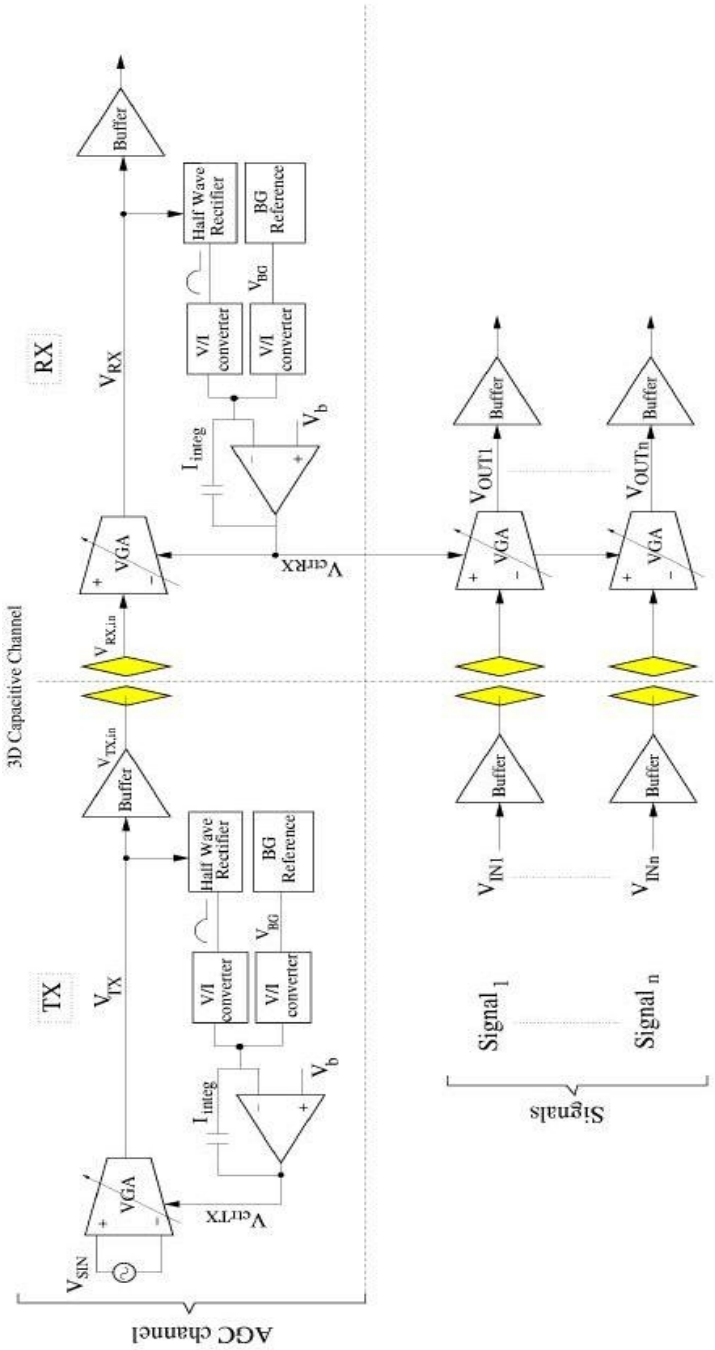



Fig 3.2 Block diagram of the system for 3D capacitive transmission in the general case of analog signals. (Buffers are unity gain amplifiers)

## CHAPTER 3: 3D CAPACITIVE TRANSMISSION OF ANALOG SIGNALS

The proposed scheme includes a variable gain amplifier (VGA) circuit and it relies on the availability of a Si band-gap reference cell in both the transmitter (TX) and the RX.

The attenuation factor H3D in the n channels is compensated by variable gain amplifiers (VGA) whose control voltage $V_{VCTRR\text{-}X}$ is generated by the AGC circuit. In the TX part of the AGC circuit, a sinusoidal voltage $V_{SIN}$ at a fixed frequency is applied to the VGA, whose output is half-wave rectified. The output voltage of the rectifier is converted into a current, compared with a current proportional to the band-gap reference voltage $V_{BG}$ and low-pass filtered by the integrator. It turns out that in steady state the amplitude of $V_{TX}$ is equal to $\pi/2\ V_{BG}$, where $\pi/2$ is due to the rectifier. Because of the capacitive attenuation factor, the amplitude of received signal $V_{RX,in}$ differs from the transmitted one. A loop in the RX of the AGC circuit with the same block diagram as in the TX, but possibly implemented in a different technology, forces the amplitude of $V_{RX}$ to be equal to $\pi/2\ V_{BG}$. Therefore $V_{TX}$ is equal to $V_{RX}$ apart from a delay. $V_{ctrRX}$ is applied to all the identical VGAs of the n receivers forcing $V_{IN} = V_{OUT}$ in each of them. Assuming ideal behaviour of the circuit building blocks, an amplitude error between input and output voltages is only caused by mismatches in the $C_{RX}$ and $C_{CC}$ values. These mismatches can be minimized by keeping the distance between the AGC circuit and the signal channels low and, ultimately, define the maximum number of signals which can be controlled by a single AGC.

# PROPOSED SYSTEM

The developed system essentially consists in the following components:

- **Variable Gain Amplifier (VGA)**
- **Band Gap Voltage Reference Cell (BGVR)**
- **Half Wave Rectifier (HWR)**
- **Voltage to Current Converter**
- **Current Integrator**

## VARIABLE GAIN AMPLIFIER (VGA)

The real challenge in designing a variable gain amplifier is to keep high linearity within an acceptable range of signal amplitudes and at the same time maintaining wide bandwidth. The key point lies at the pre-amplifier stage. Besides maximizing the gain and sensitivity for a given bandwidth, the pre-amplifier must also have a wide dynamic range. The automatic gain control (AGC) channel includes a low-voltage low-power wideband CMOS variable gain amplifier circuit. Moreover, a fully differential topology is recommended to reduce the sensitivity to noise induced by digital signals. The architecture of the VGA is sketched in Fig. 3.3 [5]. It consists in a cascade of a linear transconductance amplifier and a linear transimpedance amplifier.

## CHAPTER 3: 3D CAPACITIVE TRANSMISSION OF ANALOG SIGNALS

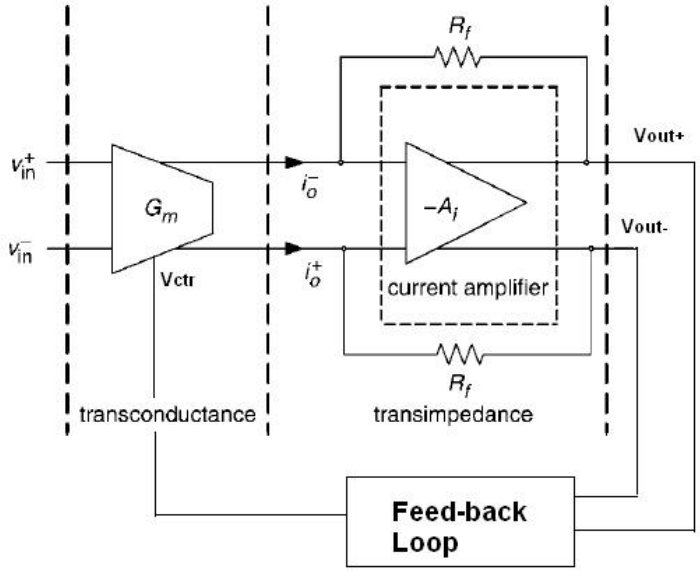



Fig.3.3 VGA architecture: pre-amplifier stage and transimpedance stage with feed back loop

The usual voltage amplifier is replaced with a current amplifier. Doing so realizes a constant bandwidth that is ideally independent of the transimpedance gain, thus it allows constant bandwidth when varying the voltage gain [5]. The schematic circuit of the proposed VGA is sketched in Fig. 3.4. The current amplifier is formed by the transistors M5 – M8 and the shunt-feedback resistors $R_f$. It converts the output current into the output voltage, keeping voltages $V_X$ and $V_Y$ nearly constant. The value on those nodes is sets by the common-gate stage; it is adjustable and can be tuned by varying the bias voltage Vbias. The transconductor stage is composed of $M_1$ - $M_4$, with $M_C$ operating in the triode region used as a variable conductance. The voltage gain of the VGA circuit is the product of the transconductance gain ($G_m$) and the transimpedance gain ($R_m$).

## CHAPTER 3: 3D CAPACITIVE TRANSMISSION OF ANALOG SIGNALS

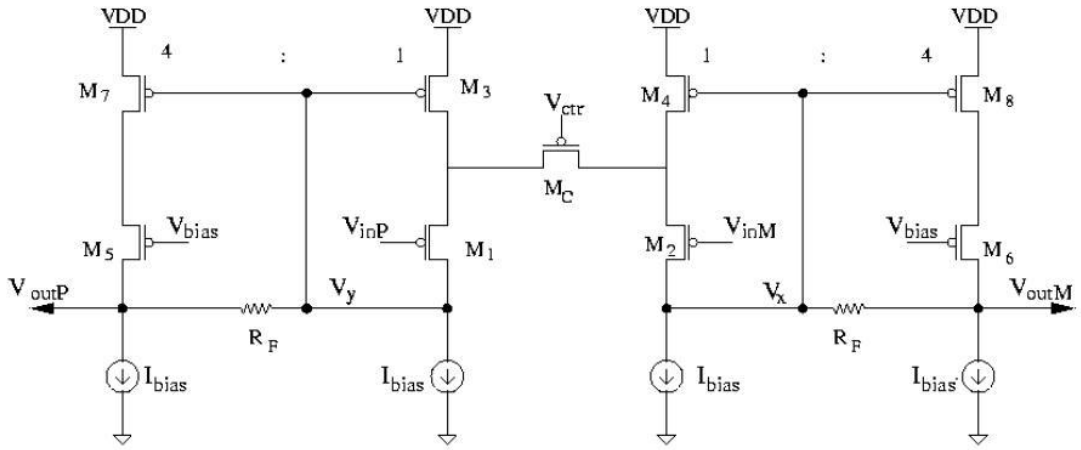



Fig. 3.4 Proposed Variable Gain Amplifier circuit

The transimpedance gain is given by (1), where $R_{in}$ and $A_i$, respectively, are the input resistance and the current gain of the current amplifier. Note that, when $A_i >> 1$, it is $R_m = -R_f$ (1).

$$R_m = -\left(\frac{R_f A_i - R_{in}}{1 + A_i}\right)$$

Transistors M1 and M3 perform two roles simultaneously, both as part of the transconductance amplifier and as an input section of the current amplifier. This saves power dissipation and area. As part of the transconductance amplifier, M3 (M4) provides negative feedback to force a constant current flowing through M1; thus the differential input voltage appears across MC and the differential output current flows in M2. As an input section of the current amplifier, M1 and M3 (M2 and M4) provide low input impedance for the differential current signal. The previous considerations establish the project parameters to sizing the input transistors; small channel width can affects the gain performance, on the other hand increasing the size results in a huge input capacitance on the receive node abating the transmission efficiency. The transistor

size has been chosen after SPICE simulations in order to have the best compromise for the trade-off.

The output stage realizes a wide-swing cascode current source with Ibias and α equal 15uA and 4, respectively. The values of Ibias and α (the current gain) together with Vbias value determine the DC value on the output node, thus the operating point of the following circuits. The parameters have been calculated so that DC value is fixed to 1.5V, when no input signal is forced. The choice gives some advantages when implementing the other components of the systems.

The transimpedance gain is given by expression (2), where alpha represents the current gain. When alpha >> 1, the low-frequency transimpedance gain is equal to -Rf and the poles are at $\omega_{p2}$ = - (gm1/Cgs1) and $\omega_{p2}$ = - (gm2/Cgs2), which are at the cut-off frequencies ($F_t$) of the transistors. This allows the VGA circuit to have wide operating bandwidth (2).

$$R_m = \frac{g_f - \alpha\, g_{m2}}{g_f\, g_m 2(1+\alpha)} * \frac{1}{1+s[(\alpha/1+\alpha)(C_{g1}/g_{m2})]}$$

$$+S^2[(C_{gs1}/g_{m1})(C_{gs2}/g_{m2})]$$

Therefore, the dominant pole of the transimpedance amplifier will be determined by the output load capacitance ($C_L$). Varying $R_f$ also affects the output pole and thus the bandwidth of the VGA. Thus, while designing the VGA, the values of $R_f$ and $C_L$ should be chosen first according to the bandwidth requirement and resistance of MC is used to adjust the voltage gain while keeping the bandwidth constant. The load capacitance depends by the input capacitance of the cascaded circuits that appears on the VGA output node. The mentioned components are not the same for all the VGA devices used within the system; in fact in Transmitter the VGA drives the electrodes and the half wave rectifiers, while in the Receiver it is connected to the half wave rectifiers and the output buffer stages, finally in the Data Channels appears only the input capacitance of

the output buffer stages. Different load capacitances results in different cut-off frequencies, thus it provokes frequency dependent behaviour. To avoid such error, small buffers have been implemented. Each VGA output is connected to a buffer which drives the following stages, so that the load capacitance is equal to the input capacitance of the buffer and all above it is the same for every VGA. The buffer stage is realized with operational amplifier in unit gain configuration. It is not needed to drive wide loads, while it is required to not affect the bandwidth and the linear swing range of the whole system. Since other system components present more critical characteristics, those specifications can be easily accomplished. The input capacitance is expressed as:

$$C_{in} = \frac{\epsilon_{ox}}{T_{ox}} WL$$

where $\epsilon_{ox}$ is the oxide permittivity equals to:

$$\epsilon_{ox} = \epsilon_0 * \epsilon_r = 8.854187 \times 10^{-12} \text{ F/m} * 3.9 = 34.53 \times 10^{-12} \text{ F/m}$$

Tox is the oxide thickness which for 90nm standard CMOS technology is 1.6nm, but in the case of 2.5V power supply voltage option it values 5nm.

In order to reduce the effects of mismatch in transistors realizations, L is not equals to the minimum channel length $L_{min}$ but it values 0.5um, while the channel width W is 5um.

The load capacitance value is worked out by inserting those parameters in the previous expression:

$$C_L = C_{IN} = (34.53 \times 10^{-12} \text{ F/m} * 2.5 \times 10^{-12} \text{ m2}) / 5 \times 10^{-9} \text{ m} = 17.27 \times 10^{-15} \text{ F}$$

## CHAPTER 3: 3D CAPACITIVE TRANSMISSION OF ANALOG SIGNALS

Tacking in account that the system is projected to operate at frequencies up to 20 MHz, the value of $R_f$ can be calculated by the formula:

$$F_0 > F_t = 1 / 2\pi R_f C_L$$

$$\text{thus } R_f < 1 / 2\pi F_t C_L = 461\ K\Omega$$

Reducing the value of $R_f$ results in an increase of current flow on the branch and it reduces the linearity of the system. So that $R_f$ should be greatest as possible maintaining the desired bandwidth. The resistance $R_f$ can be implemented with triode connected PMOS transistors. Nevertheless, the resistors have been realized in poly-silicon series resistances, since it guarantees higher linearity. The total resistance value deals with the trade-off between minimum bandwidth required, high linearity constrain and area occupation; $R_f = 150\ K\Omega$.

The controllable linear resistance can be implemented using either a NMOS or a PMOS device. The resistance value is controlled by varying the control voltage source ($V_{ctr}$). To maintain excellent control on gain variation, the project of the following stages should be carried out in order to guarantee that Mc operates in triode region. The transistor should be modelized so that the desired gain variation corresponds to a defined voltage swing. A wide voltage range is meant a spare linearity, on the other hand a short range imply a poor control in gain variation.

SPICE simulations demonstrate that the VGA operates with high linearity in the required gain ranges from 2 to 4. The maximum output voltage swing is up to 1V as sketched in Fig. 3.5.

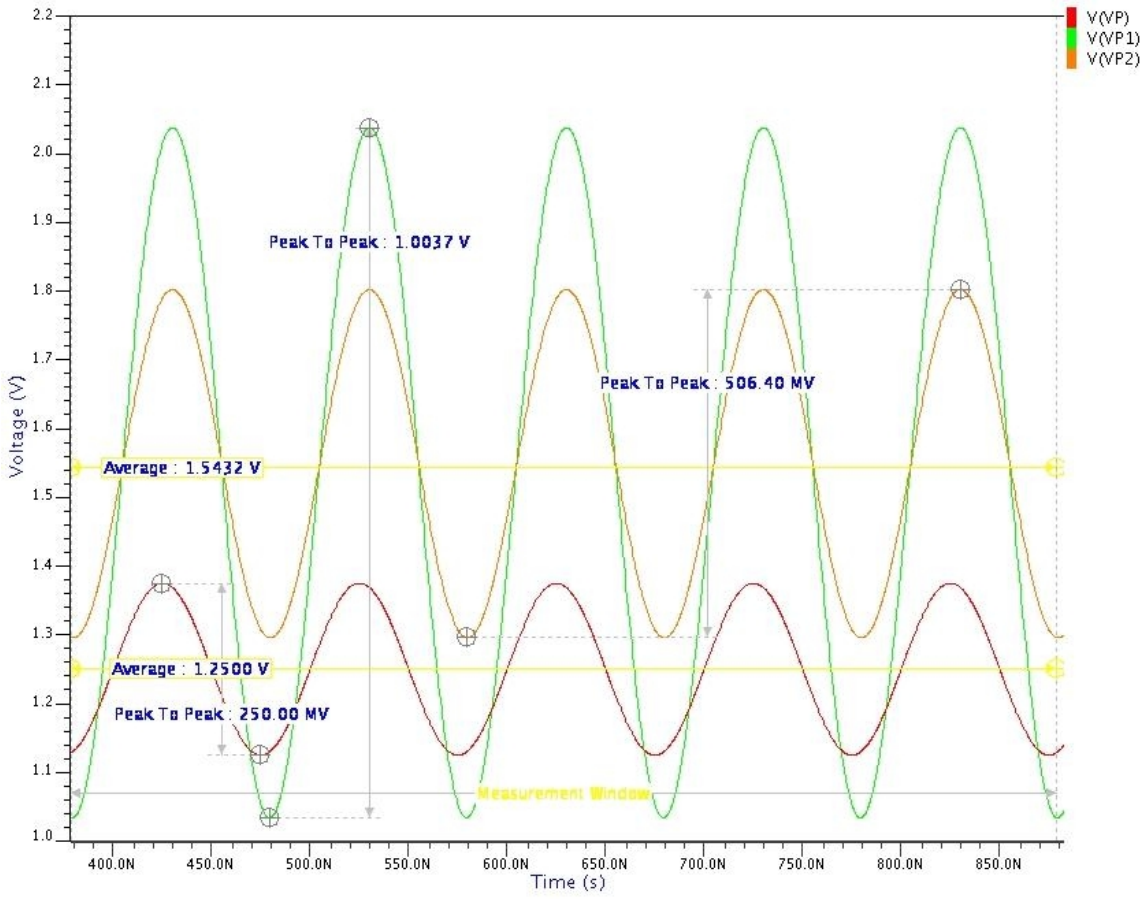



Fig. 3.5 Variable Gain Amplifier: VP input signal, VP1 and VP2 output waveforms in the case of gain equal to 2 and 4 respectively. Control voltage sources set to 0.1V and 0.85V

Fig. 3.6 shows the behaviour of the circuits for the same input signal amplitudes while varying the operating frequency. Within the 100 KHz - 20MHz range, the gain does not depend by the frequency.

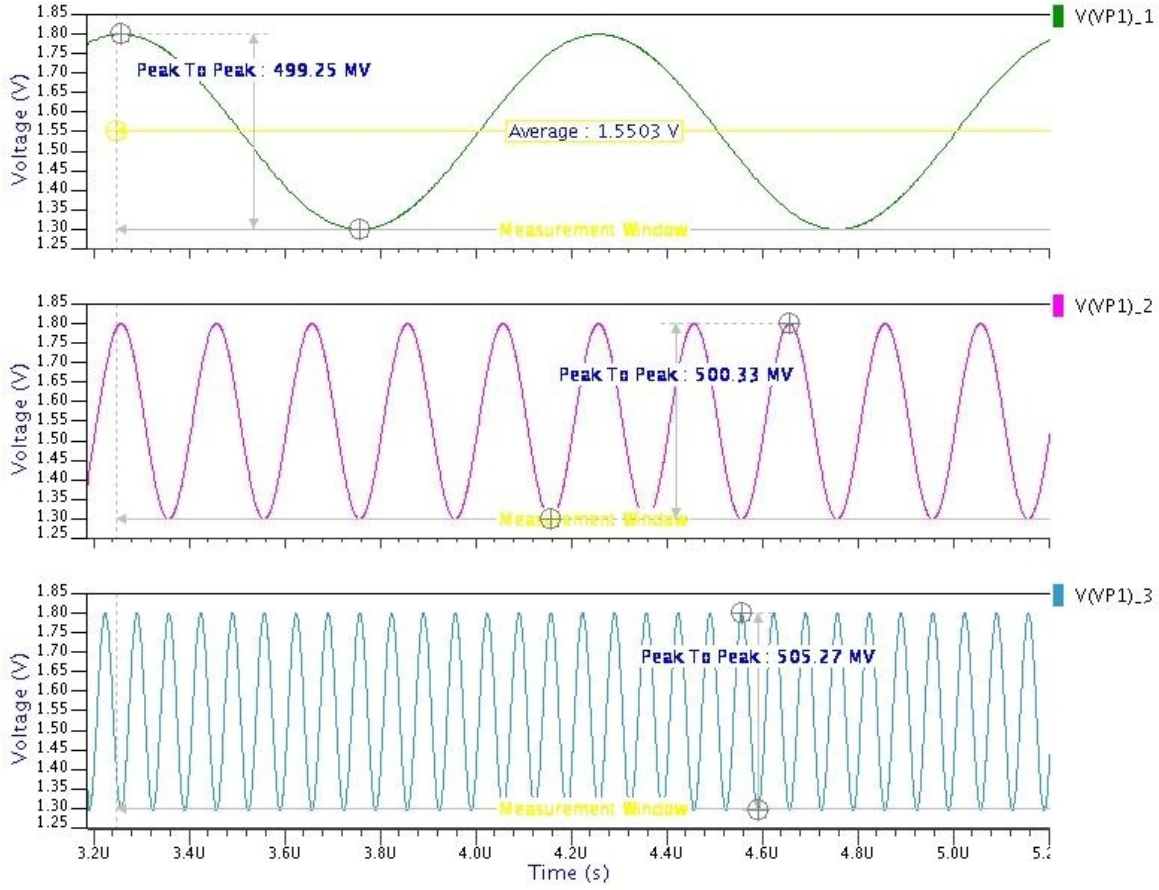



Fig. 3.6 Variable Gain Amplifier: gain vs. frequency. The VGA gain has been fixed to 2. The input signal frequency is equal to 1MHz (top), 5MHz (middle), 15MHz (bottom)

Finally, Fig. 3.7 reports the small signals frequency response of the VGA, showing constant bandwidth greater than 20MHz.

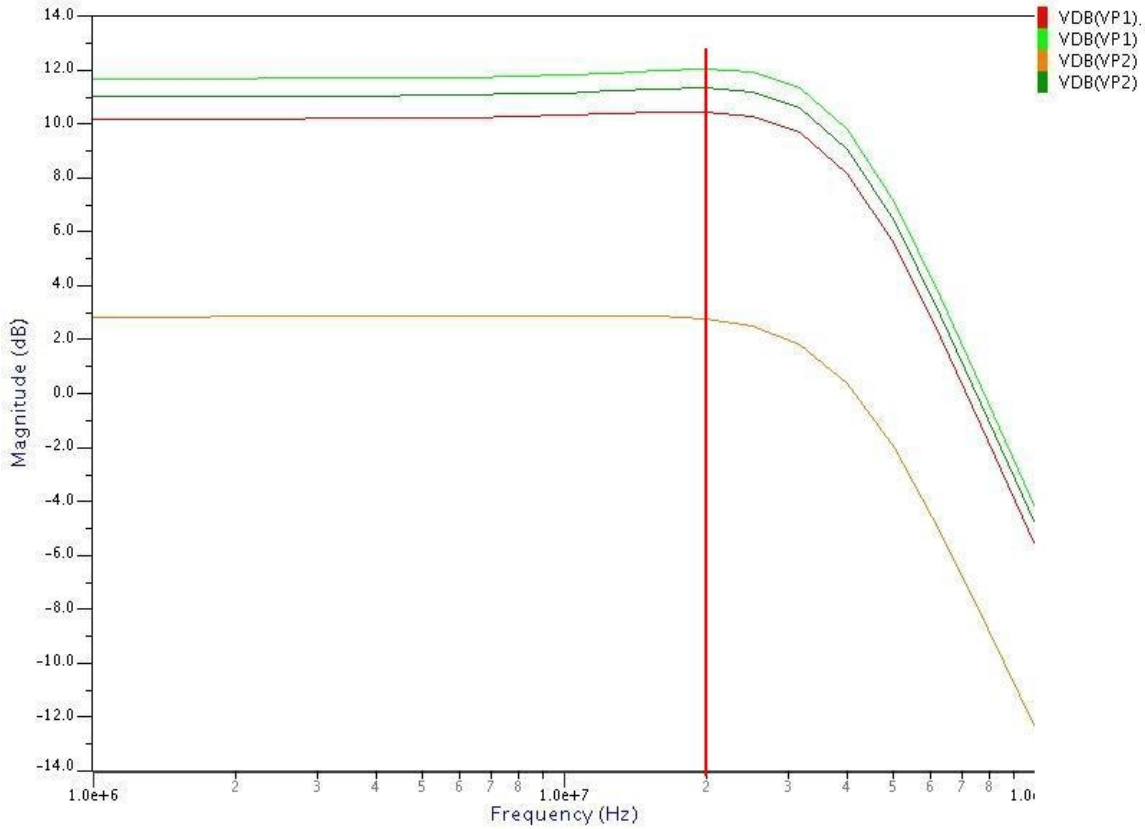



Fig. 3.7 Small signals response of the Variable Gain Amplifier for different gain values within 2 to 4 with load capacitance equals to 50fF. The marker shows the approximately flat bandwidth up to 20 MHz

## BAND GAP VOLTAGE REFERENCE CELL

A band gap voltage cell is a voltage reference circuit widely used in integrated circuits usually with an output voltage around 1.25V, close to the theoretical 1.22eV band gap of silicon at 0K. Voltage references are generally present as basic cells in the VLSI CMOS libraries. Carefully design and specific correction techniques are used in order to assure high accuracy, the output voltage presents variations in the range of few mill volts with temperature dependence typically smaller than 8ppm/°C in a temperature range from -40 to +85°C. Dedicated standard cells are realized in every technology node and available for different process options.

## HALF WAVE RECTIFIER

A rectifier is an electrical device that converts alternating current (AC) to direct current (DC), a process known as rectification. Half wave rectification removes one half the input signals and follows the other part. The ideal behaviour is sketched in Fig. 3.8.

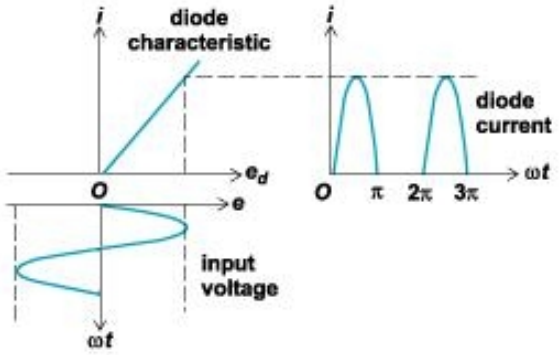



Fig. 3.8 Ideal rectified waveform

## CHAPTER 3: 3D CAPACITIVE TRANSMISSION OF ANALOG SIGNALS

In the described Automatic Gain Control system, the two single-ended components $V_{tx,p}$ and $V_{tx,m}$ of the VGA differential output voltage $V_{tx}$ are applied at the inputs of two half wave rectifiers. The proposed circuit is showed in Fig. 3.9; where the rectifier on the left allows the output voltage VP to follow the positive half of the AC wave, while during the negative phase the node is kept at a constant value. The rectifier sketched on the right follows the negative half of the AC input wave, while during the negative phase the node is kept at a constant value.

The transistors triode connected M1 and M2 woks as rectifying diodes. In order to operate in saturation region, the maximum value on $V_{TX,P}$ is limited to $V_{TX,P,max} = VDD - V_{ds,sat} - V_{gs,sat}$ (reciprocal consideration hold true for $V_{tx,m}$). It should be reminded that the system operates with 2.5V power supply, so that every component has been implemented using Thick Oxide Device. The process option presents voltage threshold in the range of 0.5V - 0.6V; resulting in a considerable loss in the maximum admissible signal swing in the presented configuration.

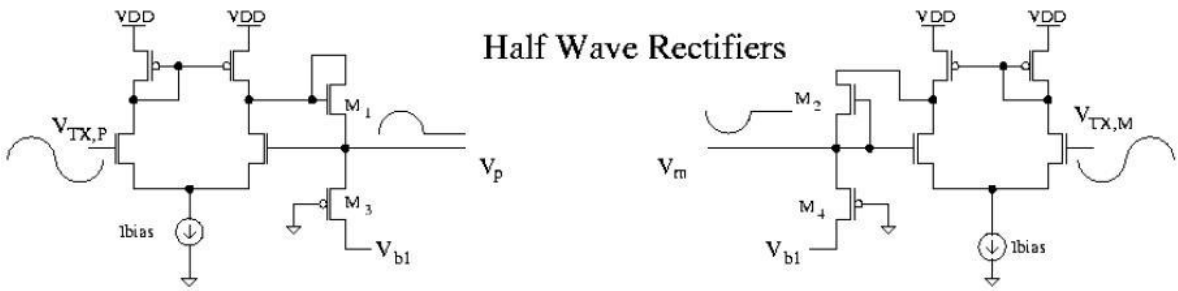



Fig. 3.9 Half-wave rectifier circuit

To minimize the diode threshold, thus assure high voltage swing, the circuits have been implemented using two standard thin oxide NM0S (M1 and M2) with low threshold voltages; 0.12V and 0.18V respectively. It is worth to notice that the

maximum admissible power supply for this type of transistors is 1.25V, nevertheless they are forced to operate in over voltage condition only for limited time. Furthermore, several SPICE simulations have been done to verify that tension $V_{GS}$ is never critical, thus it is reasonable that the stress cannot damage the device.

The bias voltage $V_{b1}$ must exactly be equal to the DC value of $V_{TX,P}$ and $V_{TX,M}$; poor accuracy will provoke error in comparison stage limiting the system precision. Vb1, a well as Vb2, is obtained by means of a replica of one half of the VGA circuit with constant input voltage applied.

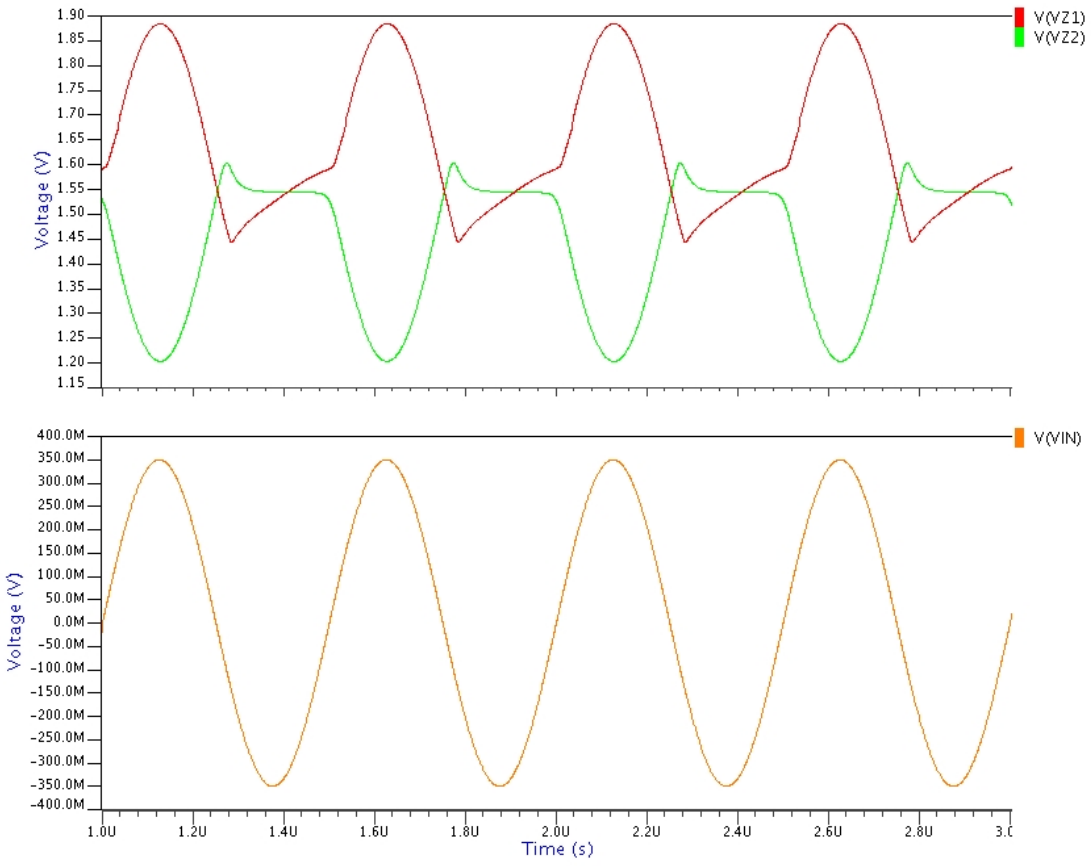



Fig. 3.10 Half Wave Rectifier: VIN input signal, VZ1 positive rectified waveform, and VZ2 negative rectified waveform

No negligible problems arise from the parasitic capacitive coupling between the gate and source of the transistor. During the negative phase, the rectified output node is pushed below the fixed threshold (1.55V). Moreover, maximum/minimum peaks slightly differ from the respective input values. The simulated waveforms are sketched in Fig. 3.10.

The superimposed parasitic effects affect the evaluation stages. In fact the rectified output voltage VP - VM is applied to one of the two V/I converters (discussed in the next section) whereas a constant voltage proportional to *VBG* is applied to the other one. The difference of the output currents of the two converters is integrated by the capacitance $C_{INT}$ and the output voltage $V_{CTR}$ fed to the VGAs. It is clear that any error within rectifier stage will appear in the signal compensation. To better explain, let us express the relation (3) between the rectified voltage, the generated current and the derived voltage $V_{CTR}$ (expressed as average voltage)

$$I = A_I \sin(\omega_0 t)$$

$$I = \frac{1}{T}\int_0^{T/2} A_I \sin(\omega_0 t)\, dt = \frac{\frac{A_I}{T} * T}{2\pi}[\cos(\omega_0 t)] = \frac{A_I}{\pi} \quad ,$$

$$where \qquad \omega_0 = 2\pi f \quad \wedge \quad A_I = 2V_{in} \;,\; A_I = \Delta VBG * \pi$$

The last expression defines the stable condition reached thank to the loop (see next section).

Introducing the error expressions in formula (3), it results:

$$\frac{1}{T}\int_0^{T/2} A_{err}\sin(\omega_0 t)\,dt + \frac{1}{T}\int_{T/2}^{T} K_{err2}\,dt =$$

$$\frac{1}{T}\int_0^{T/2} A_I\sin(\omega_0 t)\,dt + \frac{1}{T}\int_0^{T/2} K_{err1}\,dt + \frac{1}{T}\int_{T/2}^{T} K_{err2}\,dt$$

Where $K_{err1}$ and $K_{err2}$ arise from peak detection and parasitic capacitive coupling respectively

$$error = \frac{1}{2}(K_{err1} + K_{err2})$$

$K_{err1}$ and $K_{err2}$ are not related each other, thus the total error can be minimized with an opportune sizing of the diode connected transistors and tuning the differential pair gain.

## VOLTAGE TO CURRENT CONVERTER

A voltage-to-current converter changes the electric attribute carrying information from voltage to current. It acts as a linear circuit with transfer ratio $k = I_{OUT}/V_{IN}$ [mA/V] having dimension of conductivity. The voltage-to-current converter has been implemented as transconductance amplifier. The output current is a function of Rrr and the amplifier gain. R should be large enough to assure high linearity, while not occupying too area. The resistance it has been realized using four series connected

NMOS with gate connected to VDD. The circuit implements two V/I converters with the output connected each other as shown in Fig. 3.11. Thus, the rectified voltage VP - VM is applied to one of the two V-I converters whereas a constant voltage proportional to VBG is applied to the other one. As a consequence the resulting positive or negative current flows through the node till the integrating capacitance $C_{INT}$.

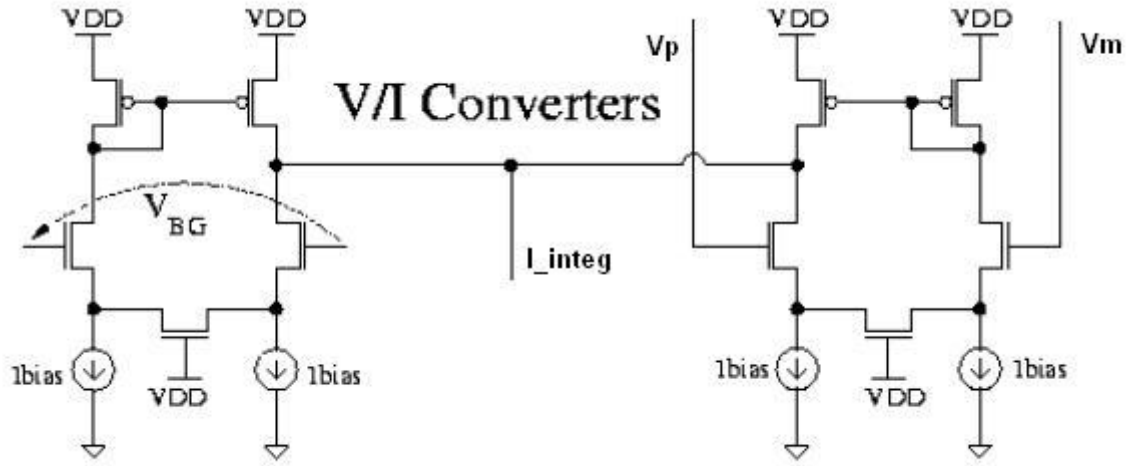



Fig. 3.11 Proposed voltage to current converter

## CURRENT INTEGRATOR

It has been implemented the standard current integrator sketched Fig. 3.12; where the AC current generated by the two voltage to current converters is integrated across capacitor C1. Here, the op-amp circuit would generate an output voltage $V_{ctr}$ proportional to the magnitude and duration that an input signal has deviated from $V_{ref}$. In fact, left-hand side of the capacitor is held to a voltage reference, due to the "virtual ground" effect. The output voltage $V_{ctr}$ is simply the voltage across C1 and it will feed to the VGA to automatically determine the correct value of gain needed to compensate

the signal attenuation. The rate of voltage change over time depends by the operating frequency, discharging time, and the size of capacitor; thus those parameters should be accurately choose to do not waste too area, while still having negligible ripple on the output voltage. The operating frequency of the VGA is not a project specification. The value should be chosen in order to optimize the behaviour of each component.

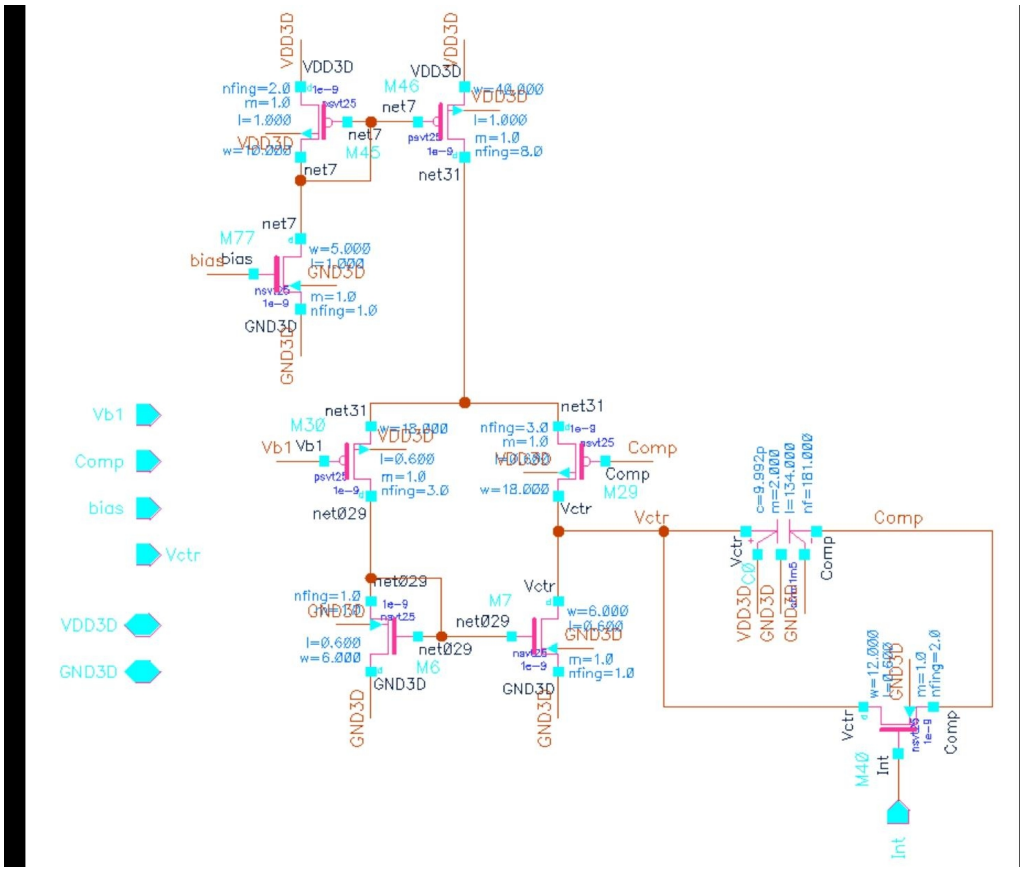



Fig. 3.12 Proposed Current Integrator

Trade-off exists, in fact: high frequencies imply smaller integration capacitance, while the condition is critical for the half wave rectifier since it is more difficult to follow the input signal. On the other hand low frequencies do not affect the rectification process, but require huge capacitors to perform fairly well integration. The operating

frequency has been chosen equal to 3MHz. It results in a capacitance in the range of pico farad. Greater the capacitance smaller the voltage ripple. In order to have a variation below 10mV the capacitance is equal to 9.992pF. The capacitor is implemented using two MOM capacitors formed by 181 fingers from Metal1 to Metal5, each fingers with length of 134um.

# TEST PROTOTYPES

A test chip devised for realization of 3D prototypes by stacking identical dice has been constructed in 0.09μm CMOS technology. For the first validation a simple system has been realized, it is formed by one Automatic Gain Control (AGC) channel that drives only two Data Transmitter (DT) channels. In further applications it is possible that one AGC could control several DT, anyway it is needed to determine the inter-chip distance variation, thus capacitance variations, along the horizontal axis due to process imperfections. So that it will be possible to establish the maximum number of DT channels per AGC. The photo of the chip obtained at microscope is sketched in Fig. 3.13, where the several parts of the prototype can be easily recognized.

Basically it consists in:

The so called “3D-Communication System”
The so called “2D-Communication System”

A third part has been added, it contains a copy of the single circuits used in the system. It is useful during the preliminary test phase to characterize the variable gain amplifier, the half wave rectifier, the output buffer, the voltage to current converter and the current integrator. It allows detecting unexpected behaviours in the whole system.

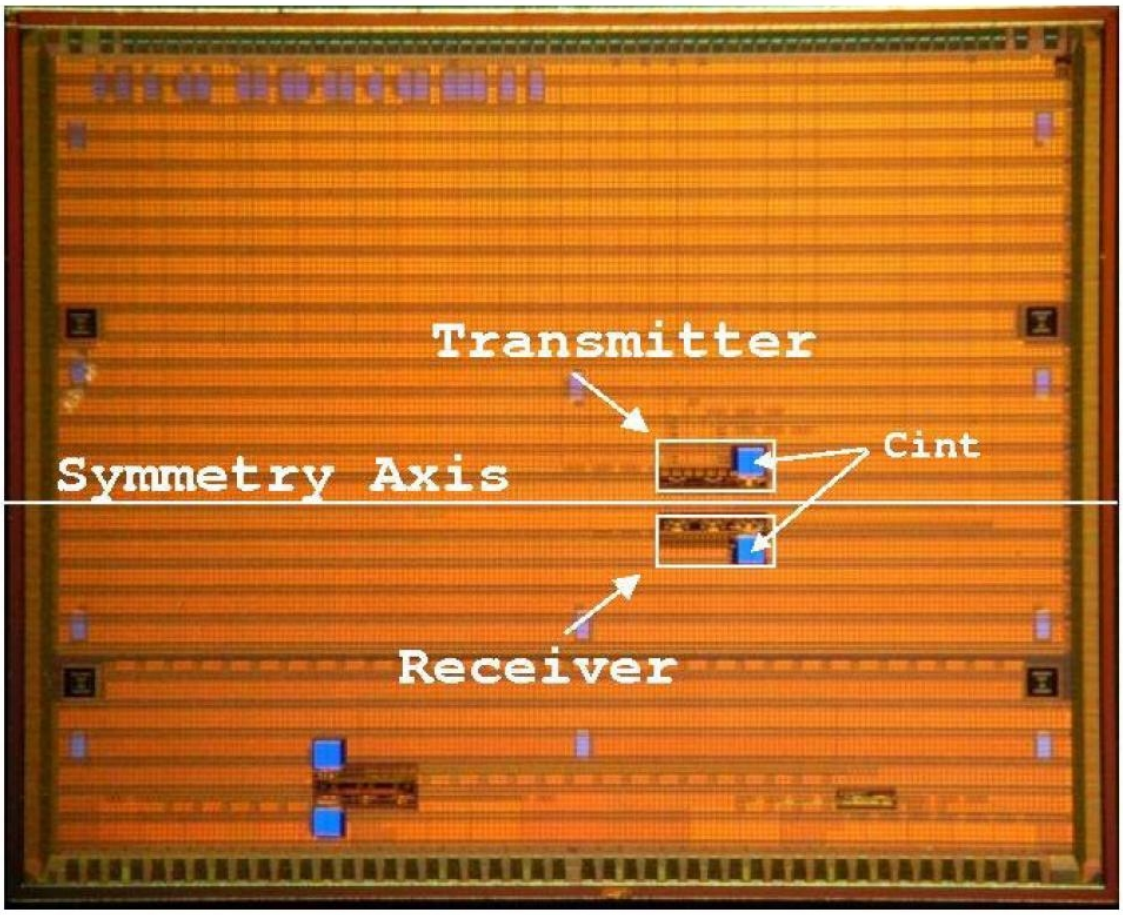



Fig. 3.13 Realized test chip, 90nm technology

## 2D-COMMUNICATION SYSTEM

It is almost a replica of the circuit realized for the chip-to-chip analog signal transmission described in the first section. The main difference is at capacitors level, where the electrodes are realized between metal3 and metal5, instead of in the upper metal layer, to provide the capacitive interconnection. Those two metals layers have been chosen because they space each other of 1um, more or less the same distance of two chips staked. This way allows to simulate the system behaviour on a single chip,

and to demonstrate the validity of the architectural approach described above. Furthermore it is really useful during the test, since it permits to have more output signal available. In fact, it should be reminded that in order to realize a chip-to-chip communication, two dices are staked one above the other one. The dices are opportunely misaligned such that the PADs are accessible; anyway the number of PADs available is much less than in normal planar configuration. It means that only the signals strictly needed (i.e.: input signals and output signals) can be visible to the output, thus there are no ways to access to other internal nodes. The 2-D Communication System offers the possibility to observe the behaviour of the devices inside the circuit. It permits to better understand unexpected results and eventually correct errors.

In order to choose the correct biasing values and external circuitry interface, in addition too the SPICE simulations, a preliminary test on the 2D device has been performed before proceed with the test on the 3D Prototype.

## 3D-COMMUNICATION SYSTEM

The main system is placed along the axis of symmetry as shown in Fig. 3.13. So that, when two identical dices are stacked face-to-face, the electrodes realized on the upper metal layer are perfectly aligned providing the capacitive coupling. In order to obtain high alignment accuracy, the process is performed optically. Thus it is necessary to insert four Align reference layer.

# TEST INTERFACES AND SETUP

The power supply unit furnishes the +5V, 0V and -5V voltage supplies used by all the external interface circuitry. The 2.5V core supply needed for the chip is also derived by the external power supply or otherwise it can be obtained from a voltage adjustable battery, the reasons will be explained later in the dissertation.

The Input signals were generated by Waveform Generator. The Agilent Technologies 33250A Function/Arbitrary Waveform Generator use direct digital synthesis (DDS) techniques to create a stable, accurate output signal for clean, low distortion sine waves. It can generate a single sine wave output up to 80MHz, with amplitude range from Vpp = 10mV to Vpp = 10V. At least two 33250A are needed in order to generate two distinct signals. The former drives the Reference Channel; operating at constant frequency (3MHz) and fixed amplitude. The other gives the input to the Data Channels, with different amplitudes and frequency in the range from 100 KHz to 20MHz. It is also possible to test the behaviour of the two data channels operating with different timing and amplitude by using a third waveform generator. Since the system has a fully differential implementation, the input signals should be in phase opposition, while having the same amplitude. Three Power Splitter/Combiners ZSCJ-2-2 showed in Fig 3.14 have been used, the behaviour will be briefly explained in the next section.

The output signals were observed with an Infiniium MSO8104 Oscilloscope. It tightly integrates four analog channels in the same acquisition system to provide time-correlated viewing and triggering. It normally operates up to 1GHz frequency bandwidth. Moreover it implements a low pass filter function at 20MHz. This function applies a real time digital filter to the source waveform. This filtering feature enhances the ability to examine important signal components by filtering out unwanted frequency components due to the superimposed noise.

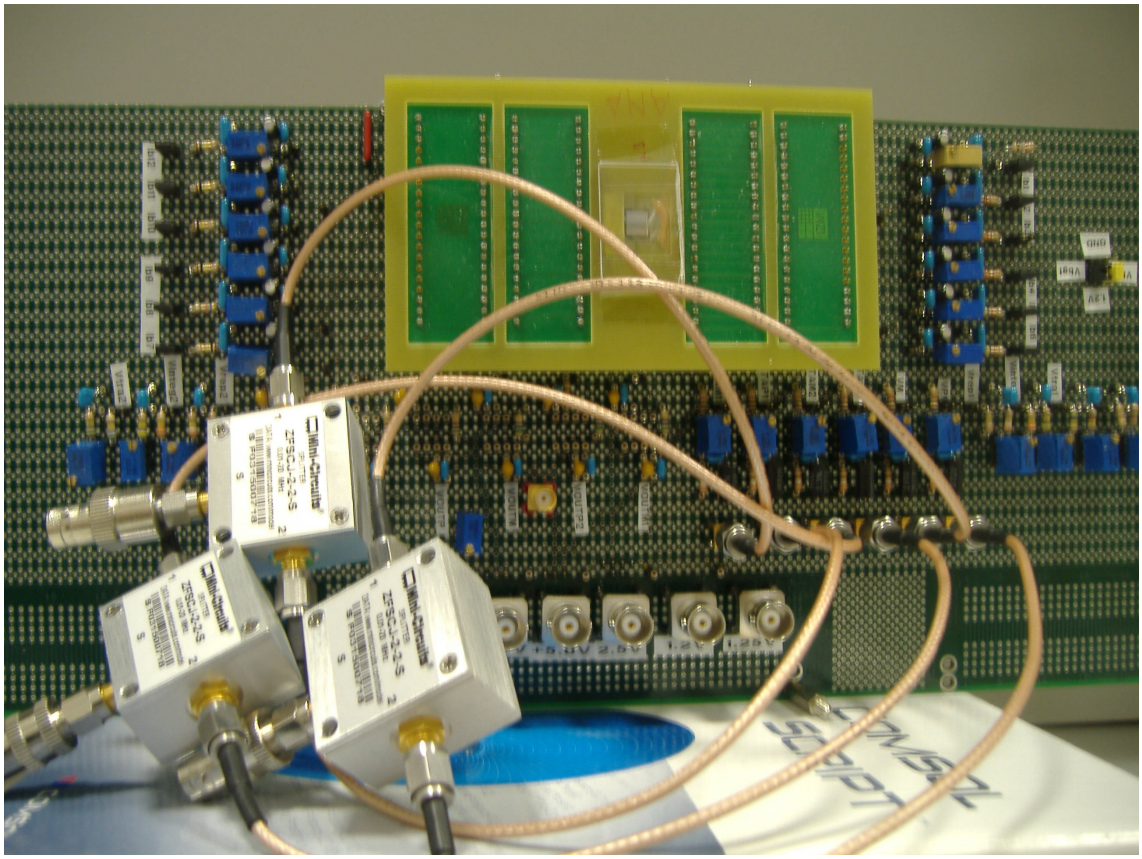


Fig. 3.14 Test board with the Power Splitter/Combiners

## POWER SPLITTER/COMBINER

Fig. 3.15 shows the block schematic of a two-way power splitter. It consists of a divider section, a matching section, a resistor R1 and a capacitor C1. The function of the divider section is, as the name implies, to divide the input signal into two parts. The matching section together with capacitor C1 performs wide-band matching at all ports. Resistor R1 helps provide wide-band isolation between the two RF output ports. The matching and divider sections form the heart of the splitter. They are realized using magnetic cores. The matching and divider sections are integrated into a single unit, and

are herein afterwards called the splitter.

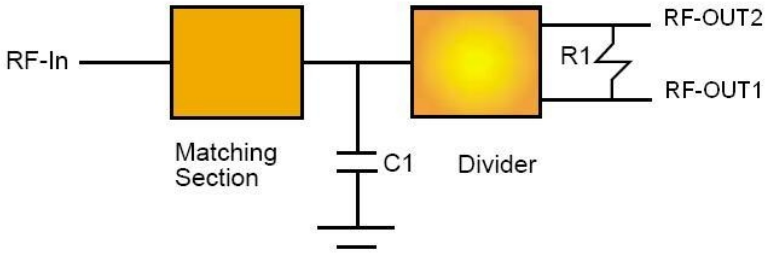



Fig. 3.15 Block schematic of the power/splitter combiner showing the matching section and the divider stage

Among several type of power splitter/combiner, it has been chosen the ZSCJ-2-2 splitter made by Mini-Circuits [6] since it covers the desired frequency range of 0.01 MHz to 20MHz. Moreover it presents insertion loss extremely low, typically 0.5 dB above the 3 dB split over the band, as it is shown in Fig. 3.16.

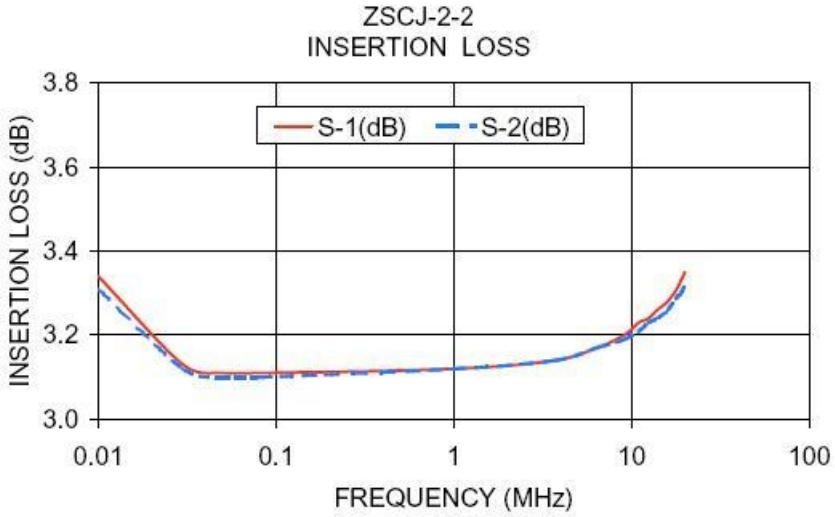



Fig. 3.16 Frequency dependence of the insertion loss

While the ZSCJ-2-2 splitter specifications are summarized in Table III.

**Table III**

**Splitter/Combiners Specifications**

| FREQ. RANGE (MHz) | ISOLATION (dB) | | | | | | INSERTION LOSS (dB) ABOVE 3.0 dB | | | | | | PHASE UNBALANCE (Degrees) | | | AMPLITUDE UNBALANCE (dB) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | | M | | U | | L | | M | | U | | L | M | U | L | M | U |
| $f_L$-$f_U$ | Typ. | Min | Typ. | Min | Typ. | Min | Typ. | Max. | Typ. | Max. | Typ. | Max. | Max. | Max. | Max. | Max. | Max. | Max. |
| 0.01-20 | 35 | 25 | 30 | 25 | 25 | 18 | 0.3 | 0.8 | 0.2 | 0.5 | 0.3 | 0.6 | 1* | 2 | 2.5 | 0.1 | 0.1 | 0.2 |

L = low range [$f_L$ to 10 $f_L$] M = mid range [10 $f_L$ to $f_U$/2] U= upper range [$f_U$/2 to $f_U$]
* Phase unbalance is 3 degrees max from $f_L$ to 3$f_L$

## WIRE WRAP BOARD

In order to perform the 2D-Communication System test and the 3D-communication System test dedicated wire wrap board have been realized. The test board used for 3D test is sketched in Fig. 3.17. Originally also a PCB board had been realized. The PCB solution presents advantages in terms of capacitive parasitic and inductive effects. On the other hand it is a less flexible approach and it left few possibilities to change a defined configuration. Since it was the first time that such kind of design was tested, the former solution was preferred. It is well know that analogical signals have high sensitivity to noise, thus a carefully project of the external circuitry interface is essential.

It basically consists of three parts:

**1] Input interface**

**2] Bias circuitry**

**3] Output interface**

In the follow sections, the system requirements, the devices specifications, and circuitry solutions are explained.

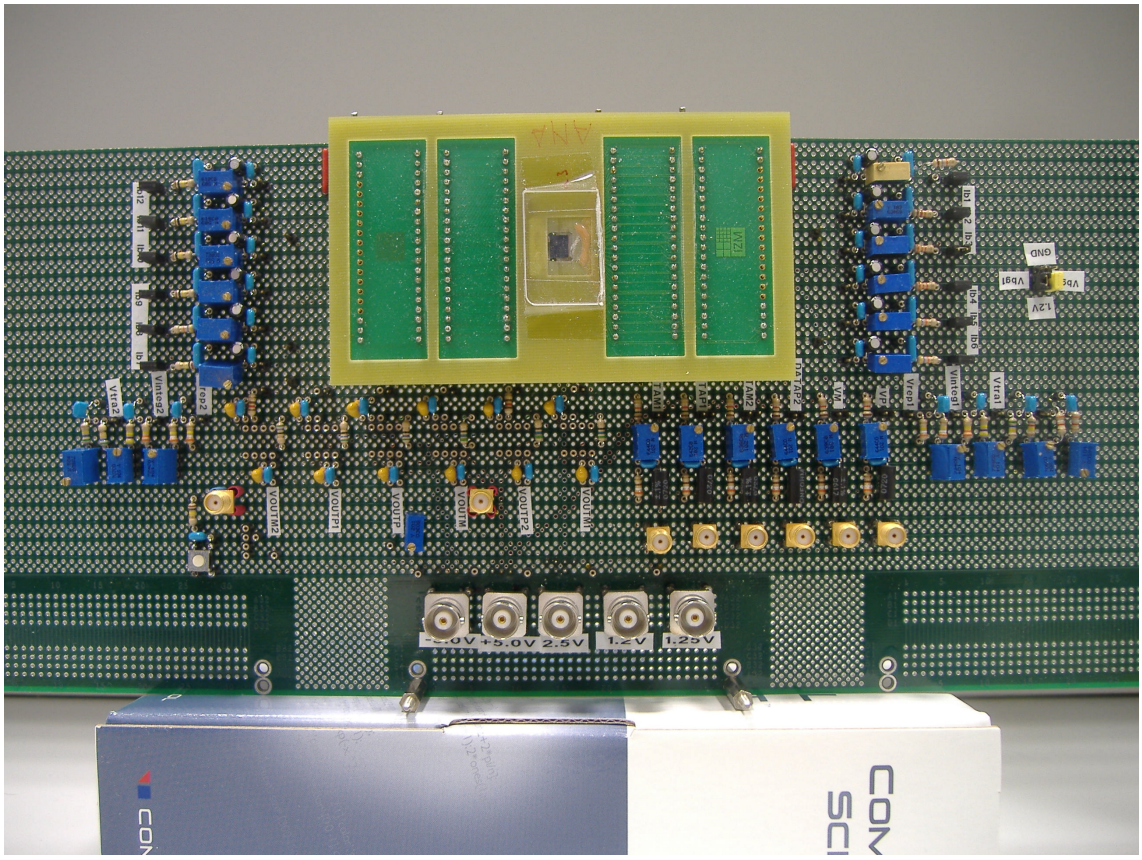



Fig. 3.17 Test Board for the 3D chip-to-chip communication system

# INPUT INTERFACE

The sinusoidal voltages Vsin-i are externally generated by the AGILENT and the Power Splitter/Combiners ZSCJ-2-2. The Splitter is designed to operate in a 50 ohm system. Let us consider the basic lumped element power splitter/combiner circuit of Fig. 3.18. The output transformer has an equal number of turns from the center tap to each end. Therefore, as an auto-transformer (2 to 1 turns ratio) the impedance across the output ends is four times larger than the impedance across the center tap to one end.

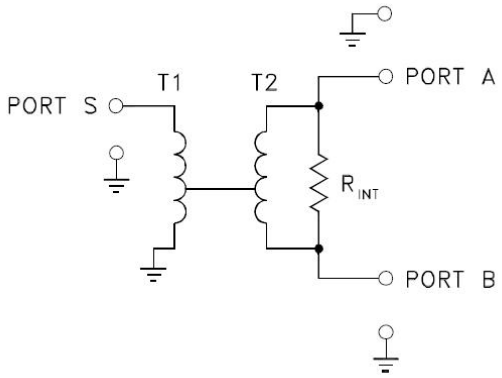



Fig 3.18 two-way splitter/combiner schematic

A 2 to 1 (50 to 25 ohm) input transformer (not shown in the figure) is added at port S so that the port S impedance is matched to the 50 ohm input system. As previously mentioned, the output transformer has a 4 to 1 impedance ratio, so that it has 100 ohm impedance. The resistances at port A and B appear across the output transformer in series, they are required to be equal to the transformer impedance for optimum power match. It is meant ports A and B are each terminated in 50 ohm. Moreover, in order to obtain maximum isolation $R_{in}$ equals the output transformer impedance. As it is shown in Fig. 3.19, a 50 ohm high precision resistance (R1) has

been chosen, followed by a de-coupling capacitor (C1).

After the de-coupling, a periodic signal having a zero mean value is obtained. However the under-test system is projected to work with signals having a mean value equal to half the internal power supply, i.e. 1.25V. The mean value is restored through a voltage divider referenced to ground. The voltage input (i.e. external power supply equals to 5V) is applied across the series resistances $R_{up}$ (i.e. R4 + T1) and $R_{douw}$ (i.e. R5).

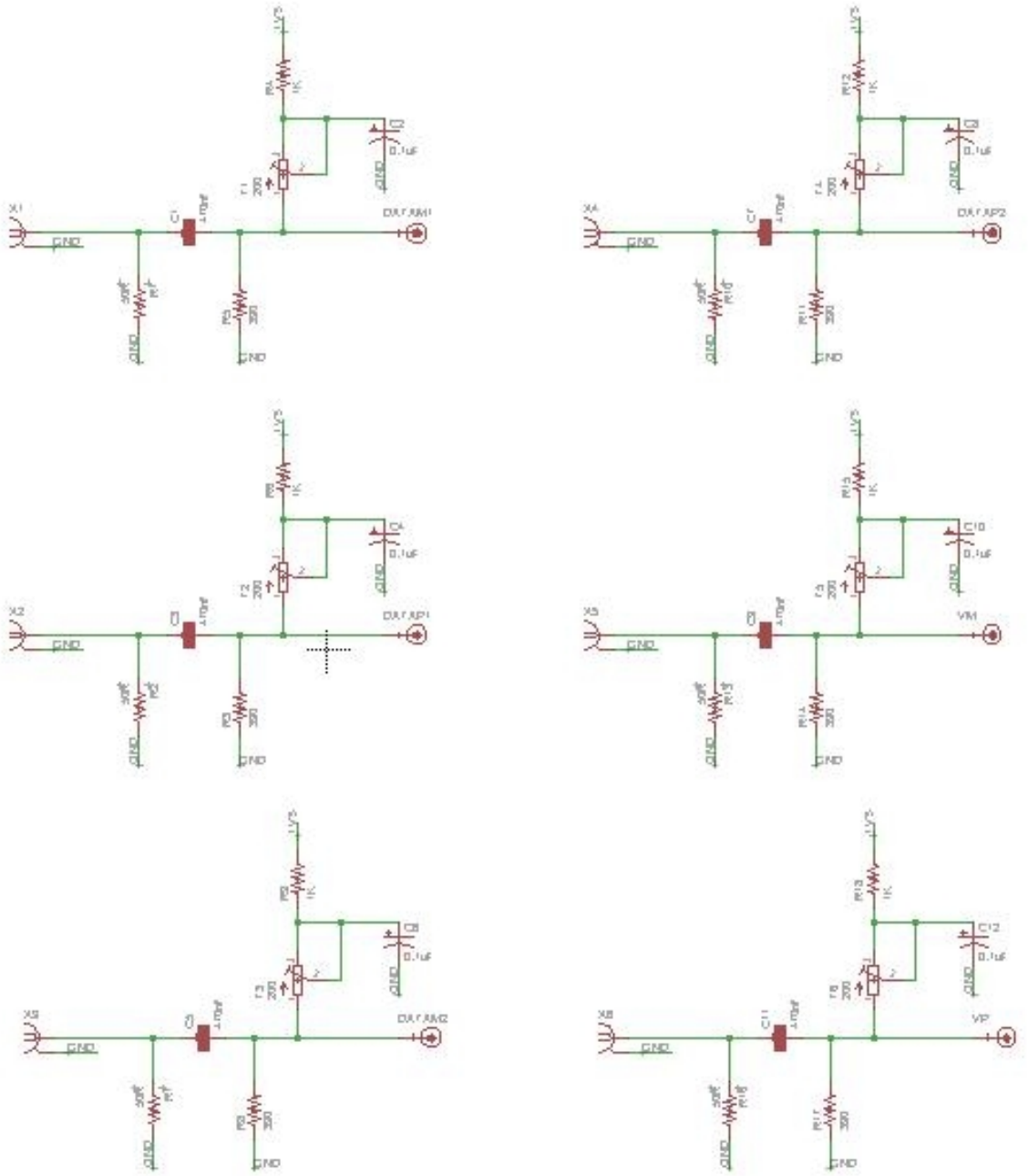


Fig. 3.19 Input interface

Applying Ohm's Law, the relationship between the input voltage, $V_{in}$, and the output voltage, $V_{out}$, can be found (5):

$$V_{out} = \frac{R_{down}}{(R_{up} + R_{down})} V_{in}$$

The input circuit interfaces are placed as closer as possible on the test board. Anyway it can happen that there are different drop voltages due to undesired effects as bad made contacts, different wire lengths and so on Fig. 3.20.

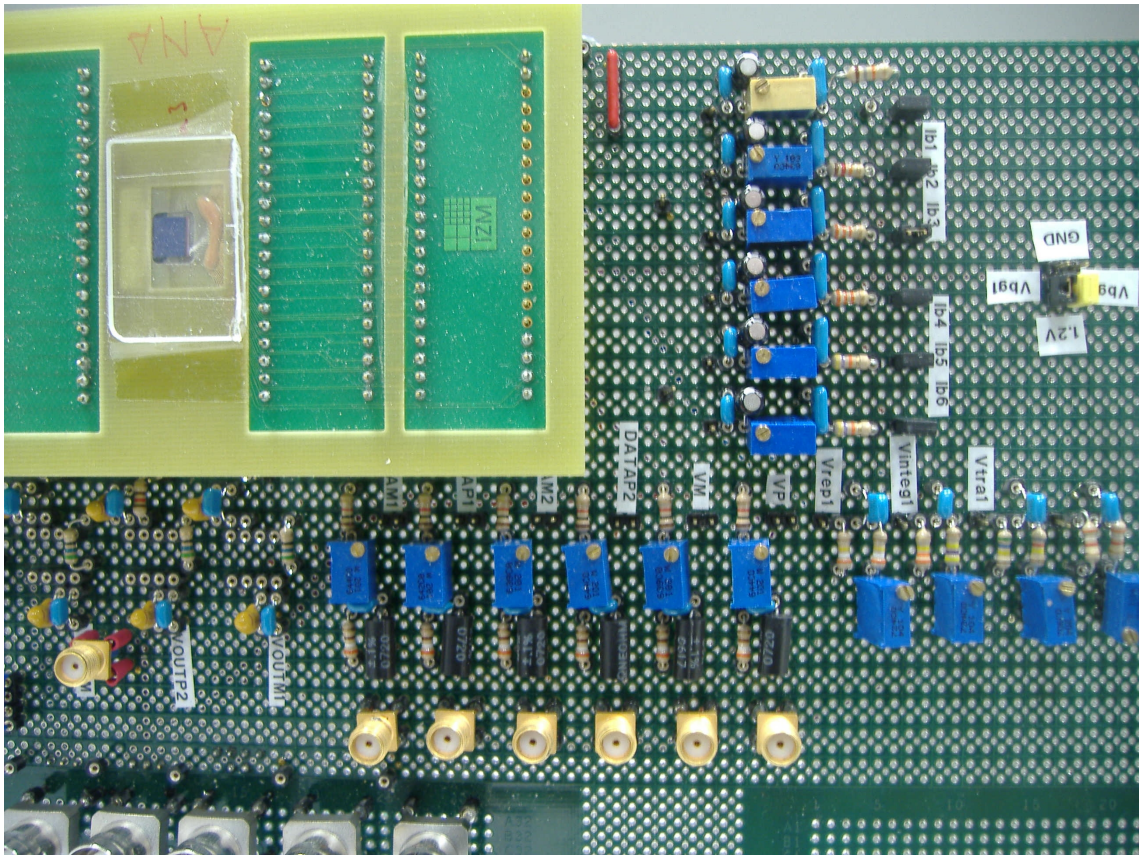



Fig. 3.20 Input interface: BNC, input load resistance 50 Ohm, biasing circuits

In order to guarantee the same mean voltage value on all the inputs, the resistance R4 is implemented by using a resistor plus a variable resistor. It has been chosen to use also a standard resistor to allow the possibility to add a RC filter. It is worth to notice that adequate power supply filtering is critical to reduce noise injection. Since the break frequency is determined by the time constant, it is possible to determine values of R4 and C2 from the expression (6)

$$f_t = \frac{1}{(2\pi\tau)} = \frac{1}{(2\pi R C)}$$

Then, the value of R5 is obtained by insert the values of R4 = 1KΩ, Vin = 5V, Vout = 1.25V in (5). So that R5 = 390Ω.

Finally it is possible to choose the value of C1 = 470uF, remembering that the capacitor should be great to well de-coupling and also it should be:

$$f_t = \frac{1}{(2\pi\tau)} = \frac{1}{(2\pi R_5 C_1)} \ll 100\ KHz$$

Since the system aims to operate in a range from 100 KHz to 20 MHz, the cut-off frequency of the CR high-pass band filter formed by C1 and R5 should be far from 100 KHz.

## BIAS CIRCUITRY

Table IV and Table V report the list of the bias voltages and bias currents with respective values and briefly description.

**Table IV**

**Summary of the used bias voltages**

| | VALUE | DESCRIPTION |
|---|---|---|
| Vinteg1,2 | **1.55V** | **Reference voltage of integrator circuit. Both in TX and RX stages** |
| Vtra1 | **1.50V** | **VGA bias voltage in Transmitter stage** |
| Vrep1 | **1.25V** | **Replica circuit constant input signal, in Transmitter stage** |
| Vtra2 | **1.55V** | **VGA bias voltage in Receiver stage** |
| Vrep2 | **1.25V** | **Replica circuit constant input signal, in Receiver stage** |
| Vpol* | **1.25V** | **Bias voltage of Receiver node** |
| VBG1** | **1.20V** | **Band Gap Voltage Reference in Transmitter stage** |
| VBG2** | 1.20V | **Band Gap Voltage Reference in Transmitter stage** |

* Vpol is directly furnished from the external power supply unit or it is generated within another interface that implements an adjustable voltage battery as it will be explained in the next section.

** I final implementation, VBG1 and VBG2 are generated within the system by the basic Voltage Reference Cells as previously mentioned. Although in the current prototype, since it aims to validate the architecture, the Cells are not integrated, thus VBG1 and VBG2 will be furnished from the power supply unit.

It is worth to notice that the reference voltage for the integration stage is the same both in transmitter circuit and receiver circuit. While, in the case of the VGA bias voltage (as well as for Replica constant input signal), the voltages are derived

separately for the TX and RX, although they theoretically have the same value. The reason is to allow more possibilities to compensate errors by differently tuning the bias voltages. Obviously, it is necessary only in test applications in order to better understand the behaviour of the prototype.

**Table V**

**Summary of the bias currents**

| | VALUE | DESCRIPTION |
|---|---|---|
| IBIAS1 | **13.0uA** | **Transmitter VGA** |
| IBIAS2 | **10.0uA** | **Transmitter Buffer Stage** |
| IBIAS3 | **12.5uA** | **Transmitter Half Wave Rectifiers** |
| IBIAS4 | **12.5uA** | **Transmitter Voltage Shifter** |
| IBIAS5 | **12.0uA** | **Transmitter V-I Converters** |
| IBIAS6 | **12.0uA** | **Transmitter Current Integrator** |
| IBIAS7 | **12.0uA** | **Receiver VGA** |
| IBIAS8 | **12.0uA** | **Receiver Buffer Stage** |
| IBIAS9 | **12.0uA** | **Receiver Half Wave Rectifiers** |
| IBIAS10 | **12.0uA** | **Receiver Voltage Shifter** |
| IBIAS11 | **12.0uA** | **Receiver V-I Converters** |
| IBIAS12 | 12.0uA | **Receiver Current Integrator** |

The considerations expressed above hold true also in that case. The several bias voltages are a fraction of the same input voltage (2.5V). In the same manner, the bias currents are determined through a voltage divider, after calculation of the corresponding voltage values. The Vbias and Ibias values have been evaluated by SPICE simulation; anyway it is expecting that the behaviour of the real device slightly differ from it. Thus it is important to allow modifying the biasing values separately. It is simply accomplished by tuning the variable resistors used in each row of the voltage divider. Each bias current refers to a single specific component [i.e. Ibias1 refers to VGA circuit, Ibias2 refers to Buffer Stage, Ibias3 refers to Half wave rectifier, ecc…]. Every bias current is applied to the several stages within the System through the dedicated

current mirrors. So that, the total bias current is calculated as the number of current mirrors multiplied for the bias current value required by each component. The voltage values, current mirrors number, thus total current values and finally associated resistance values are displayed in Table VI.

**Table VI**

**Values of the bias currents and voltages and the related resistance values**

| BIAS CURRENT | n. Current Mirrors per each component | | | TOTAL CURRENT | EQUIVALENT VOLTAGE | R(Ω) | TRIM RANGE |
|---|---|---|---|---|---|---|---|
| | 3D | 2D | TEST | | | | |
| **IBIAS1** | 1 | 4 | 2 | 91.0uA* | 1.817V* | 47K* | 10K* |
| **IBIAS2** | 3 | 6 | 4 | 130.9uA | 1.525V | 12K | 10K |
| **IBIAS3** | 1 | 2 | 1 | 50.0uA | 1.600V | 33K | 5K |
| **IBIAS4** | 1 | 2 | 1 | 48.8uA | 1.636V | 33K | 5K |
| **IBIAS5** | 1 | 2 | 0 | 35.8uA | 1.856V | 47K | 5K |
| **IBIAS6** | 1 | 2 | 0 | 36uA | 1.852V | 47K | 5K |
| **IBIAS7** | 3 | 0 | 0 | 17.2uA* | 2.189V* | 100K* | 50K* |
| **IBIAS8** | 3 | 0 | 0 | 36uA | 1.852V | 47K | 50K |
| **IBIAS9** | 1 | 0 | 0 | 12uA | 2.276V | 120K | 50K |
| **IBIAS10** | 1 | 0 | 0 | 12uA | 2.276V | 120K | 50K |
| **IBIAS11** | 1 | 0 | 0 | 12.4uA | 2.264V | 120K | 50K |
| **IBIAS12** | 1 | 0 | 0 | 12.2uA | 2.278V | 120K | 50K |

| BIAS VOLTAGE | VALUE | R_UP | R_DOWN | TRIM. RANGE |
|---|---|---|---|---|
| **Vinteg1,2** | 1.55V | 330K | 470K | 100K |
| **Vtra1** | 1.50V | 330K | 570K | 100K |
| **Vrep1** | 1.25V | 390K | 330K | 100K |
| **Vtra2** | 1.55V | 330K | 570K | 100K |
| **Vrep2** | 1.25V | 390K | 330K | 100K |

* It should be noted that the bias current values required on the silicon implementation generally slightly differ from the values previously estimated with SPICE simulations. However the consideration is not valid in the cases of Ibias1 and Ibias7 (i.e. VGA bias current in Transmitter and Receiver respectively), since a great discrepancy is proven. In fact the theoretical value is in the range of 11-13uA, while the value used during the test should be around 6-9uA. When the nominal value is applied to the prototype, the VGA operating point settles to 1.1V which is largely lower on the estimated value. In that condition, the cascade stages (i.e. half wave rectifiers, V-I converters) cannot operate properly. Thus the value of Ibias1 (Ibias7) should be tuned till reaches the correct operating voltage value.

In order to reduce the noise injection from power supply 1uF capacitors improve low-pass band filters. Moreover extra capacitors are required to stabilize the nodes VDD (+2.5V) or VSS (+5V) or VEE (-5V). A parallel combination of 100uF, 10uF, 0.01uF is recommended. The whole Bias Interface is sketched in Fig. 3.21.

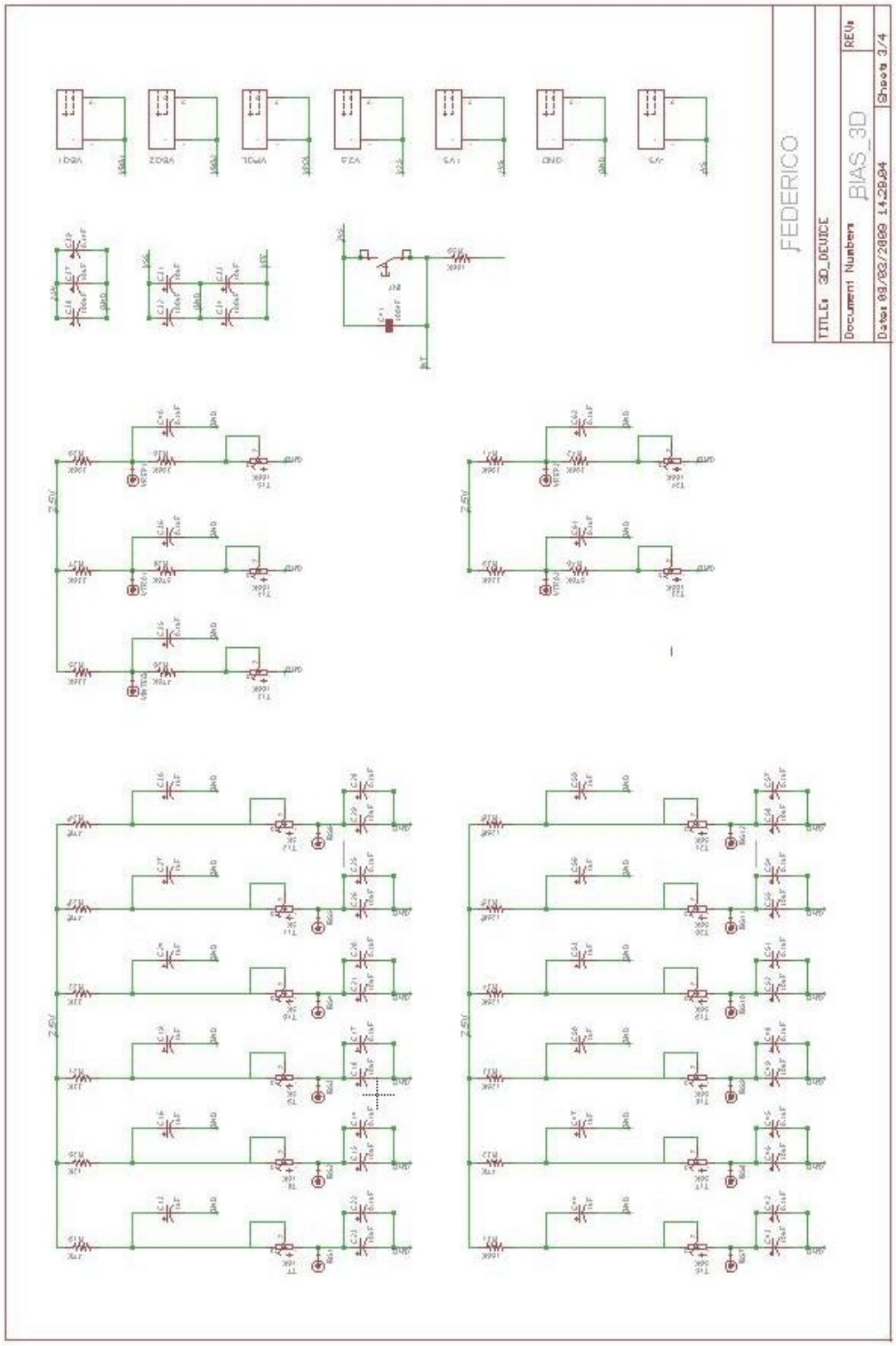



Fig. 3.21 Current and Voltage polarization of the proposed system

# OUTPUT INTERFACE

The Output interface connects the Chip's PADs with the external measurement devices. The proposed system was already furnished with internal buffers. Anyway they were projected to drive purely capacitive loads on the range of few pF. The external interface has to drive, without distortions, the output load consisting in probes' input load, cables impedance, test-board parasitic capacitance. Moreover it should allow driving resistive loads. It is worth to notice that the Infiniium MSO8104 Oscilloscope can be used either in 50 ohm input configuration or in high impedance input configuration. Anyway the interface has been implemented to allow to perform spectral analysis. In the case that a Spectrum Analyzer will be used, it obligatory requires a 50 ohm input load. Also if we do not operate in high frequency domain, it should be noted that the system aims to cover the frequency range from 100 KHz to 20 MHz, thus a wide small and large signal bandwidth as well as a low distortion are needed. For these reasons the suitable operational amplifier has to: drive both capacitive and resistive loads, have small input capacitance (-pF), exhibit large signal bandwidth greater than 20 MHz. The presented specifications are fairly well satisfied by the AD9631 integrated. The AD9361 is furnished by Analog Devices [7]. It is a very high speed and wide bandwidth amplifier that combines many of the best characteristics of both current feedback and voltage feedback amplifiers. Since their open-loop frequency response follow the conventional 6dB/octave roll-off, his gain product is basically constant. Increasing the closed-loop gain, results in a corresponding decrease in small signal bandwidth. The AD9631 is unity gain stable; the chosen non-inverting configuration is sketched in Fig. 3.22. The value of the feed-back resistor is critical for optimum performance at minimum stable gain (+1). After several experiments on the current set-up, it has been fixed to 57 ohm, such value slightly differ from the suggested one Rf = 140 ohm. This resistor acts as a parasitic suppressor against oscillations that can occur due to lead (input, feed-back) inductance and parasitic capacitance.

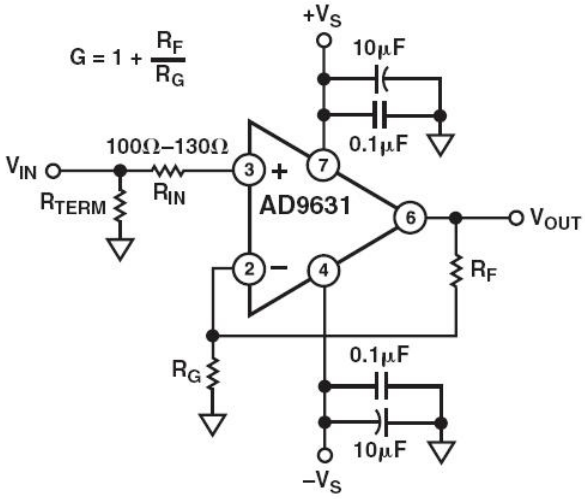



Fig 3.22 AD9631: non-inverting configuration

The chosen value of Rf provides the best combination of wide bandwidth, low parasitic peaking, and fast settling time. For the same reasons, a 120 ohm resistor should be placed in series with the positive input. The AD9631 large signal frequency response in non inverting configuration is sketched in Fig 3.23.

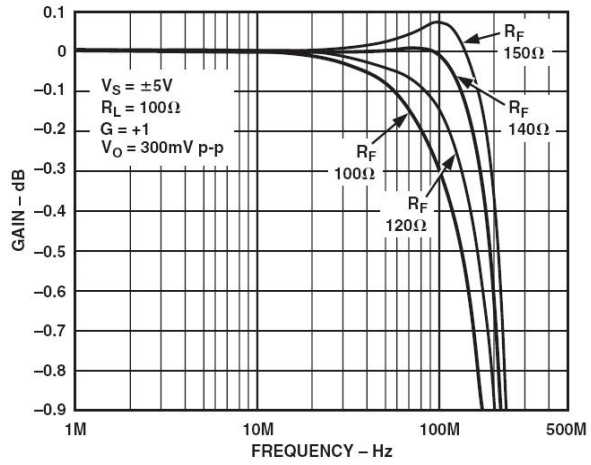



Fig 3.23 AD9631: large signal frequency response

## CHAPTER 3: 3D CAPACITIVE TRANSMISSION OF ANALOG SIGNALS

Adequate power supply bypassing can be critical when optimizing the performance of a high frequency circuit. Inductance in the power supply leads can form resonant circuits that produce peaking in the amplifier's response. In addition, if large current transients must be delivered to the load, the by-pass capacitors typically greater than 1uF will be required to provide the best settling time and lowest distortion. In order to cover a wide operating frequency range a parallel combination of 0.1uF and 10uF capacitors has been used.

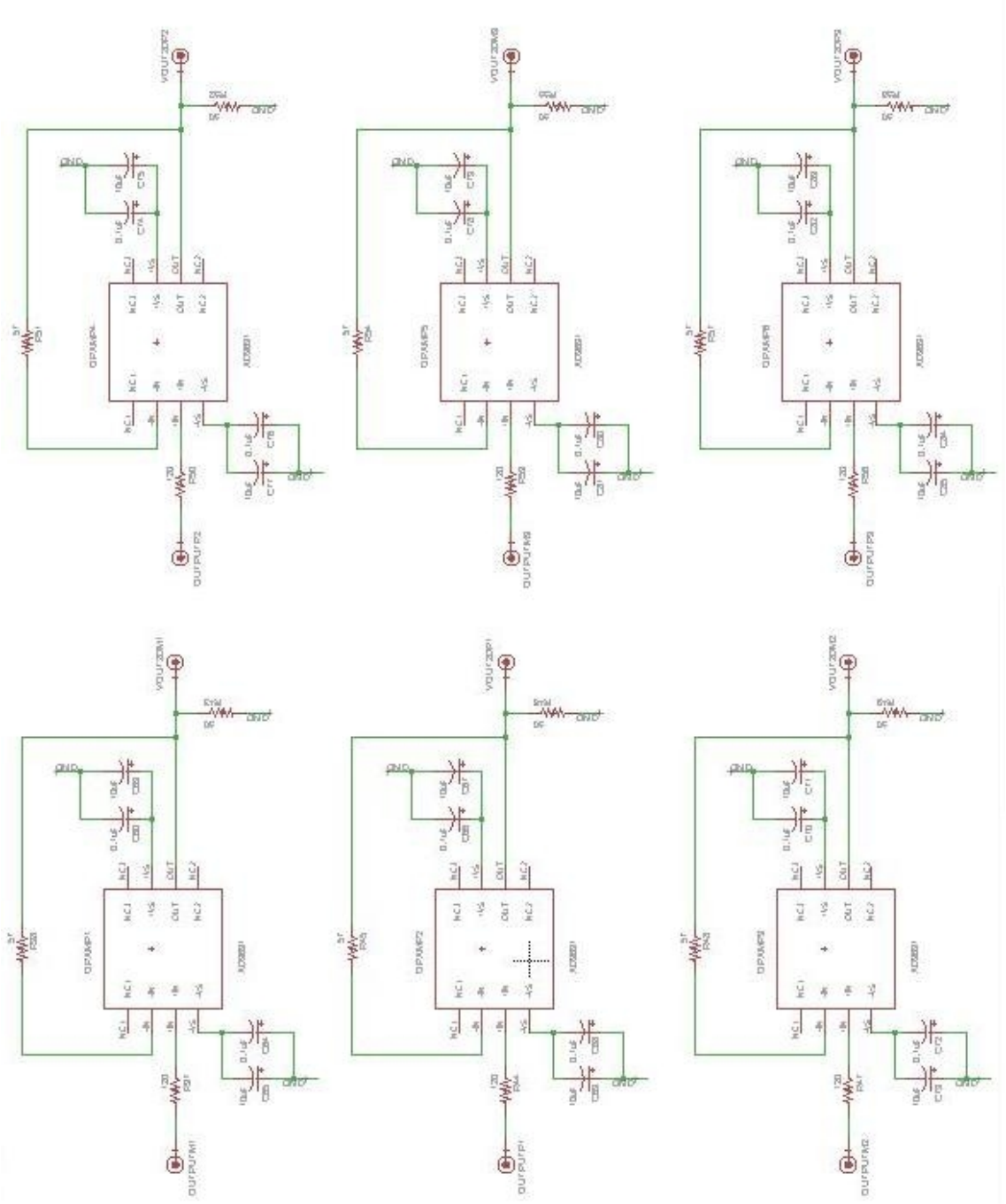


Fig. 3.24 Output interface

## ADJUSTABLE VOLTAGE BATTERY

Dirty power originates from the external AC mains can affects the quality of the unit power supply; transient pulses and associated noise can flow to the facilities. Moreover fluctuation from electrical equipment can cause cumulative and equally damaging power hazard. Since the test prototype is very sensitive, even a small appliance can cause problems. Obviously every measurement equipment is furnished with power filters and surge protectors that guard against spikes power surges and other forms of damaging electrical disturbance. Despite that glitches can occur during the operate mode. The proposed system is robust; however it has a weak point in the Receiver polarization node. The biasing circuit consists in a MOSFET constantly kept in off state, connected to Vbias. So that it acts as very high resistance.

It is the simplest solution and works well in normal operating conditions. Under external over-excitation it can turn on, thus it injects current on the receiver node. As immediate consequence the voltage drops from 1.25V to 0.7V or less, resulting in a disease of the system. Anyway these phenomena are transient effects. The voltage slowly recovers to the nominal value, since the process has a transient time constant RC, where R is around hundreds mega-ohms. But in worst case, when a spike occurs, the generated over-current can irreversibly damage the device, so that the prototype is not longer utilizable. In order to avoid these events the standard power supply unity it has been replaced with an adjustable voltage battery. Constraints that affect power supplies are the amount of power they can supply, how long they can supply it without needing some kind of refueling or recharging, how stable their output voltage or current is under varying load conditions, and whether they provide continuous power. The regulation of power supplies is done by incorporating circuitry to tightly control the output voltage and/or current of the power supply to a specific value. The specific value is closely maintained despite variations in the load presented to the power supply's output, or any reasonable voltage variation at the power supply's input. The stabilized power supply is sketched in Fig. 3.25; it basically consists in a standard PP3 battery

(9V) and a LM117 3-Terminal adjustable output.

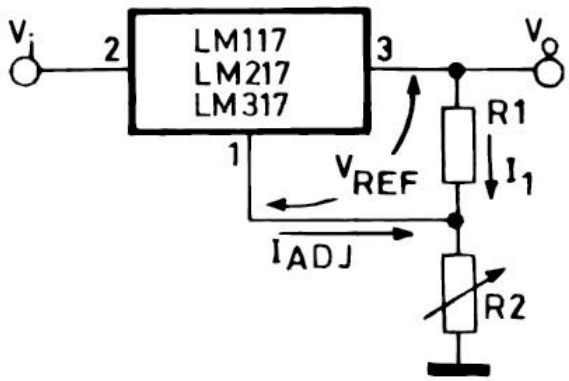



Fig. 3.25 Voltage regulator circuit using the LM317.

The LM317 positive voltage regulator is capable of supplying in excess of 1.5A over a 1.2V to 37V output range. In operation, the LM317 develops a nominal 1.25V reference voltage VREF, between the output and adjustment terminal. The reference voltage is impressed across program resistors R1 and, since the voltage is constant, constant current I1 then flows through the output set resistor R2, giving an output voltage Vout:

$$V_{out} = 1.25\ (1 + \frac{R_2}{R_1}) + I_{ADJ}(R_2)$$

Since the 100uA current from the adjustment terminal represent an error term, the LM317 is designed to minimize Iadj and make it very constant with line and load changes. To do this, all quiescent operating current is returned to the output establishing a minimum load current requirement. If there is insufficient load on the output, the

output will rise. The LM317 is capable of providing extremely good load regulation, a few precaution are needed to obtain maximum performance. The proposed value of the current set resistor connected between the adjustment terminal and the output terminal is 240 ohm. The resistor should be tied directly to the output case of the regulator rather near the load, in order to eliminate line drops from appearing effectively in series with the reference and degrading regulation. The values of R2 used to derive the core voltage supply (2.5) and the bias voltage (1.25) are calculated from formula (11), respectively 240 ohm and 0 ohm. Moreover a 1.0uF tantalum input bypass capacitors is added to reduce the sensitivity to input line impedance. The adjustment terminal may be bypassed to ground improve ripple rejection. This capacitor $C_{ADJ}$ prevents ripple from being amplified as the output voltage is increased. An output capacitance Co in the form of 10uF aluminum electrolytic capacitor is required for stability.

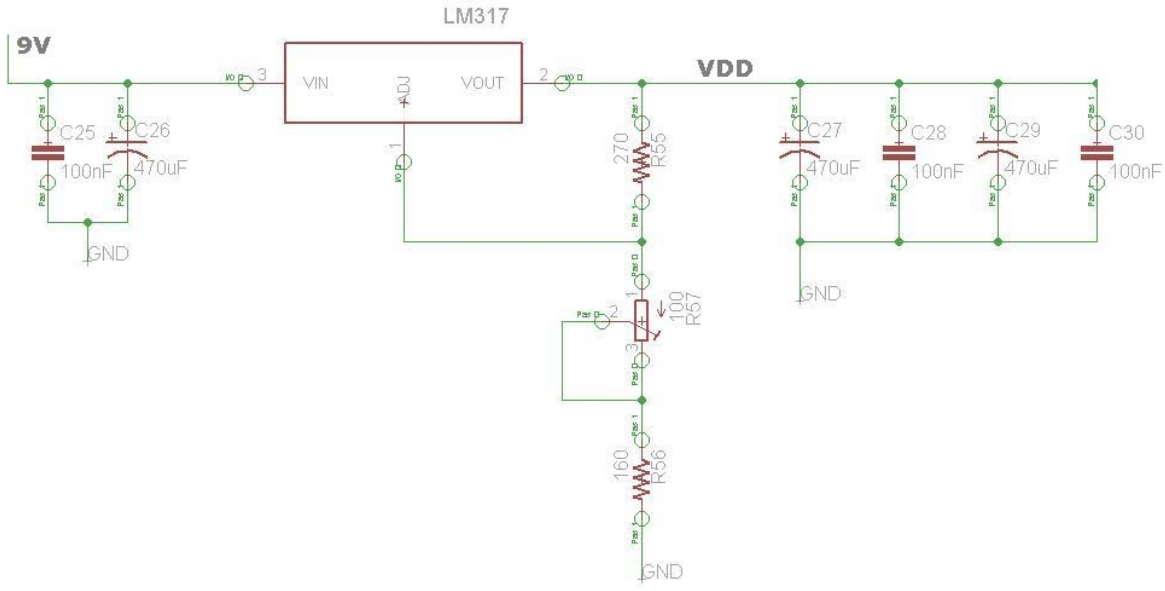



Fig. 3.26 Implemented adjustable voltage circuit with LM317

# RESULTS

A test chip devised for realization of 3D prototypes by stacking identical dice has been constructed in 0.09μm CMOS technology (Fig. 3.13). Thick oxide transistors were used in order to allow for a 2.5V supply. The design aims at 20MHz bandwidth. The operating frequency of the AGC circuit was chosen at 3MHz. A higher frequency would make the design of the rectifiers more critical, while a lower frequency would require a larger size for the integrating capacitor. Two signal channels and one AGC channel were included in the prototype. The size of the electrodes is 30x30μm$^2$. The 20pF integration capacitor $C_{int}$ is implemented by an MOM structure. The only analog option is the unsalicided poly used in the resistances $R_F$ of the VGA. The sinusoidal voltage $V_{SIN}$ and the band-gap voltage are externally generated. In Fig. 3.27-left the measured input and output waveforms of the AGC channel are shown for changing sinusoidal signal amplitude. As expected, the output amplitude settles to a constant value independent of the input one.

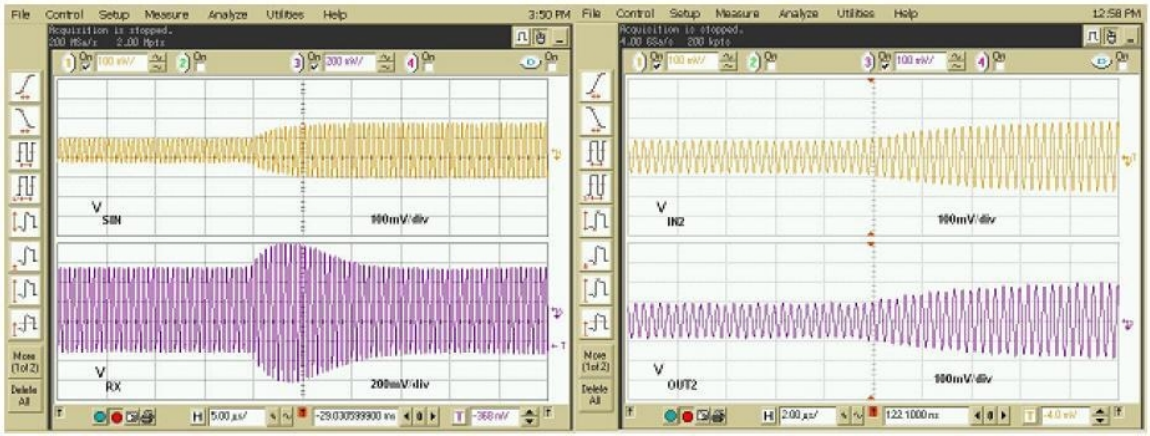



Fig. 3.27 Measured input output waveforms when a sinusoidal signal of varying amplitude is applied: AGC Channel (left) signal Channel2 (right)

When the amplitude of the input signal changes, the AGC automatically varies the $V_{cntr}$ that feeds the Variable Gain Amplifier, decreasing the transconductor gain. Thus, the decrease of amplification will compensate the increase of the input signal amplitude. The control channel output signal amplitude remains constant, with his value fixed by the voltage reference cells as fraction of the Band Gap voltage .

Fig. 3.27-right shows an example of the measured input and output waveforms of one of the two signal channels demonstrating the functionality of the analog transmission. The VGA gain is automatically changed in order to follow the amplitude variations on the input signal.

The in-band gain @3MHz vs. input signal amplitude for the two channels is shown in Fig. 3.28.

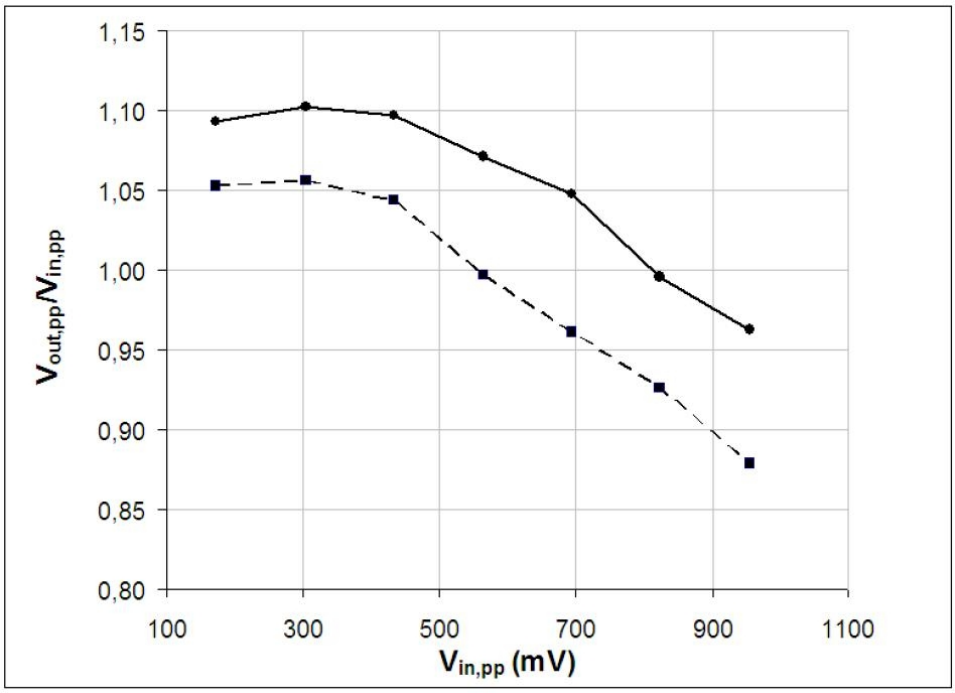



Fig. 3.28 Voltage gain vs. differential input peak-to-peak voltage amplitude for signal 1 (dot) and 2 (square) @3MHz

The gain difference between the two channels is entirely due to mismatches between the two attenuation factors, while the gain error with respect to the ideal value of one is due to mismatches between each signal channel and the AGC channel and to the non-ideality of the AGC circuit. The non-linearity can mainly be ascribed to defects in the rectifiers and the VGA.

The errors introduced by the rectifier circuits have already been described in the dedicated section and they can be mitigate.

In Fig. 3.29 the behaviour of the single VGA is depicted for different frequencies and different signal amplitudes.

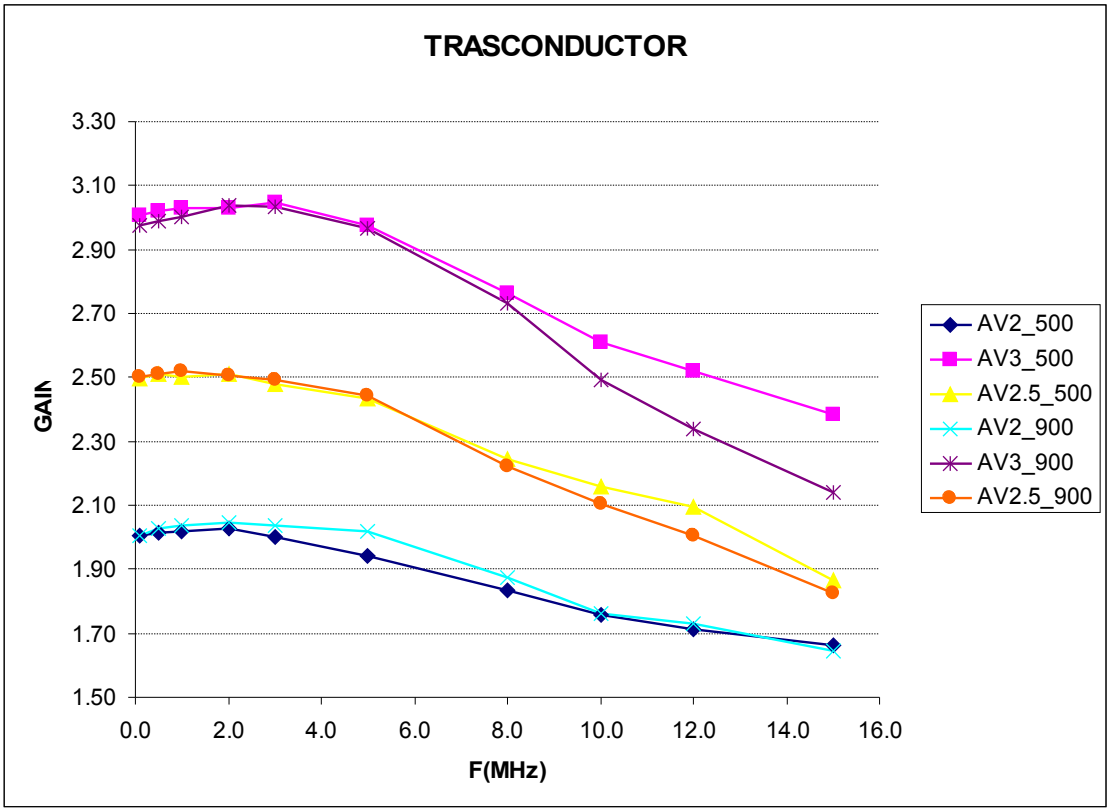



Fig. 3.29 Ratio (Vout/Vin) for the VGA circuit with different forced gains (i.e. equals to 2, 2.5, 3) and different input signals amplitudes, on the frequencies range from 100KHz to 15 MHz. It shows the dependence from the frequency and gain

The measurements have been performed on the single VGA component implemented within the structure test, thus apart from the system in feed back loop. The component is the same used inside the AGC circuit but in this case the control voltage is not automatically tuned but it is fixed by an external signal. It allows to observe the behaviour of the cell for different frequencies, amplitudes with forced amplification gains. Fig.3.29 shows that the VGA introduces a distortion depending by the operating frequency.

The VGA response is also related to the signal amplitude as sketched in Fig. 3.30. Measurements have been performed on the single VGA circuit (in Test cell) and on the output of the 3D system. It has been used the same input signal (f = 3MHz), varying the signal amplitude. Fig. 3.28 already shown the relation between system gain and amplitude, but this behaviour is mostly due to the VGA.

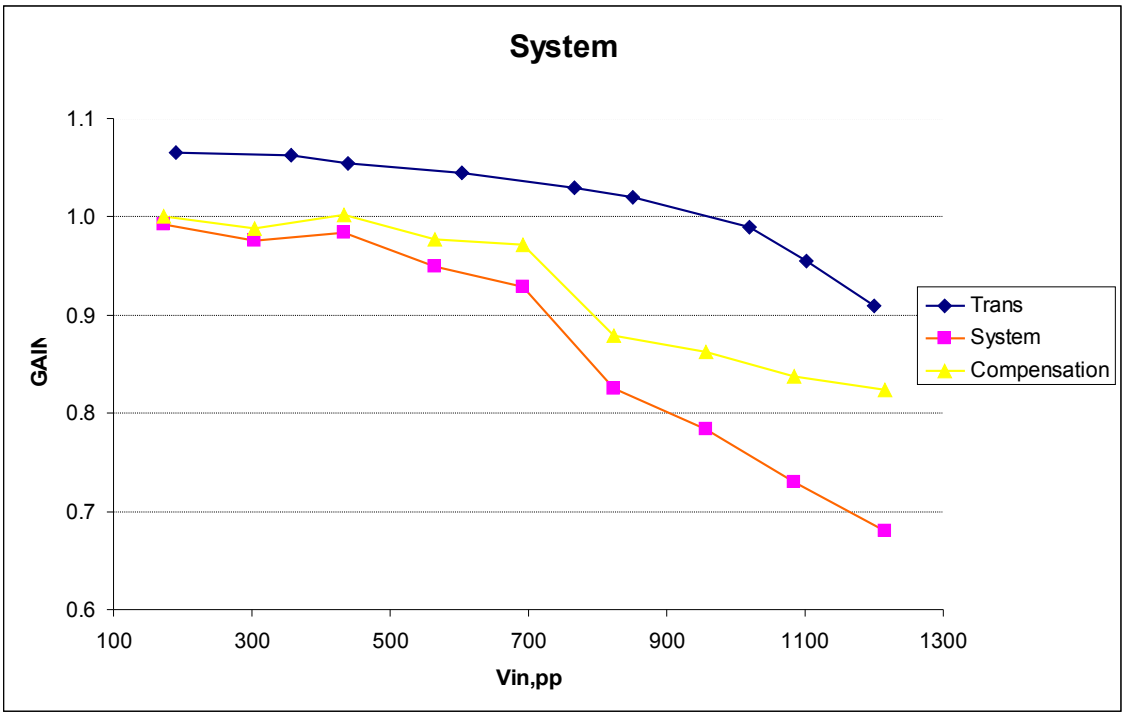



Fig. 3.30 In-band gain @3MHz vs. input signal amplitude for single Variable Gain Amplifier, System, and System taking in account the error introduced by the VGA

Fig. 3.30 reports the gain vs. input signal amplitude for single VGA (blue line), System (orange line) and System compensation (yellow line). The system compensation does not mean a compensation technique implemented on the device, but it is realized after the measurements tacking in account the data obtained and the errors introduced by the VGA. Thus, we think that in further realizations, it will be necessary to re-project only the rectifier stages and the VGA circuit in order to reduce the inaccuracies and obtain better performances. Anyway it has been implemented and validated the principle and the architecture necessary to chi-to-chip transmission the analog signals, with 3D techniques based on capacitive coupling.

Finally, in Fig. 3.31 the gain of channel 2 vs. frequency is shown for two input amplitudes with the AGC circuit on and off.

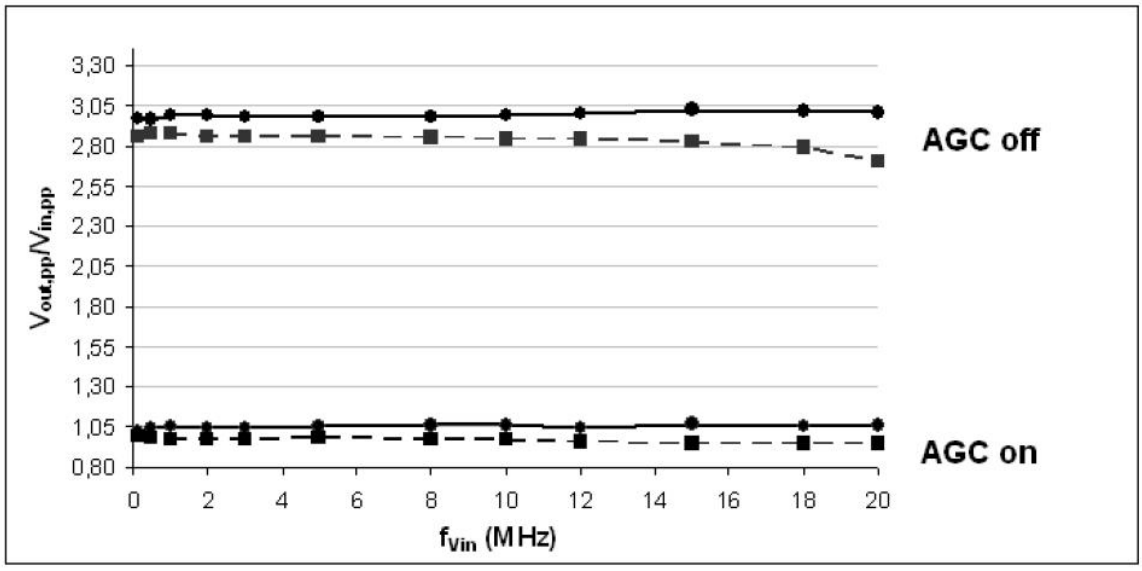



Fig. 3.31 Voltage gain vs. frequency (100KHz to 20 MHz) for signal Channel2 with AGC on and off. Voltage amplitude Vin2 is equal to 500 mV (dot) and 700 mV (square)

## CHAPTER 3: 3D CAPACITIVE TRANSMISSION OF ANALOG SIGNALS

Without any calibration the ratio between input and output signals will set to a random value. While using the calibration system the ratio is maintained around 1. Thus, it is demonstrated that it is possible recover the initial signal after the inter chips transmission.

The coupling between the AGC and the signal channels is responsible for two output tones at 2MHz and 4MHz of amplitude less than 0.5mV and 2mV, respectively.

The inter-channel cross modulation $V_{OUT,i}$ / $V_{IN,j}$ with $i \neq j$ is less than 0.002 (measurements realized with the Spectrum Analyzer).

The area of the AGC circuit (equal in TX and RX) including the 3D electrodes is 138x191µm2 and is mainly due to capacitance $C_{INT}$ (Fig. 3.14). The area of each signal channel (90x30µm$^2$) is dominated by the electrodes. The power of the AGC-TX and RX circuits is 3.6mW at 2.5V, while 1mW is required by each signal channel.

# THROUGH SILICON VIA (TSV) TECHNOLOGY

The ever storage capacity demanding products as MP3, mobile phones, digital still camera... are asking for smaller size packages and higher memory density. In response to the performance and density requirements, the semiconductor industry has definitively moved from 2D to 3D style packages with shorter electrical connections. 3D stacking of die with TSV (through Silicon Via) connection as well as wafer level packaging of CMOS image sensors (CIS) are becoming very hot topics. Wafer level packaging technologies, e.g. CSP with redistribution layers or flip chip mounted devices on wafer, are already introduced in high volume production. Currently, different technologies which use Through Silicon Vias (TSV) in active or passive silicon devices are in development to satisfy the need to increase performance and functionality while reducing size, power and cost of the system. Today, there are two mainstreams to realize TSVs. One is the implementation into the front-end CMOS process and the second is a post front end process (via first/via last process). Both scenarios have pros and cons and

the selection depends on application and infrastructure. The post front end process allows the realization of compact 3D system architectures as a packaging task with complete tested device wafers independent from the device wafer manufacturer. Key process technologies enabling 3D architectures with TSV interconnect include:

- Via formation with high aspect ratio,
- Isolation, barrier and seed deposition,
- Via metal filling, redistribution lines (RDL),
- Wafer thinning,
- Thin wafer handling and transfer processes,
- Assembly: wafer/chip alignment, adjusted bonding

Most of those 3D technologies are quite new to the packaging industry and require a FE/BE infrastructure. That's why 3D-IC architectures are today still at the R&D stage, even in the largest IC companies, but they are in focus as a potential solution with a high priority. Many of the key technical issues and challenges for TSV interconnects are not fully resolved yet. There are also a number of alternative technologies, e.g.:

- Process integration: via-first vs. via-last
- Via filling: materials (e.g. poly Si, Cu, W, conductive polymer, metal paste) and techniques (e.g. electroplating, CVD, polymer coating),
- Wafer level assembly: chip-to-chip chip-to-wafer or wafer-to-wafer,
- Bonding: soldering, direct Cu-Cu, adhesive, direct fusion

The development of selected technical parameters for TSVs is given in Table I. This data are provided by ITRS and postulates volume production.

Table 1: Key technical parameters for stacked architectures using TSV

| *Year of Production* | *2007* | *2008* | *2009* | *2010* | *2011* | *2012* | *2013* |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| **Numbers of stacked die using TSV** | 3 – (8) | 6 | 9 | >9 | >9 | >9 | >9 |
| **Minimum TSV pitch** | 10.0 | 8.0 | 6.0 | 5.0 | 4.0 | 3.8 | 3.6 |
| **TSV maximum aspect ratio**** | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 |
| **TSV exit diameter (µm)** | 4.0 | 4.0 | 3.0 | 2.5 | 2.0 | 1.9 | 1.8 |
| **TSV layer thickness for minimum pitch** | 50 | 20 | 15 | 15 | 10 | 10 | 10 |

**This applies for small diameter vias. The larger diameter vias will have a smaller aspect ratio

# TSV FABRICATION PROCESS STEPS

## DRIE PROCES

3D stacking of die through TSV can be done with different strategies: from the "Via first" approach (front-end process) to the "Via last" approach (back-end process). Each of these different ways of elaborating the Vias has its advantages and drawbacks in terms of electrical performances, refilling materials and cost, but they have in common the need to etch the Vias. The D.R.I.E. (Deep Reactive Ion Etching) process is used for the definition of the different via shapes, depths, and sizes. The most common used process is known as the "Bosch" process and it alternates short step of SF6 plasma for the fast but isotropic removal of silicon, with short C4F8 plasma deposition step for the sidewalls protection. This process has the potential to deliver very high and selective etching rate.

The fabrication of TSV includes tree major steps: i) silicon via etching, ii) via insulation and iii) via metallization. Depending upon the positioning of the TSV sequence within overall wafer manufacturing, there are many different process flows.

The main standards can be summarized as:

- “Via First prior to FEOL” approach
- “Via First after BEOL” approach
- “Via Last after BEOL” approach

## “Via First prior to FEOL” TSV

Via first prior to FEOL is the strategy where the Vias are done on the blank Si wafer prior to any CMOS process (Fig. 4.1). Due to the step position in the wafer fabrication process, it can be realized by the CMOS manufacturer, or even the wafer supplier, but not by the packaging company.

Fig. 4.1 “Via First prior to FEOL”. From left to right: D.R.I.E., Oxidation, Poly-Si refilling

Due to all the further CMOS manufacturing steps, the refilling material has to be resistant to all the thermal processes, usually at higher than 1000 degrees. That's why the mostly used refilling material is poly-silicon. As the poly-silicon refilling process can be achieved only in narrow features, the etched features have a width lower than 5μm. One advantage of such process is that it doesn't require any seed layer, and the isolation layer can be easily done using a traditional oxidation process. As the wafer will be thinned down at the latest stage to a thickness around 150μm, the less than 5μm wide features have to be etched down to at least 150μm, which means a more than 30 aspect ratio. The Aspect Ratio (A.R.) of a given feature, like a trench, is defined as the ratio between the depth of the trench and the width. Typical Via First prior to FEOL etch is sketched in Fig. 4.2.

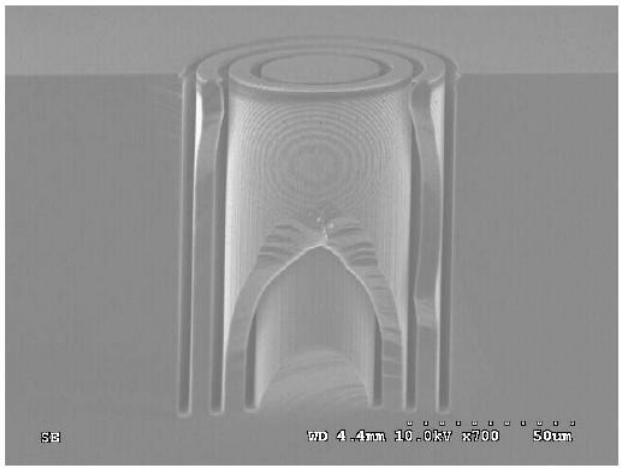



Fig. 4.2 Typical "Via First prior to FEOL etch"

"S.H.A.R.P." (Super High Aspect Ratio Process) process [1] can be used to achieve the best etch performances. It is demonstrated (Alcatel patent) that an Aspect Ratio as high as 110 is achievable.

Similar processes from a technological standpoint has been developed by Silex [2]. It

provides down to 30 μm pitch for through wafer connections in up to 600 μm thick substrates. The described via process is primarily a "via-first" process as it needs a high temperature step during filling. The schematic cross section of via substrate feature is sketched in Fig. 4.3.

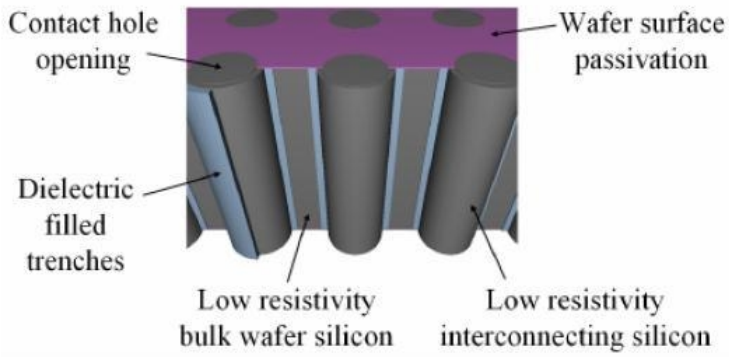



Fig. 4.3 Schematic cross section of Silex process

The "via-first" approach is in most cases very suitable for a MEMS processing perspective. In fact, in most often there is a desire to minimize additional wafer level post processing and handling of for example released MEMS structures since these can be rather delicate. Moreover the completed via substrate allows more or less unlimited processing of the substrate, even at temperatures exceeding 1000°C. This means the Via substrate wafer will be able to enter the MEMS device manufacturing flow as part of the starting material.

## “Via First after BEOL” TSV

In this approach, the TSV will be realized once the CMOS device will be completed and before the grinding process for wafer thinning (Fig. 4.4). The advantage of this strategy is that the CMOS structure is already finished and passivated. At this stage, since the wafer will no longer be exposed to high temperature cycles, the use of via copper refilling material with much better electrical and thermal properties compared to poly-silicon is allowed. Typical features size for those “Vias first after BEOL” are diameters ranging from 10 to 40µm etched at depth of 70-120µm.

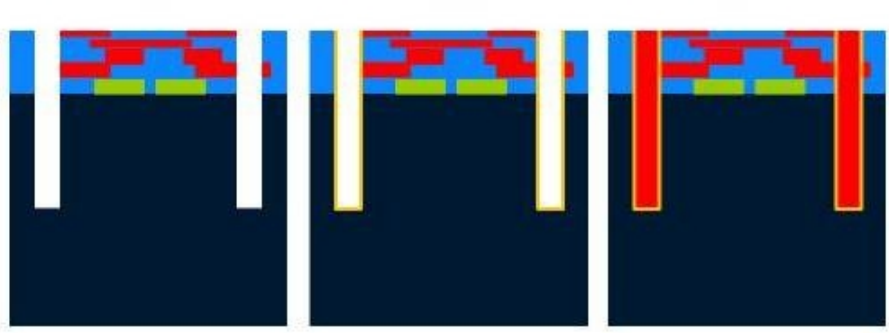


Fig. 4.4 “Via First after BEOL”. From left to right: D.R.I.E., Oxidation, Copper refilling

After etching and before refilling the via with copper, a dielectric layer and later a barrier and a seed layer have to be applied using CVD and PVD techniques (Fig. 4.5). For a better coverage of the metal layer a “Open Mouth” profile has been developed.

This approach consists in making a facet at the top part of the hole, enlarging the entrance of the Via hole. Starting with the same mask design used in standard flow, the process consists in a first step during which a highly tapered profile is applied through the use of mixed SF6 + C4F8 plasma chemistry and then to apply a pure anisotropic

process as shown on the Fig. 4.6. Fig. 4.7, shows the result of such a combined process.

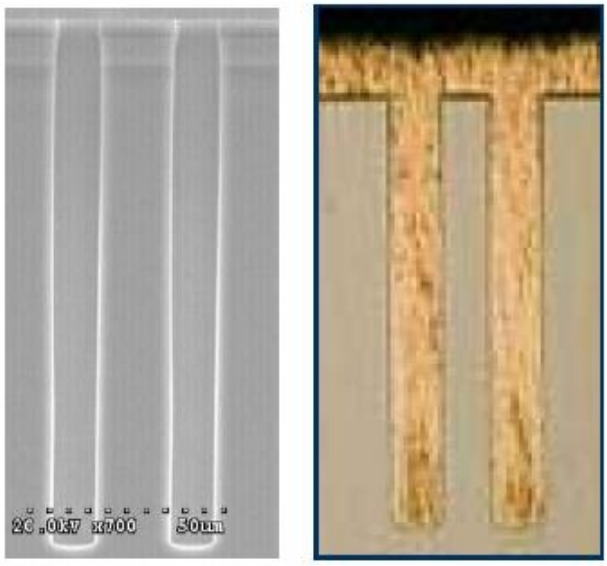



Fig. 4.5 Left: “Via Firs after BEOL” after D.R.I.E. Right: “Via Firs after BEOL” after refilling step

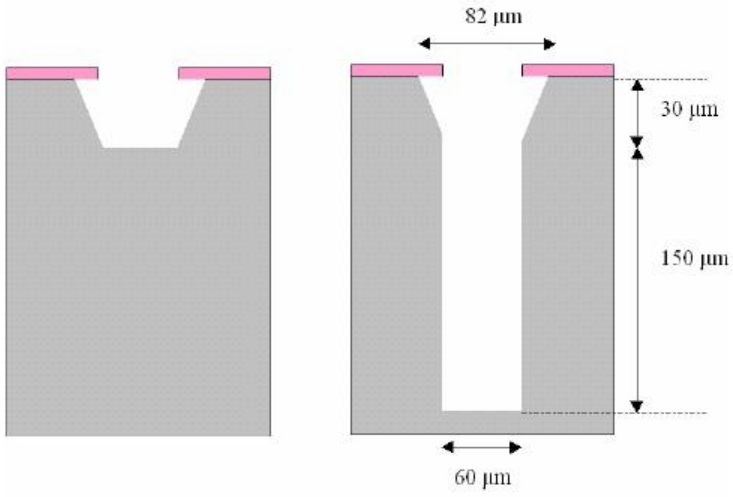



Fig. 4.6 Mouth profile: tapered etching step, followed by vertical etching step

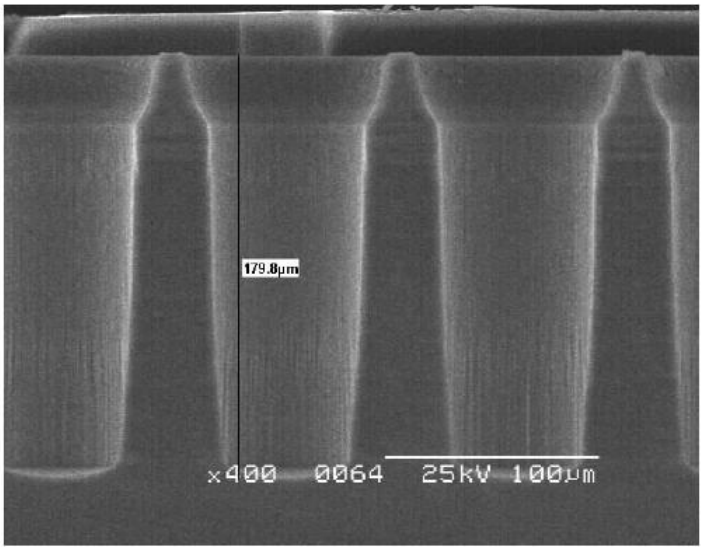



Fig. 4.7 SEM profile of Via Hole

## "Via Last after BEOL" TSV

In the "Via last" approach, the device wafer is grinded and thinned to its final thickness prior to the Via formation (Fig. 4.8). For that operation, the device wafer is temporary bonded onto a wafer carrier, usually made of glass. It will remain attached to this carrier until the stacking of the wafer onto another wafer, or until the transfer of the wafer onto a dicing frame, depending upon the attachment technique: wafer to wafer or die to wafer. Because the wafer carrier is little bit larger than the silicon wafer for a robust wafer, the overall handling system has to accept wafer substrates with diameter 2 mm larger than the standard wafers. Another major difference is on the capability of the wafer chuck to be able to electro statically clamp Silicon on Glass (SOG) substrates.


Fig. 4.8 “Via Last after BEOL”. From left to right: D.R.I.E., Oxidation, Copper refilling. A temporary wafer carrier is used from the via formation until the complete refilling of the via

From a process perspective, the fact that the wafer is temporary bonded with adhesive onto a wafer carrier, implies that all the further process steps have to be limited at temperatures between 180 °C to 200 °C. The DRIE “Via last” etching is very similar to the “Via First”, after BEOL approach in term of via dimensions and profiles.

# APPLICATIONS FOR THROUGH SILICON VIA (TSV) TECHNOLOGY

Devices realized with through via technology have many advantages over their single-layer counterparts. The short vertical interconnects allow them to operate at higher speeds with a lower power budget. In addition, TSV technology allows disparate elements to be processed on separate wafers for simpler production and greater optimization.

One of the most unmistakable examples of via process implementation is wafer level packaging of MEMS devices. Metal routing and contact pad area in many cases consumes a large portion of many MEMS die layouts. Incorporating via technology for interconnect and wafer level packaging significantly reduces the form factor while getting simplifying a large share of the traditional “back-end” process. As previously mentioned the “Via First prior to FEOL” approach is the most suitable for those applications since the Via already exist within the wafer before the steps required to generate the MEMS devices.

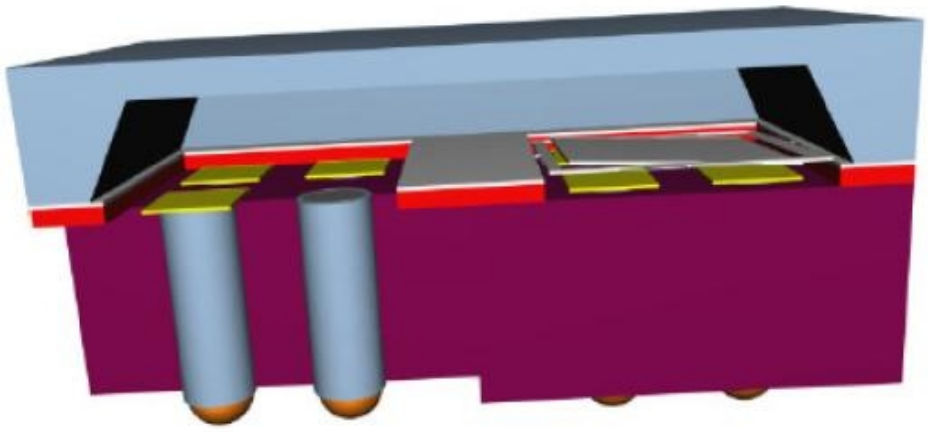


Fig. 4.9 Example of MEMS device with 3D interconnect

An interesting and quite new area is logic-on-memory; that is creating a high bandwidth memory interface to logic. The demand for memory bandwidth is growing rapidly; this is due to the increased use of multi-core processors. Multi-core processors can continue to provide historical performance growth on most modern consumer and business applications. However, power efficiency of these cores must also be improved to stay within reasonable power budgets. To maintain overall performance growth with stunted per-core and per-thread performance, growth rates will require an even more rapid increase in the number of cores per die. Growing performance by increasing the

number of cores on a die at this rate, however, puts unprecedented requirements on the corresponding growth of off-chip bandwidth. 1-TFlops multiprocessor die with 1 TB/s of off-chip bandwidth and a 32-Core CPU is target for 2010 [3]. Right now, Intel ® Tera-scale Computing Research Program is working on interconnect, memory and software technologies, where "Tera-scale" refers to terabytes of data to be handled by platforms with teraflops of computing performance [4]. With a Tera-scale computer one could create studio quality, photo-realistic 3D graphics in real time or also games will become more realistic. Particularly graphic applications tend to be more parallel in nature and scale well across a large number of core or execution threads, so that if the number of core is doubled, the execution time or compute throughput is roughly halved. But the computing power will be useless if it is also not guarantee a tera-bytes/s of bandwidth to on die-cache. A workable way to satisfy such requirements is by using the "via after BEOL" approaches since the copper interconnections present good electrical properties.

TSV technology also enables great improvements in memory technology and allows seamless integration of differing substrates. There are system advantages to thinning and staking multiple memory die such that the aggregate memory has the same and form factor as one memory package. This application requires TSV approaches where micron scale connections are achieved.

## TEZZARON® TECNOLOGY AND MEMORIES INTEGRATION

We had the possibility to work with Tezzaron TSV technology [5]. The first key breakthrough in their 3D development was the Super-Via™, a vertical copper structure that adapts standard process flow wafers to Tezzaron's 3D stacking process. The process belongs to the family of "Via after" approaches, in fact Super-Vias can be post processed into any wafer merely by adding copper metallization and additional

dielectrics, so they do not require a full manufacturing line or direct involvement of an outside foundry. Further, the Super-Via interconnect provides alignment marks, thinning control, interconnect, and bonding surfaces in a single structure. It further adds intrinsic cooling capabilities with its vertical copper structures. Tezzaron's interconnect methods enable high-density interconnect with extremely reduced pitches for true circuit-level 3D integration as it is shown in Table II.

**Table II**

| | Super-Via™ | 2nd Generation | Face-to-Face |
|---|---|---|---|
| Size | 4.0 µ x 4.0 µ | 1.2 µ x 1.2 µ | 1.7 µ x 1.7 µ |
| Minimum Pitch | 6.08 µ | <4 µ | 2.4 µ (1.46 µ) |
| Feed-Through Capacitance | 7 fF | 2-3 fF | << |
| Series Resistance | <0.25 Ω | <0.35 Ω | < |

Tezzaron® technology permits to create 3D memory devices with vast densities and superior speeds without any radical changes to design. Memory devices can contain many layers of memory cells stacked on top of one control and interface layer, greatly increasing memory density, reducing cost and power requirements. Staking memory dies to create super-memories is a field of great interest either for industrial and general purpose consumer either for advanced applications (i.e.: military, space environment). In fact the devices are much denser and their short vertical interconnects allow them to operate at higher speeds with a lower power consumption. The reliability of 3D integration using TSV technology is demonstrated and examples of 3D-SRAM already exist. Thus it was necessary to determine an application which should has involved memories, taken advantages by using the Tezzaron 3D integration technology and achieved original results from research standpoint. We focused our attention on the study of memory architectures for mitigation of Latch up phenomena. The idea is to realize and validate a new scheme to detect and remove the SEL for memories. This

research activity implies the necessity to analyze the latch up phenomena and develop a model to study and simulate the effects on CMOS technologies through circuital simulators. The equivalent circuit should take in account the SEL characteristics, the parameters specific of the technology (i.e. Chartered 130 nm) and the parameters depending by the given layout design. Using the developed model is possible to study and realize a logic that automatically interrupts the latch up event when it occurs. The logic is implemented within a standard SRAM, making it robust to SEL. After the validation of the proposed scheme for 2-D devices, which alone is a remarkable result, the modified memory is integrated in stacked layers in order to evaluate various issues concerning 3D systems. Specifically the prototypes aims to evaluate the effects of latch up on 3D integrated memories and on standard memories. In fact, stacked memories occupy reduced area compared with memories of the same capacity, thus there is a reduced probability of ion's strikes. On the other hand, it has not still investigated the possibility that a single particle passes through two or more layers damaging different areas of the same device. This study is important when implementing redundant architectures.

The analysis performed on Latch-up phenomena and its effects on devices, the project specifications, the proposed logic to detect and remove SEL, and the implementation on SRAM in 130nm node will be explained in the following chapters. Unfortunately, while the memory implementation is already done and the results will be presented, the integration of the modified memory within a 3D stack is still on going. Thus, in the presented dissertation it is not still possible give some results about the 3D integration with TSV technology.

# LATCH UP PHENOMENA

Latch-up has long been a bane to CMOS IC applications; its occurrence and theory have been the subjects of numerous studies and articles. Latch-up is a failure mechanism of CMOS integrated circuits characterized by excessive current drain coupled with functional failure, parametric failure and/or device destruction. It may be a temporary condition that terminates upon removal of the exciting stimulus, a catastrophic condition that requires the shut down of the system to clear or a fatal condition that requires replacement of damaged parts.

As microelectronic devices have been reduced in size and charge storage capacity, they have become vulnerable to radiation effects from single ions in the space-radiation environment [1]. Accurate estimates of the probability of single-event effects (SEE) in microelectronics used in spacecraft systems are important for system design. There are a multitude of effects due to individual interactions of single heavy ions and protons with microelectronic devices. Models have been developed by the radiation-effects community to describe the physical mechanisms and estimate the probability of the effects occurring in a given space environment. SEE occurs via stochastic processes

driven by random incidence of ions of various species, energies, and angles of incidence in the space environment. All SEE are initiated by localized disturbances in electronic charge, with consequences that depend on charge generation in sensitive volumes (SV) within microelectronic devices and collection at sensitive circuit nodes. Single-event upset (SEU) is the most common and most studied effect. SEU occurs in both memory circuits and logic circuits. The impact of SEU in logic circuits is highly system-dependent and requires considerations for timing and interconnections. Processor circuits can experience a functional interrupt from a single event. For high-density circuits, charge sharing from a single ion can lead to multiple-bit upset (MBU). The mechanism for SEU is charge generation along the path of a heavy ion and collection at a circuit node that can lead to regenerative action in a circuit latch and change of logic state. Permanent damage can be caused by single ions in some types of devices, resulting in single-event latch-up (SEL). Single-event burnout (SEB) is a concern primarily for power MOSFET-s and bipolar power transistors. Under conditions of high electric field, passage of a heavy ion through an insulator results in catastrophic damage due to rupturing of the gate dielectric, resulting in single-event gate rupture (SEGR). This occurs primarily during write or erases operations in non-volatile static random access memories (SRAM-s) or electrically erasable programmable read-only memories (E2PROM-s) and in power MOSFET-s. SEU and SEL have been observed in space systems and all the effects have been observed and studied in ground simulations using accelerators. Physics models of the interaction mechanisms have been developed, with differing levels of maturity, and used to develop engineering models for prediction of rates in space.

# DESCRIPTION OF THE HEAVY ION EFFECT

The starting point of all single-event effects is the filament of electron-hole pairs (EHP) generated by the energetic ion along its track. A detailed description of the physical mechanism leading to this charge deposition is given in [2]. The ion progressively transfers its energy to the target material until it stops when its energy is low enough to permit direct interaction with the atom's core. After a time in the one pico-second range, the description of the interaction of EHP is considered valuable. An average energy of 3.6eV is needed to create one EHP in silicon. The energy loss of a heavy ion is characterized by a Linear Energy Transfer (LET) function, usually expressed in MeV cm2/mg, or after multiplying by the target material's density (2.3 g/cm3 for silicon) in MeV/cm, or in pC/pm if the electron charge and the average energy is taken into account (1pC/pm = 100 MeV-cm2/mg for Silicon). The penetration of the ion is characterized by the range, expressed in um. LET and range can be obtained from special computer codes such as TRIM or from Ziegler tables [3] that allow the calculation of the charge deposited along the ion's track for a given structure. Given the ion's LET, the number of EHP per um is known; but the initial EHP density of the filament depends on its diameter. Some papers deal with this concern and indicate that the diameter, decreasing along the track, always ranges under 1um [4]-[7]. Let us assume that the filament can be modelled immediately after the ion strike by a DIRAC function column [11]. The dominant process is then diffusion, and the EHP density must comply with the simplified conservation equation (1).

$$\frac{(\partial n_{ehp})}{(\partial t)} = D * \Delta n_{ehp} \qquad (1)$$

where D is the diffusion coefficient for the particles and n the electron-hole pairs

concentration.

The solution of this equation is (2)

$$n_{ehp} = \frac{LET}{(4\pi Dt)} * \exp(-r^2/4\mathrm{Dt}) = n_0 \exp(-r^2/r_0^2) \qquad (2)$$

With *t* in ps, *D* in cm2/s, *ro* in um, LET in EHP/pm and *no* in cm-3

$$r_2^0 = 4 * 10^{-4} Dt \quad \wedge \quad n_0 = 10^{12} \frac{LET}{(\pi r_2^0)} \qquad (3)$$

*D* is not really known because very large EHP densities are involved. A common value for electrons and holes must be taken because the neutrality of the column is globally self-maintained (if not, the resulting electric field would speed-up holes and slowdown electrons). Then *D* may be approximated by (4):

$$D = \frac{kT}{q} * \frac{(\mu_n + \mu_p)}{2} \approx 25 cm^2/s \qquad (4)$$

For a LET of 50 MeV.cm2/mg (3.2 x lo6 ehp/um or 0.51pC/um) one obtains

at t = 1ps : r0 ~ 0 .1um, *n0* = 10exp(20) cm-3

at t = 9ps : r0 = 0.3um, n0 = 10exp(19)cm-3.

Dipolar rearrangement occurs quasi-instantaneously (electron relaxation time scale) and changes the electric field pattern in order to comply with the high conductivity column. To summarize, the filament immediately acts as a conducting path, possibly as a wire. Afterwards, the filament evolution depends not only on radial diffusion, but also on local conditions. Obviously, the above equations are no longer valid when the decrease in EHP density is governed by the electric field or by recombination. The perturbations caused by the heavy ion in electronic devices are, thus, the various consequences of the creation of the conducting path. In simple cases, the perturbation is only a prompt current flow along the path. In many cases, the completely modified potential distribution along the track induces more important changes in the device's operation. In some cases, an indirect disturbance may occur when part of the deposited charge finally reaches distant sensitive regions. Alone or together, these effects must be considered to explain the specific effects related to each device type.

# SINGLE-EVENT LATCHUP IN CMOS TECHNOLOGY

The latch-up phenomenon appeared with the development of CMOS technology in 1962. It stems from parasitic bipolar transistors that are structurally inherent to this technology. These transistors correspond to a parasitic PNPN structure. In normal operating mode, this structure is in a high-impedance state and has no effect. However, it can be switched to an ON state in various ways. If the power supply can provide the associated high current flow**,** the circuit can be thermally destroyed. Since the 1990's**,** more than half of the integrated circuit market is served by CMOS technologies. Low power consumption and high noise immunity contribute to the large development of these technologies. During the same years that CMOS was assuming an important role in VLSI technology, several techniques for controlling latch-up were being investigated.

## CHAPTER 5: LATCH-UP PHENOMENA

Latch-up can be triggered by both electrical and radiation excitations, including heavy ion effects. Although today's CMOS technologies provide some degree of latch-up protection, present and future development in sub-micron technology will find the problem more difficult because size reduction, unfortunately, induces higher susceptibility. For the heavy ion concern, one may notice that the, dimension of the ion-perturbed area stays unchanged while the dimensions of the device are made smaller and smaller. The perturbation area can then overlap an increased part of elementary transistors.

It is advantageous to briefly review the basic theory of SCR structures inherent in CMOS input and output circuitry. Both p- and n-type background materials are necessary to obtain N- and P-channel MOSFET-s in CMOS technology. This is commonly accomplished by creating n-type wells in a p-type wafer or vice versa. Many modern CMOS technologies use an epitaxial layer to reduce the substrate resistivity. Let us assume the more common arrangement, i.e.: a p-type substrate with an n-type well. The parasitic PNPN structure is a result of the transistor source and drain diffusions. Fig. 5.1 shows the cross section of a CMOS inverter and schematically presents the parasitic bipolar transistors and some associated resistors. Tentatively, discrete resistors are drawn to represent the current paths between the terminals and/or the collector and base regions of the parasitic transistors. Obviously, bases and collectors are also diffused; and their locations in the figure are based on junction proximity and electrical working considerations. The N+ sources and drains of the N-channel MOS devices act as the emitters (and sometimes collectors) of lateral NPN devices; the P-substrate is the base of this device and collector of a vertical PNP device. The N-well acts as the collector of the NPN and the base of the PNP. Finally, the P+ sources and drains of the P-channel MOS devices serve as the emitter of the PNP. For simplicity, it is assumed that the vertical PNP (VPNP) as well as the lateral NPN (LNPN) transistors have a common collector, as shown in Fig. 5.1. The N-well is normally connected to VDD (it is polarized to the most positive voltage) via a N+ diffusion tap while the substrate is terminated at GND through a P+ diffusion. These power supply connections involve bulk or spreading resistance to all points of the substrate and N-well.

## CHAPTER 5: LATCH-UP PHENOMENA

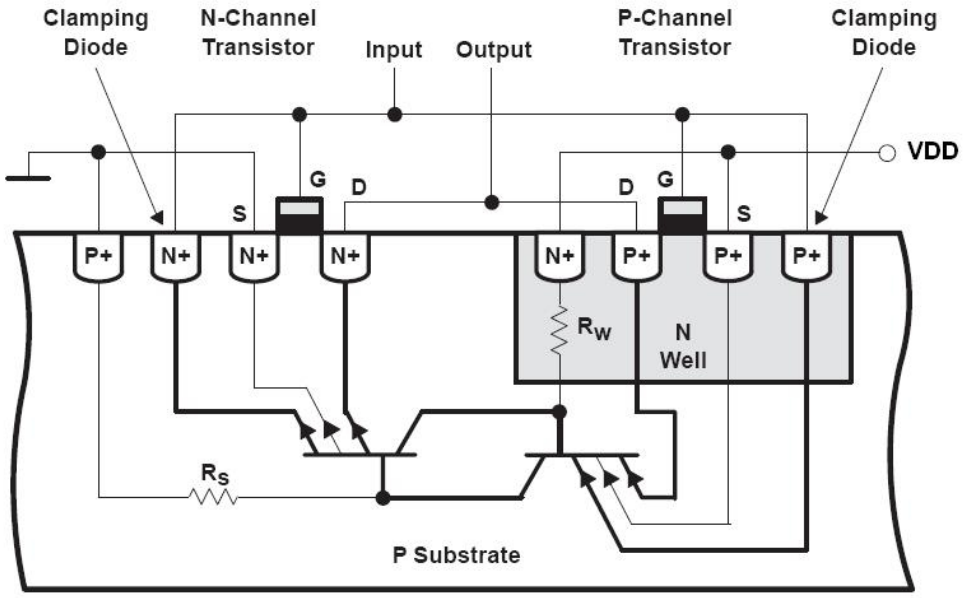



Fig. 5.1 Schematic CMOS inverter cross-section showing the parasitic bipolar transistor and the discrete resistors

In order to understand what latch-up is, one must notice that part of the VPNP collector current can reach the LNPN base and, similarly, that the LNPN collector current can activate the LNPN base. An unstable loop is, thus, inherent to the CMOS technology. Normally, only a small leakage current flows between the P-substrate and N-well causing only a minute bias to be built up across the bulk due to the resistivity of the material. In this case the depletion layer formed around the reverse-biased PN junction between P-well and the substrate supports the majority of the VDD-GND voltage drop. As long as the MOS source and drain junctions remain reverse-biased, CMOS is well behaved. So that, under normal operating conditions, the parasitic bipolar transistors are in high impedance states because their respective emitter and base terminals are shorted. The feedback loop is masked. In the presence of intense ionizing radiation, thermal or over-voltage stress, however, current can be injected into the PNP emitter-base junction, forward-biasing it and causing current to flow through the

substrate and into the N-well. Thus due to an external excitation (electrical or radiation), one parasitic bipolar transistor may be forced into conduction, activating this unstable loop condition. A self-maintained low-impedance path is opened between the supply terminals, VDD and ground, causing a circuit malfunction that may be followed by thermal failure if adequate protections have not been taken. Latch-ups involving Vout terminals are dependent upon the output logic state. In modern CMOS designs, many combinations of parasitic vertical and lateral bipolar transistors can contribute to the latch-up phenomenon.

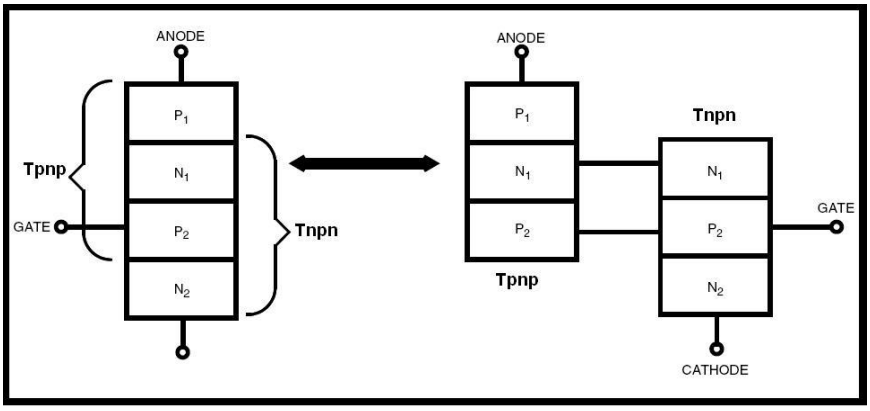



Fig. 5.2 Four Layer SCR structure (PNPN)

The circuit complexity tends to obscure the main physical mechanisms of latch-up. Basically, the latch-up phenomenon only requires the PNPN loop existence and can occur with only two parasitic transistors. This is why studies often use an elementary four-terminal PNPN structure. This simplest necessary structure is shown in Fig. 5.2 associated with the equivalent circuit of Fig. 5.3. Assuming that the supply is still applied at the VDD and GND plugs, this PNPN structure can be characterized by the current versus voltage (I-V) curve obtained between the remaining p+ and n+ plugs. A typical (I-V) characteristic is shown in Fig. 5.4 where the n+ plug is grounded.

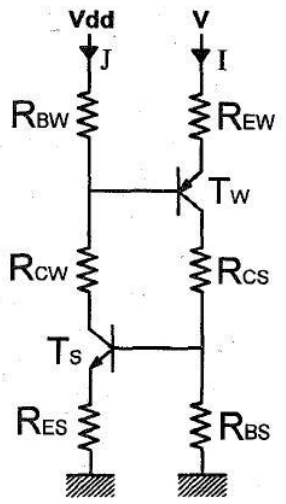



Fig. 5.3 (Equivalent circuit)

The two possible states for the normal VDD bias are represented, and two key critical points are identified. They are the triggering point (VTRIG, ITRIG) and the holding point (VHOLD, IHOLD) which delimit the unstable negative resistance region. In electrical triggering mode, the high conducting regime is initiated if the voltage on VDD is increased above VTRIG and needs a bias decrease under VHOLD to be stopped. Although the ion's susceptibility cannot be directly derived from the static (I-V) characteristic, its triggering and holding points are good indicators of the sensitivity of the device [8]. They are dependent upon the values of the various resistive paths inside the structure. In order to get tractable relations between the parameters determining the latch-up behaviour, the equivalent circuit of Fig. 5.3 is analyzed with simplified assumptions for the bipolar junction transistors. In linear mode, the current gain is sufficient to neglect base current and emitter-base voltage variation, i.e. the emitter-base potential drop is constant and equal to VBETH (~ 0.8V). In saturation mode, the collector and emitter voltage are equal, and VBE is still equal to VBETH. Let us suppose that the TW and TS transistors of Fig. 5.3 (respectively the VPNP and LNPN transistors) are operating in linear mode.

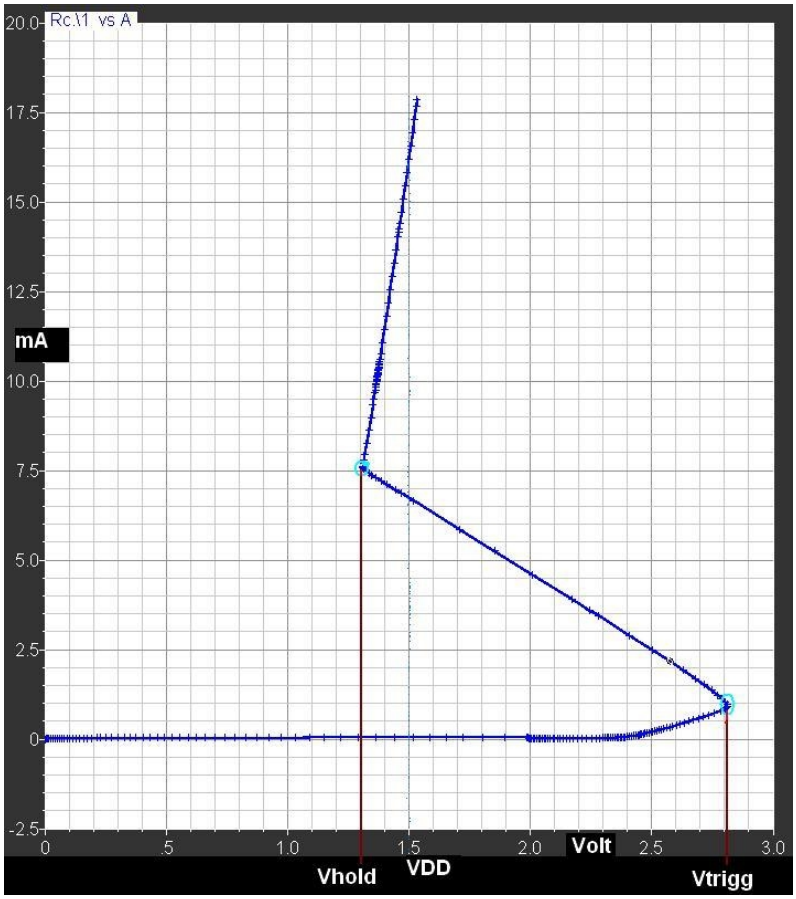



Fig. 5.4 Typical (I-V) characteristic of the PNPN structure showing the two possible working points

With the above assumption, one obtains:

$$V - R_{EW} I - V_{BE,th} = V_{DD} - R_{BW} J \qquad (5)$$

$$R_{ES} J = R_{BS} I - V_{BE,th} \qquad (6)$$

Solving (5) and (6) with respect to I leads to

$$V = V_{DD} + V_{BE,th}\left(1 + \frac{R_{BW}}{R_{ES}}\right) - R_{EW} I \left(R_{BW}\frac{R_{BS}}{R_{EW}} R_{ES} - 1\right) \qquad (7)$$

Equation (7) corresponds to the negative slope portion of Fig. 5.4. It clearly shows that instability (i.e., negative resistance) is only possible if the open-loop gain [here (Rbw * Rbs) / (Rew * Res)] is higher than unity. It must be noticed that Rew and Res are normally lower than the other resistances. In the same way, triggering and holding points can be calculated. Triggering occurs when Ts is just turned on, i.e. when I = (Vbe,th / Rbs) = Itrig.

Then (8)

$$V = V_{DD} + V_{BE,th}\left(1 + \frac{R_{EW}}{R_{BS}}\right) = V_{trig} \qquad (8)$$

The holding point corresponds to the end of the linear region, that is, at the saturation point of $T_S$ or $T_W$.

It can, thus, be deduced from the conditions (9), (10)

$$J = \frac{V_{DD}}{(R_{ES} + R_{CW} + R_{BW})} = J_{sat} \qquad (9)$$

if $T_S$ saturates prior to $T_W$

$$I = \frac{V_{hold}}{(R_{EW} + R_{CW} + R_{BS})} = I_{hold} \qquad (10)$$

if $T_W$ saturates prior to $T_S$.

.

It can be derived that $T_S$ saturates prior to $T_W$ if (11)

$$\frac{(\frac{R_{BS}}{R_{CS}} - \frac{R_{ES}}{R_{CW}})}{(1 + \frac{(R_{ES} + R_{BW})}{R_{CW}})} > \frac{V_{BE,th}}{V_{DD}} \qquad (11)$$

In this case

$$V_{hold} = V_{trig} - J_{sat} R_{BW} (1 - \frac{(R_{EW} R_{ES})}{(R_{BW} R_{BS})}) \approx V_{trig} - \frac{R_{BW}}{(R_{BW} + R_{CW})} V_{DD} \qquad (12)$$

If $T_W$ saturates prior to $T_S$

$$I_{hold} = \frac{(R_{ES} V_{DD} + V_{BE,th} (R_{ES} + R_{BW}))}{(R_{CS} R_{ES} + R_{BS} R_{BW})} \qquad (13)$$

that is

$$I_{hold} \approx \frac{(R_{ES} V_{DD} + V_{BE,th} R_{BW})}{(R_{CS} R_{ES} + R_{BW} R_{BS})} \qquad (14)$$

and

$$V_{hold} \approx \frac{(R_{BS} + R_{CS})}{(R_{CS}R_{ES} + R_{BS}R_{BW})}(R_{ES}V_{DD} + V_{BE,th}R_{BW}) \qquad (15)$$

Finally, when both $T_S$ and $T_W$ are saturated, the (I-V) characteristic is a straight line with a slope 1/R such as (16).

$$R = (R_{ES}//R_{BS} + R_{CS}//R_{CW})//R_{BW} + R_{EW} \quad ,$$

$$\frac{1}{R} = \frac{1}{R_{BW}} + \frac{1}{R_{CS}} + \frac{1}{R_{CW}} \qquad (16)$$

Although the above calculations have greatly simplified the behaviour of real transistors, they give realistic values as long as the transistor gains are actually higher than unity. Reliable tendencies for the trigger and holding points can then be derived from the simplified model. Obviously, a more sophisticated calculation is necessary when accurate results are needed, especially in the case of very low transistor gains. Normally, VTRIG is approximately equal to VDD + Vbe,th (except if Rew is particularly large). Condition (11) is usually satisfied unless Rcs and/or Rbw are particularly large.

Then:

$$V_{hold} = V_{BE,th} + \frac{R_{CW}}{(R_{BW} + R_{CW})}V_{DD} \qquad (17)$$

It should be noted that latch up holding current has traditionally been used as an indicator of the susceptibility of a device structure to latch up. The choice of holding current as an indicator is reasonable, since latch up is a bipolar phenomenon and bipolar devices are current driven. Also holding voltage is important as it is well understood that at a voltage smaller than the holding voltage exists no stable, high-conductivity state besides the off-state. It is worth to notice that if it is implemented a design where latch up holding voltage raises higher than the supply voltage, then it can be considered latch up free. In the next chapter it will be described the SEL mitigation architecture for memories implemented at CMP laboratory. The proposed scheme basically consist in latch-up detection stage and latch-up removal stage. The system aims to detect the voltage drop on VDD lines induced by the SEL; the value is strictly related to the value of hold tension, thus it is essential determine such operating point for the given technology. Moreover, the correction stage indirectly acts on the intrinsic parameters of the device (i.e. Tap resistances) and it results in modifying the latch up susceptibility. Specifically it changes the parameter REW expressed in formula (8), thus in formula (12), reducing the probability of self-biasing the parasitic Thyristor.

# TRIGGERING THE PARASSITIC SCR STRUCTURE OF CMOS DEVICES

The parasitic thyristor can be triggered in various ways:

- If there is a voltage at the input or output of a circuit that is more positive than the supply voltage, or more negative than the ground connection (or, to be precise, more negative than the connection to the substrate), current flows into the gate of the thyristor.

- An electrostatic discharge can trigger the parasitic thyristor. Even if the electrostatic discharges have a duration of only a few tens of nanoseconds, when this happens, the complete chip may be flooded with charge carriers, which then flow away slowly, resulting in the triggering of the thyristor. The parasitic thyristor can be triggered by a rapid rise of the supply voltage. This effect often was observed in earlier generations of CMOS circuits.
- Additionally, the thyristor might be triggered by a high supply voltage; far higher than the value given in data sheets. In this case, the supply voltage must be increased up to the breakdown voltage of the transistors. When in breakdown, the current in the parasitic transistors, which should be blocking, increases in an avalanche process, so that activation of the thyristor must be anticipated.
- In space and avionic environments or in any situation involving components that operate close to a source of high-energy radiation latch-up is normal initiated by ionizing radiation.

The effects of temperature must also be noted at this point. Increasing temperature will cause an increase in both the leakage current through the SCR and in the current gains Beta1 Beta2 of the two bipolar transistors. As such, the magnitude of the driving force required to turn the SCR on will decrease with increasing temperature. In other words, the SCR will be more easily triggered as temperature increases for any of the triggering mechanisms described.

Regardless of the triggering source, the latch up event can be described (thus simulated) with a voltage pulse applied at the VDD (or GND) tap of the device or equivalently with a current pulse applied between the base terminals of the parasitic bipolar transistors. Considerations about the pulse amplitude and duration, and analysis of the back draws related to the different approaches will be explained in the next section.

## CHAPTER 5: LATCH-UP PHENOMENA

After the triggering of the parasitic thyristor, various reactions can be observed:

- The parasitic thyristor triggers very rapidly and enters a very low-resistance state. The source of the supply voltage is short circuited as a result of the circuit that has been affected. A very high current flow, which, in a very short time, leads to destruction of the component. The thyristor can be switched off again only by switching off the supply voltage.
- The thyristor triggers in the manner previously described, but, in this case, the thyristor has a comparatively high forward resistance. The result is that only the supply current increases, but this increase usually is quite large. Because of the high power dissipation in the circuit, the component can be damaged. The thyristor usually switches off only after the supply voltage has been switched off.
- In some cases, the thyristor has a very high resistance. The high forward resistance limits the current to values below the holding current of this thyristor. In this case, the supply current increases. The supply current sinks to normal values when the trigger current at the gate of the thyristor (as a result of an over-voltage at the input or output of the integrated circuit) is switched off.

The behaviours mentioned above are depicted in Fig. 5.5. It has been performed a parametric simulation using different values for P-substrate resistance (i.e. R P-substrate equals to 150 Ohm, 850 Ohm and 1.55K Ohm respectively). The parameters used in SPICE simulations (i.e. resistance values, gains of parasitic bipolar transistors) depend on layout design and on specific technology process. For this qualitative analysis high, medium and small values of resistance have been used (as it will be explained the values extracted in our case are R P-substrate = Rew = 600 ohm and R N-well = Rbw = 1200 ohm).

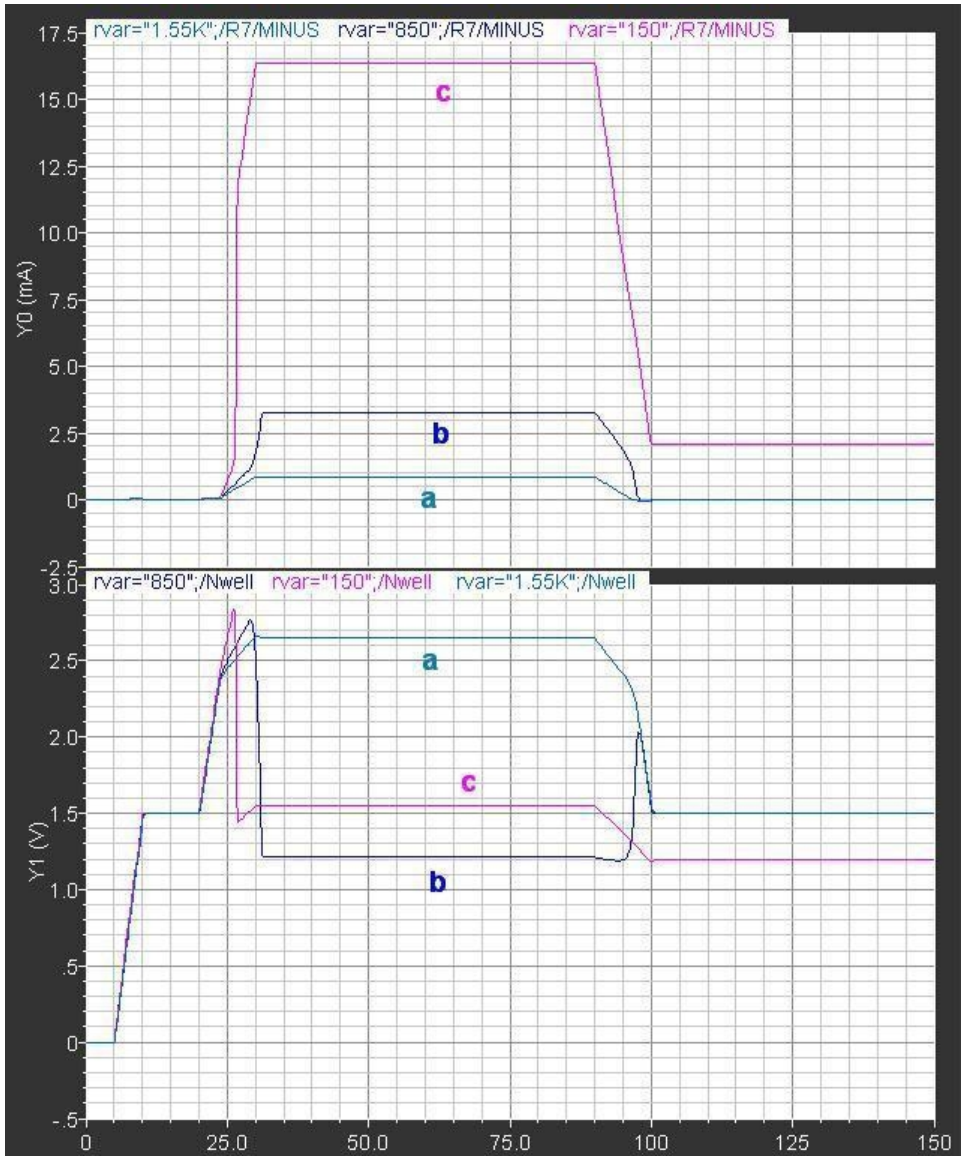



Fig. 5.5 a. Latch-up is not triggered, b. Triggered but not held, c. Self-biased condition is achieved

In Fig. 5.5-bottom are depicted the voltage raise on the VDD line (i.e. the nominal value is 1.5V), while Fig. 5.5-top shows the related currents induced by latch up event.

It results in three cases:

a) Very high resistance. The latch-up event is not triggered, there is not any induced current flow through the substrate. When the tension decreases to the nominal value (VDD) the nominal operating condition are restored
b) High resistance. The parasitic thyristor is triggered but it does not reach the self biased condition (the available current is limited by the resistance). The transient current flow can affect the data stored in the bit cell (or it can have other effects on the logic values), but the event disappears without irreversibly damage the device.
c) Low resistance. The inherent SCR structure switches to on state and the self-biased condition is established. The value of current flow through the substrate depends only by layout parameters. The device is not immediately damaged by the energy injected from the ion's strike, but in the lapse of time of tents or hundreds nano seconds (depending by the structure) it will be destroyed due to Joule effect.

# EXPERIMENTAL FEATURES

There are two latch-up tests to be performed according to the SRC specifications:

1) Over-voltage at power supply plug
2) Current injection at base-emitter junction

The over-voltage test is meant to mimic power glitches and/or high voltage tests like burn-in. In this test, the forced voltage is generally x1.5 the maximum operating level of the power supply plug. The corresponding clamp current depends on the

nominal power supply current drawn during latch-up setup ($I_{NOM}$) while internal qualification procedures state the clamp level at which the current should be clamped above $I_{NOM}$. This stress mode is emulated in the standard SCR structure by sweeping the voltage of the anode (P+ PMOS tap) and the supply shorted together, while grounding the cathode (N+ NMOS tap) and the substrate tap. In this stress mode, the parasitic SCR is triggered by the avalanche of the N-well to substrate junction. A small series resistor is added to P+ plug; the consequent voltage drop is sufficient to permit to observe the negative resistance region without direct switching from triggering condition to holding condition. The resistor fits with the estimated resistance value for VDD lines and related contacts. The circuit used for SPICE simulations is sketched in Fig. 5.6.

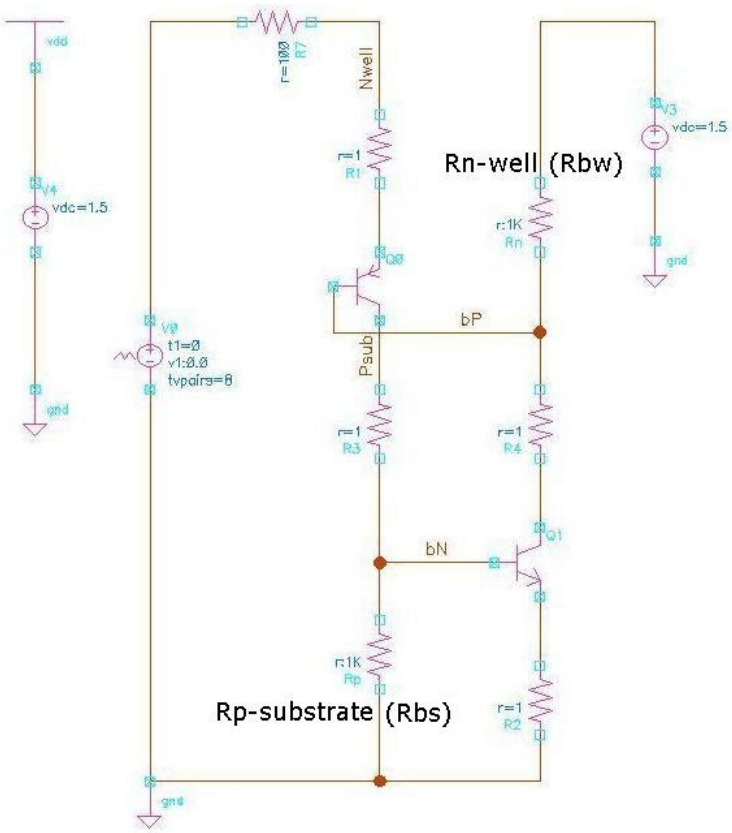



Fig. 5.6 SPICE circuit to study latch-up events induced with over-voltage conditions

The triggering voltage of the parasitic SCR in this stress mode is above 2.8V and largely exceeds the I/O over-voltage for the given technology; sets at 2.25V. It means that it is useful to extract fundamental parameters as triggering and holding points, but on the other hand it is not suitable to be implemented to study the effects of latch up occurrence during the normal function of a device as a memory. Since logic values, power consumption, delays in signal propagation, write and read operations will be affected by an non-realistic operating condition.

Current injection-test is meant to mimic over/undershoots at the signal pads due to a variety of causes (reflections, noise, external injection…). It does not permits to work out the triggering current and voltage for the given technology, since the circuit passes suddenly from the off condition to the self-biased operating point. However, it allows to operate with the nominal supply voltage value so that it makes possible to study the behaviour of a device when latch-up occurs. A current pulse is added between the two base terminals of the parasitic bipolar transistors pair. It should be noted that in literature there are no detailed models or exhaustive tables which permit to correlate the energy released from the ion to the device with the current pulse generated by the induced electron-hole regeneration process. The model should consider the particle energy, the impact area, the length of the filament, the characteristic parameters of the specific technology end layout design in order to correct shape the equivalent current pulse. However, the soft error events (i.e. SEL involving limited energy, thus events not destructive) are usually described with Gaussian functions with amplitude from fractions of mill amperes till one mill ampere. Some works also referred to use Dirac function column current pulse with height of tents mill amperes for latch-up triggering. The duration of the pulse is related to the initial EHP generation process and it can be estimated to be on the range of pico seconds. The probability of triggering the parasitic Thyristor depends by the duration and the amplitude of the pulse (i.e. the total energy amount). The minimum value of the forced current can be derived by the over-voltage test and it is around 3mA. Moreover, parametric SPICE simulations show that there are negligible differences in applying pulses from 5mA to 100mA or greater.

## CHAPTER 5: LATCH-UP PHENOMENA

The stress mode is emulated in the standard SCR structure by injecting current in base-emitter junction, with the N-well tied to VDD and the substrate tap tied to ground. The injected holes collected at the substrate tap contribute to increase the voltage drop across the Rbs (or Rbw)**,** which shunts the Base-Emitter junction of the lateral NPN (or of vertical PNP), eventually turning on this transistor. The equivalent circuit used for simulations is sketched in Fig. 5.7.

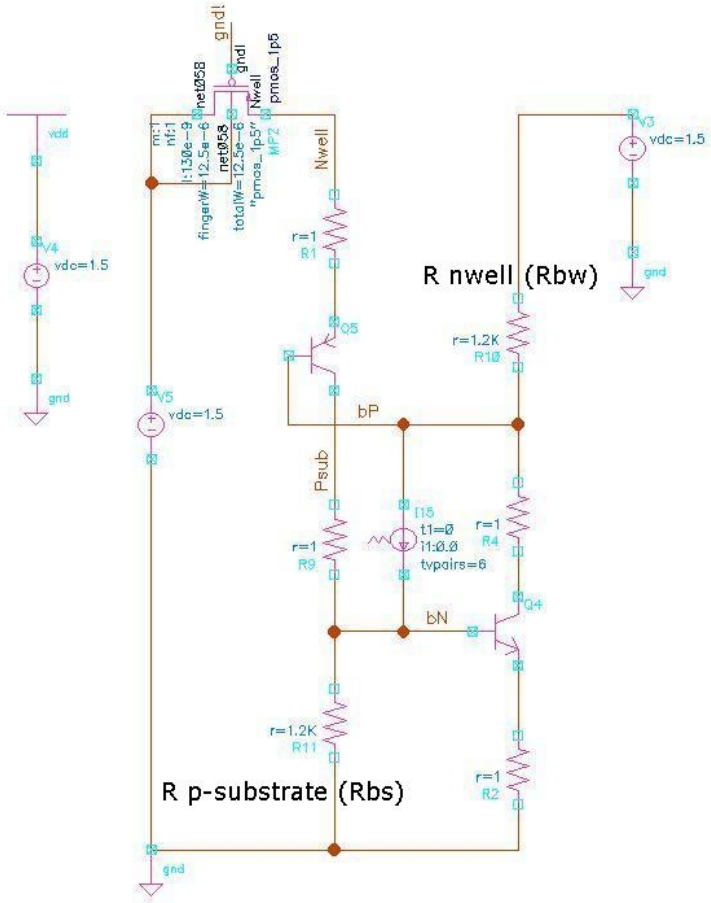



Fig. 5.7 Equivalent circuit of the PNPN structure for current induced Latch-up events

Fig. 5.8. depicts the confront between the two different methods.

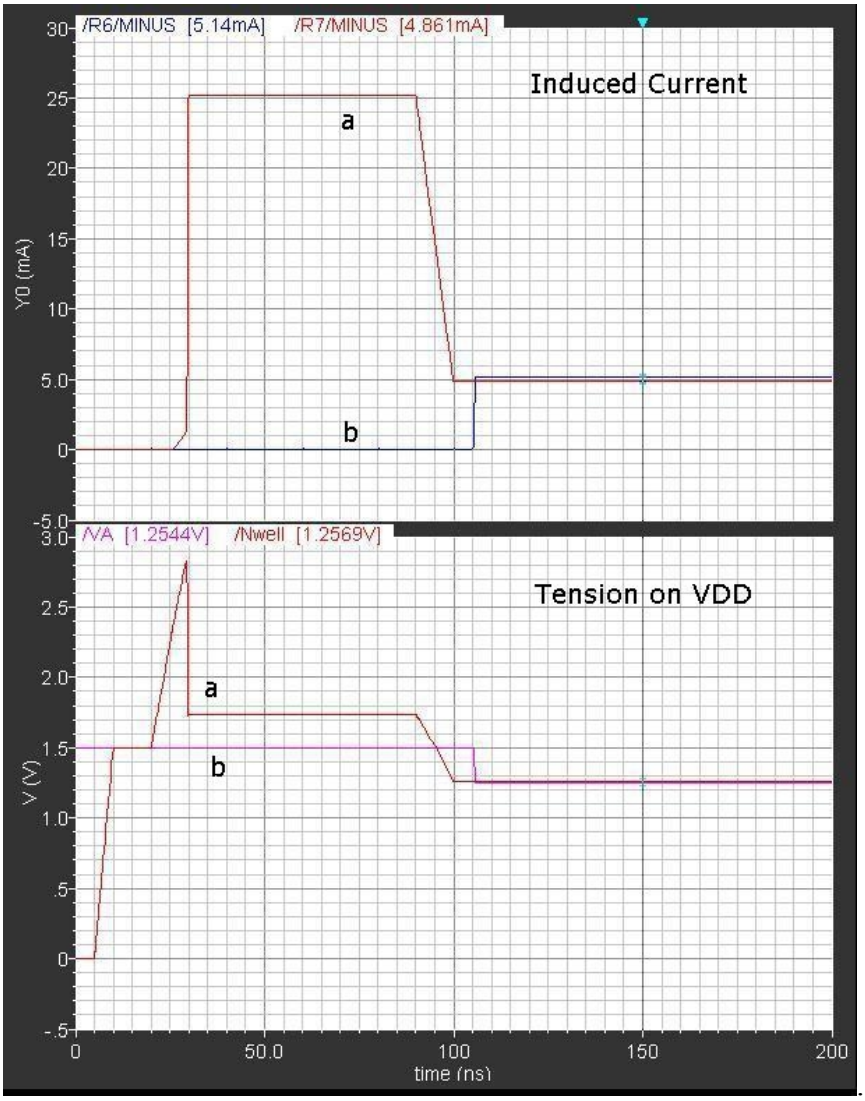



Fig. 5.8 Confront between the current injection mode (b) and the over-voltage latch up triggering mode (a). Top: current flowing through the substrate. Bottom: voltage on supply line

Curves (a) result from the over-voltage method (i.e. using the circuit sketched in Fig. 5.6). By drawing the induced current in function of the related voltage, the typical I-V characteristic (sketched in Fig. 5.4) is obtained, from where it is possible to extract both the latch-up triggering value and the holding value. While in the case of current injection approach there is the direct switching between the low- and high-current regions (i.e. voltage drop from 1.5V to 1.2V without any increase in curve(b)), thus the triggering condition is masked and only the hold value is obtained.
Since we aim to realize a logic to detect latch up events, we are interested to determine the hold tension value, while the triggering point is useless. In fact it gives only information about the susceptibility of the technology to latch up.

Although if, during the latch up triggering phase the there are different values of current and tension in the two cases, both the circuits present the same behaviour when the Thyristor self biased condition is achieved (i.e. latch up hold region). Fig. 5.8. shows that either in the over voltage triggering either in the current injection mode, the lath–up event induces voltage drop of few hundreds mill volts on supply line and the related current flows through the substrate are on the range of 4-5 mill Amperes. The logic that will be presented in the next chapter will response to the voltage decrease in order to detect the latch up event.

# MODEL PARAMETERS

To perform quantitative prediction of latch-up, various two-dimensional (2-D) or 3-D device simulators have been developed [9], [10].The numerical simulators used for latch-up modelling, e.g. PISCES or DAVINCI, require the knowledge of greater number of technological and physical parameters than analytical models. Furthermore, numerical models impose heavy demands on accuracy of parameters definition that is not always possible. However, the results for standard process derived by numerical

simulators can be used within an analytical model of latch-up that takes into account the dependence of parasitic bipolar transistors current gain on a current flowing across these transistors. The model is based on methods of non-linear dynamic systems theory. The expressions for threshold dose rate Pth (in the case of pulsed irradiation) and for threshold LET Tth (for the SEL) derived from the model only depend on physical parameters of CMOS structure and semiconductor substrate. These are the minority carrier lifetimes, thickness of parasitic transistors base, distributed resistance of well and substrate, and temperature.

However, for the stated purposes it is not relevant determine the effects of Ion's strike under disparate conditions as: vertical or inclined ion track, long or short track, strikes close or not to P+ plug, structures implementing epitaxial layer and guard ring. In fact, to implement a latch up mitigation scheme and to work out the related system specifications, it is a matter of interest define the relations or an equivalent model which describes the parameters used in numerical simulators with parameters and net-lists compatible with circuital simulators and point out the conditions that are critical. For example, the models furnished for the bipolar transistors used in standard processes widely differ from the models needed to describe the parasitic thyristor. In this case the equivalent net-list has been obtained thanks to the cooperation with IROC Technologies Corp. It has been made possible to compare the results obtained with parametric SPICE simulations to the results obtained with numerical simulators for similar test conditions. It permits to work out the parameters of the equivalent circuit used to emulate the parasitic bipolar pair. The parameters inherent the extracted BISM4 models includes the values of bipolar transistors gain, values of the resistances Rew, Rcs, Res, Rcw depicted in the schematic of Fig. 5.3. While the Rbw and Rbs resistances are considered separately since their contributions are very important to characterize the latch up event within a specific technology node and layout design. The most important values are the gains of the lateral NPN and vertical PNP bipolar transistors since they determine the Latch-up susceptibility. They respectively are Bnpn = 9.7 and Bpnp = 4.7. While, the latch-up hold condition rather depends on the values of diffused resistances. Since the BSIM4 model already presents values for base, emitter and collector resistances it is

possible to take Rew = Rcs = Res = Rcw = 1 ohm. Thus the attention is focused only on the shunting resistors Rbs and Rbw across the base-emitter junctions of the bipolar transistors: these resistors divert base drive from the bipolar transistors and as a result increase both the trigger current and holding current levels required for the structures to participate in latch-up. Rbs represents the substrate spreading resistance and it strongly influences the trigger current ITRIGG that is more or less equal to Vbe,th/Rbs. As consequence, it determines the trigger voltage value VTRIGG. On the other hand Rbw represents the N-well resistance and it takes part in value of VHOLD.

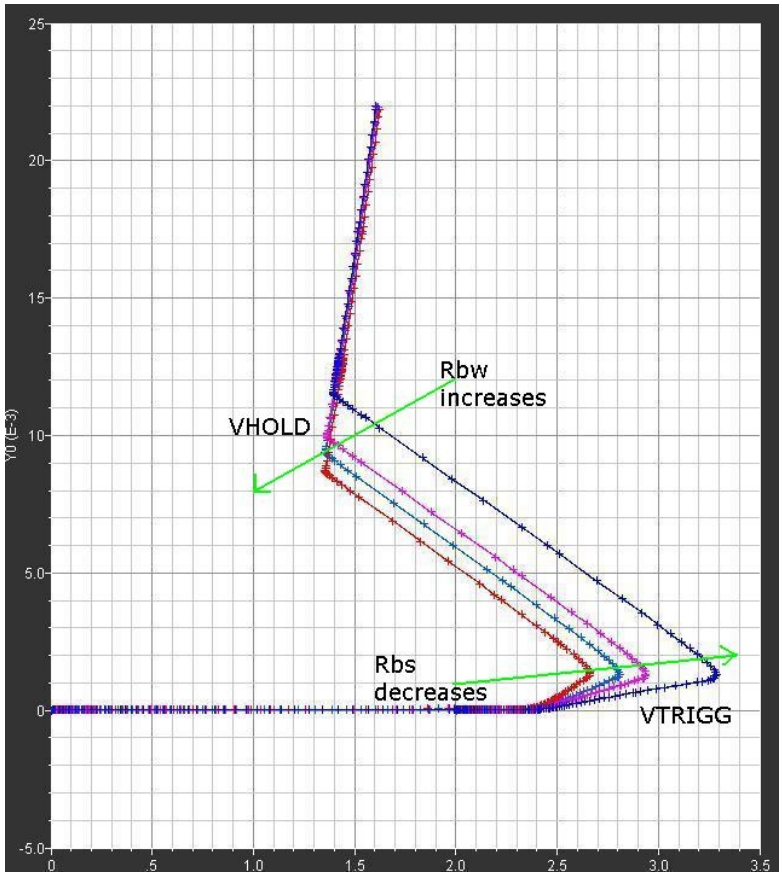



Fig. 5.9 Parametric I-V characteristic, for different values of Rbs and Rbw. When Rbw (R N-well) increases VHOLD decreases, while decrease of Rbs (R P-substrate) results in increasing VTRIG

An example of different I-V characteristics, obtained by varying Rbs and Rbw, is given in Fig. 5.9. The values evinced for Chartered 130 nm technology node and the layout of the specific memory are Rbs = 600 ohm (resistance associated to P-substrate), Rbw = 1.2K ohm (resistance associated to N-well).

# THE DYNAMIC FEATURES OF THE HEAVY ION-INDUCED LATCH-UP

The latch up triggering approaches previously mentioned does not give any information on time dependence of the latch-up process. Simulation analysis of carrier densities after an ion strike provides increased insight on this point. Assuming, for instance, that the ion's impact is located at the n-well of the structure shown in Fig. 5.2, between the n+ and p+ plugs, and that the deposited energy is large enough to induce the latch-up phenomenon. Fig. 5.10 [11] gives the current at the VDD plug.

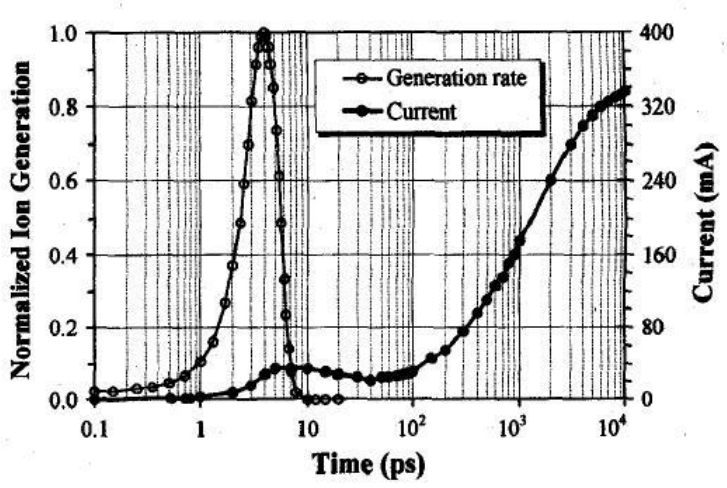



Fig. 5.10 Typical latch up current versus time evolution after the ion’s impact. The ion’s EHP generation is also shown

The ion's effect is simulated by a Gaussian EHP generation rate, as shown in the figure. From simulations, the dynamic conditions involved during latch-up can be described as follows. First, the potential is changed along the ion's track. VDD is no longer dropped at the n-well/p-substrate junction, but over a larger width. As a quasi-instantaneous result, the potential distribution into the n-well is no longer uniformly equal to VDD. The potential decrease toward the n-well bottom is such that the PNP base becomes forward-biased. Consequently, the vertical PNP transistor is activated; and holes are injected into the p-substrate. But, because the electron base current is only supplied by the excess electrons that have been generated by the ion, the PNP transistor tends to return to its off-state. This early step is visible in Fig. 5.10 and corresponds to duration in the 30-40ps range. Excess holes in the p-substrate are pulled out by the p+-substrate plug and diffuse in the substrate. Since the substrate resistance is high enough, the hole-current flow increases the base potential of the lateral NPN transistor. This transistor is, thus, increasingly activated by collection of holes arriving in the n+ terminal region. Electrons are injected into the n-well and can, thus, sustain the PNP transistor activation. The regenerative process is going on and is clearly visible in Fig. 5.10. The delay between the ion’s strike and the final current increase depends on the layout and size of the geometry. At a time in the few ten nanoseconds range, the steady state current is reached and injection by the two emitter types results everywhere in high carrier density. The current flows from VDD to ground resulting in heating the N-well and P-substrate junction. Although if it could not affect immediately the logic behaviour of the component, this situation is potentially destructive, resulting in damage to bond wires or metal supply tracks on the die due to localized overheating.

An inverter circuit with the associated SCR structure is sketched in Fig. 5.11.

Fig. 5.12 shows the effect of a SEL on the mentioned circuit during its normal operating conditions.

## CHAPTER 5: LATCH-UP PHENOMENA

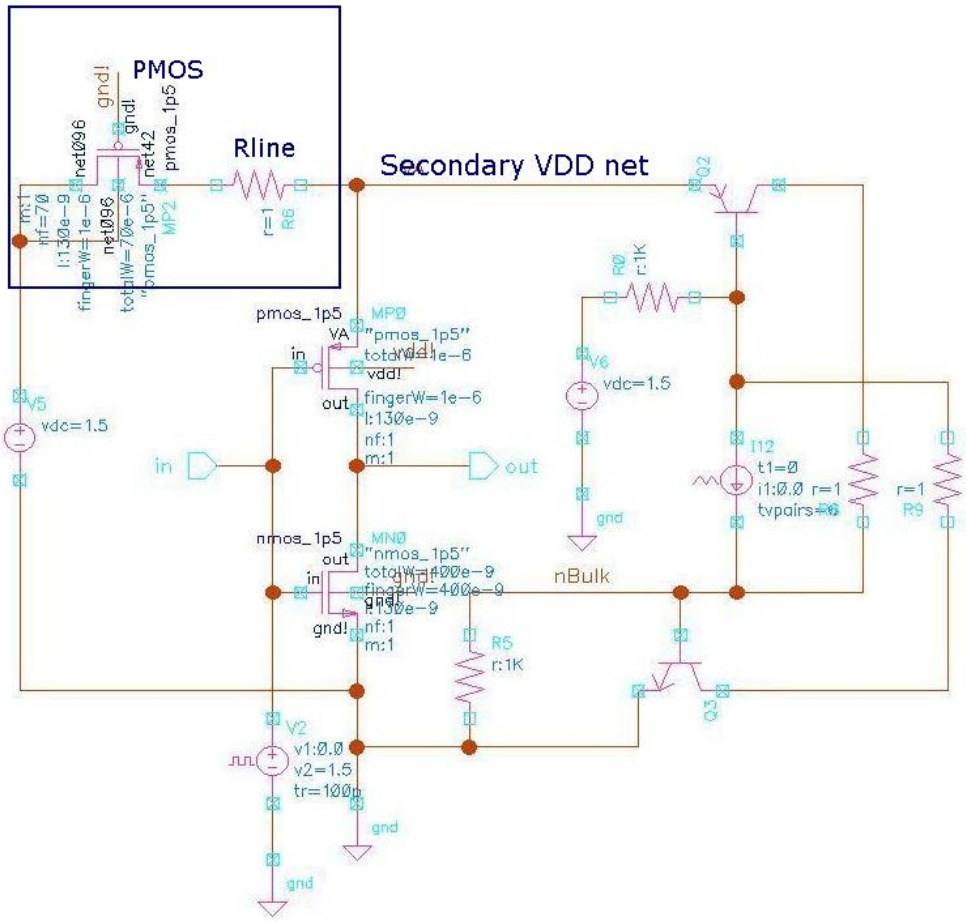



Fig. 5.11 CMOS inverter and the associated SCR structure to study the effects of latch-up event during the normal operating mode. The PMOS and Rline have been added to represent the secondary VDD line as it will be explained in chapter 6

Latch up event does not immediately damage the component. At first time, it slightly affects the device's behaviour (Fig. 5.12 bottom) since the voltage variation is on the order of few hundreds mill volts (Fig. 5.12 middle). However it generates a low impedance path between VDD and GND causing excessive current to flow.

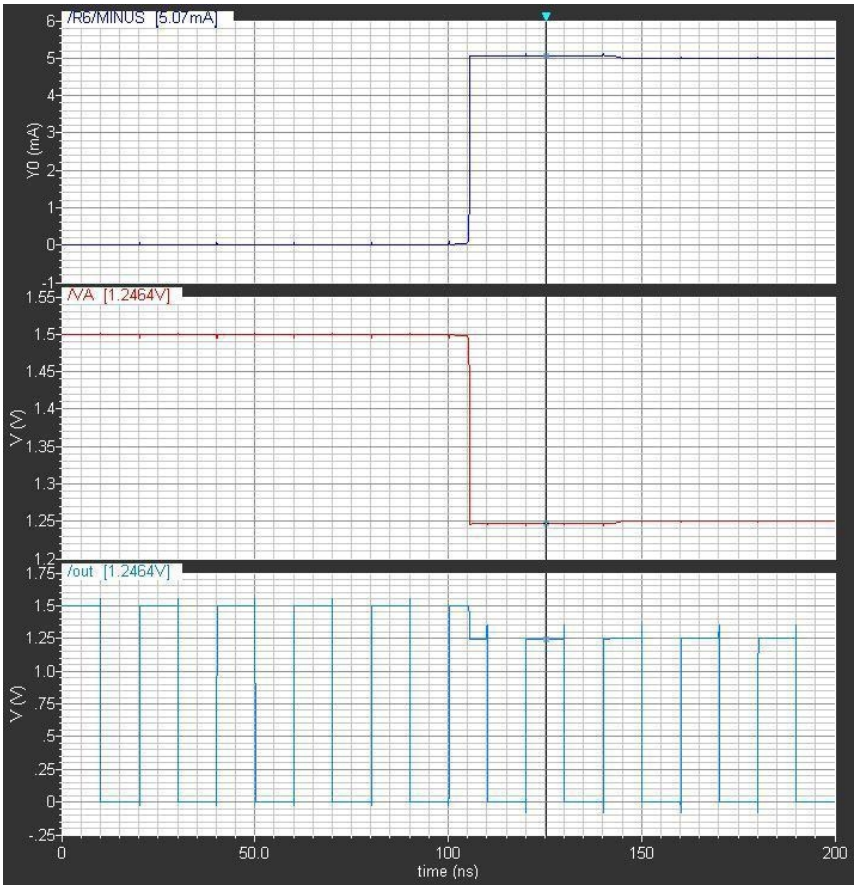



Fig. 5.12 Simulated waveform of SEL occurrence during normal operating condition in a CMOS inverter

Current is on the range of several mill ampere (Fig. 5.12 top) is drained by the parasitic bipolar transistors through the substrate. SPICE simulator cannot emulate the effect of thermal heat on the device junctions, but it is obvious that they will be irreparably damaged. Thus, every architecture which aims to detect and then avoid the latch-up phenomena should react in the time of tents nano seconds, otherwise the excessive thermal increase in the substrate will completely destroy the devices.

# CHAPTER 5: LATCH-UP PHENOMENA

# SINGLE-EVENT LATCH-UP (SEL) MITIGATION SCHEME

As a matter of fact, latch up mitigation may be required not only for space and other radiation hostile environments but also for commercial electronic products related to applications running at ground level. There are three existing latch up mitigation solutions:

I. External current sensors
II. Guard rings
III. Technology modified process

The first approach uses current sensors at the board level to detect the excessive current induced by the latch up, provides shut down of the power supply of the affected device and re-establish the power after a pre-specified laps of time. The main drawback of this approach is that the circuit state is destroyed.

The second approach [1] uses guard rings consisting on adding n-type and p-type regions that break the parasitic thyristor structure. This solution is very efficient for preventing latch up. However, it results on excessive area cost which may exceed duplication.

The third approach [2] [3] consists on modifying process parameters for reducing latch up sensitivity. One drawback of this approach is that it may result in an imperfect protection. Also, process modification is undesirable as it incurs significant cost. Finally, process modification may impact circuit performance (in particular for logic parts). Thus, it may be unacceptable in SoCs that include both memory and logic cores.

As a matter of fact, today there is no efficient latch up mitigation solution that has minor impact on circuit size, circuit performance and cost to be acceptable at commercial ground-level applications. The aim of the follow study is to provide such a solution.

# PRINCIPLE OF THE PROPOSED SCHEME

The new scheme combines three principles:

- Latch up detection,
- Latch up removal,
- Correction of latch up induced errors

Latch up detection: When a latch up occurs, the first action of the proposed scheme is to detect its occurrence.

Latch up removal: The second action of the scheme is to limit the current flowing through the activated parasitic thyristor in order to prevent circuit destruction (in case of catastrophic latch up), then the circuit feeds back to normal operating conditions. This action is performed by using current-limiting transistors. These transistors are placed on

the power grid, and can be turned off upon latch-up detection to block the abnormal current induced by a latch-up.

Correction of latch up induced errors: the actions mentioned above bring the circuit back to its normal electrical condition. However, they can not recover probable errors induced by a latch up. Error recovery in memories is usually achieved by means of error correcting codes. However, error correcting codes have limited correction capabilities. For instance, the Hamming code can detect any double error and correct any single-error. As a matter of fact, the use of such codes cannot guaranty correction of errors induced by latch up, since the resulting short between the VDD and ground lines may alter the state of whole memory blocks. To correct the latch up induced errors, the new Hamming code, with intelligent distribution of current limiting transistors along the power lines (for instance the VDD lines) of the memory cell array, in a manner that, the errors are contained into cell domains that can be corrected by the Hamming code.

To summarize the proposed scheme aims to: i) limit the current to prevent circuit destruction, ii) eliminate the latch up condition to bring the circuit back to normal electrical state, iii) contain the errors in cell domains correctable by an error detecting code, like the Hamming code.

A new latch up mitigation scheme has been developed [4] [5]. The proposed scheme is illustrated in Fig. 6.1.

Standard ground power grid is implemented, while the VDD lines of the memory cell-array are implemented vertically, and each vertical VDD line feeds cells that can belong to adjacent columns (vertical VDD lines). These lines are referred as secondary VDD2 lines. There is a trade-off concerning the number of columns connected to each VDD2 line: greater it is the number of columns connected to the same VDD line and greater it will be the number of cells affected by the latch up event when it occurs, on the other hand each VDD2 line requires to implemented a dedicated logic for each line resulting in larger area occupation.

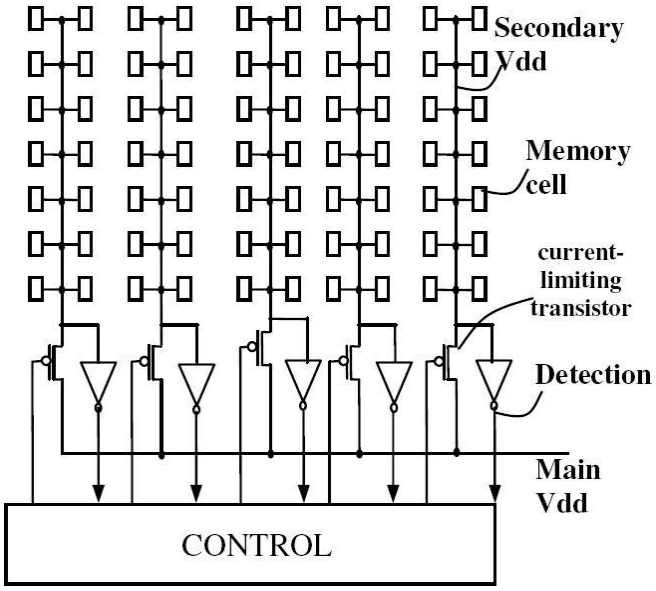



Fig. 6.1 Generic proposed architecture

Each secondary VDD2 line is connected to the primary VDD line (horizontal VDD line forming the external power ring) by means of a PMOS transistor maintained on the on-state by applying on its gate the low level.

When a SEL occurs, the capacitance of the secondary VDD2 line is discharged by the current flowing through the substrate, and its electrical level drops to GND. This is because the resistance of the current limiting transistor is much higher than the resistance of the "short" between VDD and ground created by the latch up. When a latch up short a secondary VDD line to a secondary ground line, either the voltage level of the VDD principal line either the voltage level of the principal ground line is not affected due to the current-limiting PMOS transistor, which limits the maximal current flowing between the principal VDD line and ground. This can be understood easily by considering the fact that power lines in ICs can be designed to drive currents flowing through hundreds and often thousands of gates, without experiencing adverse voltage drop. The current flowing through a single transistor (the current limiting transistor) is

drastically lower than the current flowing through thousands of logic gates. As a matter of fact, no particular care is required to size the principal VDD line in a manner that it does not experience noticeable voltage drop. By maintaining a low voltage drop on the principal VDD and ground lines, we guaranty that the latch up could affect only the state of cells fed by the same secondary VDD2 line. Since memory arrays are usually implemented with column multiplexing, cells belonging to the same word are distant. For instance, for a 1-out-of-16 column multiplexing, the minimum distance between cells of the same word is 16 cells. As a matter of fact, errors in the cells of two neighbour columns will affect a single cell per memory word. Also in the rare case where a single particle traverses several cells, the affected region will necessarily be limited in the neighbour of few cells. Again, due to column multiplexing the errors will affect a single cell per memory word. Therefore, Hamming code will be able to correct the errors, making system operation transparent to a latch up occurrence. Considering circuit damage, the current flowing though the activated thyristor discharges the secondary VDD2 line, and then its value is limited by VDD/Ron (the current flowing through a current limiting transistor), where Ron is the on-state resistance of a current-limiting transistor. This current is quite low and does not threat device integrity, if the latch up condition is quickly removed. As concerning latch up deactivation we can use current limiting transistors of small (e.g. minimal) width. According to the circuit characteristics, the resistance Ron of these transistors could be sufficiently large so that the current VDD/Ron is lower than the critical current required to maintain the latch up. In this case, after initial charge of the secondary ground line, the latch up current falls to the current VDD/Ron. Since this current is lower than the latch up critical current the latch up is deactivated. However, using current-limiting transistors of small width can impact memory performance in unacceptable manner. In this case, we have to use larger current-limiting transistors, which width may not reduce the current at a value lower than the latch up critical current. For these reasons, the latch up elimination scheme has to implements built-in current sensors (BICS) which monitor each secondary VDD2 line. When a SEL occurs, it is detected by the BICS which monitors the corresponding VDD2 line. When a BICS is activated in response to latch up occurrence, the control

circuit drives off the corresponding current-limiting transistor. Then, after a predetermined short lapse of time necessary to deactivate the latch up, the control circuit brings this transistor back to the on-state. During all this time, the circuit operates without interruption and erroneous values read from the affected columns are corrected by the error correcting code.

# LATCH-UP MITIGATION SCHEME ARCHITECTURE

The block diagram of the proposed scheme is sketched in Fig. 6.2. Some annotations are briefly introduced in order to clearly explain the purpose of each component.

The VDD lines are implemented vertically (while the GND grid is not modified). Each VDD line feeds cells that belong to the same BIT column, so that SEL will affects only one memory column. It allows to minimize the impact of latch-up events on the whole memory, at cost of negligible increase of area occupation. The lines are referred as secondary VDD lines (called VDD2). Each VDD2 is connected to the primary VDD line (horizontal VDD line) by means of a PMOS transistor maintained on the on-state by applying on node Q (its gate) the low level (it is represented by cell PNMOSLVT in fig. 6.2). When a Latch-up occurs, the parasitic thyristor related to a bit cell shorts the corresponding VDD2 line with GND (the event is simulated with cell LATCHUP). Current flows from VDD2 line to GND, and it results in a voltage drop of hundreds mV on the resistance associated to the VDD2 line and PMOS. The voltage drop is detected by a High Threshold Inverter (represented by cell RDETECTOR). The Detector output is connected to a Toggle flip flop (within the cell FF_STAGE), thus the toggle input (T) switches from logic ‘0’ to logic ‘1’ and the T-flip flop enters the toggle mode.

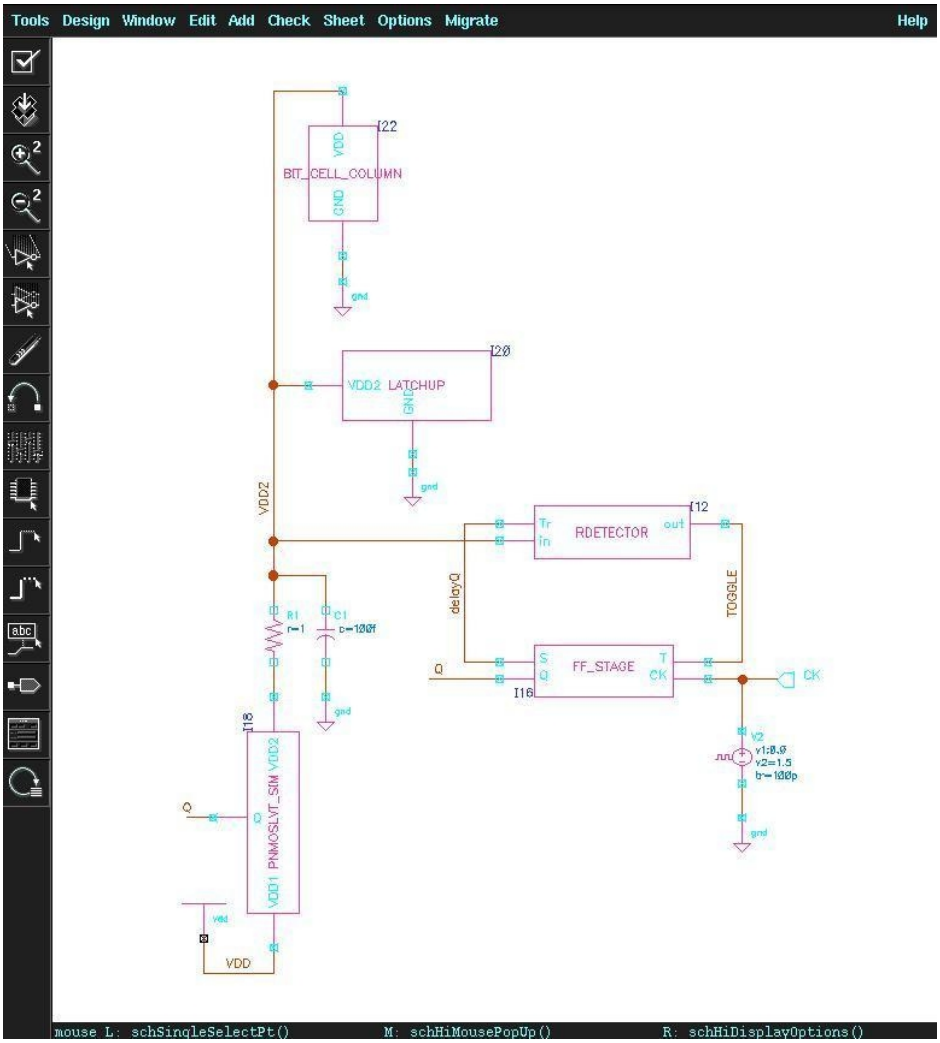



Fig. 6.2 Latch-up Mitigation Scheme Architecture

At the next clock raise edge, the output node Q raises to logic ‘1’, and the value is kept for one clock period. When Q is on high level the PMOS is turned off, so that the VDD2 line is maintained in high impedance and there is no charge to maintain the latch-up condition. After a clock period the node Q toggles to logic ‘0’, thus a low level is applied at PMOS gate and the normal operating conditions are restored.

# CHAPTER 6: SINGLE-EVENT LATCH-UP (SEL) MITIGATION SCHEME

In order to perform SPICE simulations of SEL and validate the proposed latch up mitigation scheme, it has been realized a system composed by:

- **LATCHUP block**
- **PNMOSLVT block**
- **RDETECTOR block**
- **FF_STAGE block**

Let us briefly explain the logic function of each component.

**LATCHUP:** it is the schematic circuit of the PNPN structure used for latch up phenomenon studies. It mimics the SEL in the CMOS technology as it has been explained in the previous chapter. The parasitic Thyristor is realized with NPN and PNP bipolar transistors; the gains are determined thanks to IROC and they are 9.7 and 4.7 respectively. The diffused resistances are represented by ideal resistances: $R_{P\text{-}SUBSTRATE}$ = 600 ohm, $R_{N\text{-}WELL}$ = 1200 ohm. The values of those resistances have been calculated for Chartered 130nm node and they also depend by the layout design.

Different values result in:

- Different voltage triggering values (VTRIG). It determines the susceptibility of the technology to latch up. It is not interesting for the analysis and the development of our scheme.
- Different voltage hold values (VHOLD). It represents the value needed to maintain the thyristor self biased condition when SEL occurs. As consequence it determines the voltage level that should be detected on the secondary VDD2 line (around 1.2V).

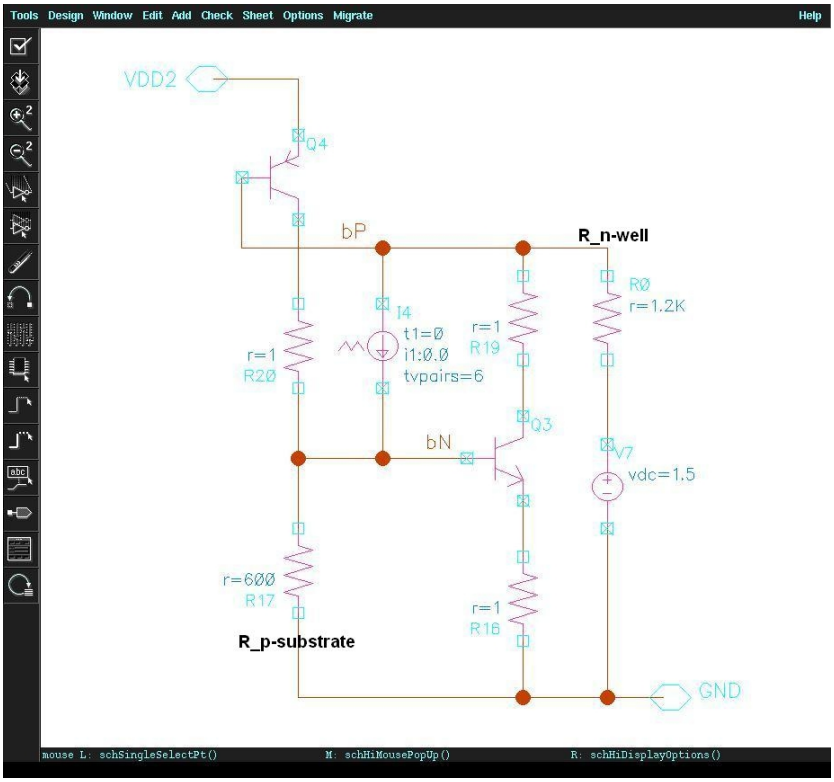



Fig 6.3 Equivalent circuit of PNPN structure used for Latch up SPICE simulations

The extra current due to electron-hole generation process is simulated by applying a current pulse (Dirac function column) of 20mA amplitude and 10ps duration, at the base terminal of the bipolar transistors.

**PNMOSLVT:** it is formed by one PMOS transistors connected with the primary VDD line and the secondary VDD2 line, and one NMOS transistors connected with VDD2 line and the ground ring. Both the transistors are driven by Q node.

To decide the size of the PMOS transistor, it should be noted that the transistor limits he current flowing on VDD2. It acts as current limiter either in normal operating conditions either when SEL occurs. For that reason it should be sized in order to

guarantee the respect of memory performance, thus it is several times greater than the minimum size. On the other hand, relatively small channel width values do not allow to achieve the latch up self-biased condition.

**TABLE I**

**IDS current and Ron of PMOS transistors (Standard and LVT processes) for different values of channel width**

| WIDTH | PMOS STANDARD | | PMOS LVT | |
|---|---|---|---|---|
| | IDS | Ron | IDS | Ron |
| **6 um** | 1.305 mA | 1150 Ohm | 1.907 mA | 785.7 Ω |
| **12 um** | 2.609 mA | 574.9 Ohm | 3.814 mA | 392.8 Ω |
| **18 um** | 3.914 mA | 383.2 Ohm | 5.721 mA | 261.9 Ω |
| **24 um** | 5.218 mA | 287.5 Ohm | 7.628 mA | 196.4 Ω |
| **36 um** | 7.827 mA | 191.6 Ohm | 11.443 mA | 130.9 Ω |

Table I presents the values for IDS currents and Ron resistances associated to PMOS transistor for different width values, pointing out the difference between the standard process (|Vt| = 0.56V) and the Low Voltage Threshold process (LVT, |Vt| = 0.37V).

Fig. 6.4 sketches the parametric simulation of pre-charge process of one VDD2 line for different values of PMOS channel width. The clock frequency of the memory is fixed to 50MHz, thus from Table I and Fig. 6.4, it is worked out that using PMOS with channel width equal to 12 um does not affect the memory performance.

With SPICE simulations it is possible to determine a range of values for the PMOS channel width where the memory performance are not significantly affected and reasonably the latch up cannot sustains since there is not enough current available. Let us examine Fig. 6.5 to better explain the three possible operating regions. A Dirac function current pulse (height 20mA, t = 10ps) has been applied to the structure to generate the latch-up event.

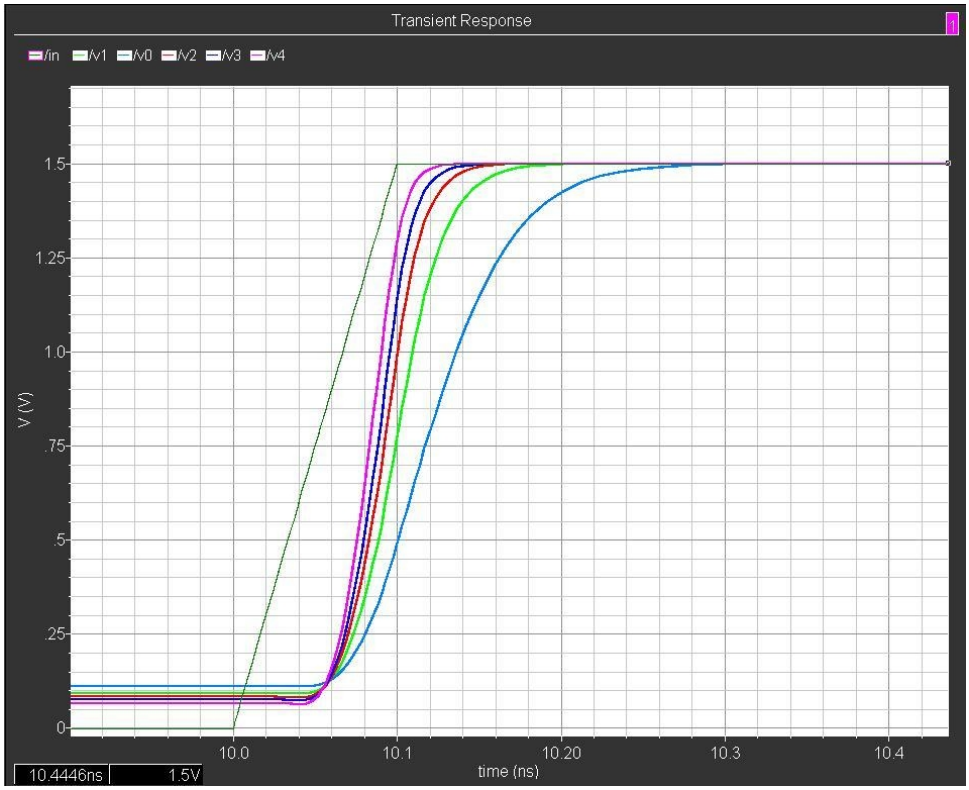



Fig. 6.4 Pre-charge of secondary VDD2 line.

- Case A (Fig. 6.5-bottom): PMOS channel width = 6 um (small). The latch up produces a voltage drop, but the corresponding current flow does not reach the self biased condition. Thus the SEL disappears after fractions of nano seconds, without activate the latch up removal logic and without affects the data stored in the memory. This is the optimal condition to avoid SEL, but the relatively small PMOS size does not guarantee the respect of memory performance.
- Case B (Fig. 6.5-middle): PMOS channel width = 12 um (medium), it does not affect the performances. The parasitic thyristor self biased is not achieved, after a laps of time on the range of few nanoseconds (2-4ns) the latch-up disappears.

- Case C (Fig. 6.5-top): PMOS channel width = 24 um (large), it does not affect the performances. The latch-up event produces a current flow of some mill amperes (4 mA) and the logic related to VDD2 line will be activated to remove the latch-up condition.

Case A should be rejected since it impacts on memory performances.
Case B seems to be ideal since it does not affect the memory performance and it guarantees to avoid latch-up events before that they are established. However, it results critical to be implemented in a prototype that aims to be a demonstrator. In fact, it is not possible to exactly estimate the time duration of voltage drop induced by SEL before that the partially turned on parasitic thyristor switches off again so that the normal electrical conditions are restored. Thus it is not determined if it suffices to activate the correction logic or not. Moreover it is necessary to widely investigate the effects on the data stored in the bit cell. There is the risk that when the SEL is activated, it changes the logic value stored in the bit cell, than it disappears without destroy the device but also without be detected.

We think that it will be possible to answer to those questions after an accurate latch up characterization of the device. In further implementations, it will be necessary to perform physical simulations on the given bit cell (considering the specific layout) and analyze its response to different pulse shapes. It will permit to establish if some results, thus threshold values of the PMOS channel width, are general and inherent to a technology node or on the contrary they also depend by the design. For the moment, the prototype aims to validate the realized latch up detection and removal logic, maintaining wide generality, thus we decided to oversize the PMOS channel width to be sure of operating in the conditions mentioned in case C.

Oversize the PMOS results in a smaller equivalent resistance thus a larger amount of current available and increased probability of achieve the thyristor self-biased condition. For this reason it has been taken W = 24um, which corresponds to 80 times the minimum channel width ($W_{MIN}$ = 0.3 um). Moreover, it has been used the LVT option since the resistance (corresponding to same channel widths) associated to the

transistors is smaller than in standard process as it is shown in Table I.

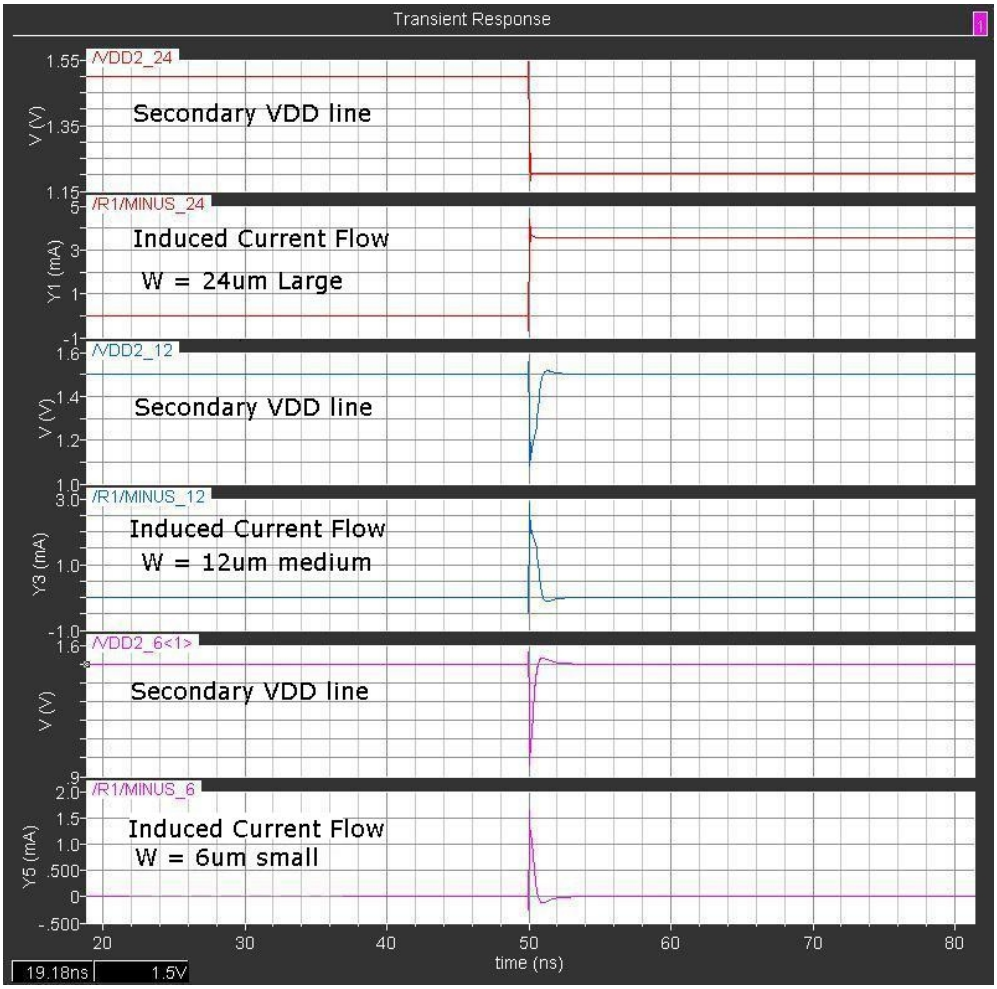



Fig. 6.5 Voltage drop on VDD2 line and latch-up induced current for different values of PMOS channel width size.

Fig. 6.6 shows the layout of the realized PMOS and NMOS transistors. It should be noted that the Drain of the PMOS (top) is connected to the secondary VDD2 line while the Source (bottom) is connected to the primary VDD line. The NMOS transistor connected to VDD2 line and ground ring is important to discharge the VDD2 line when the latch up event is detected. In fact, when node Q rises from low level to high level,

the secondary VDD2 line is in high impedance, also if it will be slowly discharged by leakage currents. During this phase, glitches due to the parasitic capacitance Cgd of the huge PMOS or any other kind of spikes can produce a slightly, temporary rise on VDD2 and results in a commutation of the inverter connected to VDD2 line (the first stage of the latch up detector, as it will be explained in the next section). If it occurs during a clock rise edge, the toggle flip flop within the latch up removal logic enters the idle mode (i.e. the toggle mode is maintained only for one clock period instead of two periods) so that node Q is kept to logic '1' till the memory reset (T = 0, Q = 1, no TOGGLE, VDD2 line is disconnected). The added NMOS discharges the VDD2 line when Q rises to high level avoiding any possible problem, at the cost of adding a minimum width size NMOS transistor.

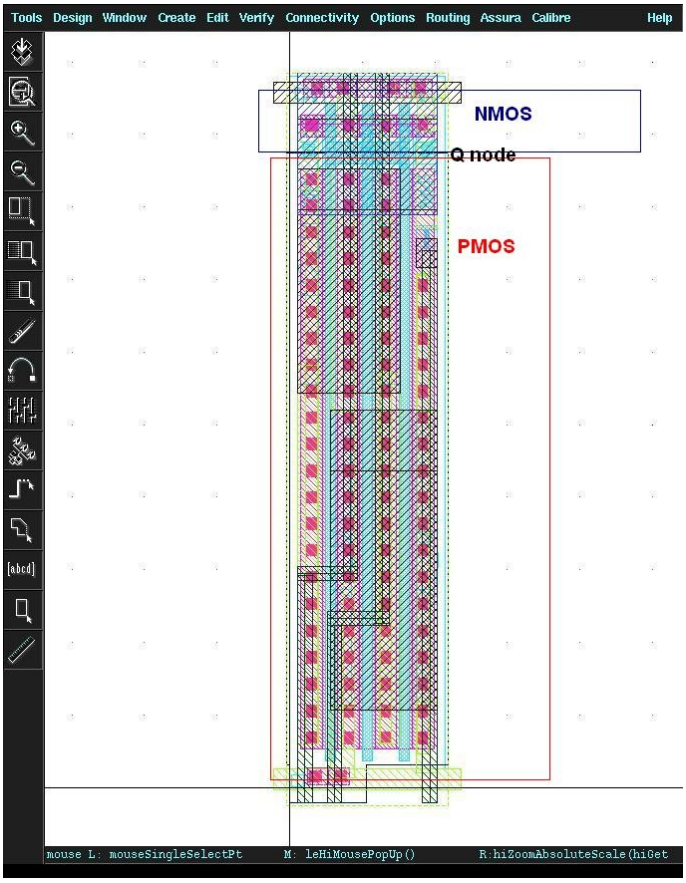



Fig. 6.6 PNLVTMOS layout

**RLINE:** the VDD2 line is represented in Fig. 6.2 by equivalent resistance and capacitance. The secondary supply lines are realized in METAL5 which is a top metal in this process, so that it presents a very low resistance per square. In order to reduce the line resistance, we used the maximum width (1.92 um) allows by the BIT column width (the layout constrains will be explained in the dedicated section). The line length is around 1300 um, since it coincides with the memory height. Thus the resistance is calculated to be 1 mill ohm in the case of a bit cell connected near to the PMOS, and 20 ohm in the case of the farthest bit cell.

**RDETECTOR:** it refers to the logic needed to detect the voltage drop produced when a SEL occurs. The voltage value on VDD2 line during latch up event is directly related to the value of VHOLD (the means of hold tension has been explained in the previous chapter). Taking in account the technology parameters for Chartered 130 nm node and the parameters depending by the memory layout (i.e. dimensions of the bit cells, distance to N-tap and P-tap), SPICE simulations show that in our case the value of VHOLD is 1.2V. When the parasitic thyristor achieves the self biased condition, the corresponding voltage drop on VDD2 line is in the order of 300 mV. Thus, to detect the latch-up by voltage detection it is needed to implement a voltage comparator or realize High Threshold Inverter. The first solution requires the use of a voltage reference cell at 1.3V, and the circuit is surely more complex and presents greater area occupation.

To realize a HT inverter with threshold above 1.2V, it is mandatory to use the LVT process. The PMOS transistor is oversized ($W_p$ = 12 um) while the NMOS transistor is minimum sized ($W_n$ = 0.3um) and it is realized with standard process.

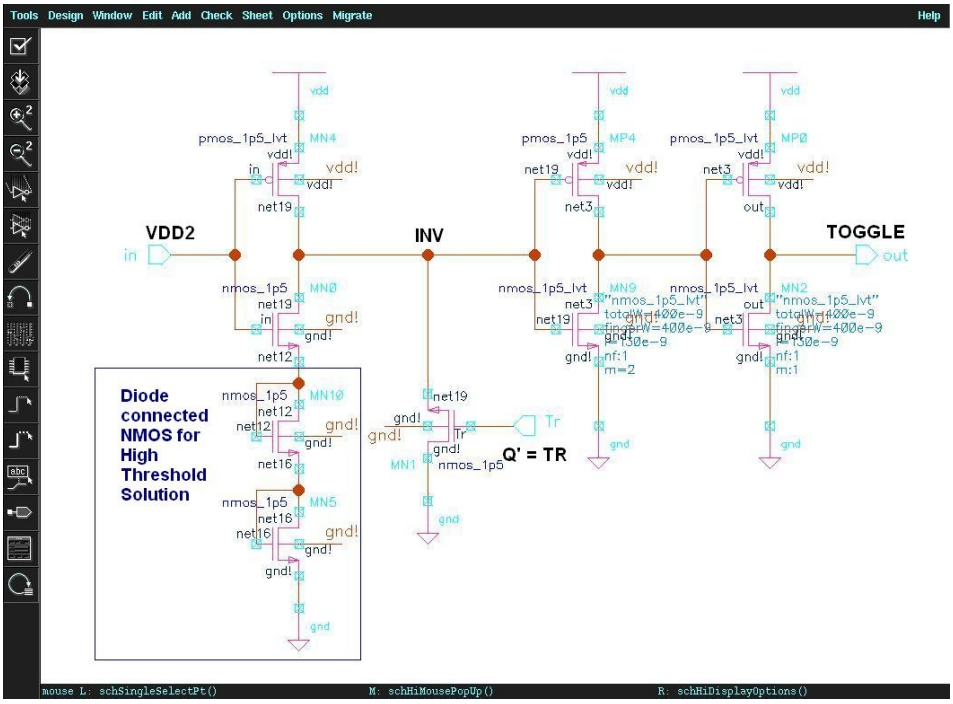



Fig. 6.7 RDETECTOR cell schematic

The realized circuit is sketched in Fig. 6.7. The HT inverter is followed by a small buffer that mixes standard transistors and LVT transistors in order to optimize sensibility and minimize the time needed to logic switch (detection), thus to activate the following stages.

Fig. 6.8 represents a parametric simulation of the Detector response for different voltage drop values. The initial voltage value on VDD2 line (this node coincides with the inverter input) is equal to 1.5 V, then at t = 20 ns (see marker) a negative pulse is applied. The voltage values on VDD2 line are 1.18V, 1.22V, 1.26V, 1.3V, 1.34V and 1.38V. The HT inverter detects the variation on his input node and consequently switches the output node (TOGGLE). The times needed for the commutation depend by the voltage value on VDD2 lines. They respectively are: T1 = 1.25 ns, T2 = 2.6 ns, T3 = 7 ns, T4 = 20 ns, T5 = 80 ns and T6 = no commutation.

We estimated that the voltage value on VDD2 decreases to 1.2 V so that it is detected in the time of 2 ns. Anyway the realized circuit can detect higher values on VDD2 line (up to 1.3V, i.e. smaller voltage drop induced by latch up event) and activate the recovery logic in a time smaller than the clock period (fixed at 20 ns, f = 50MHz). The system is also robust against glitches on VDD since spikes smaller than 200 mV or with duration of few nano seconds do not cause a commutation of the inverter.

It is worth to notice that when the VDD2 line is restored to 1.5V, the inverter output discharges very slowly and due to the circuit architecture the switch on the output node from logic '1' to logic '0' will take hundreds of nano seconds

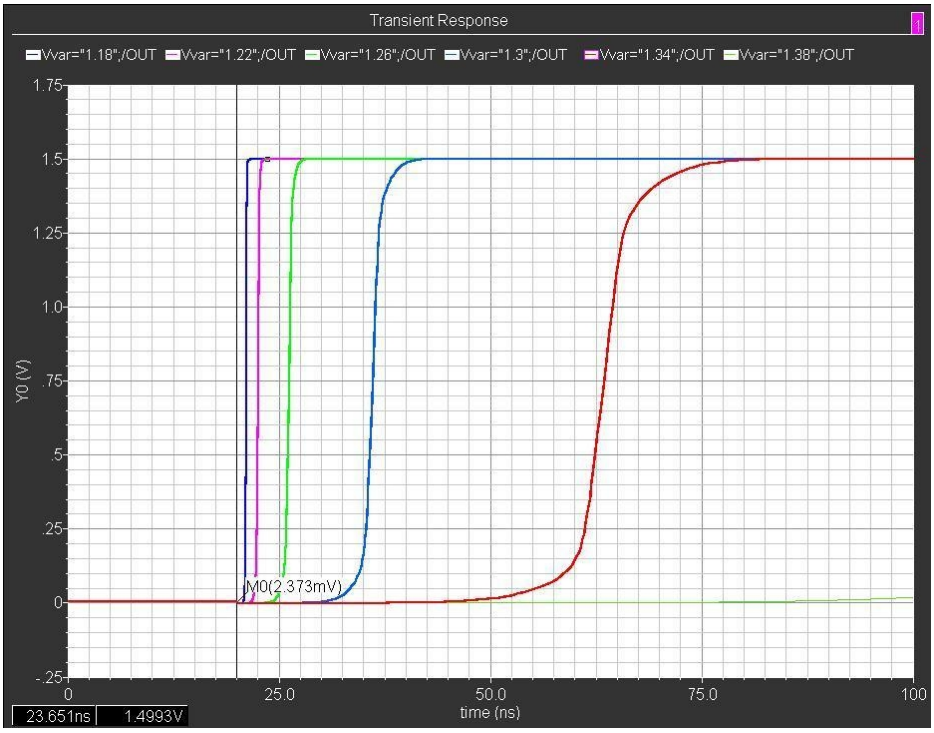



Fig. 6.8 Parametric simulation of the Detector response: commutation on TOGGLE node for different amplitude pulse on VDD2

.

Fig 6.9 represents the behavior on INV node (bottom) and TOGGLE node (top). It has been performed a parametric simulation with different voltage drop values (i.e. voltage on VDD2 line equals to 1.18V, 1.22V, 1.26V and 1.30V). Fig. 6.9-top shows that the voltage drop is detected and the output node (TOGGLE) switches from logic '0' to logic '1'. Then, in correspondence of the 3rd marker (T = 60 ns), the input voltage is restored to 1.5V. Fig. 6.9-bottom shows that the voltage on INV node decreases very slowly (hundreds of ns). It can results in an error: in fact the Toggle flip flop controlled by the TOGGLE node is kept in Toggle mode longer than necessary.

So there are two possible conditions:

CK samples T = 0 when Q = 0 => correct
CK samples T = 1 when Q = 0 => next clock period: T = 1 Q = 1 => T = 1 Q = 0 => T = 1 Q = 1 => … depending by the unpredictable discharge time CK will sample T = 0 when Q = 1 ERROR or CK will sample T = 0 when Q = 0 correct

Thus, depending by the discharge time and depending by when the clock signal rises, we not only waste clock cycles before restore the correct operating condition, but there is also the risk to income in a error

It is necessary to guarantee that the INV node discharges within a clock period. It can be accomplished by sizing the buffer stage in order to respect the discharge timing constrain (i.e. $T_{discharge} < Tck$). The disadvantage of this solution is that it makes worsen time needed to detection. Otherwise, it is possible to optimize the circuit for the detection action and then add a NMOS transistor, connected to INV node and to ground, to force the discharge of INV node when the tension on VDD2 line is recovered to the correct value. The signal (Q' = TR) to control the NMOS transistor is generated by a simple logic which will be illustrated in the next section.

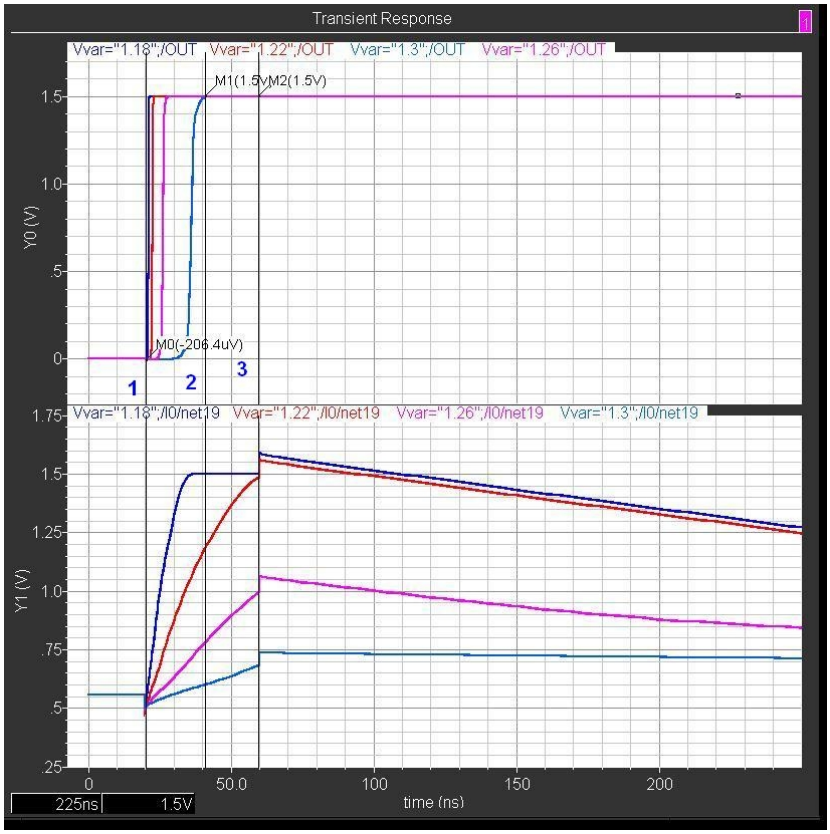



Fig. 6.9 Detection stage: discharge of INV node

**FF_STAGE:** this cell is composed by a Toggle flip flop and a Master Slave flip flop.

When the SEL is detected, the only way to interrupt the latch up phenomenon is to keep the supply to a value below the value of VHOLD. This is achieved in the proposed architecture by cutting the connection between the primary VDD line and the secondary VDD2 line. The signal on PMOS gate rises to a high level, then, when the extra charge produced by the ion's strike is collected to ground, it is necessary to restore the normal operating conditions. In order to achieve this function the circuit implements a toggle flip flop. The T-flip flop is formed by a toggle latch (sketched in Fig. 6.10) followed by a simple latch stage. The truth table shows that when T = 0 the flip flop is in Idle mode

and the output node Q is kept at the previous value, while when T = 1 the T-flip flop operates in toggle mode and at each clock raise edge Q toggle from '1' to '0' then from '0' to '1'.

.

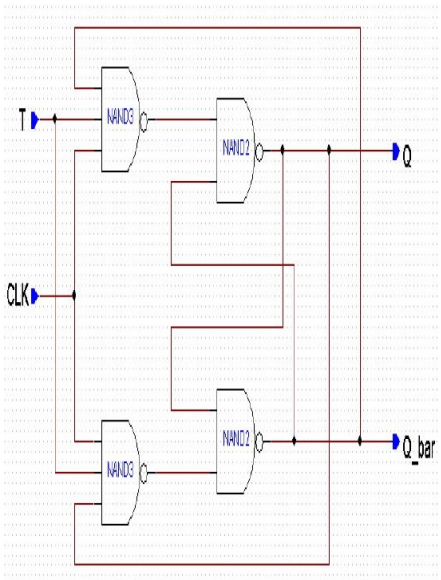



Fig. 6.10 Toggle latch stage

**T-flip flop truth table**

| T | Q' | State |
|---|---|---|
| 0 | Q | Idle |
| 1 | Not Q | Toggle |

When the SEL is detected on the TOGGLE node, the toggle input T switches to logic '1' and the T-flip flop enters the toggle mode. At the first clock raise edge, the output node Q, normally kept in low state, rises to logic '1'. VDD2 line is disconnected by VDD and it is discharged to GND, thus latch up event is interrupted. After one clock period, node Q toggles to logic '0' and VDD2 is charged to 1.5V, thus T is switched to logic '0' and the T-flip flop enters the idle mode (T = 0, Q = 0, no toggle, VDD2 re-charged to VDD). This solution guarantees that the secondary supply line is disconnected for one clock period, enough to dissipate the extra charge, than the normal operating conditions are restored

Standard toggle flip flop cell cannot be implemented in our design since it does not respect the layout constrains. For these reason it has been proposed a circuit that achieves the toggle function using a reduced number of transistors, saving area and all above reducing the routing complexity. The schematic of the proposed T-flip flop is sketched in Fig. 6.11. The first stage realizes the toggle stage, while the second stage is a normal latch stage. The circuit also internally generates the inverted signal NOT Q, the internal clock CI and the internal reversed clock CN. The clock phases should be regenerated within each cell to avoid skew problems. It worth to notice that the operations of logic cells are not synchronous, since they are activated when a SEL occurs on the related bit column. Thus, if the timing is respected inside the single cell, it is not necessary to implement a complex tree for external clock transmission. The PMOS transistors of the second stage have been oversized (2 times the minimum channel width) in order to drive the input capacitance of the PMOS on VDD line, without the necessity to use a buffer. The whole circuit uses 16 transistors instead of the 28 transistors of the classical schemes.

The output node Q drives the PMOS connected to VDD line and VDD2 line, but it is also entered a master slave flip flop (it is realized by using a similar schematic).

The MS flip flop simply stores the value of Q node and presents it on his output delayed by one clock period. The output signal (TR) controls the NMOS connected with the inverter output node (as mentioned in the previous section). When Q decreases to a low level, VDD2 line is recharged to 1.5V, thus INV node slowly discharges. The

NMOS is turned on by TR (= Q') in high level and it guarantees the discharge of the INV node.

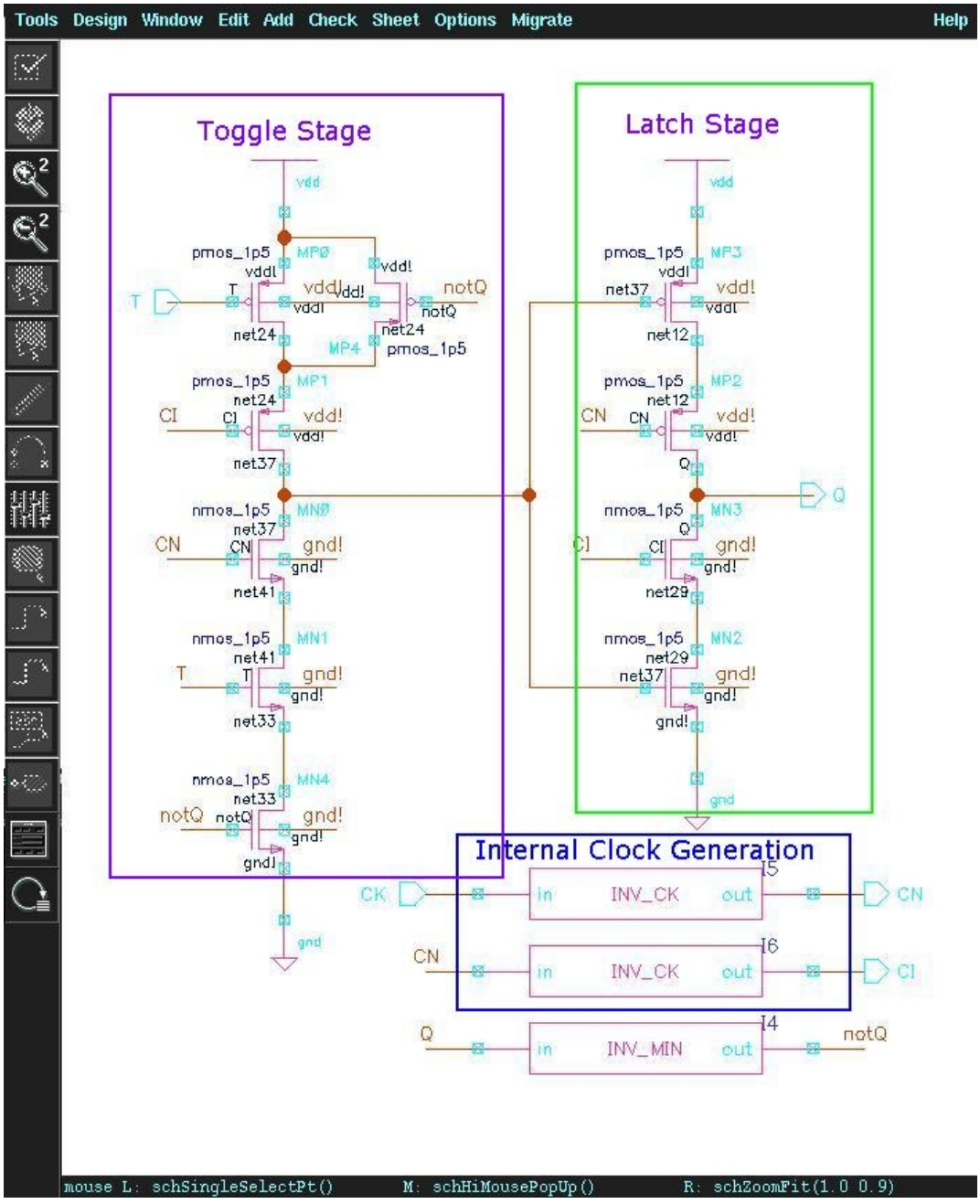



Fig. 6.11 Proposed T-flip flop schematic: Toggle stage, latch stage. The positive and reverse clock phase is generated internally

# LOGIC FUNCTIONS DESCRIPTION

This section explains the operation of the implemented LOGIC, carrying out the time constrains. We refer to Fig. 6.12 to describe each phase of the latch up event detection and correction, pointing out the function of each component mentioned above.

It should be noted that the clock period is equal to 20 ns and it is fixed by the memory that have to be modified, but it does not constrain the proposed architecture. In fact the detection stage operates asynchronously and the removal logic can operate at higher frequencies.

1) t: 0ns – 25ns. Initial set up.
T = 0, Q = 0, T-flip flop is in idle mode, TR (delayed Q signal) = 0, secondary VDD2 lines are charged to VDD = 1.5V.

2) t: 25ns – 50ns. Normal operating conditions.

3) t = 50ns (1$^{st}$ marker), latch-up event. A current pulse (Dirac function not showed in figure) is applied on the base terminals of the NPN and PNP bipolar transistors. The parasitic thyristor is turned on connecting VDD2 line to GND. The formed low impedance path results in a current flow in the order of 4mA. The current flow produces a voltage drop on the VDD2 line resistance and the voltage on VDD2 line decreases to 1.2V (appreciatively equals to VHOLD). It activates the detector stage.

4) t = 57ns, SEL detection. The High Threshold Inverter does not immediately detect the voltage drop: it takes 7ns (worse case) to commutate the output node from logic ‘0’ to logic ‘1’. When the switch is completed the T-flip flop enters in toggle mode (T = 1).

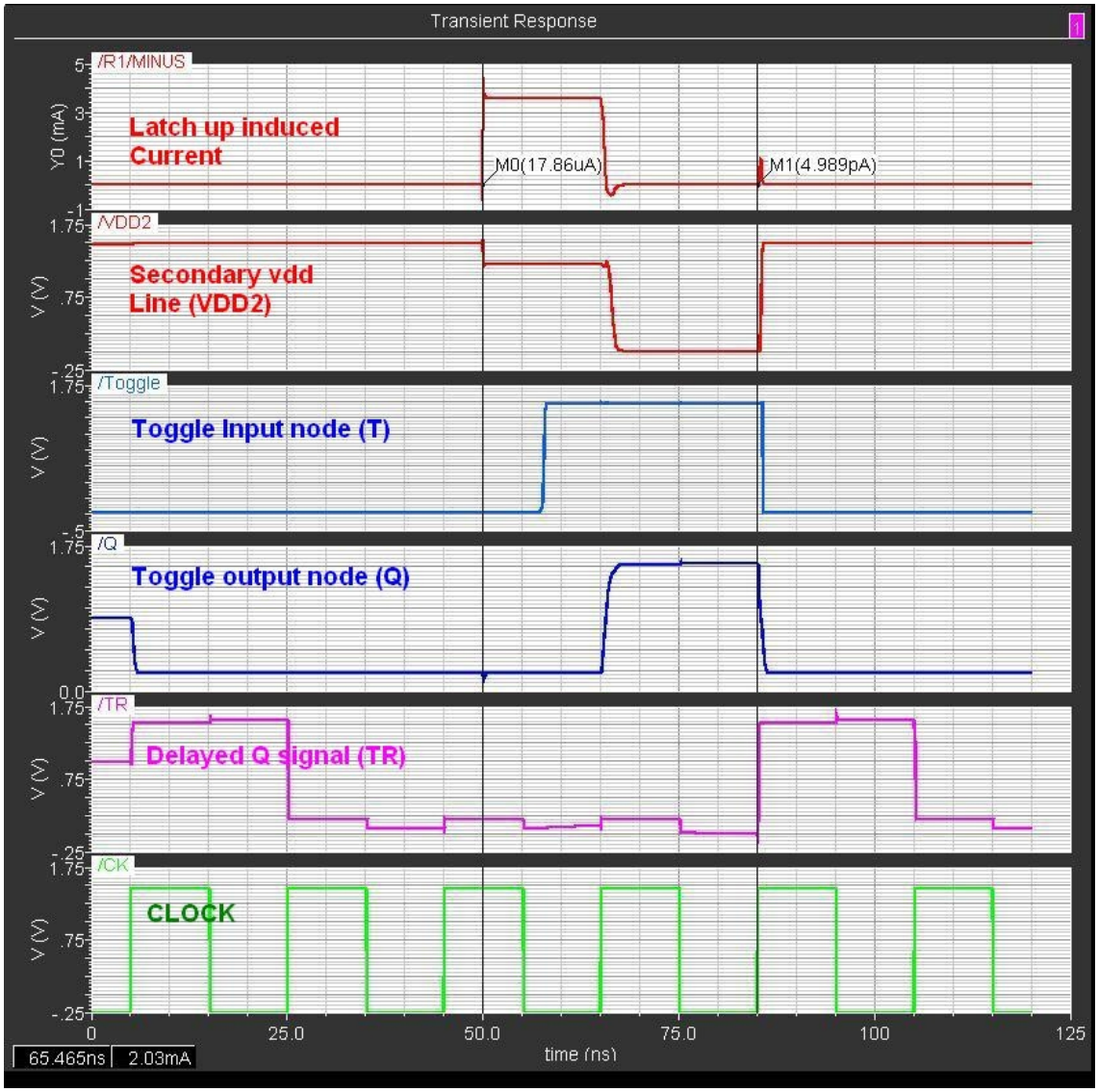



**Fig. 6.12 SEL occurs. Latch-up is detected. The secondary VDD2 line is disconnected from the primary VDD line, thus Latch-up event is interrupted. VDD2 line is re-charged to 1.5V and the normal operating condition is restored.**

5) t = 65ns, latch-up correction. During the first clock raise edge, the node Q toggles from logic ‘0’ to logic ‘1’. Thus the PMOS connected to VDD and VDD2 lines is turned off. VDD2 line is discharged to ground through the small NMOS transistor

driven by Q. There is not any more charge available to maintain the Latch-up condition, thus the induced current I-latch-up decreases to 0uA.

6) t = 85ns, recovering to normal operating conditions. The clock raise edge samples T = 1 (toggle mode), the output node Q switches from logic '1' to logic '0'. VDD2 line is again connected to VDD ring and re-charged to 1.5V. TR node switches to logic '1' (it follows Q delayed by one clock period) and turns on the NMOS which helps to discharge the output node of the HT inverter. The T-flip flop input (T) decreases to logic '0', so that the next clock raise edge will sample the T-flip flop in Idle mode and the output Q will be maintained to logic '0'.

Finally let us define:

- $T_{INV}$: time needed by the HT-Inverter to switch from logic '0' to logic '1'
- $T_D$: time between the commutation on T-flip flop input node (T) and the first clock raise edge

Worse case for $T_{INV}$: the voltage on VDD2 line is fixed by the latch up hold tension VHOLD, but it also slightly depends by the resistance of VDD2 line. The whole resistance consists in the resistance of the PMOS and the resistance of the VDD2 metal line ($R_{LINE}$). The PMOS resistance is equal for every cell connected to the bit column. $R_{LINE}$ depends by the position of the bit cell: it varies from 1mill ohm for the nearest cells to 20 ohm for the farthest cells. It result in a slightly variation of the value of the voltage drop on VDD2 line. If the SEL affects a cell near to the PMOS the voltage drop is smaller and it will be harder and slower to be detected. Anyway the difference is on the range of 40mV, so that $T_{INV}$ is reasonably in the range of 1ns – 7ns.

Worse case for $T_D$: it occurs when the clock raise edge is just before the commutation on the T-flip flop input node (T). In this case the latch up event is detected but it is necessary to wait one whole clock period before discharge the secondary VDD2

line ($T_D$* = Tck).

Worst case to detect and stop the latch-up event:

$T_{TOT} = T_{INV}* + T_D* = 7ns + 20ns = 27ns$

# LAYOUT DESIGN

The proposed architecture has been realized in order to be implemented within a given SRAM. Thus there are layout constrains that should be respected and some modifications to the original memory layout have been necessary.

We briefly describe the Memory architecture before explain the layout design solutions.

## MEMORY COMPILER

The presented prototype contains two SRAMs realized on different layers. Afterwards, the memories are connected forming a 3D integrated memory using the technology process developed by Tezzaron Semiconductor. The prototype is realized in standard CMOS 0.13um Technology provided by Chartered. Since the memories are generated using the Generic Single Port RAM Compiler (GSPRAM), which is based on the Generic Design Kit GPDK 0.18um CMOS technology, a further effort to pass from a technology node to the other one has been done.

The RAM compiler is a tool which offers the service of generating the required views of a RAM block according to the parameter values supplied by the user.

The output of the compiler consists of:

- The layout (full layout and abstract)
- The transistor level description (for physical verification)
- The behavioural model (functional and timing)

The full layout of the RAM block is produced by creating a mosaic of pre-defined tile cells. Unlike standard cells, these tile cells do not have the same height, but all cells in a row will have the same height. Another difference is that each tile cell alone does not necessarily pass the DRC and LVS checks, because it relies on its neighbours for the well contacts.

Routing the internal signals of the RAM is entirely done by abutment, each tile cell containing the necessary wire segments. The routing of internal signals is made antenna-free by construction. Protection diodes are included for the input signals, thus the external routing to the input pins does not require antenna precaution. The RAM uses metal layers up to Metal4 included; Metal5 and Metal6 are available for free routing on top of the RAM block. Power and Ground lines are routed as a redundant grid, to providing low impedance supply to any region. In addition power rings are placed all around the RAM block. It worth to notice that the described power distribution scheme has been modified and carefully replaced in order to implement the Latch-up mitigation scheme as previously mentioned.

The RAM is made of fully static CMOS. This means that power is consumed only during the signal transitions. While the input signals are stable (in any state), only a process-dependent leakage current is drawn from the power supply.

# ARCHITECTURE

The bit cell is the most critical element because it is repeated a big number of times, having the greatest impact on the block dimensions. The dimensions are a concern not only in terms of silicon area cost but also in terms of speed, because long metal lines crossing the block have significant parasitic capacitance. The RAM is based on the classical “six transistors” bit cell sketched in Fig. 6.13.

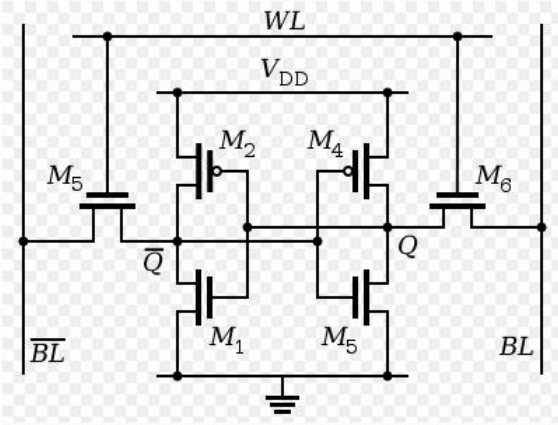



Fig. 6.13 Standard 6-T Bit Cell

This cell is devised to be placed in a 2D array where word select lines are routed “horizontally” and data buses “vertically”. The vertical data buses or bit lines BL and not_BL are used for writing and reading. They are redundant, i.e. they carry the same information complemented. This redundancy is necessary because the N-channel MOS pass transistors are efficient only to drive low state signal. The capacitance of these bit lines may be significant, so it is necessary to pre-charge them to the same voltage before reading, otherwise they could cause the bit cell to flip unexpectedly to the opposite state during reading.

In a simple array, the number of columns is equal to the word size WS (number of bits per word), while the number of rows is equal to the number of words NW. But in most cases this array would have a very unsuitable aspect ratio, its height being very great compared with its width. The usual solution is to reduce the number of rows by multiplexing multiple words on the same row. This is done by splitting the array into a number of blocks. In order to have an efficient control on the block aspect ratio, the bit cell array is split into NB blocks, each block being one word wide (WS columns). The number of rows being NR, the capacity of the RAM is NW = NB * NR words. So far a given NW, there is a number of possible splitting combinations giving different aspect ratios, offering the opportunity to adapt the RAM aspect ratio to the chip's floorplan.

In order to occupy all the space available on silicon prototype the used memory is formed by thirty-three Blocks (NB = 33), four hundred ninety-two rows (NR = 492). The whole word size is twenty-two bits, in fact it consist in 16 word bits plus five bits to implement the Hamming code, and the parity bit . It has been realized a sparse memory (i.e. capacity $\neq 2^n$ ) in order to occupy all the area available on chip, since there will be higher probabilities to have ion's impact during the dedicated tests. The whole capacity is:

Capacity (words): 16236

Capacity (bits): 357192

## LAYOUT OF THE PROPOSED SCHEME

It has been decided to connect each secondary VDD2 line to only one memory column. There are two solutions to realize the corresponding SEL mitigation scheme:

- Realize a single logic stage after the detection stage and multiplex the output of the logic stage in order to properly control each VDD2 line
- Realize a dedicated detection stage and logic stage for each bit column

The second solution presents a higher flexibility since it is not related to the size and the shape of the memory. The logic cells are added to each bit column, and in the case of changing the memory configuration it is not needed to change the cells routing or any multiplexer stages.

Since the VDD2 line supplies only one BIT column, the related control logic should be implemented below each column. The layout design is constrained by the column tile width which is equal to 2.14 um.

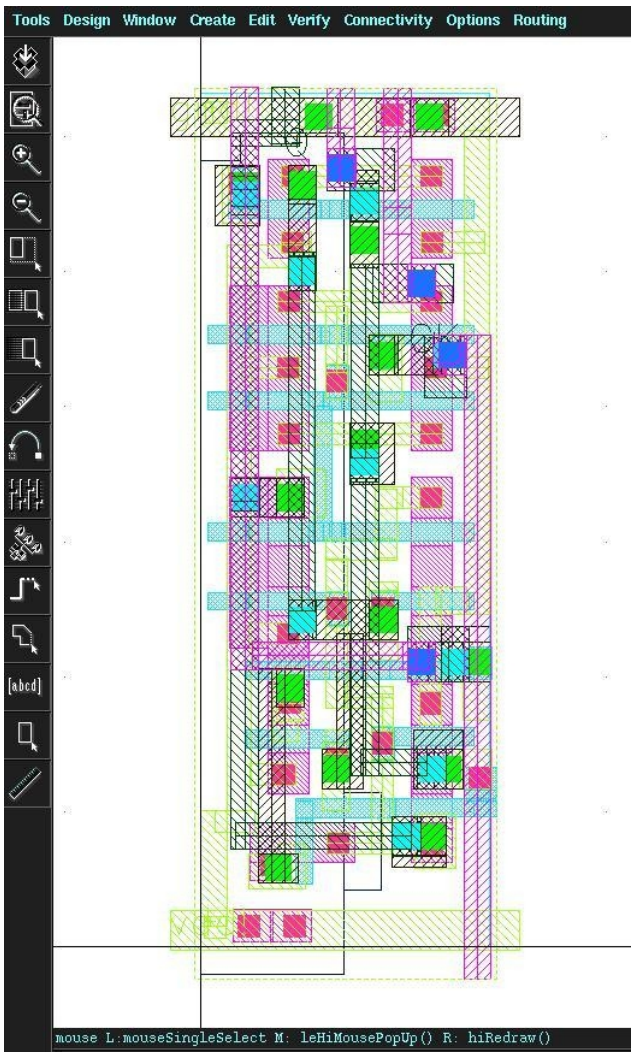



Fig. 6.14 FF_STAGE layout

Fig 6.14 shows the layout of the FF_STAGE (i.e. T-flip flop plus M-S flip flop). The internal connections are made in Metal1, Metal2, and Metal3.While Metal4 is used only for the routing of the external clock signal and the T-flip flop output node Q which should be connected to several parts of the logic that stands above.

The internal VDD and GND lines are made in Metal1 and Metal2 respectively, so that when the cells are realized within the memory, the power nets are already connected horizontally. Moreover, adding vertical Metal1 and Metal2 strips in correspondence of N-tap and P-tap tiles implemented in the original memory, it is also possible to automatically have vertical power connections. It worth to notice that the logic does not consume great amount of energy, in fact a cell is activated only when the related bit column is affected by a latch up event, and the logic remains active for one clock period.

Fig. 6.15 shows the layout of the whole implemented logic. From the bottom to the top is possible to distinguish: the T-flip flop, the M-S flip flop, the huge PMOS connected to primary VDD line and to secondary VDD2 line and finally the Detector stage. On the top of the cell it is added a tile formed by two metal2 strips and called BG cell. So that, in the memory is possible to replace the original BG tile with this modified tile in order to directly implement the logic in the memory design.

From Fig. 6.15 it is worked out that the cell width is 2.14 um (as previously mentioned) and its height is 28.425 um.

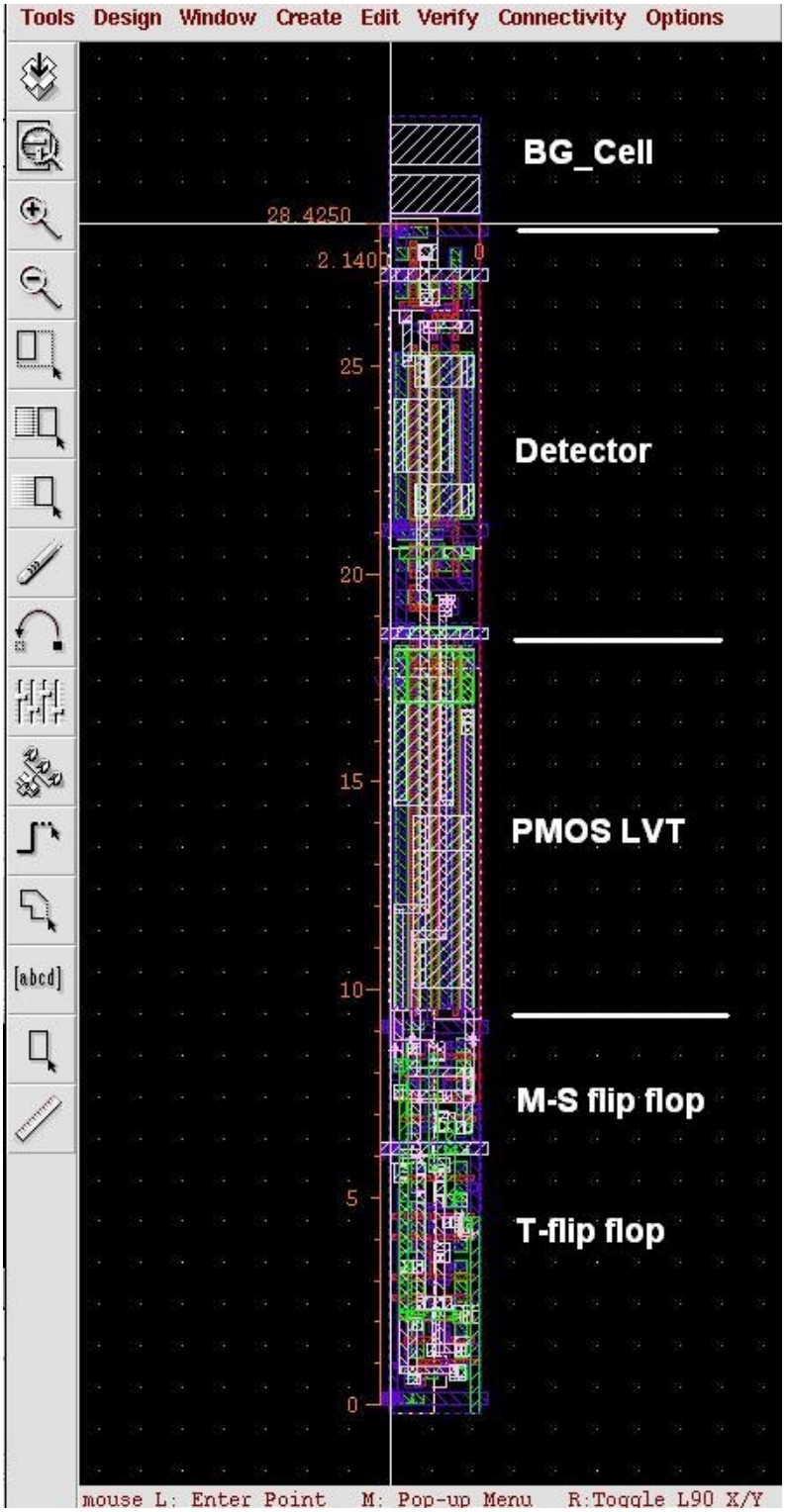



Fig. 6.15 Logic layout

## LAYOUT MODIFICATIONS OF THE ORIGINAL MEMORY

The memory is organized in three parts:

- Block Decoder
- Row Decoder
- Word Block

Every component links tiles contained in the GSPRAMtiles library. The memory layout is modified by replacing the original<TILE> with the corresponding <TILE>_MOD. In the next implementation, it will be possible to directly link the <TILE>_MOD instead of <TILE> when the GSPRAM compiler generates the memory, so that the Modified Memory can be directly obtained without any extra work It worth to notice that some tiles contain the same layout and have the same purpose, but they are called in different ways depending by where they are collocated (i.e. on the top or bottom of the memory, or in Decoder or Word or Block cells).

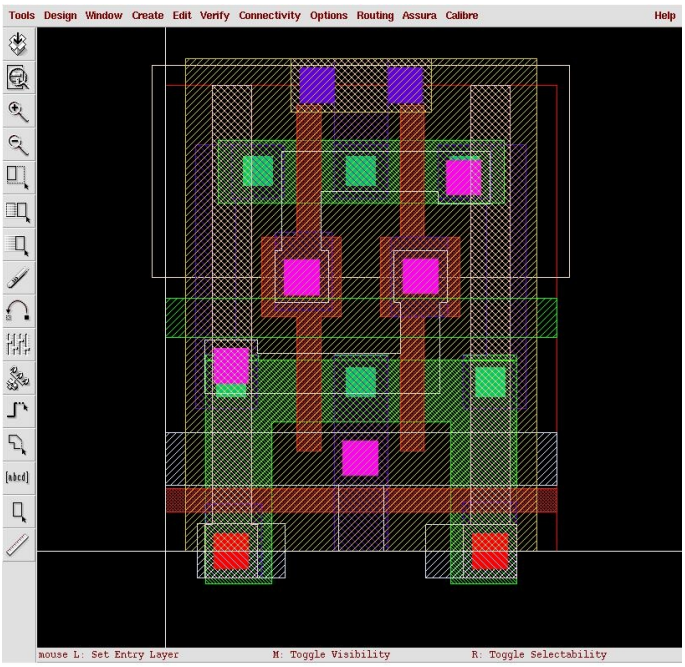
Let us explain the main modifications made to each tile and the related scope.

LOGIC_OPT (or BG_MOD): It contains the Metal2 strips corresponding to VDD and GND power rings. It adds the realized logic.

BI_MOD: (in Decoder) added Metal5 vertical strip to realize the secondary VDD2 line

BM_MOD: (in Block) Bit cells. The horizontal VDD line (metal2) has been removed and replaced with vertical VDD2 line (metal5) and Via1 to Via5. It is not mandatory, but a rectangle of metal2 it has been added on the bottom; it realizes a vertical connection between GND line to have a better power distribution.

Fig. 6.16-top and Fig. 6.16-bottom sketch the bit cell layout before and after the modification



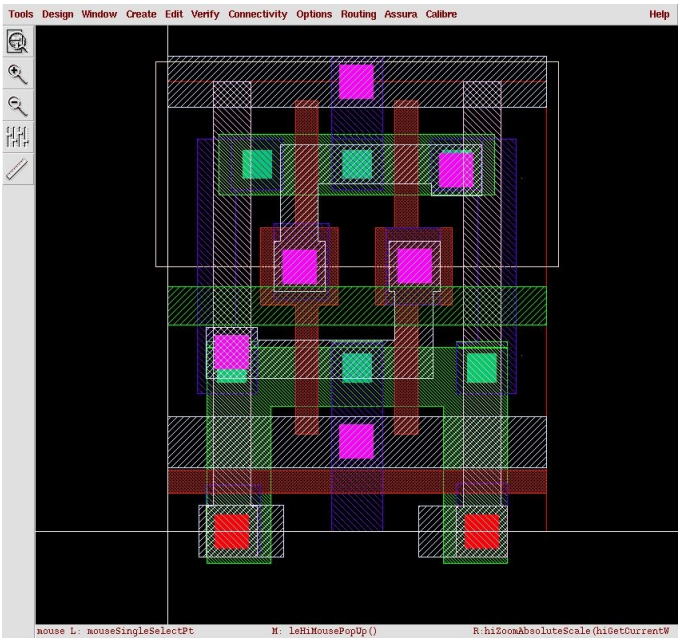



Fig. 6.16 Bit cell layout. Top: original cell (horizontal VDD line in Metal2). Bottom: modified cell (vertical VDD line in Metl5)

CA0_MOD & CA1_MOD: (in Decoder) normally those cells should not be modified. Anyway due to the shrink from 180nm node to 130nm the metal2 and metal3 rectangles placed under Via2 and Via3 respectively are not well aligned. It results in a DRC error (space <0.21 um). The rectangle have been aligned with the metal4 strip (i.e. also with Via1, Via2 and Via3)

In some regions, the cells rows are flipped upside-down every other row, in order to share N-wells and substrate taps. The N-taps or P-taps are always connected to VDD or GND with vertical or horizontal lines or both. When we cut the horizontal VDD lines some taps inside the memory are floating, so that the N-wells are not connected to VDD. In order to solve the problem it is necessary to modify some tiles: all the N-taps are connected to VDD vertically, while the P-taps are connected to GND horizontally. The tap interested in the process are marked as P<name> or S<name>, where <name> depends by the position of the tile in the memory architecture.

SG_MOD & PG_MOD: (in Decoder) connect the vertical line to the external power ring. Via2 placed on the top instead of on the bottom to connect the metal3 line to VDD instead of to GND.

PN, SN, PA, SA have metal3 and metal2 strips but they don't have any Via contact so the should not be modified.

PI_MOD: (in Block) Via2 placed on lines 1, 6, 7 and 14 (counting from top to the bottom) instead of lines 3, 5, 9 and 12. Add Via1 on lines 5 and 12. To vertically connect the taps to VDD instead of to GND.

SI_MOD: Idem. Via2 and Via1 on lines 1, 4, 5, 8 instead of on lines 2, 3, 6,7, and Via1 added on lines 2 and 7.

PT_MOD & ST_MOD: (in Block, top tiles) Via2 placed on bottom strip instead of on top strip to connect VDD instead of GND.

PM_MOD & SM_MOD: (in Word) Via2 placed on top strip instead of on bottom strip to connect VDD instead of GND. Via1 both on the top strip and bottom strip.

LOGIC cells contain strips to connect the cells to VDD or GND. The connection is only horizontal. VDD nets are realized in metal1, while GND nets are realized in metal2. However, it is possible to automatically realize a GND grid or VDD grid with vertical connections. To do that, all the tiles marked as P<>_MOD or S<>_MOD contain vertical metal1 strip and vertical metal2 strip (width: 0.29um, length: 28.5um).

CG_MOD: (in Decoder). In the standard memory this cell is sometimes placed between two columns. It has been added horizontal metal1 and metal2 strips on the bottom to guarantee the horizontal VDD or GND connections of the added Logic.

# LAYOUT AND CHARACTERISTICS OF THE MODIFIED MEMORY

The Logic is implemented at the bottom of the memory, directly connected with the VDD and GND power rings. It is completely realized externally to the original memory foot prints. It is possible to implement a part of the Logic within the memory, in the space between the Block Decoder and the Word block. In this case the layout design of the logic is not more complex, but it presents two disadvantages:

1] Several internal memory tiles have to be modified, since the standard memory architecture is not perfectly regular: there are particular cells dedicated to protection diodes, some cells are flipped sideways every other column in order to share contacts or vias and sometimes the space between N-well is not respected.

2] The memory compiler automatically generates the Decoder signals, whose number depends by the number of Blocks (NB) and word size (WS). This determines the space available inside the memory to implement the Logic. Thus this solution is related to a specific memory size and shape and it is not so flexible.

The realized memory is sketched in Fig. 6.17. From the figure it is not possible to distinguish the layout of the cells. Anyway the parts forming the memory are visible. Moreover, the red line highlights the original memory size, while the blue line highlights the added logic. Fig. 6.18 is focalized on the regions containing the new architecture. The red rectangle highlights the original tiles containing the block decoder cells, the generated signals and the power rings. The blue rectangle highlights the cells forming the latch up mitigation scheme.

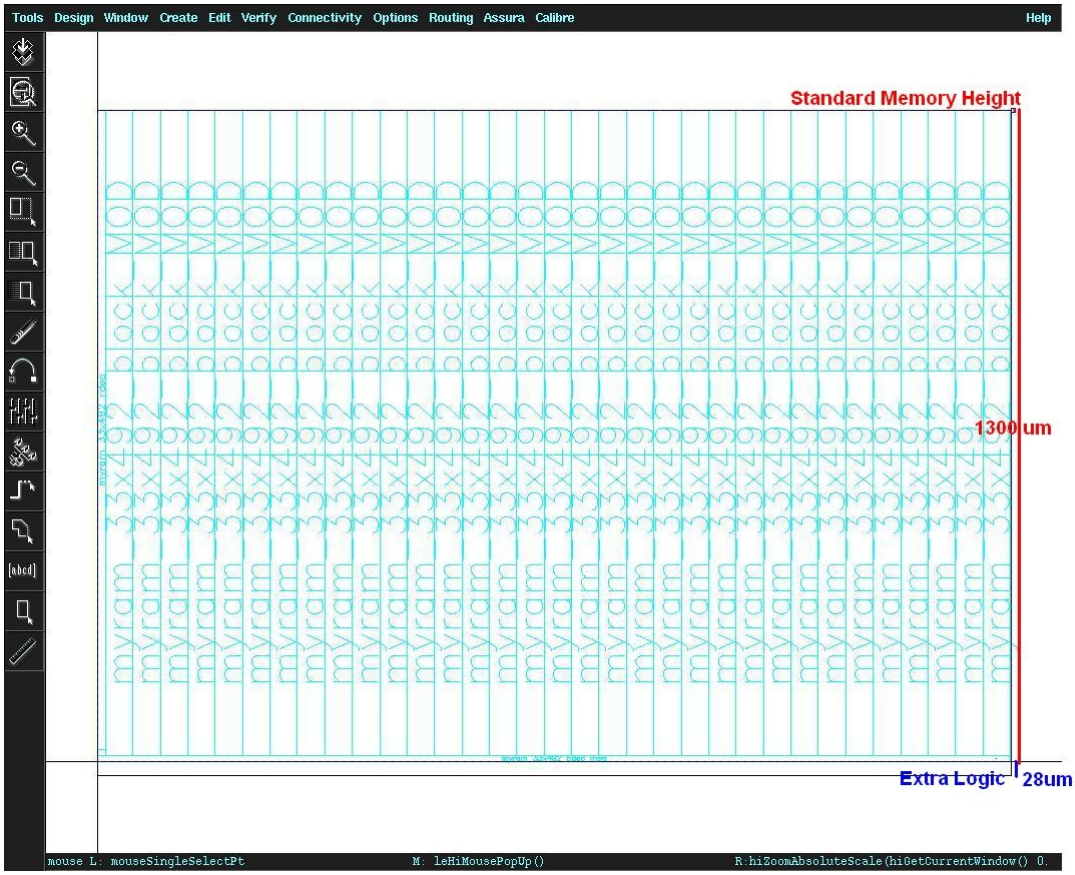



Fig. 6.17 Memory footprint

Finally, the characteristics of the original memory and the new memory are summarized and compared.

**Original Memory**

Word Size (bits): 22

Capacity (words): 16236

Capacity (bits): 357192

Width (micron): 1811.89

Height (micron): **1293.58**

Area (square micron): 2343833

**Modified Memory**

Width (micron): the same

Height (micron): **1322**

**Added Logic size:**

Width: 2.14 um (for each column)

Height: 28.4 um

**Ratio** (standard SRAM/modified SRAM) = 1322/1294 = **1.022**

**Over head: 2.19%**

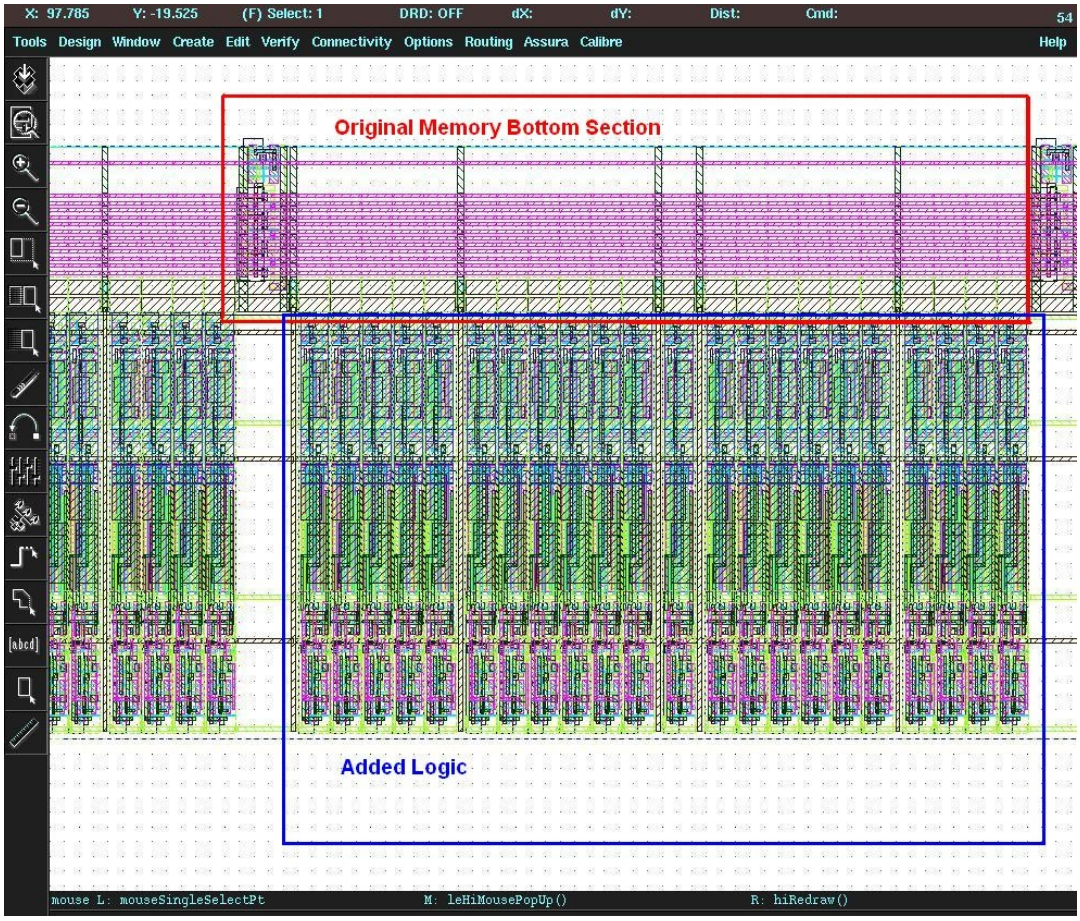



Fig. 6.18 Layout of the modified memory

It should be noted that almost half of the area of the logic is occupied by the

PMOS connected to VDD2 line (W = 8um x 3 fingers) and by the PMOS within the detector stage (W = 4um x 3 fingers). They are both oversized; it is possible to reduce their dimensions saving the 20% of area in the logic cell.

# CONCLUSIONS

The latch-up phenomena have been analyzed to develop an fairly accurate equivalent model compatibles with SPICE simulator. A low-cost SEL mitigation scheme for memories has been proposed and realized in Chartered 130nm technology. The scheme presents high flexibility since it is not related to SRAM characteristics (i.e. size, shape, word length, timing), but only to parameters inherent the technology (i.e. substrate resistance, n-well resistance…). The logic has been implemented within a SRAM generated with a general Memory Compiler Tool. The architecture does not affect the original memory performances, does not increase the power consumption (when the latch-up event occurs, it is only needed to re-charge the VDD line related to one bit column) and slightly impacts the area occupation (area over head < 2.2%).

The modified SRAM it has been realized in order to be implemented within a 3D stack by using the Through Silicon Via Technology furnished by Tezzaron. The project aims to validate various issues concerning the implementation of 3D systems and the use of memories 3D stacked for applications in space environment.

# LIST OF PUBLICATIONS

[1] R. Canegallo, A. Fazzi, L. Ciccarelli, L. Magagni, F. Natali, PierLuigi Rolandi, E. Jung, L. Di Cioccio, R. Guerrieri. “3D Capacitive Interconnections for High Speed Interchip Communication”, Custom Integrated Circuits Conference, 2007. CICC ‘07. IEEE 16-19 Sept. 2007 Page(s): 1-8.

[2] A. Fazzi, R. Canegallo, L. Ciccarelli, L. Magagni, F. Natali, E. Jung, Pierluigi Rolandi, and R. Guerrieri. “3-D Capacitive Interconnections With Mono- and Bi-Directional Capabilities”, IEEE Journal of Solid State Circuits, Vol. 43, no. 1, January 2008.

[3] R. Canegallo, L. Ciccarelli, F. Natali, A. Fazzi, R. Guerrieri, PierLuigi Rolandi. “3D Contactless Communication for IC design”, Integrated Circuit Design and Technology and Tutorial, 2008. ICICDT 2008. IEEE International Conference on 2-4 June 2008 Page(s): 241-244.

[4] A. Fazzi, R. Canegallo, L. Ciccarelli, L. Magagni, F. Natali, E. Jung, Pierluigi Rolandi, and R. Guerrieri. “3-D Capacitive Interconnections With Mono- and Bi-Directional Capabilities”, Solid-State Circuits Conference, 2007. Digest of Technical Papers. IEEE International 11-15 Feb. 2007 Page(s): 356-358.

**PUBLICATIONS**

[5] M. Nicolaidis, K. Torki, F. Natali, K. Belhaddad, D. Alexandrescu. “Implementation and Validation of a Low-Cost Single-Event Latch-up Mitigation Scheme”. IEEE 2009, Workshop on Silicon Error in Logic-System Effects, Stanford University, March 24$^{th}$-25$^{th}$.

# BIBLIOGRAFY PER CHAPTERS

## INTRODUCTION

[1] International Technical Roadmap of Semiconductors, TWG A&P, www.itrs.org.

## CHAPTER 1

[1] W. Rhett Davis, J. Wilson, S. Mick, J. Xu, H. Hua, C. Mineo, A. M. Sule, M. Steer, and Paul D. Franzon, “Demystifying 3D ICs: The Pros and Cons of going vertical”, IEEE Design and Test of Computers, 2005.

[2] R.M. Lea et al., “A 3-D Stacked Chip Packaging Solution for Miniaturized Massively Parallel Processing,” IEEE Trans. Advanced Packaging, vol. 22, no. 3, Aug. 1999, pp. 424-432.

[3] B. Black et al., “3D Processing Technology and Its Impact on iA32 Microprocessors” Proc. IEEE Int. Conf. Computer Design (ICCD 04), IEEE CS Press, 2004, pp. 316-318.

**BIBLIOGRAFY**

[4] Baik-Woo Lee, Jui-Yun Tsai, Hotae Jin, Chong K. Yoon, Rao R. Tummala, “New 3-D Chip Stacking Architectures by Wire-On-Bump and Bump-On-Flex”, IEEE Transaction on Advanced Packaging, Vol. 31, no. 2, May 2008.

[5] N. Miura, H. Ishikuro, K. Nitsu, T. Sakurai, T. Kuroda, “A 0.14 pJ/b Inductive-Coupling Transceiver With Digitally-Controlled Precise Pulse Shaping”, IEEE Journal of Solid State Circuits, vol. 43, no. 1, January 2008.

[6] Said F. Al-sarawi, Derek Abbott, Paul D. Franzon, “A Review of 3-D Packaging Technology” IEEE Transaction on Components, Packaging, and Manufacturing Technology, part B, vol. 21, no. 1, February 1998.

# CHAPTER 2

[1] A. Fazzi, L. Magagni, R. Canegallo, S. Schmitz, and R. Guerrieri, “A 0.14mW/ Gbps high density capacitive interface for 3D system integration”, in CICC2005, September 2005, pp. 101-104.

[2] B. Charlet et al. “Chip-to-Chip Interconnections Based on the Wireless Capacitive Coupling for 3D Integration”, In MAM2006, March 2006.

[3] R.J. Drost et al. “Proximity communication”. In IEEE J. Solid-State Circuits, Volume 39, Issue 9, pp.1529-1535, September 2004.

[4] K. Kanda et al., “1.27Gb/s/pin 3mW/pin wireless super-connect (WSC) interface scheme”. In ISSCC2003, February 2003, pp.186-187.

**BIBLIOGRAFY**

[5] Mari Inoue et al., “Daisy Chain for Power Reduction in Inductive-Coupling CMOS Link”. In Symposium on VLSI Circuits 2006, June 2006.

[6] N. Miura et al., “A 1Tb/s 3W Inductive-Coupling Transceiver for Inter-Chip Clock and Data Link”. In ISSCC2006, February 2006, pp.424-425.

[7] N. Miura et al., “A 195Gb/s 1.2W 3D-Stacked Inductive Inter-Chip Wireless Super-connect with Transmit Power Control Scheme”. In ISSCC2005, February 2005, pp.264-265.

[8] Jian Xu et al., “2.8 Gb/s Inductively Coupled Interconnect for 3-D ICs”. In Symposium on VLSI Circuits 2005, February 2005, pp. 352-355, June 2005.

[9] D. Mizoguchi et al., “A 1.2Gb/s/pin Wireless Super-connect based on Inductive Inter-Chip Signalling (IIS)”. In ISSCC2004, February 2004.

[10] R. Wieland et al., “3D Integration of CMOS transistors with ICV-SLID technology”, In Proc. of 3D Architectures for Semiconductor Integration and Packaging 2005, Technology Venture Forum, June 2005.

[11] H. Gan et al., “Pb-free micro-joints for the next generation micro-systems: the fabrication, assembly and characterization”, Electronic Components and Technology Conference 2006, San Diego, 2006.

[12] T. Etzaki et al., “A 160Gb/s Interface Design Configuration for Multi-chip LSI”, In ISSCC2004, February 2004

[13] L. Luo et al., “3gp/s ac-coupled chip-to-chip communication using low-swing pulse receiver”. In ISSCC2005, pages 522–523, February 2005.

**BIBLIOGRAFY**

[14] A. Fazzi, R. Canegallo, L. Ciccarelli, L. Magagni, F. Natali, E. Jung, P. L. Rolandi, and R. Guerrieri “3D capacitive interconnection with mono- and bi-directional capabilities”. In ISSCC2007, February 2007.

# CHAPTER 3

[1] N. Miura et al., “An 11Gb/s Inductive-Coupling Link with Burst Transmission”, ISSCC Dig. Tech. Papers, pp. 298-614, Feb., 2008.

[2] A. Fazzi, L. Magagni, R. Canegallo, S. Schmitz, and R. Guerrieri, “A 0.14mW/ Gbps high density capacitive interface for 3D system integration”, in CICC2005, September 2005, pp. 101-104.

[3] A. Fazzi, R. Canegallo, L. Ciccarelli, L. Magagni, F. Natali, E. Jung, P. L. Rolandi, and R. Guerrieri “3D capacitive interconnection with mono- and bi-directional capabilities”. In ISSCC2007, February 2007.

[4] W. Yang, B. Afshar, T-Y. Cheng, V. Gaudet, A. M. Niknejad. “A 2.5 mW Inductorless VGA with Dual Feedback DC-Offset Correction in 90nm CMOS technology”. Proc. of 2008 RFIC Symposium, pp 91-94, April. 2008.

[5] A. Thanachayanont and P. Naktongkul, “Low-voltage wideband compact CMOS variable gain amplifier”, ELECTRONICS LETTERS 20th January 2005 vol. 41 no. 2.

[6] Mini-circuits power splitters/combiners Data Sheet

[7] Analog Devices AD9631 Data Sheet

**BIBLIOGRAFY**

# CHAPTER 4

[1] M. Puech, N. Launay, N. Arnal, P. Godinat, JM. Gruffat, “A novel plasma release process and super high aspect ratio process using ICP etching for MEMS”, MEMS/NEMS seminar December 3rd 2003.

[2] Edvard Kälvesten, Thorbjörn Ebefors, Niklas Svedin, Pelle Rangsten, Tommy Schönberg, “Elektriska anslutningar i substrat / Electrical connections in substrates”, Swedish Patent No.:0300784. Several parallel International patent application publications (in English) e.g. WO2004084300 published 30.09.2004.

[3] H. Peter Hofstee, “Future Microprocessors and Off-Chip SOP Interconnect” IEEE Transaction on advanced packaging, vol.27, no. 2, May 2004.

[4] J. Bautista, “Tera-Scale Computing and Interconnect Challenges”. Intel-Microprocessor Research.

[5] www.tezzaron.com (logic on memory applications)

# CHAPTER 5

[1] B. L. Gregory, “CMOS Latch-up and Prevention”, Sandia Laboratories, Alburquerque, NM, June 1975.

[2] R. G. Useinov, “Analytical Model of Radiation Induced or Single Event Latch up in CMOS Integrated Circuits”, IEEE Transaction of nuclear science, vol. 53, no. 4, August 2006.

**BIBLIOGRAFY**

[3] J. F. Ziegler, J. P. Biersack, and U. Littmark, “The Stopping and Range of Ions in Matter”, New York: Pergamon, 1985.

[4] W. J. Stapor and P. T. McDonald, “Practical approach to ion track energy distribution” J. Appl**.** Phys., vol. 64, no. 9, pp. 443W34, 1988.

[5] W. J. Stapor, P. T. McDonald, A. B. Campbell, **A.** R. Knudson, and B.G. Glagola, “Charge collection in silicon for ions of different energy but same linear energy transfer (LET),” IEEE Trans. Nucl. Sci., vol. 35, no. 6, pp. 1585-1590, 1988.

[6] R. C. Martin, N. M. Ghoniem, Y. Song, and J. S. Cable, “The size effect of ion charge track on single event multiple-bit upset,” IEEE Trans. Nucl. Sci., vol. NS-34, no. 6, pp. 1305-1309, 1987.

[7] T. L. Criswell, D. L. Oberg, J. L. Wert, P. R. Measel, and W. E. Wilson, “Measurement of SEU thresholds and cross sections at fixed incidence angles,” IEEE Trans. Nucl. Sci., vol. NS-34, no. 6, pp. 131G1321, 1987.

[8] G. J. Hu and R. H. Bruce, “A CMOS structure with high latch-up holding voltage,” IEEE Electron Dev. Lett., vol. EDL-5, pp. 211-214, 1984.

[9] R. L. Holpelhorn and B. D. Shafer, “Radiation induced latch up modelling of CMOS ICs”, *IEEE Trans. Electron Dev.*, vol. 34, no. 6, pp. 1396–1401, 1987.

[10] P. E. Dodd, M. R. Shaneyfelt, E. Fuller, J. C. Pickel, F. W. Sexton, and P. S. Winokur, “Impact of substrate thickness on single-event effects in integrated circuits”, *IEEE Trans. Nucl. Sci.*, vol. 48, no. 6, pp. 1865–1871, 2001.

[11] G. Bruguier and J-M. Palau, “Single Particle-Induced Latch-up”, IEEE Transactions on Nuclear Science, vol. 43, no. 2, April 1996.

**BIBLIOGRAFY**

# CHAPTER 6

[1] D.B. Estreich, A. Ochoa, and R.W. Dutton, “An Analysis of Latch-up Prevention in CMOS IC’s Using Epitaxial Buried Layer Process”, International Electron Device Meeting, 1978, pp. 76-84.

[2] R.R. Troutman, “Latch up in CMOS Technology: The Problem and its Cure”, Kluwer Publications, Boston, MA, 1986.

[3] M. Nicolaidis, “Circuit Intégré Protégé Contre les Court-Circuits et les Erreurs de Fonctionnement Suite au Passage d’une Radiation Ionisante”, French patent application, Mai 2005.

[4] B. Gilles, M. Nicolaidis, Ch. Papachristou ”An Efficient BICS Design for SEUs Detection and Correction in Modern Memories”, Design Automation and Test in Europe (DATE), March 2005, Munich.

[5] M. Nicolaidis, K. Torki, F. Natali, K. Belhaddad, D. Alexandrescu. “Implementation and Validation of a Low-Cost Single-Event Latch-up Mitigation Scheme”. IEEE 2009, Workshop on Silicon Error in Logic-System Effects, Stanford University, March 24$^{th}$-25$^{th}$.