POLITECNICO DI TORINO

Master of Science in Electronic Engineering

Master's Thesis

Development Of An Innovative n-LD-MOSFET in BCD10 Technology

Supervisor:
Prof. Gianluca Piccinini
Ing. Giuseppe Croce

Candidate:
Danilo Covello

2020, April

Acknowledgments

I wish to express my sincere gratitude to Andrea Mario Torti and Paolo Gattari who convincingly guided and encouraged me to be professional and to continue investigating when the road became tough. Without their continuous help, the goal of this project would not have been achieved.

Moreover, I want to communicate my appreciation for the physical and technical contribution of STMicroelectronics that supported and funded this project.

Table of contents

List of figures

Abstract

Nowadays, more and more electronic systems and applications fields, ranging from the automotive sector, energy management and distribution to IT and consumer industry, require devices with the ability to drive high current loads along with the ability to sustain a high voltage drop, when they are *ON* and *OFF*. Furthermore, the need for a very small power dissipation is becoming rapidly a crucial point in the design of new transistors or complex systems.[1] As a consequence, it is necessary firstly, to increase the power transfer efficiency and, secondly, to limit the heat generation. Integrated power transistors are born with the idea of combining all these requirements to have the best trade-off among high current, low *ON*-resistance, wide operating frequency range, low static consumption, good thermal stability, high reliability, and small size.[2]

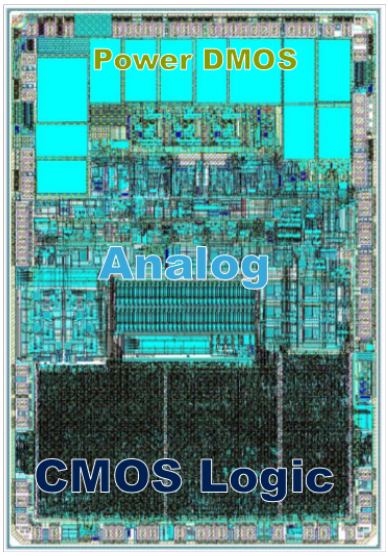
In the last 30 years, the market demands showed an unstopped growth due to the increase in the number of interested fields, produced units, and of new complex and powerful applications with higher power demand.[3] Consequently, the research has investigated new roads. New substrate materials[1] have been studied like GaN (Gallium Nitride) for optoelectronic, high power and/or high-frequency applications, SiC (Silicon Carbide) for high power and/or high-temperature applications, GaAs (Gallium Arsenide) for microwave applications, and many other III-V compounds. Materials that have a higher band-gap to provide high voltage breakdown, lower *ON*-resistance and, in the end, a much lower power dissipation. At the same time, new device architectures have been attempted and structures like *Vertical Diffused MOSFET (VD-MOSFET)* or *Isolated Gate Bipolar Transistor (IGBT)* have been introduced in business and optimized massively.[4][5]

Similar efforts have been applied to find solutions to integrate power MOSFETs into a Smart Power platform such as *BCD* (Bipolar-CMOS-DMOS) technology. The integration of power MOSFETs in advanced technology nodes was driven by the needs to integrate denser digital cores for signal and data processing. It had to face

and overcome many challenges concerning the limitations coming from the scaling of some critical dimensions (oxide thickness, spacer dimension). Contrary to the digital section, the power section does not follow the scaling of the operating voltages and its integration in advanced technology nodes could penalize the performance of the existing solutions. Therefore, this work will be dedicated to the study of a new *Lateral Drift MOSFET* architecture (LD-MOSFET) intended for low-voltage applications (up to 30 V) aimed to overcome these limitations. With these objectives, an all-in-active *LD-MOSFET* will be combined with metal field plate technology to obtain new devices with higher possibilities to be technologically and economically competitive. Studies, analysis, simulations, and experimental measurements will be reported and detailed in order to characterize this innovative solution.

Introduction

Today, power devices are used in almost all electronic fields and applications. One can easily mention thousands of examples such as automotive systems, battery chargers or power management units. A power device, however, is not only characterized by power functions and power components because, generally, it must also perform some digital processing or require accurate digital control that may need a CMOS section or even a μP. Also, a device may require some analogical functions, hence, again, a bipolar section is compulsory.[2][6]



Let's consider a simple and clarifying example: an Hard Disk Drive (HDD) system.[7] It requires all three previous mentioned sections:

1. A power section for spindle and voice coil driving.

2. An analogical section including some high bandwidth and low noise components as the pre-amplifiers and some actuators to improve precision in the head positioning.

3. A logic section, involving a μP and a ROM memory, devoted to perform read-write operations and to control head positioning.

The three sections are designed following different development and optimization criteria, but they must work together to obtain the correct system functionality, i.e. every single section is useless if left stand alone.

To get the best performance and the lowest self-heating due to dissipation, the different sections must be put as closest as possible. To meet this constraint, SiP

(System in Package) or MCM (Multi-Chip Module) are largely used for many applications but they are not the optimal solutions. Surely, it is possible to achieve better results if the different parts can be integrated into the same chip.

Following this last guideline, in mid-eighties, the BCD technology was born. It became possible the realization of the so-called *Systems On Silicon* (SoS)[7], i.e. complete systems, made of different electronic components, unified in a single die. The name BCD summarizes the intention to integrate into the same chip, with unique process flow, all necessary devices: *Bipolar* transistor to realize complex analogical functions, *CMOS* transistor to have fast and good digital switches, and *DMOS* transistor for the power section. Figure[1] 1 summarizes the purpose proposed by the BCD technology, previously described.

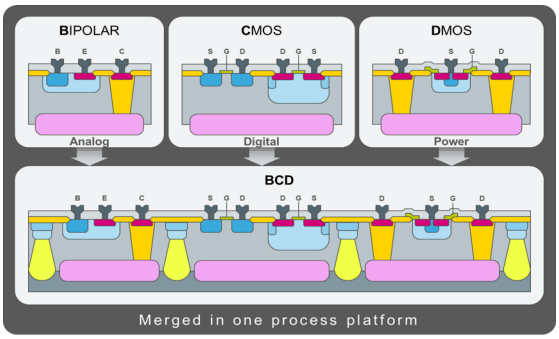


Figure 1: Integration evolution of BCD.

BCD technology was an innovation under many points of view. It removes the necessity to interconnect different dies at the package level with external lines. This reduces the resistance and the capacity of the lines that can be translated into enhanced general performance. Moreover, the suppression of external lines reduces ElectroMagnetic Irradiations (EMI), lowering the assembly cost and improves the overall reliability.[7]

The next sections of this work are organized as follows. Chapter one, will complete the introduction to the BCD technology, reporting the main steps of the process

[1]Image taken from [8], BCD10 description section.

flow together with the main integration challenges and possible insulation schemes. In chapter two, it will be described briefly the most important power architectures for which classification will be provided; next, we focus on the n-LD-MOSFET studying its architecture and its main characteristics. The last paragraph, instead, will be devoted to the description of the TCAD tools used for simulations and each step required to perform them. Then, in chapter three, we will analyze a reference device to describe the relationships between technology parameters and electrical ones. For the latter ones, some analytical models based on the semiconductor device theory will be also proposed and their results compared with the simulation and experimental ones. Then, in chapter four, we will conclude the description of the LD-MOSFET discussing the ReSURF effect. It is the most important effect for this kind of structure that makes it possible to reach so high breakdown voltages. We will describe also the field plate technology that is the architectural change that we introduce to overcome the limitations to the overall performances when designing low-voltage devices with the all-in-active architecture. Finally, the last chapter will be dedicated to the description of the field plate integration and the experimental and simulative analyses to fully characterize this solution.

Chapter 1

BCD Technology

The BCD technology from its introduction by the STMicroelectronics in the mid-eighties continues to evolve day by day. On one hand, it follows the scaling predicted by Moore's law, particularly for CMOS technology. During the last thirty-five years, the technology node scaled down from 4 µm of the first BCD generation (BCD1) to 90 nm of the last generation (BCD10). On the other hand, due to the increasing number of targeted applications and fields, it increases the number of different devices and voltage classes that are possible to integrate. To date, when asking for a new BCD product, it is possible to design a complex system made of CMOS, bipolar transistors, DMOS optimized for low-voltage classes (< 40 V), intermediate voltage classes and high voltage classes (> 600 V), passive components such as resistors, capacitors and transformers, and non-volatile memories. Due to this huge number of different functionalities that require different performances, the BCD roadmap was split according to three different integration objectives: High-Voltage, High-Density, and High-Power (figure 1.1).[7]

High-Voltage BCD concerns all the products where the requirements about leakage and parasitic capacitances are critical. Generally, to satisfy the specifications about them, it is used an SOI substrate instead of a bulk one. High Power-BCD concerns all the devices that must bear high current density and hence they put less stringent requirements about the area, for example. Finally, High-Density BCD regards VLSI and should be compatible with advanced CMOS.

One of the main challenges of BCD technologies is the definition of a process flow that allows the integrability of a wide range of different devices ensuring the electrical performances of all of them. However, often, what is desirable to optimize for a certain device is not required or might degrade another one. Let's consider, for

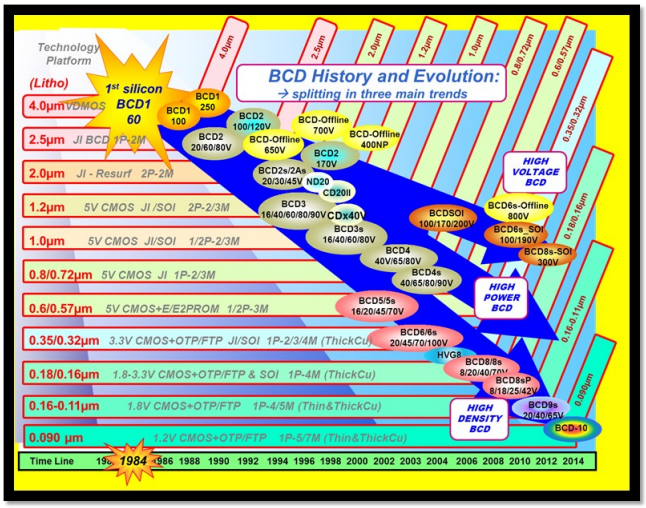


Figure 1.1: BCD roadmaps.

example, one of the most common and most difficult situations the BCD must face: the integration of a CMOS and DMOS transistors into the same die. This situation highlights the first conflicts:

- The Gate Oxide (*GATOX*) is realized through advanced thermal oxidations to obtain silicon-oxide interfaces with the best electrical properties. However, as we will see in the next chapter, the DMOS and some devices intended for analog functions (3.3 V or higher) require a thick oxide while the digital CMOS require thin oxide (1.8 V for 180 nm or 1.2 V for 90 nm). The need for having more than one *GATOX* thickness is an issue that the BCD flow must face and solve.

- The DMOS uses low doping levels and large areas to sustain high voltage and supply large currents. Moreover, some geometrical dimensions are defined by the diffusions induced by the many thermal steps. The DMOS requires high thermal budgets. The CMOS, instead, uses high doping levels and small areas, it requires so small thermal budgets. The compatibility of the thermal budgets is hence another issue that the BCD process flow must solve.

The development of a so complex flow must face many of these issues of incompatibility, some of these of utmost importance, others of a lesser one. The solution

proposed by the BCD flow puts together the optimization of each single technology step and the study of the order with which they are executed. So, returning to the previous examples, the first issue concerning the different *GATOX* thicknesses is solved by dividing the oxidation into two or more steps interspersed by a masking step and an etch. The second issue, linked to the different thermal budgets, instead, can be solved by properly ordering the implantations of the well and the annealing steps. The order with which we perform those implantations becomes hence crucial: the wells must be implanted starting from the ones that require the highest thermal budget to the ones that require the lowest thermal budget so that CMOS wells do not see the high and potentially destructive thermal budgets needed by the high-voltage wells. Since not all issues can be solved without conceding something, at the end the BCD flow is tuned to reach the best trade-off and so the best final performances between all structures that are competing.

1.1 Isolation Schemes

Besides the issues related to the development and optimization of the process flow, another challenge of utmost importance concerns the development of techniques to properly isolate different sections of the chip or simply neighbour transistors[9] Issues regarding the isolation of similar devices have existed in every technology since their birth and, along the years, solutions have been provided and successfully integrated, but the need for robust and performing isolation between devices operating with so different operating conditions was completely new. This assumes particular importance when sensible devices such as bipolar transistors intended for accurate analogue functions or small CMOS for fast digital computation, are integrated near power devices demanding high currents.[10] In the following paragraph, a short description of the most used isolation schemes is provided.

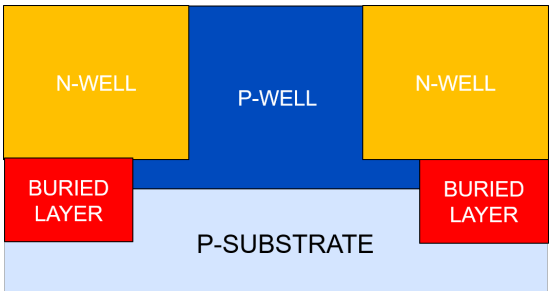


Figure 1.2: Junction Isolation scheme.

The Junction Isolation (JI) provides isolation through a reverse-biased p-n junction. [10] Its main advantage is the very low cost since it requires only a small number of additive implantations. But, this technique is characterized by high leakage and large areas demanded. Furthermore, the p-doped regions intended for isolating different n-wells become the base of many parasitic bipolar transistors that, under particular operating conditions, can trigger on and connect devices that should have been isolated. Nevertheless, the JI is often used to provide lateral isolation, especially when the performance requirements are not critical, while it is almost the only technique for vertical isolation[2]. An exception regards the SOI (Silicon On Insulator) wafer because this latter one can be provided by the Buried Oxide (BOX). Figure 1.2 shows two active n-wells that have been isolated exploiting the

just described JI scheme. The lateral isolation is provided by a p-well shorted to the substrate.

Deep Trench Isolation (DTI) uses, instead, a deep trench to provide lateral isolation.[11] The trench is dug into the silicon until the substrate, then the sidewalls are oxidized. Finally, the trench is filled with oxide or with highly p-doped polySilicon to realize a substrate contact. The advantages of this solution are manyfold: less area, almost absent leakage, no lateral parasitic NPNs, and immunity to latch-up.[12] Moreover, the allocation of a substrate contact ensures immunity from the cross-talk between different wells. The only disadvantage is the increment of the fabrication cost, hence, as for the SOI wafers, it is used only for high voltages or when the performances are critical. Figure 1.3 shows again two n-wells that are now isolated exploiting the DTI module. Particularly, in the picture, the deep trench is filled with p-doped polysilicon. The figure on the left side uses a buried layer to provide vertical isolation according to the JI scheme while the picture on the right side, being integrated in a SOI wafer, can use the BOX. [13]

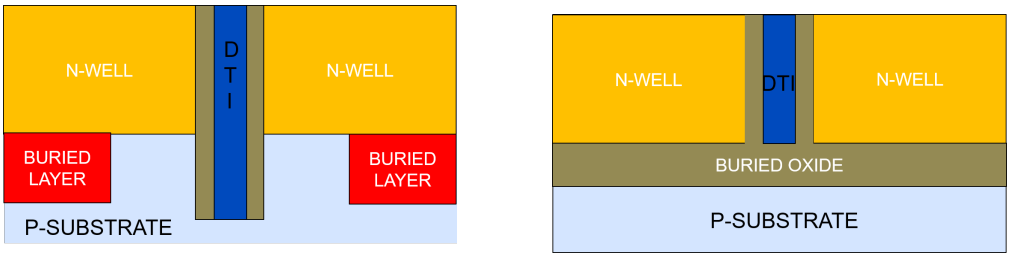


Figure 1.3: DTI scheme and BOX scheme.

The lateral oxidation is taken from the CMOS flow and aims to avoid the current flows between neighbour active areas of not critical devices by an oxide barrier. In ancient time, the barrier was created oxidizing locally the silicon (LOCOS). The wafer had to be masked with a patterned silicon nitride before starting thermal oxidation until to obtain a thick layer of oxide. For advanced technology nodes, the loss of area due to the smooth transition from the active area to the oxide barrier becomes soon no longer negligible, therefore, the LOCOS approach was substituted with a different one called Shallow Trench Isolation (STI). It allows realizing deeper

barriers and almost vertical sidewalls, thus reducing considerably the size of the isolation regions. The realization of the STI is slightly more complex than the one of the LOCOS, it requires indeed more steps, hence is more expensive. Firstly, the substrate must be cover by a patterned oxide-nitride mask (Hard Mask HM). The trench is so dug by a chemical etch and the walls are then oxidated to enhance the adhesion for the subsequent oxide deposition that fills the trench. Finally, a Chemical Mechanical Polishing (CMP) is performed to planarize the surface. Figure 1.4 shows the lateral isolation achieved by an STI module.

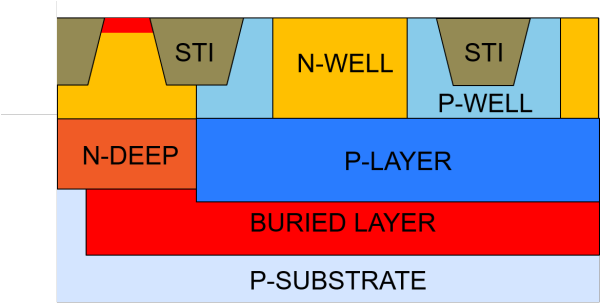


Figure 1.4: Lateral Isolation achieved by STI module.

1.2 BCD Process Flow

The BCD process flow[6] is based on the integration of a DMOS flow into CMOS flow. In this work, the bulk substrate is a thick highly p-doped silicon wafer on which a thin slightly p-doped layer is grown with epitaxy. The doping level of the substrate is chosen, not too low, to have a small substrate resistance, and not too high to maintain the crystalline form and good electrical properties. Indeed, when currents are injected into the substrate, the small resistance ensures fewer noise and higher electrical stability since the parasitic bipolar transistors defined by the substrate and the isolation rings are triggered on by higher currents.

A description of the process flow, divided for simplicity and clearness in several steps, will follow. Moreover, the description of the various steps will be escorted by many pictures that show what happens to the silicon substrate each time. Particularly, for this example, the cross-sections will allow following the integration of an LD-MOSFET. These cross-sections, as any other one, was made using the process simulator, but the preparation and the execution of the simulations, as well as the extraction of their results, will be described in chapter 2.2.

1. Implantation of the Buried Layers (BL).
   The BL is a highly n-doped region that is created by implanting antimony (Sb) ions very superficially. Figure 1.5 shows the result of the process simulation after its implantation and the subsequent annealing steps. In the picture, the hot colours indicate a n-region while the cold ones a p-region. At the very beginning of the BCD technology, the BL was the drain of the vertical power structures but it lost this function with the introduction of the lateral architectures in the mid-nineties. Today hence, it is used to ensure vertical isolation according to the JI scheme. The BL becomes hence the collector of parasitic NPNs and avoids that the current spreads toward or from the substrate when the bipolar transistors are triggered.

2. Growing of an epitaxial layer.
   A new layer of slightis epitaxially grown above the BL to host the active areas. Figure 1.6 shows the resulting cross-section of this step: the upper well is now

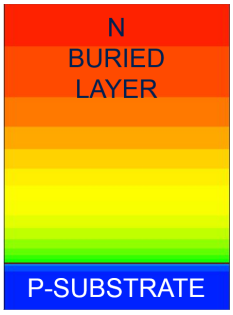


Figure 1.5: Cross-section after the implantation of the buried layer.

clearly visible above the BL and the substrate. The design of this layer (EPI) is quite critical starting from the choice of the type of doping ions. As it is shown in the cross-section, for low-voltage devices, we grow up a p-doped layer to obtain better CMOS compatibility. Its thickness and resistivity are tuned to ensure that the vertical breakdown voltage guarantees the maximum breakdown voltage target of the technology for n-DMOS.[9] The higher are the operating voltages, the thicker and less doped the layer should be and vice-versa. For high-voltages devices ($> 100\,\mathrm{V}$), the n-type of doping often substitutes the p-type since the BL is automatically connected and becomes easier the integration of deeper EPI.

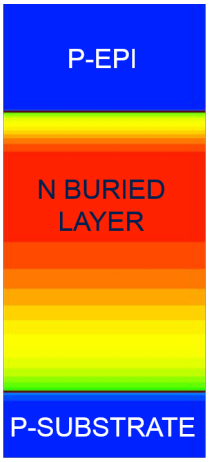


Figure 1.6: Cross-section after the epitaxial growth.

3. Realization of the Deep Trench Isolation (DTI).

As explained in the previous paragraph, DTIs are realized to provide lateral isolation between different die sections where performances and noises rejection are critical.

4. Definition of the Active Areas (AA).
   The previous steps are all dedicated to the integration and isolation of power devices, pure CMOS integration can safely skip them. This step, instead, is taken directly from the CMOS flow and it is used for both digital and power sections. The active areas define the active regions of an integrated circuit. According to the technological node and the specific design rules, the active area can be intended for a single device or multiple ones. In any case, the current flows between different AA must be carefully avoided. With this purpose, this technological step aims to realize the lateral isolation of the AA. It is ensured by oxide regions that can be realized locally oxidizing the silicon (LOCOS) or adopting the Shallow Trench Isolation (STI). Figure 1.7 shows only the upper part of the previous structure after the definition of the active area.

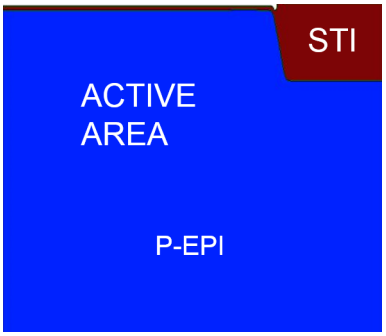


Figure 1.7: Cross-section after the AA definition.

5. Implantation of high-voltage (HV) wells.
   We implant the wells that need the greatest thermal budget. Particularly, they are the isolation wells, the wells of the LD-MOSFET, and the body of the high voltage n-MOS and p-MOS. After the implantations, the substrate does the required thermal budget into a furnace annealing.

6. Implantation of low-voltage (LV) wells.

We implant the wells that need the lowest thermal budget, i.e. the CMOS wells. After the implantations, the substrate does the required thermal budget by Rapid Thermal Annealing (RTA). In figure 1.8, it is possible to note the result after that all the wells of the LD-MOSFET were implanted.

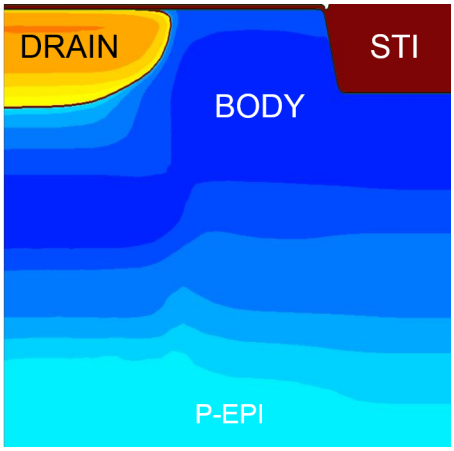


Figure 1.8: Cross-section after the implantation of the high voltage wells.

7. Deposition of the gate oxide (GATOX).
   We grow the GATOX of all MOSFET transistor through the In-Situ Steam Generation (ISSG) process. Firstly, it is grown the thickest oxide for the high-voltage devices. It is then masked and etched away where it is not necessary. Finally, it is grown also the thin oxide for the CMOS, partially growing also in the high voltage region to achieve the final thickness.

8. Definition of the gate electrode.
   We depose a thick polySilicon layer that, after being doped, is patterned (See figure 1.9). At this point, it is possible to insert other steps to realize multiple gate to integrate Non-Volatile Memories (NVM) such as EEPROM.

9. Realization of the spacers.
   The spacers are realized by the succession of an oxide deposition, a nitride deposition and a chemical etch. Both DMOS and CMOS use the spacers inherited by the CMOS flow. After the formation of the spacers, the LDD (Lightly-Doped Drain) and the pockets are implanted. As happens for the implantation of the wells, the HV-LDD are implanted first and the substrate

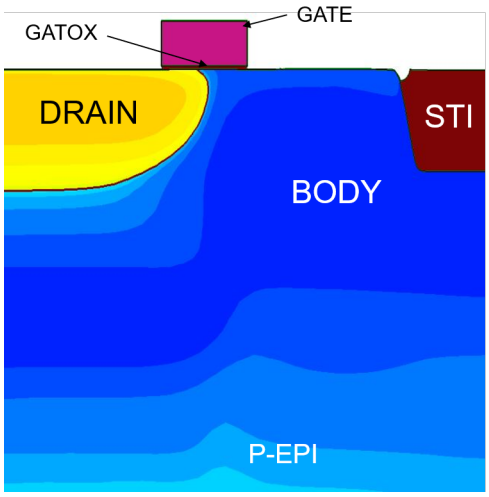


Figure 1.9: Cross-section after definition of the gate.

does the required RTA before the formation of the CMOS spacers and the implantation of the LV-LDD so that these latter do not see the thermal budget of the former (see figure 1.10).

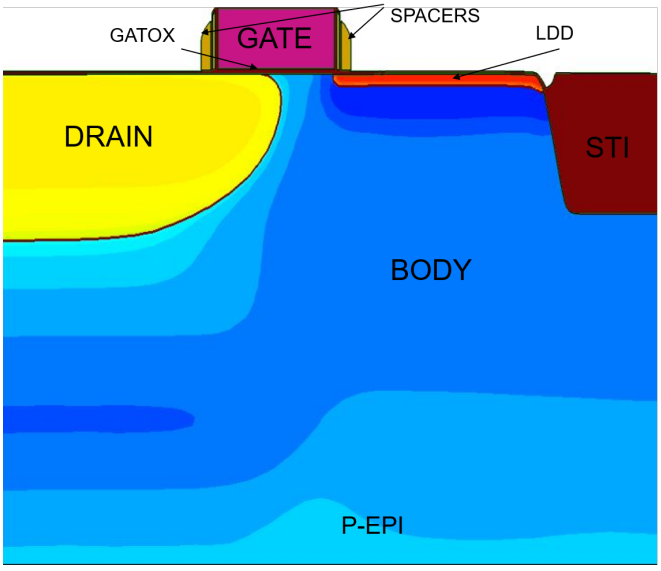


Figure 1.10: Cross-section after the spacer formation and LDD implantation.

10. Definition of the highly-doped regions.
    High doses of boron or arsenic are implanted where the n+/p+ regions must

be created.

11. Realization of the silicide.
    A silicide layer must be formed where the contacts should land to reduce the contribution of the metal-semiconductor junction to the overall resistance. Since the silicide layer is a metal-like, we manage to change the semiconductor-metal junction into a metal-metal junction. The silicon surface is so masked by a patterned oxide/nitride mask called SiPROT, to protect some regions from the silicide formation. Metal like cobalt is then deposed and a new RTP is performed to allow the reaction of the metal ions with the silicon. In figure 1.11, the cross-section up to this point is shown, the simulator does not simulate or emulate the formation of the silicide, so, there are no clear indication of the silicide layers. Anyway, as described before, any region that is not cover by the SiPROT layer was subjected to the silicidation process.

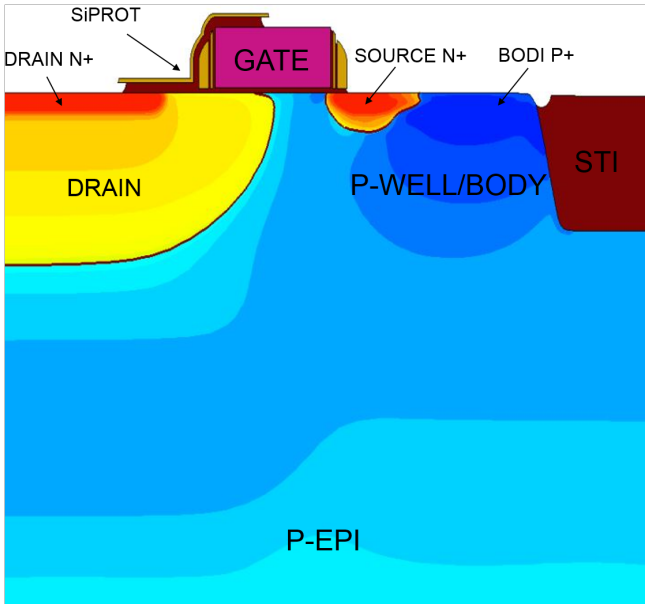


Figure 1.11: Cross-section after the n+/p+ implantation, the SiPROT definition and the silicide formation.

12. Processing of the Back End Of Line (BEOL).
    The Pre-Metal Dielectric (PMD) is deposed and the contact trenches patterned and etched. Then, they are filled with tungsten (W) immediately after

the sputtering of the barrier, finally, the planarity is achieved through a tungsten CMP. Figure 1.12 shows the cross-section of the LD-MOSFET after the definition of the contacts. The remain steps of the BEOL allow the integration of the metal interconnections. In the most advanced technology nodes, the metallizations are made with copper damascene that substitutes the aluminium for its better electrical properties: lower resistance, high robustness to the electromigration and thermic stress. The BCD can not follow only the BEOL design rules of the CMOS flow but they must take care of the need of high voltage and high current of the power section, particularly the ones concerning the thickness of the Inter-Metal Dielectric (IMD) and the metal lines themself. Generally, there are some level of thin metal lines intended for fast digital operations and a thick metal line specially designed for power operations.

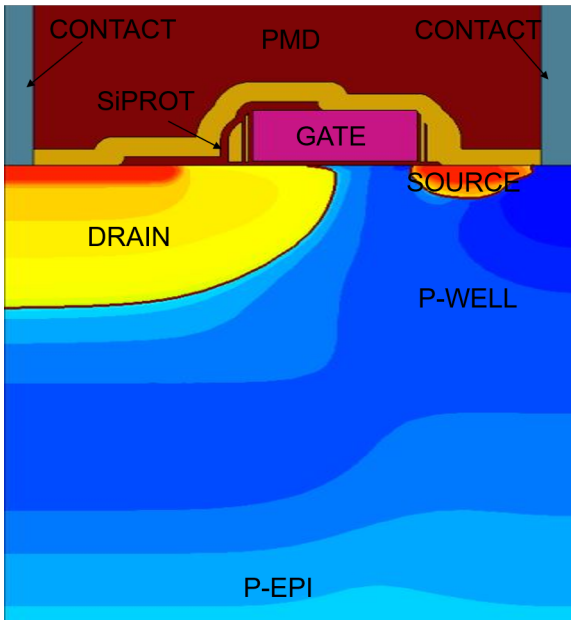


Figure 1.12: Cross-section at the end of the flow.

Chapter 2

Power Architectures

Power devices have been studied since the birth of the first bipolar transistor in the late '40s and, over the years, they have become even more powerful, reliable and capable to sustain very high voltage and current. Today, there are, in fact, power devices that can bear thousands of Volts and thousand of Amperes. Just as the intense study of digital transistors was driven by the massive development of the microprocessors, the development of power transistors was guided by the need for active devices for new and powerful applications that require higher voltage or current capability and less loss of power. Most of these transistors are intended to be a switch between two different systems or between a system and its power or ground lines; they must therefore ensure the maximum efficiency for the energy transfer to the load. Another frequent destination is for protection of a sensible section of the circuit against, for example, Electro-Static Discharge (ESD), voltage or current spikes or short of the load. Finally, they are also often used for power converters and rectifiers, I/O interfaces and so on.

As mentioned, power architectures were born together with the introduction of the first bipolar transistor and made a big step ahead towards the end of the seventies with the introduction of the first power MOSFETs. The modern power architectures can be classified as bipolar-based, MOS-based or Bipolar-MOS-hybrid according to the family they derive from. The thyristors in all their variants such as, for example, the Gate Turn-Off thyristors (GTO) or the Gate-Assisted Turn-off thyristors (GATT) and the Darlington configuration belong to the bipolar-based family. In a nutshell, the thyristor is a diode where the direct conduction is possible only applying a proper signal to the control electrode called Gate. Also the Darlington configuration belongs to the same family. It is realized with two bipolar transistors connected in a way that the current amplification factor is the product

of the two single ones. So, they have the emitters shorted together and the base current of the second stage is driven by the output current of the first stage. Generally, bipolar-based devices can reach very large voltage and current ratings and have smaller output impedance, however, they are also very slow, particularly for the on-off transition, and they are characterized by high dissipation due to the high drive current and low input impedance.[14]

We needed to wait for the introduction of new architectures based on the MOSFET one to overcome these two negative aspects. In the year 1969, it was introduced the V-groove MOSFET (V-MOSFET). Approximately 10 years later, it was substituted by the Vertical Diffused MOSFET (VD-MOSFET) that was born as an evolution of the previous structure. In the same years, also the Lateral Diffused MOSFET (LD-MOSFET) was developed. The main advantages of the architecture MOS-based are faster transition, larger Safe Operating Area (SOA), much higher current gain, high input impedance, and null drive current.

Finally, researchers studied the possibility to combine devices of both families to take the best characteristics from each one. As a result, we cite only the most important structure: the Isolated Gate Bipolar Transistor (IGBT). It is similar to the Darlington with the difference that the input transistor is a power MOSFET. In this way, the IGBT combines all the advantages of the bipolar transistors, such as a low ON resistance, and all the advantages of the MOSFETs, such as a very high input impedance.[14]

2.1 LD-MOSFET Architecture

At the very beginning of BCD technology, vertical architectures were used to realize integrated power devices but, as early as in the nineties, the lateral architectures substituted the vertical ones. On the one hand, the former are realized with less technological steps and, therefore, they are cheaper. On the other hand, since the current still flows on the surface, they are easily compatible with the planar VLSI (Very Large Scale Integration) processes. In the vertical structures, instead, the current flows vertically toward the deep drain, so, to ensure the same VLSI compatibility, there is the need to create highly-doped and deep plugs to bring the current from the drain region to the superficial metal lines. For these reasons, we will dedicate this work to the study of a lateral power structure, besides they have been historically preferred among the MOS-based architectures in the BCD smart power integration.

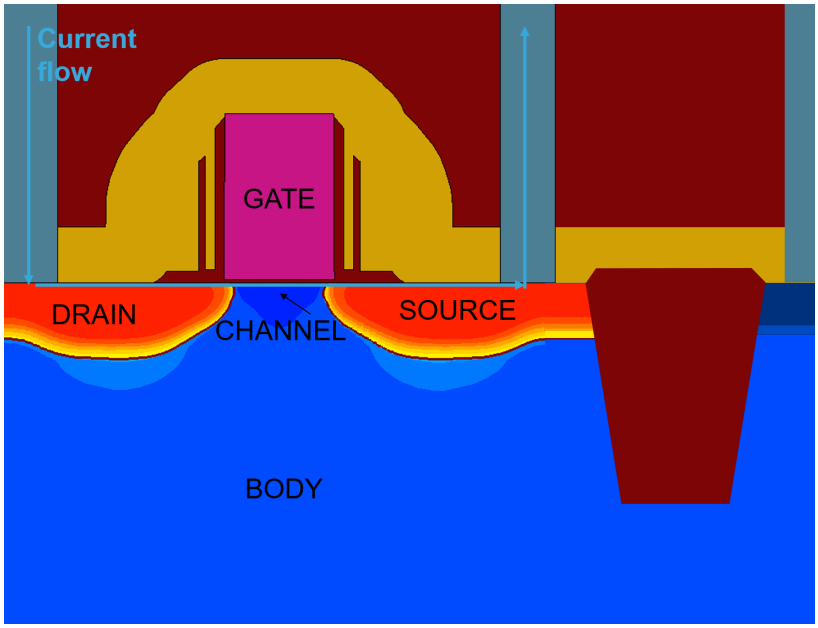


Figure 2.1: Cross-section of a standard digital n-MOSFET.

To better introduce the n-LD-MOSFET structure, we will start from the well-known n-MOS architecture, highlighting the issues which appear when we want to move to power applications and that forced us to move to different architectures.

Then, we will explain the architectural solutions that, applied to the MOSFET structure, allow to overcome the afore-described issues and define a power lateral architecture. Finally, we will provide a description of several possible lateral architectures used to target different voltage classes. Figure 2.1 shows the cross-section of a n-MOSFET with all the meaningful regions highlighted. Just for coherence with the subsequent LD-MOSFET cross-sections where the drain and source regions will be no longer symmetric, the drain is placed on the left side of the structure and the source on the right one. In the cross-section, there is shown also the body strap separated from the source by an oxide trench.

The idea behind the MOS structure is surely well-known: biasing the *gate* electrode it is possible to switch on or off the current flow between the *drain* and the *source* regions that act as low resistive electron tanks. When the *gate* is grounded, there is no electrical connection between those regions and no current can flow. On the contrary, when the *gate* is properly polarized, it forces the bending of the silicon energy levels at the interface with the oxide allowing the creation of a superficial layer of electrons. These electrons are so confined in the *channel* region that is defined as the region between the electrostatic barriers made by the GATOX and the bent conduction band edge. In this condition, it exists an electrical path between the *drain* and the *source* through the channel and therefore, imposing a voltage drop between those regions, it is possible to force a current flow. In conclusion, a MOSFET system is a device in which the *gate* electrode can control the current that flows between the *drain* and the *source* terminals. Particularly, what we have described up to now and what we are going to use for the rest of this work is the so-called 'enhanced MOSFET' that must not be confused with the 'depletion MOSFET'. In the former, in fact, the *gate* can switch on or off the current flow creating or destroying the channel, while in the latter, it can only control the conductivity of the channel, i.e. it can only modulate the current intensity. Finally, the *body* of the component can be used to have a second control over the channel or, equivalently, over the threshold voltage if it is accessible from the outside, i.e. it has an independent contact.

This architecture is fast, cheap and works well for analog applications where it

acts as an amplifier and for digital applications which drive its development, optimization, and scaling. However, this architecture became soon incompatible with the demand of power applications. The scaling of digital devices is followed also by the scaling of the used voltages while the devices intended for power applications must maintain the same voltage capability since the operating voltages do not change over time. This has two important consequences: firstly, digital and power devices start to follow different scaling policies and optimization criteria, and secondly, the architectures intended for power applications start to diversify to meet the aforementioned requirements[1]. The necessity to satisfy at the same time the constraints imposed by the scaling policies and the need to bear the same voltage stresses is the main reason that forced the introduction of different architectures for power devices starting from the '70s.

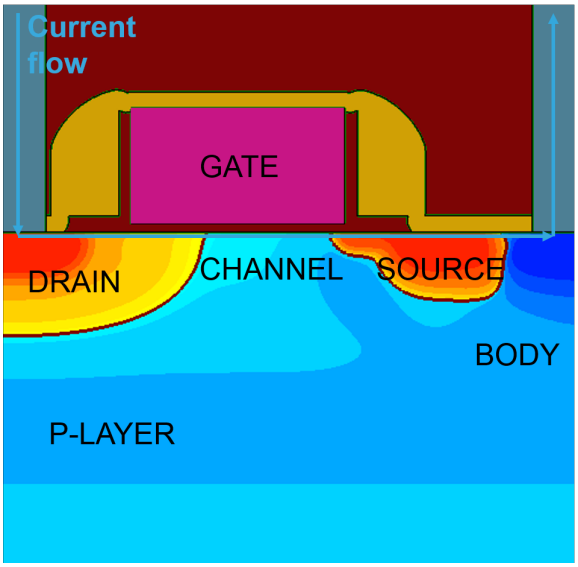


Figure 2.2: "C-MOS based" LD-MOSFET cross-section.

Let's consider an *OFF* working condition in which the drain electrode is high while the other electrodes are grounded. In a MOS structure like the one drawn in figure 2.1, the entire voltage drop is localized only inside the depletion region of the drain-body junction. Moreover, the gate electrode forces inside the silicon a perpendicular field that locally increases the absolute electric field. This leads to

the increment of the drain leakage due to the GIDL (Gate Induced Drain Leakage) phenomenon, to the increment of the maximum value reached by the electric field and to the related reduction of the breakdown voltage. To partially overcome this problem and improve the voltage capability a dedicated mask is used to realize the drain, that becomes larger and less doped. Figure 2.2 shows the cross section of this power architecture, typically used to target low-voltage application: the large and low doped drain region allows to improve the breakdown voltage with relatively low worsening of the ON-resistance and the gate-drain capacitance. Nevertheless, voltage capability of this kind of architecture (CMOS-based) scales together with the technological node, because of it this architecture is limited by the gate oxide thickness and the spacer dimension, i.e. the distance between the gate and the drain highly-doped region, becoming useless in advanced technology node.

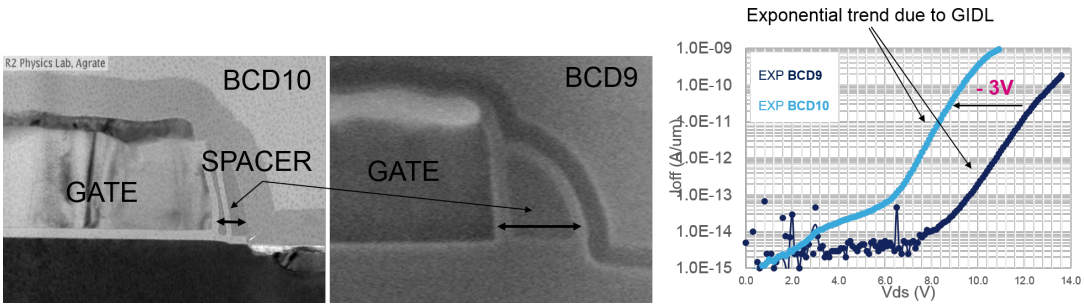


Figure 2.3: BCD9 vs BCD10 spacer dimension and comparison of the related output characteristics.

To better explain this limitation let us consider a real example. The left and the central pictures of figure 2.3 show two SEM images of two devices integrated into the most recent BCD technology node and the previous one respectively. Moving to the new technology node, the spacer dimension has been almost halved. Doing so, the electrostatic effect induced by the gate has been increased with the consequence that the OFF-state electrical performances result degraded. If we compare the output characteristics measured with the transistor biased to work in the OFF-state, i.e. with the gate and the source grounded (see the right graph of figure 2.3), we note, as expected, that the curve shifts toward left, the exponential trend due to GIDL anticipates so that the *OFF* current is, at the same voltage stress, many orders

of magnitude higher. In conclusion, the scaling and the performance requirements for digital applications penalize the voltage capability and the performance of this CMOS-based power architecture. So, while digital transistors have followed their way, power transistors have diversified and have started to follow proper scaling rules. To overcome the afore-described voltage limitations, power transistors must hence maintain thick oxide layers at the drain side and a low doped drift region between the highly-doped drain region and the gate.

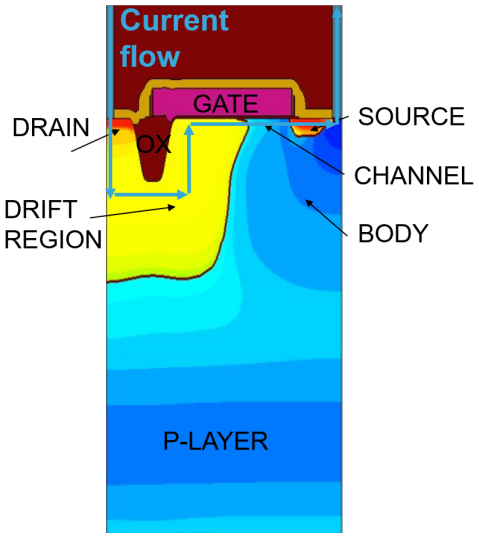


Figure 2.4: Typical LD-MOSFET cross-section, using FOX as drain extension region.

The typical architecture used to achieve this result is reported in figure 2.4. The thick field oxide (LOCOS or STI as in figure 2.4), used as isolation in the CMOS technologies, is leveraged to realize the drain extension region of the power MOSFET, separating the highly-doped drain region from the gate. The lateral dimension of the field oxide and the drain doping concentration can be optimized to target a wide range of voltage capability (from tens to hundreds). The cost paid for this solution is a longer current path since the current must flow around the oxidation, that can penalize the R_{ON} performance in particular in the low-voltage range ($< 30\,\mathrm{V}$). For low-voltage application, 'all-in-active' architectures

still remain desirable. In order to overcome the limitation related to the CMOS-based architecture (see figure 2.2), the highly-doped drain region can be separated by the gate simply through mask pattern and an hard mask (SiPROT, i.e. Silicon Protection) is used to protect the low-doped region, realized with a dedicate drain mask, from the silicide formation. Figure 2.5 shows the cross-section of the so-described all-in-active architecture. The main challenges of this kind of architecture are, on one side, the scalability that can limit the competitiveness in the very low-voltage range (up to 10 V) and, on the other side, the maximum voltage capability. Purpose of this work is to study the introduction of a new feature, the metal field plate, to optimized the drain extension region of this all-in-active architecture to increase its maximum voltage breakdown up to 30 V while securing competitive overall electrical performances.

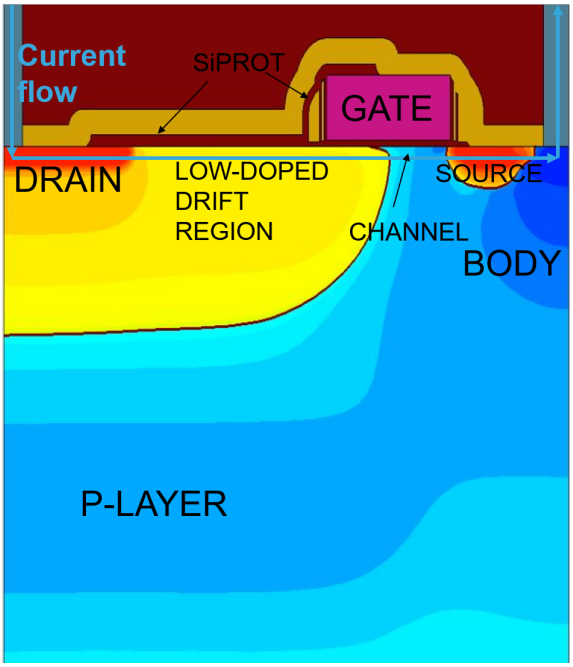


Figure 2.5: "All-in-active" LD-MOSFET cross-section.

To conclude this section we want to summarize what does it means to optimize a power device. The choice of the drain extension architecture, used for the simplified classification reported above is only a starting point; across all the reported

architectures optimized a power means:

1. To correctly size the lateral dimensions, and in some cases the vertical one where the integration constraint allows to do it. The lateral dimension is the pitch of the device, i.e. the distance between half source contact and half drain contact. It includes an active part that allows to sustain the required voltage and does not scale with the technology node and a passive one that can benefit of the tighten rules of the advanced technology platform, e.g. contact width, contact to poly...

2. The drain engineering is a key factor to bring out the best from a specific architecture. In all the advanced power devices the Reduced Surface Field Effect (ReSURF) is exploited by adding a deep implant of the opposite type of the one of the drain, i.e. boron implantation for n-channel MOS to realize the p-region called 'P-Layer', it allows to use higher doped drain and therefore to achieved better performance. The entire chapter 4 will be dedicated to the description of this very important effect.

3. The body engineering allows to realize very short channel length, significantly improving the device performances and the electrical ON-state Safe Operating Area (SOA). This can be achieved with different techniques: through the P-body approach, where the channel is realized by the lateral diffusion of a self-aligned implantation done after a dedicate poly etch, or leveraging the advanced CMOS approach, i.e. by using the standard wells defined by the lithography and the highly-doped self-aligned pocket implantation that are realized together with the drain/source low-doped implantations (LDD). The voltage capability is, however, typically less dependent on the body engineering. The body optimization and the device isolation are out of the scope of this work that it is instead focus on the optimization of the drain extension region.

A further classification of the device architecture can be done based on how the drain is realized: firstly, the drain can be realize complementary to the body region (we will call it 'Drain-not-everywhere'). The drawback of this architecture is that the critical lateral dimensions, as for example the channel length, depends both on

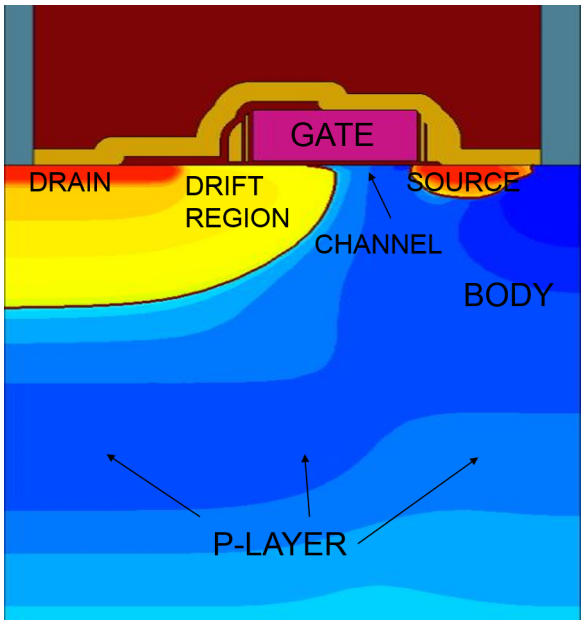


Figure 2.6: Cross-section of an all-in-active drain-everywhere LD-MOSFET.

the drain and the body alignment to the Poly. The advantage is that the drain and body concentration are independent except for the boundaries effects. This architecture is typically used for the very low voltage range (up to 10-15V) where high drain doping concentration is required to obtain very small ON-resistance. Secondly, the drain can be realize everywhere (for this reason we will call it 'Drain-everywhere') and the body defined through it by doping compensation. Figure 2.6 shows the cross-section of this integration choice, comparing this cross-section with the one shows in figure 2.5, the only appreciable difference concerns the edge of the p-layer: in the first picture, it is only below the drain region while in the second one it covers the whole device. Besides this difference, the advantages of this architecture are to avoid a mask alignment and, on advanced power MOSFET, to better isolate the drain thanks to the P-layer that is so implanted everywhere as well. On the other side, the body doping concentration depends on the drain one and this limits the maximum doping concentration that can be used for it. Finally, the last option is to have the drain implanted inside the body, opened everywhere, as it is made for the C-MOS integration. For advanced BCD, this architecture is attractive only for very low voltage application were the drain extension region are

small and the drain doping concentration high. In this work we will focus mainly on the drain-everywhere architecture that is the most attractive to target a 30 V voltage capability.

Finally, for completeness, the figure 2.1 shows the cross-section of a p-LDMOS, the dual structure, where the regions are the same but they are doped in a complementary way.

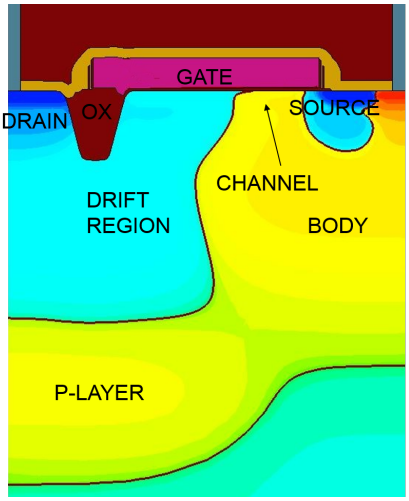


Figure 2.7: p-LDMOS cross-section.

2.2 TCAD Tools and Simulation Flow

Simulations are the core of almost every analysis, therefore, a paragraph dedicated to explaining how they are carried out and how images and graphs are produced seems compulsory. This section describes the different tools needed to produce, for example, a trans-characteristic of a transistor starting from a pure silicon substrate, before the device is integrated into it. To make more clear the entire flow, an example on a well-known device like a digital MOS will guide this description step by step.

For clearness, it is convenient dividing the simulation flow into five steps:

1. Preparation of the process simulation.
2. Process simulation.
3. Preparation of the electrical simulation.
4. Electrical simulation.
5. Display and analysis of the results.

Each of these steps is performed by a different TCAD tool, each one with its own syntax and its own inputs. A detailed analysis of each step will be carried out in the following sections to clarify, for the entire work, where each of the results, images or graphs proposed comes from and how to read it correctly.

2.2.1 Process Simulation

The process simulation tries to reproduce what happens to a silicon substrate after one or several technological processes such as ion implantations, depositions, or thermal diffusions. It requires three input files: the layout of the structure, a merging file with the instruction to transform the layout layers in a mask set and the list of the various technological steps to be simulated. The programs used to prepare the simulation are Virtuoso for the layout generation, and Ligament for the generation of the process flow and for the definition of the simulation domain. Then the simulation is performed by SProcess. The latter two are provided by Synopsys company while the former belongs to Cadence group.

The layout is the drawing used to generate all the photo-lithography masks that allow the integration of the device into a silicon substrate. Each mask allows the exposition of the resist to the UV light in some precise regions. Then, after the development of the resist, the silicon substrate is exposed only in the previously impressed region or only in the complementary regions as a function of the type of resist. In this way, it is possible to limit a well-defined process step to those exposed regions without involving the whole surface.

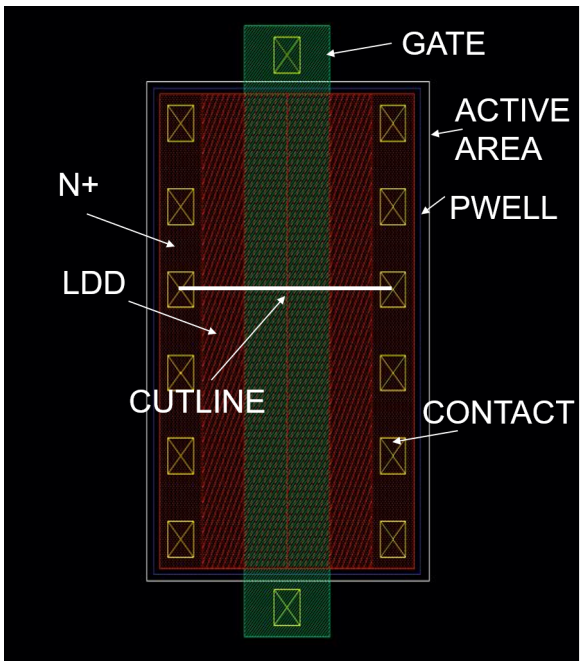


Figure 2.8: Example of the layout of a digital MOS.

Figure 2.8 shows the layout of a symmetric MOSFET as drawn in Virtuoso's environment. Each rectangle or square represents the dark-field or the light-field of a mask. In particular, as also labelled in the figure, it is possible to recognize:

- `POLY` mask: the rectangle coloured in green, placed exactly in the centre of the structure. It is a dark-field mask used together with a negative resist that allows etching and removing the polySilicon from everywhere except inside the drawn region.

- `CONTACT` mask: the squares coloured in yellow. In the picture there are the two arrays for the drain and source contacts and the two gate contacts placed in the head of the transistor. It is again a dark-field mask but it is used together with a positive resist since we need to remove the PMD in the indicated regions and create hence the trenches to host the tungsten for the contacts.

- `N+` mask: the rectangles filled with red dots placed in the same region where there are the contacts. It allows the implantation of arsenic to realize the highly-doped n+ drain and source regions.

- `NLDD` mask: the rectangles filled with red lines, placed between the `POLY` and the `N+` masks. It allows the implantation of the drain and source LDD and the pocket implantation.

- `AA` mask: the white and greatest rectangle. Outside it, a FOX is realize to provide lateral isolation to the device.

- `PWELL` mask: the blue rectangular that cover almost the whole device. As explained, the CMOS body is implanted everywhere and source and drain regions are realized into it.

On top of this layout, a further layer that defines the simulation domain must be added. SProcess simulates the region indicated by this layer for a certain depth of the substrate and not for the entire depth to save simulation time and memory space. So, drawing a 1D cutline will result in a 2D simulation, while drawing a square-like region, a 3D simulation. For this example, let's consider a 2D simulation that ranges from a drain contact on the left to the respective source contact on the right. Once the layout is ready, a file containing the list of operations that SProcess will have to simulate, completes the preparation for the process simulation. This second file is simply a list of instruction, each of them made of at least two fields. The first one contains the name of the process and the second the values of the variables. For example, a diffusion process needs at least three parameters: the initial temperature, the final temperature and the diffusion time; while thermal oxidation needs also the oxygen concentration or its flow rate.

At this point, the process simulation can begin. SProcess solves each instruction of the process flow. First, it generates an accurate and adaptive mesh on which it solves a system of partial differential equations with the finite element method (FEM). While a detailed description of the math of the solver is unnecessary, it is important underlining that the tool was previously calibrated personalizing each parameter of each model to properly fit the experimental data.

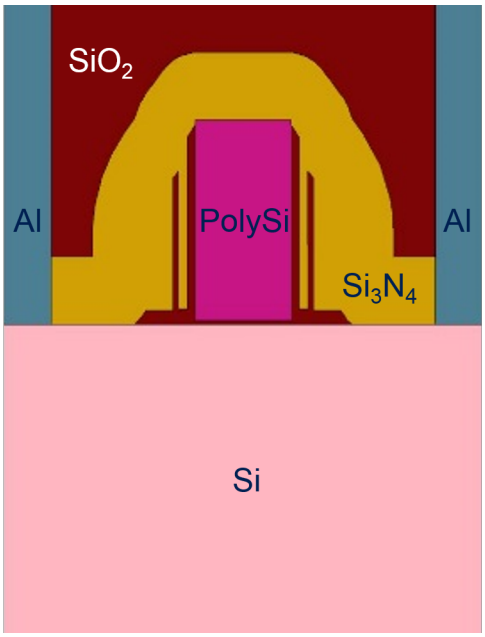


Figure 2.9: Process simulation output.

The output of this stage is 2D interactive picture that shows the cross-section along the previously drawn cutline. Figure 2.9 shows the starting output of this simulation step. In it, one can distinguish only the regions characterized by a different material, particularly, the ones that are used in this work and that appear in the picture are:

- Grey: aluminum(Al) or any other metal used for interconnections, vias or contacts.
- Brown: silicon dioxide(SiO_2).

- Yellow: silicon nitride(Si_3N_4).
- Pink: silicon(Si).
- Purple: polysilicon($PolySi$).

Observing this figure, one can recognize the two contacts for the drain and source regions placed at the edge of the simulation domain; the gate, exactly in the centre, the spacers, the silicon substrate and the pre-metal-dielectric (PMD). The body contact is not simulated to save space and time of the simulation. Nevertheless, it will be added, as for the substrate contact, on an edge of the domain.

Together with the morphology of the device, the output of the process simulation includes many other useful information. Indeed, even if the gate is visible, is this structure truly a MOS system with the source, drain and channel regions well-defined? Or is it possible that, for example, over-diffusions or wrong n+ implants prevented the channel formation? Consequently, it is certainly more useful observing also at least the doping distribution inside the silicon. This distribution, after the process simulation, is well-known and it is present inside the same file. The visualizer SVisual gives the possibility to superimpose on the cross-section the physical quantities that the simulator has already evaluated. Those quantities, at this stage, are the doping distribution relative to the single or all species, the stress and the strain. It allows also drawing new cutlines to see those distributions as a function of a geometric axis in a 2D graph. These are the reasons why these pictures were defined as 'interactive'.

Figure 2.10 shows in the left picture the total doping distribution inside the silicon of the same architecture and in the right picture the absolute doping concentration along the highlighted cutline, i.e *C1*. Now, it is clearly discernible that the MOS structure is correct and that the various regions are well-defined. As it is understandable from the figure, `SVisual` uses hot colours to identify high values, and cold colours for low or negative values. So, in the image, it is easy to recognize the n-region of source and drain coloured with colours from red to yellow and the p-region coloured with different blues.

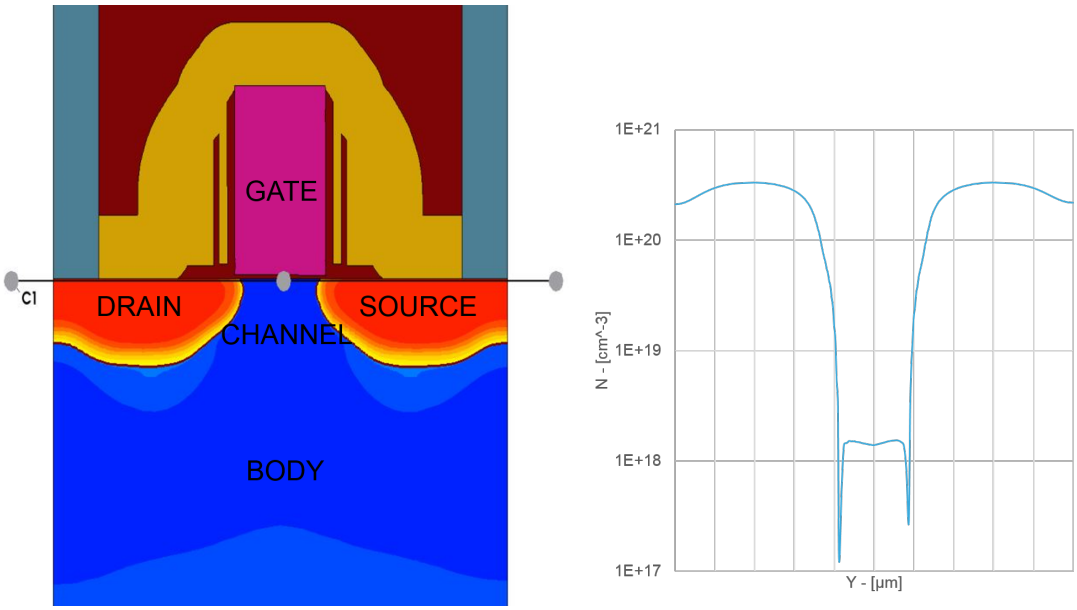


Figure 2.10: Doping distribution inside the MOS structure.

2.2.2 Electrical Simulation

Once the structure is correctly integrated on silicon, it is possible to perform the electrical simulation to visualize the I-V characteristics, the distribution of the electric field and so on. To prepare the simulation, three important operations must be performed in advance:

1. Add contacts to impose on them the boundary or stress conditions for the device.

2. Define a new and targeted mesh, more refined where it is important for the electrical simulation. So, for example, more refined on the channel for a threshold measurement or on a junction to see its breakdown.

3. Define the simulation and so choose the physical models to be used, the solving models, specify the outputs to be saved, and the voltage or current stress to be applied to each contact.

These operations can be made in SProcess by script or in Sentaurus Structure Editor (SDE) which has an easier graphic interface. Then, the SDevice

tool performs the electrical simulation taking as input the so-prepared cross-section. All the aforementioned tools are always provided by Synopsys. The figure 2.11 shows the result of this preparation: the three contacts, black for the drain, white for the source and blue for the gate have been added. Besides them, also the body and substrate contacts have been added also if they are not visible in the cross-section since they are placed deep. A mesh, correctly refined around the channel region, has been defined as well. Finally, once defined, the simulation can start. For this example and coherently with the generated mesh, let's consider a threshold simulation with the transistor working in the linear region, ramping firstly the *drain* contact up to 0.1 V and then the *gate* contact up to 5.0 V.

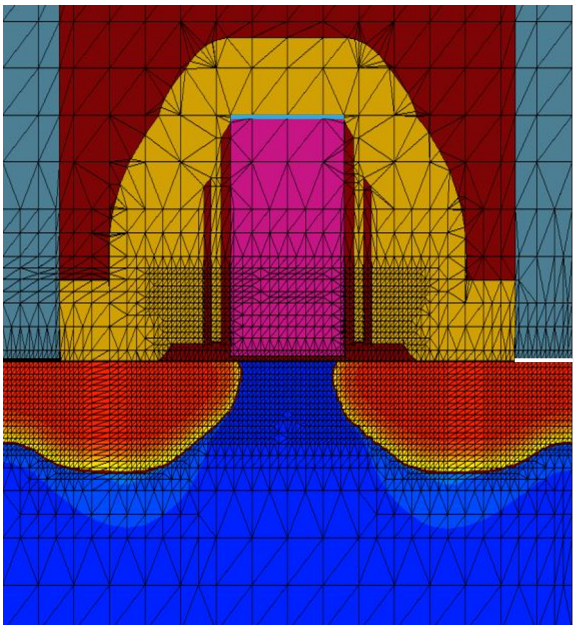


Figure 2.11: Cross-section ready for the electrical simulation.

SDevice produces two output files. The first one contains all electrical data linked to each contact and required as output during the preparation of the electrical simulation. So, it is possible to plot a graph putting in X and Y axes the needed electrical quantities. In the example, to see a trans-characteristic, the gate voltage is placed along the x-axis and the drain current along the y one, but it is possible to choose any combination of voltage, current, and charge. Figure 2.12 shows the

result.

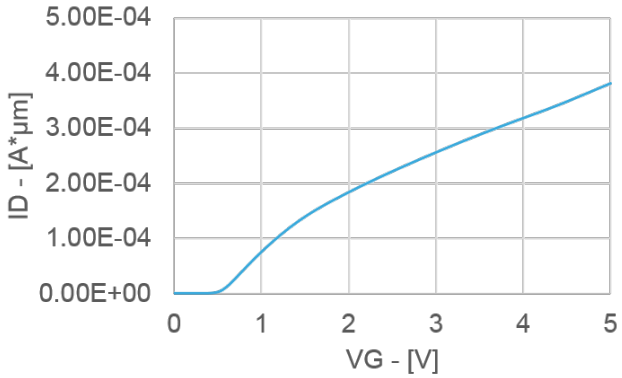


Figure 2.12: Trans-characteristic of the MOS.

The second output file, instead, is always a cross-section where it is possible to see, superimposed to the structure, the chosen quantities like, for example, the current distribution, the electric field, the impact ionization, the trapped charge and so on. The tools described in this paragraph, together with their inputs and outputs, are summarized in table 2.1.

Tool	Aim	Inputs	Outputs
Virtuoso	Layout definition	-	Layout
Ligament	Flow definition	Tech. Process Flow	Compilated Process Flow
SProcess	Process Simulation	Layout, Process Flow	Device Cross-section
SSE	Contact and Mesh	Device Cross-section	Cross-section
SDevice	Electrical Simulation	Cross-Section	I-V curves, E...
SVisual	Visualizer	Simulation Outputs	-

Table 2.1: Summary of simulation tools

Chapter 3

Power LD-MOSFET Electrical Parameters

During the design of a power transistor, the engineer takes decisions and optimizes each single process step and the order with which each step must be executed, i.e. he decides the temperatures of the thermal steps, the duration of the etching, the energy of the implantations and so on. This is particularly true for discrete components, where the process flow can be optimized for only those devices. For components that must be integrated into a BCD platform, the engineers have less degrees of freedom since the entire flow is tuned to make possible the integration and to grant the performance of the various structures.

Anyway, any change made to the process flow changes the electrical performances of the transistor.[15] Some changes have a direct impact on one or more of them, while other ones affect them as a side effect. Particularly, in our work and for our purposes, we will change the doses and the energies of the dedicated HV-wells. Examples of electrical parameters that can be affected are the ON-resistance (R_{ON}), the breakdown voltage when the transistor is OFF (BV_{OFF}), the gate-drain capacity (Q_{GD}), the threshold voltage (V_T), without forgetting the reliability and defectiveness aspects that will also be affected. Thanks to these characteristics, it is possible to compare different process solutions and choose the best one, namely the one that guarantees the best compromise between all the electrical parameters. This implies that there cannot be an optimal transistor under every point of view and for any application, but it is instead possible to optimize a device according to specific targets coming from the customer requirements or market benchmarks, but remaining always into the limits imposed by the chosen technology platform. Optimizing, therefore, means finding the best trade-offs between all electrical features,

improving certain aspects by sacrificing others.

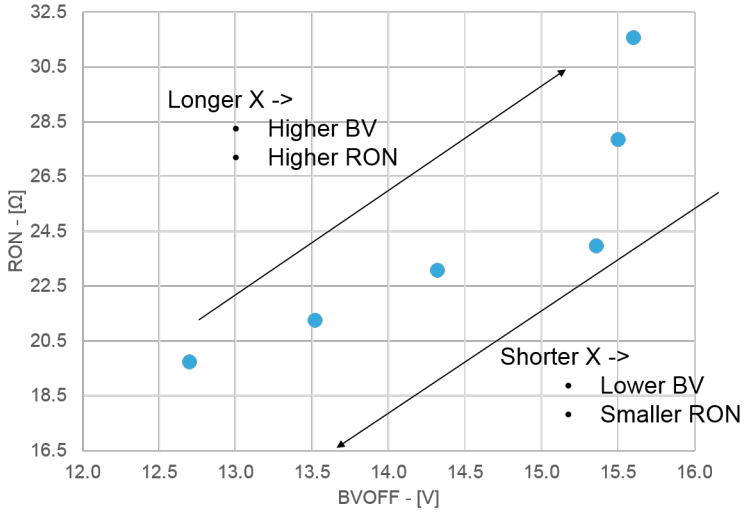


Figure 3.1: BV_{OFF} - $R_{ON} \cdot A$ trade-off as a function of X.

A power device is often used as a switch. A good switch must bear high voltages during the OFF-phase and have the smallest possible resistance during the ON-phase to have the highest power transfer to the load. Both parameters are a function of many technological and geometrical parameters such as the dose and energy of the drain and p-layer implantations. Once an architecture is set, we can change one parameter and see what happens to the electrical performance. Let's choose, for example, the length of the drift region X. In figure 3.1), there are plotted into a R_{ON}-BV_{OFF} graph the experimental data measured on a test-chip for a set of all-in-active architectures which differ only for the drift length. It results that we can increase the breakdown voltage to target higher voltage classes but with the consequence to penalize the resistance. On the contrary, to try to obtain better performance, i.e. lower resistance, we penalize the BV_{OFF}. With this example, we have introduced one of the most important trade-offs for a power device. The analysis of this trade-off and the consequent study and optimization of the drain side will be at the centre of the remaining chapters. The other electrical parameters are also important but their optimization can be performed later acting on different technological aspects.

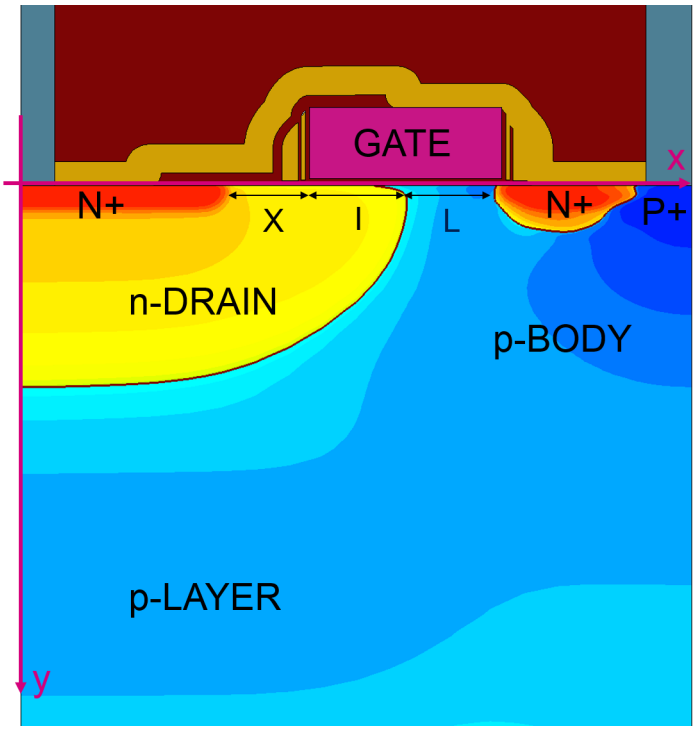


Figure 3.2: Cross-section of the reference structure (POR) chosen for the analysis.

Let's consider more in detail now the third point of the previous graph. The cross-section of the architecture under analysis is shown in figure 3.2. From now on, the x- and y-axes are directed as they are drawn in the picture. Moreover, we identify the drift region, the accumulation region and the channel respectively with L, I, and X capital letters that, for the device that we have considered are equal to:

Channel Length	$L = 0.2\,\mu m$
Gate Drain Overlap	$I = 0.2\,\mu m$
Drain Extension	$X = 0.2\,\mu m$

For the definition and the computation of the resistance and the breakdown voltage, also the following implantations characteristics are fundamental.

DRAIN	
Dose	$N_1 = 8 \times 10^{12}\,\mathrm{cm}^{-2}$
Energy	$E_1 = 150\,\mathrm{keV}$
Dose	$N_2 = 4 \times 10^{12}\,\mathrm{cm}^{-2}$
Energy	$E_2 = 80\,\mathrm{keV}$
P-LAYER	
Dose	$N = 2.2 \times 10^{13}\,\mathrm{cm}^{-2}$
Energy	$E = 360\,\mathrm{keV}$

On this structure, the breakdown voltage, the threshold voltage and the R_{ON} are measured and the results are reported in table 3.1 besides the previous graph.

V_T	R_{ON}	$R_{ON} \cdot W$	BV_{OFF}
1.81 V	23.071 Ω	1.846 kΩ µm	14.3 V

Table 3.1: Electrical performances of the POR structure.

Let's consider this particular architecture as a starting point. From it, we start changing geometry and technological parameters to obtain better electrical performances. However, it is important to understand ahead which changes allow to increase the voltage capabilities or to reduce the resistance and among them which shall be preferred and why. As a consequence, the next paragraphs are dedicated to describe in more detail the set of electrical features we have highlighted and to understand which and how the technological variables affect them. Analytical models are provided to estimate and clarify the links between the technology level and the electrical one, moreover, measurements and simulations will be massively exploited to support and verify the theoretical analysis.

3.1 Threshold Voltage VT

The threshold voltage is one of the main electrical parameters of a transistor. It is particularly important for analogical applications since it influences the linearity of the device discriminating the linear region from the interdiction region. For digital and power applications, instead, the transistors are mainly used as switches, i.e. they are driven with logic discrete levels so that the gate electrode is biased with null bias or with the maximum one allowed by the technology; as a consequence, the threshold voltage assumes less importance. Qualitatively, besides its usage into the analytical expressions of the drain current as a function of the gate voltage, it describes mainly the performances and the static consumption of a transistor. Generally, both are inversely proportional to the threshold voltage, so, smaller is the V_T, greater is the leakage and faster is the transistor for switching.

It exists several definitions of the threshold voltage[16] that are here reported:
Definition 1: the threshold voltage is the gate voltage that makes the inversion charge nothing.
Definition 2: the threshold voltage is the gate voltage that makes the superficial minority carrier concentration equal to the majoritarian carrier concentration in the substrate.
Definition 3: for a gate voltage greater than a certain value, it exists a linear region for the inversion charge as a function of the gate bias. The intercept of the line that fit the curve in this region is the threshold voltage.
The first definition is quite arbitrary, since, in principle, it is possible to choose any value of the inversion charge, moreover it differs from the usual definition used to create circuit models. For these reasons, it is discarded. The second and the third definitions are, instead, equivalent and linked to the MOS capacitor. Since they are directly referred to how the V_T is measured, they will be the ones that will be used.

From [17], the theoretical threshold voltage expression is here reported:

$$V_T = V_{FB} + |2\Phi_p| + \frac{\sqrt{2\epsilon_s qN(|2\Phi_p| - V_B)}}{C_{ox}} \tag{3.1}$$

where

$$V_{FB} = \Phi_m - \Phi_s \qquad \text{Flat band voltage}$$

$$V_B \qquad \text{Body bias}$$

$$\Phi_p = \frac{E_f - E_{f_i}(+\infty)}{q} = \frac{K_B T}{q} \cdot ln(\frac{N}{n_i}) \qquad \text{Silicon potential at} + \infty$$

$$C_{ox} = \frac{\epsilon_0 \epsilon_{ox}}{T_{ox}} \qquad \text{Gate capacitance}$$

$$Q_D = \sqrt{2\epsilon_s q N(|2\Phi_p| - V_B)} \qquad \text{Depletion charge}$$

Substituting the numbers related to the device shown in figure 3.2 and that are reported in the table 3.2, the result is

$$V_T = 1.519$$

Parameter	Value
Φ_m	4.05 eV
V_B	0 V
N	5×10^{17} cm^3
T_{ox}	130 Å
T	300 K

Table 3.2: Personalized parameters for the V_T evaluation.

Concerning table 3.2, Φ_m is the work function of the n-polysilicon that was fixed equal to the electronegativity of silicon itself. T_{ox} is the thickness of the gate oxide and is extracted from the process flow or SEM images while the body bias and the temperature come from the measurement conditions. To justify the chosen doping level, instead, a more accurate explanation is needed. The equation [3.1] is true only for a channel with uniform doping, but, in our structures, the doping profile is not constant, especially along the x-axis. Figure 3.3 shows the doping profile along a cutline parallel to the x-axis just below the interface with the oxide. Considering only the channel region it is possible to note that the absolute concentration increases from drain to source. The responsible for this increasing

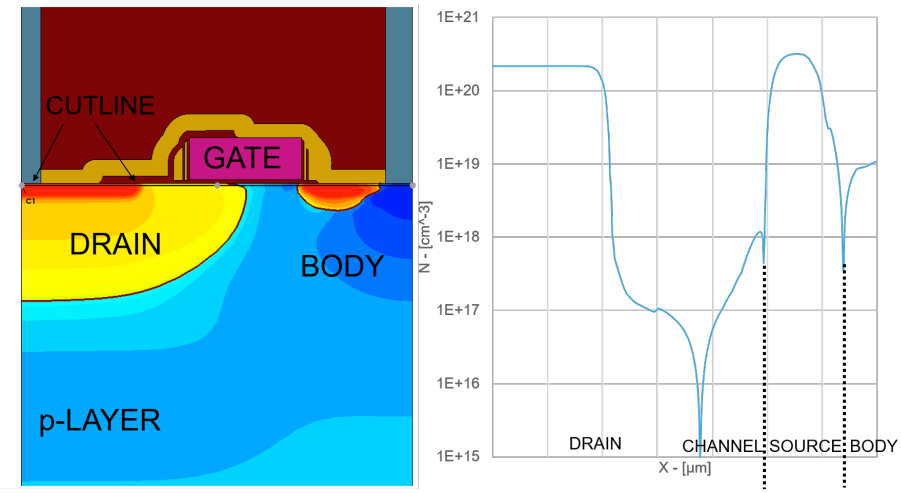


Figure 3.3: Channel doping distribution along the drawn cutline.

profile is the presence of asymmetric pocket implantation. The pocket implantation is used to introduce additional atoms of boron near the edges of the channel to compensate the Short Channel Effects (SCE) and to adjust the V_T value.[18] In a standard CMOS, the pocket is implanted at both gate sides, and this results in a bell shape doping profile[18]; while in our device it is implanted only at source side. We can approximate the implanted charge with the two following assumptions:

1. The concentration at the drain side is due only to the p-well implantation.

2. The concentration at the source side is due only to the pocket implantation.

These assumptions are also justified looking at the results of the simulator. The doping level reported in table 3.2 is the average value between these two corners.

Figure 3.4 shows the experimental trans-characteristic obtained measuring the structure on a test chip. On an experimental point of view, it is possible to follow two different roads for the measurement of the threshold voltage[19]. The first method is the easiest and fastest one: the threshold voltage is the gate voltage that allows a certain drain-source current. Graphically, it consists of drawing a horizontal line in the trans-characteristic at the current chosen as the threshold, find the intersection with the curve and read the voltage value at that point. This method is much used for digital transistors where great accuracy is not required and the only important

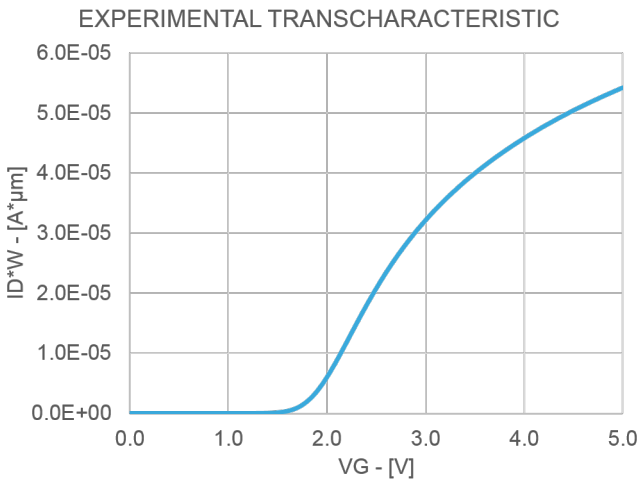


Figure 3.4: Experimental trans-characteristic.

thing is to find the gate voltage that induces a current able to start degrading the logic voltage levels[19]. For example, if we use a threshold current density of $I_D \cdot W = 1\,\mu\text{A}\,\mu\text{m}$, it is possible to extract the V_T as

$$V_T \approx 1.7V$$

Figure 3.5 shows the procedure before described under a graphical point of view, for clearness, the trans-characteristic is plotted with a logarithmic scale and only for a smaller range of the gate voltage. The threshold current has been chosen as the value for which we consider the transistor ON.

Alternatively, it is possible to consider the equation of the drain current and solve to find the unknown V_T knowing the trans-characteristic. Generally, the trans-characteristics are evaluated biasing the transistor in the linear region[1]. Particularly, the 3.4 is obtained biasing the drain electrode at 0.1 V. From the semiconductor device theory, the drain current in the linear region and neglecting the modulation

[1]It is also possible to compute the V_T with the transistor biased in the saturation region. It is often used to analyze the short channel effects that are, in this way, accentuated by the Drain Induced Barrier Lowering (DIBL).

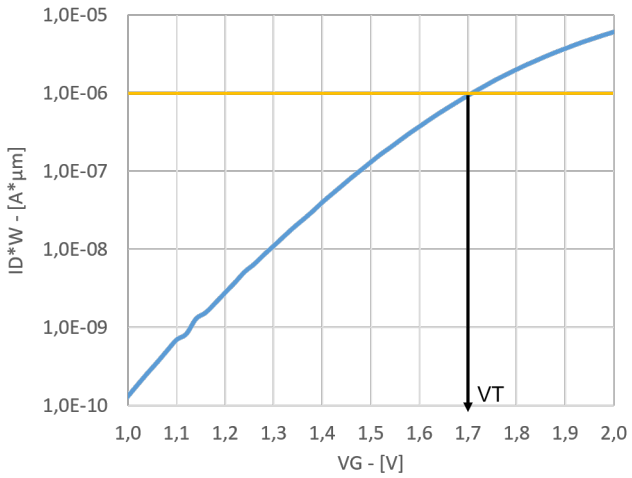


Figure 3.5: V_T evaluation - first method.

of channel length is expressed by:

$$I_D = k \cdot (V_G - V_T) \cdot V_D \qquad \text{for } V_D << 2 \cdot (V_G - V_T) \tag{3.2}$$

where the variable k is equal to

$$k = \frac{\mu_n \cdot C_{ox} \cdot W}{L} \tag{3.3}$$

The equation [3.2] states that I_D is linearly dependent on VG with two unknown parameters: the variable k, technology-dependent, and, as anticipated, the threshold voltage V_T. So, it is possible to define a system of two equations in two variables substituting two different points taken from the $I_D - V_G$ curve. Choosing, for example, the points shown in figure 3.6, this system will follow:

$$\begin{cases} 1 \times 10^{-5} = k \cdot (1.86 - VT) \cdot 0.1 \\ 5 \times 10^{-6} = k \cdot (1.68 - VT) \cdot 0.1 \end{cases} \Rightarrow \begin{cases} k = 3.125 \times 10^{-4}\,\mathrm{A\,V^{-2}} \\ V_T = 1.82\,\mathrm{V} \end{cases}$$

Graphically, it is possible to arrive at the same result following several steps, listed in the following and show graphically in 3.7. [19]

1. Draw the trans-characteristic and the trans-conductance curve.

2. Find the maximum of the trans-conductance curve and draw a vertical line in

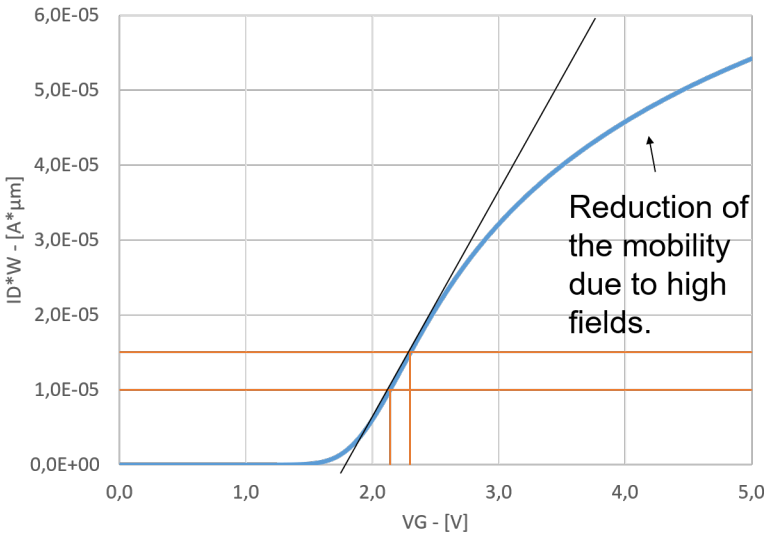


Figure 3.6: Numerical evaluation of the linear V_T - second method.

that point.

3. Find the intersection between the vertical line and the I_D curve and draw the tangent to this last curve in that point.

4. The intersection between the tangent and the x-axes is the threshold voltage.

As it is clear from figure 3.7, the linearity relation between I_D and V_{GS} holds only in a very small interval around the maximum of the trans-conductance. This also means that the solution of the linear system aforementioned is practically independently on the chosen points until they belong to that small interval. Outside, the drain current deviates from linearity. For higher gate voltages the high electric fields inside the structure increase the scattering events and consequently the electrons channel mobility decreases. This phenomenon, joined with the voltage drop on the series resistance of the drain extension that, due to the high current, is no longer negligible, causes the deviation from the ideal linear behaviour. In the same way for lower gate voltages, the sub-threshold effects and the interdiction of the transistor cause the deviation from the linear behaviour. According to this routine and consistently with the previous analytical solution, the threshold voltage is again 1.8 V.

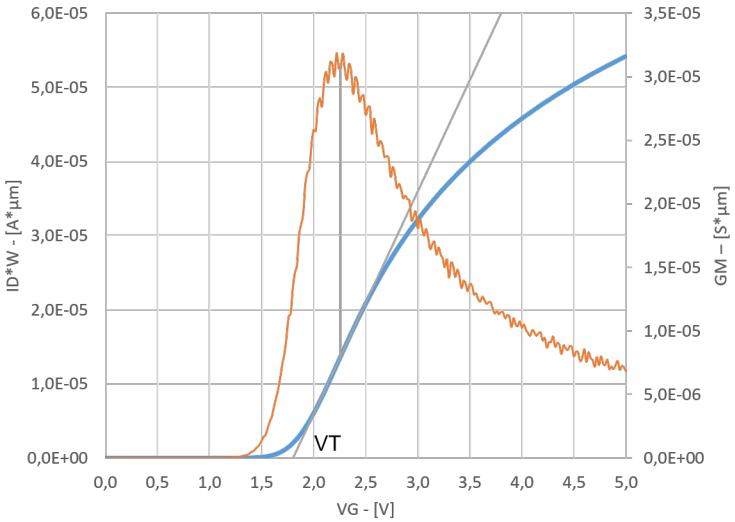


Figure 3.7: Graphical evaluation of the linear V_T - second method.

Comparing finally the experimental trans-characteristic with the simulated one it is possible to see that there are two significant differences (see figure 3.8). The first one regarded a shift of the threshold voltage and the second one a shift of the ON-resistance that, since it will be analysed in the next paragraph, here will be not considered. The ΔV_T might come from a wrong estimation of the channel doping due to a not correct modelling of the diffusion phenomenon; particularly, the phosphorus diffusion model that SProcess uses, overestimates the diffusion of the source region. The visible effect is a partial compensation of the pocket implantation charge, a consequently decreasing of the channel doping and finally a smaller threshold voltage for the simulated curve. Alternatively, it is also possible that there are some extra charges at the interface silicon-GATOX. Table 3.3 summarizes all the results we found.

	VT
THEORETICAL	1.5 V
SIMULATED	1.5 V
MEASURED	1.8 V

Table 3.3: Comparison of VT results.

Often, some extra charges are considered to align the simulated and theoretical

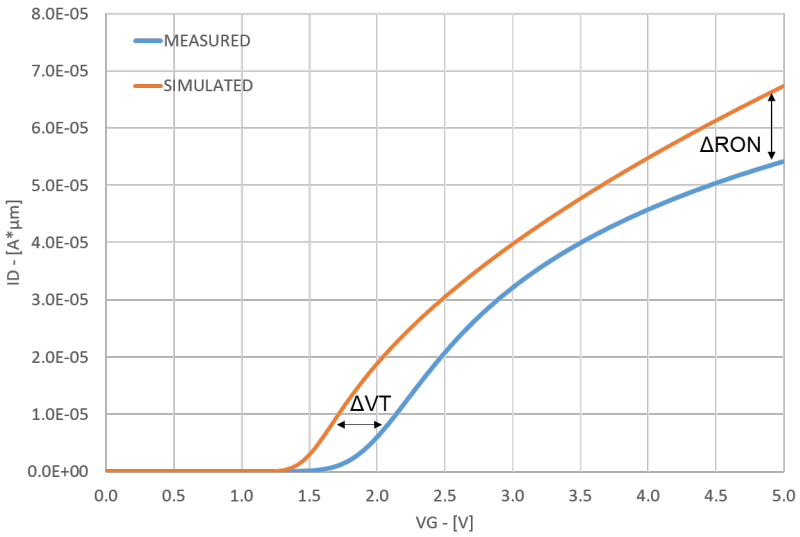


Figure 3.8: Comparison between experimental and simulated trans-characteristic.

results with the experimental one. It can be inserted into the [3.1] as a simply addictive term that we call Q, or we can force the simulator to consider a certain value of addictive charges at that interface.

$$V_T = V_{FB} + |2\Phi_p| + \frac{\sqrt{2\epsilon_s qN(|2\Phi_p| - V_B)} + Q}{C_{ox}} \tag{3.4}$$

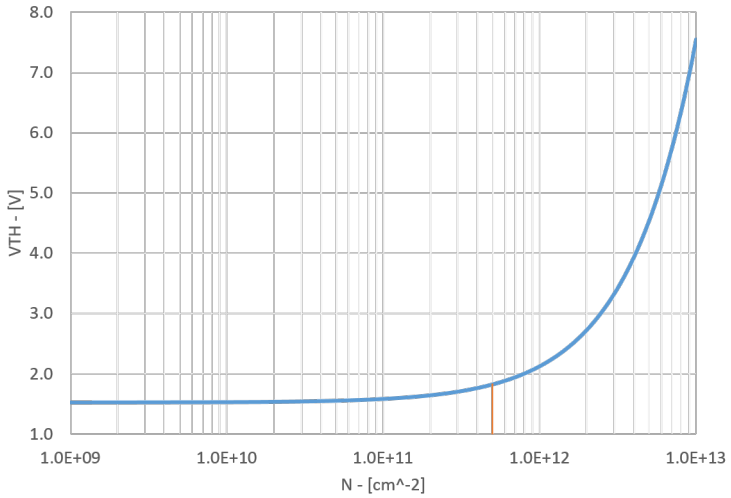


Figure 3.9: V_T curve as a function of the extra charge.

To find the value of Q to line up the theoretical and simulated values with the experimental one, the V_T trend as a function of the extra charges is plotted. For a small value of the extra charges the threshold voltage is dominated by the channel doping and is almost insensible to them. On the contrary for a high dose of extra charges the dependence is very strong. To line up the results an extra charge equal to $Q = 5 \times 10^{11}\,\mathrm{C\,cm^{-2}}$ must be added. The same value of extra charges is needed for both theoretical and simulated results.

3.2 ON-Resistance

The ON-resistance (R_{ON}) is the resistance evaluated between the drain and source contacts when the transistor is in the ON-phase. Qualitatively, it describes how much the transistor is far from being an ideal switch. The ideal switch is characterized by null resistance between its terminals when closed, while, in a real switch, it is impossible to avoid a resistive contribution. This contribution causes a voltage drop across the switch and a loss of the overall power. For these reasons, a designer aims to minimize the R_{ON}. It is measured at the gate bias used to turn on the transistor with the drain electrode biased to work in the linear region. In our measurements, we are used to fixing the gate and drain bias to $V_G = 5V$ and $V_D = 0.1V$ as we did for the threshold voltage computation.

Starting from the linear trans-characteristic (see figure 3.4) to evaluate the resistance is enough to find the current value at the maximum gate voltage and then applied the [3.5].

$$R_{ON} = \frac{V_D}{I_D(V_{G_{max}})} \tag{3.5}$$

Next to this absolute value, two other resistance values are usually computed: the resistance per width and the resistance per area. Since it is possible, in principle, to reduce the resistance how much we want by increasing the transistor width, the resistance per width is an important figure of merit that discriminates which architectures or process changes are the best. The device with the smallest resistance per width can reach, in fact, a certain resistance with a smaller lateral dimension that means a smaller area and smaller capacitances. The resistance per area, instead, is another figure of merit that is related to the dissipated power by the switches while delivering current to the load. They are defined as follow:

$$R_{ON} \cdot W = \frac{V_D}{I_D(V_{G_{max}} = 5V)} \cdot \frac{W}{1000} \tag{3.6}$$

$$R_{ON} \cdot A = \frac{V_D}{I_D(V_{G_{max}} = 5V)} \cdot \frac{W \cdot P}{1000} \tag{3.7}$$

Where W and P are respectively the width and the pitch of the transistor. The pitch is the distance between half drain contact and half source contact. The factor one thousand is used only to normalize the results according to the following units of measurement:

$$R_{ON} \rightarrow [\Omega]$$
$$R_{ON} \cdot W \rightarrow [\mathrm{k}\Omega\,\mu\mathrm{m}]$$
$$R_{ON} \cdot A \rightarrow [\mathrm{m}\Omega\,\mathrm{mm}^2]$$

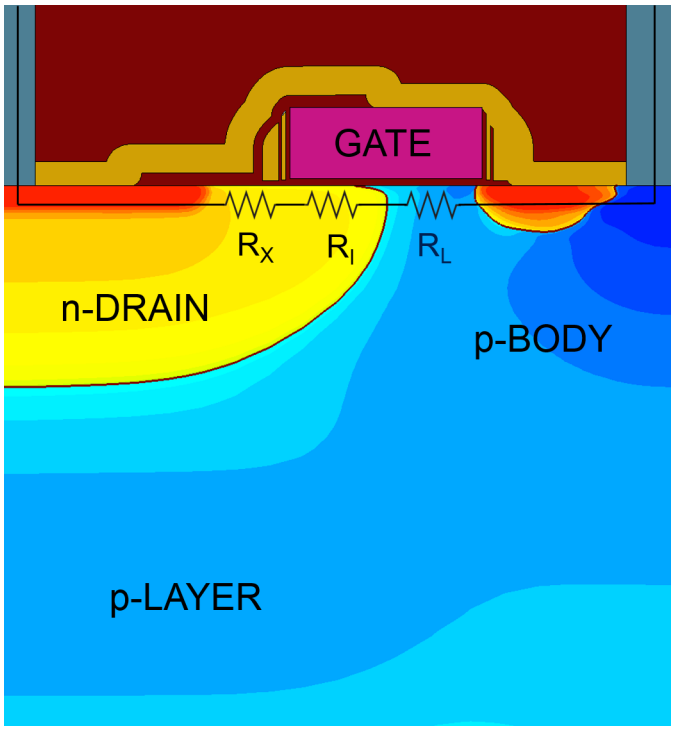


Figure 3.10: Main contributions to the R_{ON} highlighted in the reference cross-section.

Considering the cross-section of the device (Figure 3.2) we can state that the total resistance is the sum of several contributions: metal lines, metal-silicide junctions, highly-doped source and drain regions, channel, accumulation and drift regions. Silicidated regions are completely similar to metals so that their contributions are

very small. Furthermore, all the regions contributing to passive pitch are reduced as much as the technology node allows. The metal lines, instead, besides having too a very small resistance, show a not negligible contribution when powers have very large total width (The resistance of the power lines can be in the tens of mΩ range). However, we will focus only on the 'Silicon' contribution to the total resistance. It is so possible to consider three main contributions to the total resistance: the channel resistance, the gate-drain overlap resistance and the drain extension (see figure 3.10). The other contributions can be grouped and considered introducing an offset term in the complete expression of the resistance [3.8]. [20][21]

$$R_{ON} \cdot W = R_{L_S} \cdot L + R_{I_S} \cdot I + R_{X_S} \cdot X + R_{off} \tag{3.8}$$

3.2.1 Channel resistance

In figure 3.10, we called the contribution of the channel region to the total resistance R_L . Its accurate evaluation is often a challenge since the uncertainty linked to some technological and physical parameters is very high. Its analytical expression [4] can be derived with some trivial mathematical manipulations directly from the expression of the drain current for the linear region. [20][21]

$$I_D = k \cdot (V_G - V_T) \cdot V_D \rightarrow \qquad R_L = \frac{V_D}{I_D} = \frac{1}{k \cdot (V_G - V_T)} = \frac{L}{C_{ox} \cdot W \cdot \mu_n \cdot (V_G - V_T)}$$

Analyzing the equation, we can make various considerations. The gate voltage and the oxide capacitance are known: the former is fixed to 5 V by the measurement conditions of the ON-resistance that we have already discussed; while the latter is computed starting from the oxide thickness that is known through direct measurements on SEM images, besides the thermal oxidation with which we grow the gate oxide is well-controlled and much robust against process variation. The transistor width W will disappear since we will work with the resistance per unit width and never with its absolute value for the aforementioned reasons. The threshold voltage, the electron mobility and the channel length are, instead, unknown or known with great uncertainty. We have already discussed the threshold voltage computation in the previous paragraph and the difficulties related to the choice of the average channel doping and of the extra charges, so here, without repeating the whole discussion, we will use only the previous result. The electron mobility in an inversion region like the channel of a MOSFET is different from the one that we can extract from tables or models that are valid for doped semiconductors, particularly, three new phenomena can not be longer neglected: phonons scattering, surface roughness scattering and Coulomb scattering.[22] There are plenty of scientific articles [21][22] that report models for the electron mobility in an inversion layer more or less complex, but here we want to follow a different road. Having a lot of experimental data, we want to extract the value of the mobility from them to then compare it with the literature and the simulator results. Finally, the channel length is defined at the layout as the distance between the edges of the drain and the gate masks but its actual value depends on the relative diffusion of the ions of phosphorus and

boron during the thermal steps. To take into account this variability, we express the channel length as the sum of a nominal length and an unknown ΔL. In the end, we can write the following expression for the ON-resistance that has two unknowns: ΔL and μ_n. Moreover, we must add an addictive term K to take care of the other contributions that do not depend on the channel. This shrewdness is necessary to obtain consistent results since the experimental data consider the overall resistance and not only the channel resistance.

$$R_L \cdot W = \frac{L + \Delta L}{C_{ox} \cdot \mu_n \cdot (V_G - V_T)} + K \tag{3.9}$$

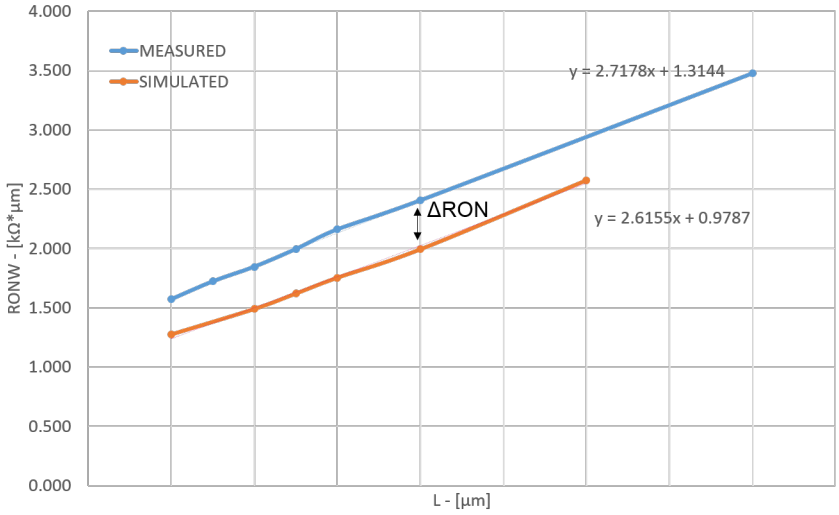


Figure 3.11: Simulation and experimental results of R_{ON} measurements on real devices.

As anticipated, to extract the values of the unknowns we need the experimental data that we will immediately discuss. In the graph 3.11 two curves are plotted: the blue line reports the experimental resistance per unit width evaluated in the electrical laboratory for several structures those differ for only the channel length; the orange line, instead, reports the simulation results of the same structures. For our purpose, we would need to only the experimental data, nevertheless, it is interesting to note also how the calibration procedure used on the simulator before starting the simulations brings very good results: the simulator can correctly evaluate the impact of the channel length on the resistance since the two lines have almost the

same slope. However, it is still present an almost constant gap, ΔR_{ON}, that can depend on an inexact modelling of one or some process steps such as the diffusions and/or on the contributions of the passive pitch of the device or measurement set-up. In any case, the simulations provide very accurate indications about the sensibilities, the parameters trends and the distribution of physical quantities[2]. Now, focusing on only the experimental results, we can write a set of linear systems with two equations in two unknowns. In each system, we substitute a different pair of points taken from 3.11 to set up, in the end, twenty-one different systems.

$$\begin{cases} R_L \cdot W|_i = \dfrac{L_i + \Delta L}{C_{ox} \cdot \mu_n \cdot (V_G - V_{T_i})} + K \\ R_L \cdot W|_j = \dfrac{L_j + \Delta L}{C_{ox} \cdot \mu_n \cdot (V_G - V_{T_j})} + K \end{cases}$$

To the term K, we assign the intercept of the best fitting line of the experimental data. We are now ready to solve all the linear systems we set up previously. Since each of them provides a value for the pair ΔL-μ_n, in the end, we obtain two vectors with twenty-one entries: one for each unknown. Considering their averages, we find

$$\begin{aligned} \mu_n &= 426.1234\,\mathrm{cm^2\,V^{-1}\,s^{-1}} \\ \Delta L &= 1.9\,\mathrm{nm} \end{aligned}$$

The electron mobility result is reasonable, it is very similar to what evaluated in [21] at equal boundary condition, i.e. gate bias, and also the simulator anticipates a value much near to it. In figure 3.12, there is plotted the simulated mobility along a cutline perpendicular to the channel. As it is possible to see, in the channel region the value we find with the theoretical model that we described before, and the one the simulator find are practically the same. The decreasing trend of the mobility at the source side is due to the pocket implantation that increases the doping level making, as a consequence, the mobility smaller. The value of ΔL, instead, is less realistic. From the cross-section reported in figure 3.12, we can immediately observe that already in the simulation, the channel edges are not 'vertical' at all, but rounded due to the diffusions. Particularly, relying on the simulation results, the actual L is

[2] About the correctness of a simulation, someone told me that a perfect agreement between simulation and experimental results means that the simulation is surely wrong!

around ten nanometers longer than its nominal value and consistently, the I region is around ten nanometers shorter than its nominal value. The error we made concerns the average operation that computes a so small result due to the presence of some negative values among the positive one. Both positive and negative ΔL span from few nanometers to tens of nanometers. Considering hence the average value of only positive values or only negative ones, we obtain

$$\Delta L \approx \pm 10\,\text{nm}$$

Neglecting the result with the minus sign on the basis of the simulations, also the variation of the channel length is in agreement with what observed in the cross-sections.

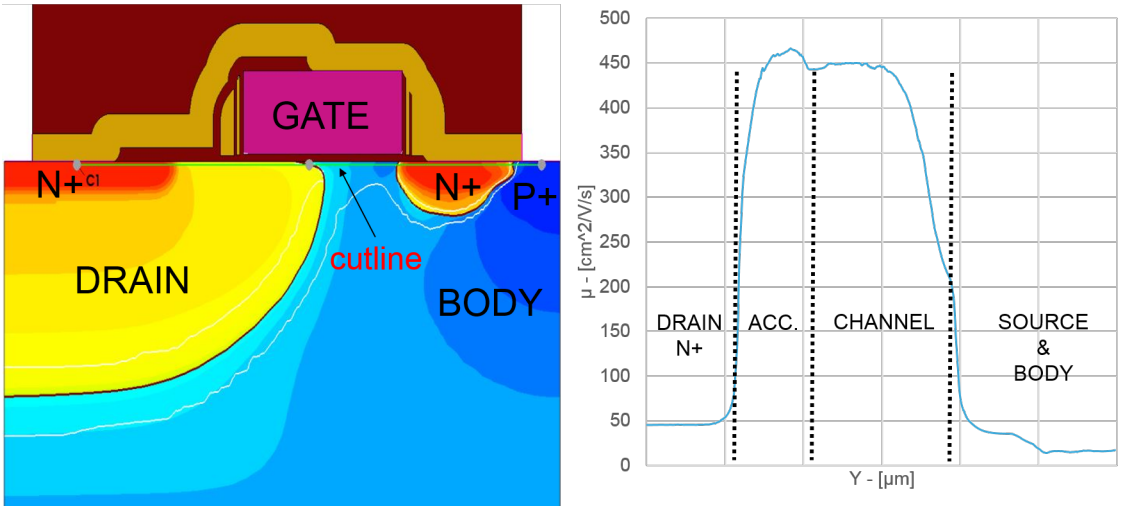


Figure 3.12: Channel mobility at the interface with the oxide.

It is also possible to compute the R_{L_S} term of the [3.8] simply deriving the equation [3.9] respect to the channel length.

$$R_{L_S} = \frac{\partial R_{ON} \cdot W}{\partial L} = \frac{1}{C_{ox} \cdot \mu_n \cdot (V_G - V_T)} \tag{3.10}$$

Substituting the previous results and the values expressed in table 3.2, we find:

$$R_{L_S} = 2.761\,\mathrm{k}\Omega \tag{3.11}$$

In the end, table 3.4 summarizes all the results we found highlighting the correctness of the theoretical model and the values we use.

	Theoretical	Simulated	Experimental
R_{L_S}	$2.761\,\mathrm{k}\Omega$	$2.616\,\mathrm{k}\Omega$	$2.718\,\mathrm{k}\Omega$

Table 3.4: Experimental vs Theoretical results

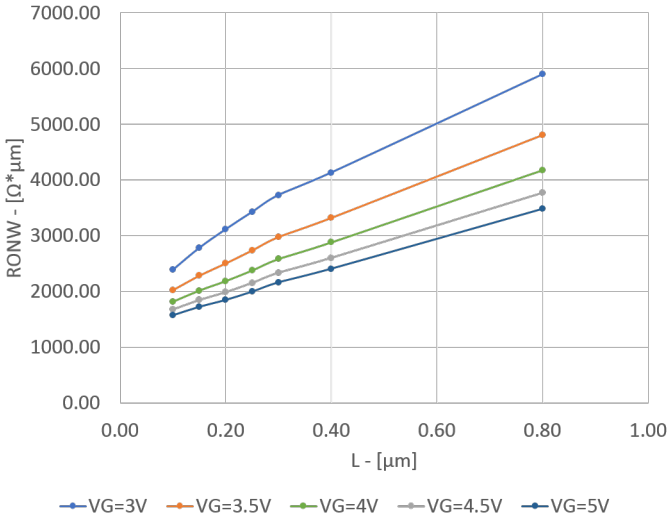


Figure 3.13: $R_{ON} \cdot W$ at different V_G.

Another important information that we can extract from the experimental curves concerns the variation of the electron mobility as a function of the gate voltage. Firstly, we plot the $R_{ON} \cdot W$ evaluated at different gate voltages (see figure 3.13) for the different structures. Then, we solve the equation [3.10] respect to μ_n for each different curve and finally we plot the results in a V_G-μ_n graph (figure 3.14). The results show a decreasing trend of the mobility that is mainly due to the growth of the scattering phenomena.[21]

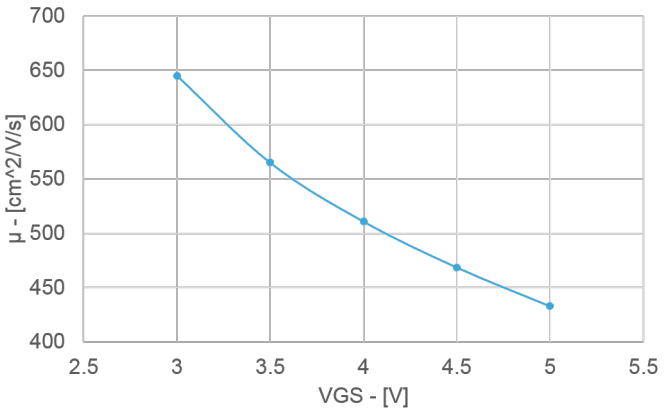


Figure 3.14: μ_n as a function of V_G.

3.2.2 Gate-Drain overlap region resistance

In figure 3.10, we call the contribution of the gate-drain overlap region to the total resistance R_I. The resistance of this region is determined more than by the fixed charge we introduce with the drain implantation, by the accumulated superficial layer. This accumulation of electrons is due to the capacitive effect induced by the gate that, in the measurement conditions, is biased to 5 V. For the description of the analytical model, we start reporting the second Ohm's law [3.12] that links the resistance to geometrical and physical parameters.

$$R = \rho \cdot \frac{I}{WT} = \frac{I}{WT \cdot q\mu n} \tag{3.12}$$

I, W, T are the physical dimensions of the region under analysis namely the length, the width, and the thickness or depth respectively, while ρ is a proportional coefficient called resistivity. From the semiconductor devices theory, the resistivity of the silicon can be expressed as a function of the carrier mobility, the elementary charge, and the number of carriers per unit volume n. All the MOSFET devices, included the LD-MOSFETs, are unipolar, i.e. the current is due to only electrons or holes. Particularly, the n-LD-MOSFETs are characterized by a current due to a flow of electrons from the drain region to the source one, while the p-LD-MOSFETs by a current of holes from the source region to the drain one. From now on, we will use 'carriers' and 'electrons' as synonyms and, for the same reason, every quantity we will compute or will mention must be consider as referred to the electrons, e.g. electron mobility, electron density... if not differently specified. As we did previously, we cancel the width since we always work with the resistance per unit width.

$$R \cdot W = \frac{I}{T \cdot q\mu n} \tag{3.13}$$

Generally, n is assumed to be equal to the concentration of doping impurities neglecting the intrinsic electrons concentration since the latter is several orders of magnitude smaller for standard operating conditions, i.e. room temperature and high doping level. In this case, instead, the accumulation of electrons forced by the gate can not be neglected. For this reason, we need to use for this region a more

complex mathematical model that, together with some sensible and simplifying assumption, allows us to obtain significant results.

Almost every model[21][23] of this region considers separately the superficial and accumulated layer from the deeper and neutral layer. For greater clearness, in the following, we refer to these two regions with the subscripts 'acc' and 'neu' those stand for 'accumulated' and 'neutral' respectively. From an electrical point of view, this system, so divided, can be seen as the parallel of two different resistors. The resistance of the neutral region can be estimated with the [3.12] since that region is not subjected by the coupling effect of the gate electrode and it can be assimilated to a neutral piece of doped silicon. Of course, each variable of the [3.12] must be referred to that layer and so the carriers concentration becomes the average doping level of that region and so on.

$$R_{neu} \cdot W = \frac{I}{T_{neu} \cdot q\mu_{neu} N_{neu}} \tag{3.14}$$

The resistance of the other layer can be also estimated with the [3.12] with the shrewdness to consider all variables referred to that layer and to include the electrons due to the accumulation. We can add the accumulation charge as a simple additive term.

$$R_{acc} \cdot W = \frac{I}{q\mu_{acc} T_{acc} \cdot (n_{fix} + n_{acc})} \tag{3.15}$$

where, T_{acc} is the thickness of the superficial layer, n_{fix} and n_{acc} are respectively the electrons induced by the fixed charges and by the gate coupling. This last term can be evaluated very easily considering the fundamental relation of a planar capacitor that links the charge to the applied voltage through the capacity.

$$n_{acc} = \frac{C_{ox}(V_G - V_{FB})}{q \cdot T_{acc}} \tag{3.16}$$

Substituting the [3.16] into the [3.15], we obtain:

$$R_{acc} \cdot W = \frac{I}{\mu_{acc} \cdot (qT_{acc}N_{acc} + C_{ox} \cdot (V_G - V_{FB}))} \tag{3.17}$$

The parallel between the [3.17] and the [3.14] is the resistance we are looking for. A consideration can be made on the fact that the equation [3.17] returns to be the [3.12] if there is no capacity coupling ($C_{ox} = 0$) or the gate bias is the one that brings the MOS system to the flat band condition ($V_G = F_{FB}$).

$$R_I \cdot W = \frac{I}{\mu_{acc} \cdot (qT_{acc}N_{acc} + C_{ox} \cdot (V_G - V_{FB})) + T_{neu} \cdot q\mu_{neu}N_{neu}} \tag{3.18}$$

As done for the channel resistance, also here we have to make some clarifications about the terms of the so developed model. The only well-defined variables are the gate capacitance and the gate voltage, while all other variables are still unknown. Among them, the thickness of the accumulated layer is the first thing to define to separate the two layers and start to analyze them individually. From the semiconductor theory, it is known that the accumulation of majority carriers as a function of the spatial coordinate decays with an exponential trend ruled by the Debye length (L_D). Due to this high-varying trend, the electrons density is reduced by a factor e after only a single Debye length, after few of them the accumulation charge becomes negligible respect to the fixed charge and can be safely neglected. Figure 3.15 shows the electrons density (blue curve) and the doping concentration (orange curve) along a cutline near the centre of the I region as shown in the cross-section on the left side. At the oxide-silicon interface, the electrons density due to the accumulation has a peak of some orders of magnitude larger than the fixed charge. According to the TCAD, the width of the peak is around 15 nm. The Debye length assuming room temperature and an average value of the superficial doping profile is, instead, equal to

$$L_D = \sqrt{\frac{\epsilon_0\epsilon_{ox}KT}{N_{acc}q^2}}\Bigg|_{N_{acc}=1.3\times10^{17}\,\mathrm{cm}^{-1}} \approx 6.6\,\mathrm{nm}$$

Comparing the Debye length and the electrons distribution obtained with the simulator, we have a confirmation of what said when we introduced the Debye length. Now we can affirm that $2L_D$ are enough to make negligible the accumulation term. So, we fix the thickness of the accumulation layer to two times the Debye length.

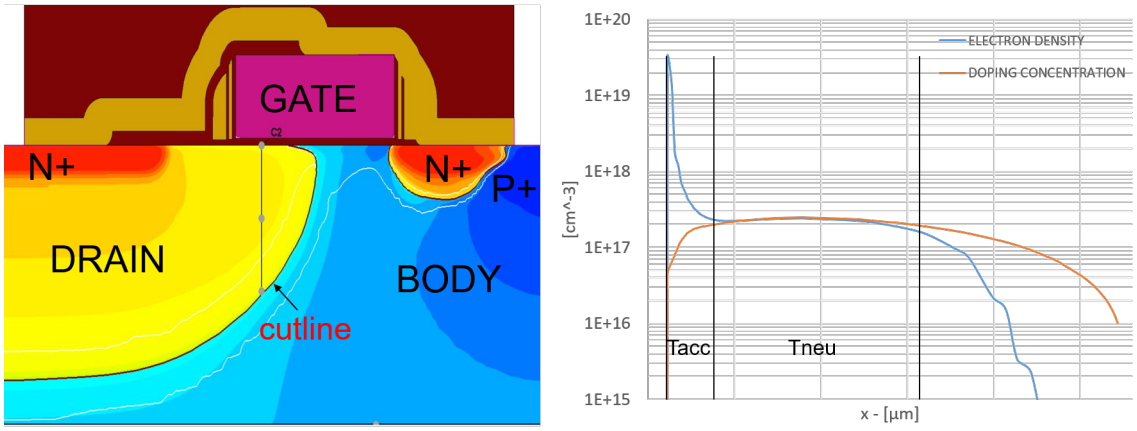


Figure 3.15: Drain doping distribution and electron density of the I region.

Moving on along the cutline, after the superficial and accumulated layer, there is the neutral layer where the electrons density returns equal to the fixed charge as for any doped semiconductor. Finally, for completeness, there is the depletion layer of the junction between the drain and the body where the electrons density becomes ideally null or very small.

Regarding the other terms of the 3.18, the flat band voltage can be derived assuming the same average doping level just used to evaluate the Debye length. The doping level of the neutral layer can be derived averaging the result of the simulation and with it also the mobility can be extracted with the model discussed in [24]. As before, I has a nominal value determined by the intersections of the drain and gate masks and actual value that depends on the various diffusions. Finally, the mobility in an accumulated layer and the thickness of the neutral layer are the only variables still unknown. The former, like the mobility in an inversion layer, is strongly dependent on the effective electric field or, equivalently, the gate bias and can not be approximated with the mobility computed for only doped silicon. The latter, instead, is determined by the edges of the accumulation layer and the depletion layer of the deep junction. The first one is approximately constant while the second one changes from zero at the interface with the channel to a few hundreds of nanometers according to the position where it is evaluated. A reasonable approximation of this depth can be computed as follows. Let's write the expression

of the resistivity used when there is a non constant doping profile along the depth and the length.

$$R \cdot W = \frac{I^2}{q \cdot \int_0^I \int_0^T \mu(x,y) \cdot n(x,y) dxdy}$$

The doping profile and consequently the mobility can be assumed constant basing on the simulation results and they can hence bring outside the integrals.

$$R \cdot W = \frac{I^2}{q\mu n \cdot \int_0^I \int_0^T dxdy}$$

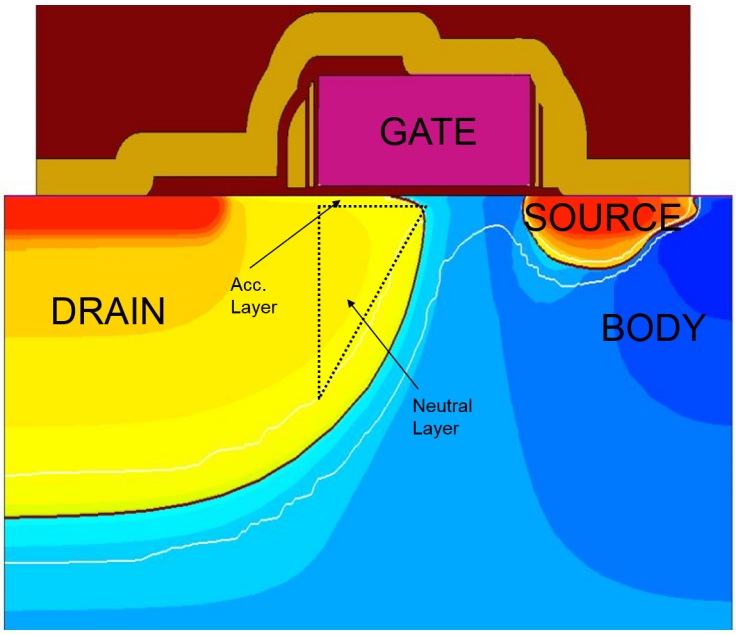


Figure 3.16: Geometrical approximation of the I region.

At this point, the integrals is equal to the area of the neutral region. It is possible to approximate its shape with a triangle as shown in figure 3.16. Then, we can use the mean value theorem for integrals to find an equivalent or effective thickness that we can use in the [3.18].

$$\int_0^I \int_0^T dxdy = \frac{I \cdot T_{max}}{2} = I \cdot \frac{T_{max}}{2} = I \cdot T_{eff}$$

We can hence rewrite the [3.18] in the following way:

$$R_I \cdot W = \frac{I + \Delta I}{\mu_{acc} \cdot (qT_{acc}N_{acc} + C_{ox} \cdot (V_G - V_{FB})) + T_{eff} \cdot q\mu_{neu}N_{neu}} + K \qquad (3.19)$$

Where we have added also the term ΔI to take care of the variation of the I length due to diffusions and the term K to consider all contributions that do not depend on this region. Further consideration can be done on the accumulation charge since it is perfectly in agreement with what computed by the simulator. Solving the [3.16] substituting the numbers we have already defined and that are reported in table 3.5 for clearness, we obtain

Variable	Value
L_D	$6.6\,\mathrm{nm}$
T_{acc}	$13.2\,\mathrm{nm}$
N_{acc}	$1.3 \times 10^{17}\,\mathrm{cm}^3$
V_G	$5\,\mathrm{V}$
V_{FB}	$-0.15\,\mathrm{V}$
C_{ox}	$2.66 \times 10^{-7}\,\mathrm{F\,cm}^{-2}$
N_{neu}	$2.3 \times 10^{17}\,\mathrm{cm}^3$
$T_{neu_{max}}$	$257\,\mathrm{nm}$
T_{eff}	$128.5\,\mathrm{nm}$
μ_{neu}	$529.5\,\mathrm{cm}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$

Table 3.5: Variables values.

$$n_{acc} = \frac{C_{ox}(V_G - V_{FB})}{q \cdot T_{acc}} \approx 6.5 \times 10^{18}\,\mathrm{cm}^{-3}$$

This result can be obtained also computing the integral average of the electrons density distribution reported in figure 3.15.

$$n_{acc} = \frac{\int_{x_1}^{x_2} n(x)dx}{x_2 - x_1} \approx 6.6 \times 10^{18}\,\mathrm{cm}^{-3}$$

Now, we can extract the value of the unknown mobility from the experimental

results. Figure 3.17 shows the trend of the $R_{ON} \cdot W$ as a function of the length of the I region. Regarding the left decreasing section, we note how the reduction of the I, makes the resistance bigger; for this architecture, this trend seems in contrast with the second Ohm's law that states that the resistance opposed by a semiconductor is directly proportional to its length. At shorter and shorter I, the $R_{ON} \cdot W$ increases fast due to the shape of the equipotential lines until the channel is electrically disconnected. For this reason, we neglect such part and we focus on only the right section. The linear fit of this part is also shown in the graph.

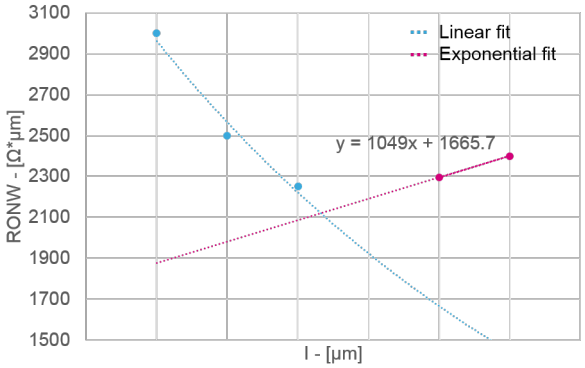


Figure 3.17: $R_{ON} \cdot W$ as a function of the I length.

Repeating the same procedure used in the previous paragraph, we can write two equations, one for each point we have. Solving them and averaging their results, we find the following value for the electron mobility.

$$\Delta_I = 10\,\text{nm} \qquad K = 1665.7\,\Omega$$
$$\mu_{acc} = 468.6951\,\text{cm}^2\,\text{V}^{-1}\,\text{s}^{-1}$$

The term K is chosen as the intercept of the best fitting line as already done in the previous analysis, while Δ_I was fixed to the average value we have obtained for the variation of the channel length with, of course, opposite sign.

Now, it is possible to compute also the slope of the linear dependence on the I

substituting all values we have discussed into the [3.20] that is obtained deriving the [3.18].

$$R_{S_I} \cdot W = \frac{1}{\mu_{acc} \cdot (qT_{acc}N_{acc} + C_{ox} \cdot (V_G - V_{FB})) + T_{neu} \cdot q\mu_{neu}N_{neu}} = 1106.4\,\Omega \tag{3.20}$$

From figure 3.12, we can extract the accumulation mobility computed by the simulator that is practically the same than the one we computed (See table 3.6).

	Theoretical	Simulated
μ	$469\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$	$\approx 460\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$

Table 3.6: Experimental vs Theoretical results

Finally, it can be interesting to understand what is the relative effect of the neutral layer respect to the accumulation one. With this purpose, we can solve the [3.18] respect to the mobility as we have done before, this time there is no need to make any assumption regarding the thickness of the neutral layer.

$$\mu_{acc} = 661.2316\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$$

The result is of course larger than the one we obtained before, the error due to the simplification is about 40%. This is also compatible with what stated in [21] where in the model a factor two is added to the denominator since the contribution of the neutral layer is estimate as 50%. Moreover, the error is a function of the drain doping level. The dependence on the latter is obvious looking the equations and can be studied more in details plotting the various contribution as a function of the drain doping. In figure 3.18, there are shown four different curves. The flat blue line is the resistance of the accumulation region that, as already explained, practically does not depend on the doping distribution. The red decreasing curve, instead, is the resistance of the neutral layer that, as for any piece of doped silicon, shows a strong dependence on doping. The parallel between the previous two curves is the sheet resistance and it is drawn in grey. As expected, when the doping level is small,

the accumulation charge dominates and the other contribution can be completely neglected; on the contrary, for high doping level, who dominates is the contribution of the neutral part. For doping layer in between both contribution must be considered to achieve significant results. Finally, the yellow curve, plotted in the right axis, indicates the relative error that we compute using the aforementioned simplification. In conclusion, the contribution of the neutral part can not be neglected.

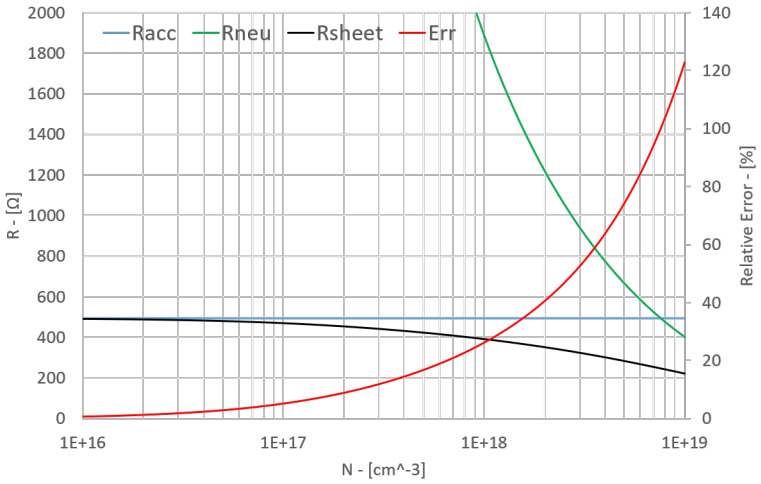


Figure 3.18: Resistance contributions as a function of the doping concentration.

3.2.3 Drift Region Resistance

The contribution to the total resistance of the drift region is called R_X in figure 3.10. In the first analysis and neglecting the border effect of the gate contact, this region can be seen as a simple piece of doped silicon whose resistivity, and so the resistance, depends only on the doping distribution and on its dimension. The analytical expression of its sheet resistance per unit width is

$$R_X \cdot W = \frac{X}{q \cdot \mu \cdot N \cdot T} \tag{3.21}$$

And the corresponding sheet resistance is

$$R_{S_X} \cdot W = \frac{1}{q \cdot \mu \cdot N \cdot T} \tag{3.22}$$

Where X is the distance between the n+ region and the gate left edge that, as for any other contribution, has a nominal value depending on the intersections of the drain mask with the gate and n+ masks and a real value depending on the diffusions of phosphorus or arsenic we have introduced with the n+ implantation. The n+ region is defined as the region with a so high doping concentration that the resistivity becomes similar to the one of a metal. T is the vertical section where the current flows and so the distance between the silicon surface and the edge of the depletion layer of the junction with the p-layer. N is the average doping level of that drain part and μ the electron mobility.

However, this is true only if the current density is constant for the whole depth of the drain; we already know that the greatest part of the current is concentrated on the surface when exits from the accumulation region to enters in the drift region. There, the electrons starts spreading since there are no reasons to remain confined, no electrostatic barriers or capacitive coupling. As it is possible to observe in figure 3.19, the current density is different according to how far from the gate we are. To understand better, this analysis is made on a structure with longer X, all the other things are instead equal to the POR structure we have used till now. The cutline *C1* summarizes how the current is distribuited when it exits from the accumulated region I. Moving toward the drain to the second cutline, it is possible to appreciate

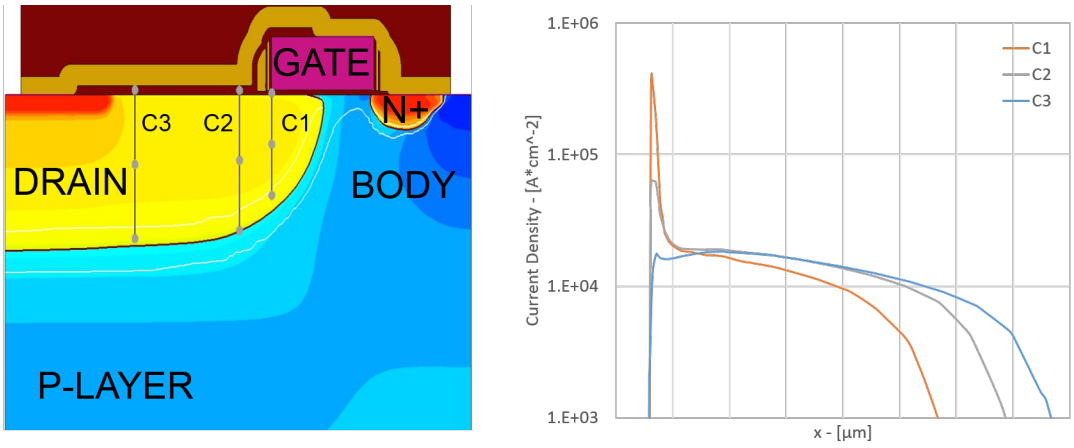


Figure 3.19: Current density as a function of the depth at different distance from the gate.

how the peak is reduced while the current spreads inside the drain. At a certain distance from the gate the current flows can be considered constant. So the previous model works only in this far sector, while in the first one we need to consider also the superficial accumulation.

The experimental and simulation results were obtained as already done and explained in the previous paragraphs. Figure 3.20 shows those results in a R_{ON} vs X graph. The orange and blue lines are referred respectively to the experimental data and to the simulation results; the two dotted lines are instead the best linear fitting whose slope is the sought sheet resistance. It is possible to see that the simulator and the experimental data are almost perfectly in agreement. The sheet resistance is the same while the small difference in the absolute value can be due to the offset of the measuring setup.

Again, we can model this region dividing the superficial layer from the deeper one. Concerning the deep layer, it is possible to evaluate its resistivity considering the average value of the doping profile that can be extracted from the simulation as well as the thickness of this layer. In this case this latter can be assumed constant and equal to the depth that the drain has farther from the gate, there the thickness is quite smaller but also the electrons are still concentrated on the surface. The

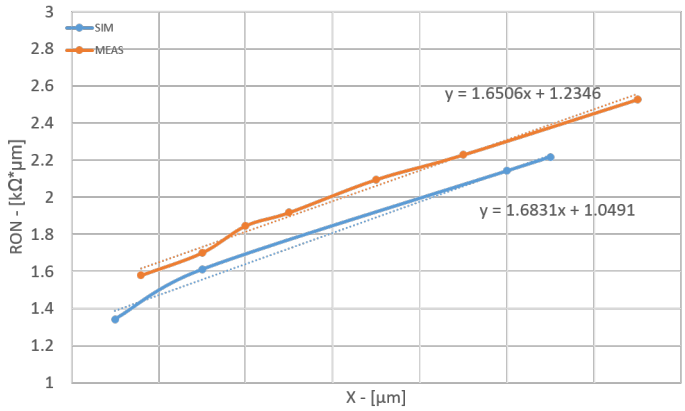


Figure 3.20: ON resistance function of the drift region extension.

mobility is again evaluated with the model described in [24].

$$R_{deep} = \frac{1}{q \cdot \mu \cdot N_{deep} \cdot T_{deep}} \tag{3.23}$$

Concerning the superficial layer, as we did before, we can consider the electron density distribution that has a non-negligible contribution together with the fixed charge. Since more of these following steps have been already described in the previous analysis, here they will be only reported or mentioned. Firstly, we can use the same accumulation charge density to evaluate all the electrons that enter at the right edge of the superficial drift region.

$$N_{acc} = \frac{C_{ox}(V_{GS} - V_{FB})}{q \cdot T_{acc}} \approx 6.5 \times 10^{18}\,\mathrm{cm}^{-3}$$

The accumulation of electrons spreads quickly into the drain and since the electrons are the majoritarians we can use again an exponential decay ruled by the Debye length to describe its reduction as we move farther from the gate. it is also used to evaluate the thickness of this superficial layer that is again fixed as twice the Debye length.

$$N_{sup}(x) = N_{acc} \exp^{\frac{x-X}{L_D}} + N_{fix}$$

Again, we can use the integral mean value theorem to extract its average value.

$$N_{sup_{avg}} = \frac{\int_0^X N_{sup} dx}{X}$$

Finally, we can use this value to evaluate the resistance of this superficial layer. The parallel between the sheet resistances of the two layers is the overall sheet resistance we are looking for. Substituting into the [3.24] all the variables we have just discussed and that are reported in table 3.7 for clearness, we find

Variable	Value
L_D	$6.6\,\text{nm}$
T_{acc}	$13.2\,\text{nm}$
N_{fix}	$1.3 \times 10^{17}\,\text{cm}^3$
V_G	$5\,\text{V}$
V_{FB}	$-0.15\,\text{V}$
C_{ox}	$2.66 \times 10^{-7}\,\text{F cm}^{-2}$
N_{deep}	$1.8 \times 10^{17}\,\text{cm}^3$
T_{deep}	$336.8\,\text{nm}$
μ_{deep}	$597\,\text{cm}^2\,\text{V}^{-1}\,\text{s}^{-1}$

Table 3.7: Variables values.

$$R_{S_X} = \frac{1}{q \cdot \mu \cdot N_{deep} \cdot T_{deep} + q \cdot \mu \cdot N_{sup_{avg}} \cdot T_{acc}} = 1694.0\,\Omega \tag{3.24}$$

In table 3.8, we report a comparison of the results we obtain.

	Theoretical	Simulated	Measured
R_{S_X}	$1694.0\,\Omega$	$\approx 1683.1\,\Omega$	$\approx 1650.6\,\Omega$

Table 3.8: Experimental vs Theoretical vs Simulated results

At the end, a summary of this chapter on the resistance can be made listing the experimental sheet resistances for the various contribution (see table 3.9) and trying to use them to estimate the resistance per unit width of, for example, the structure we have described at the beginning of this chapter. It is important to note that the largest contribution is due to the channel that has a relative weight also larger than the drift region, while the smallest one is due, of course, to the accumulation region. Finally, we can extract the weight of the offset term we have added into the [3.8]

solving it for the structure we have defined as our starting point. As expected, its contribution is very small but not completely negligible.

	R_{S_L}	R_{S_I}	R_{S_X}
EXPERIMENTAL	$2718\,\Omega$	$\approx 1049\,\Omega$	$\approx 1651\,\Omega$

Table 3.9: Experimental contributions for each active region to the overall resistance.

$$R_{ON} \cdot W - R_{off} = R_{L_S} \cdot L + R_{I_S} \cdot I + R_{X_S} \cdot X + R_{off} = 1.8266\,\mathrm{k}\Omega\,\mu\mathrm{m}$$

$$R_{off} = 0.019\,\mathrm{k}\Omega\,\mu\mathrm{m} \approx 20\,\Omega$$

3.3 Breakdown Voltage

Breakdown voltage is, together with the ON-resistance, the most important electrical parameter for a power transistor.[15] In any datasheet, there are reported two indications about the voltage limitations: the Maximum Operating Voltage (MOV) and the Absolute Maximum Rating (AMR). The MOV, as the acronym says itself, represents the voltage class of the device, i.e. the maximum voltage drop between the drain and the source with which the transistor work properly and the degradation of the aforementioned parameters is negligible or expected by accurate analytical models. The AMR, instead, is the maximum voltage drop between the drain and source allowed but only for a very short time. It is not fixed but it is often agreed between the technologies developers and the systems designers; the latter ones know which voltage spikes the transistor must support and, based on their indications, the former ones try to optimize the technology. If the transistor works with voltages greater than the MOV for a long time or even greater then the AMR, it starts degrading its electrical performances or it breaks.

The breakdown is defined as a big increment of the current due to an uncontrolled generation of electrons-holes pairs due to the avalanche mechanism. The high current, in turn, generates an important self-heating. Until the transistor does not change its structure due to the high heat, the breakdown phenomenon is recoverable also if with degraded performances. This kind of breakdown mechanism is common for p-n junctions or doped silicon itself[3]. Another more critical and always unrecoverable breakdown mechanism is the perforation of an insulation layer and the consequent formation of a low resistance path inside it. The voltage stress at which the first breakdown mechanism is reached at a certain point of the device is the breakdown voltage (BV).

The primary role of the engineer, under this aspect, is to ensure that the BV is in any case greater than the AMR you look for. Let's consider, for example, to want to realize an n-drift LDMOSFET with the following requirements: MOV and AMR

[3]The silicon can bear an electric field until some tens of kV per centimetre before the avalanche breakdown.

equal to 10V and 12V respectively. At the design stage, the AMR and the MOV are very little relevant, we need, instead, to know the minimum BV that this device should have to ensure the previously mentioned AMR and to choose consequently the correct integration solutions. The minimum BV can be computed starting from the AMR and considering two fundamental aspects.

- Voltage de-rating in temperature. The device will be certified within a certain temperature range, let's consider for this example the standard range for commercial devices: $[-40\,°\text{C} \div 125\,°\text{C}]$. The worst condition for the breakdown is at the lower temperature edge. Indeed, the energy with which the electrons bump themselves is higher since the mean free path increases as the temperature decreases. As a consequence, the breakdown anticipates. Therefore, the BV variation due to the temperature can be considered introducing a percentage factor computed simply multiplying the maximum temperature variation from the room temperature and a thermal coefficient. The thermal coefficient can be measured and verified each time with experimental measurements at different temperature. For this example, we use an empirical coefficient.

$$\frac{\partial BV}{\partial T} \cdot (T_{worst} - T_{env}) = -0.1\frac{\%}{°\text{C}} \cdot (-40 - 28) = 6.8\%$$

- Process variabilities. This contribution is much difficult to be evaluated so we introduce an arbitrary percentage factor equal to 10%. Again, this percentage factor comes out from the experience on many older or similar devices.

In the end, the minimum breakdown voltage should be:

$$BV_{OFF_{min}} = 12V \cdot (1 + 6.8\% + 10\%) \approx 14V$$

We have already discussed the trade-off between R_{ON} and area, so, here, we start to introduce a second trade-off between R_{ON} and the breakdown voltage when the device is in the OFF working condition (BV_{OFF}), trade-off that will escort us for the remaining of this work. As we will see better in the next chapters, generally the changes intended to improve the BV_{OFF} have the negative effect to worse the R_{ON}. This means that to target the optimal structure we have to design the device, not

with the best possible BV_{OFF} but with the best R_{ON} and maintaining the BV_{OFF} over a certain minimum value.

3.3.1 OFF Breakdown

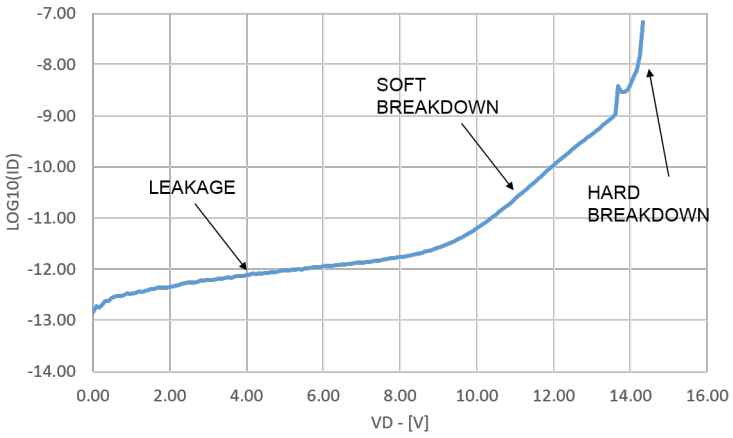


Figure 3.21: Output characteristic ($@V_G = 0V$).

The OFF-breakdown (BV_{OFF}) is the maximum V_D allowed when the transistor is OFF, i.e. with null gate bias. Its measure is destructive for the structure since we have to ramp only the drain electrode until the device breaks. Generally, to try to save the device, we use a feature of the parametric analyzer that stops the measurement when the drain current reaches a certain value that we choose according to the rule of thumb of one nano-ampere for each micron of transistor width (The limit for which a transistor can be considered OFF). Figure 3.21 shows the experimental output characteristic, plotted in a semi-logarithmic scale. That output characteristic has three well-distinguish regions:

- For low drain bias, what we see is the leakage of the transistor for standard operating condition and it shows a small dependence on the drain voltage. The leakage depends on several contributions such as the reverse current of the junction and the sub-threshold current.

- For intermediate drain voltages, what we see is a linear increment of the current in the semi-logarithmic graph. This exponential growth is due to the gate induced drain leakage (GIDL) phenomenon where the gate, increasing locally the electric field, raise the leakage of the drain-body junction. This kind of breakdown is often called 'soft breakdown' since it is not destructive.

- At the BV, the current increases with an almost infinite slope until to return to zero when the transistor breaks definitely. As we said, the measurement is automatically stopped when the current reaches the value of 80 nA (the width of the transistor is indeed 80 µm), so, the physical breakage does not occur and it is not visible. This breakdown is so often defined as 'hard breakdown' since it is definitive.

At this point, it is necessary to clarify a concept that, as presented until now, could be misleading. The BV_{OFF}, from an engineering point of view, is defined as the voltage drop, in OFF-state, that allows a leakage of $1\,\mathrm{nA\,\mu m^{-1}}$. In most cases, it is coincident with the hard breakdown while in others where GIDL is important, the BV_{OFF} can intercept the soft breakdown.

Besides the BV_{OFF} value and the shape of the output characteristic, another important information concerning where the device breaks must be found. The LD-MOSFET architecture has three weak points: just below the drain spacer, at the player-drain junction and near the edge of the n+ region (see figure 3.22).[15][25] They are defined weak because there the electric field has a local maximum. The increment of the electric field due to the growth of the voltage stress is different at the various spot and depends strongly on technological and geometrical parameters, but, anyway, the first spot that reaches the critical value of the electric field causes the avalanche breakdown mechanism. In the following, we describe why the electric field has local peaks in those critical regions highlighting the relations between the peaks and the design parameters. This is generally very difficult since the problem is intrinsically bi-dimensional and all the electrostatic effects that act on the drain such as the gate coupling should be considered at the same time. Nevertheless, we aim to uncouple the weak drain spots and to provide for each one a simplified model that more qualitatively than quantitatively can answer to the previous question and clarify the effect that a process change has on the distribution of the electric field.

Let's consider, initially, the first weak point. There, the electric field, and particularly its x-component, has a peak, i.e. a local maximum. To better understand, let's assume the drain region as a rectangular with a constant doping profile (fig.

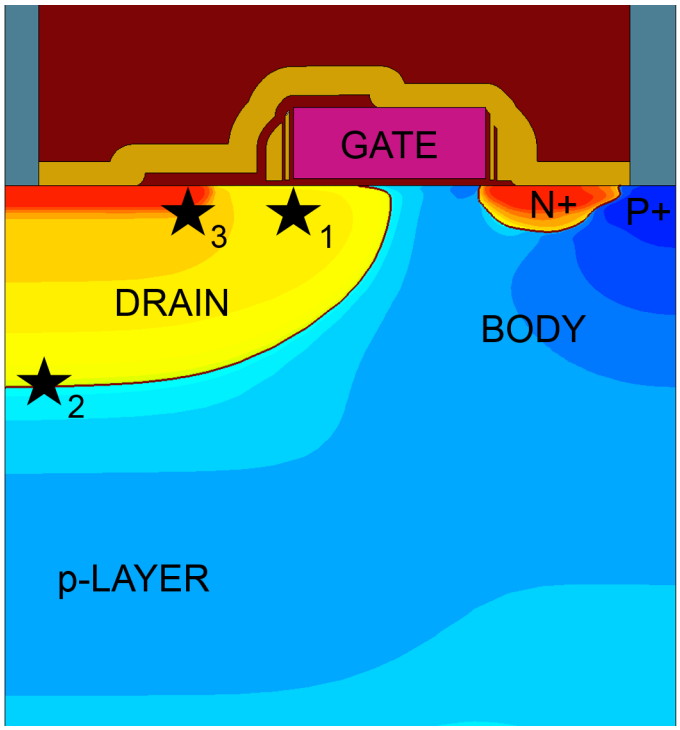


Figure 3.22: POR cross-section with the weak points highlighted.

3.23). The right and left coordinates are respectively the coordinates of the gate left edge and of the n+ drain region. To measure the BV_{OFF}, we apply voltage stress at the drain contact or, equivalently, at the left edge of our simplified model. Doing that, we have assumed that there are a negligible voltage drops on the resistances characterizing the very highly-doped regions. If we consider only the top side of the drift region of the drain without including the gate-drain overlap, the total electric field can be assumed as determined by only the tangential component and the problem can be faced exploiting a 1D model.[25] We can use so the 1D Poisson's relations between charge density, electric field, and electrostatic potential, to approximate the electric field distribution at the interface with the oxide and to estimate the its peak value. To solve this second order differential system we need also two consistent boundary conditions, i.e. the potential at the two sides of our model. The potential in 0 can be assumed as the voltage stress itself, the potential in X, instead, can be assumed as zero to simplify the computations. Actually, according to the simulator, this is not completely true since the greatest part of the voltage stress drops in the

drift region, while the small remaining part drops on the I region.

$$\begin{cases} \phi(x) = \phi(0) - \int_0^x E(x)dx \\ E(x) = E(0) + \int_0^x \frac{\rho(x)}{\epsilon_s}dx \\ \phi(0) = V_D \\ \phi(X) = 0 \end{cases}$$

$$E(x) = \frac{V_D}{X} - \frac{qN}{2\epsilon_s}X + \frac{qN}{\epsilon_s}x$$

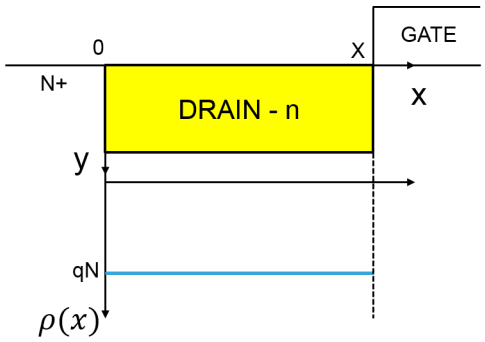


Figure 3.23: Simplified and schematic model of the drift region.

Its maximum value is achieved exactly below the spacer and it is equal to

$$E_{MAX} = \frac{V_D}{X} + \frac{qN}{2\epsilon_s}X \tag{3.25}$$

Moving further along the same cutline inside the I region, the electric field starts decreasing due to the electrostatic effect of the gate electrode that acts as a field plate (see chapter 4). Then it has a new peak at the drain-body and at the body-source junctions. Figure 3.24 shows the absolute value of the electric field just below the interface with the oxide at the breakdown, it confirms what just said and reports some real numbers.

Solving the equation 3.25 respect to the V_D, it is possible to estimate the breakdown voltage since all other variables are known. The X dimension is fixed not

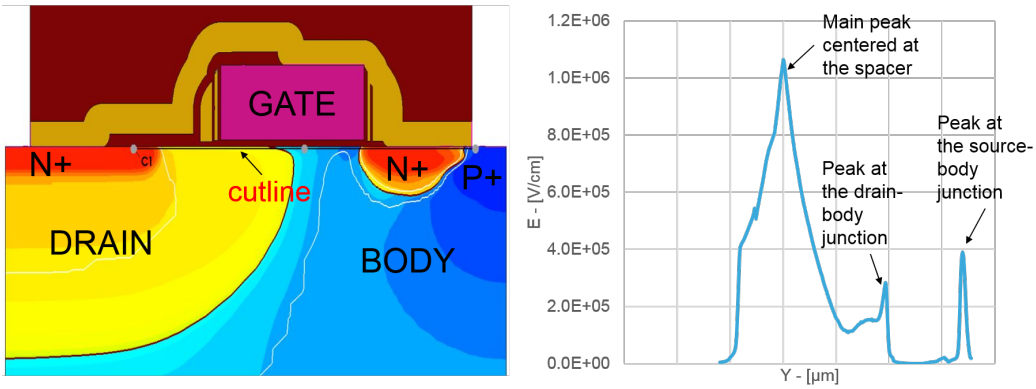


Figure 3.24: Absolute electric field at the breakdown condition.

to its nominal value but to the extension of the depletion region that is the real and meaningful physical quantity since the voltage stress drops only inside it. The doping level is chosen as the average value of the doping distribution of the drain extension region. Finally, the critical electric field can be extracted from the simulation as well as the doping distribution and the width of the depletion region. The used values are reported in table 3.10.

$$BV = 14.15\,\mathrm{V}$$

Variable	Value
X_{dep}	$160\,\mathrm{nm}$
E_C	$1.09 \times 10^6\,\mathrm{V\,cm^{-1}}$
N	$1.66 \times 10^{17}\,\mathrm{cm^3}$

Table 3.10: Variables values.

The summary of the results obtained from the simulation, the measurement and the analytical model are reported in table 3.11.

	MEASURED	THEORETICAL	SIMULATED
BV_{OFF}	14.3 V	14.15 V	13.95 V

Table 3.11: Results comparison.

Let's consider, now, the second weak point placed exactly at the junction between the drain and the p-layer. There the electric field has again a local maximum whose value depends mainly on the doping distribution of the two sides of the junction and on the applied bias. We can adopt a scheme similar to the one that we have already used for the previous analysis to extract a simplified model for this breakdown. The scheme is so changed adding a second domain representing the p-layer region characterized by certain constant doping layer (see figure 3.25). The electric field, considering an abrupt junction, along the vertical axis has a triangular shape and it is defined by the following set of equations.

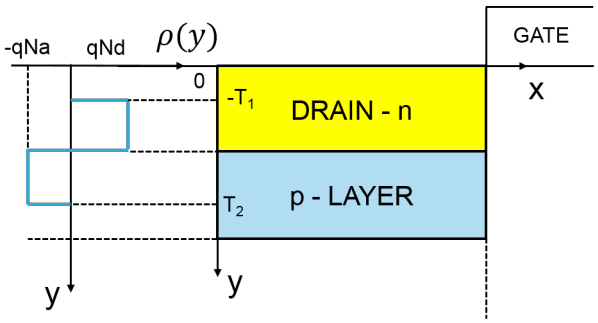


Figure 3.25: Simplified model of the drift region - 2nd version.

$$\begin{cases} \frac{qN_D}{\epsilon_s}(y + T_1 & -T_1 < y < 0 \\ \frac{qN_D}{\epsilon_s}T_1 - \frac{qN_A}{\epsilon_s}y & 0 < y < T_2 \\ 0 \quad \text{elsewhere} \end{cases}$$

Where T_1 and T_2 are the edges of the depletion layer of the p-n junction.

$$E_{MAX} = \frac{2(\Phi_i + V_D)}{T_1 + T_2} \tag{3.26}$$

Figure 3.26 shows the simulated absolute electric field and the doping distribution along the drawn cutline. As it is possible to see, the doping level changes of more than a decade, so the assumption of abrupt junction with constant doping distributions at both sides can not be considered valid. Moreover, the electric field has a shape that is more similar to a bell than a triangle. In conclusion, if the model

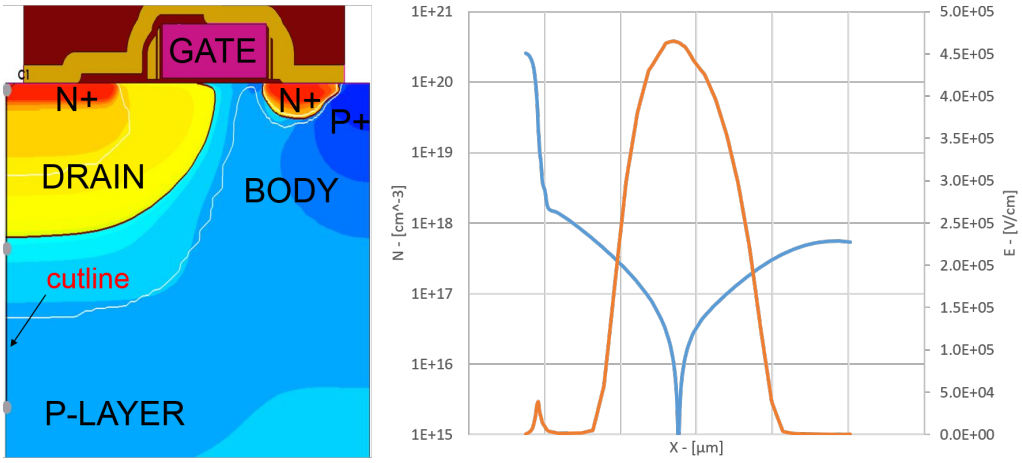


Figure 3.26: Doping concentration and absolute electric field vs x.

we have extracted for the first weak point can provide quantitative and reasonable results besides a qualitative description of the electric field distribution, the simplified model for the vertical junction can only provide a qualitative description. To obtain significant results it is possible to modify the model to adapt it once we know the simulation results. Firstly, it is possible to neglect the Φ_i contribution inside the [3.26] since the voltage breakdown is much higher than the intrinsic drop of a junction.

$$E_{MAX} = \frac{2V_D}{T_1 + T_2} \tag{3.27}$$

Then, instead of considering the drain voltage as the area of a triangle, it is possible to consider the area of a parabola. Using the Archimede's theorem, we find

$$E_{MAX} = \frac{3V_D}{2(T_1 + T_2)} \tag{3.28}$$

Variable	Value
$T_1 + T_2$	390 nm
V_D	13.95 V

Table 3.12: Variables values.

Now, we can extract the width of the depletion region and the drain bias at

the breakdown from the simulation to evaluate the maximum electric field at the junction. The values of the variables are reported in table 3.12 while the simulation and analytical results are reported in table 3.13. The electric field at the vertical junction is smaller than the one we have considered for the first weak point and this result is a further proof that for this particular architecture the breakage occurs earlier at the spacer.

	THEORETICAL	SIMULATED
E_{max}	$5.4 \times 10^5\,\mathrm{V\,cm^{-1}}$	$4.7 \times 10^5\,\mathrm{V\,cm^{-1}}$

Table 3.13: Results comparison.

Finally, it is possible that the device breaks in the third weak point, when this happens, it is said that the transistor goes in reach-through. The electric peak starts to increase at the highly-doped region when the depletion region approaches it. Of course, the higher is the drain dose that we implant, the more difficult is that the structures start to be limited in this point. Moreover, to target the low-voltage classes at the center of the next chapters, we will never use so low drain doses since the resistance becomes soon no longer competitive. It can occur both toward the p-layer or toward the gate but, as said, in our structures is always hidden by the breakdown in the other two weak points.

Chapter 4

Reduced Surface Field (ReSURF) Effect

As told at the end of chapter 2, the drain engineering is a key factor to bring out the best from a specific device architecture. This means, in advanced power devices, to optimize the Reduced Surface Field effect (ReSURF) that allows to achieved better performance. To better understand what is the ReSURF, let's consider the 'all-in-active' architecture introduced in chapter 2 (Figure 2.5) and let us assume to not have a p-layer (i.e. a dedicate implant below the drain region) as a degrees of freedom of our drain optimization. As we have seen in the previous chapter, an architecture like this has three weak points. Not having the p-layer, the vertical junction is surely not critical. If the structure is limited at the n+ region, it is sufficient to increase the doping level of the drain to overcome this limitation, that increment, moreover, is always desirable since it lowers also the ON-resistance. However, we start soon to be limited by the electric field below the spacer. Figure 4.1 shows the impact ionization distribution at the breakdown condition. The impact ionization is the physical process in which an electron with enough kinetics energy hits another electron transferring enough energy to promote it to the conduction band. In the regions with a high electric field, there will be more electrons with enough energy to create other electron-hole pairs. Moreover, the just generated electrons and holes can, in turn, be accelerated by the same critical field and, hitting other bound electrons, create new carriers. When this happens, this phenomenon becomes self-sustaining, the number of carriers grows rapidly and uncontrollably exactly like an avalanche, hence the name of this breakdown mechanism. The impact ionization parameter is, therefore, a good figure of merit to visualize the regions of the structure where the electric field and the number of carriers are high enough to start the avalanche breakdown.

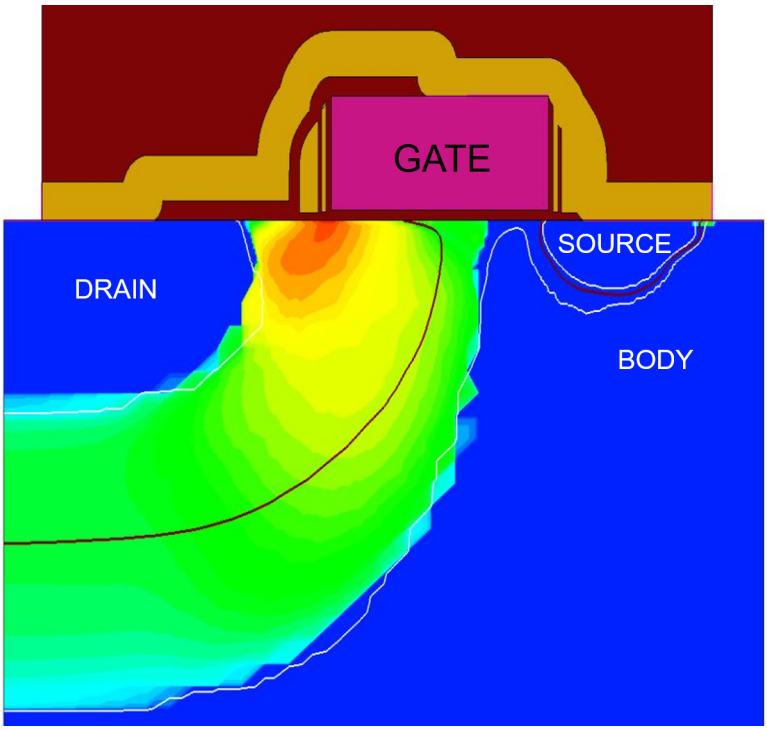


Figure 4.1: Impact ionization at the BV_{OFF}.

The picture shows that the point where the avalanche breakdown starts, is located below the spacer as expected. There, as we have seen in the previous chapter, the electric field peak is due mainly to the tangential component. In conclusion, we can not dope the drain too little since we are limited by the n+ region or we obtain a non-competitive resistance, but we can not dope the drain too much since we are limited at the spacer. We can optimize the drain tuning its implantation dose and energy to reach the best BV_{OFF}, but, for any choice, the voltage capabilities are heavily limited and it is very difficult to realize devices with competitive ON-resistances. Therefore, what we described raises new questions such as: does the drain dose that maximizes the breakdown voltage secure the best BV_{OFF}-R_{ON} trade-off for that geometry? Or do we have other ways to obtain higher breakdown voltage or smaller resistance? Before answering these questions and describing two possible architectural solutions, a better description of the electric field distribution and a comprehension of which distribution secures the maximum breakdown voltage is needed.

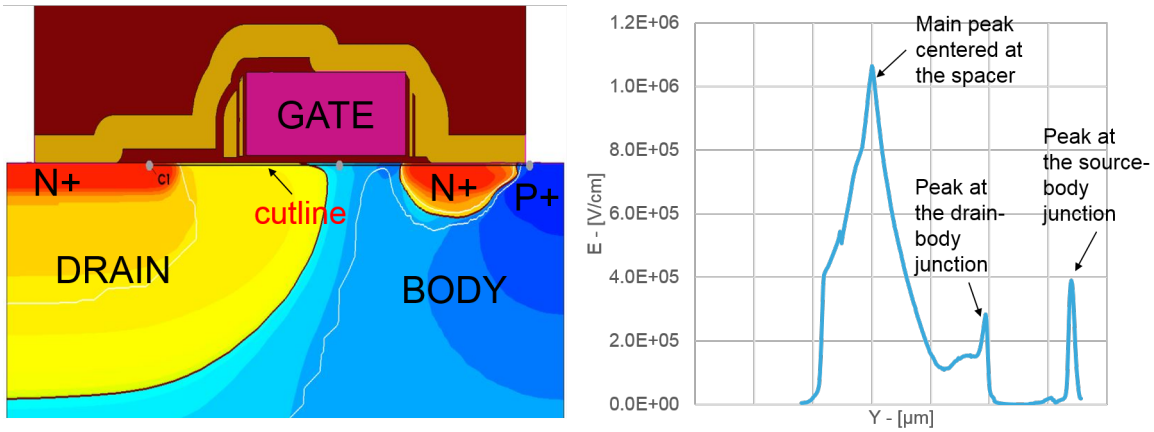


Figure 4.2: Absolute electric field at the breakdown condition.

As we have seen in the previous chapter, the electric field along a cutline just below the interface with the oxide has a triangular shape whose maximum value is reached exactly below the spacer (We report again in figure 4.2 the electric field distribution we are talking about). For the Poisson's equation, the area below the curve is the drain voltage stress so, if we want to enhance the voltage capabilities, we have to increase that area. The height of this triangle can not change since it is fixed to the maximum field that the silicon can bear that is constant if we neglect its variability as a function of the doping level. A possibility might be to enlarge the base, for example making longer X, but the breakdown voltage soon saturates and stops increasing as we will see shortly, moreover, the increment of the X worsens the performances since the ON-resistance increases. The last alternative is to find a way to modify the shape of the electric field distribution changing it from a triangular shape to a flat one. The constant field distribution is, in fact, the one that has the maximum area for a certain base and height. The answer to the first question is now only partially clear: we understand that there is, at least under a mathematical point of view, an ideal electric field distribution, i.e. the flat one, that allows maximizing the breakdown voltage for a fixed geometry but at this point, we can neither state if that distribution is feasible nor if exploiting it we can reduce further the resistance. Anyway, this ideal distribution corresponds to the

mathematical expression [4.1].[26]

$$\frac{\partial E(y)}{\partial y} = 0 \tag{4.1}$$

As we did in the previous chapter, the electric field inside the drift region can be approximated by applying the 1D Poisson's equation [4.2]. However, if we compare the [4.1] and the [4.2], it is clear that they are not compatible if we do not make the drain to be intrinsic. Therefore, our target seems not reachable and the answer to the second question should be negative.

$$\frac{\partial E(y)}{\partial y} = \frac{\rho(y)}{\epsilon_s} \tag{4.2}$$

To solve this apparently unsolvable issue the LD-MOSFET architectures uses massively the so-called Reduced Surface Field (ReSURF) effect. Advanced techniques that exploit the ReSURF principle are employed in advanced technologies and allow the integration of power devices with voltage classes also greater than 1000V.[27] The idea behind the ReSURF effect is very simple and, at the same time, very effective: to reduce the tangential electric field we can exploit a perpendicular one. The ReSURF is used to overcome the voltage limitation due to the peak at the spacer side and so move farther the breakdown voltage.

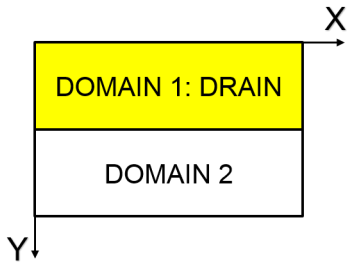


Figure 4.3: Generalization of the RESURF effect.

To explain, initially theoretically, the ReSURF effect, we need to consider a perpendicular component of the electric field. Let's assume, therefore, that we have a certain domain with which we can force inside the drain that component (see figure 4.3). Now the problem can no longer be treated as a monodimensional one but it becomes bidimensional, hence, we are forced to use a 2D Poisson's equation

to appropriately describe the electric field inside the drain.

$$\frac{\partial E(y)}{\partial y} + \frac{\partial E(x)}{\partial x} = \frac{\rho(x,y)}{\epsilon_s} \tag{4.3}$$

Now, the [4.3] and [4.1] are compatible and the optimal field distribution we are looking for exists, again at least under a mathematical point of view[26]. Particularly, if we manage to introduce a perpendicular field that is equal to the charge distribution divided by the silicon dielectric constant, the tangential electric field becomes automatically flat.

$$\frac{\partial E(x)}{\partial x} = \frac{\rho(x,y)}{\epsilon_s} \rightarrow \frac{\partial E(y)}{\partial y} = 0 \tag{4.4}$$

Having introduced this second domain, the absolute electric field distribution in the drain becomes a vector that can be decomposed into a tangential component E_x and a perpendicular one E_y. The tangential component is the only one that accelerates the electrons and gives them enough energy to start the avalanche breakdown at the spacer. As described before, the introduction and the increment of a perpendicular electric field component can reduce the tangential one. However, the perpendicular component can also become critical. If it increases too much, it can create new breakage conditions located somewhere else. In conclusion, the maximum breakdown voltage for a certain geometry is reached with an optimal combination of the tangential and perpendicular components of the electric field which assures that the breakdown is reached contemporarily in both directions.

4.1 Junction ReSURF

In the description of the ReSURF effect, there is a last open point regarding the second domain that is still undefined. The most used and simplest way to force a perpendicular field inside the drain is to exploit the so-called 'Junction RESURF'. It concerns the creation of a p-n junction by the implantation of a p-layer just below the drain. Therefore, we can now substitute that second domain with a p-layer and carry on the analysis (see figure 4.4). As we have already seen when talking about the breakdown at the vertical junction, in the depleted region of the junction there

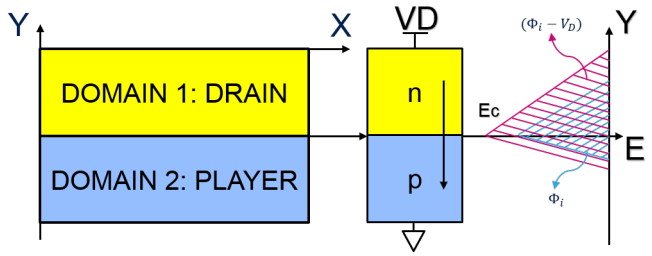


Figure 4.4: Generalization of the RESURF effect, modeling of the junction RESURF.

is an electric field directed toward the y-axis that has its maximum value exactly at the junction coordinate. Changing the implantation doses or energies of one or both regions, it is possible to tune the distribution of E_y and so the effectiveness of the ReSURF. Once we have defined the second domain, we can eventually clarify where the breakage condition is reached when the structure starts to be limited by E_y itself. If we try to force a stronger E_y field, the reverse-biased vertical p-n junction goes in the avalanche breakdown.

What we have introduced is called also 'Single ReSURF', because only two domains are involved: the first one is where we want to force an electric field and the second one is who forces that field. Advanced ReSURF techniques exploit the possibility to use more domains to increase further the BV. To better understand, let's assume that we cannot obtain the maximum effectiveness for our device because we are limited by the electric field at the vertical junction. Nobody prevents us to use a third domain to deplete the p-layer reducing consequently the electric field that limited us. Of course, this introduces a new junction and so a new limitation, but we can repeat the same strategy introducing a fourth domain that acts on the third one and so on. According to how many times we apply the ReSURF, the structure can be defined as 'Double ReSURF', 'Triple ReSURF'... Generalizing, the 'Junction ReSURF', independently on the number of domains involved, is a particular case of the category of the 'Periodic ReSURF'. Indeed, to reach the best BV, from a mathematical point of view, we can put an infinite number of domains alternating each time the doping type. Today, many high voltage devices use the 'Triple ReSURF'.[28][29][30] An example of the application of the double ReSURF can be

found in [31], where the drain depletion is induced also from the top of the drain by the implantation of very superficial p-rings.

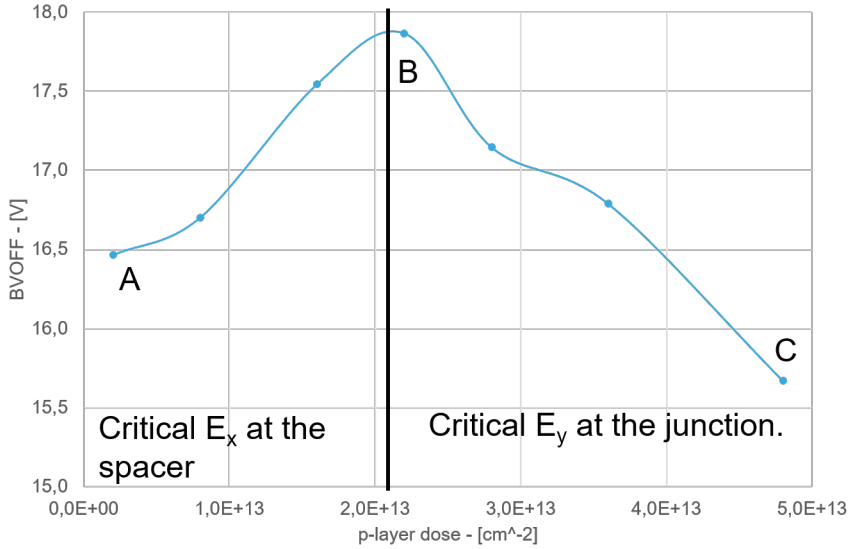


Figure 4.5: p-Layer ReSURF curve.

Up to now we have analyzed the ReSURF effect from a pure theoretical point of view. Now, our goal is to better describe the effect of the p-layer on the distributions of the electric field and the impact ionization through the simulations. The ReSURF curves have a bell shape like the one drawn in figure 4.5. In this qualitative graph, the BV_{OFF} of the device is drawn as a function of the p-layer implantation dose, two completely different regions can be identified.

- For low p-layer dose, the ReSURF effect is weak, in fact, the depletion region will extend almost only in the p-layer so the perpendicular field we can force inside the drain is small. As a consequence, the structure is still limited by the tangential field at the spacer.

- As the dose increases, the BV_{OFF} increases as well, up to a maximum value that is achieved when the vertical breakdown and the breakdown at the spacer occur at the same time.

- Increasing further the p-layer dose, the BV_{OFF} starts to decrease since the junction breakdown anticipates more and more.
- in case of low drain doping or short X the voltage capability could be limited by reach-through. It occurs when the drain depletion approaches the n+ region. However, this situation, for this particular structure, is always masked by the second region.

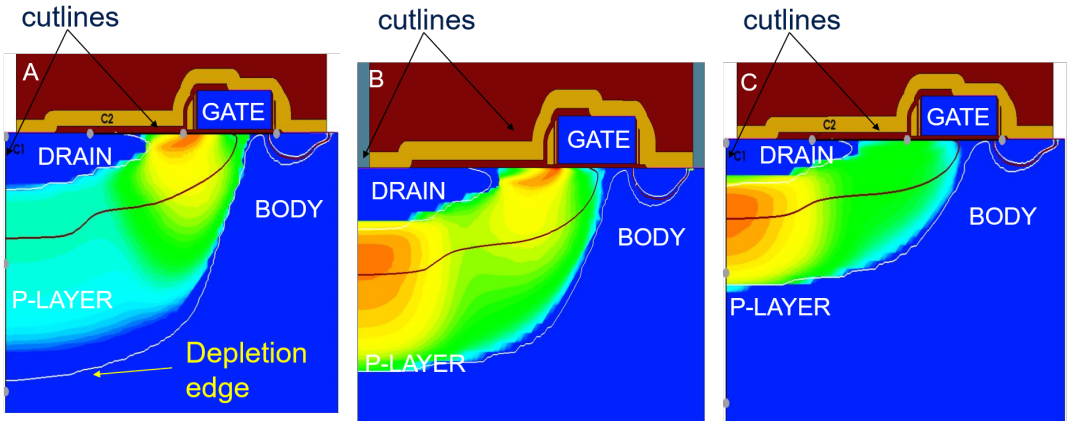


Figure 4.6: Impact ionization distributions for the A,B, and C architectures.

Figure 4.6 shows the impact ionization distribution inside the same structures (fixed geometry and drain/body doping) for the three p-layer: A, B and C of figure 4.6. Moving from left to right, it is possible to see how, increasing the dose of the p-layer, the critical point moves from the spacer to the vertical junction. The first structure is limited by the spacer, the last one by the vertical diode, while the third one seems optimal having a well-balanced impact ionization on the two sides.

Concerning the electric field along the two cutlines highlighted in the previous cross-sections, we can make several considerations. In the right side of figure 4.7, the electric field is plotted as a function of the vertical axis along the vertical cutline. It is also highlighted the position of the p-layer drain junction that is placed, of course, at the correspondence of the peak. Considering initially the width of the depletion region, it is clear that for low p-layer dose, the depletion is almost completely contained into the p-region. For intermediate and high doses, the depletion of the

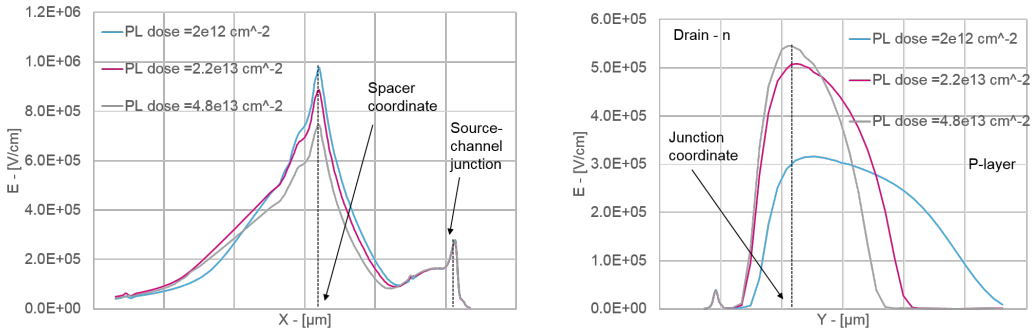


Figure 4.7: Electric field distribution along the two highlighted cutlines.

p-layer decreases gradually. The depletion of the drain, instead, increases up to a certain value determined by the edge of the highly-doped drain region. Moreover, as stated before, when the depletion approaches the n+ the electric field in that region starts to increase quickly. The small peak on the left side of the grey curve is a proof of this effect. Regarding the peak value, what we see is its significant increment that is proportional to the increment of the dose of boron. In the graph on the left side, instead, the electric field is plotted as a function of the tangential coordinate along the other cutline. The only remarkable thing is the decreasing trend for the peak value. It is also important to note that also if for the two highest doses we do not see any visible benefits on the depletion width, the peak at the spacer continue to decrease without saturating. This consideration is important under two different points of view.

Firstly, to reduce the BV_{OFF} variability, it is better that the critical point is at the vertical junction instead than at the spacer since it is less affected by process variations. Secondly, also if the BV_{OFF} does not change, a smaller tangential electric field improves the overall reliability. At the oxide interface, indeed, less electrons have high enough energy to be trapped in the oxide, modifying the resistance of the device.

At this point, we can complete the answer to one of the questions we have posed at the beginning. In figure 4.8 we try to combine the changes of the p-layer and drain. We plot hence different ReSURF curves as a function of the p-layer dose and each of

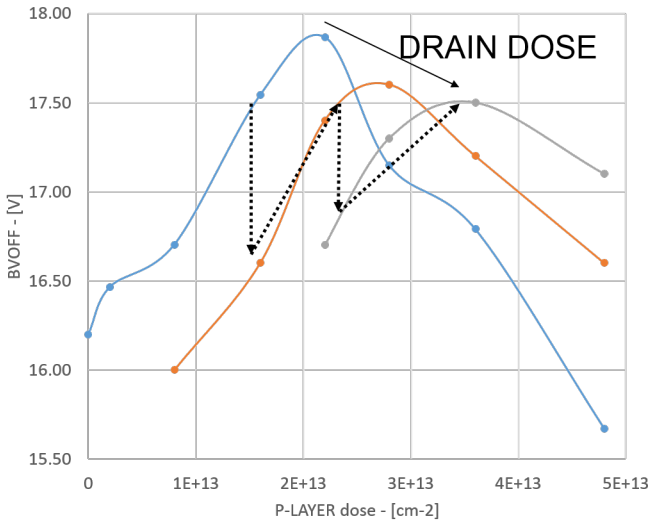


Figure 4.8: ReSURF curve combining drain and p-layer doses changes.

them obtained with a different drain dose. What we note is that, increasing the drain dose, the curve shifts toward right and the maximum value decreases. This latter is justified since, with a more charged drain the vertical breakdown anticipates. The former, instead, happens since with a more doped drain the breakdown at the spacer is anticipated and a stronger perpendicular field or higher p-layer dose is necessary to reach the top of the curve. So, if we aim to realize a device with a BV_{OFF} of 17.5 V, for example, we can follow the path shown by the arrows on the graph. We start from the blue curve that allows that BV_{OFF} with a certain resistance. Then we increase the dose of the drain to reduce the resistance but, doing this, we loose in BV_{OFF}. So we have to increase the dose of the p-layer to gain again the desired BV_{OFF} with no or minimum impact on the ON-resistance. We can repeat this step many times until we reach the maximum drain dose and so the smallest resistance for that geometry with which we can secure the target BV_{OFF}.

4.2 Field Plate Assisted ReSURF

The usage of a p-layer has, however, several limitations: the effectiveness of the depletion induced by the p-layer is poor, especially for high doped drain, moreover, the depletion region does not directly insist on the critical point that is instead

very superficial. The overall consequence is that it is possible to realize devices designed on an all-in-active architecture that secure different low-voltage classes but with poor overall electrical performances. As explained, we aim to find another integration solutions to realize competitive devices with voltage classes that range from 5 V to 20 V. Therefore, our first purpose, before trying to optimize also the R_{ON}, is to understand how we can reach a breakdown voltage of 30 V.

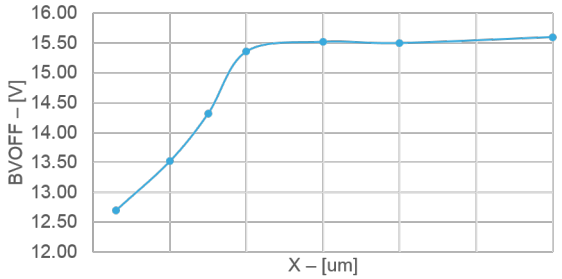


Figure 4.9: Experimental BV_{OFF} as a function of the X.

With this target, without changing the all-in-active architecture and with what we know until now, there are only two possible roads we can follow: to enlarge the drift region or to make the drain less charged. Both roads have, however, drawbacks. Figure 4.9 shows the experimental measurements of the breakdown voltage for several structures that differ only for the length of the drift region.[15] It demonstrates how the first road becomes very soon a cul-de-sac. The BV_{OFF} saturates and any further increment of X does not bring any improvements but only a worsening of the resistance. The reason for this saturation is hidden behind the fact that the real important parameter is the extension of the depletion region and not of the whole drift region. It is inside the depletion region that the voltage stress drops and there is a non-null electric field. So, considering a fixed voltage stress, the larger is the depletion the smaller is the critical field since the area must remain constant. In conclusion, the first road can be a possible solution only if we have an effective way to deplete the drain at the same time. The second road might be used to support the first one since the drain depletion increases if its dose diminishes.

However, it is not practicable: a too low drain dose is counterproductive since the structure begins to be limited at the n+ region and the R_{ON} increases earlier too much as well.

Following only these two roads as possible solutions to achieve our target, i.e. a much higher BV_{OFF} while still securing competitive ON-resistance, is not possible for the aforementioned reasons, so we started to develop a relatively new idea to improve significantly the ReSURF effect. We looked for a way to increase further the perpendicular electric field and consequently the depletion region and the BV_{OFF} without being limited by the junction between p-layer and drain. Moreover, this solution shall not penalize the resistance too much. To introduce this architectural change, we resume the simplified model we have used to introduce the 'Junction ReSURF' and we describe, initially theoretical, the concept of the so-called 'Field Plate Assisted ReSURF'.

The 'Field Plate Assisted ReSURF' is already used for very high-rated structures where we need to reach a BV_{OFF} of many hundreds of Volt. We will take this idea and transport it on our low voltage structure with the necessary modifications. This kind of ReSURF, instead of a p-n junction, uses a MOS system to force the perpendicular electric field inside the drain.[27] The MOS system is made of a metal field plate, hence the name, and a thick oxide.[32] So, this time we try to see what happens if we substitute the second domain with a metal and oxide layer instead of a p-layer (see figure 4.10). Looking at the theoretical electric field shape, we can see that also if the global shape is different, the field inside the drain has the same shape of the one obtained for a 'junction ReSURF'. There are also many other analogies between the two ReSURF solutions. From an analytical point of view, we can compare the analytical expression of the depletion width inside the drain and the maximum field reached in the silicon. The following expressions can be taken by almost any semiconductor devices handbook.

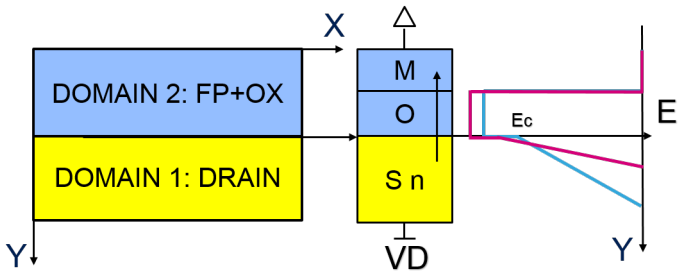


Figure 4.10: Generalization of the RESURF effect, modeling of the field plate assisted RESURF.

Junction ReSURF

$$E_{MAX} = \frac{qN_D x_n}{\epsilon_s}$$

$$x_n = \frac{N_A}{N_A + N_D} \cdot \sqrt{\frac{2\epsilon_s(N_A + N_D)(\Psi_i - V_D)}{qN_A N_D}} \approx \sqrt{\frac{2\epsilon_s(\Psi_i - V_D)}{qN_D}}$$

Field Plate Assisted ReSURF

$$E_{MAX} = \frac{qN_D x_n}{\epsilon_s}$$

$$x_n = \sqrt{\frac{2\epsilon_s(V_D - V_{OX})}{qN_D}}$$

The two expressions of the maximum electric field into the silicon are equal. The other expressions are also much similar especially if we assume that the p-layer is much more doped than the drain ($N_A >> N_D$) as usually happens to obtain the best ReSURF effectiveness. In conclusion, the 'Field Plate Assisted ReSURF' is another way to reach the aim of the ReSURF that is perfectly similar to junction ReSURF. It will, therefore, be characterized by similar ReSURF curves, it will modify the electric field shape of the drain displacing the critical point among the weak points, it can be optimised to reach the optimal BV_{OFF} in a similar way to what we have done for the p-layer.

We have already seen a ReSURF induced by an FP in the previous chapter. At

that time the ReSURF effect was still unknown, so we passed over it quickly and without detailing it. When we talked about the first weak point, we said that the electric field increases until it arrives at the gate edge while after it decreases. The reason is that the gate acts as an FP, it forces a perpendicular electric field inside the I and, consequently, according to the ReSURF principle, the tangential component, responsible of the peak, decreases quickly. Moreover, the peak at the drain-body junction is always smaller than the peak at the spacer because the ReSURF induced by the gate.

The solution involving the metal field plate (FP) has many advantages. Here, we list only two of them postponing the description of the others to the next sections: firstly, the MOS system is created above the drain and so the field is induced directly on the critical side of the drain, and secondly, the voltage limitation is no longer linked to a breakdown of a p-n junction. When we added the p-layer, we have seen that together with the benefits of the reduction of the tangential fields, we have introduced also a new weak point that in some cases can limit the structure. According to the analogies we insisted before, also the FP brings many benefits that we will see shortly, but also a new critical point. Particularly, the fields across the oxide will never be a limiting factor since the thickness of the oxide is quite high. However, we can not say the same about the electric field inside the silicon at the interface with the oxide, that can become critical, especially for high drain doses where the width of the depletion region is small.

Figure 4.11 shows a cross-section to visualize how the so introduced and described FP can be integrated into a standard all-in-active architecture. In the picture, besides the drain, source, gate and p-layer regions, it is labelled also the new architectural element. The FP is made by a continuum metal layer that covers most of the X. The FP contact is also shown, it allows to electrically connect the FP electrode to the metal lines and then to the outside.

What we propose is not to integrate the field plate substituting the p-layer, as could seem, but to integrate and to optimize both solutions at the same time, in order to have the highest effectiveness and elasticity as possible. The use of a FP

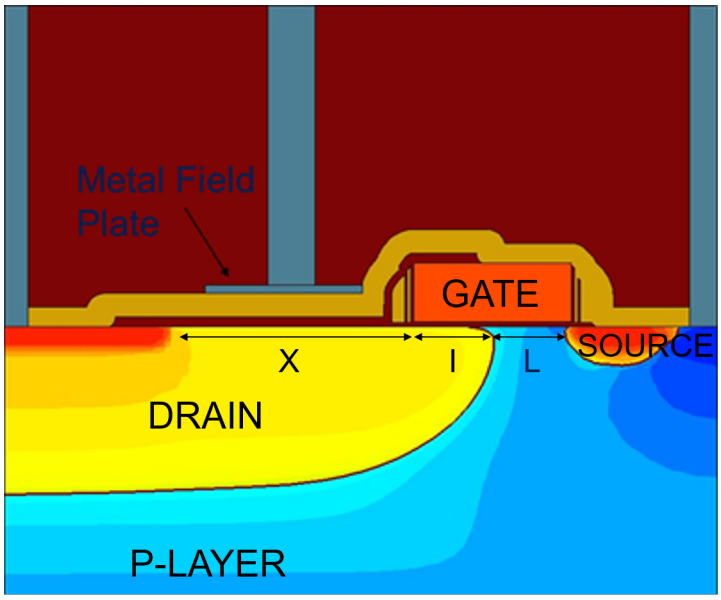


Figure 4.11: Cross-section of an all-in-active architecture implemented also the FP as introduced theoretically.

introduces, in fact, further degrees of freedom for the optimization of the architecture since, as we will see in the next chapter, the electrical performances are strongly affected by the geometrical parameters and the shape (continuum FP or discrete one). To understand the potentialities of a structure that implement a field plate, we take again the ReSURF curve we have drawn during the analysis of the variations of the p-layer dose, we add a metal field plate above the drift region and we perform again all the simulations. The results are shown in figure 4.12. The blue curve is the same one we have seen in figure 4.5, the pink one, instead, is obtained from these new simulations. When the structure is limited by the vertical junction the FP does not bring any improvements; but for structure limited by the spacer, it can further reduce the field and increase the BV_{OFF} by many Volts. It is so possible to increase again the drain dose and consequently reduce further the ON-resistance.

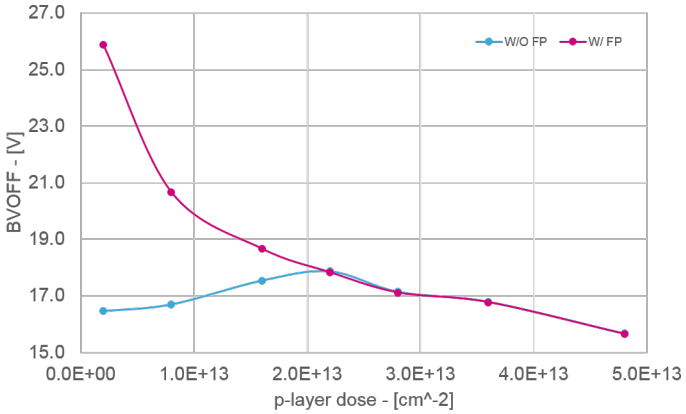


Figure 4.12: BV comparison between structure w/ and w/o FP for different p-layer dose.

Chapter 5

Metal Field Plate Technology

In the last chapter, we introduced the idea to improve the ReSURF of the all-in-active architectures with the application of a metal field plate. The graph in figure 4.12 showed that the FP can bring many benefits to the voltage capabilities in certain combination of drain and p-Layer. This final chapter will show, experimentally, the electrical results of a prototype metal field plate onto an LD-MOSFET architecture in BCD10 technology. The aim is the comprehension of how the FP acts on the overall electrical performances: what is its effect on the distribution of the electric field, how the critical points move and finally how the ON-resistance is changed. The first paragraph will be dedicated to FP integration.

5.1 FP Integration

A great effort was made to design and integrate this new architectural solution into a test chip. Particularly, as previously explained, these devices must be realized with the BCD flow avoiding that the performances of other structures are negatively affected. Many integration schemes are possible but some solutions cannot be adopted without dedicated studies about the effects that the changes to the process flow have on the whole system. The first test chip was hence designed to provide the first experimental data related to this new architecture and to justify the investments of further analysis, time and money. The experimental data, particularly, are needed to confirm the results of simulations and tune the calibration, if required. To this extents, we have chosen the integration scheme described in detail hereinafter, the geometries (i.e. the layouts), the doses and the energies of the drain and p-layer regions starting from TCAD simulations that were massively used to design the structures and to allow an extensive analysis of all the parameters related to FP.

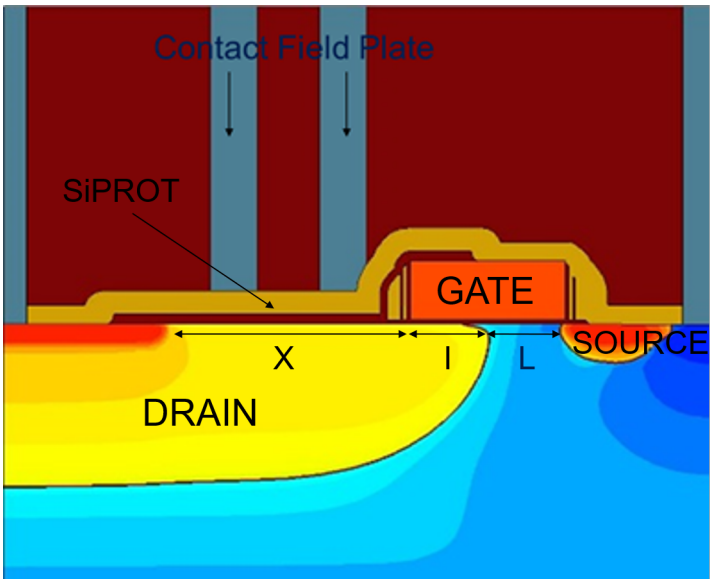


Figure 5.1: Cross-section of an all-in-active architecture with contacts as FP.

The idea of the integration of a metal field plate is to force a perpendicular electric field inside the drain drift region by the realization of a MOS system. Differently from the structure used in the preliminary TCAD activity, where the FP was a continuum metallic layer (see figure 4.11), we chose to realize the FP through an array of contacts, that are the standard ones available in the BCD10 platform. This solution allowed to minimize the integration activity and to avoid the need of additional masking level. The SiPROT stack was optimized to stop the contact etch and to target the desired FP height, i.e. the dielectric thickness. Figure 5.1 shows the cross-section of an all-in-active architecture which integrates this FP implementation. The main elements that compose the metal field plate are summarized here below:

- A redesigned SiPROT layer that, together with its usual function, i.e. to avoid the formation of the silicide in the drain drift extension, must also stop the contact etch and secures a certain thickness of the dielectric between FP and the silicon surface.
- An array of contacts to realize the FP metallization located above the drift region. The particular pattern is designed to achieve the best effectiveness as we will see in the following.

- A first metallization level (Metal 1) to short all the FP contacts. It is designed to cover the whole drift region until the heads of the transistor.

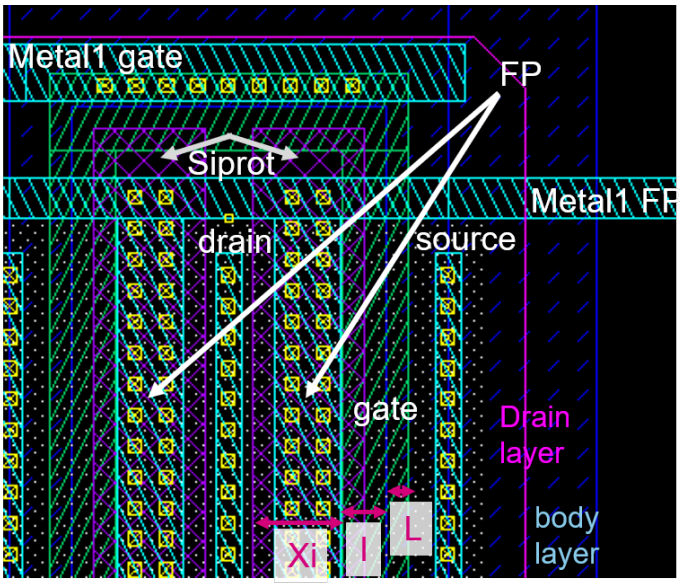


Figure 5.2: Layout of the an all-in-active architecture with contacts FP.

A detail of the layout is reported in figure 5.2, where only the layers that are significant for this chapter are made visible while all the others, such as the LDD or N+, are hidden to avoid to make the picture unreadable. Moreover, to make the reading more clear, several labels are added to the picture. They identify all the layers concerning the FP that we have already described, the gate, the active area, the source and the drain. Finally, we have denoted the lengths of the active regions so that they are immediately recognizable. Some clarifications about this layout seem necessary at this point. The gate has a doughnut-like shape in which the drain is placed in the centre and the sources at the two external sides so that the overall transistor width is doubled. The drain layer (purple edge) embraces all the active areas since the device is 'drain everywhere'. Then, concerning the active regions, it is possible to note how the nominal channel length L is defined from the edge of the gate mask on the right side to the edge of the body mask on the left. Next, the accumulation region I is defined as the remaining length of the gate that is not covered by the body layer, i.e. from the body mask edge to the left gate edge.

Finally, the drift region X extends from the left gate edge to the edge of n+ region that is covered by the SiPROT mask.

Using a discrete pattern as a contacts array, will it be as effective as a continuum metal layer? We answered this question by TCAD simulations and including dedicated structures in the test pattern, and we proved, as we will see later, that if the spacing between neighbour contacts is sufficiently small, the FP effectiveness is guaranteed and it behaves as a continuum metal layer of length equal to the distance between the beginning of the first contact and the end of the last contact (FP_{eq} reported in picture 5.5). In this specific case, we can leverage a feature of the advanced BCD10 technology node, i.e. a tight contact spacing.

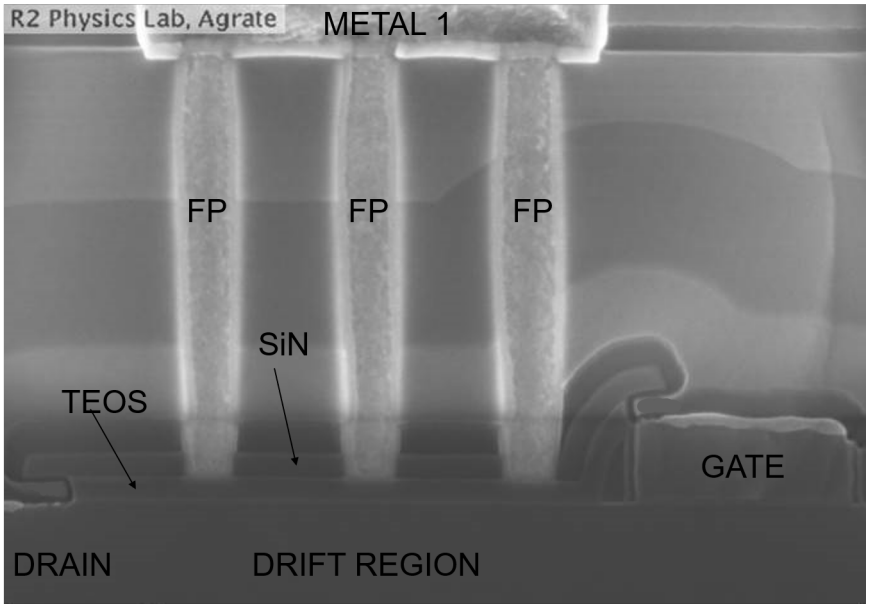


Figure 5.3: SEM cross-section of the all-in-active architecture with a FP.

After the definition of the layout, several other morphological trials have been performed to be sure that, after the contacts realization, i.e. the etching of the trenches, the sputtering of the barrier and the subsequent tungsten filling, the contacts land maintaining a certain height from the silicon surface (since this is a critical technological parameter). In this sense, the SiPROT layer was re-engineering to

allow the integration of two different FP heights, labelled 500 Å and 700 Å. As previously said, these values were identified with deep TCAD analyses together with the implantation doses and energies, besides the geometries. Figure 5.3 shows the cross-section obtained with the Scanning Electron Microscopy (SEM), the gate electrode and the contacts are perfectly defined and visible. In the upper part, there is the first metallization. The TEOS and nitride layers of the SiPROT are labelled while the bright region at the drain contact and above the gate represents the silicide.

5.2 FP Experimental Study and Optimization

In the previous chapter, we have seen how difficult is the optimization of an all-in-active LD-MOSFET to achieve the best $BV_{OFF} - R_{ON}$ trade-off. When we move from a standard architecture to one that implements a FP, the number of parameters in play increases a lot since we have to consider the FP geometry besides the drain and p-layer variations. Therefore, to analyze the whole system, many studies have been performed to understand the effects that the modification of a single variable or a few of them have on the electrical performances and field distributions. These analyses will be soon exploited to realize a Design Of Experiments (DOE) by which it will be possible to know the best combination of the design variables to reach determined performances. The experiments, namely the list of devices integrated on silicon and the drain implantations, have been designed to be sure that any benefits can be imputed only to the quantity under test and, more importantly, that there are no other limitations that can hide both positive and negative effects.

More in detail, the geometries and the drain doses were chosen to pursue mainly two goals:

1. To realize structures that allow, thanks to the FP architecture, the improvement of the R_{ON} performance, maintaining the same voltage capability (same BV_{OFF}), particularly in the low-voltage range, where the challenge coming from the technology node scaling becomes greater.

2. To realize structures that allow, thanks to the FP integration, the increment of the BV_{OFF} of all-in-active architecture up to 30 V, with best in class R_{ON}

performance, overcoming the limitation described in section 4 (Figure 4.9).

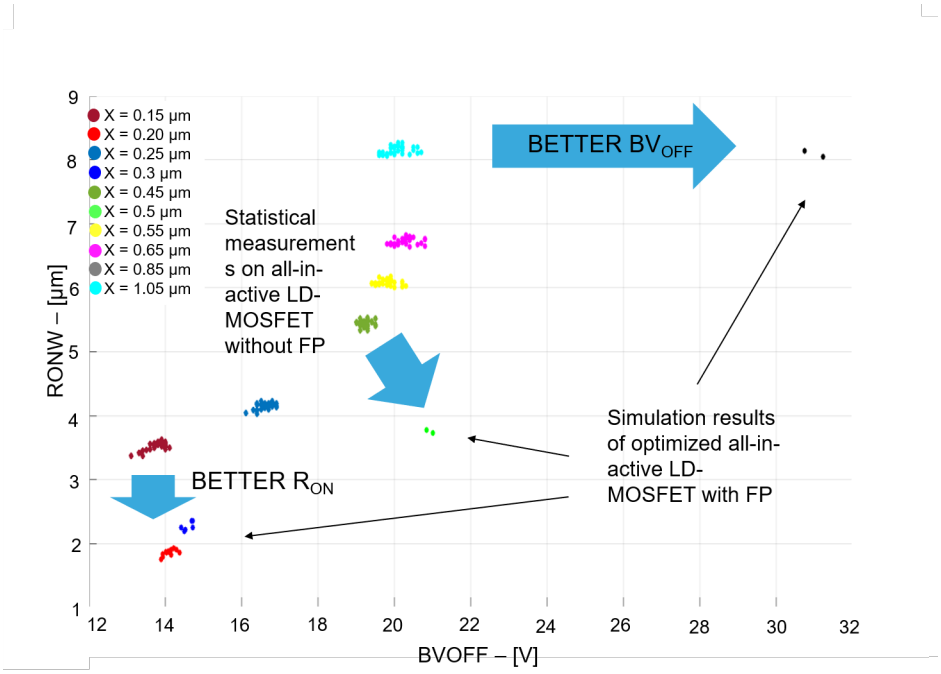


Figure 5.4: Main purposes that are targeted with the addition of a FP.

To better understand these points, i.e. what it means to optimize a power device, let us consider the figure 5.4. It shows the statistical measurements on architectures without FP that differ only by the length of the drift region X. Similarly to what described about the standard all-in-active architectures, according to the figure 4.9, the benefit on BV_{OFF}, coming by the increment of the drain extension region X, is limited and for large X the only effect is a worsening of the R_{ON}. The challenge of our optimization activity is exactly to populate the right/bottom side of the graph. The data points outside the experimental curve are the results of the preliminary TCAD activity that confirms that the objective stated above can be achieved thanks to the introduction and the engineering of a metal FP. Particularly, they come from an accurate optimization that involves all FP and structure design parameters, e.g. the X, the field plate height, etc..

To understand how the geometry affects the electrical performances, a lot of structures have been simulated and, based on that, drawn and then integrated on

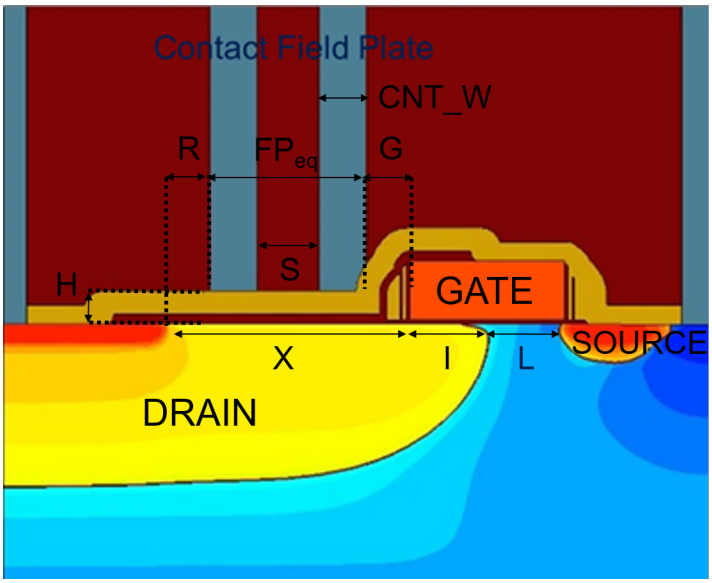


Figure 5.5: Main purposes that are targeted with the addition of a FP.

silicon. For each parameter, therefore, we realized many structures to investigate the effects of its variation in a certain interval and for different combinations of the other ones. In figure 5.5 we report the cross-section of an architecture that implements a contact FP in which we highlighted all the meaningful parameters for the subsequent analysis. The analyses will focus mainly on all geometrical parameters related to the FP: the thickness of the dielectric or the height of the FP from the surface, the distance G between the gate and the FP near-most edge, the distance R between the n+ and the FP near-most edge, the spacing between the contacts used as an FP and the number of them that is directly linked to the total equivalent FP extension, FP_{eq}. These parameters will be analyzed in details hereinafter. Before, let us spend two comments on L and I. The channel length L and the gate-drain intersection region I have been analyzed too, but are not discussed in this work. The channel length optimization is independent from the FP introduction and from the drain engineering and it has been fixed to the minimum allowed value, the best to improve the R_{ON}. The I, as well, has been minimized to reduce the gate drain capacitance, an important figure of merit for the power device related to the power losses during the switching from OFF-state to ON-state and vice-versa. In fact, another benefit coming by the FP is that it is a terminal separated by the gate and its overlap with

the drain does not affect the gate drain switching power losses. Let us, now, move to the deep analysis of the geometrical parameters strictly related to the new FP architecture.

- H: the height of the FP, i.e. the thickness of the dielectric of the MOS system realized by the FP. However, the effectiveness of the FP is linked to how strong is the electric field that we can force inside the drain; it does not depend only to the H but the real and important parameter is the capacity of this MOS system. Particularly, the SiPROT layer is made by an oxide layer and a nitride one, hence the capacity can be evaluated considering a series of two capacitors.

$$C_{FP} = \frac{\epsilon_{TEOS}\epsilon_{SiN}}{\epsilon_{TEOS}T_{SiN} + \epsilon_{SiN}T_{TEOS}}$$

From which, we can define an equivalent dielectric thickness that is used as the independent variable of the following graphs.

$$T_{eq} = \frac{\epsilon_{TEOS}}{C_{FP}}$$

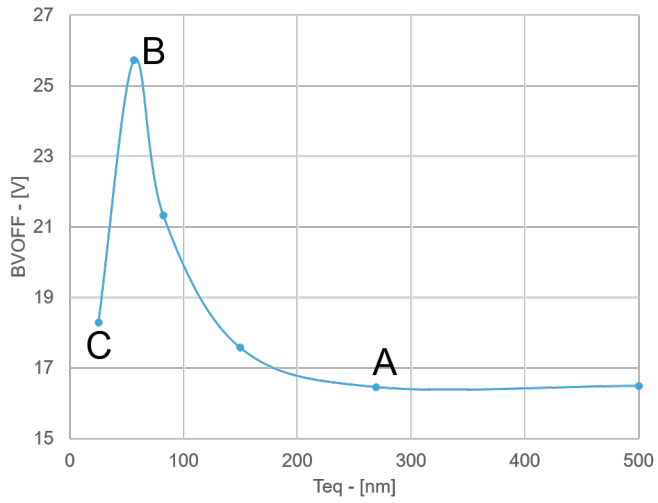


Figure 5.6: Breakdown voltage as a function of the dielectric equivalent height.

The graph in figure 5.6 shows the simulated breakdown voltage as a function of the so defined equivalent thickness. Let us consider as reference a structure

without the FP, with fixed geometry and doping concentration. This structure breaks at 16 V, corresponding to the right-most point in the graph, there the BV_{OFF} is limited by the high electric field below the poly edge (spacer), that is the hot point that we want to improve with the FP introduction. As the FP approaches the silicon surface, the capacity increases and, consequently, the induced electric field increases as well. When the FP is very far from the surface (A architecture), the ReSURF effect induced by the FP is very small or negligible. The BV_{OFF} hence saturates to the breakdown voltage of the same architecture without the FP as shown in the graph. Then, reducing the dielectric thickness, the BV_{OFF} increases up to a maximum value (B architecture) before dropping down again. In the junction ReSURF, the decreasing trend of the BV_{OFF} when the perpendicular field becomes too strong is due to the breakdown of the vertical diode. In the Field Plate Assisted ReSURF, a very similar thing happens but the anticipated breakage is due to the electric field at the silicon very close to the interface with the oxide. To sustain the same voltage stress, indeed, the architecture with thinner oxide has a smaller drop across it and consequently a larger one inside the silicon. The C structure is so the first example of a structure that is limited by the FP itself.

Let's consider now the distribution of the electric field along a cutline at the interface with the oxide for three different height of FP. In figure 5.7, the cross-section is related to the B architecture where the contacts are substituted by a continuum FP to simplify the analysis. The blue curve is the electric field in the absence of the FP. The other curves are related to the A, B, and C architectures. As the FP approaches the silicon surface, the electric field below the spacer decreases according to the ReSURF principle while the electric field in the drift region increases following the increment of the perpendicular component. As before, the structure with the greatest BV_{OFF} is characterized by an electric field distribution that has the two peaks almost at the same level, i.e. the distribution that is nearer to the flat one. If the effectiveness of the FP increases too much the field on the n+ side explodes. The shape of the curve of figure 5.6 depends also on the drain extension geometry (X and FP_{eq} in

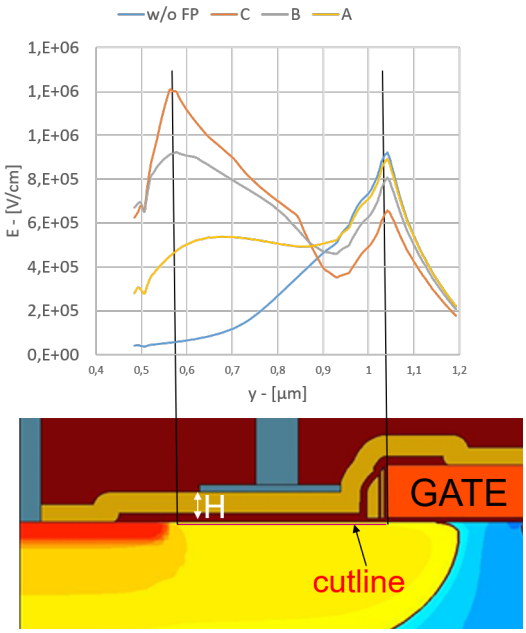


Figure 5.7: Electric field distribution at different FP height.

particular) and on the drain doping concentration. After the deep preliminary TCAD activity, we finally identified in the range between 50 nm and 70 nm the optimal thickness to target a BV_{OFF} of 30 V. The final dielectric stack integrated in the electrical lot, labelled 500 Å and 700 Å in section 5.1, are the output of this analysis.

Figure 5.8 shows the comparison between TCAD and experimental results for the final structures integrated on silicon (using the contact array as field plate) with the two different dielectric stacks, corresponding respectively to 47.5 nm (labelled 500 Å) and 58 nm (700 Å). The device geometry and doping concentration used are the one optimized to target a BV_{OFF} of 30 V. The prediction of the TCAD simulation, fully confirmed by the experimental measurements, prove that the metal FP integration allows to extend the maximum voltage capability of an all-in-active LD-MOSFET up to 30 V that was one of the main objective of this work.

- S: the spacing between neighbour contacts. The contacts spacing is a further

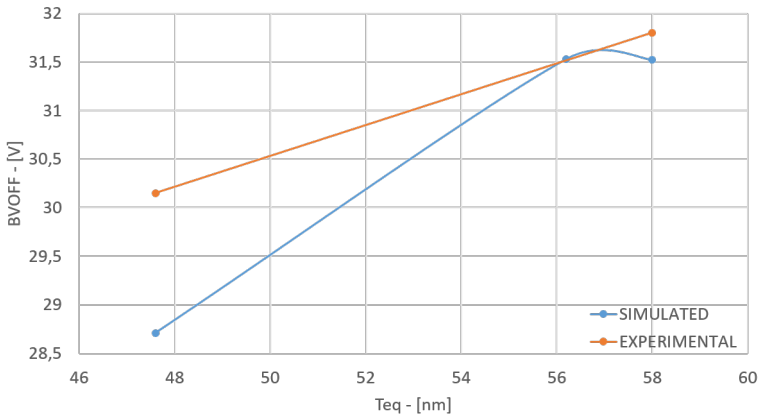


Figure 5.8: Breakdown voltage as a function of the dielectric equivalent height - comparison between experimental and simulated structures.

optimization parameter that can be used. Qualitatively, until the lateral silicon depletions induced by neighbour contacts touch each others, the reduction of the effectiveness due to the spacing is negligible. On the contrary, when the contacts remain independent, the overall effect is similar to have only the contact nearest to the gate. Figure 5.9 reports the breakdown voltage as a function of the relative spacing. The relative spacing is evaluated starting from the minimum spacing S_{min} defined by the design ruled of the BCD10 (90 nm) as follows

$$\epsilon_S = \frac{S - S_{min}}{S_{min}} \cdot 100$$

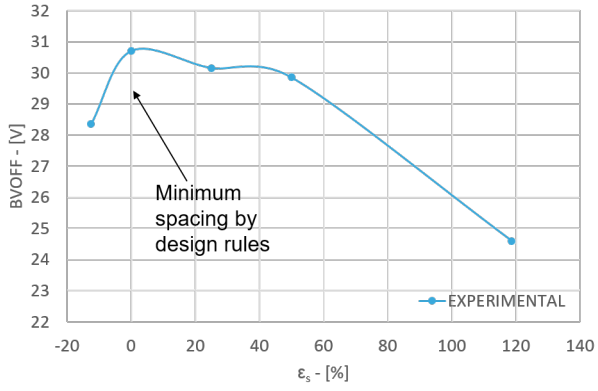


Figure 5.9: Breakdown voltage as a function of the S.

As before, these results are obtained through measurements on the same electrical lot. As explained qualitatively before, we can note that to see a sensible reduction of the BV_{OFF}, the spacing must increase above the double of minimum distance. In conclusion, the BV_{OFF} variation is a function of two different and contrary effects: on one side, spacing the contacts allows to cover a larger portion of the drain extension, so the BV_{OFF} is enhanced but, at the same time, the depletion in the region not covered by the contacts and that are depleted only by the lateral depletion continue to worsens, so the BV_{OFF} tends to decrease again. For this reason, the curve remain approximately constant except for very short spacing or very large one.

- G: the distance between the gate edge and the FP edge. This distance is quite important since the effectiveness of the FP on the spacer side is strongly affected by it. To obtain the best effect the FP should be placed as close as possible to the poly edge and so this distance must be minimized. However, the size of the spacer can be a practical limitation also if for the drain doses that we use, we are never limited in this sense. The graph in figure 5.10 shows the simulation results of the BV_{OFF} as a function of a relative distance from the gate. Let's call G_{min} the distance between the gate and the edge of the spacer that we have chosen as reference working point. The independent variable we use to plot is so evaluated.

$$\epsilon_G = \frac{G - G_{min}}{G_{min}} \cdot 100$$

In this way the first point correspond to a FP egde that is coincident with the gate edge, i.e. $\epsilon_G = -100\%$. Until the FP is sufficiently close to the poly edge, it is effective to bend the equipotential lines parallel to the silicon surface, reducing the tangential electric field component in the silicon below the poly edge. In this case, the FP can works correctly and BV_{OFF} depends on the FP length. On the contrary when FP is to far from the poly edge, i.e. G is too large, it is no more effective to bend the equipotential lines in that critical region, so the BV_{OFF} rapidly drop down to the value of the same architecture without FP.

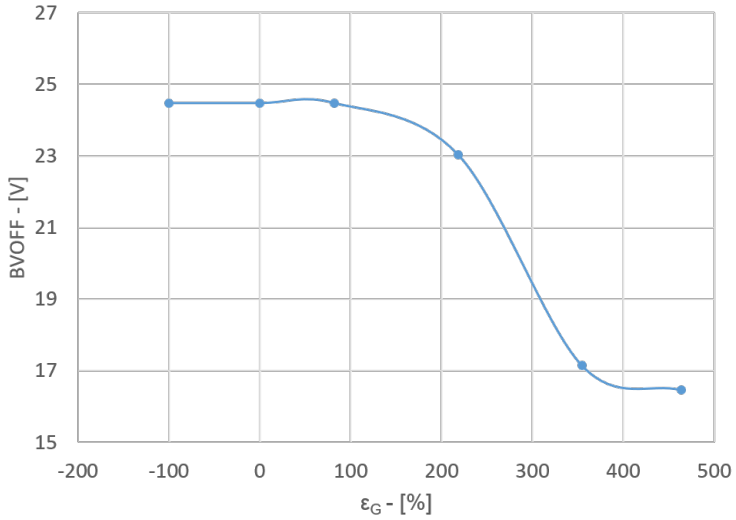


Figure 5.10: Breakdown voltage as a function of the G.

- FP_{eq} or N: the equivalent total extension of the FP. They are related as follows

$$FP_{eq} = N \cdot (S + CNT\_W) - S$$

Moreover, it is often considered another important dimension that indicate the total length of the drift region covered by a FP, so including also the G.

$$O = FP_{eq} + G$$

All these dimensions provide indication on the total FP width and, for our purposes can be used almost indifferently since, except for the dedicated structure used to investigate the S and the G, all the other trials have both variables fixed to their minimum value that, as we have seen, ensures the best performances. Qualitatively, the FP extension is strongly linked to both the critical points, i.e. the spacer side and the n+. Increasing the FP extension allows depleting a larger region so the critical field at the spacer decreases but if the FP becomes too close to the n+, i.e. R becomes too small, the breakage at the n+ anticipates. Due to the FP integration scheme, i.e. exploiting the contacts, we can not choose all possible values of FP_{eq} but we can only decide the number of contacts (The S and G are fixed), for this reason we prefer to

use this parameter instead of the other two, i.e. O and FP_{eq}. Secondly, the distance from the n+ R and the N are also related by the following expression:

$$R = X - O$$

As a consequence, increasing one of them will result to reduce the other. The only way to improve both of them and therefore the BV_{OFF} is to increase the X worsening the R_{ON}. Again, the optimal structure is the one with the shortest X for which we secure the BV_{OFF} target.

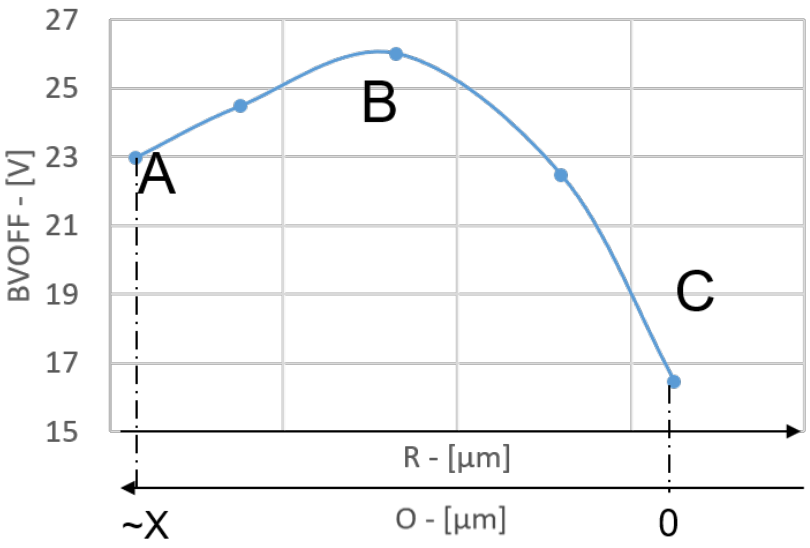


Figure 5.11: Breakdown voltage as a function of the R and O.

The figure 5.11 reports the simulated breakdown voltage as a function of the R and of the O. As explained, these two parameters can not be discussed separately. As for the height of the FP, also these simulations have been carried out from a reference structure that has a different drain architecture and a continuum FP. The curve has a bell shape where. The C structure does not have the FP, therefore, the O is null while the R is fixed to the entire extension of the drain drift region. The A architecture, instead, has a FP that covers part of the n+. This curve explained directly what previously described

showing the best performance in the B point, where R and O are similar.

However, when the electric field at the n+ increased too much, besides a reduction of the breakdown voltage, a second issue appears. The soft breakdown becomes very accentuated and bring an important increment of the leakage of the transistor at relatively low voltage. Figure 5.12 reports the output characteristics of the A, B and C architectures. The blue curve is always related to an architecture without the FP (A structure). The orange one is obtained with the best R-O trade-off (B structure) while the grey one is related to an architecture in which the FP is partially overlapped to the n+ region (c structure). As it is possible to note, the orange one is what we look for: the leakage does not change but the BV_{OFF} is moved many Volt farther, the grey one, instead, highlights the accentuated soft breakdown we introduced before.

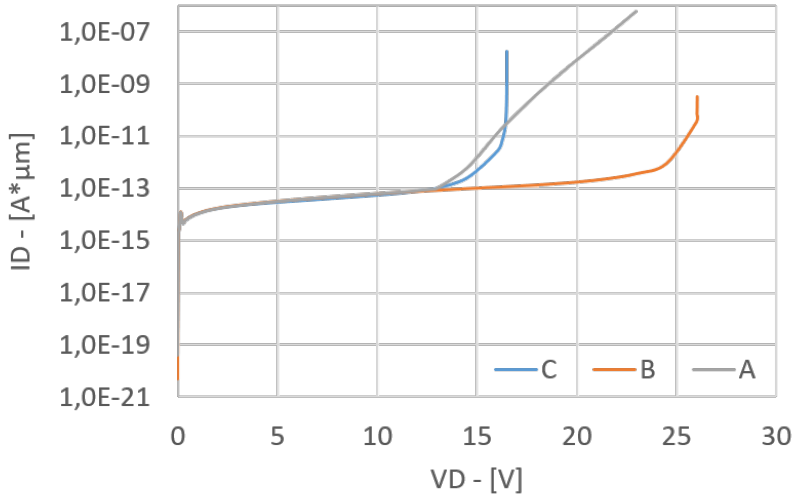


Figure 5.12: Output characteristics of the A, B and C architectures.

In the previous description, we gave an overview of the main FP geometrical parameters and their effects on the voltage capability of the architecture. Now, we will provide further investigations about combinations of the parameters involving also the drain implantation and the length of the drift region. Concerning the latter ones, we focus the attention on the graph reported in figure 5.13. All the data

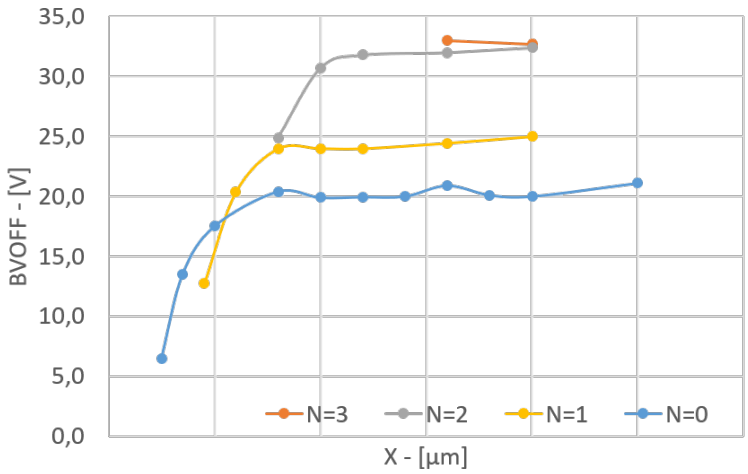


Figure 5.13: BV_{OFF} as a function of X for different FP length (O).

are obtained from experimental measurements on many of the structures we have designed. The blue curve is obtained from the reference structures that do not implement the FP. As we have already seen and explained in previous analyses, the BV_{OFF} increases until it saturates. Then, we apply a single contact that acts as an FP (yellow curve). The contact acts as field plate introducing positive ReSURF action close to silicon surface, so the BV_{OFF} limit increases. Of course, to allow the integration of the contact, the X should be long enough. This is the reason why as the number of contacts increases, the curve starts at larger X. A single contact is however not enough to achieve the best ReSURF effectiveness. Its action, indeed, is very limited once X becomes larger and larger. Therefore, we add a second contact (grey curve), a third one (orange curve) and a fourth one (since this one is overlapped with the orange curve it is not shown) to increase the overall FP length. This picture shows us three important concepts:

- The first point of the structure with one contact is worse than the one of the structure without the FP. In that point the device is limited at the n+. Since the very short X, the contact is placed very near to the highly-doped drain, i.e. R is too small. Therefore, the depletion is very narrow and the electric field increases too fast. The introduction of an FP is not always positive to increase BV_{OFF} capability.

- The breakage that causes the saturation of the blue, yellow and grey curves occurs at the spacer. The depletion induced by the FP increases allowing to bear higher voltage stress until a maximum value. There, the further increment of X is not followed by an increment of the depletion so the BV_{OFF} saturates. To overcome this limitation, it is possible to stretch the FP adding one or more contacts until the R remains large enough to avoid limiting the structure as in the previous point.

- This positive trend can continue until the structure starts to break in an other point, i.e. different from the spacer. Actually, the breakdown occurs at the vertical junction, so, a further increment of the number of contacts is useless. Moreover, if we continue to increase the FP length, at fixed X, approaching the n+ region, the hot point changes again and it moves to the highly-doped drain region, therefore, BV_{OFF} starts again to decrease as we will see with the next graph.

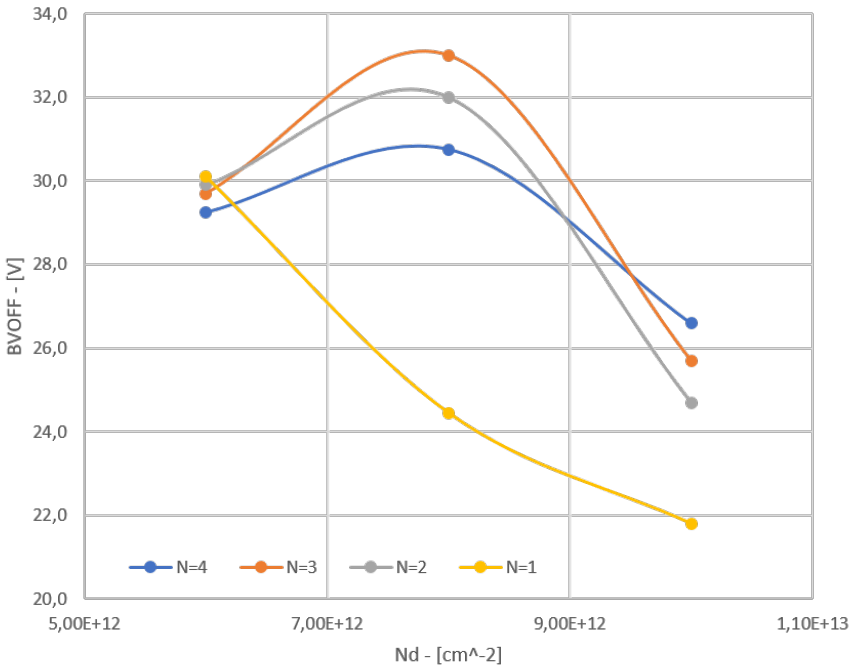


Figure 5.14: BV_{OFF} as a function of the drain dose.

However, using the highest number of contacts is not always the optimal solution. In fact, looking at the figure 5.14 in which the BV_{OFF} is plotted against the drain dose, the best number of contacts depends also on the drain dose. This graph is obtained as a cut of the previous one, the X is hence fixed to the last but one point. Again, the yellow, grey, orange and blue curves are related to the structures with one, two, three, and four contacts, respectively. Except for the yellow curve, the other ones have a bell shape. In the right decreasing section, with higher drain doses, the devices are always limited by the electric field below the spacer. At low drain dose the situation is even reversed. The FP depletion effect is faster (same extension occurs at lower drain bias) and the n+ region is reached earlier, so, an higher number of contacts results in lower BV_{OFF}.

In the end, it is not possible to optimize each parameter independently from the others. All design parameters have to be considered at the same time to design a device with the best trade-off between the breakdown voltage and the resistance.

5.3 Field Plate and ON-Resistance

The last analyses concern the impact the Field Plate Assisted ReSURF has on the ON-resistance. We analyze in section 3.2 the R_{ON} dependence on the drain dose and on the length of the drift region. However, after the application of an FP, the discussion about the resistance becomes more complex. As said, the FP realize an MOS system. According to the bias we give to the FP electrode, the depletion inside the silicon changes. Three possible scenarios are hence possible:

- $V_{FP} - |V_{FB}| = 0$. When the FP bias is equal to the flat band voltage of that MOS system, the FP looses any effectiveness on R_{ON}, i.e. the structure has the same performances of the one without FP.
- $V_{FP} - |V_{FB}| < 0$. In this condition the FP forces the depletion of the silicon. The extension of such depletion reduces consequently the effective thickness of the drift region and it is proportional to the increment of drain sheet resistance and thus the total R_{ON}.
- $V_{FP} - |V_{FB}| > 0$. In this condition the FP forces an accumulation inside the silicon. As it happens in the I region, the accumulation improves the ON-resistance reducing the sheet resistance of the drift region.

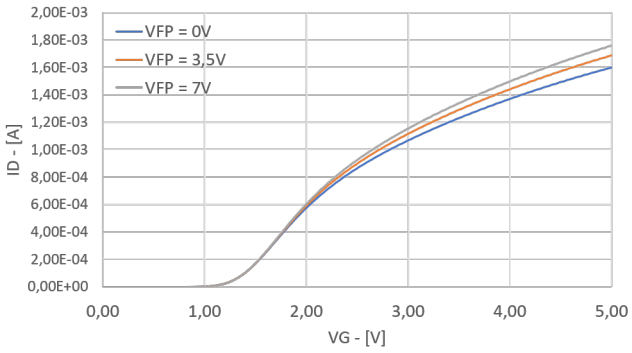


Figure 5.15: Experimental trans-characteristics at different FP bias.

Figure 5.15 shows three experimental trans-characteristics obtained biasing the FP with three different voltages. As it is possible to note, the threshold voltage

does not change while the R_{ON} has three well-distinguish values. The smallest one is, of course, related to the highest positive FP bias. From all the experiment measurements, the percentage variation of the resistance ranges from few percents to 20% for a FP bias that changes from zero to seven Volts. It depends mainly on two factors: the drain dose and the extension of the FP with respect to the drift region. However, biasing the FP has consequences also on the breakdown voltage. Indeed, it tunes the electric field that it is responsible for the ReSURF. Figure 5.16 shows three output characteristics obtained biasing the FP with three different voltages. As expected, the start of the exponential growth of the leakage can be modulated by the applied FP bias. It regulates the local electric field in silicon. As a consequence, also the hard breakdown voltage (last point of each curve) anticipate or delay according to FP bias.

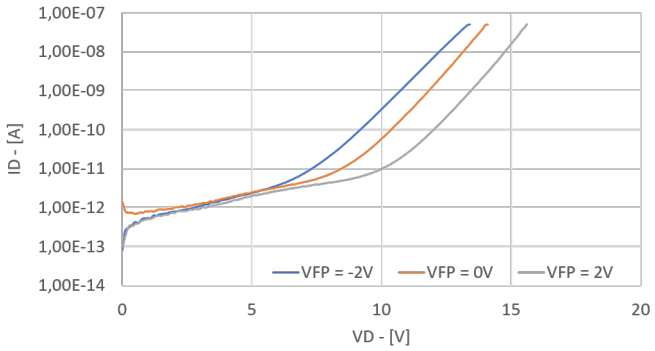


Figure 5.16: Experimental output characteristics at different FP bias.

In principle, the FP electrode can be drawn independent and accessible from outside, or shorted to the gate or the source. The first condition is the most flexible and effective, however it needs a complex control circuit which may be expensive. At system level, a control circuit can be designed to give a proper bias to the FP in the different working conditions. During the ON-phase, the FP should be positively biased minimizing the R_{ON} while during the OFF-phase it should be biased to maximize the BV_{OFF}. Shorting the FP to the gate, on the contrary, can be seen as an easy way to obtain at least improved R_{ON} without addressing externally the FP. In fact, in this way the FP is biased at 0V during the OFF-state and to 5V during the ON-state. However, this solution increases the gate capacity and consequently, the speed of the transistor and, more important, the switching losses. Finally, having

an FP shorted to the source has the advantage of not require any additional circuits and it does not affect the gate capacity. With a small penalization in the R_{ON} the metal field plate, even at 0 V bias, has a huge positive effect on breakdown capability, overcoming the limitation imposed by all-in-active LD-MOSFET without it.

Conclusion

The scaling of power devices does not follow the rules of the digital core. As described in chapter 2, power architectures have been differentiated in order to overcome the limitations due to the technology scaling with the goal to improve or at least maintain, generation after generation, their performances.

The goal of this work was to study the introduction of a new architectural element, i.e. the metal Field Plate, in a traditional all-in-active device with two main goals:

1. Improve the performance in the very low voltage range ($< 15\,\mathrm{V}$), where all-in-active devices are typically required to be competitive and where the limitations coming from the technology node scaling are more relevant.
2. Extend the voltage capability of all-in active architectures up to $30\,\mathrm{V}$, with performances competitive with respect to the traditional architectures that use a field oxide in the drain extension region.

The activities done consisted in:

- A preliminary huge TCAD activity to study and size the Field Plate geometry that was the input for all the following steps.
- The definition of the target dielectric stack thickness required as input for the integration activity. The 'smart' use of a contact array as field plate identified as the better solution to speed up the realization of the first prototypes on silicon.
- The design of the structures on a real masks set, i.e. the definition of their layout.
- The definition of the implantation conditions to be used on the electrical lot.
- The electrical characterization to prove the effectiveness of the proposed solution.

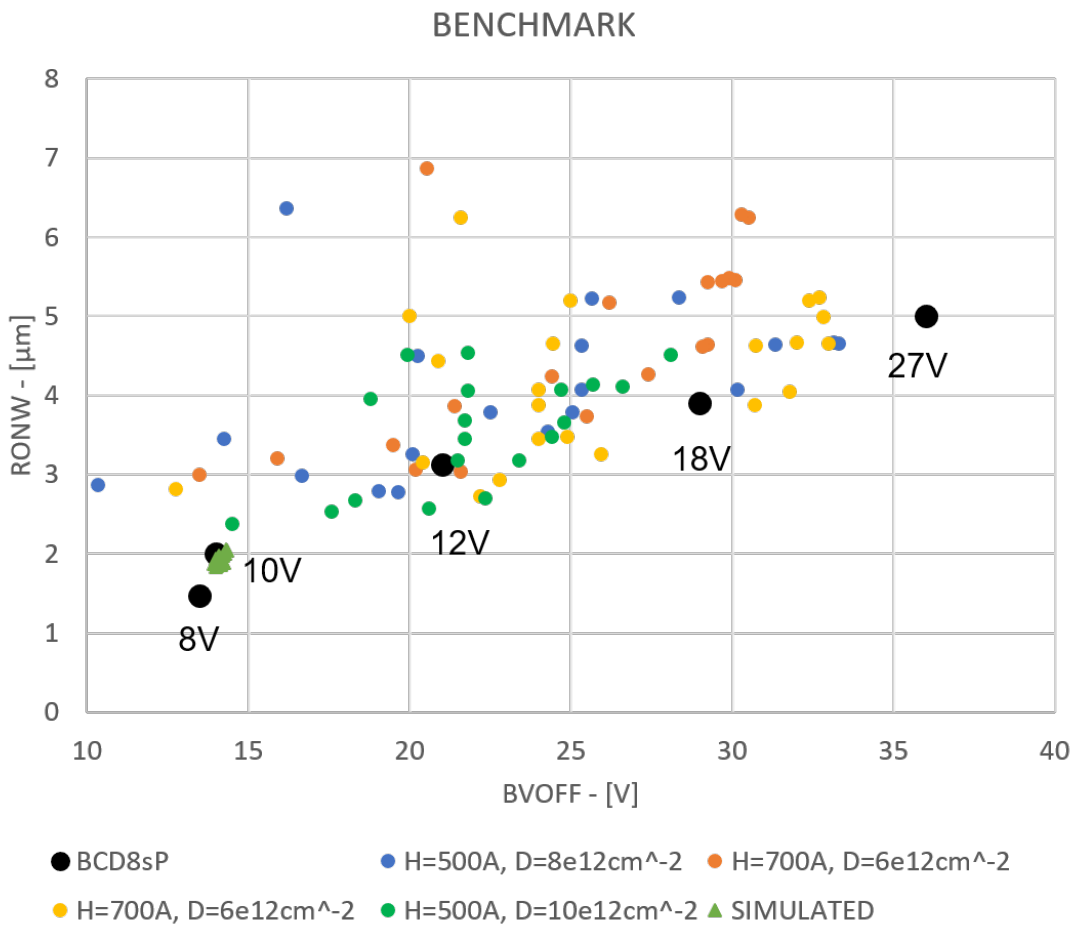


Figure 5.17: Benchmark with the low-voltage devices belonging to the BCD8sP technology.

Figure 5.17, showing the figure-of-merit $R_{ON} \cdot W$ vs BV_{OFF}, is a synthesis of the achieved results. The performances of the new architectures are benchmarked against the performances of the BCD8sP power devices, still considered as the best-in-class. For the new structures with metal Field Plate, all the experimental results are reported, divided for clearness among the different trials of FP heights and drain doses. As explained in the previous chapter, a structure is better than another one if it is in the bottom right part of the chart, i.e. if it has lower R_{ON} for a fixed BV_{OFF}, or if it has higher BV_{OFF} with the same R_{ON}.

Therefore, we can conclude that the new power architecture allows to extend the maximum voltage capability of an all-in-active device up to 30 V, with performances that are comparable with, and in some cases even better than, the BCD8sP ones in the full range of voltages. This is an important result in particular if we consider that these are the first prototypes.

Moreover, it is important to stress another important point that was only sketched in this work, focused on the R_{ON} optimization that it is the most important figure of merit of a power device. The structures reported in figure 5.17 are obtained, as described in chapter 5, with the overall gate length, and in particular the gate-drain overlap extension, that is half of the one used by the traditional architectures of BCD8sP. This means that the gate capacitance will be significantly smaller than the BCD8sP one. This is strictly related to another important figure of merit of the power devices, the charge that it is required to switch on and off the device (called Q_G), and it is a measure of the power losses during the power transition. In high frequency applications these losses could be comparable to the one related to the R_{ON}, i.e. to the losses during the power transfer to the load when the transistor is ON.

We can conclude that the integration of a metal field plate in an all-in-active power device promises to be a competitive solution for advanced BCD platforms. This work is only a starting point, but it puts the basis for further optimization of the device. Many challenges must be faced starting from the reliability assessment that was out of the scope of this work. However, the effectiveness that the metal Field Plate have to reduce the electric field in the critical region of the device, make us optimistic.

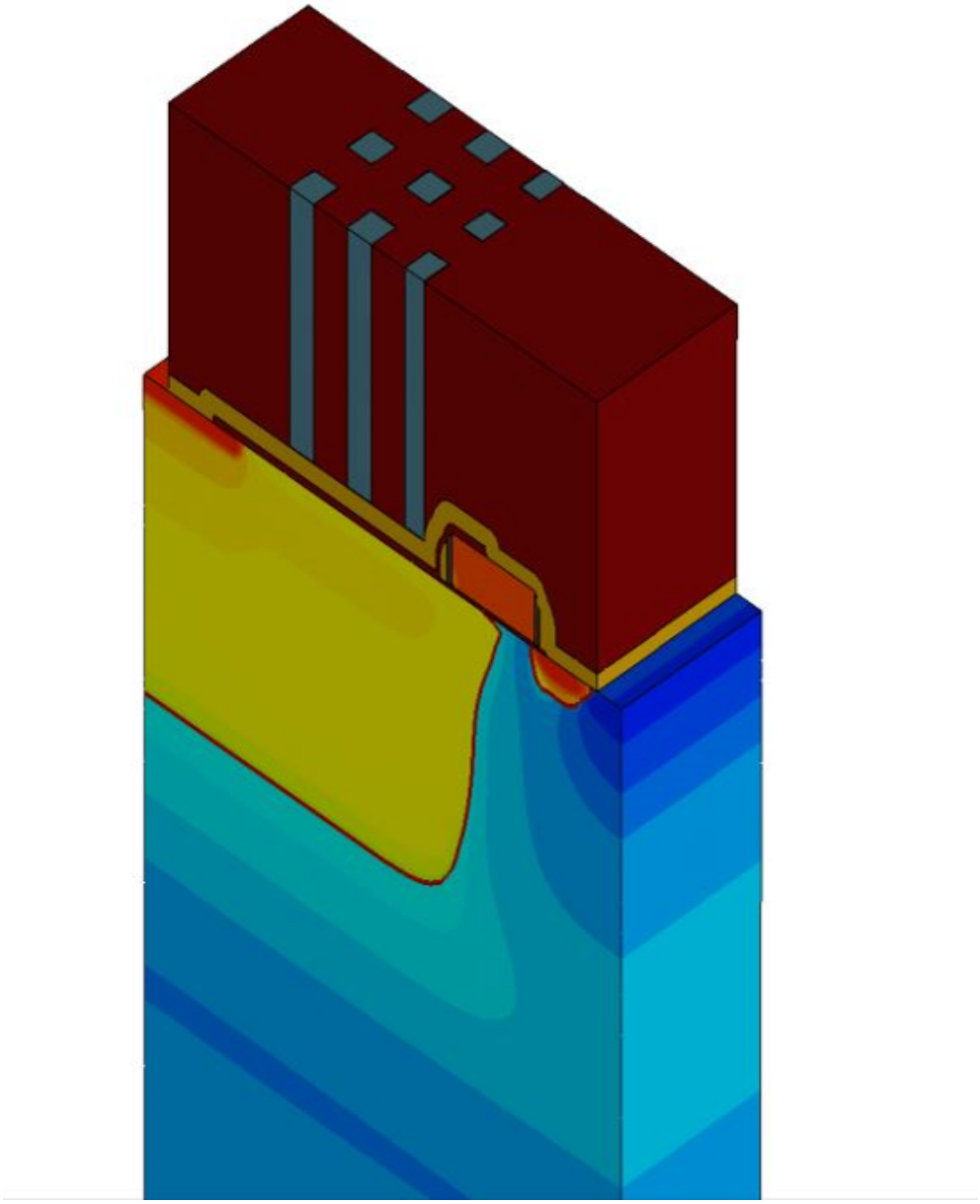


Figure 5.18: 3D picture of the LD-MOSFET architecture we have studied and developed.

Bibliography

[1] G. Majumdar. Recent technologies and trends of power devices. *International Workshop on Physics of Semiconductor Devices*, pages 787–792, Mumbai, 2007.

[2] A. Baiocchi. New developments in mixed bipolar/CMOS/DMOS technology for intelligent power application. *IEEE Colloquium on Integrated Power Devices*, pages 2/1–2/4, London, UK, 1991.

[3] P. Gueguen. How power electronics will reshape to meet the 21st century challenges? *IEEE 27th International Symposium on Power Semiconductor Devices IC's (ISPSD)*, pages 17–20, Hong Kong, 2015.

[4] 47th annual device research conference. *IEEE Transactions on Electron Devices*, 36(11):1519–1523, 1989.

[5] E. C. Niehenke. The evolution of transistors for power amplifiers: 1947 to today. *IEEE MTT-S International Microwave Symposium*, pages 1–4, Phoenix, AZ, 2015.

[6] C. Contiero A. Andreini and P. Galbiati. A new integrated silicon gate technology combining bipolar linear, CMOS logic, and DMOS power parts. *IEEE Transactions on Electron Devices*, 33(12):2025–2030, Dec. 1986.

[7] S. Sueri A. Russo B. Murari, R. Garibildi. Smart Power Technologies Evolution.

[8] Stmicroelectronics website: https://www.st.com/content/st_com/en.html.

[9] S. Pendharkar. Integration of substrate-isolated high voltage devices in junction isolated technologies. *Proceedings of 35th European Solid-State Device Research Conference (ESSDERC)*, pages 485–488, Grenoble, France, 2005.

[10] E. Aloni A. Eyal Y. Choi E. O. Arad, A. Parag and S. Shapira. Junction isolation for high voltage integrated circuits. *IEEE 27th Convention of Electrical and Electronics Engineers*, pages 1–4, Israel, Eilat, 2012.

[11] K. Oyama et al. Effect of field area on disturbance propagation through silicon substrates in SOI-BCD process. *International Symposium on Electromagnetic Compatibility - EMC EUROPE*, pages 1–5, Angers, 2017.

[12] M. N. Chil et al. Advanced 300mm 130nm BCD technology from 5V to 85V

with Deep-Trench Isolation. *28th International Symposium on Power Semiconductor Devices and ICs (ISPSD)*, pages 403–406, Prague, 2016.

[13] Piet Wessels. Smart power technologies on SOI.

[14] B. J. Baliga R. A. Kokosa M. S. Adler, K.W. Owyang. The Evolution of Power Device Technology. *IEEE Transaction on Electron Devices*, pages 20–21, 1994.

[15] P. A. Govindacharyulu S. R. Marjorie and K. L. Kishore. 2D analysis of self aligned LDMOS structures in terms of breakdown voltages. *IEEE Conference on Modeling of Systems Circuits and Devices*, pages 86–91, Hyderabad, India, 2019.

[16] Z. Yu and X. Zhao. A semi-analytical approach to the evaluation of threshold voltage in depletion MOS's with nonuniformly doped substrates. *IEEE Transactions on Electron Devices*, 35(7):993–998, July 1988.

[17] Richard S. Muller, Theodore I. Kamins. *Device Electronics for Integrated Circuit - 3rd edition.*

[18] F. Ferdous M. H. Bhuyan and Q. D. M. Khosru. A Threshold Voltage Model for Sub-100 nm Pocket Implanted NMOSFET. *International Conference on Electrical and Computer Engineering*, pages 522–525, Dhaka, 2006.

[19] Alessandro Diligenti, Francesco Pieri. Appunti di Strumentazione e Misure per la Microelettronica. *Servizio editoriale Universitario di Pisa*, 2007.

[20] S. C. Sun and J. D. Plummer. Modeling of the on-resistance of LDMOS, VDMOS, and VMOS power transistors. *IEEE Transactions on Electron Devices*, 27(2):356–367, Feb. 1980.

[21] J. C. W. Ng and J. K. O. Sin. Extraction of the Inversion and Accumulation Layer Mobilities in n-Channel Trench DMOSFETs. *IEEE Transactions on Electron Devices*, 53(8):1914–1921, Aug. 2006.

[22] E. J. Z. M. Sathi and Q. D. M. Khosru. An accurate model of inversion carrier effective mobility considering scattering mechanisms for nanoscale MOS devices. *International Conference on Electrical Computer Engineering (ICECE)*, pages 243–246, Dhaka, 2010.

[23] O. Gonzalez-C E. A. Gutierrez-D and R. S. Murphy-A. Electron transport through accumulation layers and its effect on the series resistance of MOS transistors. *Proceedings of the 1998 Second IEEE International Caracas Conference on Devices, Circuits and Systems. ICCDCS 98. On the 70th Anniversary of the*

MOSFET and 50th of the BJT, pages 51–54, Isla de Margarita, Venezuela, 1998.

[24] M. Severi G. Masetti and S. Solmi. Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon. *IEEE Transactions on Electron Devices*, 30(7):764–769, July 1983.

[25] C. Andre T. Salama Zahir Parpia. Optimizationof RESURF LDMOS transistor:an analytical approach. *IEEE transaction on electron devices*, 37(3):789–796, March 1990.

[26] R. J. E. Hueting A. Heringa J. Schmitz A. Ferrara, B. K. Boksteen and P. G. Steeneken. Ideal RESURF Geometries. *IEEE Transactions on Electron Devices*, 62(10):3341–3347, Oct. 2015.

[27] A. W. Ludikhuize. A review of RESURF technology. *12th International Symposium on Power Semiconductor Devices ICs. Proceedings*, pages 11–18, Toulouse, France, 2000.

[28] X. Zhou Z. Li M. Qiao, Y. Li and B. Zhang. A 700-V Junction-Isolated Triple RESURF LDMOS With N-Type Top Layer. *IEEE Transactions on Electron Devices*, 35(7):774–776, July 2014.

[29] Xin Zhou Jun Wang Zhuo Wang, Muting Lu and Bo Zhang. A novel Triple-RESURF SON LDMOS and its analytical model. *IEEE International Conference on Electron Devices and Solid-State Circuits*, pages 1–2, Chengdu, 2014.

[30] M. Qiao et al. Analytical Modeling for a Novel Triple RESURF LDMOS With N-Top Layer. *IEEE Transactions on Electron Devices*, 62(9):2933–2939, Sept. 2015.

[31] Bo Zhang Jie Wu, Jian Fang and Zhaoji Li. A novel double RESURF LDMOS with multiple rings in non-uniform drift region. *Proceedings. 7th International Conference on Solid-State and Integrated Circuits Technology*, 1:349–352, Beijing, China, 2004.

[32] A. S. Kluchnikov and A. Y. Krasukov. Application of Field Plate to Increase Breakdown Voltage of DMOS. *8th Siberian Russian Workshop and Tutorial on Electron Devices and Materials*, pages 107–108, Altai, 2007.