ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

INGEGNERIA E TECNOLOGIA DELL'INFORMAZIONE PER IL MONITORAGGIO STRUTTURALE E AMBIENTALE E LA GESTIONE DEI RISCHI - EIT4SEMM

Ciclo 36

Settore Concorsuale: 01/B1 - INFORMATICA

Settore Scientifico Disciplinare: INF/01 - INFORMATICA

MODELS DEVELOPMENT OF NEW CELLS PROTECTION FOR ELECTROSTATIC DISCHARGES INTEGRATED IN SMART POWER TECHNOLOGY

Presentata da: Laura Zunarelli

Coordinatore Dottorato

Luca De Marchi

Supervisore

Susanna Reggiani

Co-supervisore

Elena Gnani

Esame finale anno 2024

To all women in STEMM who are also daughters, partners and mothers.
May you do your best without giving up anything.

UNIVERSITÀ DI BOLOGNA

Abstract

ARCES-DEI

Doctor of Philosophy

Models development of new cells protection for electrostatic discharges integrated in Smart Power technology

by Laura Zunarelli

The demand for multiple functions to be implemented has increased the popularity of portable electronics and "smart-devices" featuring several sensing capabilities such as temperature, motion, current, light etc. Thus, a more sizable portion of the semiconductor industry finds the need for dedicated analog technologies, high power, high speed and high density logic. Electrostatic Discharge (ESD) protections have proven to play an important role in this chain of technology revolution since they represent a big portion of potential return causes the ICs manufacturing chain. The irreversible damage that a sudden release of current causes to an IC is of out most importance for the sake of reliability. Several different ESD clamps are available in literature, providing different solutions in terms of Safe Operating Area (SOA) depending on the final application. Bipolar-based solutions are widely used for ESD applications, due to their ability to exactly taylor triggering and holding voltages, while featuring excellent failure currents. However, power scaling is a major concern as the pulse width of the applied stress increases. Moreover, due to the presence of several diffusions tied to each other in the same structure, there is an high possibility to build parasitic paths to be identified and investigated during the device development and characterization. Similar hassles occur in pnpn based structures such as Silicon Controlled Rectifiers solution. These structures are obtained by adding a high-doped p-type diffusion within the drain well of a MOSFET, and the modulation of the drain extension can be further exploited to reach high triggering conditions. SCRs are capable of switching from a very high impedance state to a very low one, thus reaching very low holding voltages. Even though this characteristics makes it very , it also suffers from latch-up problems. Moreover, this structure is easly damaged by Kirk effect under transient high current, forming a strong electric field in corner regions and leading to breakdown damage.

In this thesis, a TCAD approach for the investigation of the safe operating area and the power to failure performances of trading ESD protection is presented. The

proposed approach is used to investigate power to failure condition for very short stress times, under 1ns, comparing the theoretical expected trends to the predicted and measured ones of the device under test. The physical reasoning of the lack in performance experiences in the proposed devices is presented together with a possible solution. The same proposed approach is applied also to other devices to study their holding voltage condition in correlation with the failure. In the first part of this thesis, a detailed review of the ESD models and characterization techniques is presented together with the physical models that are peculiar of and ESD event. In addition, ESD physics and operational mode of basic components is explained. In the second part, the operating conditions of a new ESD clamp based on BJT transistors is presented. The main physical mechanisms that drive the bipolar transistors into failure conditions during a very fast stress are analyzed and further investigated through geometrical modifications. In the third part, a conventional SCR-LDMOS ESD clamp is investigated as case study to learn the role played by the physical mechanisms and the two bipolar transistors when the device is in holding-voltage conditions. Several layout and geometrical modifications have been performed on the standard device to study and show the effect over the IV curve and the resulting failure conditions. The obtained performance of the device is compared with available literature data with by comparing their figures of merit to demonstrate and quantify the obtained advantages.

As a conclusion my thesis project has demonstrated the fundamental role played by the physical interpretation of the modern technology devices under high injection regime, self-heating and avalanche. This non-standard operating regimes have been poorly investigated before thus their physical reasoning is rarely available in literature. Moreover, this analysis enables us to produce and collect important information of these devices behavior in all stress regimes, from vfTLP to EOS conditions, characterizing the Power-to-failure conditions of the devices.

Contents

List of Figures

List of Tables

List of Abbreviations

BCD	**B**ipolar **C**MOS **D**MOS
CMOS	**C**omplementary **M**etal **O**xide **S**emiconductor
DMOS	**D**ouble-diffused **C**omplementary **M**etal **O**xide **S**emiconductor
DRIE	**D**eep **R**eactive **I**on **E**tching
ESD	**E**lectrostatic **D**ischarge
EOS	**E**lectrical **O**verstress
EMP	**E**lectromagnetic **P**ulses
IC	**I**ntegrated **C**ircuit
IoT	**I**nternet **o**f **T**hings
LDMOS	**L**ateral **D**iffused **M**etal **O**xide **S**emiconductor
LOCOS	**L**ocal **O**xidation **S**ilicon
NIL	**N**-type **I**mplant **L**ayer
NBL	**N**-type **B**uried **L**ayer
MOS	**M**etal **O**xide **S**emiconductor
PIL	**P**-type **I**mplant **L**ayer
RF	**R**adio **F**requency
SoC	**S**ystem **o**n **C**hip
TCAD	**T**echnology **C**omputer **A**aided **D**esign
TLP	**T**ransmission **L**ine **P**ulse

Chapter 1

Introduction

The success of portable consumer electronics (smart-phones, smart-watches, tablets, navigation devices etc.) has continuously incremented the analog functions to be implemented in such devices. Thus, customized analog functionalities are becoming a considerable cut of the semiconductor industry [1]. When dealing with these kinds of smart power technologies, high voltages are involved. The reliability issues related to such devices are enormous: starting from fabrication to their final application a large number of issues need to be addressed in order to prevent premature failure of devices.

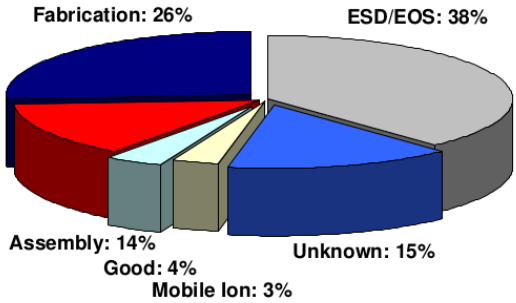


Figure 1.1: Field distribution return causes estimated in the IC device production market [2]

As shown in Figure 1.1, one of the main return causes in IC production is the occurrence of an Electrostatic Discharge(ESD)/Electrical over-stress(EOS) event. ESD is a subset of the broad spectrum of the EOS family, with EOS being defined by events including lightning and Electromagnetic Pulses (EMP), even though most of the device physics and the analytical modeling can be applicable to both ESD and EOS events. More generally, the ESD phenomenon is the sudden release of electricity from one charged object to another one when the two objects come into contact. It can arise from human handling or contact with machines, as it will be explained later on. These kinds of stresses are commonly known for their destructive effects on very large-scale integrated (VLSI) chips [3]. During an ESD event, the device is driven into high electric fields and high current densities in a relative small area.

Thus, the IC can be lead into breakdown of the insulator or even worse permanent thermal damage [4]. Therefore, the suppression of ESD negative effects has become high-priority precaution for all phases of the IC production and use. For this reason, most of IC chips are protected against these kinds of events by on-chip protection circuits, customized to rapidly discharge the high current associated with an ESD event, as show in Fig1.2.

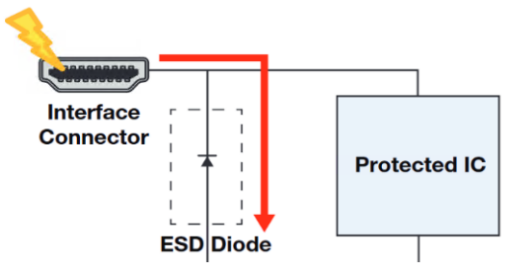


Figure 1.2: ESD strike with ESD protection circuit [2]

In this way, the design of the integrated circuits needs to address the ESD-protection as well. When the scaling of the technology started to take hold, almost 30 years ago, advances in protection designs were made by testing and failure analysis, creating the need of more tests and trials instead of speeding up the time-to-market. Lately, interest in ESD protection circuit design has grown, rising scientific effort in device modelling and electro-thermal simulations for protection effectiveness. However, several other features should be investigated for complete reliability. For example, the protection to different kinds of stresses. The generation of electrostatic charge inside ICs can mainly be related to four basic physical mechanisms. The first one is called tribo-electric charging and it occurs among objects with different Fermi energy where the object at the higher energy attracts an electron, acquires it, leaving the other object electrically charged; if the recombination is not immediate, this causes additional instances of contact and separation, thus increasing the total amount of charge, which builds up a higher voltage. Other environmental factors may influence this damaging mechanism.

The second process is called ionic charging. It occurs only if the flow of ionized gas molecules is not properly balanced or adjusted to the charging properties of the individual manufacturing process step; it can be easily controlled if the the the cell does not reach too high voltages. The third process is called direct charging and it consists of the direct flow of charges from a cable into an IC; it depends on the duration of the pulse induced inside the IC. The circuit might react in different ways, depending on the voltage differences reached along the charged path. Finally, the field-induced charging mechanism: it occurs when a charged object is brought near an uncharged one, electrically conducting, such as a piece of metal. The force of the nearby charge

due to Coulomb law causes a separation of these internal charges. For an IC, what happens is that an external electrostatic charging will eventually overexcite the device causing a sudden discharge if the IC touches another conductive object. This might cause a sudden release of current into the IC [4].

The tool to perform simulations of ESD events, is a conventional Technology Computer-Aided Design (TCAD) environment [5] which correctly predicts the stress level at which failure occurs due to breakdown or thermal runaway. Thus, this kind of software tools plays a key role for the development of new rugged devices. TCAD tools can be adopted to numerically simulate for the main steps of the semiconductor processing, the prediction of the electrical and thermal device operation and the interconnect characterization. It is used as a tool helping to design and optimize semiconductor processing technologies at any design stage. It supports a broad range of applications such as CMOS, power, memory, image sensors, solar cells, and analog/RF devices [6]. Moreover, it provides tools for interconnect modeling and extractions, providing critical parasitic information for optimizing chip performance. Overall, the TCAD tools provide a good way to explore new device structures, to select viable process and device development pathways, to optimize process modules and integration by fully exploring the process parameter space while reducing the number of experimental wafers and development cycles, to capture and analyze the impact of process variation on device performance, and to increase process capability, robustness and yield. For the purpose of this work, the electro-thermal characterization of a power devices up to their failure limits is needed. Sentaurus Synopsys is the most flexible and advanced platform for simulating electrical and thermal effects in a wide range of power devices such as IGBT, power MOS, LDMOS, thyristors, and high-frequency high-power devices based on wide band-gap materials [7].

Simulations of ESD protection cells allows to reconstruct the device Safe Operating Area (SOA) (Fig1.3). The ESD operating Area defines the edge barriers of the active operation of a device under ESD conditions. The ESD protection must shield the circuit even in un-powered OFF state without interfering with the normal behavior of the component. More specifically, a few key points should be extracted from the I-V curves, i.e., the breakdown voltage V_{t1},the holding voltage V_h, which is at the end of the snap-back regime and the R_{ON} of the protection cell under high current regimes; usually this resistance is measured by interpolating different values along the curve. Going further, the current increases until it reaches the second breakdown voltage V_{t2}; after this point the device reaches the thermal runaway and, eventually, starts to melt. The starting point of the thermal runaway onset is indicated in Fig. 1.3 as I_{T2}.

In this thesis, numerical simulations of new-generation ESD devices have been

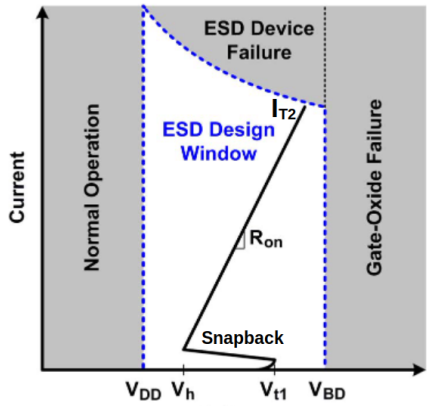


Figure 1.3: ESD design window defined by the power supply voltage (V_{DD}) of the IC, the failure level of ESD protection device, and the gate-oxide breakdown voltage (V_{BD}) of the MOSFET [2].

compared against experiments to quantitatively validate the proposed approach. The investigation of the failure mechanisms has been tackled by simulating their electrical characteristics in the framework of the Synopsys TCAD tool. Mainly, the physical models and thermal behavior of bipolar transistors internally stacked forming ESD protections has been addressed and thoroughly checked. The particular features of the Smart Power technology and the device geometry have been properly reconstructed. The core of my research activity is to build a TCAD methodology that can be adopted for all kinds of ESD tests, able to characterize and improve ESD protections, fully understand the physical events that lead the device into trigger, holding and thermal failure at different stress conditions. To this purpose, different pulse stress times were simulated in order to investigate the whole power failure characteristic, identified in the Wunsch Bell trend [8], as in Fig 1.4 and find any possible solution to up-rise their performances without the need of any fine tuning of the TCAD structure [9]. In Figure 1.4 the Power-to-Failure (P_f) dependencies with stress time can be briefly explained as follows: the adiabatic model ($P_f \propto \frac{1}{t_f}$) , typically below 10ns, the Wunsch-Bell model ($P_f \propto \frac{1}{\sqrt{t_f}}$) from 10 to 100ns and the Dwyer [9] ($P_f \propto \frac{1}{\log t_f}$), where the dependence of the failure power with respect to the stress time is practically constant.

More specifically, two different kinds of test-structures have been studied finding possible correlations and understanding the possible final application for each of them, depending on the achieved SOA. The first structure under test is a bipolar transistor. These devices are widely used in ESD application due to their ability to exactly tailor triggering and holding voltages, while featuring excellent failure currents, above 5 mA/µm. However, the failure level not always follow the expected

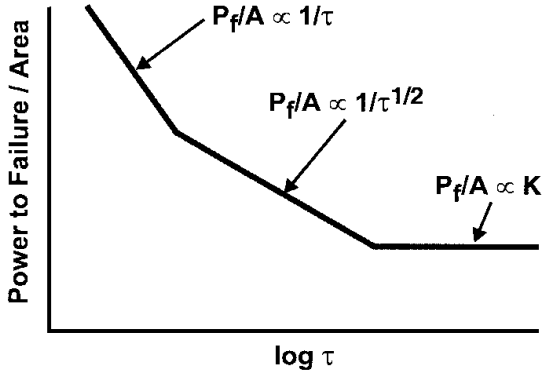


Figure 1.4: Wunsch-Bell curve: power-to-failure versus log pulse width general trend

power scaling. The main concern is related to the very fast stress, achieving relatively low failure current with respect to expected Wunsch-Bell trends. Moreover, several devices suffer from the latch-up risk that can be identified as a regenerative switching along a pnpn path, a common failure mechanism that is encountered in several CMOS ICs. The second test structure is a Silicon-Controlled-Rectifier (SCR) integrated into an LDMOS (drain extended device), obtained by adding a p-type diffusion within the drain well extension. In this way, a simple pnpn structure is formed. Its peculiar characteristic is due to the combination of a bipolar transistor followed by an SCR regime. The holding voltage levels of this device are analyzed through TCAD analysis.

Therefore, this thesis is organized as follows:

- **Chapter 2** : semi-classical description of the charge transport in semiconductor devices addressed by the drift-diffusion model. Main generation-recombination process will also be discussed, focusing on impact-ionization generation and all other correlated physical mechanisms. Finally a brief general discussion regarding numerical simulations is given.

- **Chapter 3** physical and operational characteristics of ESD protection elements are summarized in order to have a better understanding of the behavior of semiconductor devices under a high current stress event into known-topology. Industrial stress tests performed over ESD protection devices to meet the requirements are explained. Transmission line pulse characterization is described.

- **Chapter 4** The most relevant tests are implemented into TCAD tool, focusing mainly on the Transmission Line Pulse (TLP) approach. In particular, the study of two internally stacked BJT used as ESD protection under very fast transients.

- **Chapter 5** : Another ESD protection commonly known as SCR, based on a LDMOS structure, is simulated and its SOA is characterized. The focus is on the relation between snap-back condition, low holding voltage and high breakdown current. Several modification of the structure are simulated in order to understand and overcome the physical mechanism that disable any degree of freedom in the increase of the holding voltage.

- **Chapter 6**: final summary and conclusions on the activity.

Chapter 2

Preliminaries - Semi-classical transport models for semiconductors

2.1 Chapter overview

The description of the semi-classical charge transport models for semiconductor devices is reported with a special focus on the drift-diffusion (DD) model. Starting from Maxwell equations the description of the behavior of electric and magnetic fields and how they relate to each other is assumed. Provided that the current density is considered as the sum of the drift and diffusion contributions, the drift-diffusion equation can be derived [10]. The first one simply accounts for the Ohmic conduction and it is thus proportional to the electric field. The second contribution reproduces the movement of charges from the region with higher to the region with lower carrier concentration. A specific discussion is driven into the main generation-recombination processes, focusing on the impact-ionization responsible for the strong increase of the leakage current in reversed biased pn junctions. The analysis is mainly focused on the thermal behavior of the semiconductor devices due to the frequent occurrence of thermal effects on ESD protections.

2.2 Drift-Diffusion model

The drift-diffusion model can be derived from the Maxwell equations [11]:

$$\nabla \cdot D = \rho \tag{2.1}$$

$$H = \frac{\partial D}{\partial t} + J \tag{2.2}$$

$$\nabla \cdot B = 0 \tag{2.3}$$

$$E = -\frac{\partial B}{\partial t} + J \tag{2.4}$$

with:

$$B = \mu H \tag{2.5}$$

$$D = \varepsilon E \tag{2.6}$$

where ρ is the charge density, J is the current density and μ and ε are the magnetic permeability and dielectric permittivity in the material, respectively. Taking the divergence of Eq. (2.2) and remembering that $\nabla \cdot \nabla \times H = 0$, we obtain the continuity equation of the carrier density:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot J = 0 \tag{2.7}$$

with $\rho = q(p - n + N_D - N_A)$, where N_D and N_A are the donor and acceptor dopant concentrations, respectively, and n (p) is the electron (hole) concentration in the valence band. In writing the expression of ρ we are implicitly assuming that we are dealing with a crystalline semiconductor. Total current density J is the sum of the electron and hole current densities, namely $J = J_n + J_p$. Hence, one can split Eq. (2.7) in two different equations, one for each type of carriers [10], [12]:

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla \cdot J_n = W_n \tag{2.8a}$$

$$\frac{\partial p}{\partial t} + \frac{1}{q}\nabla \cdot J_p = W_p \tag{2.8b}$$

W_n is the total generation rate for the electrons, namely the difference between the number of electrons entering and leaving the conduction band. It can be expressed as the difference between the generation and recombination rates of the electrons $W_n = G_n - U_n$. Similar considerations can be applied to $W_p = G_p - U_p$ for holes. Assuming that the different generation-recombination processes are uncorrelated and taking the steady state approximation for the trap populations, one finds:

$$U_n - G_n = U_p - G_p = U_{SRH} + U_A + U_{II} + U_D \tag{2.9}$$

where U_{SRH}, U_A, U_{II} and U_D represent the net recombination rates associated to thermal recombination, the Auger recombination, impact-ionization generation and photon induced direct transitions, respectively. U_D will be neglected since optical direct transitions are not considered in this study.

In the drift-diffusion picture, the electron and hole current densities can be expressed as:

$$J_n = q\mu_n nE + qD_n n \tag{2.10a}$$

$$J_p = q\mu_p pE - qD_p p \tag{2.10b}$$

where μ_n (μ_p) is the electron (hole) mobility and D_n (D_p) is electron (hole) diffusion coefficient and, in principle, both depend on the electric field. Under non-degenerate conditions (the Boltzmann statistics can be used as an approximation of the Fermi-Dirac distribution), the Einstein relationship relates these two quantities:

$$D_{n(p)} = \mu_{n(p)} \frac{k_B T}{q} \tag{2.11}$$

where k_B is the Boltzmann constant and T the absolute temperature of the lattice.

In writing Eq. (2.10) we are implicitly neglecting the magnetic effects ($B = 0$), as no specific role is assumed to be given to the magnetic field application. Otherwise, there would be an additional term containing the magnetic field in the expression of the current densities. The set of equations 2.1, 2.6, 2.8, 2.10 are non-linear first-order partial differential equations in the unknowns E, D, n, p, J_n, J_p. They must be recast in a more compact form to be solved. To do so, the quasi-static approximation is introduced to write the electric field as follows:

$$E = -\nabla\varphi \tag{2.12}$$

where φ is the electric potential. This approximation is valid in semiconductor devices at the typical operating frequencies. Secondly, we introduce the quasi-Fermi potentials φ_n and φ_p. They are auxiliary functions that generalize the concept of Fermi potential under non-equilibrium conditions. In the equilibrium limit they coincide with the Fermi potential φ_F. In this way we can express the electron and hole concentrations in terms of the quasi-Fermi potentials:

$$n = n_i \exp[\frac{q(\varphi - \varphi_n)}{k_B T}] \tag{2.13a}$$

$$p = n_i \exp[\frac{q(\varphi_p - \varphi)}{k_B T}] \tag{2.13b}$$

where n_i is the intrinsic carrier concentration.

Using the Einstein's relations (2.11) and the definitions (2.13), the Poisson equation and the current density equations for electron and holes can be recast in a more suitable form for the numerical simulations of semiconductor devices:

$$-\varepsilon\nabla^2\varphi = q(p - n + N_D - N_A) \tag{2.14}$$

$$\frac{\partial n}{\partial t} + \nabla \cdot (\mu_n n \varphi_n) = W_n \tag{2.15}$$

$$\frac{\partial p}{\partial t} - \nabla \cdot (\mu_p p \varphi_p) = W_p \tag{2.16}$$

The set of these three equations constitute the drift-diffusion model. It is worth mentioning that, since very high fields are involved, so that electrons may become hot (i.e. acquire high kinetic energy), other conduction models for transport, such as the hydrodynamic model (HD), may prove effective, in principle. The HD model is a more complex and detailed model that takes into account additional effects such as momentum transfer, energy exchange, and carrier-carrier scattering. However, its computational complexity and high-performance computing requirements makes it not suitable for routine devices simulations [13]. However, the first degree of approximation of the drift-diffusion model is reliable.

2.3 Approximated thermal model

In addition to thermo-electric power, another important thermal effect in semiconductors is thermal conduction. The heat flow Q is driven by the temperature gradient in a diffusion type process as follows:

$$Q = -k\frac{dT}{dx} \tag{2.17}$$

where k is the thermal conductivity with two major components for both electrons and holes:

$$k = k_L + k_M \tag{2.18}$$

as k_L is the component of phonon (lattice) conduction, while k_M is the mixed free-carrier conduction. The lattice contribution is carried out by diffusion and scattering of phonons. The scattering events include different typed of occurrences such as phonon-to-phonon, phonon-to-defects, boundaries and surfaces and so on and so forth [12].

If we consider a simple Silicon-based pn junction in reverse bias condition, most of the voltage drop is across the junction itself. Considering the situation as plane source at junction x=0 in an infinite medium where both thermal conductivity and specific heat rate are proportional to temperature, an approximated thermal model

can be derived [14]. Thus, the general one-dimensional heat equation is:

$$\frac{dQ}{dx} - \rho C_p * \frac{dT}{dt} = 0 \tag{2.19}$$

where Q is the heat flow in 2.17 k is the thermal conductivity in 2.18 and C_p is the specific heat. However, semiconductor actual junction geometries strongly depend on the specific device. Moreover, the failure mechanisms for the real junction shape is not necessarily consistent. Thus, after some approximation assumptions, the temperature variation equation is described by:

$$\frac{d_2 T}{dx^2} - \frac{1}{\alpha}\frac{dT}{dt} = 0 \tag{2.20}$$

The solution to 2.20 is the temperature failure of the junction [8].

2.4 Generation Recombination processes

Inter-band generation-recombination processes allow the transition of electrons from the valence to the conduction band and vice-versa. We can distinguish between direct processes and trap-assisted. In the former case, the transition occurs directly between the two bands (Auger recombination, impact ionization generation), while in the latter it is assisted by the presence of trap levels into the band gap which originate from defects in the lattice crystal of the semiconductor (Schockley-Read-Hall theory) [15].

2.4.1 Schockley-Read-Hall recombination

The Schockley-Read-Hall (SRH) recombination describes the trap-assisted thermal recombination assuming the presence of a single trap level with energy E_T and a steady-state condition. With the presence of more than one level, the contributions of each level should be summed at the end of the calculation. However, it can be shown that the most efficient value of E_T is located approximately at the mid-gap, neglecting the other trap levels. It is usually expressed as:

$$U_{SRH} = \frac{np - n_{eq}p_{eq}}{\tau_{p0}(n + n_B) + \tau_{n0}(p + p_B)} \tag{2.21}$$

where n_{eq} (p_{eq}) and τ_{n0} (τ_{p0}) are the electron (hole) equilibrium concentration and lifetime, respectively, and n_B and p_B are given by the following expressions:

$$n_B = \frac{n_{eq}}{d_T} \exp(\frac{E_T - E_F}{k_B T}) \tag{2.22a}$$

$$p_B = p_{eq} d_T \exp(\frac{E_T - E_F}{k_B T}) \tag{2.22b}$$

where E_F is the Fermi energy and d_T the degeneracy coefficient of the trap. Recombinations is larger than generation (and vice-versa) when there is an excess of np with respect to equilibrium [12].

2.4.2 Auger recombination

Two electrons into the conduction band may collide and exchange energy. If at the end of the collision one of the electrons exhibits a loss of energy equal or greater than the energy gap, it experiences a transition into an empty state of the valence band. The other electron acquires the same amount of energy and goes into a higher energy state in the conduction band. This process is usually called Auger recombination initiated by electrons. Analogue considerations can be made for a collision between two holes into the valence band. The Auger recombination for electron and holes can be written as:

$$U_{An} = c_n n^2 p \tag{2.23a}$$

$$U_{Ap} = c_p p^2 n \tag{2.23b}$$

Where c_n (c_p) is the Auger recombination coefficient for electrons (holes). The probability that two electrons in the conduction band or equivalently two holes in the valence band can collide is usually low at room temperature since the probability of a collision of a charge carrier with phonons is much higher. The Auger recombination is dominant in heavily doped regions with a high charge density, and as a consequence, low electric field. For this reason it is very important in our analyses since the electron and hole concentrations are always high together with the electric fields of interest in high injection conditions.

2.4.3 Impact-ionization generation

Impact-ionization transitions occur when an electron whose initial state is in the conduction band at high energy, collides with another electron in the valence band. After the collision, the first electron is still in the conduction band but with a lower energy, while the second one acquires enough energy to transit into the conduction band.

This process is called impact-ionization initiated by electrons. The impact-ionization generation for electrons and holes can be written as:

$$U_{IIn} = -I_n n \tag{2.24a}$$
$$U_{IIp} = -I_p p \tag{2.24b}$$

I_n (I_p) is called impact-ionization generation coefficient for electrons (holes). The carrier generation due to impact-ionization occurs only when an electron or a hole acquires an energy larger than the energy gap. This happens only in presence of a strong electric field which provides a sufficient amount of kinetic energy to the charge carrier over a distance much shorter than the mean free path in the material. As a consequence, impact-ionization is dominant in regions with low carrier concentrations and a high electric field and for this reason it plays a relevant role in our analyses.

Electron-hole pair production due to avalanche generation requires a certain threshold field and the possibility for the carriers to accelerate, that is, wide space-charge regions. If the width of a space-charge region is greater than the mean free path between two ionizing impacts, charge multiplication occurs, which can cause the electrical breakdown. The reciprocal of the mean free paths are called impact-ionization coefficients α_n and α_p, for electrons and holes, respectively. Their mathematical definition follows by assuming a condition far from equilibrium, when impact-ionization dominates over the other generation-recombination processes. Hence, $U_n - G_n = U_p - G_p \cong U_{II} \cong -I_n n - I_p p$. Under steady-state conditions, the continuity equations become:

$$\nabla \cdot J_n = -qnI_n - qpI_p \tag{2.25a}$$
$$\nabla \cdot J_p = qnI_n + qpI_p \tag{2.25b}$$

Assuming that the conduction term is dominant over the diffusion term due to the high electric field yields to $J_n \cong q\mu_n nE$ and $J_p \cong q\mu_p pE$. Rewriting the current densities as $J_n = J_n e$ and $J_p = J_p e$ where $e = E/|E|$, Eq. 2.25 become:

$$-\nabla \cdot J_n = \alpha_n J_n + \alpha_p J_p \tag{2.26a}$$
$$\nabla \cdot J_p = \alpha_n J_n + \alpha_p J_p \tag{2.26b}$$

where the impact-ionization coefficients are:

$$\alpha_n = \frac{I_n}{\mu_n |E|} \tag{2.27a}$$

$$\alpha_p = \frac{I_p}{\mu_p |E|} \tag{2.27b}$$

2.5 TCAD-based numerical simulations

The drift-diffusion and thermo-electrical models described so far can be solved analytically only in a few simple cases. On the contrary, when the complexity of the device geometry increases, numerical simulations are fundamental to understand the effects of each physical model on the behavior of the device. Different software suites are today available to reproduce the different levels of abstraction, including: process, device and circuit simulations. Process simulations allow to reproduce ion implantation, dopant diffusion and oxidation, etching and lithography steps. However, when the details of the fabrication processes are not relevant to the device analysis, as in our case, the structure of the device can be recreated in the simulator without process simulations, provided that the geometry and doping configuration are known. This is usually achieved by using specific tools available into the software suite. The recreated device is then used as input for the device simulations in which the drift-diffusion model is solved numerically. The device is divided into a discretized structure called mesh. Thus each differential equation of the system is solved at each vertex of the mesh. The Poisson equation (Eq. 2.14), the transport equations for electrons and holes (Eq. 2.15 and Eq. 2.16, respectively) are solved iterative until the convergence is achieved. The current-voltage (I-V) characteristics can thus be extracted and compared against measurements performed on real devices, having a better understanding of the physical mechanisms that take place inside the devices. In this work, the commercial suite provided by Synopsys Inc. has been used [7]. Process simulations have not been considered since all the information about geometry and doping were given by the technology property. Sentaurus Structure Editor has thus been used to define the device and the mesh structure. The electrical characteristics have been simulated through Sentaurus Device solving numerically the drift-diffusion and heat flux models.

2.6 Concluding remarks

In this chapter, the main generation-recombination processes, focusing on the impact-ionization are reported. Since the most relevant transport mechanisms in ESD protection devices is expected to be the avalanche generation due to the electron impact-ionization, the physical mechanism causing the breakdown of the parasitic bipolar transistors is analyzed, along with other contributions. The TCAD tool allows us to have a complete physical representation of the transport properties since many models are already implemented on Silicon. The latters do not require any particular parameter fitting since they have been previously calibrated on Silicon.

Chapter 3

ESD models and structure peculiarities

3.1 Chapter overview

A general introduction of the ESD world is given in this chapter. The research and development groups of the main industrial chip fab competitors have studied and disclosed the ESD features of well-known components and ushered standard stress condition under which all new ESD clamps should comply. These test models provide an essential guideline not only for ESD protection designers but for all circuit designers.

3.2 ESD models and characterization techniques

The main mechanisms of electrostatic charge generation have been explained briefly in Chapter 1. In this section the stress models used to emulate electrostatic voltages are delineated. The latter ones are standard configurations used in Companies for testing purposes on ESD cells.

3.2.1 ESD industrial stress

The stresses induced into ICs by an ESD event have been classified considering the mechanism that causes the stress. The circuit is set to be un-powered while the sudden release of current takes place. The phenomena that are emulated by the injected spike of current are explained in the following sections:

CDM - Charge Device Model [16]

The conventional model used to characterize ESD stresses is the Charged Device Model (CDM), with operating conditions around 250V to 500V. It is the most frequent discharge mechanism in an automated handling environment. The device package is positioned upside-down and previously charged by a field charging plate; the parasitic capacitance that are created between the plate and the IC package are gradually discharged over a pin called "pogo pin". In such a situation, a very short and high current spike may occur due to the capacitance of the device, the inductance of the pin, and the resistance of the ionized channel, consequently causing damage to the circuit. Pulses of few ns duration and amplitudes up to 10A are generated in most cases.

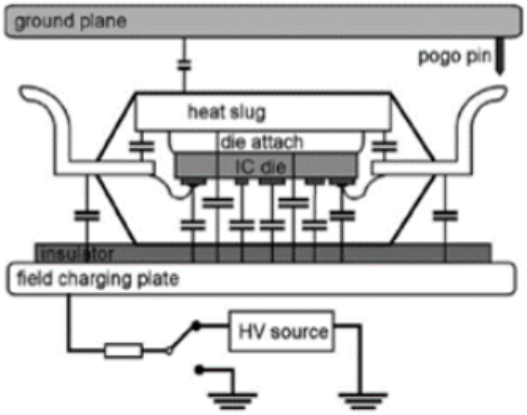


Figure 3.1: Charge Device Model illustration [16]

For a more immediate understanding, this event is exemplified in Fig 3.1. The device is in contact with both a charged relay and a discharged relay (Field charging plate and ground plate); this model is called contact mode. The capacitance value is highly influenced by the package and by the presence of an air gap (or other dielectric) between the package and the ground plane. The voltage difference is determined by the resistance generated by the voltage source and the one generated by the isolation between the chip the ground plane. If the device was previously discharged, the voltage is determined by capacitive voltage suppression. The area that surrounds the discharge loop, consisting on bond wire, pin and relay, affects the inductance of the discharged path.

HBM - Human Body Model [16]

Human Body Model (HBM) is the idealisation of the phenomenon related to the touch of an IC by an operator to plug it into a board. In other words, it might happen

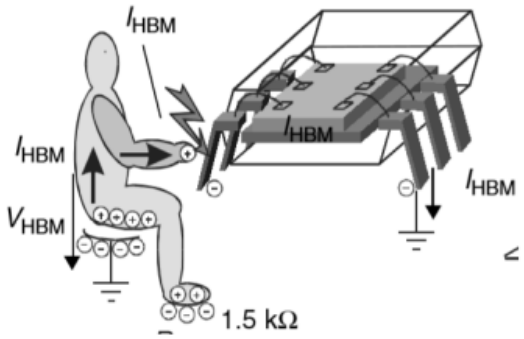


Figure 3.2: Human Body Model set of situation [16]

that the charged capacitance of the person's body discharges through the IC. This circumstance is shown in Fig 3.2. Without any protection, the gate oxide facilitates the IC voltage increase, which can reach the breakdown region of the devices. This is modeled by the HBM test, which is given by a current pulse injected into the circuit with specific raise and fall times and a minimum bandwidth of 350MHz, for single shot events. The circuit models an ideal pulse of 150 ns duration. The amplitude of such signal for the HBM operating conditions is in the range of 1000 V – 2000 V.

MM - Machine Model [16]

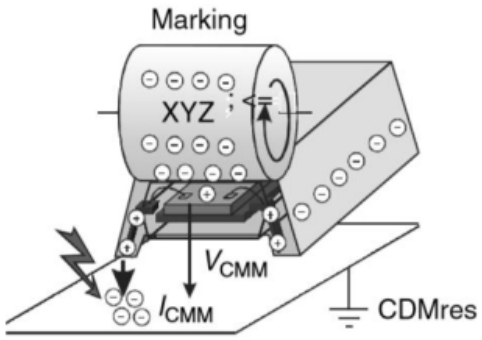


Figure 3.3: Machine Model circumstances [16]

Machine Model (MM) emulates IC stresses caused by sources with high current discharges that come from the operation of a machine over the device during fabrication; an example is shown in Fig 3.3. These events are not necessarily classified as ESD ones, so this mechanism is not representative for ESD in an automated handling environment. Both HBM and MM models involve at least two pins and typically generate power-related failures in pn-junctions. With respect to HBM model, the MM is related to failure in the pn-junction which occurs at pre-charged voltages lower than threshold failure. For this reason, HBM tests need be carried out after the MM, as

if MM fails, HBM will fail as well. The difference of the three models in terms of injected pulse into the protection in un-powered conditions is shown in Fig 3.4.

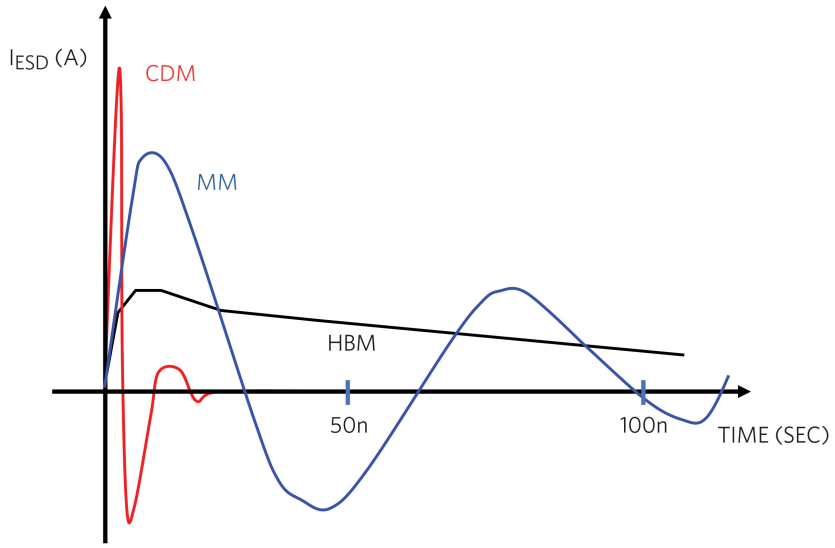


Figure 3.4: CDM, HBM and MM pulses [16]

It can be guessed that the spike that the CDM stress injects can cause significantly differing effects in the system with respect to the HBM and MM ones. It has also to be considered that wave-forms are usually unintentionally altered by different environmental sources, like, e.g., interference between probe and signal, dispersion in cables, electromagnetic interference etc. Thus, a difficult challenge for simulation purposes is to reproduce the same effect considering the worst case scenario given by all possible kind of environmental factors. On top of all, each different type of protection has a different kind of side effect with different levels of sensitivity to the stress.

3.2.2 Transmission Line Pulse characterization [17]

Transmission Line Pulsing (TLP) is a technique which allows to measure the dynamic and the quasi static behavior of a single device, I/O cells or the entire IC behavior applying square pulses of different amplitude and length. TLP is used for deriving the device isothermal quasi-static I-V characteristics in combination with the evolution of the leakage current, for increasing amplitudes of the square pulses. Typically, a 100-ns pulse width is considered, with 10-ns rise time. The main outcome given by the TLP I-V curves is usually correlated with HBM even though the TLP-tester provides excellent handling and the analysis of significantly more data can be carried out to better understand the behavior of the device under test.

The circuit set-up is shown in Fig 3.5: it should ensure a constant injected current.

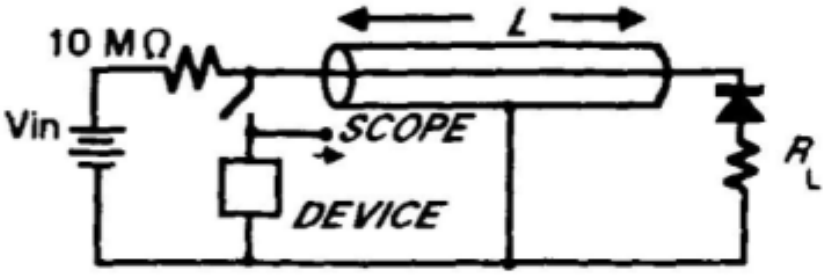


Figure 3.5: Simplified TLP setup circuit [18]

Moreover, in order to generate a square pulse of prescribed duration and amplitude, the distributed capacitance of the Transmission Line (TL) is charged and discharged. For this reason, in order to allow the injection into the circuit of a controlled current or voltage pulse, a coaxial wire of a suitable length, depending on the pulse duration, is needed [18].

Pulsed characterization techniques are necessary since, otherwise, a DC characterization causes strong self-heating. Indeed, electrostatic discharge cause a high current, a fast event with a complex waveform. The transmission line pulses are very similar to the HBM stress tests and have been demonstrated to be extremely useful in understanding the device behavior during an ESD event. Each pulse must be designed carefully to analyze the device over time until failure occurs.

In this way, key parameters like trigger voltage, holding voltage, failure voltage and failure current can be extracted with high accuracy. More specifically, after inserting the device under test (DUT) into the test fixture or contacting the wafer, the leakage current needs to be measured before and after each pulse. The increase of leakage current is considered as a failure criterion in measurements. Moreover, it is crucial to select the bias voltage carefully in order to avoid additional stress. Depending on the final goal, the whole wafer can be tested, according to this technique.

When dealing with simulation, the same setup can be used, simplifying the injection of current as current generator simulating different pulses of different lengths and amplitudes (see Fig 3.6). This technique is commonly used even though it can be time-consuming since each point of the characteristic is a point that has to be simulated.

3.2.3 From very fast TLP (vf-TLP) to EOS stress pulses

Depending on the length (t) and rise time (t_r) of the pulse, the behavior that is expected from the tested device is obviously different and related to the correlated industrial stress tests described in 3.2.1. The attention given in the literature community to current densities and pulse widths as overcome the consideration attributed

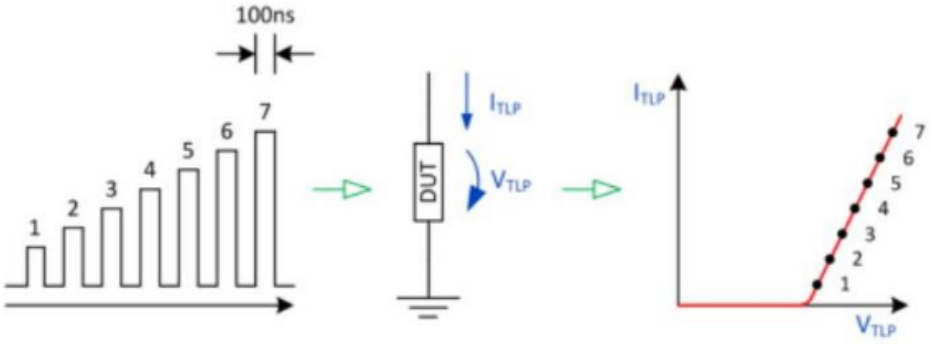


Figure 3.6: TLP simulation setup

to impact of the t_r . As a matter of fact, fast transients can lead to different failure modes than pure HBM and MM stress. These rise-time effects can be attributes to the devices' dynamics that can build overshoots at the beginning of the pulse (at t<t_r). Even though in the majority of the cases these effects are not harmful for the device, several situations the rise time effects may actually damage vulnerable circuitry; thus, it has to be verified that the overshoot given by fast transients is not dangerous on the IC. In [19], several experiments were performed for various TLP and vf-TLP configurations with data taken on wafers. With the proposed solutions, it was possible to determine the suitability of protection devices for fast ESD transients as CDM, system level ESD and real HBM.

3.3 Macroscopic Physical models

3.3.1 Thermodynamic physical model of ESD stresses

The macroscopic physical model to detect the failing conditions are reported. The main mechanisms by which a semiconductor junction may fail due to an ESD even are avalanche breakdown around the surface junction and internal avalanche breakdown through the junction within the body of the device [4]. These are the failure mechanisms which are supposed to turn on ESD protection devices in order to prevent the system from reaching thermal failure: after the onset of avalanche breakdown, the junction is likely to increase the current significantly causing, as expected, self-heating up to the melting point of the crystalline material. As a matter of fact, in internal body breakdown, the destruction mechanism apparently results from changes in the junction features. Due to high temperatures the junction area is either totally destroyed or its conduction properties are drastically changed. The localized points where this event takes place are identified as hot spots. For this reason, the theoretical treatment of the semiconductor devices in failure conditions can be reduced to a

thermal analysis.

Failure levels can be certainly significantly affected also by other factors, such as skin effect, non-homogeneous local fields or ion migration, but hot spot is considered as the most relevant factor [9]. Moreover, in order to get a more general insight on the issue, it is necessary to find a set of equations which does not directly depend on the geometry of the specific device. Thus, it is possible to establish a general law which all devices should ruffly follow.

3.3.2 Wunsh-Bell model

One of the most common theories used to describe ESD thermal behavior is the well-known Wunsh-Bell model [8]. One of the most important features of the Wunsh-Bell model is that it describes the relationship between the thermal runaway and the current pulse injected in the system, using some approximated model to generalize the failure mechanism in semiconductor devices. In fact, the failure level (failure power and failure energy) in a semiconductor structure due to thermal run-away strongly depends on the pulse width [20]. Besides the failure level, also the on-resistance (R_{ON}), which directly affects the transient response of the ESD protections, is a parameter that needs to be considered: since it measures the intrinsic conduction losses of a device, it is usually measured to check the health and efficiency of a device, testing deterioration due to the heating of the device [21].
For the Wunsch-Bell model, the junction failure condition strongly depends on the duration of the pulse as it is shown in the following equation:

$$\frac{P}{A} = \sqrt{\pi k \rho C_p}(T_m - T_i) * t^{-\frac{1}{2}} \tag{3.1}$$

where P is the power, A is the area, $\frac{P}{A}$is the power per unit area, k is the thermal conductivity, ρ is the mass density, C_p is the specific heat, T_m is the junction failure temperature and T_i is the initial junction temperature, t is the stress time. The adiabatic dependence $\frac{1}{t_f}$ is derived for TLP simulation setups considered as standard case scenario for failure stress conditions.

In the Wunsh-Bell model, the thermal equation is used to predict the semiconductor junction failure as it provides the time dependence to failure. From a measurement point of view, the time to failure is obtained by a step voltage pulsing the junction until failure or when a degradation of the device occurs in reverse mode. Consequently, localized melting occurs, especially across the junction where the main voltage drop is experienced, or localized moving current filaments originate from the hot spots. Thanks to experimental measurements, it is possible to compare results obtained with expected result given by previous equation and test the effective efficiency of

a specific device-under-test. Considering equation (3.1) equation as a general time dependence, it is possible to build a model based on three cases:

- Heating from 25 ° C to 675 ° C
- Heating from 700 ° C to 1415 ° C, which is roughly Silicon melting temperature.
- Identical from the second case except the current density is assumed to go through local hot spots (so roughly one tenth of the overall area of the device).

This model results in three theoretical curves shown in Fig 3.7 providing a straight line on log-log plots and hence it allows one to quickly estimate failure levels by plotting the power versus pulse time duration for a known junction area. The form of the solution has many advantages. It allows one to compare devices of different junction areas and various junction geometries by plotting the power per unit area. Finally, it is possible to reconstruct an analytical semi-empirical solution of the heat equation and its dependency on time, as mentioned before. More specifically, the power dissipation which depends on the self-heating of the device can be plotted. This dependency changes with respect to the width of the current injected pulse. Another important issue is given by the type of stress that is induced in a charge device model stress (CDM) or Human Body model stress (HBM). In the case of ESD simulations and, more in general, EOS simulations (thus with longer stress times), this plot has an exponential behavior as shown in 3.7.

As an experimental result, it is possible to estimate the time to failure by analyzing the time-dependent behavior of the device.

3.4 Physical behavior and operational characteristics of ESD protection devices

Devices subjected to modes correlated to ESD stress are very different with respect to normal operating conditions. For this reason, it is very important to understand the phenomena taking place in the IC during an ESD event (very high current event that occurs for very short duration)and the main parameters governing their performance. In order to fully understand the behavior of each device in ESD mode, it is important to remember their safe operational working conditions first.

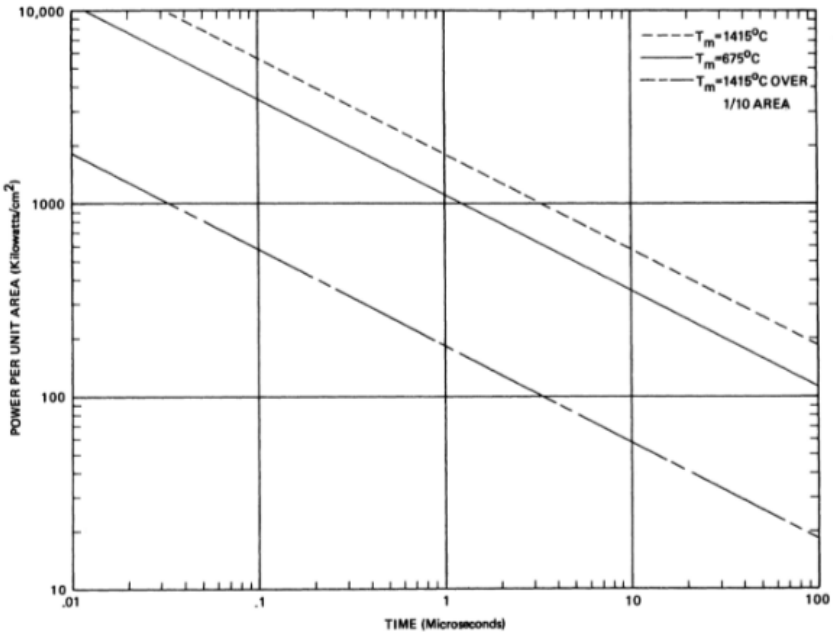


Figure 3.7: Theoretical failure curves of Silicon junctions studied in [8]

3.4.1 Resistors

A basic resistor provides a good overview on every possible operating condition that also apply to other components. Considering, for example, an N-type doped resistor, as it is the most common one for ESD applications, it is important to recall some basic equations which highlight the most important dependencies. The conductivity of a semiconductor depends on the carrier mobilities and concentrations. In general both electrons and holes take place and the equation reads:

$$\sigma = nq\mu_n + pq\mu_p \tag{3.2}$$

where

$$n = N_c e^{\frac{(E_c - E_f)}{kT}}] \tag{3.3}$$

and

$$p = N_v e^{-\frac{(E_f - E_c)}{kT}} \tag{3.4}$$

E_f is the Fermi energy, E_c the conduction band energy and E_v the valence band energy; these equations are know as the Boltzmann relationships. N_c and N_v are the effective density of states, k is the Boltzmann constant T is the temperature, $q = 1.6021917 * 10^19[C]$ is the elementary charge, μ_n and μ_p are the electron and

hole mobilities, respectively.

Assuming as an example an n-type resistor, the resulting current density depends mostly on the electrons as they are the majority carriers and the total current density J_n, which is the major one reads:

$$J_n = nqv_d \tag{3.5}$$

with $v_d = \mu_n E$ being the drift velocity of the carriers and N_d the donor (n-type) doping concentration. The drift component will eventually saturate while increasing the applied voltage, thus the applied electric field will change the behavior of the resistor from its Ohmic (linear dependence between current and voltage) to saturation regime. In the latter case, the electric field is the only parameter that is increasing. Thus, the current density equations reads:

$$J = J_{sat} = nqv_{sat} \tag{3.6}$$

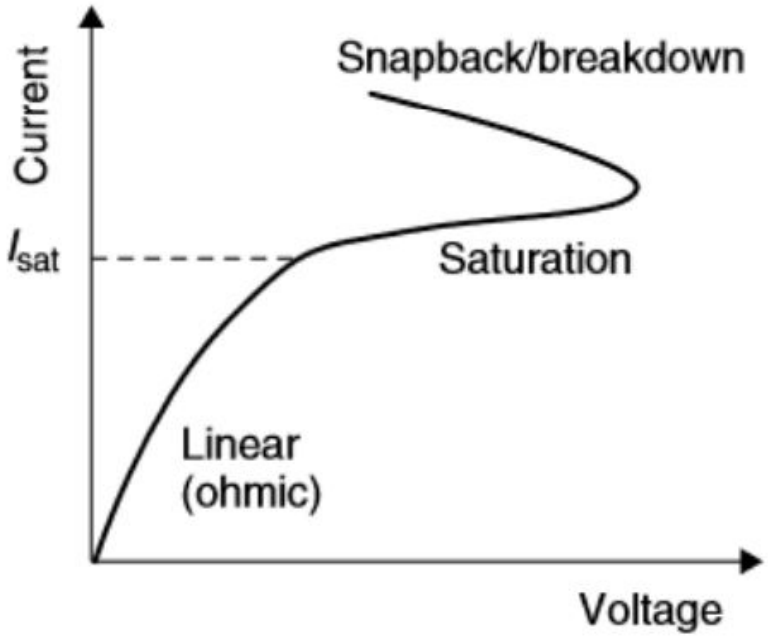


Figure 3.8: Ohmic resistor behavior [4]

If the voltage continues to further increase, it reaches the onset of the impact ionization generation. Due to the impact ionization generation, holes are created, giving rise to an opposite charge flow. As soon as the hole current increases enough to contribute to the total current, it will cause a voltage decrease creating a negative resistance characteristic, called snap-back. The explained trends are shown in Fig

3.8 up to the snap-back regime. The latter regime is typical of ESD conditions and it is usually is the adopted regime in ESD protection devices to avoid failures at large voltages. The upper limit of resistance behavior is the snap-back condition at which thermal runaway takes place.

3.4.2 p-n junction diode

In the pn-junction diode two regions are asymmetrically doped with donors and acceptors and the onset of avalanche due to impact ionization mostly depends on the doping concentration of the lightly doped region. The breakdown condition is described as follows:

$$V_{BD} = \frac{\epsilon_s E_m^2}{2qN} \tag{3.7}$$

where E_m is the value of the maximum electric field at the junction. When the impurity concentration in a semiconductor changes abruptly from acceptor impurities N_A to donor impurities N_D, as shown in Fig 3.9, one obtains an abrupt junction. In particular, if $N_A \gg N_D$ (or vice versa), one obtains a one-sided abrupt p^+-n (or n^+-p) junction [12].

At the condition of zero net electron and hole currents, the Fermi level E_f must be constant. Thus, the built-in potential reads as follows:

$$\psi_{bi} = \psi_p + |\psi_n| \approx \frac{kT}{q} \ln \frac{N_D N_A}{n_i^2} \tag{3.8}$$

where n_i^2 is the product of 3.3 and 3.4.
The depletion-layer width in the equilibrium condition can be then considered as:

$$W_D = \sqrt{\frac{2\epsilon_s}{qN}(\psi_{bi} - \frac{2kT}{q})} \tag{3.9}$$

where N is N_D or N_A depending in whether $N_A >> N_D$ or vice versa.

When applying a voltage V to one terminal of the junction, the total potential variation is given by $\psi_{bi} - V$ for a positive voltage on the p-region with respect to the n-region (forward bias) and $V - \psi_{bi}$ in the opposite condition. Substituting ψ_{bi} with the above new condition in 3.9, the depletion-layer width as a function of applied voltage is obtained.

The ideal current-voltage characteristics are based on the following assumptions [12]:

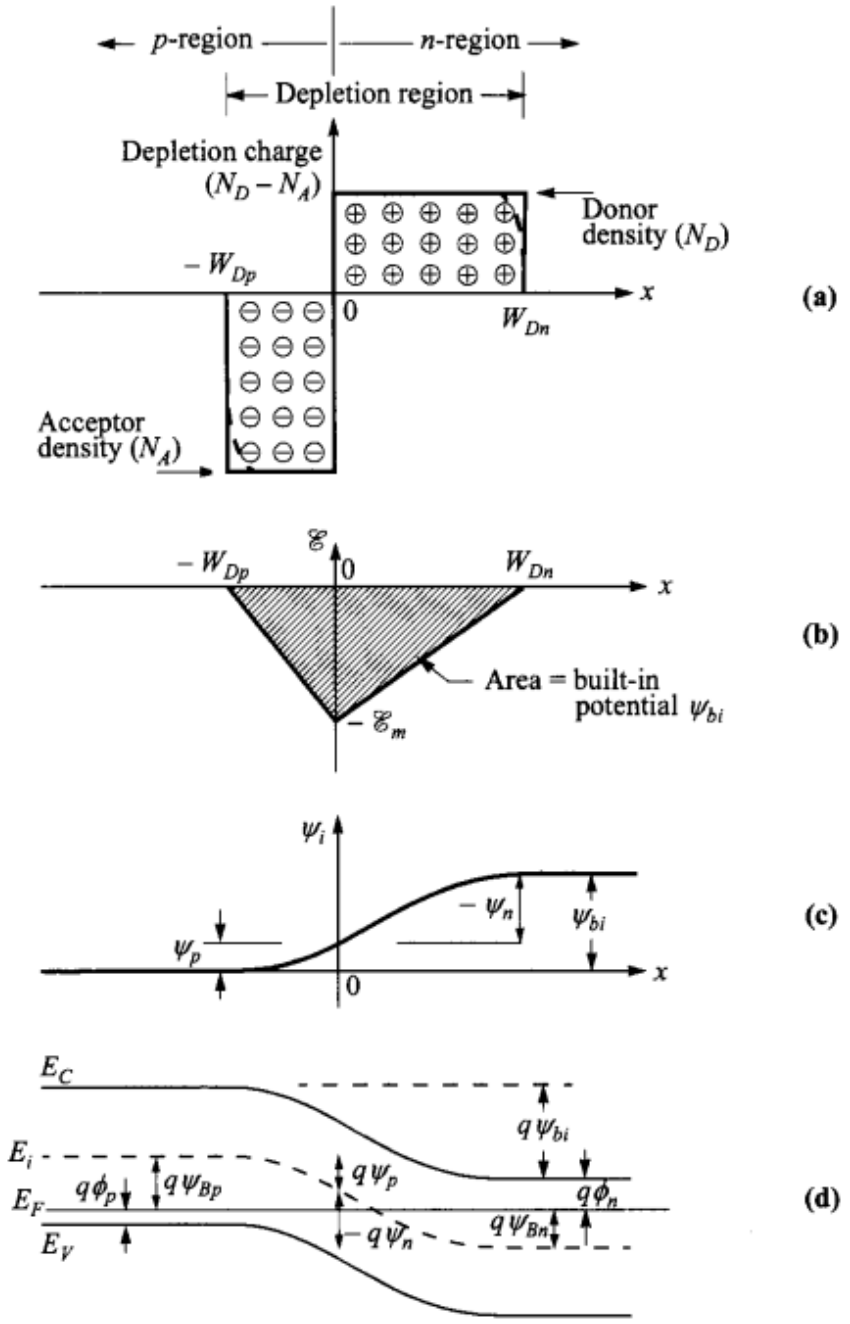


Figure 3.9: Abrupt p-n junction in thermal equilibrium.
(a) Space-charge distribution. Dashed lines indicate corrections to depletion approximation.
(b) Electric-field distribution.
(c) Potential distribution where ψ_{bi} the built-in potential.
(c) Energy-band diagram. [12]

1. the built-in potential is applied at the abrupt boundaries of the dipole; outside the semiconductors is assumed to be neutral.

2. the minority carriers injected are smaller than the carrier intrinsic density.

3. inside the depletion layer, the electron and hole currents are constant.

When applying a voltage to the device, the minority carrier densities on both sides of the junction change; thus the quasi-Fermi levels to be defined. Considering the Boltzmann relation in (3.2), we can redefine them as follows:

$$n = n_i e^{\frac{(E_F n - E_i)}{kT}} \tag{3.10}$$

and

$$p = n_i e^{\frac{(E_i - E_F p)}{kT}} \tag{3.11}$$

where
$E_F n = E_i + kT \ln \frac{n}{n_i}$ and $E_F p = E_i - kT \ln \frac{p}{n_i}$ are the quasi-Fermi levels for electrons and holes, respectively. When $(E_F n - E_F p) > 0$, the pn diode is in forward bias.

Thus, the current equation of drift and diffusion becomes:

$$J_n = \mu_n n \nabla E_{Fn} \tag{3.12}$$

$$J_p = \mu_p p \nabla E_F \tag{3.13}$$

The idealized potential distributions and the carriers concentration under forward and reverse-bias conditions can be checked in Fig 3.10 and 3.11.

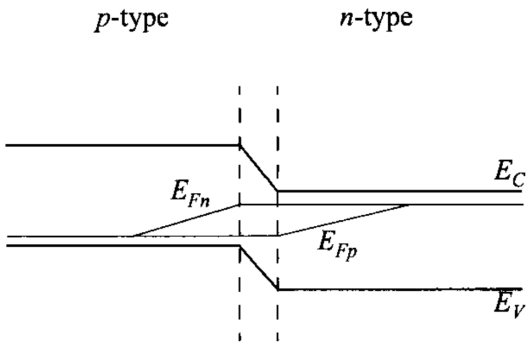


Figure 3.10: Energy-band diagram, with quasi-Fermi levels for electrons and holes in forward bias [12]

In both cases (either forward or reversed bias), the on-resistance extracted in a low current injection, will be dominated by the doping concentration of highly doped regions; this resistance will lower for high current densities due to conductivity modulation since minority-carrier density and majority one are comparable. In forward

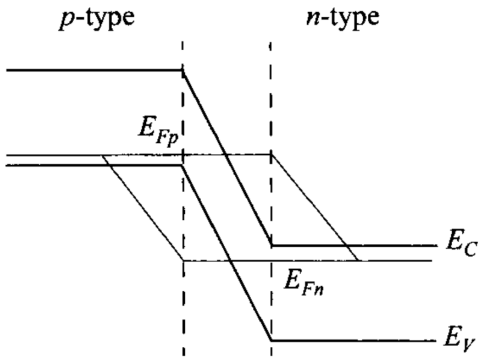


Figure 3.11: Energy-band diagram, with quasi-Fermi levels for electrons and holes in reverse bias [12]

bias conditions a positive voltage drop is applied to the anode and a negative one to the cathode. In the forward direction, the rate of carriers is constant thus the current will increase exponentially. In reverse bias condition, the current saturates at a certain current density value. In fact, in reverse bias condition, the generated current is solely due to thermally generated carriers in the depletion region; as the reverse voltage increases, carriers can gain enough energy to create electron-hole pairs which will collide with other free carriers and create additional charge (impact-ionization generation leading to avalanche multiplication). When a sufficiently high field is applied to the junction, the latter can *break-down* and conducts a very large current. This event can occur only in reverse bias condition since it needs high voltages resulting in high electric fields. The junction breakdown can lead to several intrinsic mechanisms, one of which is the thermal instability. In this case, because of the heat dissipation, the junction temperature increases [12].

3.4.3 Bipolar Transistors (BJTs)

In section 3.4.2 the theory of minority-carrier injection of a two terminal device was explained, forming the basis of the junction transistor. The bipolar transistor is a three-terminal (Collector, Base and Emitter) device usually adopted in analog electronics as a current amplifier [4]. A bipolar transistor in normal operation has a forward-biased junction (Base-Emitter BE) that enables minority carriers to be injected into the vicinity of a reverse-biased junction (Base-Collector BC). As currents in reverse-biased junctions are carried by minority carriers, the increase in minority carrier concentration at the junction edge will result in an increase in the current across the junction. The current flow across the reverse-biased junction can be modulated by controlling the injection of minority carriers from the forward-biased junction, and this forms the basis of the bipolar transistor[4]. In Fig 3.12 the series of above mentioned three junctions realizing the active area of the device is shown,

forming the so-called npn BJT.

Under normal operating conditions, the emitter junction is forward biased and the

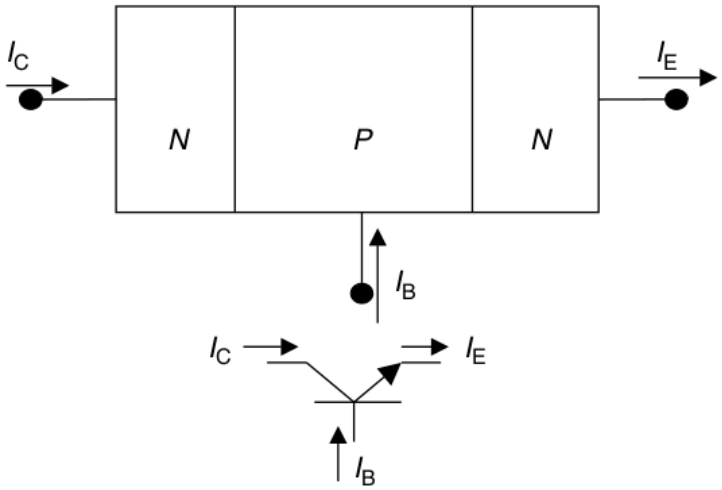


Figure 3.12: Symbols and nomenclatures of an n-p-n transistor. [4]

collector junction is reversed biased. Under these conditions, electrons are injected into the base, diffuse through the base and eventually are collected by the collector region. Electron injection takes place from the emitter n region when a positive voltage V_{be} is applied across the base-emitter junction. The base, being a *p-type* region in the npn case, represents a reverse biased pn junction. On the other hand, the base generates an hole diffusion current. However, if the injection ratio of electrons to holes is large, such as the case of the n^+ - p emitter-base junction by virtue of the difference in doping level, a current gain is realized and the transistor is turned on. All the operating regions under which the bipolar transistor can work are summarized in Fig 3.13, depending on the applied voltage.

In case of an NPN Bipolar transistor the electrons diffusion at the collector side is as following :

$$J_n = J_s[e^{\frac{qV_{BC}}{kT}} - e^{\frac{qV_{BE}}{kT}}] \tag{3.14}$$

where J_s depends on the intrinsic carrier concentration and number of impurities of the junction. The other current densities contributions at base and emitter regions can be derived accordingly to 3.14. In order to turn on the BJT, V_{BE} must be positive such that the first exponential becomes negligible and the collector current becomes $I_c = I_S e^{\frac{qV_{BE}}{kT}}$.

For the purpose of ESD conditions, it is important to highlight the current gain definition of bipolar transistor:

$$\beta = \frac{\delta I_c}{\delta I_b} \approx \frac{I_c}{I_b} \tag{3.15}$$

A bipolar transistor under ESD conditions can drain high currents but it requires

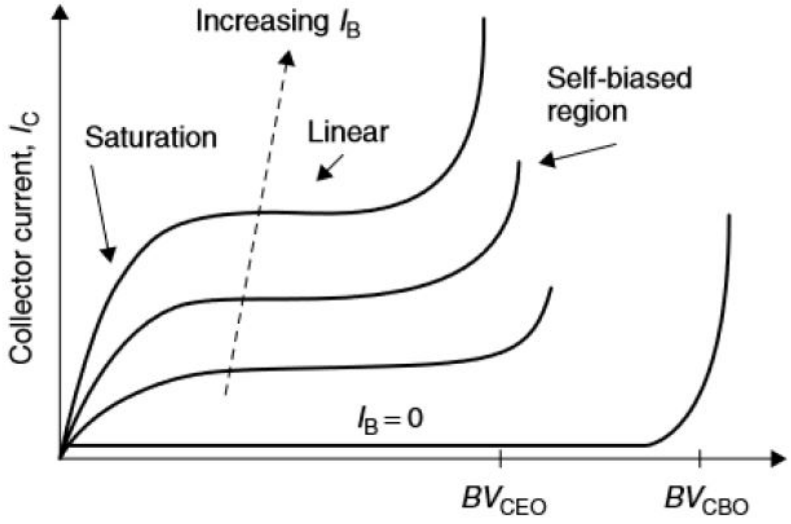


Figure 3.13: Symbols and nomenclatures of an n-p-n transistor. [4]

a high collector-emitter voltage V_{ce} since it needs to start the bipolar turn on at controlled voltage clamps. The BJT is triggered at the collector-base junction breakdown voltage; thanks to some external resistance, usually created by the fact that the base contact is not so close to the active region, the breakdown will start to conduct current across the base. This mechanism will turn on the forward junction (base-

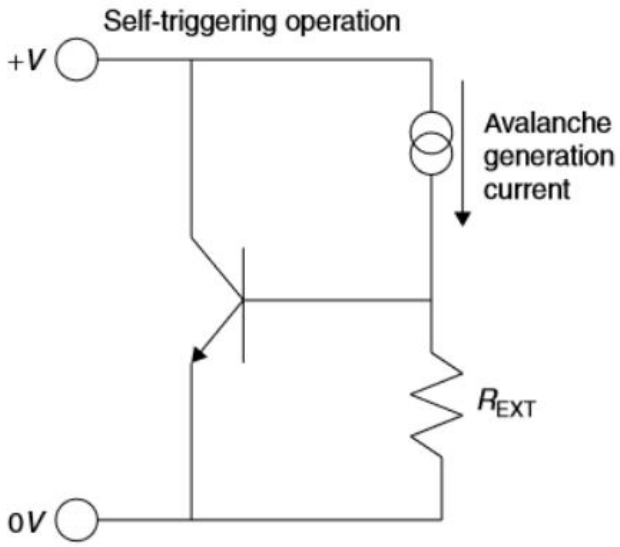


Figure 3.14: Self-triggering configuration of a BJT in CBR operating mode [4]

emitter junction), driving away the exceeding current. Different configurations can be considered (CES collector-emitter shortcut, CER resistance between collector and emitter, ECS emitter-collector shortcut, CBR collector-base resistance); an example

is shown Fig 3.14, where the configuration of the bipolar transistor with the CBR, i.e., collector-base resistance, is reported. Typically, the current at which such devices fail is around 3 to 30 $\frac{mA}{\mu m}$. The overall characteristic is shown in Figure 3.15 for the OFF-state case with I_b=0, the behavior previously described of a self-biased npn transistor under high current conditions can be recognized. Eventually, even in

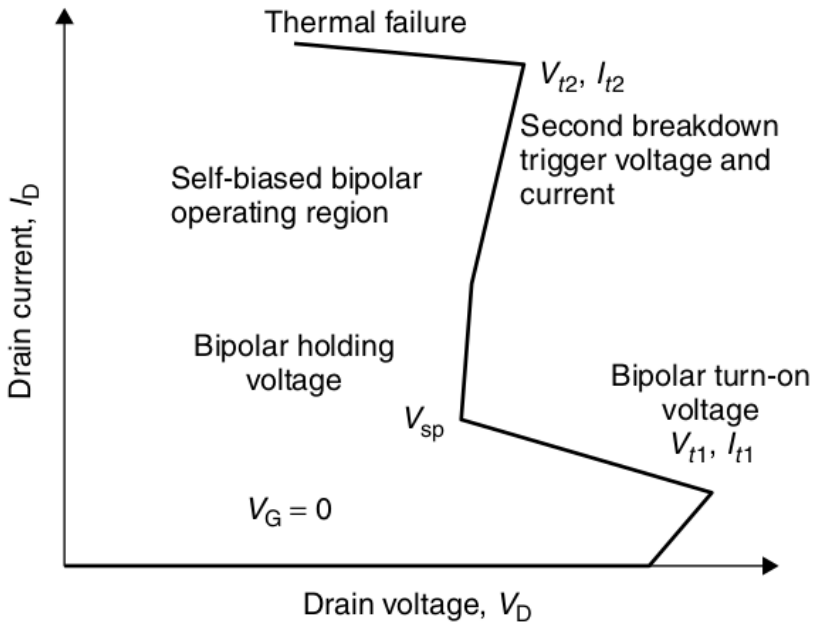


Figure 3.15: I-V output of a parasitic BJT behavior turned on by ESD conditions. [4]

normal operating conditions, with increased V_b, the BJT will be in its ON-state condition, and will reach avalanche generation of the same junction at a lower voltage (see Figure 3.13), causing the possible anticipated failure of the device depending on the initial current pulse.

3.4.4 N-channel Metal Oxide Semiconductors (NMOS) Transistor

As it is well known, the usual structure of an NMOS transistor consists of two n-type wells in a p-type substrate. During normal operation, it works in linear region or saturation region, upon application of positive drain voltage V_d . When the other contacts, gate, source and substrate, are connected to ground, no current flows until avalanche breakdown of the drain-channel depletion region takes place. By applying a positive voltage to the gate contact, $V_g > V_{th}$, with V_{th} the threshold voltage, the electric field would create a depleted region in the p-well under the gate, which will eventually become inverted, i.e., with a relevant electron concentration, creating a

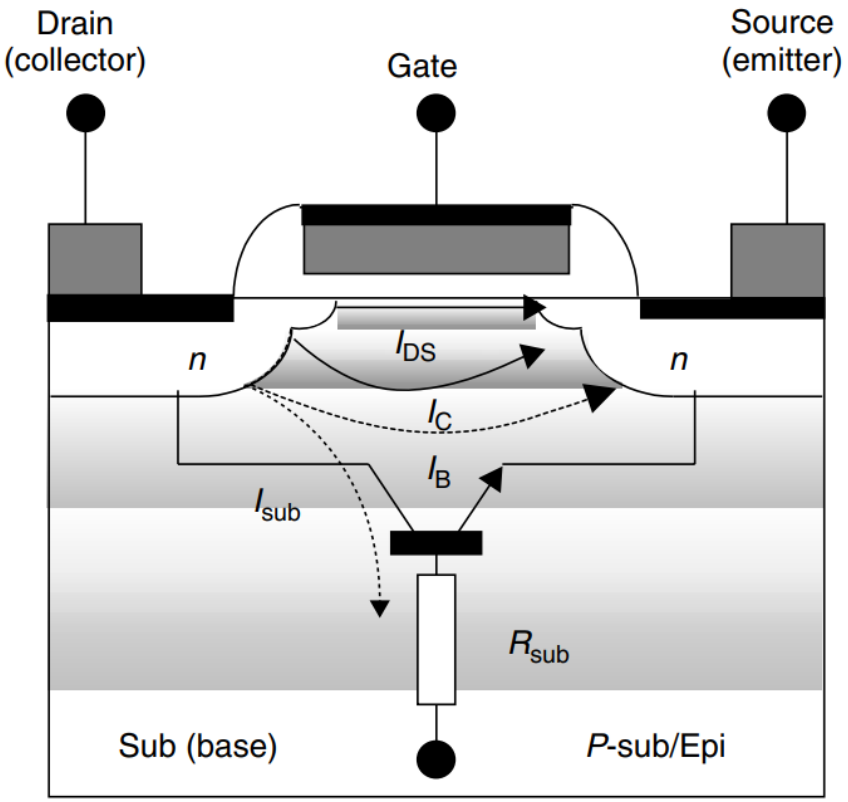


Figure 3.16: Cross section of an nMOS transistor showing the parasitic lateral npn bipolar transistor and associated currents [4]

channel between the drain and the source regions. The V_g giving rise to the inversion condition is the threshold voltage V_{th}.

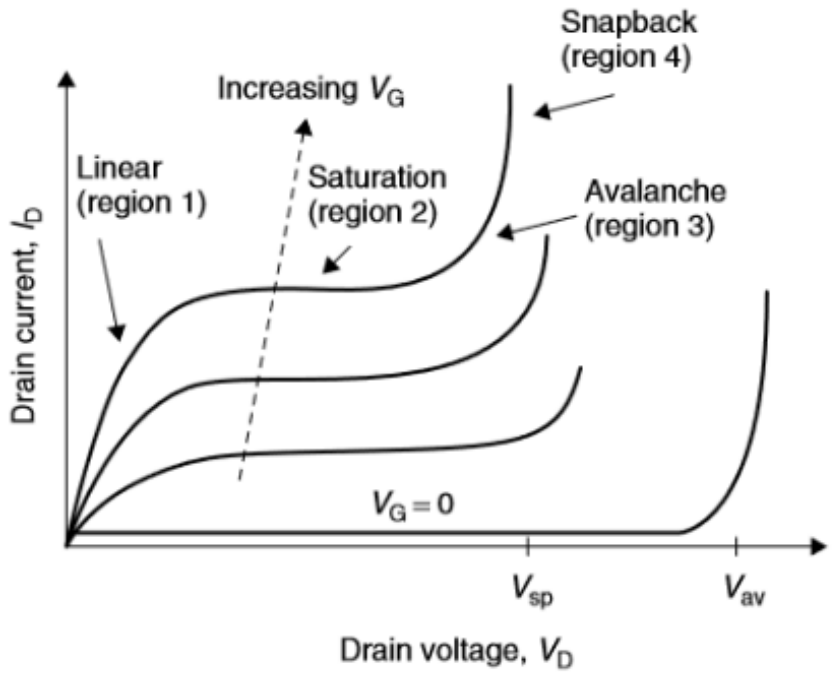


Figure 3.17: NMOS normal operating conditions with highlight of all its operating regions. [4]

The flow of the drain current with respect to the increase of drain voltage is shown in Fig 3.17. The first two regions are common operating conditions for NMOSs, where the current is turned on by V_g. For ESD applications, the current that must be sustained is much higher, so the device goes into avalanche regime and subsequently shows a snap-back behavior. In other words, it will not go through linear and saturation region, but it will operate under snap-back and avalanche. The latter behavior is given by the fact that inside a NMOS under ESD operating mode, the intrinsic junctions create a parasitic bipolar transistor which drives the MOS to behave as a bipolar under ESD conditions, while the gate voltage will not be able to control the device conductance.

In particular, the drain becomes the collector, the substrate the base and the source the emitter of the parasitic BJT. The latter configuration of the parasitic bipolar embedded in the NMOS is shown in Figure 3.16. For high current operation, the device has gate, source and substrate connected to 0 V since the device under this condition is not powered (OFF state). The drain junction is initially biased in high-impedance condition, the only current that can be considered is the current of the reverse-biased

collector-base junction. Increasing V_d, the avalanche- breakdown onset of the junction will be reached, inducing a current into the base: electron-hole pairs are generated, and the electric field drains electrons towards the drain region and holes towards the substrate. The resistive path in the substrate gives rise to R_{sub} and to a consequent potential drop across it which keeps increasing as the current I_{sub} continues to increase. By further increasing the current, the potential across the source-substrate junction increases and forward biases the base-emitter junction, pushing electrons into the substrate. The parasitic BJT in this condition is turned on: the potential at the source-substrate junction increases and forward biases this junction, emitting electrons into the substrate.

As it can observed in Fig 3.15, when the parasitic bipolar is turned on, the availability of additional carriers from avalanche multiplication decreases the drain voltage creating a negative resistance region. By further increasing the current, a conductivity modulation of the substrate will take place which reduces on resistance. If we now consider the case of a gate voltage $V_g > V_{th}$, the channel between the drain and the source will be created injecting electrons and increasing the number of multiplication events. As I_{sub} increases, the potential at the source-substrate junction increases and forward biases this junction. The drain current has now an additional contribution from the electron current density coming from the source. The parasitic bipolar transistor can be considered in on state. At the onset of thermal runaway, the base region of the bipolar, which is the key region for the full turn on of the bipolar transistor, is highly conductivity modulated, and that the point at which the carrier concentration in the drain junction is equal to the background doping concentration (the intrinsic condition) is not where the voltage collapse begins.

3.4.5 Silicon Controlled rectifiers (SCR)

The schematic view of the Silicon Controlled Rectifier (SCR) device is reported in Figure 3.18 along with the indication of the main bipolar transistors integrated in the structure. Due to its configuration, it can be considered as two bipolar transistors where the anode is the emitter of one transistor and the cathode the emitter of the other one. The schematic of the circuital behavior is shown in Figure 5.3 below.
SCRs, due to their capability to switch from a very high impedance state to a very low one, are extensively used in power device application [22]. Thus, they are an obliged solution to match automotive application requirements and area occupation. In other words, this clamp topology is efficient since there is no latch-up risk, i.e., no possibility to trigger the IC into a parasitic path but only to enable the parasitic bipolar transistor to be triggered on and protect the IC. An SCR device with low trigger

voltage, low leakage current, low parasitic capacitance, and which requires no additional process steps for ESD protection is very useful in many interesting cases, as it will be discussed more in details in the next chapters. During an ESD strike, the additional integrated SCR can produce a regenerative feedback mechanism to shunt the ESD, with low electric fields, presenting a high degree of ESD robustness.However, its relatively low holding voltage may result in serious latch up failure [23]. Thus, the latter aspect needs further study.

The device operating conditions when turned on as a ESD clamp are as follows. When a forward current pulse is applied to the anode, a forward bias forms across the SCR: the voltage difference between anode and cathode increases with current, reaching the triggering voltage (V_t). From [24], the trigger voltage for an SCR device can be expressed as:

$$V_t = BV \approx BV + V_{rep} + V_{ren} \tag{3.16}$$

where

BV is the breakdown voltage

V_{rep} and V_{rep} are the voltage drop of parasitic resistance at the emitter after avalanche breakdown.

During this process, the current flowing through the PNP is injected into the p-well, forward biasing the emitter-base junction of the NPN. When the NPN is on, it injects electrons in the nwell providing the threshold bias for the PNP. Thus, the voltage at the anode begins to decrease resulting in a negative resistance region. The voltage drop at such condition is V_h. Thus V_h can be derived from [24] as follows:

$$V_h = V_{EB,pnp} + V_{CE,npn_{sat}} or V_{BE,npn} + V_{EC,pnp_{sat}} \tag{3.17}$$

The SCR-LDMOS device is a modification of the well known lateral LDMOS-FET [25] and shows a high degree of ESD robustness. The device behaves like a normal LDMOS in its on state with a high blocking voltage (BVdss) in the off state. Under clamped inductive switching, compared to a standard LDMOS device and in addition goes into the SCR mode under extreme high voltage/high current condition encountered during an ESD strike. These features make this device an ideal self protecting output driver structure for smart power ICs. For the SCR-LDMOS device to be used as an output driver, the SCR trigger current should be more than the maximum current seen under inductive switching. Once the ESD strike is removed, the device remains latched and the SCR can be unlatched by switching on the gate of the device [25]. Further details of this specific configuration of SCR will be presented in next chapters.

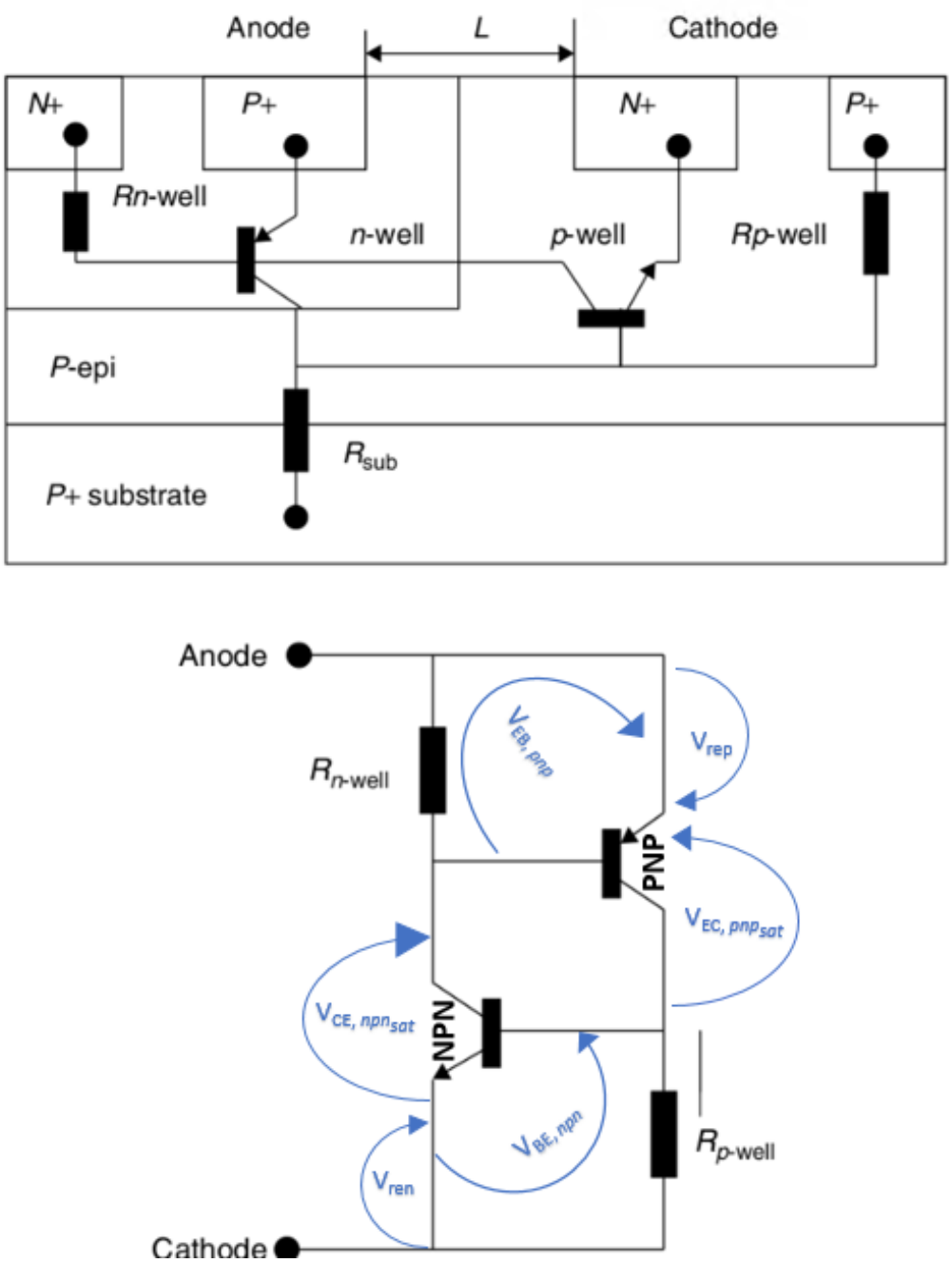


Figure 3.18: Cross section of a general SCR structure [4]

3.5 The 0.15μm BCD technology

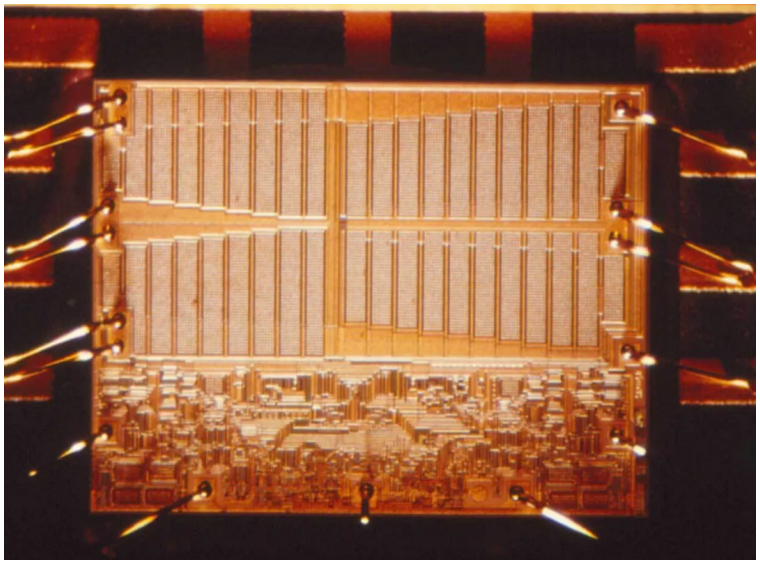


Figure 3.19: BCD technology image: STMicroelectronics' super-integrated silicon-gate process used to combine bipolar, CMOS, and DMOS technologies [26].

After ICs were introduced in the 1950s, technology variations have emerged rapidly. Some examples are bipolar transistors in 1950, CMOS in the 1960s, double-diffused metal oxide semiconductors (DMOS) in the 1970s.

In the 1980s the demand required all three kinds of chips possibly acquiring high voltage and fast switching. In 1985 Bipolar-CMOS-DMOS (BCD) was invented by SGS, now STMicroelectronics, by using the super-integrated silicon-gate process [26]. This combination of technologies brings many advantages: improved reliability, reduced electromagnetic interference and smaller chip area. The technology enables to combine power, analog and digital signal processing and makes it appealing for automotive, computer and industrial manufacturers. After STMicroelectronics, several other semiconductor industries have adopted this technology to develop new chip solutions highly powerful and efficient.

The 0.15μm BCD is a Texas Instruments high-voltage technology used for a large set of different integrated power application as, e.g, the switching power supplies. It contains 1.5 V to 5 V transistors which are used for power switching purposes. Since it is a high voltage technology, it is necessary that leakage currents are reduced to as minimum as possible values. The main power device of the BCD technology is the so-called Drain Extended MOS transistor (DENMOS), shown in Fig 3.20, with a drain region realized by two wells in order to extend it with a lightly-doped drift region. Due to low cost process and high drain breakdown voltage of DENMOS devices, they are used in a number of mixed-signal circuits like, e.g, high voltage I/O

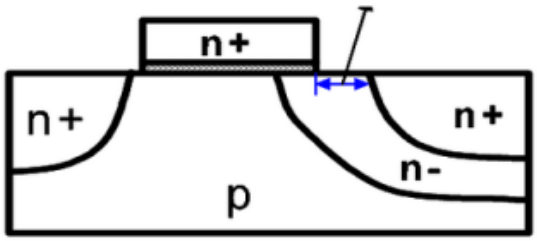


Figure 3.20: Standard DeNMOS structure considered as the main power device used in the BiCMOS technology development that is taken as test [27]

drivers and RF power amplifiers in low-power System-on-Chip applications. As a matter of fact, devices with higher doping underneath the drain diffusion region exhibit stronger bipolar triggering and higher snap-back in their breakdown characteristics, thereby sustaining higher drain current levels before device failure. This feature of the device is important for the structure. As far as the source side, it is a normal nMOS with a lower doping and, consequently, increasing the breakdown voltage. In addition, this kind of devices has a very low process cost, thus it is suitable for large scale market. Together with the DENMOS, the Lateral Double Diffused MOS (LDMOS) is another key MOSFET used in 0.15μm BCD. The LDMOS structure (see Fig 3.21 combines a short channel length with high breakdown voltage as desired for high power RF amplifiers in numerous applications. One of its main features is to optimize lateral electric field distribution using the RESURF (REduced SURface Field) effect improving significantly the breakdown voltage, as expected for an high-voltage device [27]. The RESURF principle is based on reducing the field in the current flow direction which is formed by a field plate (Gate). The field plate terminates the drift extension, killing electric field peaks caused by junction breakdown. Due to the RESURF effect, a flat and uniform electric field is observed along the drift region, preventing the vertical n-well/p-epi junction to go into avalanche [28], [29]. Moreover, the drift region is preserved as the an actual parameter to tweak the trigger voltage without the interference of the turn on of other parasitic paths. Increasing further the drift length induces an unbalance in the RESURF effect, leading the vertical field to dominate on the lateral one. It is important to notice that both devices are asymmetric in order to have different operating functions depending on the voltage applied to the electrodes: they are 5V MOS as far as the maximum gate bias is concerned; if larger voltage is applied between the oxide and the source, it will break the oxide causing failure. Thus, nominal voltages as large as 20V up to 150V can be applied at the drain contacts of such devices.

The focus of this work will be on the characterization of the 0.15μm BCD technology as far as ESD protection is concerned. More specifically, the CDM stress is

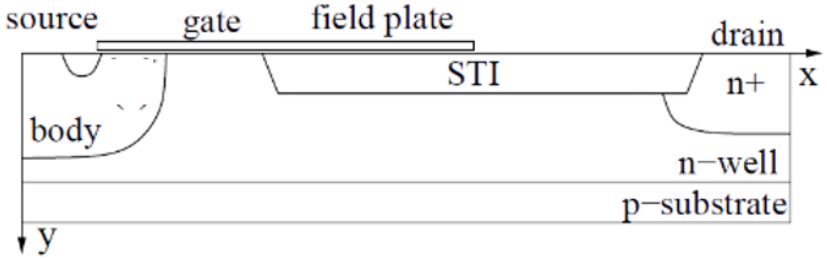


Figure 3.21: LDMOS schematic structure. All the pn junctions involved in the device can be seen from the squared regions. The structure uses an STI technology [30], even though other companies choose the LOCOS (Local Oxidation of Silicon) one [31].

addressed as it is the most critical one, together with the HBM one: the technology under test shows poor IV characteristic with respect to theoretical estimations for these kinds of stresses. According to TLP characterizations, which have been carried out by injecting current into the device through a coaxial wire, the snap-back region is most critical one: the high doping concentration of the drain region is a tipping point after the snap-back overcoming. Moreover, the junctions of the parasitic bipolar transistors located in the structure cannot be considered completely empty from mobile carriers at any time. As a matter of fact, the electric field peak cannot be considered static in the junction, but it moves creating hot spots subsequently melting the material (Silicon). This event is the so-called Kirk effect where high carrier density can cause dramatic increase in the transit time of the carriers [32]. In order to prevent this kind of events from happening inside the IC, some additional structures have been used and studied to solve the problem of this technology; their studies are shown in the following sections.

3.5.1 From cutline to TCAD tool

Sentaurus Workbench [5] helps to investigate this kind of issues through thermodynamic simulations. The latter are mostly important for self-heating structures like ESD protections. Moreover, carrying out transient simulations which show the full behavior of the structure allows to check the distribution of relevant physical quantities that cannot be measured otherwise. Then simulation results will be used to investigate more parameters than real experiments can get at different steps, in order to improve the design of the future devices. The main difficulty in the above approach is the fact that it is necessary to define and calibrate the TCAD structure following the flow of the fabrication processes in order to realize a deck which is as

close as possible to the real one. As a matter of fact, the focus of the first part of this work is to reproduce the main features of the device by defining a 2D domain of the cross-section into the TCAD tool. Once the structure has been defined, specific simulations can be run in order to reproduce the I-V characterizations up to the thermal failure condition, following its actual behavior. The results of the TCAD analysis are reported in Chapters 4 and 5.

The main purpose of this work is to run numerical simulations of the electro-thermal behavior for the most promising ESD cell. These kinds of simulations are useful to visualize the dependency between electric field and electrostatic potential. The numerical results are compared against measurements and the role of some specific parameters has been investigated, like, e.g., the impact-ionization generation and the maximum temperature reached in the device. For this reason, the simulation input has to be tuned and designed in order to investigate either the DC behavior and the TLP and VF-TLP ones. In order to run numerical simulations consistent with the characterization of the ESD cell, it is necessary to set the simulator command file by properly accounting for specific physical models available in the simulator and the set of equations. The adopted tool is Sentaurus Device (sdevice) [5]. Sentaurus Device is a device and circuit simulator for one-dimensional, two-dimensional, and three-dimensional semiconductor devices in multidimensional and mixed-mode domains. It incorporates advanced physical models and robust numeric methods for the simulation of most types of semiconductor devices ranging from very deep-subμ silicon MOSFETs to large bipolar power structures. This tool can interact with other Sentaurus tools, for example the Structure Editor, in order to use the structure obtained to perform simulations on it. Each part of the command file needs to be customized depending on the structure that needs to be simulated and on the kind of characterization that might be reproduced.

The **File** part defines the input-output files that have to be considered as source or output of the simulation.
In the **Electrode** section the contacts that have been defined in the .tdr file are listed: they are the boundary conditions as far as the electrical domain is concerned, and the corresponding voltage or current supplies has to be defined.
The **Physics** part is very important to tune and customize the stimulation with the most appropriate physical models for the device that is under investigation. More specifically it is necessary to define:

- the model of the carrier mobility, which can be tuned with different kinds of mobility models,

- a band-gap model, which describes the effective intrinsic density as a function of temperature, relevant to define the width of the space charge region,
- The SRH recombination model which defines the exchange of carriers between the conduction band and the valence band (see 3.4).
- The impact-ionization model which describes the key avalanche effect

The **Plot** section defines which kind of physical quantities are stored in the output .plt file in order to study the I-V characteristics.
The **System** section is the section that identifies the Mixed-mode part which is a tool extent to add to the structure additional standard circuitry according to the schematic of the component. The Mixed-Mode is used for multi-device simulations (Figure 3.22), specifying a circuit net list to connect the devices. In order to identify the devices to be connected, a new definition of the Electrodes as nodes must be specified for each device connectivity list. Moreover, it is possible to add passive components depending on the type of simulation and schematic construction of the system, for example resistances or capacitances or even voltage source stimuli as in the SPICE tool [33].

The **Math** section describes which kind of math solvers will be used to solve the equations expressed in the Solve section; usually, for large problem size, ParDiso solver is used since it solves large sparse symmetric or structurally symmetric systems, parallelizing the solution. In this part of the command file it is possible also to tune some tolerance on the error that might be produced while the equations are solved.

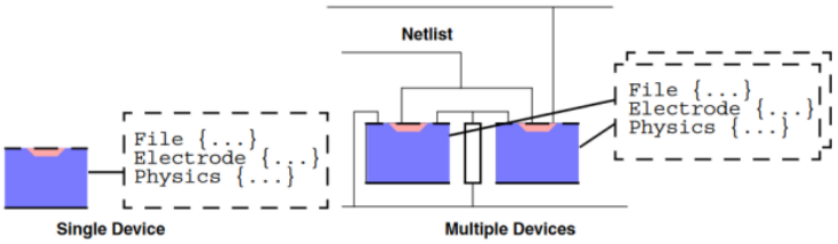


Figure 3.22: Mixed-mode simulation flow schematic. [7]

The last part is the **Solve** part where the equations used to perform the simulations are expressed and where the stimulus is given. The starting point is usually given by the Coupled solver which activates a Newton like solver (linear system solved at each step of simulation) over a set of equations usually defined by the Poisson equation and the continuity equations, resulting in the electrostatic potential and electron and hole densities.

In order to run a DC simulation, the input contact is ramped to a voltage or cur-

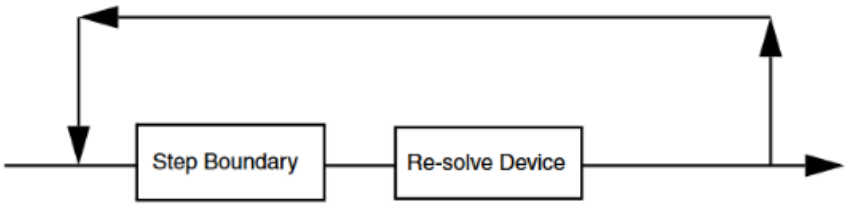


Figure 3.23: Quasi-stationary schematic iteratevely used in the TCAD

rent goal and, at each step of the simulation, the quasi-stationary solution is given. In the sdevice tool, a Quasi-Stationary command can be used to this purpose. The Quasi-stationary command ramps a device from one solution to another through the modification of its boundary conditions (such as contact voltages) or parameter values. In case of a voltage ramp (see Fig 3.23, the device ramps the selected electrode by increasing the voltage of an offset given by t * V_1 - V_0 until it reaches the voltage goal given by V_1. The ramping simulation used to emulate ESD stress is the Transient one since it has more probability of convergence for such complex devices under ESD pulse. The Transient syntax replaces a quasi stationary simulation with a slower transient ramping. The command that activates the transient ramp is a quasi-stationary like command with two optional parameter added: Initial Time and Final Time, useful to control the ramp rate.
As far as boundary conditions are concerned, electrical boundary conditions are defined in the **Electrode** section. The name given to the electrodes must match the same names specified in the structure editor file. With the **Thermode** section, the thermal boundary conditions at the contacts of the device are specified. The default configuration is considered in this work, only by specifying the initial temperature of each contact as room temperature (300K).

3.6 Concluding remarks

In this chapter an overview of several aspects of the ESD world is presented, starting from the industrial standardized stress tests to the physical behavior and operating conditions of some general devices under ESD operating conditions. A special focus is given to bipolar transistors and Silicon Controlled Rectifiers, as both devices will be deeply studied in next chapters. Moreover, the BCD technology is introduced as devices under test belong to this category.

Chapter 4

Study of stacked BJT transistors used as ESD protection cells for Ultrafast Events (CDM)

4.1 Chapter overview

The operating conditions of a new ESD clamp based on BJT transistors is presented in this chapter. The analysis is constituted by TCAD simulations in ESD operating conditions compared with available experiments. The device performances are further investigated through geometrical modifications. A study of the Power-to-failure performances is also presented as a conclusion.

4.2 Stacked NPN based clamp

4.2.1 Component characterization: the NPN 30V cell

The structure presented in this chapter is a bipolar transistor used as ESD clamp. The final purpose of the clamp is to enable the device to be protected from receiving voltage that is higher than the baseline by clamping the trigger voltage before the failure of the IC. Ideally the snap-back voltage should build a voltage discharge that can be controlled by a resistance in series. For this purpose, a series of two bipolar transistors internally stacked, as shown in Figure 4.5, is employed as the device to be investigated. In particular, the stacking of two 15V NPNs enables the entire clamp to protected from a trigger of 30V, creating a resistive path through the series of the two BJTs.

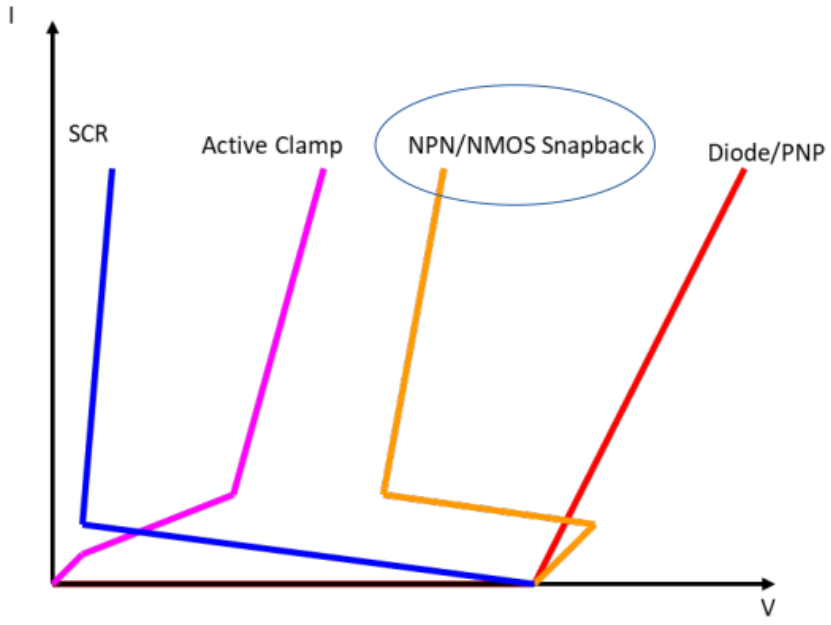


Figure 4.1: Different clamps behavior SOA under ESD conditions. The one under investigation in this chapter is here generalized as NPN/NMOS snap-back

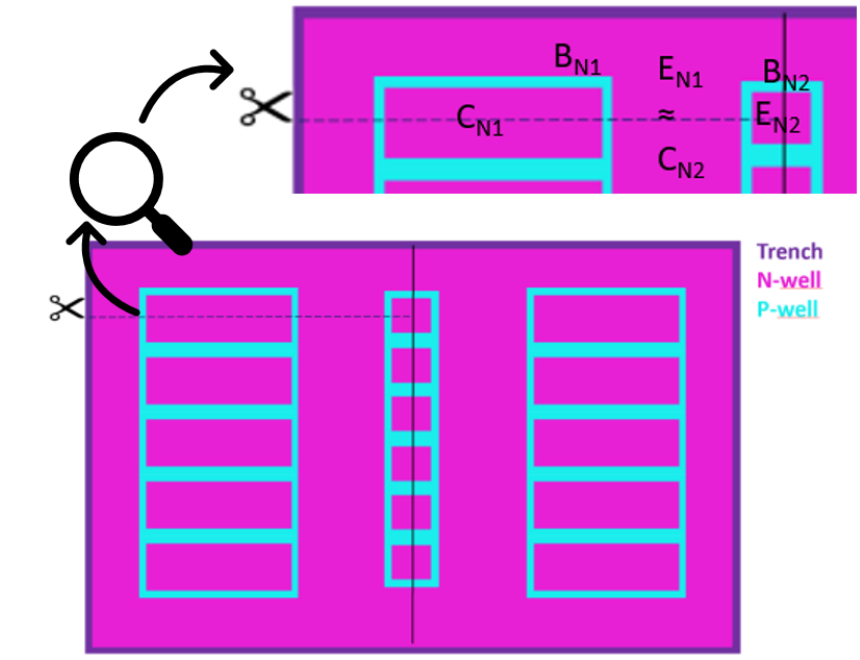


Figure 4.2: Layout of the BCD NPN 30V cell. A zoom of the single device is highlighted by the magnifier to shown the different BJT regions: collector, base and emitter

4.2.2 Component characterization: the NPN 30V cell

The entire ESD protection used for the final application constitutes of several different devices connected together to match the requirements. Thus, the NPN 30V component that is here investigated is only a part of a more complicated technology.

In Figure 4.2, the layout of the characterized component is shown. The characterization has been carried out in order to have a good starting point for all other kinds of simulations. In the layout scheme, the n-doped regions are drawn as pink zones, while the p-doped regions are drawn as green windows.

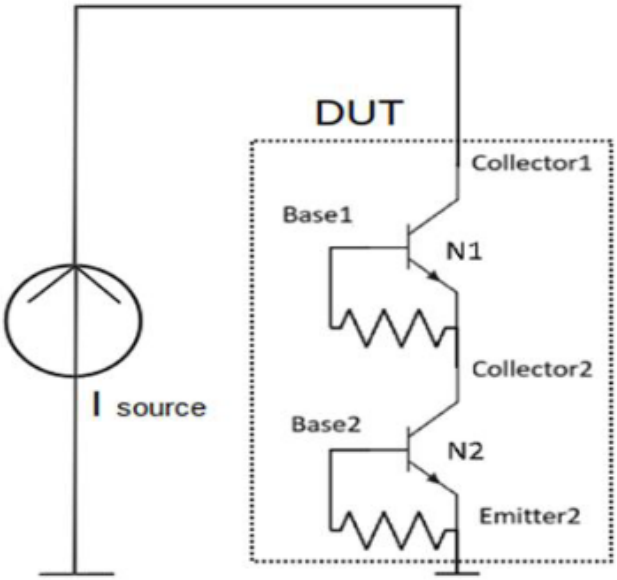


Figure 4.3: Equivalent circuit with two low-voltage coupled biased transistors internally stacked to efficiently build a high-voltage device for 30V ESD protection. BJTs are in self-triggering configuration (i.e. with biasing resistor between base and emitter).

The structure consists of two NPN BJTs in series as shown in Figure 4.3. Each one can carry over 15 Volts so the series of the two can reach the desired voltage of 30 Volts and beyond. As it can be clearly deduced, the structure is symmetric: it consists of two symmetric NPN BJTs (with B_{N1}, C_{N1}, E_{N1} indicated in Figure 4.2) in series to a central NPN BJT indicated in Figure 4.2 with B_{N2}, C_{N2}, E_{N2}. For this reason, the analysis is carried out on half structure. One of the most important characteristics of the 0.15μm BCD technology is the presence of specific features which are usually available in such kind of smart-power technologies [34]:

- The **N-type buried layer (NBL)**, mainly used to isolate different components but also to connect the substrate [35].
- **Deep Trench region**, which is used for the lateral isolation of power devices

and substrate connection. The trench is realized with Deep Reactive Ion Etching (DRIE) processes and by filling the trench with P+ poly-Si allowing a good connection to the p-substrate with a limited area. It allows to isolate different components/blocks saving significant space compared to junction isolation technologies [36].

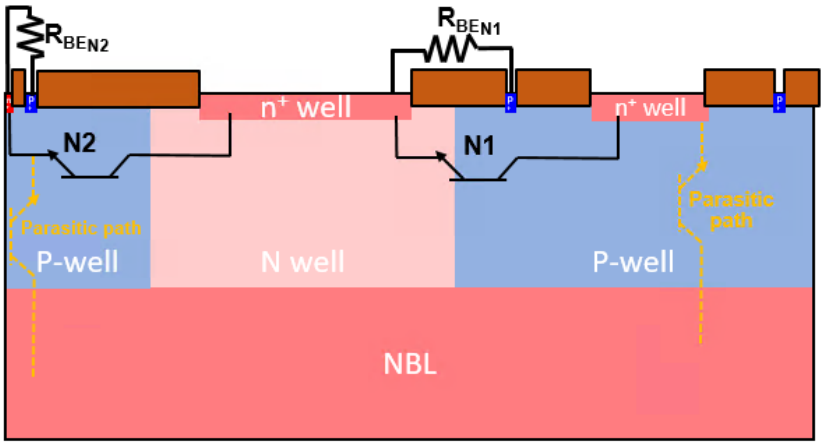


Figure 4.4: 2D cross-section of the NPN ESD cell under investigation. In yellow the parasitic paths described in the text as ccircuital indication.

The electro-thermal simulations of the NPN ESD cell have been carried out to investigate the device behavior under ESD conditions. Several parasitic paths have been identified, changing the desired flow of the current and deviate the expected I-V curve from the conventional ESD safe operating area.

The implementation of the structure is shown in Figure 4.4. Due to the complexity of the structure, there are several junctions that are not primarily involved in the series of the two BJTs paths. These parasitic paths might turn on and deviate the device from the expected performance. More specifically, the two BJTs in the structure are lateral but there might be additional vertical current paths creating parasitic vertical NPNs. Moreover, the connection of the surface of the device to NBL driven by the Deep Trench structure can carry some additional carrier paths through the NBL itself or, more dangerously, through the substrate. A possibility of this situation is shown by the yellow dashed NPNs drawn in Figure 4.4. All this kind of problems are discussed in details in the next Sections.

4.2.3 Cross section description

The simplified 2D cross-section of the ESD cell is shown in Figure 4.5. The internal architecture is realized by integrating two different sets of pn-junctions as available in the BiCMOS technology, sharing a collector-emitter region among them.
It was not possible to access the flow of the process. However, the cross section can be used to create the device main features using the TCAD tool with analytical doping profiles. In other words, from the cut line of the layout, the final deck for the TCAD simulations can be fairly obtained. The impurity concentration profiles have been inferred from process simulation results, and analytical functions have been used to easily consider layout variations of the investigated structures (as indicated in Figure 4.5). The doping profiles are also checked following the indications given by the ESD Group in Texas Instruments, considering different cut-lines extrapolated from the technology layout in Figure 4.2.

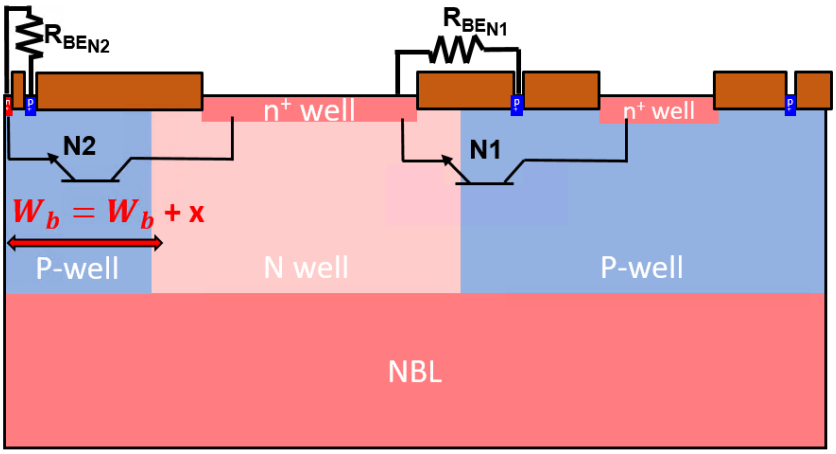


Figure 4.5: 2D cross-section of the NPN ESD cell under investigation. The series connection of the two NPNs, N1 and N2, is drawn in the 2D plot along with the indication of the width of the base well, W_b, W_{b0} is the reference width and x is the variation used to check the NPN behavior.

In Figure 4.5, the main n-doped and p-doped regions required for the realization of the BJTs under investigation are schematically shown. The n^+-well constitutes the emitter region of the N1 BJT, running along the longitudinal direction of the cross-section. Its base region is the right side p-well region. The N2 and N1 share a N-well region in the middle of the structure that is for N1 the collector region, while for N2 the emitter region. In this way the two bipolar transistor are in series connection. The N2 base region is the p-well region on the left and the collector region, connected to ground is a very small n^+ doped region on the far left side of the structure. The two transistors are internally stacked to create a series of bipolar transistors able to

sustain over 30V of trigger voltage. The stacking configuration described above is clarified by the schematic circuit in Figure 4.3.

4.2.4 Component schematic and operating conditions

The behavior of the simplified schematic circuit in Figure 4.3 is explained in this section. In particular it is considered that, under ESD condition, the stress injected into the DUT is modeled for simulation purposes by a current source. Thus, the ESD stress that the device is experiencing is a current pulse with certain rise time and length time characteristics, depending on the stress conditions that we want to consider (either HBM, CDM or MM). Under normal operating conditions, the bipolar transistors can be designed to carry high current levels despite a high collector-emitter voltage is applied, with a consequently high-power dissipation. Therefore, such devices are used as ESD protection circuits in high-voltage and high-power applications. The stacked configuration uses the internal avalanche regime of both NPNs to self-bias both devices. Due to the intrinsic differences in the adopted p-n junctions and in the geometry of the active regions, non-uniform distribution of current densities is expected in the triggering characteristics of the two NPNs. The external resistances between each base and emitter, properly tested for the purposes of the protection behavior, help improving the switching of both devices at a specific collector current, but the most significant role in the coupling of the stacked devices is given by their internal integration. Therefore, 2D simulations are needed to investigate their transient behavior [37].

4.3 Simulation of the relevant operating conditions

4.3.1 Triggering and holding regimes

Thermo-electric simulations have been carried out with default thermal boundary conditions, assuming room temperature at the metal contacts. 2D simulations are expected to give the worst-case scenario of the onset of the thermal runaway thus providing a safe margin for failure predictions. For this purpose, the thermal boundary conditions at the contact side have been properly tuned such that their contribution is negligible in high current regimes. All simulations have been carried out using the drift-diffusion model coupled with the heat-transfer equation. Shockley–Read–Hall and Auger generation–recombination models have been used along with the UniBo impact-ionization model [38]. They are turned on with default parameter values [5]. Additional details are reported in section 3.4.
To this purpose, EOS, TLP and very-fast TLP curves have been simulated using a

single pulse with 200ps rise time at different pulse lengths, applied with increasing current levels to test the ESD cell up to thermal failure. In order to simulate such stresses, the simulation depicted in Figure 4.3 is used with a current source connected to the collector of N1. Both voltage and maximum temperature at the end of each stress pulse are collected from numerical simulations. In this way is possible to determine the qualitative behavior of the device and its failure level.

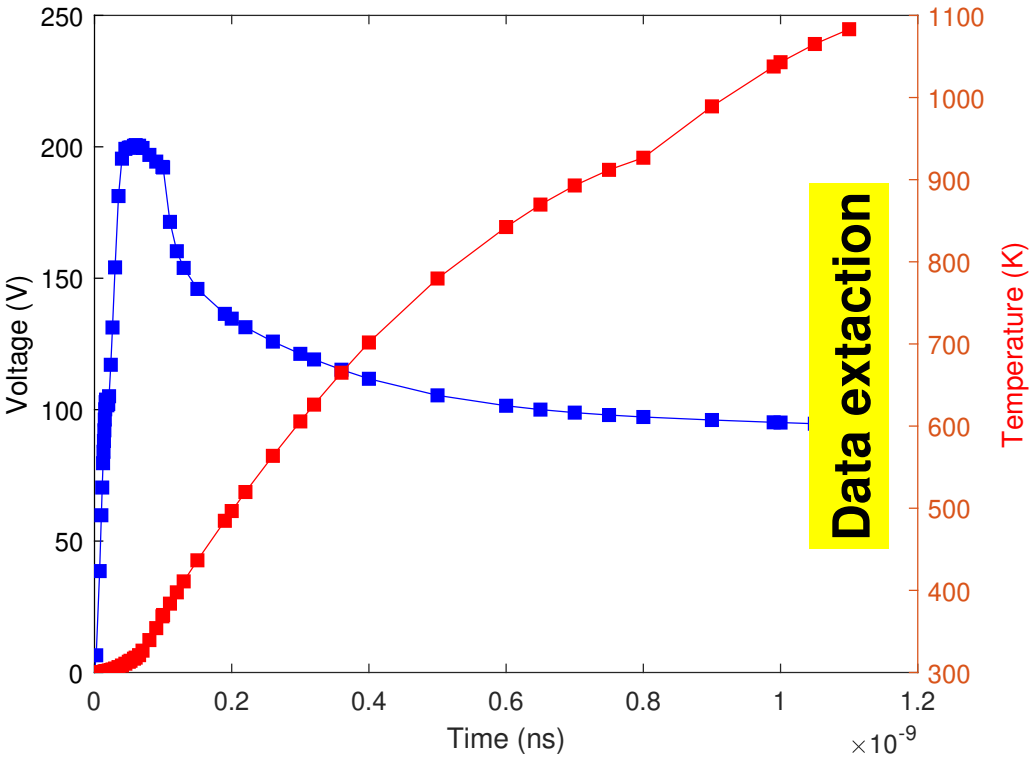


Figure 4.6: Voltage (in blue) and maximum temperature (in red, right axis) increase as a function of time during the TLP pulse. As an example, a 1ns TLP pulse is reported.

As an example, in Figure 4.6 both voltage and maximum temperature are reported for a 1ns single pulse: the voltage suddenly rises accordingly to the rising of the current signal with t_r=200 ps, then it starts decreasing at the end of the rise time due to the turning on of the BJTs, as explained in [39], and reaches a plateau at about 80% of the pulse length. The temperature increases linearly until the end of the stress where it reaches critical values; data are extracted at the end of each pulse for both parameters. The peak voltages that can be identified as overshoots are observed before the end of the TLP voltage rise, for $t < t_r$. In the investigated cases, the overshoot is not impacted by the change in the slope of the TLP pulse [19]. Thus, no calibration of the overshoot at the beginning of the stress pulse is required. The voltage overshoot in the reference device gives an electrical signature

of the NPN delay, giving rise to a larger voltage with respect to that reached at 100ns. In the structure under study, the voltage overshoot can be ascribed mostly to N2. Thus, a layout reduction of W_b is expected to increase the gain of N2 and reduce the corresponding transit time, leading to a lower voltage overshoot, a consequent lower heating and a final larger Power-to-Failure condition. These results are shown in details in the next Section with additional physical explanations.
The data extracted to trace the IV curve is taken at the end of the pulse. The simulated I-V curves are reported in Figure 4.7 and compared with corresponding TLP and vf-TLP experiments. The curves show a qualitative good agreement, allowing for the correct analysis of the corresponding transient regimes. Even though simulations do not perfectly match key data, such as trigger voltage, holding voltage, snap-back conditions and thermal runaway, they give the correct insight of the different regimes experienced by the structure. For this reason, no fine-tuning of the simulator models has been performed as the results qualitatively match the trend of the real device. The simulation results should be anyway consistent without the need of a 3D analysis, leading to reliable estimations of the hold and failure current levels. It is expected that current filamentation might be experienced in geometries with complex width layout [40], [41]. In our structure, current filamentation can be not considered and thus a 2D approach can be adopted for the full analysis.

It is important to point out that the device under test has a peculiar structure and several parasitic paths can be turned on during a transient regime, leading to premature failure for different reasons. Its peculiarity is given by the active regions of the two bipolar transistor which are not symmetric (see Fig. 4.5. N1 bipolar transistor matches a conventional integrated lateral NPN structure, with an highly doped and small collector region and a very large emitter region. On the other hand, N2 structure is closer to a parasitic bipolar device, featuring a highly doped collector region in the center of the structure. The turning ON of BJTs builds mainly a surface current density for low current pulses.
Since the analysis of vf-TLP simulations has a lack of information in literature, its analysis will be focused in this part, showing the device behavior for this kind of stress configuration. Moreover, the component shows unexpected low TLP and vf-TLP (very fast TLP) performance: after snap-back, the electric field peak moves out of the junction and the depletion region for low injection regimes is not valid anymore. ESD protection should allow the device to bear a voltage that is higher than the baseline (V_t). This behavior increases the possibility of CDM stress test failure (see theory in section 3.2) while it should not happen for this kind of high voltage devices, specifically designed to sustain high and very fast stresses.
In Figure 4.8 and 4.9, the current density distribution is reported at the triggering and

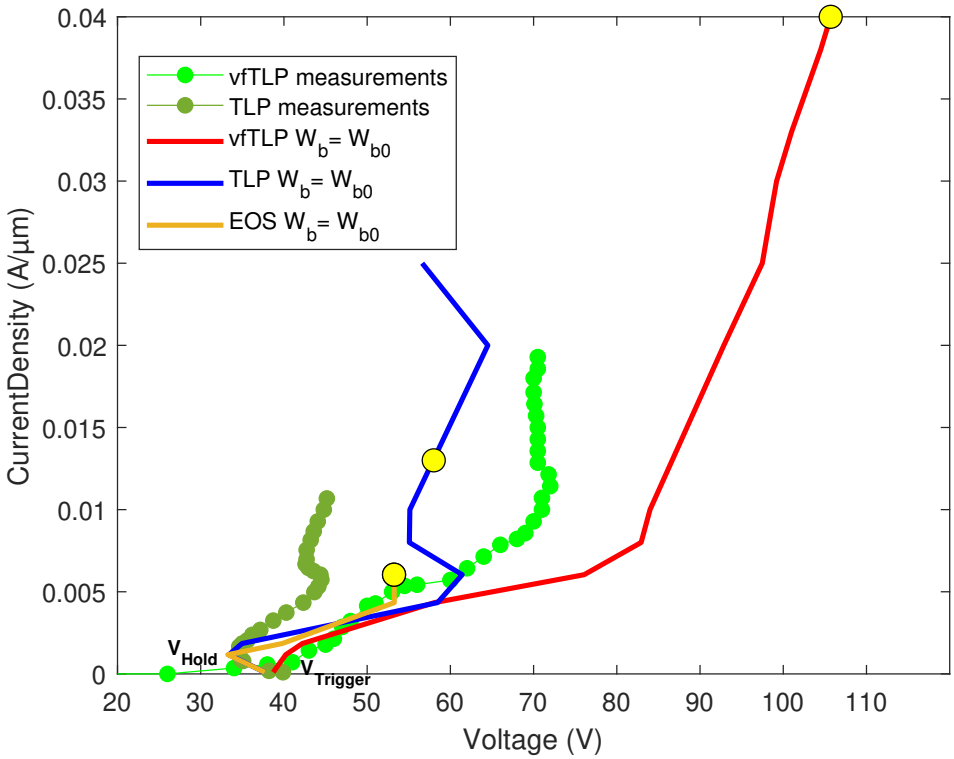


Figure 4.7: Lines with symbols: Measured I-V characteristics under TLP and Very-Fast TLP (vf-TLP) of the ESD cell. Solid lines: TCAD results for 500 ns, 100 ns, 1 ns (EOS, TLP and vf-TLP). Yellow dots: failure condition extracted from TCAD data.

holding voltage respectively under a VF-TLP conditions (1ns). In the first condition, the carriers move in a confined region close to the upper part of the devices in both NPNs. The surface current concentration eventually allows for an independent layout control of the triggering and holding voltages. More specifically, the triggering voltage is defined by the lateral current diffusion at the surface in combination with the external resistances. Vice versa, the holding voltage and the failure condition are correlated with the current density spreading towards the deeper part of the structure and the corresponding p-well and n-well doping profiles play an important role. In this condition, the impact-ionization generation and the NPN gain drive the devices into high-injection regime.

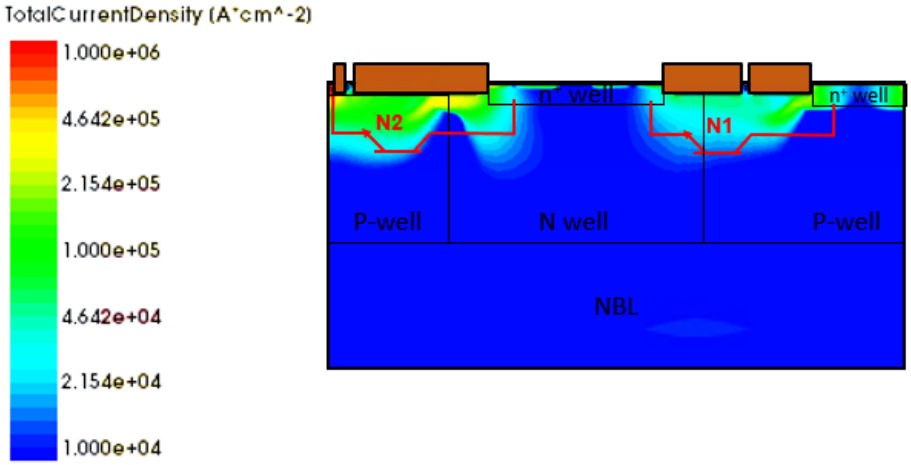


Figure 4.8: Current density at triggering voltage under vf-TLP conditions. The current is concentrated at the surface of the device, thus the volume of the pn-junctions is not entirely engaged

In 4.9 injected carriers are flooding into the entire N2 base area, deep into the structure.

4.3.2 Failure condition

When temperatures start to increase inside the device due to self-heating, the high injection condition reduces the role played by the doped regions due to the large amount of carriers into the entire volume. The definition of the thermal runaway reached through simulations has to be defined.

Experimentally the TLP characterization of an ESD device is tested at certain biasing conditions. Damage to the DUT will occur when the test pulse amplitude becomes high enough to produce enough heat to melt the structure and cause a permanent change in the device. Thus, if the current flowing trough two pins that are not involved in device protection path starts to increase, it is considered as leakage

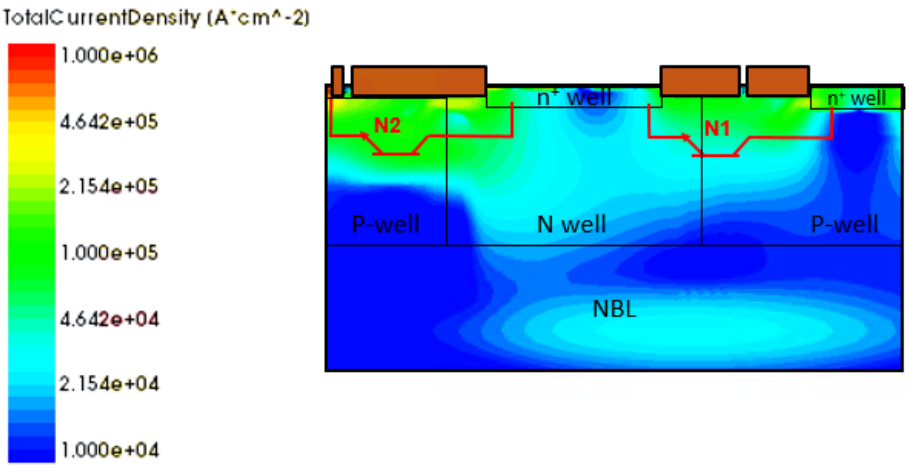


Figure 4.9: Current density at holding voltage under vf-TLP conditions. The current spreads through the entire volume, mostly for the N2 bipolar transistor

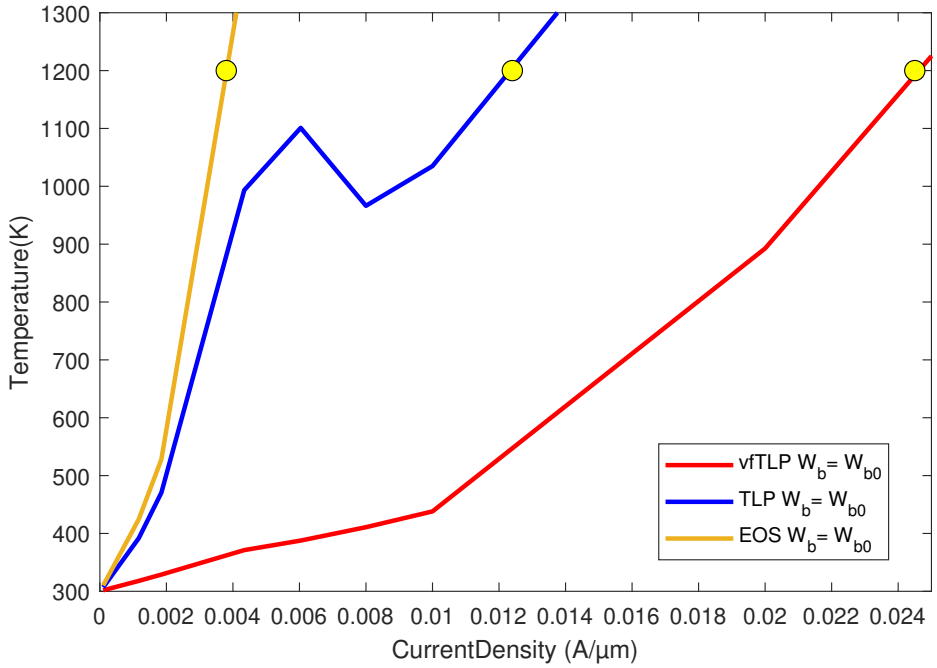


Figure 4.10: Maximum temperature (T_{max}) in the 2D domain at the EOS,TLP and vf-TLP final times as a function of the current density. The failure criterion is fixed @T_{max} = 1200K.

current derived from the permanently damaged region. Thus, during TLP measurements this event is considered as a failure criterion. In simulations it is possible to monitor the internal temperature of the device and detect accordingly when a hot spot occurs. In figure 4.10, the adopted failure criterion for simulations is shown and explained as follows: the maximum temperature in the structure is monitored at the end of the current pulse (depending on the stress under consideration, thus 500 ns, 100 ns or 1 ns for, respectively, EOS, TLP and vf-TLP stresses); when a temperature over around 1200K is reached inside the structure, the thermal runaway is considered to be reached. Yellow dots in the TCAD curves of Figure 4.7 and 4.10 correspond to the extracted thermal failure. The failure conditions indicated on each curve, clearly show a scaling of the failure current with increasing stress time. The discrepancy between the measured data and the simulation ones might be ascribed to the slight differences in the 2D simulation deck with respect to the real structure.

A complex local distribution of temperature and avalanche generation (shown in Figure 4.11) can be the cause of the poor performance of the device. In Figure 4.11 (a), it can be observed that the different geometry and overall configurations of the two NPNs lead to a faster heat up of the collector-base junction of N2 (right side) than the reversed biased junction of N1. In Figure 4.11 (b), it can be observed that a significant impact-ionization generation along the vertical junction is experienced. This behavior can be attributed to the vertical spreading of the current density in the lateral and bottom side of the pn-junction as observed in Figure 4.9.

Indeed, the full area of the N2 collector-base region is turned on and the failure of the entire device is reached. In order to compare the current density distributions that lead the failure for the vf-TLP and TLP cases, the screenshot of the current density distribution at the holding voltage under TLP condition is shown in Figure 4.12: a limited spreading of the current flow is observed in the deeper part of the N2 device for the longer stress times. Thus, the contribution of electric fields and impact-ionization generation in the N1 ad N2 regions is more balanced than for the 1ns case.

On of the most important figure of merit analyzed for the ESD protection schemes the Power-to-Failure curve. From a theoretical point of view, the power to failure of ESD clamps is expected to scale as the inverse of stress time during the adiabatic regime (i.e. vf-TLP stress) while the reported data shows a different trend (see Figure 4.13). According to the theoretical analysis, the simulated device is expected to reach a current of 1A over 100ns stress time in the Wunsch-Bell region. Therefore, the current level reached for a stress 100 times faster, thus a stress reduced to 1ns, should correspond to at least a seven times higher current. According to Dwyer

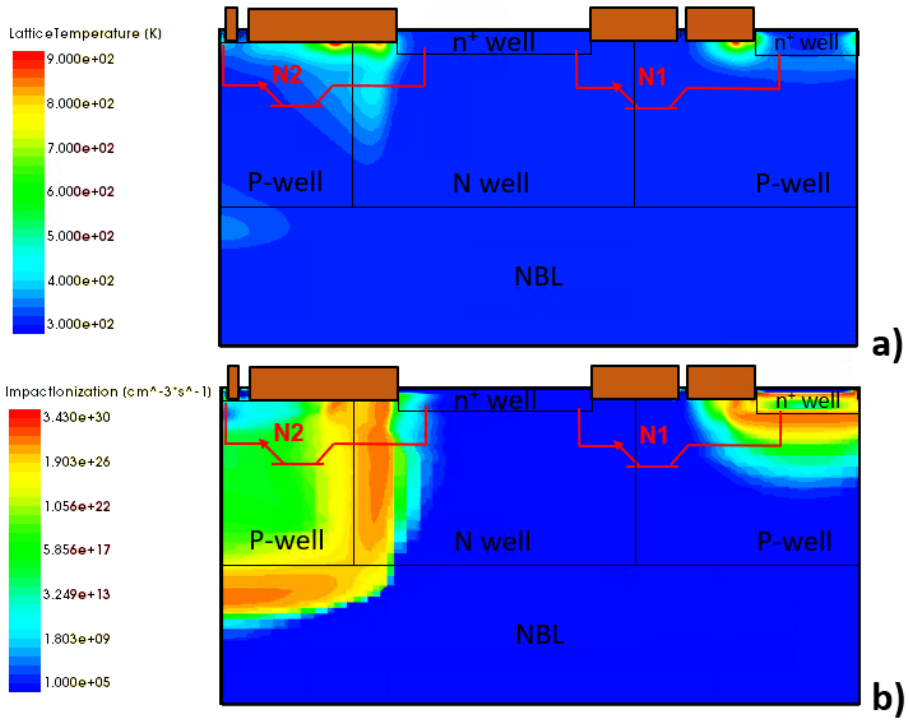


Figure 4.11: (a) Lattice temperature and (b) impact ionization generation under vf-TLP failure conditions

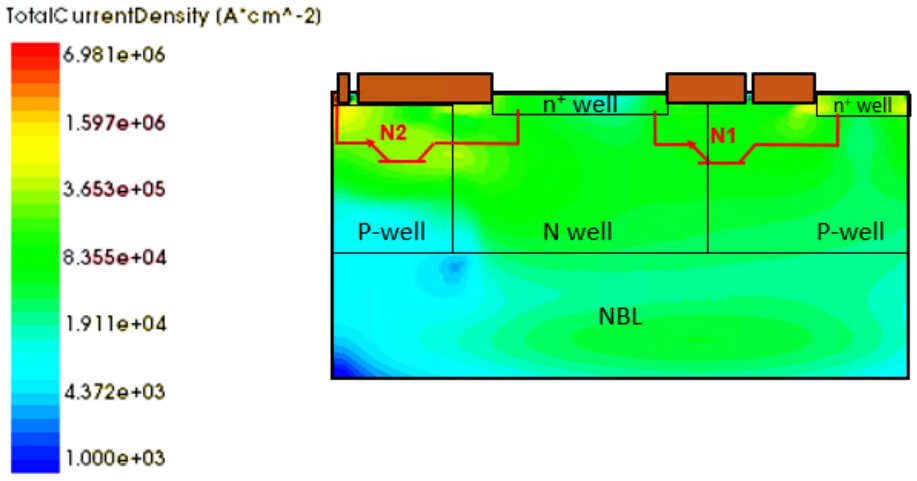


Figure 4.12: Current density at holding voltage under TLP conditions. The current involves the entire volume also of N1.

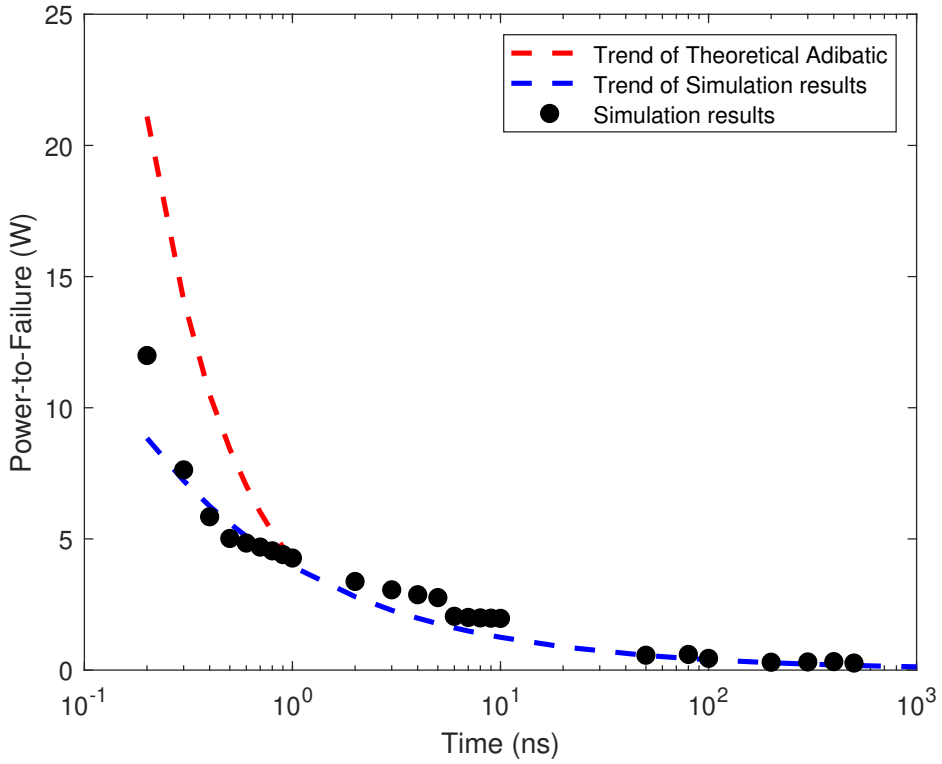


Figure 4.13: Theoretical Adiabatic trends of the Power-to-Failure curve vs. simulation results obtained from TCAD

et al. [9], the P_f scaling dependence with time can be explained though different theoretical models, i.e., the adiabatic model ($P_f = \frac{A}{t_f}$, below 1ns), or the Wunsch-bell model ($P_f = \frac{B}{\sqrt{t_f}}$ from 1ns to 100 ns), where A and B are constants derived from the general heat equation Thus, if A and B can be considered values in continuity between the two regimes, the relationships of the currents derived from these power-to-failure conditions shows at vf-TLP conditions (around 1ns) the current should be almost ten times higher that in TLP conditions (from 1 ns to 100 ns). However, measurements and simulations do not show the latter trends. This occurrence is clearly shown in Figure 4.13.

4.3.3 Proposed modification of the structure to improve the $P_f - t_f$

The overall turning-on mechanisms of the cell can be influenced by specific modifications to the layout and structure as detailed in the following. The 2D TCAD setup can be easily modified to fully understand the behavior of the coupled NPNs and to devise improvements. As an example, the current density distribution at the failure

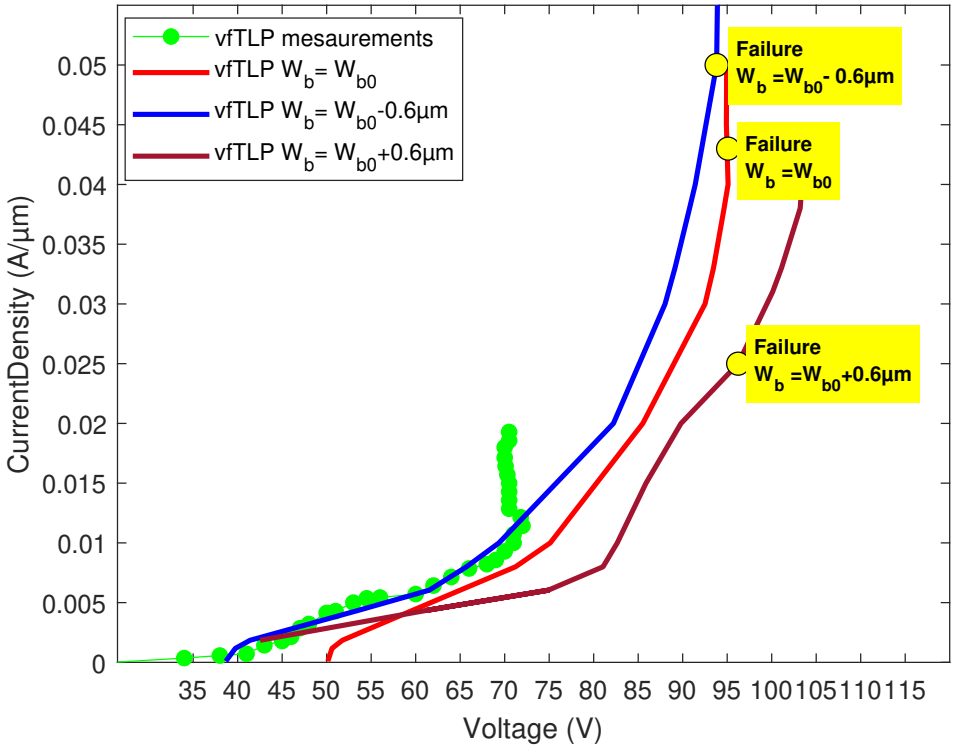


Figure 4.14: Simulated vf-TLP I-V curves of the reference cell and for two different W_b ($W_b = W_{b0} - 0.6\mu m$ and $W_b = W_{b0} + 0.6\mu m$)

may be altered by a layout modification of the deeper part of the base width. Moreover, when addressing very fast stimuli, the NPN turn-on time plays a relevant role, as demonstrated by the voltage overshoot analysis in sub-ns pulses [19]. The proposed modification is the modulation of the distance W_b, reported in the schematic view of the ESD cell in Figure 4.5. W_b is changed trough a layout adjustment of the corresponding central n-well window of doping. The latter consists of a widening or narrowing of the doping window along the depth, consequently changing the shape of the junction itself and the corresponding length of the base width, which defines the N2 gain. Numerical TCAD simulations of EOS, TLP and vf-TLP stresses were adopted to check the expected changes in the overall resistive path and N2 gain. Moreover, the influence of this layout modification on current spreading through the depth of the device is simulated to study its impact on the overall I-V curves and $P_f - t_f$.

For the sake of completeness, simulations have been carried out on devices with $W_b = W_{b0} - 0.6\mu m$ and $W_b = W_{b0} + 0.6\mu m$. In Figure 4.14 the vf-TLP curves of the proposed layout modifications are reported with the failure point, showing a reduction of the holding voltage value for shorter W_b. An improvement of the SOA is observed since also the failure condition is reached at higher current density level.

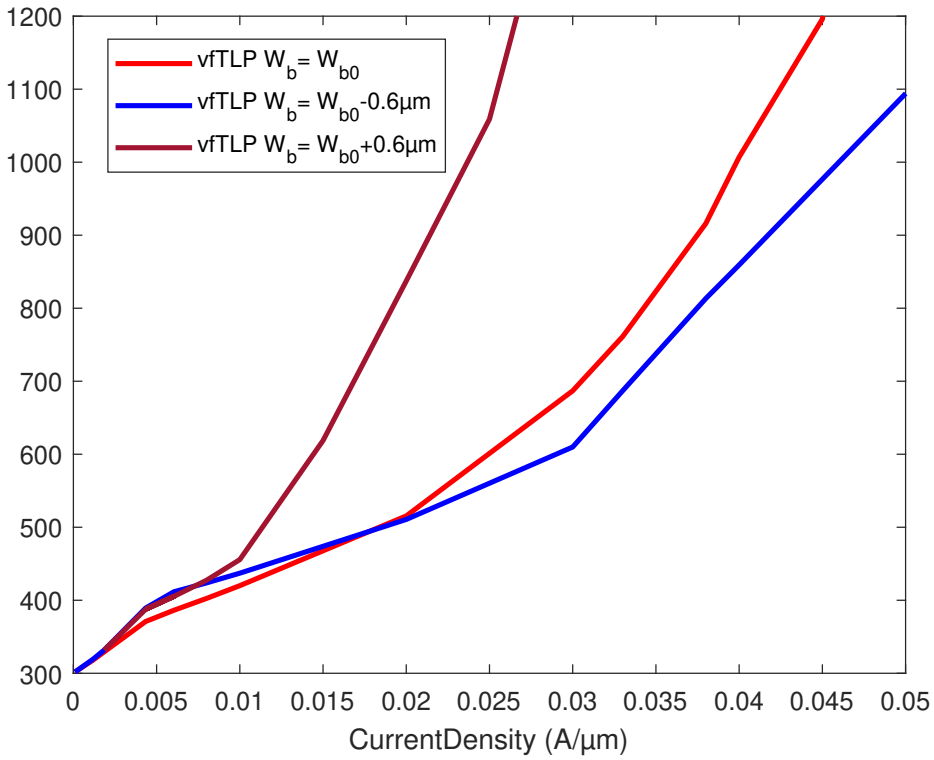


Figure 4.15: Maximum temperature (Tmax) in the 2D domain at the vf-TLP final times vs. applied current density for the reference cell and for two different $W_{b0} = W_{b0} - 0.6\mu m$ and $W_b = W_{b0} + 0.6\mu m$

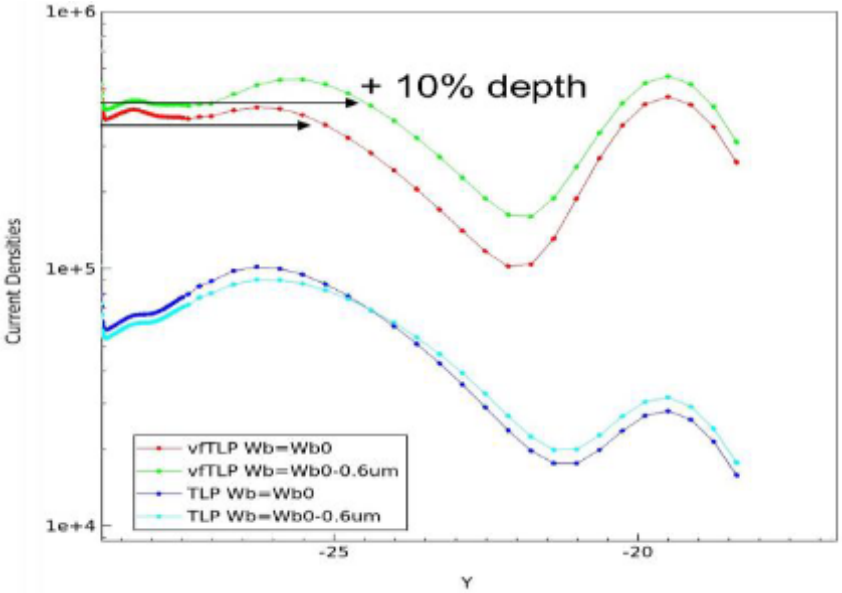


Figure 4.16: Current density along vertical cutline at the center of the ESD cell for W_{b0} and $W_b = W_{b0} + 0.6\mu m$ in vf-TLP and TLP cases.

By extracting a vertical cut line at the center of the ESD cell in the reference device (with $W_b = W_b0$) and for the proposed modifications, the corresponding current density distribution is analyzed more in details in Figure 4.16: in the TLP case, no difference in the vertical distribution of the current density is observed between the two structures. On the contrary, when a very-fast pulse is applied at the maximum current level, a larger amount of current density is observed and the shorter the W_b, the deeper the distribution, with an increase of vertical spreading of about 10%. A larger spreading leads to a reduced self-heating and larger power to failure. The maximum temperature plots in Figure 4.15 confirm the previous statements. Due to the W_b modification, the range of current levels densities goes from 0.025 to 0.05 A/μm, with an improvement of about 16% for the case with the reduced W_b.

4.3.4 Power-to-failure analysis

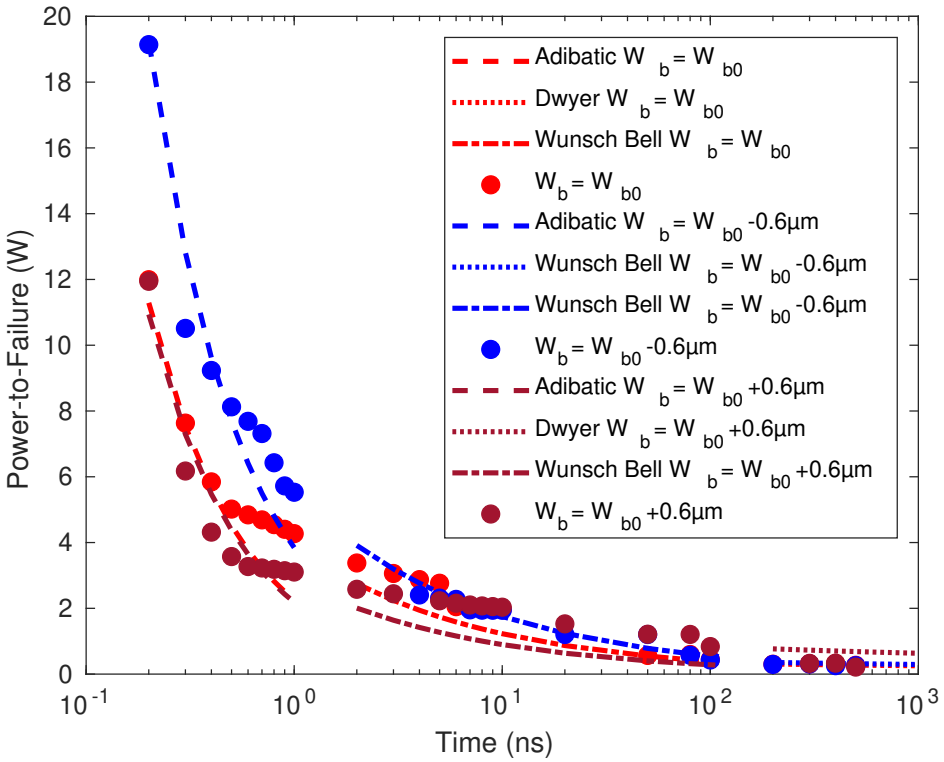


Figure 4.17: Power-to-failure vs. time-to-failure extracted from the TCAD simulations of the reference cell and for two different W_b ($W_b = W_{b0} - 0.6\mu m$ and $W_b = W_{b0} + 0.6\mu m$). Dot-dashed and dashed lines are analytical fitting curves for the adiabatic ($P_f = A/t_f$) and Wunsh-Bell ($P_f = B/\sqrt{t_f}$) contributions, respectively. semi-log scale

The expected trends for different sets of pn junctions [9] have been explained in

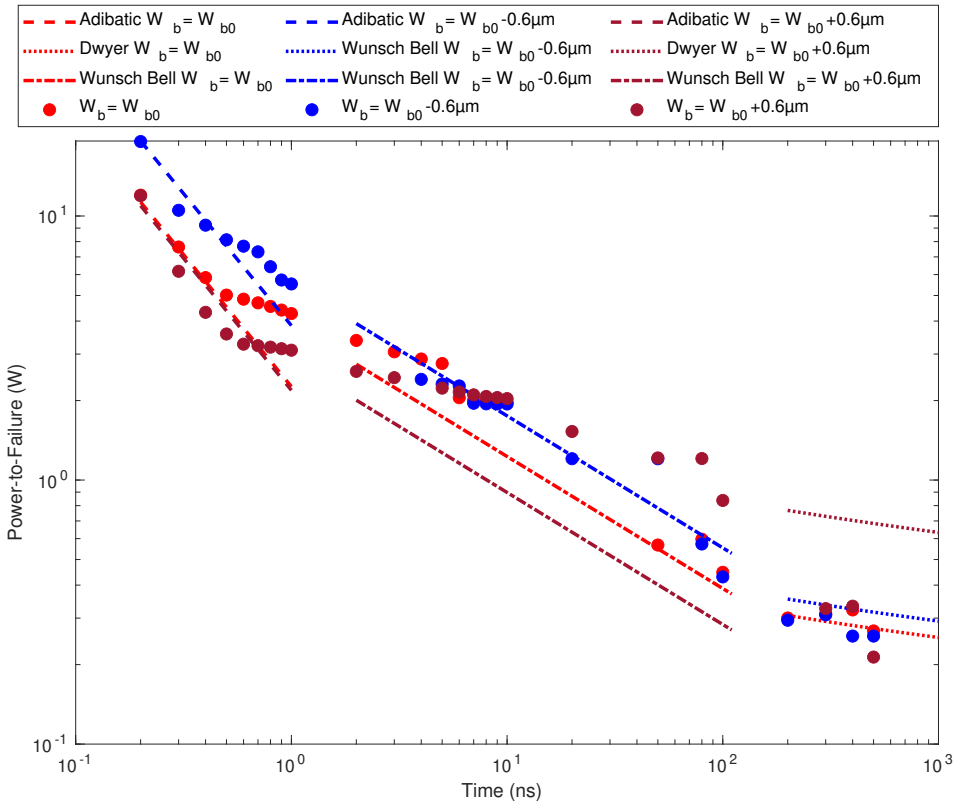


Figure 4.18: Power-to-failure vs. time-to-failure extracted from the TCAD simulations of the reference cell and for two different W_b ($W_b = W_{b0} - 0.6\mu m$ and $W_b = W_{b0} + 0.6\mu m$). Dot-dashed and dashed lines are analytical fitting curves for the adiabatic ($P_f = A/t_f$) and Wunsh-Bell ($P_f = B/\sqrt{t_f}$) contributions, respectively log-log scale

section 3.3.2 in details. TCAD results of the device under test show the expected trends, shown in Figure 4.17 and 4.18 and studied in literature but fail to give a continuity between different regions of stress conditions. The adiabatic region is expected to be extended until 1 ns of pulse length. However, the adiabatic trend is followed by the TCAD data only down to 0.3 ns stress (see Fig. 4.17). As far as the W_b modifications are concerned, as expected also from the results shown in Figure 4.14, a significant increase of the $P_f - t_f$ is obtain for times lower than 1 ns when decreasing W_b. This improvement is explained by the enhancement of the NPN gain. The corresponding improvement of the adiabatic curve is of about 40% with respect to the reference device. For longer stress times the same improvement cannot be achieved since the current mostly spreads at the surface, coupling N1 with N2. Thus, there is no junction or region of the device that can be considered as dominant for the flow of current and for the occurrence of thermal effects. In details, the TCAD data in the TLP stress region from 1 ns to 100 ns approximately do not match the analytical trend but the optimized structure with decreased W_b is the one mostly consistent with the theoretical trends. This result is confirmed by the current path for the TLP case shown in Figure 4.12, since the deeper part of the BJT volume does not influence the IV curves: current flows uniformly between the two transistors giving very similar trends with stimuli larger than 2 ns. For the EOS region, at times larger than 100 ns, very similar trends are observed for all devices following the expected Dwyer equation. As a matter of facts, this range of stress times has been widely tested also in previous literature publications and usually matches the model quite accurately both with simulations and experiments.

4.4 Concluding remarks

A TCAD approach extended to sub-ns time domain is applied to a BiCMOS ESD clamp build of a series of two internally stacked NPN bipolar transistor. The root causes of the thermal failure are identified and a simple layout modification is proposed in order to improve of almost 40% the vf-TLP performances of the device. Results also show that is possible to dive an overall framework of the device, from triggering-on to thermal failure in the whole Wunsh-Bell characteristic.

Chapter 5

Study of SCR-LDMOS for HV ESD Protection

5.1 Chapter overview

The operating conditions of a new ESD clamp based on an SCR-based configuration is presented in this chapter. The analysis is constituted by TCAD simulations in ESD operating conditions compared with available experiments. The device performances are further investigated through geometrical modifications. A study of the holding voltage modulation presented as core analysis.

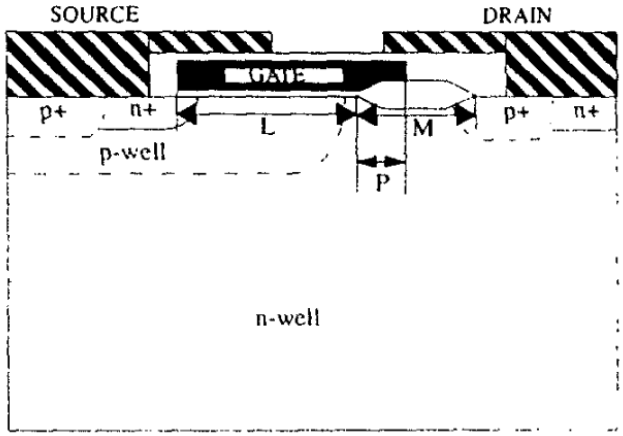


Figure 5.1: Cross section of a general SCR-LDMOS [25]

5.2 The SCR-LDMOS device

In this chapter a conventional SCR-LDMOS, a modification of the power LDMOS device in [42] with the additional p^+ diffusion into the drift region as in [25], is used

as device-under-test. In Figure 5.1 the schematic view of the device cross-section is reported, with the indication of the main geometrical features. The SCR-LDMOS is a modification of a standard LDMOS , providing a significant improvement of the SOA in terms of its ESD robustness [25]. Under high inductive switching, it has the same performance of a standard LDMOS and it goes into the SCR mode under extreme high voltage/high current conditions during an ESD event. The robustness up to failure is simulated since its low holding voltage, dependent on doping concentration and overall geometry, makes it susceptible to ESD-induced latch-up failure for high voltage CMOS/BCD technologies. One of the main drawbacks of this kind of protection is that it collapses very easily to an holding voltage of few Volts. In Figure 5.2 the SCR clamp trend is circled as the expected behavior of the DUT of this chapter.

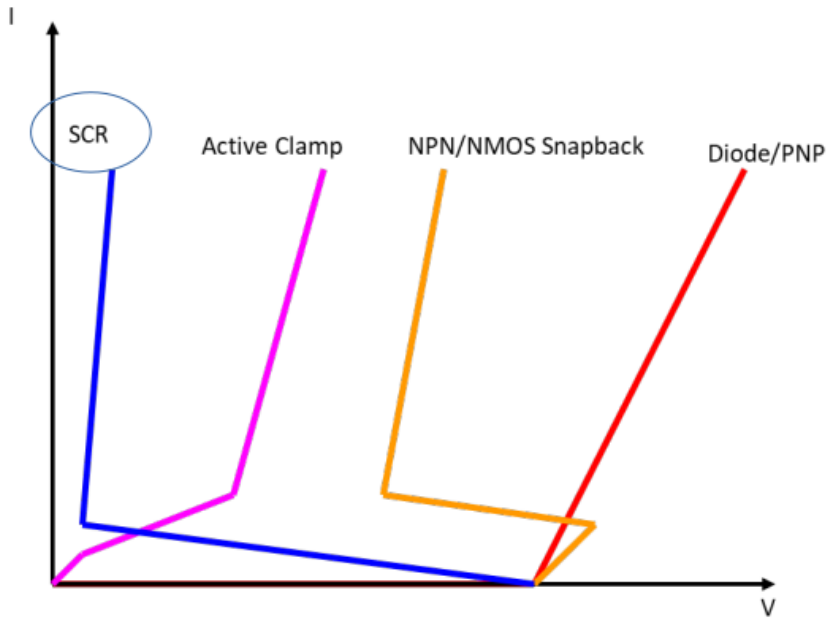


Figure 5.2: Different clamps behavior SOA under ESD conditions. The one under investigation in this chapter is here generalized as SCR

5.2.1 Cross section description

The cross-section of the reference SCR-LDMOS device is reported in 5.3. A schematic representation of the cross-coupled NPN and PNP transistors within the SCR-LDMOS domain is also shown along with the most relevant resistive paths. More specifically, the corresponding circuit reported in Figure 5.4, shows the additional resistive paths given by the drift region (R_{DRIFT}) and by the p-type epitaxial layer (R_{EPI}), reported in series to the respective collector regions. More specifically, two coupled bipolar

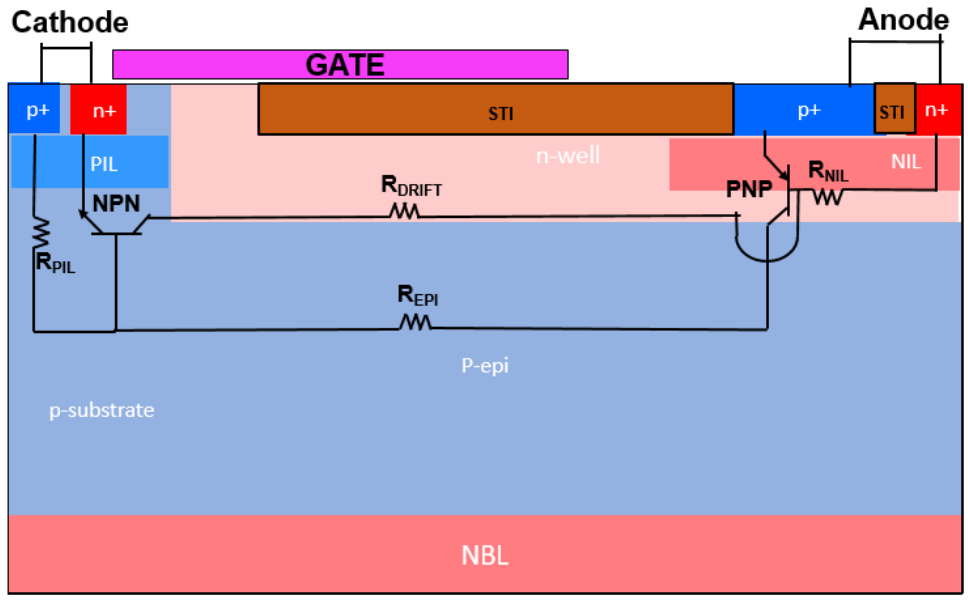


Figure 5.3: 2D cross-section of the SCR-LDMOS cell under investigation. The SCR connection of the two BJTs (PNP and NPN) is drawn in the 2D plot.

transistors are internally stacked to efficiently build the typical SCR, essentially consisting of a PNPN structure, reported in Figure 5.4. The anode of the SCR is formed by the p+ diffusion in the n-well, featuring the emitter region of the PNP BJT. An additional n-type implant well is assumed (NIL), as in [43] with the aim of modulating the holding voltage without affecting the breakdown. The cathode is formed by the n+ diffusion in the P-type implantation layer (PIL), with the aim of turning on a parasitic NPN for LDMOS standard operating conditions. The reference SCR-LDMOS device has been designed for a SOA of V_t around 120V, V_h of around 4V and an I_{t2} of about 50 $mA/\mu m$. Both NPN and PNP devices are in collector-emitter resistance (CER) configuration (i.e., with a resistor between emitter and base). In the case of the NPN BJT, the PIL, usually adopted in the body well of the LDMOS, gives the most relevant contribution to the base-emitter resistance called R_{PIL} in the scheme. Vice versa, an additional profile of the NIL is used under the p+ region, which controls the resistive contribution at the emitter-base of the PNP BJT. The NBL under the SCR structure is tied to the anode and used for isolation. A vertical NPN BJT might be switched on in parallel to the lateral one during ESD events, thus NBL engineering was discussed in few articles even if it is usually a common platform of the technology and specific adjustments should be avoided [44].

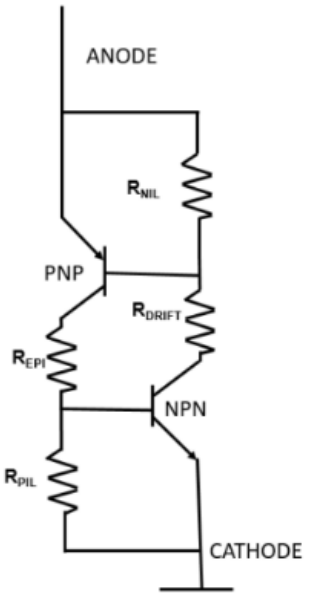


Figure 5.4: Schematic representation of the SCR structure integrated into LDMOS with all the resistive contributions given by the main doping wells of the structure.

5.2.2 Operating conditions

The schematic representation of the equivalent circuit realized in the SCR-LDMOS structure is depicted in Figure 5.4. Both BJTs are switched off during the normal operation of the LDMOS device. The NPN BJT is switched on at high voltage by the onset of avalanche condition in the drift region (the generated holes are drained out by the cathode through the PIL resistive path R_{PIL}). During an ESD event, under extremely high voltage or high current, the structure switches into SCR mode. This configuration is exploited by the LDMOS device in its off state and an ESD stress occurs: the parasitic BJT turns on building an SCR path. This is a typical configuration for bipolar devices under ESD conditions, as they are expected to turn-on at the onset of the impact-ionization regime leading to the snap-back condition. The SCR-LDMOS device has been realized in BiCMOS technology, as explained in Section 3.5. The BiCMOS technology makes use also of a buried n-well doping that can lead to the additional vertical NPN structure, giving rise to an additional vertical path as discussed in the next Sections. The buried layer has been monitored through TCAD simulations as part of the bipolar current is drained through it under high injection conditions. Due to the complexity of the structure, the features of both bipolar transistors must be taken into account simultaneously. More importantly, due to the differences in the adopted pn-junctions and in the geometry of the active regions, the whole device has to be tested in its physical behavior for any new device configuration.

5.3 Simulation of the relevant operating conditions

5.3.1 Triggering and holding regimes

2D simulations have been carried out by using the Synopsys commercial TCAD tool [5]. The impurity concentration profiles have been inferred from process simulation results, but analytical functions have been used to easily consider layout variations of the investigated structures. All simulations have been carried out using the drift-diffusion model, as explained in Section 3.4 coupled with the heat-transfer equation (see Section 2.3. Shockley–Read–Hall and Auger generation–recombination models have been used along with the Unibo impact-ionization model [45]. They were turned on with default parameter values. Thermo-electric simulations have been carried out with ideal thermal boundary conditions, assuming room temperature at the metal contacts and calibrated surface resistance according to critical regions. In Figure 5.5, the available TLP I-V measurements and the TCAD simulations are reported showing a nice agreement in any relevant ESD regime. More in detail, a pulse train with 200ps rise time and standard 100ns pulse length is applied at the anode contact with increasing current pulses. In LDMOS off state conditions, when a positive voltage is applied to the anode, an ESD stress event is induced. The voltage drop into the device is constituted by several contributions which have different impact depending on the physical conditions. The ESD cell is simulated over a range of current ratings that can give valuable information regarding the trigger voltage V_t, holding voltage V_h and failure current I_{T2}. Avalanche multiplication in the drift region causes the triggering of the parasitic NPN BJT leading to the snap-back condition (see Fig 5.5 region A). If a positive current is applied to the anode, due to avalanche breakdown, the generated electrons flow to the cathode terminal, while the holes are injected into the P-epi bulk terminal across the P-body resistor R_{PIL} (see Figure 3.18). When the voltage drop across R_{PIL} becomes higher than the threshold voltage, the parasitic NPN BJT turns on. At this current level, the main potential drop is absorbed by the R_{DRIFT} path (see Figure 3.18. At even higher current levels, the holding condition takes place (see Fig 5.5, region B): the large current flowing underneath the anode region, crosses the NIL region, behaving as the R_{NIL} of the pnp configuration and forward biases the PIL /n+ junction, turning the PNP BJT on (see Figure 3.18. After this, the structure is triggered into the double injection conductive state, and the hole injection from the parasitic PNP BJT becomes the major source of hole current. The required voltage drop between anode and cathode in such conditions is the V_h. The latter is the key parameter treated in this work.

As the current injected increase, the feedback mechanism that is established between the two parasitic bipolars leads the device to an high injection regime (see

Fig 5.5, region C), where the slight electrostatic potential drop is related to R_{NIL} and R_{PIL} mainly. The carrier crowding created by the latter condition enables the occurrence of temperature increase and brings the device into thermal runaway, hence an irreversible condition, as found in Figure 5.5, region D.

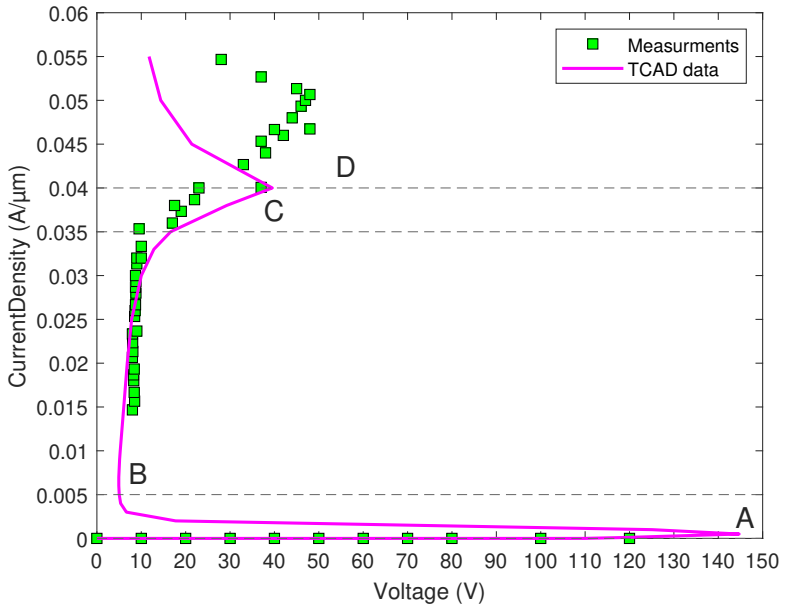


Figure 5.5: TLP I-V characteristics: measurements (dots) are compared with TCAD simulations carried out on the reference structure. Triggering point (A). holding point (B), high-current regime (C) and thermal runaway (D) are indicated

The approach used for the simulations leads to a slight underestimation of the device performances in terms of thermal runaway. However, TCAD results show the worst-case scenario, and the failure condition of the real device is reached at a larger I_{T2} current than the simulated one. No process simulation was performed since the device under test is a tried-and-tested technology. Layout variations of the investigated structures can be easily considered due to the analytical functions used for the implementation of the structure upon Sentaurus tool. Moreover, the 2D simulation is not considering additional 3D effects that might influence the overall boundary conditions. Nevertheless, since the width of the device layout is consistently larger than the length, the 2D cross section is the most relevant current path. Furthermore, 3D approaches are required for the investigation of current filamentations but in the proposed device they are not expected to play any role. Thus, in this work, 2D simulation are adopted to predict the higher levels of temperatures. The aim of the study is to detect and analyze the entire volume of the device and find the junctions involved through the cross-section in each I-V regime. As a matter of facts, even though the SCR structure the majority of the current density is concentrated at the surface of the

device while other parasitic conductive filament path are spread in the deeper part of the device.

When the voltage drop in the drift region increases due to avalanche multiplication, V_t is reached. A significant impact ionization generation, indicated in Figure 5.6 with "ii", along the vertical junction is experienced as well, due to the vertical spreading of the current density in the bottom side of the pn-junction, leading to the turn on of an additional parasitic bipolar transistor, as shown in Figure 5.6. Then the avalanche current flowing underneath the anode side in the p+ region (R_{DRIFT}) can switch on the p+/n NIL junction, triggering the parasitic PNP BJT, as it can be observed in Figure 5.7.

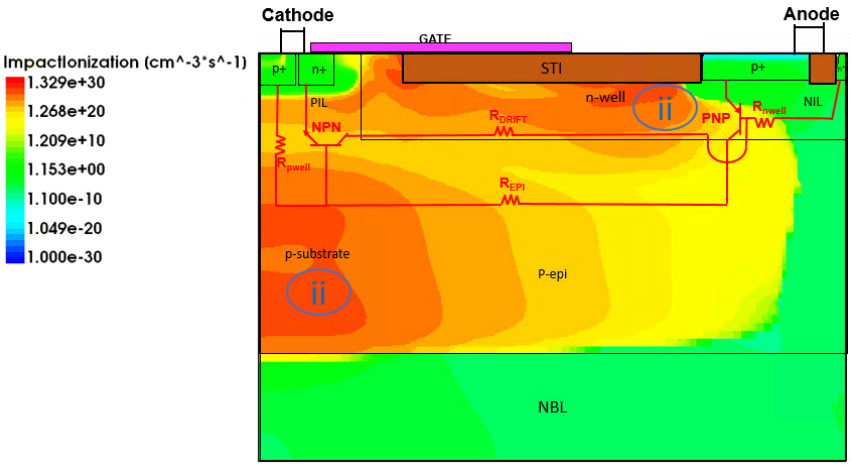


Figure 5.6: 2D plot of impact ionization (i.e. ii) at trigger voltage (Vt = 150 V, I=0.5mA/μ m). It is clear that two main region are in breakdown condition: one at the surface in the drift region and one close to the NBL, creating a vertical path below the NPN device

In Figure 5.8, a 2D plot of the electron and hole current densities at trigger level (around 0.5 mA/μm) reported. The electron current contribution creates a flow of electrons at the anode side. The avalanche current flowing underneath the anode side p+ region (R_{nwell}) can forward bias the p+/NIL junction. The hole current density flowing into the n-well triggers the base-emitter junctions of the NPN BJTs, both laterally between NIL and p-epi and vertically through the NBL triggering the additional NPN. When the device reaches such condition, it is considered triggered and the voltage drop between anode and cathode is the trigger voltage V_t. This avalanche current causes the complete turn on of the NPN transistor while the current flowing underneath the anode region can eventually turn on the PNP transistor. In the holding regime, the PNP BJT turns on, the high injection regime is reached and the feedback loop between the SCR bipolar transistor and the NPN one are created, causing the voltage drop at holding voltage. In other words, the holding voltage is defined by the amount of current that the PNP bipolar transistor needs forward-bias the emitter-base

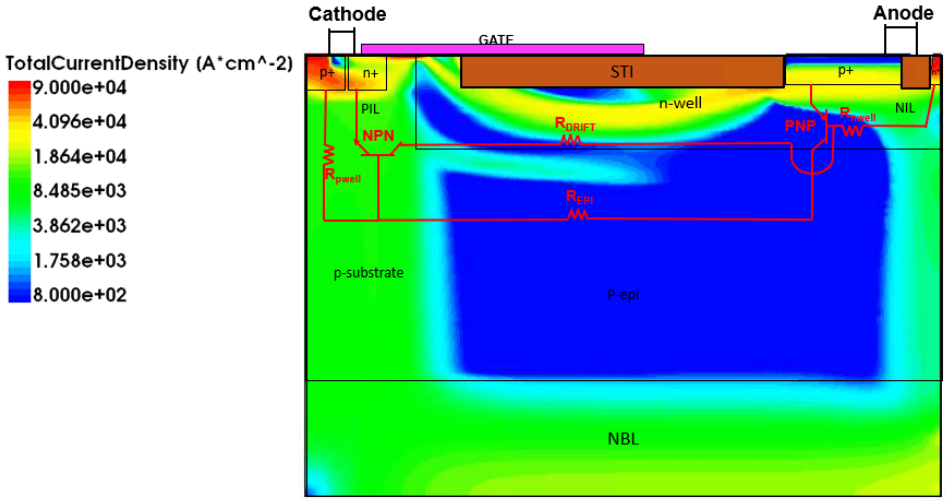


Figure 5.7: 2D plot of total current density (i.e. ii) at trigger voltage (V_t = 150 V, I=0.5mA/μm). The superficial and lateral current path and the vertical one are indicated with a blue arrow.

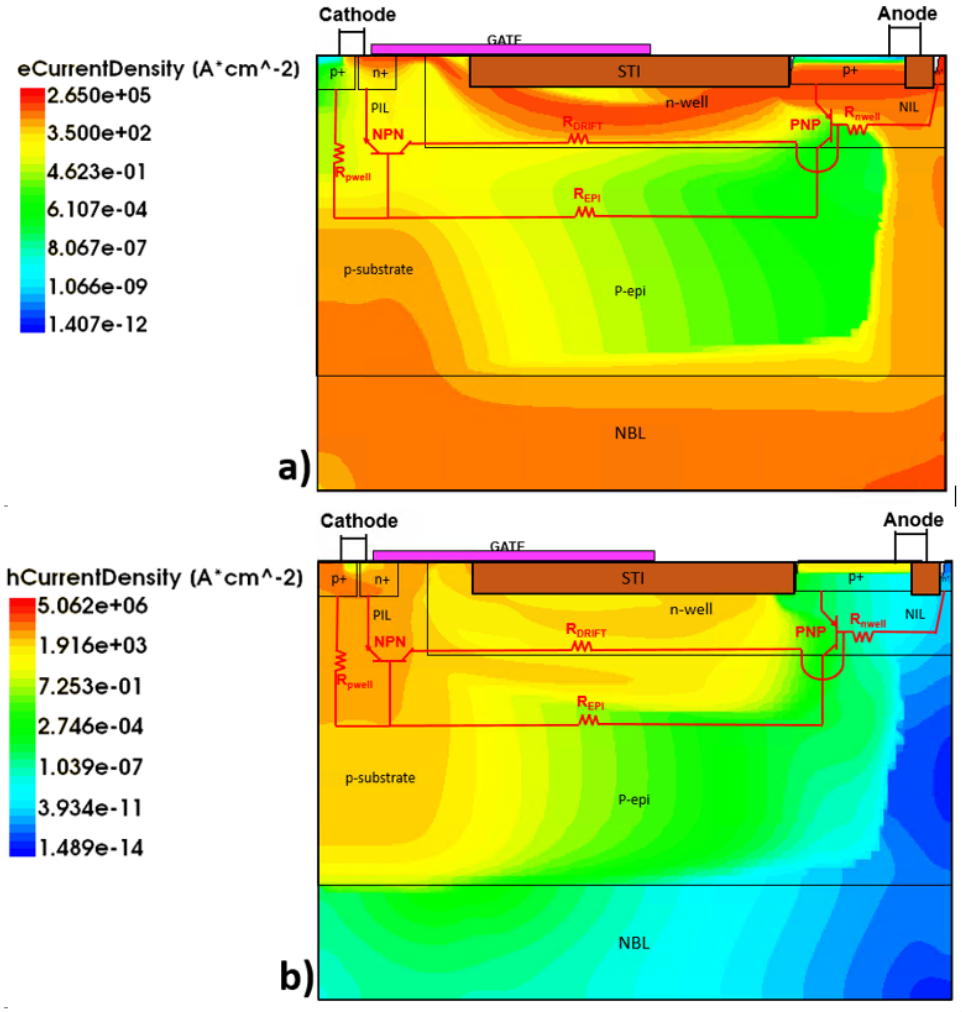


Figure 5.8: 2D plot of electron (a) and hole (b) current density at trigger voltage level. The p+ region added to anode in order to build the SCR is not influencing the current flow.

junction. The de-localization of the impact ionization peak as well as electric field peaks superficially might represent the occurrence of the Kirk effect. This effect can be observed in Figure 5.9. The device is now in high injection regime and the current starts to increase reaching a significant crowding. This congestion of carriers gives rise to relevant self-heating.

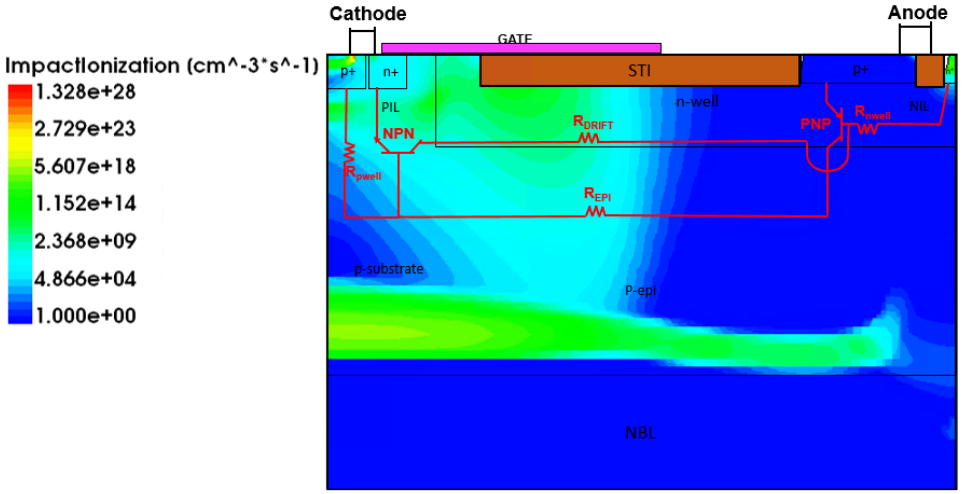


Figure 5.9: 2D plot of impact ionization at holding voltage level. After the breakdown the impact coefficient as turned on both bipolar transistors of the SCR structure.

5.3.2 Failure condition

I_{t2} level is characterized in the TCAD analysis by means of a failure criterion where a maximum temperature of 1200K is used. The current that is driven into the NPN emitter region (cathode n+) due to its small n-type doped area is not able to control the increase of temperature, thus the hot spot is created mostly at the n^+ cathode as it is shown in Figure 5.10.

It is worth mentioning that SOA measurements were not repeated with different rise-times since it is not expected to play any significant role as reported in [46] did not observe any difference. Thus, all key parameters such as trigger voltage, holding voltage and failure current have no significant dependence on the time variation under 100 ps. As current increases, the failure level can only degrade further the device. As a consequence, at a current density level of around 30mA, identified as I_{t2} in Fig 5.5, section D, the device goes into thermal failure irreversibly. As previously said, this level is underestimated by the TCAD simulations with respect to measurements since 3D effects and some process details are not accounted for in the 2D cross-section.

The technology under test is then simulated through the whole ESD spectrum (from 0.2 ns to 1000ns) to test its reliability. In Figure 5.11, TCAD results of the

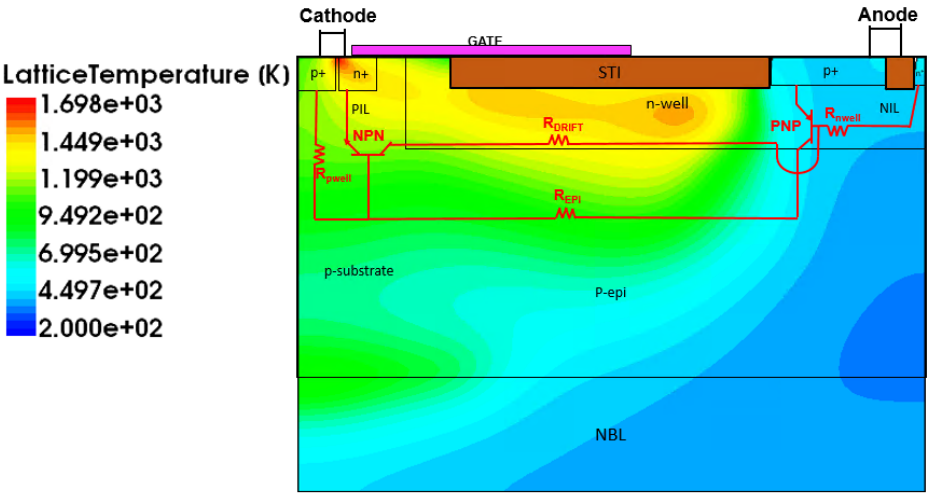


Figure 5.10: 2D plot of lattice temperature at thermal failure level. After the breakdown the impact coefficient as turned on both bipolar transistors of the SCR structure.

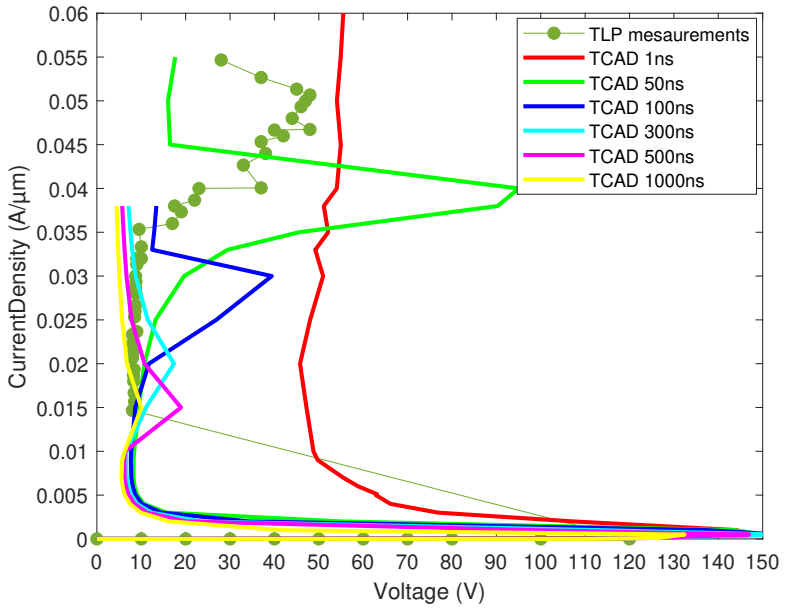


Figure 5.11: Voltage vs. current density of the reference device. The measurement data (dot green line) is given at standard TLP pulse of 100 ns. TCAD simulated curves show the lowering of the curve elbow indicating the decrease of the failure point as the pulse length increases

Pulse width (ns)	$V_t(V)$	$V_h(V)$	$I_{T2}(A/\mu m)$
1	156	50	0.06
50	156	7.6	0.04
100	156	7.6	0.03
300	156	7.6	0.02
500	156	7.6	0.015
1000	130	6	0.01

Table 5.1: Compared values of V_t, V_h and I_{T2} levels of the reference device based on the data in Figure 5.11.

simulated reference device are shown at different pulse times: trigger and holding voltage are significantly modulated at pulses shorter than 10ns due to the reactive behavior of the bipolar transistors in such time ranges. As far as the I_{t2} is concerned, it is foreseeable that the thermal runaway is reached at lower current densities along with the increase of the current pulse amplitude. In Figure 5.12 it is possible to observe the increase of temperature and the failure level using a maximum temperature of 1200K as failure criterion. The results show that the failure temperature is reached earlier as time pulse increases. At 1000 ns (EOS condition) I_{t2} is less than 15 mA/μ m. On the other hand, for stress times shorter than 50 ns the I_{t2} level increases exponentially according to the Wunsch-Bell trend in the adiabatic regime [9]. Since the analysis is not focused on the vf-TLP failure condition, the I_{T2} levels for stress times less than 50ns is not reported. The values of V_t, V_h, and I_{T2} for the whole EOS spectrum are summarized in details in 5.1.

5.4 SCRs characterization: the holding voltage issue

As clearly found with the analysis of the circuit in Figure 5.4, V_h is determined by the intrinsic feedback of the two BJTs, as the shows:

$$V_H = [R_{PIL}(I_{C_{PNP}} - I_{B_{NPN}}) + R_{EPI}I_{C_{PNP}} + V_{ECPNP}] - \mathrm{R}_{NIL}(I_{RAMP} - I_{EPNP}) + R_{DRIFT}(I_{RAMP} - I_{E_{PNP}} + I_{B_{PNP}}) \quad (5.1)$$

Thus, a preliminary study on the structure and doping configuration is carried out to shift V_h to higher values, similarly to active clamps with high holding voltages. A novel structure with adjustable layout and process parameters that sustains a V_h larger than 15 V is the goal of the following set of simulations. In addition, Wunsch-Bell criterion , provides a chance to estimate failure levels of semiconductor devices concerning different pulse durations [8]. Thus, ESD protection shall be ensured by a high second-breakdown failure current (I_{t2}) as well, which is often mostly limited

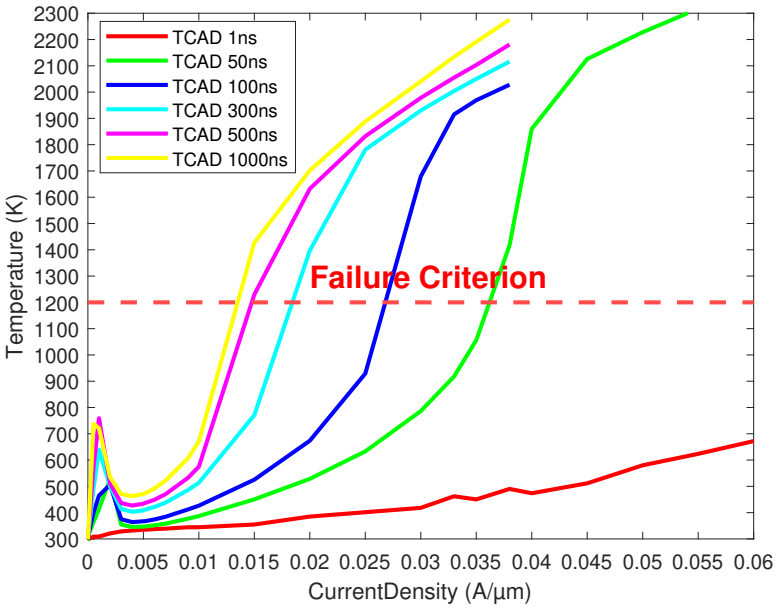


Figure 5.12: Current Density vs. temperature (K) at different pulses lengths from vfTLP (1ns) until DC level (1000ns).

by the thermal behavior [46] but also correlated to V_h.The study of the whole safe operating area is the goal of this analysis, correlating V_t, V_h, I_{t2}, together with failure condition at different stress times.

It is well known that BJT-based ESD protection structures are fairly effective as ESD protection devices because of the large current gain, thus high-voltage SCRs implemented in LDMOS structures the topology shown in Fig 5.1 have a superior ESD protection ability derived by the bipolar injection conductivity modulation [25]. As a matter of fact, the SCRs, due to their capability to switch from a very high impedance state to a very low one, are extensively used in power device applications which require latch-up immunity and reliability, in addition to high ESD robustness. As a specific example, they are mandatory solution to match automotive application requirements and area occupation. However, their holding voltages V_h are still relatively low, as previously mentioned, while V_h is required to be even larger than the voltage supply of the IC power devices to fulfill the protecting purpose. The latter issue needs to be addressed and no simple solutions are available so far, as the most relevant contribution to V_h is given by the parasitic resistors integrated in the LDMOS structure itself.

Several different solutions have been recently investigated in the literature: 2D stacked structures have been proposed in [47], which cause chip area occupied by the device to increase proportionally with the stacking number. On the other hand, a good

latch-up immunity is achieved. In [48] [49] paralleled external resistor is proposed to optimize the conducting resistance, segmenting emitter layout , dealing with a significant reduction of the I_{t2} and increased clamping voltage. Another possible solution was presented in [44], embedding a NMOSFET in the p-well of a modified lateral SCR (MLSCR) was considered another possible solution, but it alters the simple SCR layout scheme and builds more parasitics to be taken care of.

All recent studies focus on the modification of the 3D layout or on the addition of junctions. The best modification gave an increase of the V_h of roughly 20% (an increase of 4-5 Volts). The trade-off between all key parameters should be considered, leading to the definition of a figure of merit as explained in [50]:

$$FOM = V_h * \frac{I_{t2}}{S} \tag{5.2}$$

with S the surface area of the cell, considered as normalized for the sake of comparison. In the following, the conventional SCR-LDMOS solution is taken as case of study and it is modified and analyzed to find any possible correlation between the proposed modifications and the variation of the key parameters V_h and I_{T2}.

5.4.1 Trade-off between lateral and vertical current paths

An important analysis to understand the physical mechanism that triggers the SCR device, driving to V_h is the modulation of the drift length (L_{DRIFT}). In [51] the importance of the lateral dimension is associated with the base length. The modulation of the resistivity of the current path is shown along with the influence on both trigger and holding voltage. The device under test (DUT) is simulated by modifying the layout so to obtain different L_{DRIFT} values. Originally the device length was of 7μm. A vertical and lateral component of the current were detected due to avalanche current flowing through the NBL region, along with the expected surface one. The modulation of the drift region helps to understand the evolution of the vertical or lateral current path. Moreover, it is useful to study the role played by each pn junction according to their relation with L_{DRIFT}. In this way, it is possible to analyze when and how a certain area of the device is crucial for triggering, holding or thermal runaway.

In Figure 5.13 V_t and V_h are reported as functions of L_{drift}. First, the L_{drift} modulation was obtained by adjusting the field plate length accordingly. In this way, it was possible to apply the reduced surface field (RESURF) effect as previously explained in 3.5. Due to the RESURF effect, a flat and uniform electric field is observed along the drift region, preventing the vertical n-well/p-epi junction from the avalanche onset. The drift region would define the trigger voltage without the

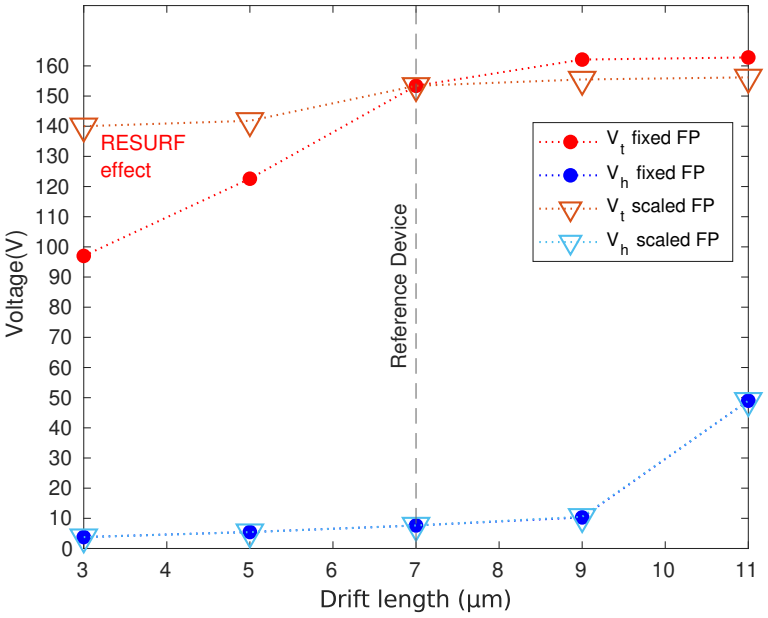


Figure 5.13: Trigger and holding voltage vs. L_{DRIFT}. The simulated results show the role of the lateral and vertical contributions to the avalanche onset (V_t). The holding voltage is quite not impacted by this modification up to very large L_{DRIFT}

interference of any other parasitic path. By increasing further the drift length, an unbalance in the RESURF effect, is obtained leading the vertical field to dominate on the lateral one. As a consequence, the p-sub/NBL junction reaches the avalanche onset at high V_{DS} and limits the V_t. When the length of the gate PolySilicon is not modified consistently, the lateral drift contribution dominates significantly at short lengths. On the other hand, the holding voltage modulation is very limited in both cases (blue curve in Fig. 5.13). An increase is obtained when the vertical current path prevails. However, the latter case is not able to push the carrier crowding from the surface to the bottom, and a temperature increase is reached faster. Moreover, the area of the device is increased consistently leading to a significant degradation of the FOM. A further study was carried out on the active area region of the NPN by modifying the overall area of the device. The purpose of this test is to observe how the base and emitter length of the NPN transistor can influence the holding voltage level, eventually without modifying all the other key values or vice versa fixing the holding voltage while increasing the failure level. More in details, the NPN emitter area and distance from the collector region is increased to find which physical mechanism dominates when changing the bipolar transistor gain and its active area at holding and failure level. In figure 5.14 the regions that are modified are highlighted with arrows. In particular, the length L_E n+ region, identified as the emitter region of the NPN transistor is initially increased. This modification improves the emitter

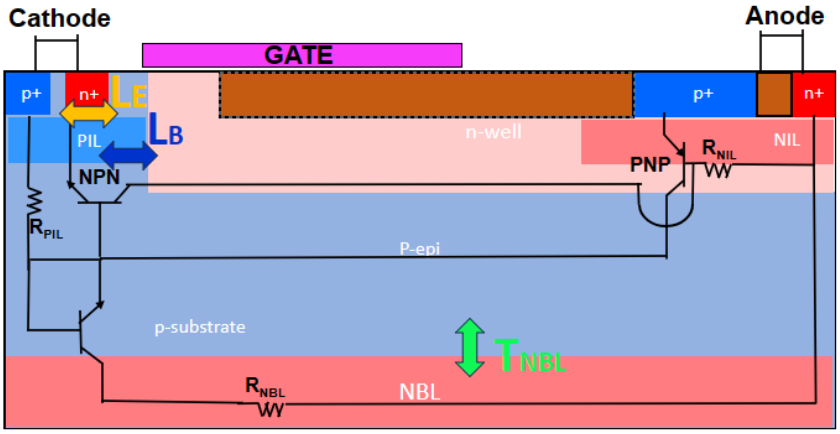


Figure 5.14: Device simplified structure with yellow arrow indicating the n+ alias NPN emitter region L_E that has been modified, blue arrow indicating the base length enlarged L_B, and green arrow the height modified of the NBL position T_{NBL}.

injection efficiency thus bringing down the current gain of the transistor. Simulations result in Figure 5.15 show how this solution cuts the gain and consequently reduces the holding voltage slightly (from 7V to 5V). In high injection regime, the critical point is, in fact, the emitter region due to its small area since it is the region where the lateral and vertical currents are collected. The area enlargement of the NPN emitter spreads the same amount of carriers into a larger area thus reaching failure point at higher current densities.

Simulations with different L_B values have been carried out in order to modify the base extension of the NPN. The structure is then stretched such that the base length increases (see blue arrow in Figure 5.14. Results in Figure 5.16 show an increase of holding voltage related to modification of the NPN transistor: the base length modulates the NPN emitter-base turn on. Through the TCAD tool, it is possible to investigate the lateral current path as well, and its limit in the increase of holding voltage. In fact, with L_B greater than 9 μm different path is turned on as well as in the NBL case, as it is explained later on. Thus, it is possible to deduce that the NPN geometry has a key role in the SCR behavior.

The vertical path observed below the cathode region in Figure 5.8, that provides a certain amount of current to the NPN bipolar transistor before reaching the holding voltage, is of some concern. The vertical path is changed in terms of area, shifting the n- buried layer closer to the surface at distances T_{NBL} in the range of ± 0.5 μm (see Figure 5.14. NBL engineering has been previously discussed in [50] where it was optimized not only as a reliable feature to isolate from the substrate but also for device performance optimization. The purpose of our simulations is to demonstrate that there is an optimal position of the NBL in order to fulfill its purpose of isolation

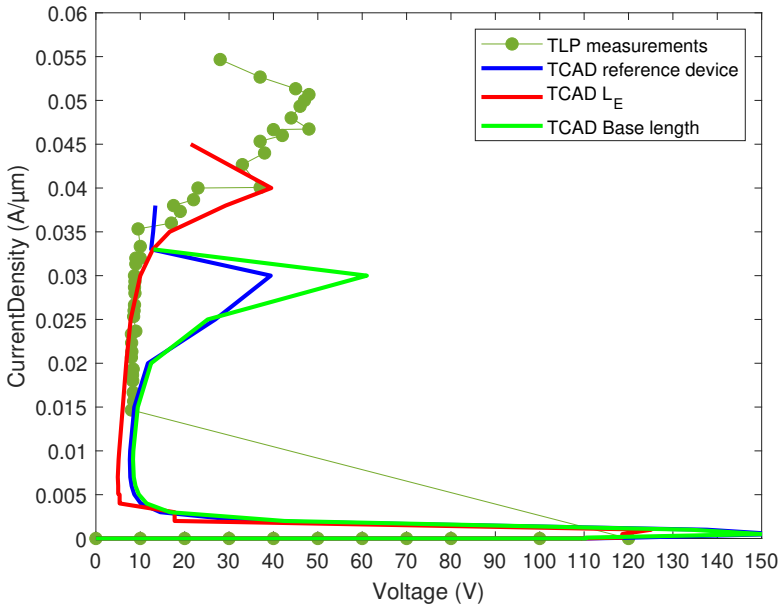


Figure 5.15: I-V characteristic of the TLP characterization compared with available measurements in three cases: black line is the reference device one.

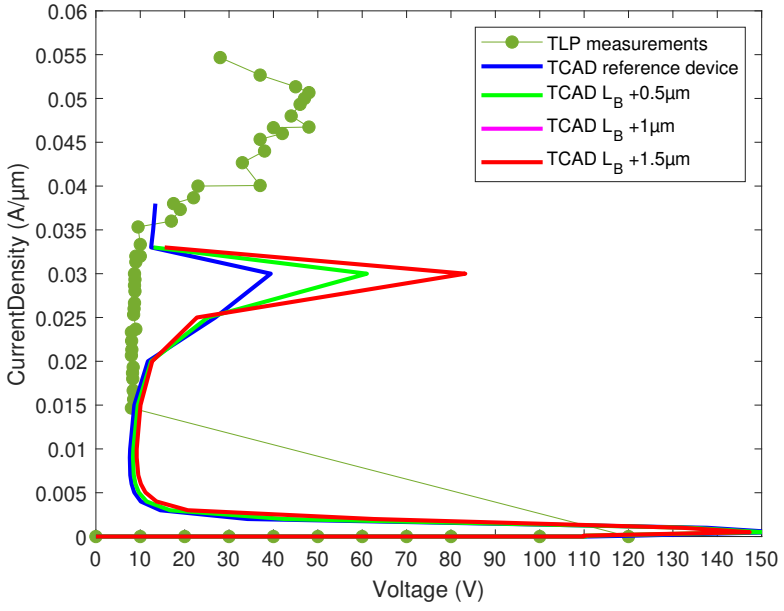


Figure 5.16: NPN base length modulation is measured in this IV curve, showing a limit a slight increase in the holding voltage until the vertical path seen in the NBL case is turned on and privileges with the normal operating conditions of the SCR.

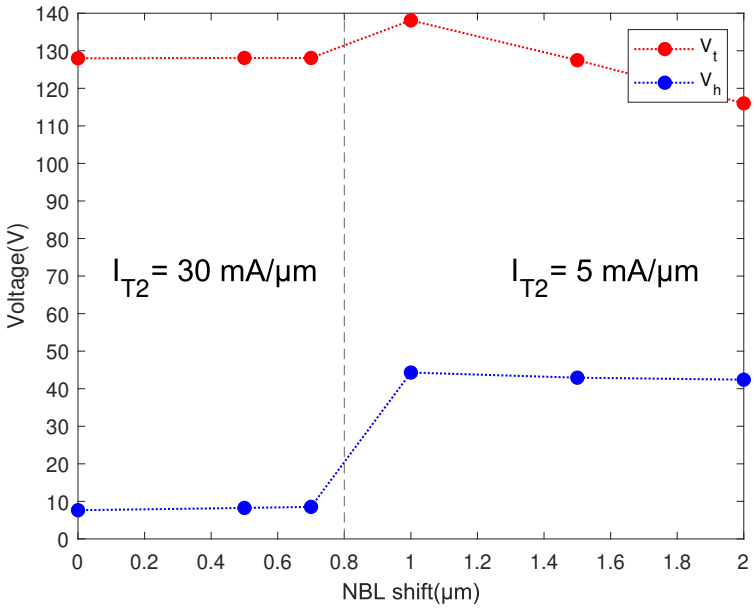


Figure 5.17: Trigger and holding and failure current density (right scale) vs. NBL depth reduction. There is a clear breaking point at around 0.7 μm where the vertical path dominates, and the failure current degrades dramatically

without having a dominant impact on the NPN parasitic bipolar transistor during ESD events. The latter creates a highly conductive vertical current path due to the heavily doped PNP/NBL region. Thus, it is important to control the vertical contribution preventing it from dominating the current flow, as explained in [48]. The limit on the position of the NBL is given by the following outcome: if it is higher of more then L_{NBL} +0.7 μm, the vertical path starts to dominate completely, turning on the additional vertical NPN and preventing the SCR feedback loop. In this condition the PNP bipolar transistor turns on at higher current densities, around 15 $mA/\mu m$; however, the device is incapable of exploiting the SCR features in terms of failure conditions. Thus, the failure point is significantly lower and it is not possible to benefit from the SCR features. The TCAD results for V_t and V_h are reported in Figure 5.17, along with the values of I_{T2}, showing the degradation of the latter while increasing the holding and trigger level of the device.

5.4.2 Effect of NBL contact on device performance

The NBL connection to the anode in Figure 5.18 is simulated to test its role at the holding voltage level. In order to avoid the shortcut between NBL through an additional contact connected to anode, a n-well implant is introduced to lower the resistance in the upper region. This n-well implant is indicated in Figure 5.18 with the

label "DEEPN". It can be inferred that the properties of the PNP and the vertical injection of current would change accordingly with the new adopted doping profile.

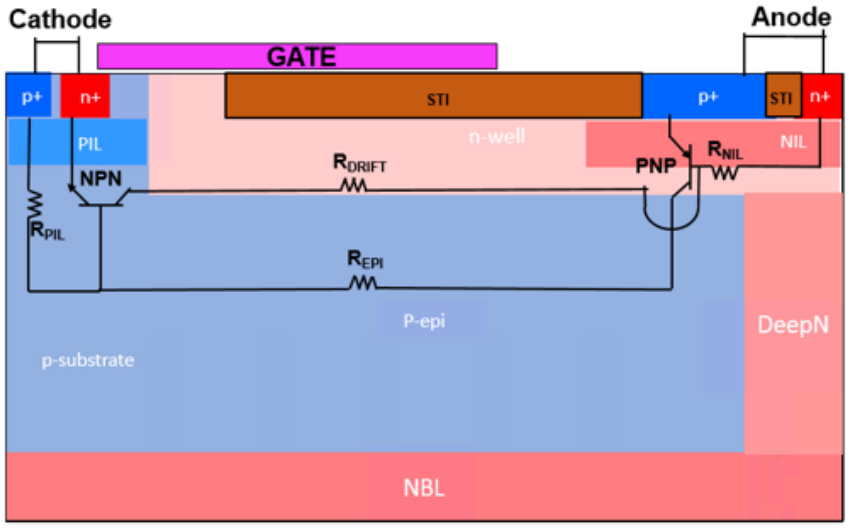


Figure 5.18: 2D cross-section of the SCR-LDMOS cell under investigation. The n-well doping that substitutes the NBL contact is indicated as DEEPN

Simulations allow to understand the role played by the DEEPN path. In the reference device, we have considered as essential the impact ionization at the NBL side. The latter increases the amount of carriers provided by the PNP transistor, giving rise to holding voltage at certain electrostatic conditions. This condition gives rise to the feedback between the two bipolar transistors at holding voltage. Without the connection of the anode to the NBL, the behavior of the SCR is localized only at the surface. The carriers injection is mainly squeezed in the upper part of the device, thus reducing the holding voltage level. However, the vertical current path can give a considerable contribution to the amount of carriers fed to the PNP base to turn on the SCR at the onset of holding conditions at larger voltages. The latter is further investigated in Figure 5.19 where the current density at the holding voltage is shown. In Fig. 5.19 (a) the NBL is directly shorted to the anode thus the current is spread both superficially and vertically through the NBL; in Fig 5.19 b) the same amount of carriers is mostly concentrated at the surface but no contribution from the vertical side is assessed.

Moreover, the electrostatic potential distribution at holding voltage change in the two cases and, in particular, the amount of carriers provided by the NBL current enables the SCR to reach higher holding voltages. Thus, in case of Figure 5.19 (b) the holding voltage is lower. This occurrence is further confirmed by the impact ionization distribution in the two cases as shown in Figure 5.20. The carriers flowing into the NBL drive into avalanche the NBL p-epi junction, providing more carriers to

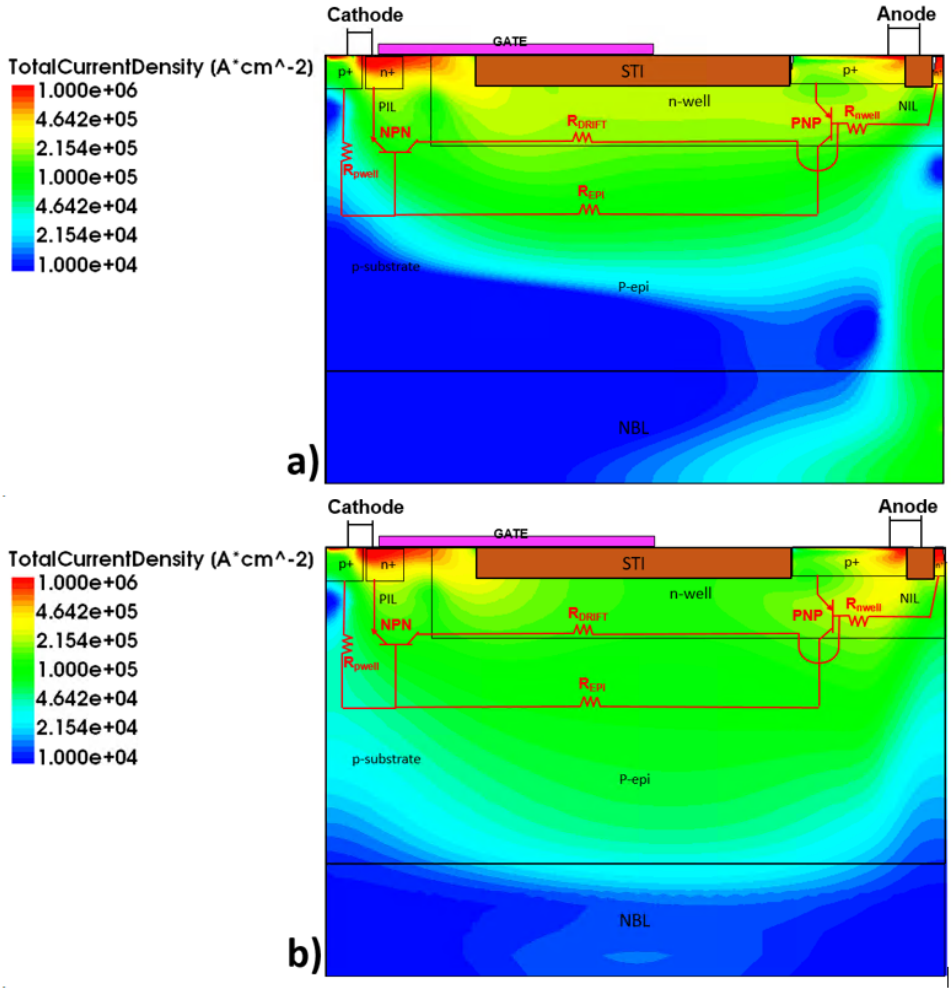


Figure 5.19: 2D cross-section of current density distribution of the SCR-LDMOS cell under investigation at holding voltage level a) with NBL contact shortcut to the anode and b) without NBl contact

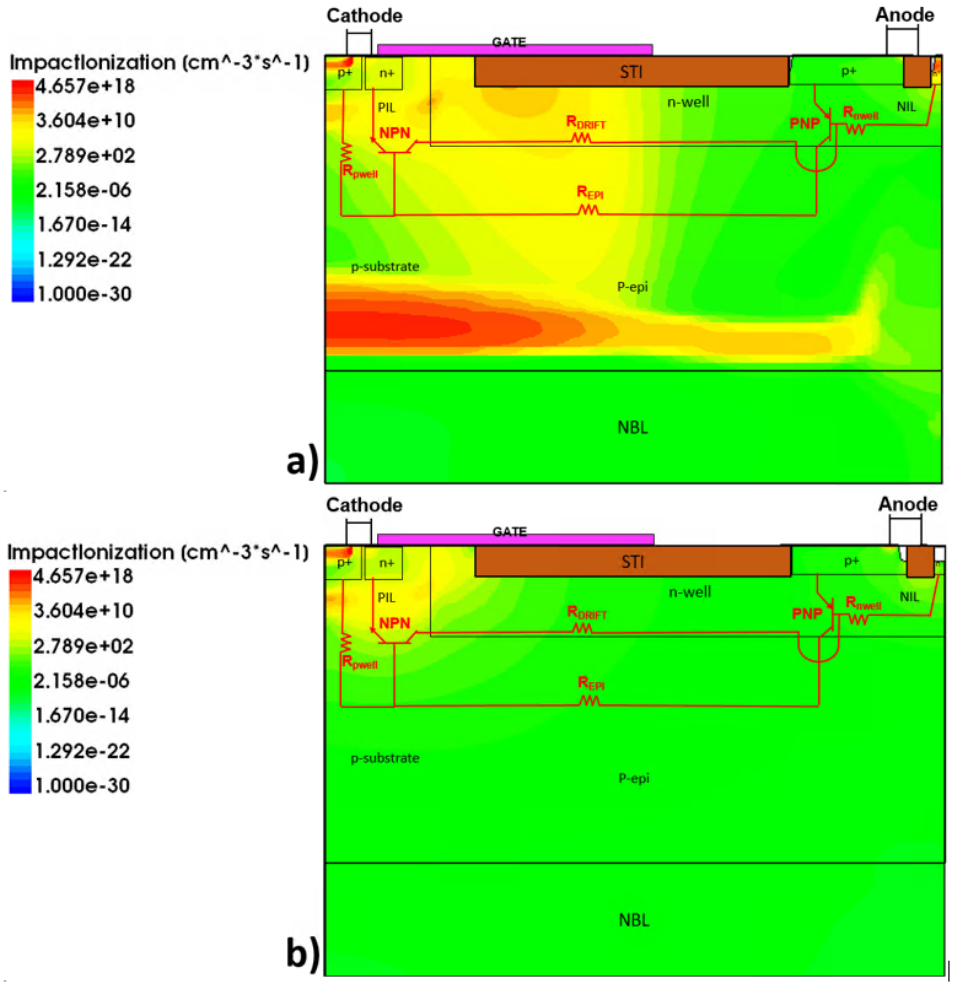


Figure 5.20: 2D cross-section of the impact ionization coefficient in the SCR-LDMOS cell under investigation at holding voltage level a) with NBL contact shortcut to the anode and b) without NBl contact

the PNP bipolar transistor (Figure 5.20 (a)). In the hypothetical absence of the DeepN connection between the anode and the NBL, the impact ionization distribution is concentrated at the surface and it leads to lower electrostatic potential (lower holding voltage conditions) as shown in Figure 5.20 (b). Therefore, in order to re-establish the connection between the drain contact and the NBL, the DEEPN resistivity is modified: if the overall doping of the DEEPN is increased, the resitivity is lowered and the carriers are able to reach the NBL without any further shorting. Hence, the NBL contact is not needed.

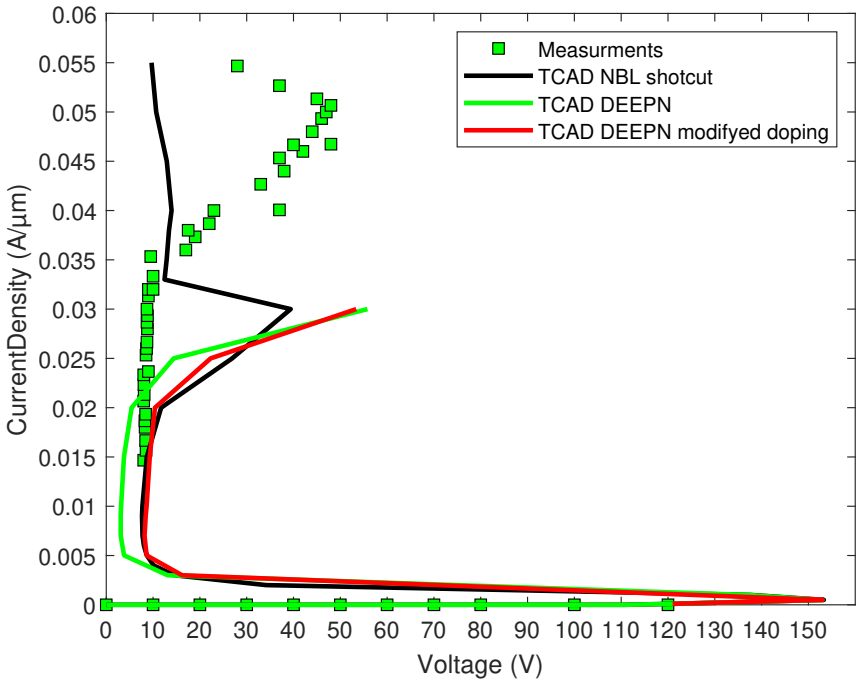


Figure 5.21: TLP I-V characteristics: measurements (dots) are compared with TCAD simulations carried out on the reference structure comparing results obtained with and without NBL contact short-cut with the anode of the device: results of TLP simulations (100 ns pulse length) show the role played by the injection of carriers from the NBL side

Results are shown in Figure 5.21 where the I-V curves obtained with the default device with the external NBL shortcut are compared with simulations carried out without shorting the NBL but adopting the DEEPN doping. The different DEEPN doping configurations lead to different V_h. The highly doped DEEPN reaches the same V_h of the shorted NBL, establishing a vertical connection between the surface and the buried layer of the device.

To further confirm that the change in the DEEPN resistivity is almost equivalent to the NBL external contact in Figure 5.22 the current density distribution in these two cases are compared. The vertical path driving part of the carriers down to the

NBL side is clearly evident in both cases, even if a different vertical distribution is still visible.

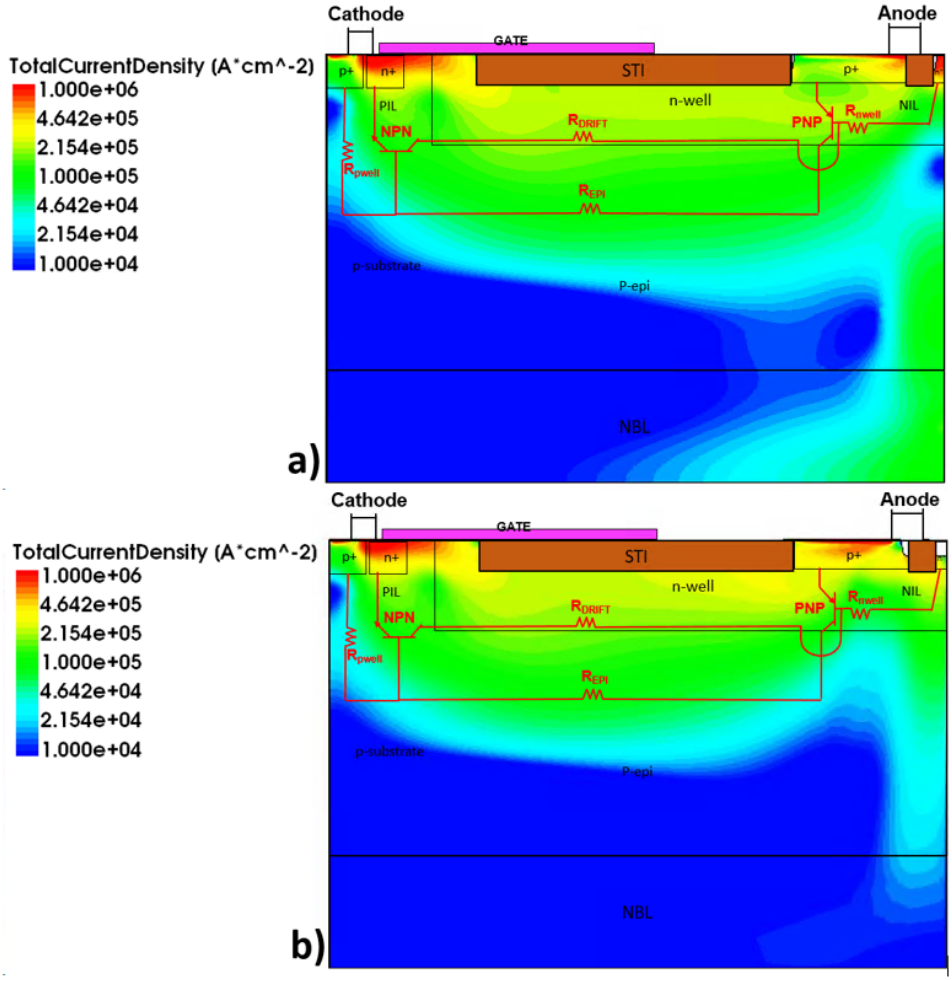


Figure 5.22: 2D cross-section of the current density distribution in the SCR-LDMOS cell under investigation at holding voltage level a) with NBL contact shortcut to the anode and b) with DEEPN doping profile modified

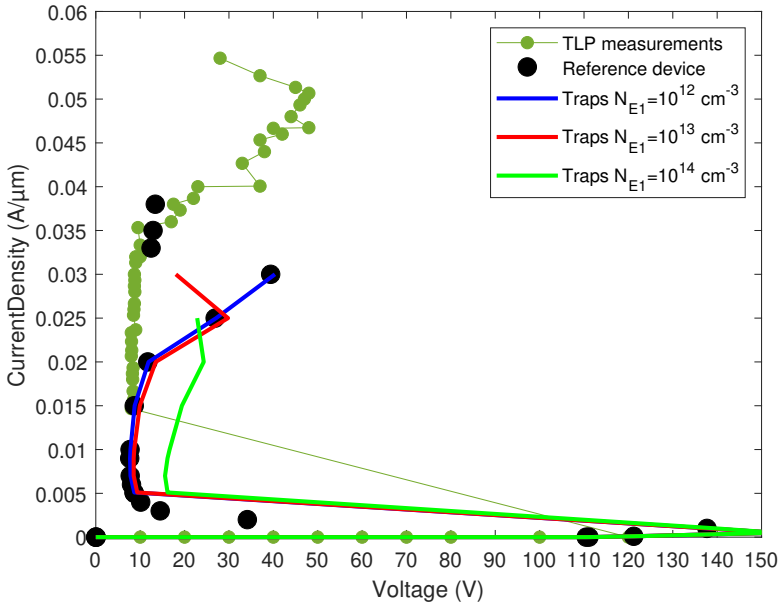
The setup that is used in the next Sections considers as optimal configuration the one with an NBL contact, since it gives the best stability in terms of holding voltage and thermal breakdown for the sake of the analysis.

5.5 Studies for the improvement of V_h

5.5.1 Holding-Voltage modulation through Irradiated SCR-LDMOS

In [52] another technique for the SCR-LDMOS improvement is proposed assuming the electron or ion irradiation as a key tool for off-line optimization of the device parameters. The proposed solution is as follows: TCAD electro-thermal simulations

have been carried out on the standard SCR-LDMOS device (see Figure 5.3) by reproducing artificially through simulations the defects originating from the irradiation with high-energy electrons as in [53]. This is the so-called carrier lifetime-control process, with major advantages in high reproducibility of the dose and stability of the tailored lifetime after annealing, best spatial uniformity and optimal concentration ratio of the dominant defects. The proposed solution was successfully applied to discrete high-voltage power devices in previous works [54] [55]. The generated traps were experimentally characterized for different irradiation conditions and accurately implemented in TCAD simulations [56]. Following [56], the role of the defect traps induced by the electron irradiation has been investigated on the SCR-LDMOS device, by running TLP simulations up to the thermal failure for different trap concentrations.



trap

Figure 5.23: TLP I-V characteristics: TCAD simulations carried out on the reference device (black symbols, no traps) and on the proposed irradiated device with three different bulk trap concentrations (N_{E1} = 10^{12}, 10^{13}, 10^{14} cm^{-3}).

Irradiation of silicon with high energy electrons generates a spatially uniform distribution of point defects acting as generation-recombination centers. The most prominent levels originate from oxygen-vacancy pairs and divacancies which are located at 0.16eV (E1) and 0.42eV (E4) below the conduction band, respectively [57]. The former dominates in the high injection regime and reduces the carrier lifetime at high carrier concentrations. The latter affects the reverse-bias condition since it increases the two-step trap-assisted generation probability. Thus, electron irradiation has been used in the device simulation setup assuming that the bulk traps are two

acceptor levels (E1 and E4), uniformly distributed with a certain concentration ratio. Process simulations are not needed for the definition of traps since the spatial and energetic features are given by experimental characterizations [56]. Simulations have been carried out considering a range of acceptor concentration from 10^{12} to 10^{14} cm^{-3}. In Figure 5.23 current densities obtained with the latter current concentration are shown. A significant modulation of the V_h has been observed for the case of acceptor concentration of 10^{14} cm^{-3} (see results in Figure 5.23, with no significant variation of I_{t2} or V_t.

In order to observe the internal behavior of the device in the different key regimes of the I-V curve, the impact of the trapped charge in the structure has been monitored. The current level is at the holding condition, corresponding to a holding voltage of 8 versus 12 Volts, respectively. The current densities at the same current pulse for no traps concentrations, and trap concentration of 10^{12} cm^{-3} are shown in Figure 5.24. The current distributions presents relevant surface crowding on the right of the structure that increases when the trap concentration reaches 10^{14} cm^{-3}. Thus, we can deduce that traps are suitable for the modulation of the V_h since it has a substantial impact on the current path. The trapping occupation in Figure 5.25 enables further considerations, showing that acceptor traps are filled in a different way with a trap concentration of 10^{12} cm^{-3} with respect to the highest concentration of 10^{14} cm^{-3}. More specifically, the trapped charge in the upper right-side n-well region gives rise to the V_h increase.

In Figure 5.26, a significant increase of the resistive path due to the presence of traps can be observed in the position 4 to 7 μm. As a matter of facts, the role of traps is to decrease the recombination lifetime in the high-injection regime reducing the excess of carriers flowing in the device. In Figure 5.27 is used to extract the carrier densities. The device is in high injection state since electron and hole concentrations are equal and much larger than the doping concentration. From 7 to 10 μm the higher trapping occupation density reduces significantly the electron and hole densities, thus increasing the resistivity.

The role played by traps can be further investigated by implementing the uniform distribution at maximum concentration (i.e., 10^{14} cm^{-3}) only on limited areas corresponding to the active regions of the NPN (left side) and PNP (right side) BJTs, respectively. As shown in Figure 5.28, with lateral extensions given by the p-well region on the left side and by the p+/n+ extensions on the right as indicated in Figure 5.26. A negligible effect of the irradiation on the left-side of the device is observed. As a matter of fact, the V_h shift is around 1% when considering only the irradiation related to the NPN area, while an increase of V_h of almost 16% with respect to the

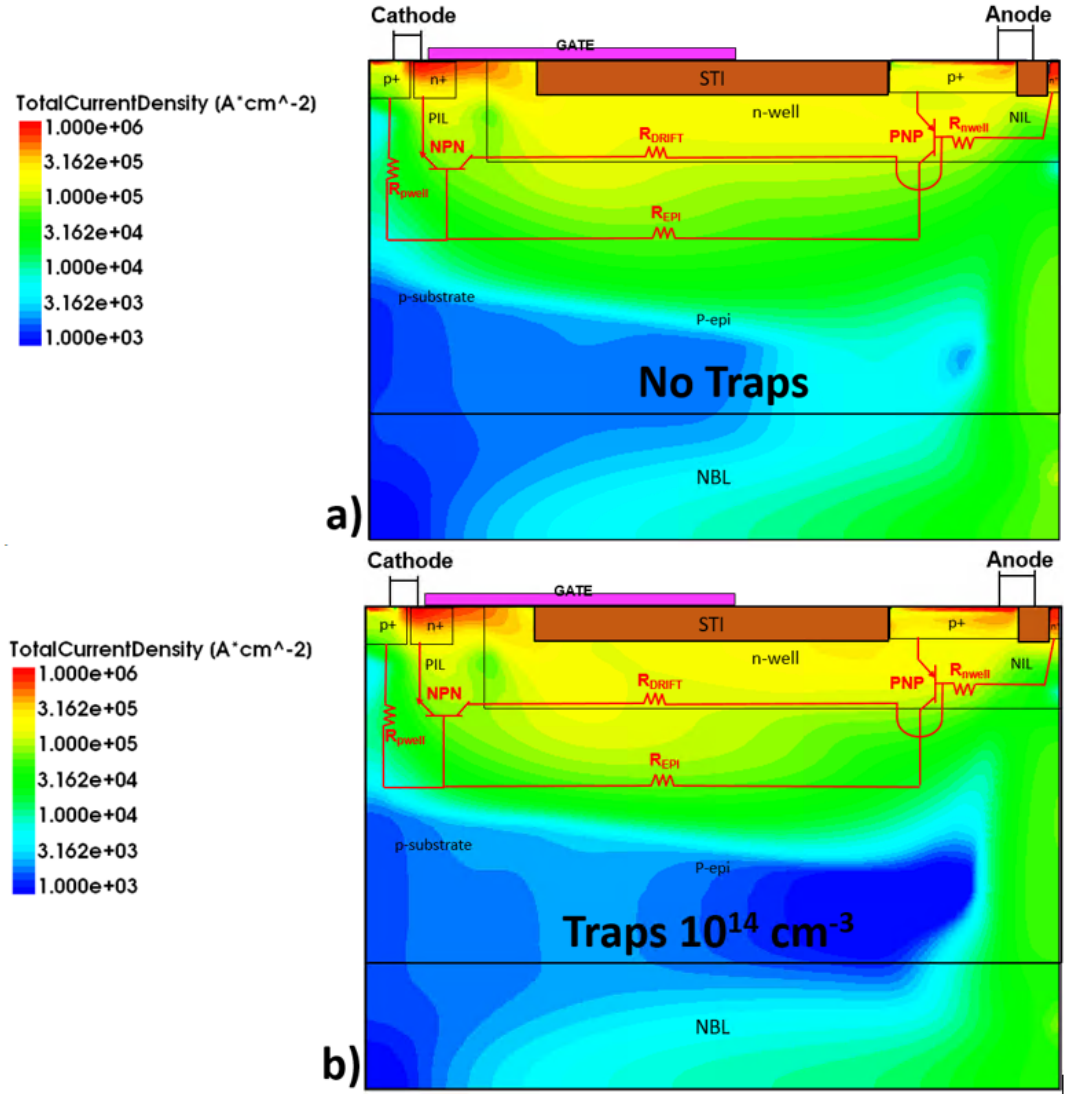


Figure 5.24: 2D plot of the total current density at holding level for the case without traps (a) compared to the case of $N_{E1} = 10^{14}\ cm^{-3}$ (b).

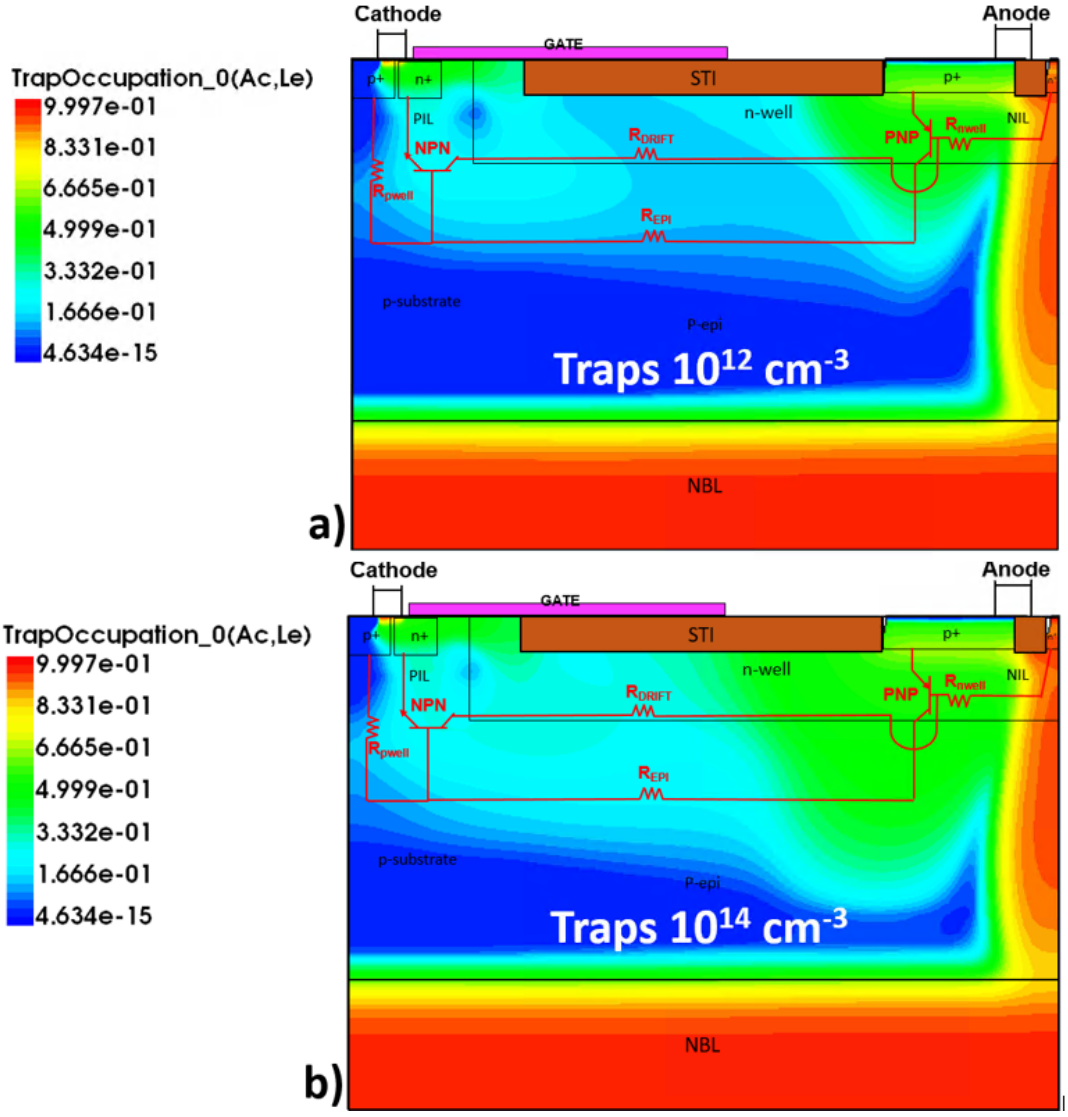


Figure 5.25: Electron trap occupation at the holding voltage in the cases with trap concentration of (a) 10^{12} cm^{-3} and (b) 10^{14} cm^{-3}.

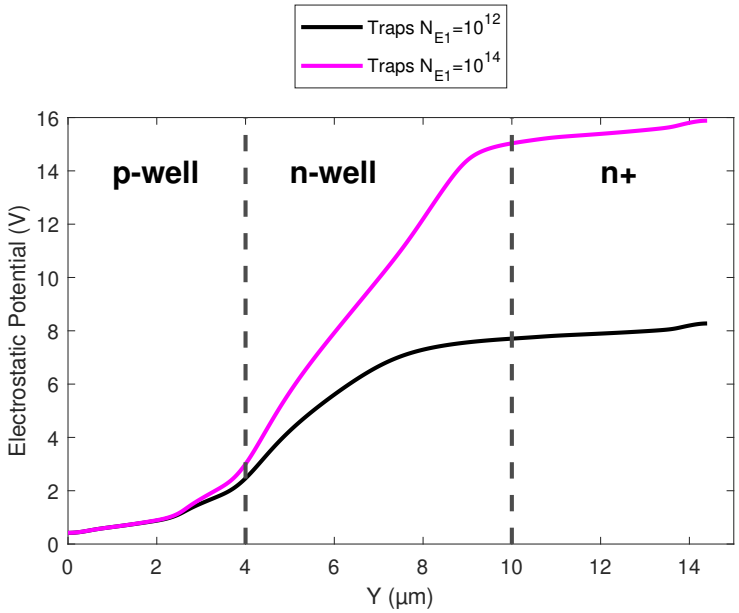


Figure 5.26: Electrostatic potential at the holding voltage in the cases with different trap concentration.

original device is detected when irradiating the PNP region. Thus, the trapping effects are effective at the base region of the PNP BJT, influencing the corresponding high-injection regime.

Finally, the self-heating effect at irradiated conditions is also tested with additional simulations performed on the reference device comparing the thermal behavior with and without traps, assuming ideal boundary conditions. The TLP pulse applied for this set of simulations is fixed at 5 mA/μm. In Figure 5.29, the results of V_h at different temperatures are compared reporting the cases with and without traps. From the previous analysis, it is known that traps start to influence the IV curves of the device after a certain concentration. Thus, a difference on the V_h is expected. The thermal behavior of the BJTs is considered the cause of the reduction of about 32% of V_h for the reference case with no traps, as self-heating causes maximum localized temperature of about 500K, mostly on the left part of the n-well region. On the other hand, the irradiated device with maximum concentration of traps has a weak de-trapping mechanism leading to a smaller variation of V_h of around 10%. This behavior is induced by the ambient temperature increase as the self-heating hot spot is still located in the left part of the n-well region, where the role of traps is less effective. Therefore, the trapping solution is less sensitive to temperature conditions with respect to the reference case.

For the sake of completeness, a thermal failure analysis has been performed as

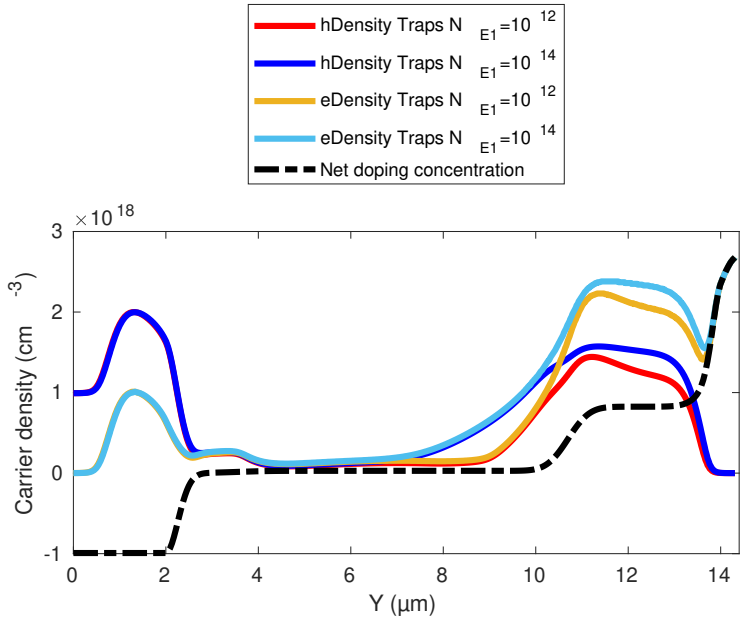


Figure 5.27: Cut-line of the device at holding voltage level for hole and electron density at different trap concentrations: $N_{E1} = 10^{12}\ cm^{-3}$ and $N_{E1} = 10^{14}\ cm^{-3}$

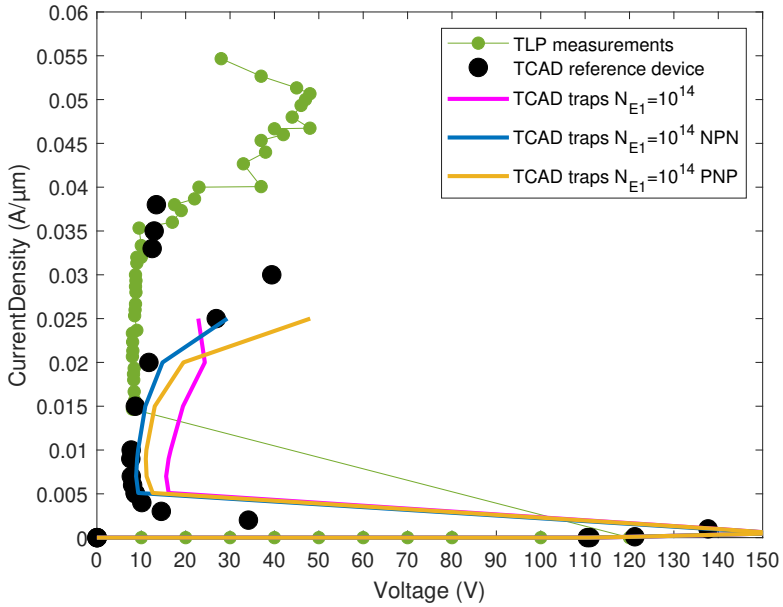


Figure 5.28: I-V characteristics of the device with traps over the whole domain compared with those with traps limited in the left and right regions, corresponding to the NPN or PNP BJTs.

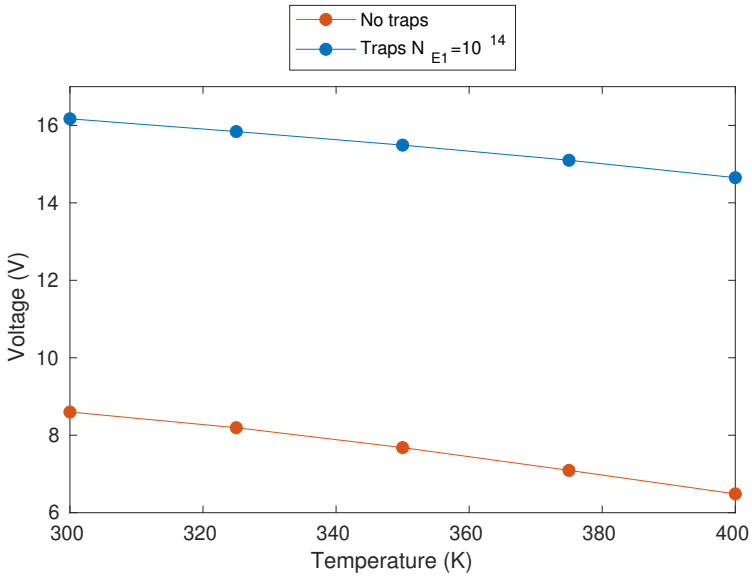


Figure 5.29: Device simulations of the V_h dependence on the ambient temperature for the reference device compared to the irradiated case with the highest trap concentration.

explained in [37]. The failure criterion adopted is the following: the maximum temperature in the structure is monitored at the end of the current pulse (100 ns), and a temperature of about 1200 K is considered at the onset of thermal runaway. In Figure 5.30, triangles in the TCAD curves correspond to the extracted thermal failure. It is clear how the trapping solution does not significantly influence the thermal heat-up of the device, thus failure degradation is preserved. In particular, the power-to-failure degradation for the irradiated case with respect to the reference one is only -6%.

5.5.2 Holding-Voltage modulation with doping and layout modifications

The first analysis carried out in [43] on the improvement of V_h in SCR-LDMOS devices, shows how an additional heavily doping N-type implantation layer (NIL), added underneath the N+ and P+ drain of the standard SCR-LDMOS, can significantly lower the emitter efficiency of the parasitic PNP BJT and so increase the holding voltage. With this approach, neither any external circuitry nor a larger layout area is necessary, while obtaining features of a proper V_t, a high enough V_h and acceptable ESD robustness. Thus, the trigger of both bipolar transistors is examined here by addressing similar approach in order to understand how it influences the feedback activation of the whole SCR structure. In particular, the PIL and NIL doping are modified and engineered. The Safe operating Area (SOA) of the device

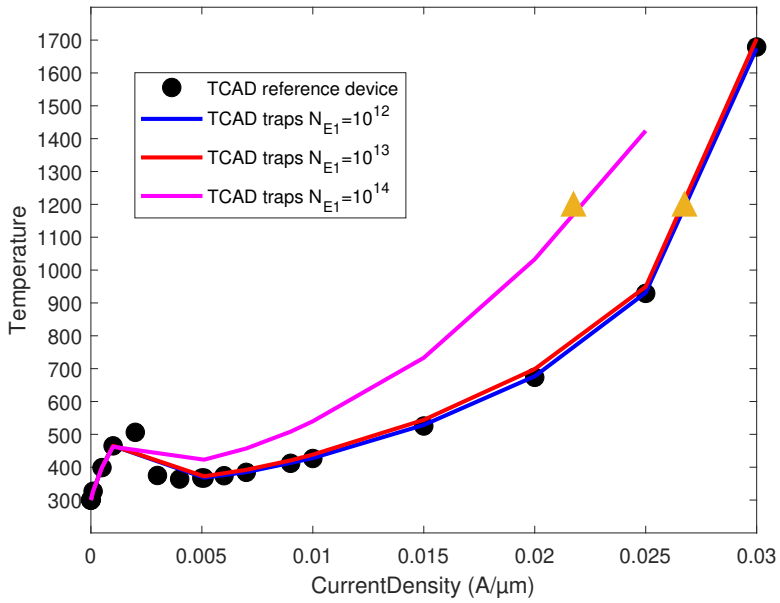


Figure 5.30: Maximum temperature (T_{max}) extracted from the TCAD 2D domain at the TLP final times vs. applied current density. The yellow triangles indicate the failure level considered to be around 1200K. The I_{t2} current corresponds to yellow triangles, indicating the failure level.

is investigated through the modification of both parasitic bipolar transistors.

Analysis of the NPN mechanism

The NPN parasitic transistor is connected in parallel to the LDMOS device, with emitter and base intrinsically shorted through the source-body contact. If impact-ionization generation takes place in the drift region at V_t, generated holes are drained away through the p-well body region eventually giving rise to a voltage drop between base and emitter. Thus, the doping level of the p-well region significantly modifies the triggering condition. More specifically, the lower the resistive path, the larger the required current to trigger the NPN leading to snap-back. The additional p-type region placed below the cathode can be refereed to the fact that this device is used as an LDMOS in normal operating conditions, hence, in that case, the buried body purpose is to reduce the effective resistance between the NPN base and emitter [58]. To this purpose, the PIL doping region was modified in order to change the gain of the bipolar transistor and test its impact on the holding voltage and thermal failure. Moreover, it is possible to reverse-evaluate the impact on the turning on mechanism at holding level until thermal failure.

Figure 5.31 shows the trends obtained from the simulations when considering an increase of the PIL doped region. The trigger voltage is quite unaffected by the PIL

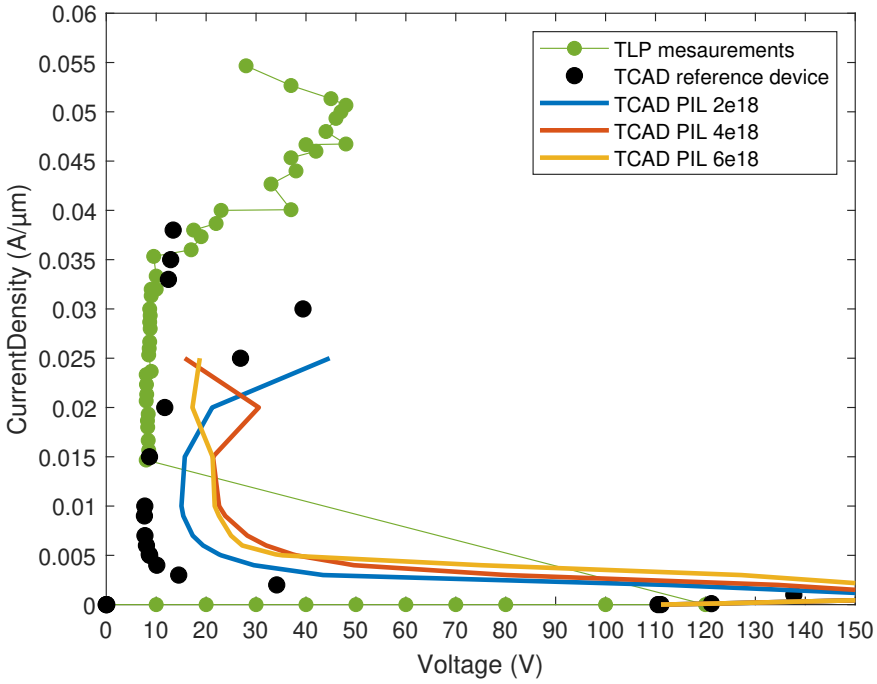


Figure 5.31: Voltage vs current at the variation of the p-type implant layer (PIL). As the holding voltage increases, the I_{t2} decreases accordingly as shown in the next figure.

modification since it can be related to the PNP and the reverse junction turn-on. Thus, V_t value varies from 153 V to 167 V. On the contrary the holding level experiences a significant increase of both current and voltage level. Indeed, the higher doping of the NPN active base region enables the NPN bipolar transistor to turn on faster than before and with different potential barriers between junction. Moreover, the increase of the doping reduces the resistive path built by the NPN base and carriers can flow to the anode side, laterally and vertically a lot easier than in the reference case. The difference in terms of current densities is shown in Figure 5.32.

In the high injection regime, which takes place at higher current pulses, the temperature hot spots become relevant. The critical point is identified in the n-type doped region at the cathode side, heating up faster as the PIL doping increases. A good trade-off is found by plotting the PIL doping of $2e+18cm^{-3}$, since the holding voltage as reached 15 V, while I_{t2} as not dropped dramatically as reported in Figure 5.33.

Analysis of the PNP mechanism

The analysis is carried on by investigating the other BJT that constructs the SCR structure: the PNP bipolar transistor. An in-depth study was done by P. Hongwei et al. in [43], adding a n-type implantation layer in the PNP base region. In their

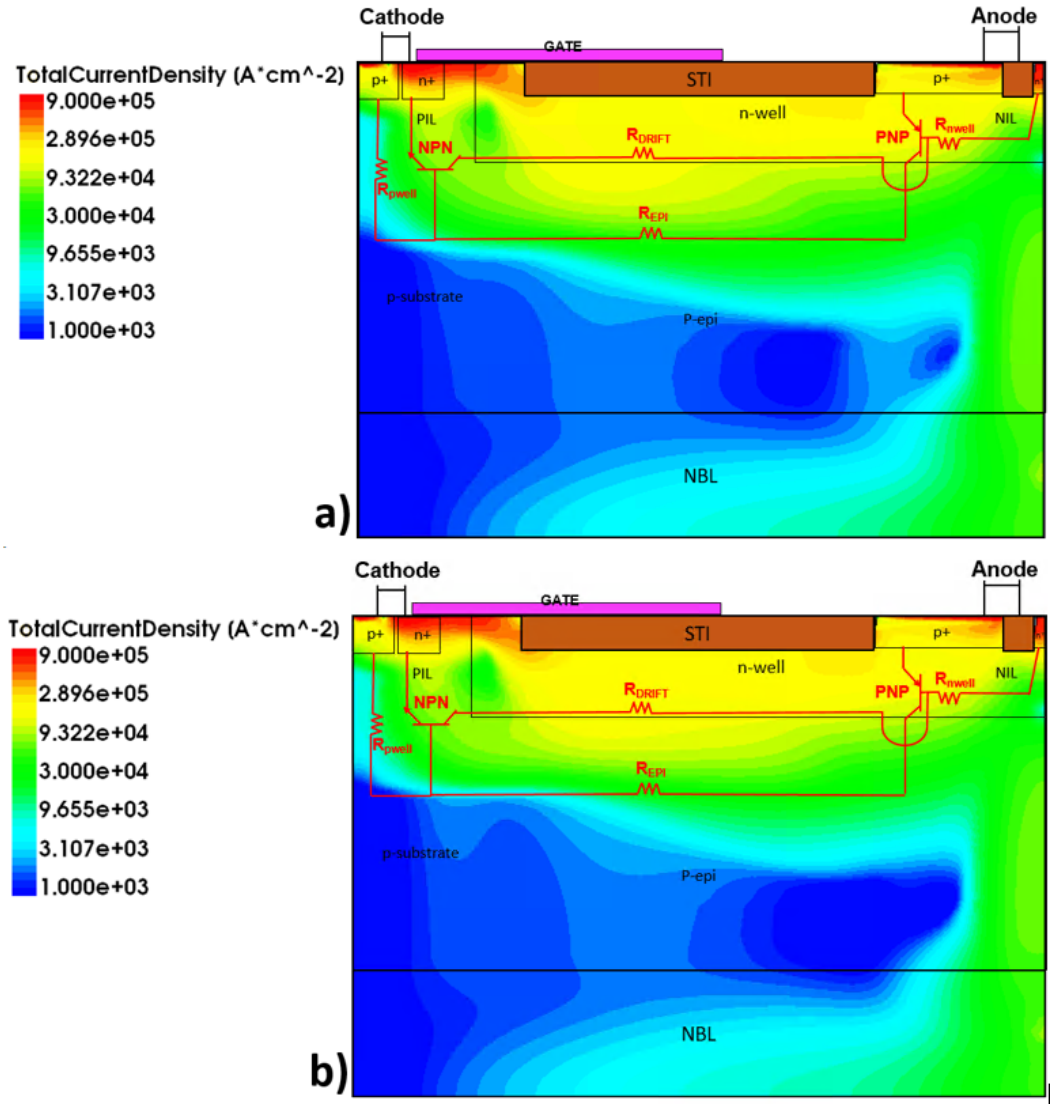


Figure 5.32: Current density distribution at holding voltage in the reference case a) and with a PIL doping variation case b)

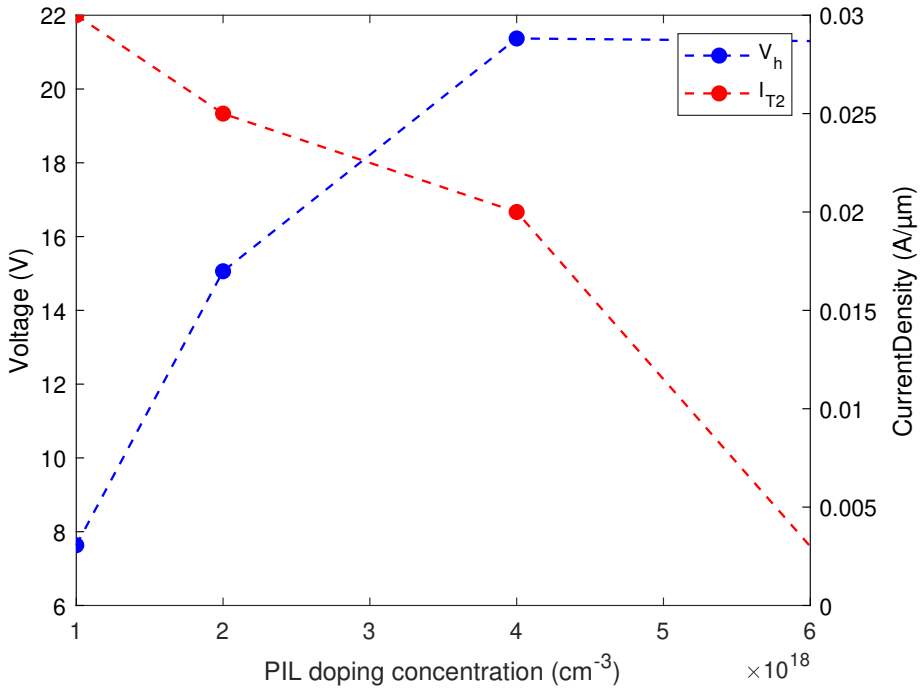


Figure 5.33: Holding voltage (left scale) and failure current density (right scale) vs. PIL doping variations.

work, it is explained how a n-type implantation layer in the PNP base is beneficial in terms of holding voltage improvement. In the device under study, the reference device already had this kind of improvement. Nevertheless, simulations of the modified doping were done in order to get a better understanding of the NPN gain and its correlation with the PIL modifications explained earlier. Results in Figure 5.34 show how its impact on the holding voltage is less significant with respect to the PIL case. However, the decrease in the resistive path of the PNP base, enables the current flow to reach the full turn on of the NPN transistor at lower current levels and shift the feedback between the two bipolar transistors path at higher potential drops. Indeed, the I-V curve show clearly the turning on of the two bipolar transistors after triggering. In addition, looking at Figure 5.35, the thermal contribution to the overall SOA can be discussed. The doping modification of the NIL region causes a carrier crowding close to p+ anode region. Consequently, avalanche multiplication occurs, causing an localized hot spot in the latter part of the structure. At that point, with an increase of the doping greater than $4 * 10^{18} cm^{-3}$, a degradation of the I_{t2} is detected as well. Thus, an optimal configuration is the one before the premature occurrence a thermal effect. After trigger, a significant amount of current is flowing superficially and, as the NIL doping increases, is kept at the surface due to the low resistive path. The doping region under study (NIL) initiates a controlled impact-ionization effect at this point, heating up the device. The occurrence of this event, the current spreads

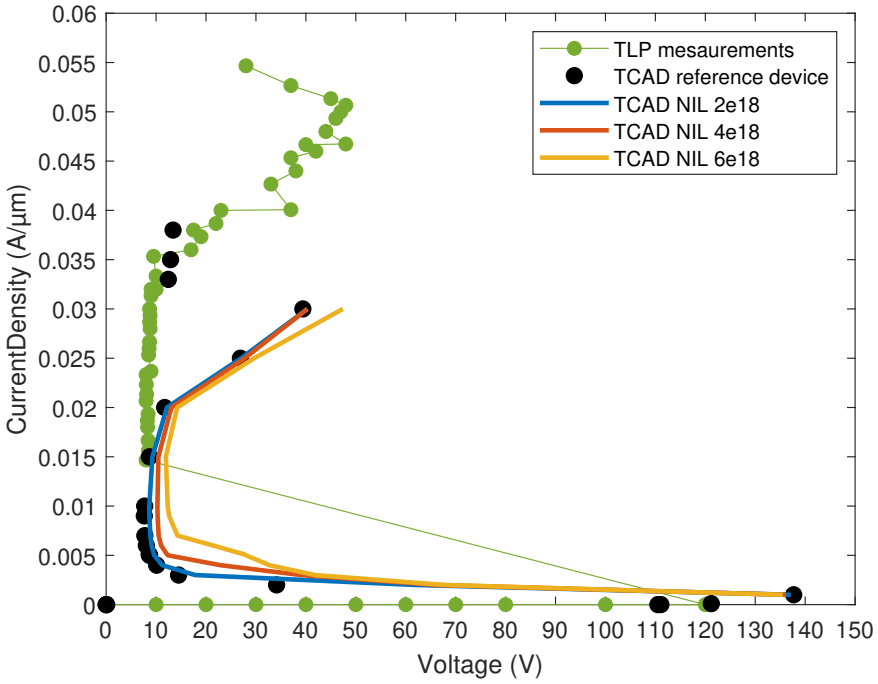


Figure 5.34: Voltage vs current at the variation of the n-type implant layer (NIL). Holding voltage is not influenced by this variation since the snap-back condition is dominated by the NPN transistor, not modified in this case.

deeper, thanks to the bipolar transistors turn on and their feedback loop, reaching the p-epi region. Thus, the electric field peak created by the avalanche generation of carriers is lowered and temperature contribution is relaxed. This it is identified as Kirk effect.

Figures 5.36 shows impact ionization and electrostatic potential after trigger, at holding voltage and after holding voltage, in order to analyze the evolution of the electrostatic potential distribution close to the the snap-back occurrence. In NIL edge, through the drift region from 9 to 11 μm, the phenomenon previously explained is evident: the impact ionization collapses as the electrostatic potential peak increases.

A similar analysis is thus carried out on the reference device by changing the peak value of both NIL and PIL Gaussian profiles. The best configuration for the device is the one that uses a trade-off between high holding voltage and a high enough failure current level. Through this investigation, it is possible to have some guidelines for future technology development, taking into account all possible physical mechanisms that occur in these p-n junction configurations. Several combinations of different PIL-NIL doping have been simulated. In this way, it was possible to observe first the increase of V_h and secondly if the I_{T2} level can be kept at a reasonable value, spreading the electric field superficially without a premature heat up of the

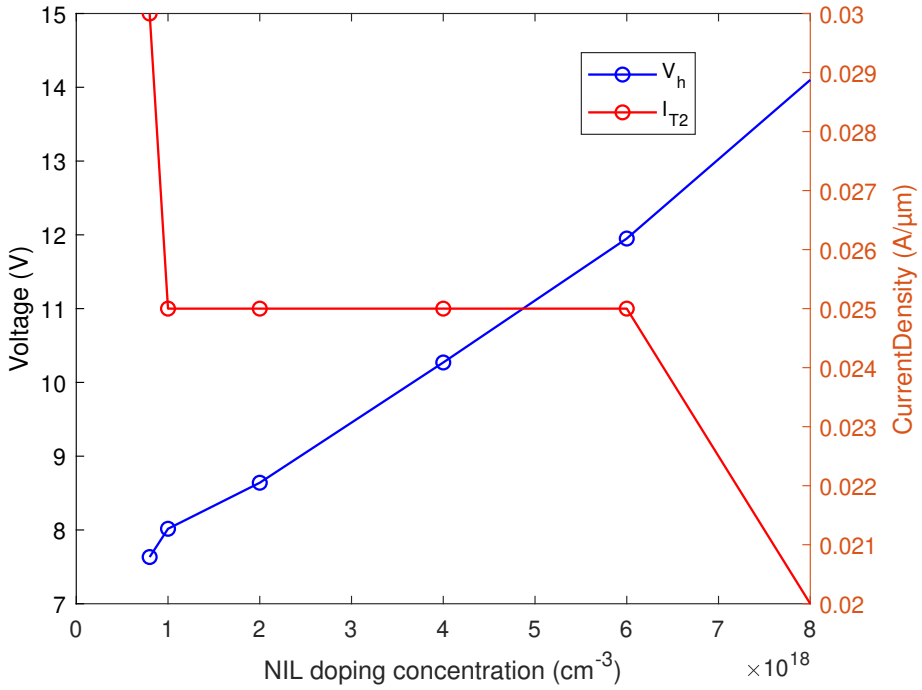


Figure 5.35: NIL doping variation vs. Voltage (holding) on the left an current density (I_{t2}) on the right. As the doping increases, the holding voltage increases accordingly until a breakthrough point were also I_{t2} goes down.

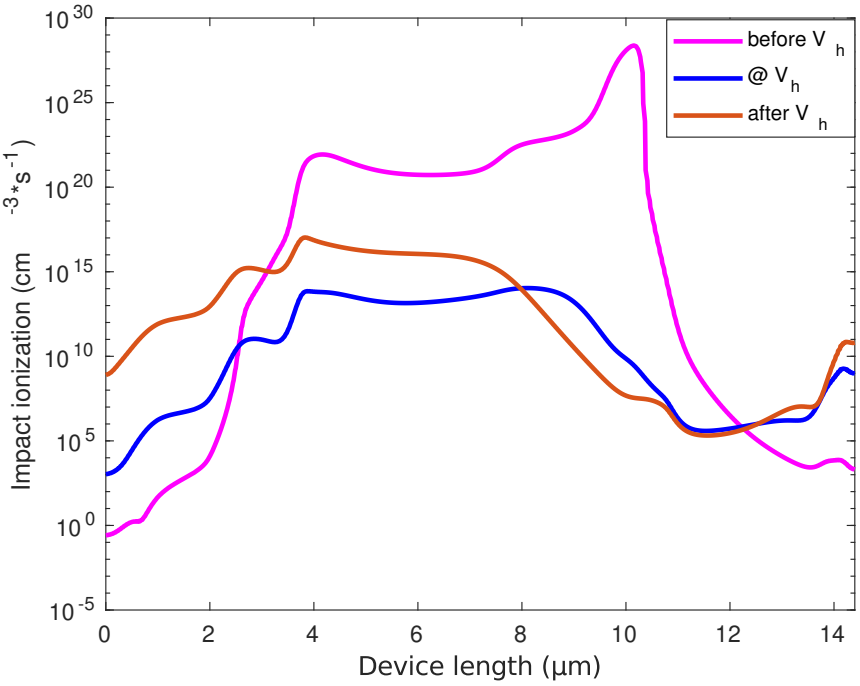


Figure 5.36: Device length vs. impact ionization coefficient in three cases: before holding (pink line), at holding voltage (blue line) and after holding (red line). The impact coefficient has a spike around the NIL edge, out forward by Kirk effect.

device. If not, that specific configuration is excluded since the overall performance of the device are deteriorated. The results of all these simulations are synthesized in Figure 5.37: the V_h and I_{t2} values extracted for different NIL doping peaks are reported showing an increase of Vh up to 10 V without significant changes in the I_{t2} ("Ref N_{PIL}" curves). V_h is mostly modified by the anticipated onset of the Kirk effect, as previously mentioned in [42]. The same effect leads to the increase of self-heating in the snap-back condition, which reaches $T_{max} > 1200$K at the largest N_{NIL}. The reported analyses clearly show that an optimization of both PIL and NIL doping concentrations can be carried out allowing to eventually gain large V_h with limited reduction of I_{t2}. In particular, V_h= 16 V can be obtained with N_{PIL}= N_{NIL}= $2x10^{18}cm^{-3}$ with a corresponding reduction of I_{t2} from 0.03 to 0.025 A/μm (Figure 5.37, "PIL+NIL" curves): a factor 2 in V_h is gained with a 17% reduction in I_{t2}.

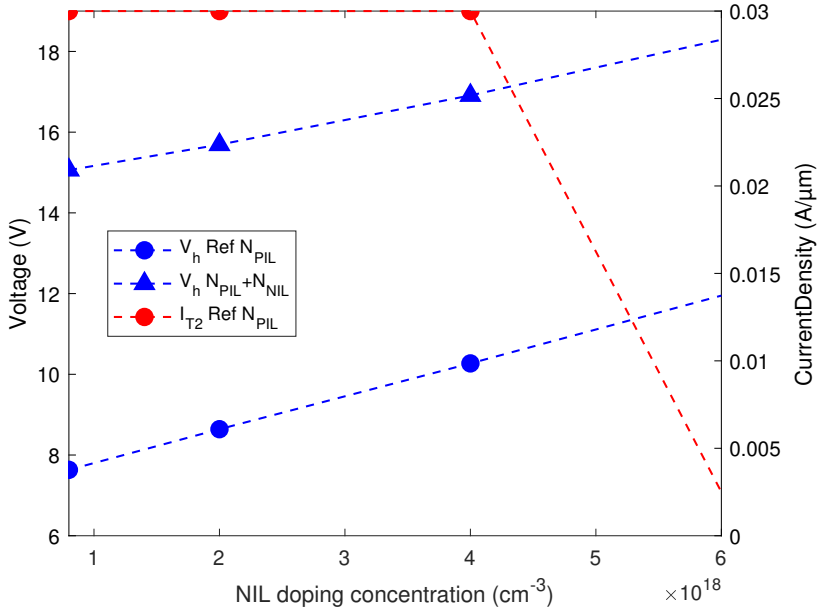


Figure 5.37: Holding voltage (left scale) and failure current density (right scale) vs. NIL doping variation for the reference PIL concentration (Ref. N_{PIL} and for the case with $N_{PIL} = 2 * 10^{18}cm^{-3}$("PIL+NIL)

In Figure 5.38 maximum temperature reached at each current level for the optimal case ("PIL+NIL") compared to the reference one and to the previous case with N_{PIL}= $2x10^{18}cm^{-3}$ is reported clearly showing the role played by the Kirk effect in increasing the temperature at lower current levels. The larger NIL doping slightly increases V_h without changing the high-injection regime.

Failure current optimization and power-to-failure study.

By analyzing the SCR-LDMOS at the failure conditions, all cases show that it is mostly related to the n+ region, identified as the emitter region of the NPN (Fig. 5.3). Simulations have been carried out by extending the length of the n+ well in

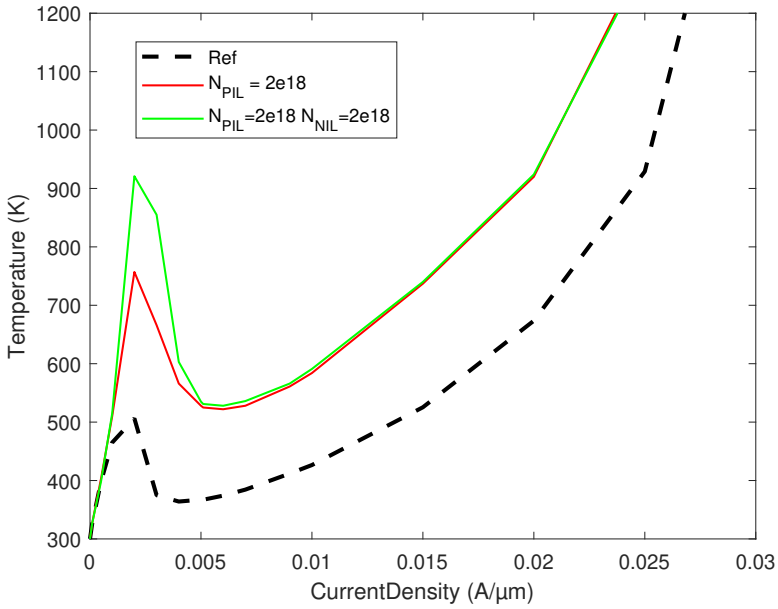


Figure 5.38: Maximum temperature as a function of the current level for the reference device ("Ref"), the case with optimized PIL doping at $2x10^{18}cm^{-3}$ ("PIL") and the case with both NIL and PIl optimizations ("PIL+NIL").

the "NIL+PIL" case, leading to a significant reduction of the current crowding. This opens to the recovery of I_{t2} from 0.025 to 0.03 A/μm, as shown in the following. Moreover, V_h and I_{t2} should be also monitored versus stress time, as modifications of the gain of both BJTs directly influence the corresponding transit times. To this purpose, they have been extracted as a function of stress time from very-fast TLP (1ns) up to long duration EOS regime (see Figure 5.39). V_h shows very large values in the short-time regime due to the transfer time of the NPN BJT: at 1 ns, the BJT is only partially turned on, leading to a larger snap-back current given by a limited current gain. I_{t2} shows a quite linear reduction with time mostly due to self-heating effects. The optimized "PIL+NIL" case leads to a slightly lower I_{t2} on the full range. The emitter extension increases the NPN BJT active area, increasing the I_{t2} with respect to the "PIL+NIL" case. It also enhances the emitter injection efficiency, leading to a slight reduction of V_h.

The analysis of I_{t2} as a function of stress time gives interesting indications on the power-to-failure. In Fig. 5.40, the power-to-failure curve of the reference device is compared with the "PIL+NIL" cases, both with and without the emitter extension. Quite similar results are obtained for the three different cases, proving that our approach has not influenced the overall performance of the structure besides the holding voltage value. The curves obtained with TCAD nicely compare with other experimental curves on different technologies, showing that SCR and PNP configurations

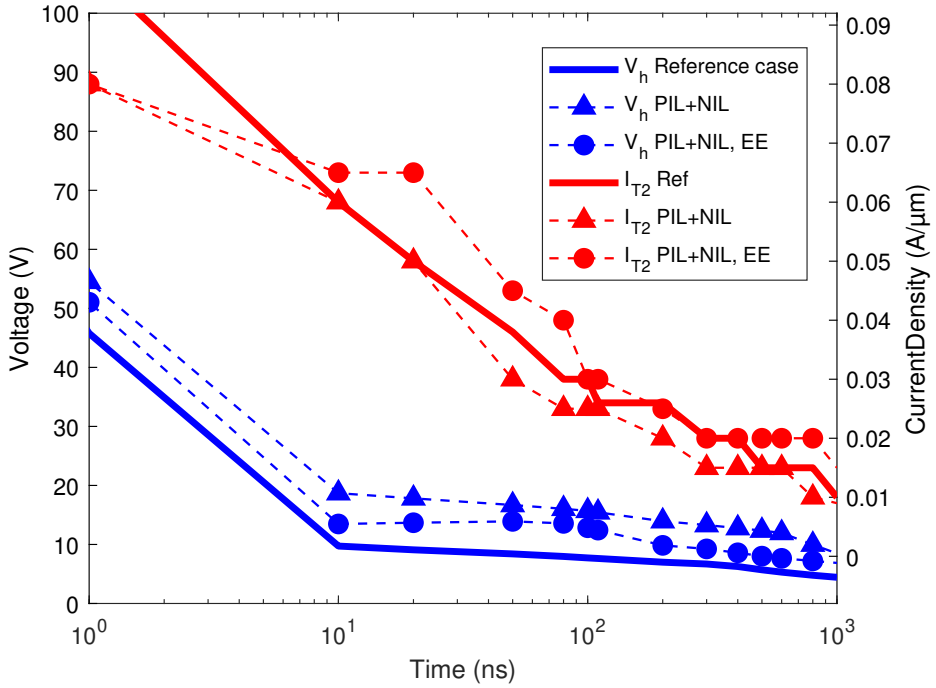


Figure 5.39: Holding voltage (left scale) and failure current (right scale) vs. time to failure variation for the reference case (solid line), the "PIL+NIL" case, both with and without the emitter extension ("EE")

lead to larger power to failure.

In Table 5.2, the FOM values of the investigated SCR-LDMOS reference device and of the modified cases are compared with the recently published solutions in [44]. The “PIL+NIL” configuration shows an improvement of about 18% with respect to the reference one, reaching a FOM similar to the best case under irradiated conditions. Thus, the proposed study gives indications for a cost-effective optimization, featuring a good solution for the V_h improvement up to 16V without impacting on the device ESD robustness.

5.6 Concluding remarks

A deep physical understanding of the conventional SCR-LDMOS structure is given, providing possible correlations between holding voltage and failure current conditions. TCAD simulations of layout and doping modifications were tested with the aim of increasing the holding voltage level without biasing the thermal runaway condition. Simulations on the reference layout condition show the limitations given by

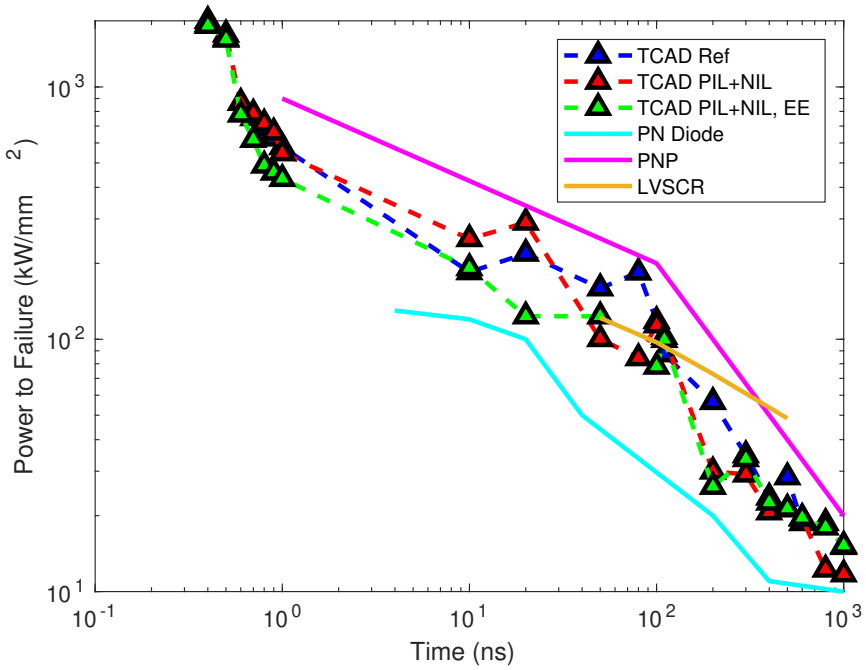


Figure 5.40: Power-to-failure time to failure TCAD results in the three cases of 5.39 compared with experimental results of Low Voltage SCR [59] and PNP data from [60] and PN diode [46]

Device Topology	$V_h(V)$	$I_{T2}(mA)$	$S\,(\mu m^2)$	FOM
HHSCR1	11.79	1.6	890	21.19
MLSCR	4.32	2.33	550	18.3
HVSCR	16.9	8084	9313	16.04
Ref Device	7.72	3	140	16.54
Ref Traps 1e14	16	2.5	140	25.14
NIL+PIL	15.69	2.5	140	28.02
NIL+PIL, EE	12.83	3	144	26.73

Table 5.2: Holding voltage, second-breakdown failure current and area of different structures proposed in [44], compared with the irradiated SCR-LDMOS devices in [52] and this work. The $FOM = V_h x I_{t2}/S$ is calculated consistently with the reported data and compared with the device under investigation for "ref", "PIL+NIL" and "PIL+NIL, EE" cases. To this purpose, TCAD simulation data are reported with a normalized width fixed to 1 μm.

the lateral and vertical path of the current and their effect on the BJT triggering conditions. The optimization of the snap-back condition has been proposed by means of simple doping profile modifications. Finally, the figure of merit of the SCR-LDMOS device has been compared with other technologies, highlighting a significant improvement on the overall performances.

Chapter 6

Conclusions

The analysis submitted in this thesis is focused on TCAD simulations applied to the study of several physical mechanisms dealing with ESD events. Moreover, a methodology to study ESD clamps is provided through the TCAD to reduce the qualification time of such protection devices. Two different sets of devices from the same BiCMOS technology have been investigated.

The first one under consideration consists of two low voltage bipolar transistor internally stacked, to realize a high voltage ESD clamp. The vf-TLP time domain is poorly investigated in literature even though new technology applications require protection for stresses in the sub-ns regime. Failure conditions in the adiabatic model are the main concern of this work since the ESD architecture shows a non-homogeneous behavior, not matching the theoretical predictions. The root cause of the thermal failure has been identified as a simple layout modification which has been proposed to improve the understanding the cell behavior. Thus, the ESD clamp under study has been simulated and deeply investigated to fully understand the anomalous behavior of the coupled NPNs and to devise improvements for its unexpected limitations in the latter regime.

The second device used as case study in this thesis is a conventional SCR-LDMOS structure. This ESD protection device embedded in laterally diffused MOSFET is very attractive due to its high performance in terms of robustness and full compatibility with CMOS technology. On the other hand, low holding voltage, given by the intrinsic feedback of two bipolar transistors, results in an increased latch-up risk. A deep physical understanding is given, providing possible correlations between holding voltage and failure current conditions. TCAD simulations of layout and doping modifications have been proposed with the aim of increasing the holding voltage without biasing the thermal runaway. Simulations on the reference layout show the limitations given by the lateral and vertical paths of the current and their effect on the BJT triggering conditions. Thus, several solutions have been investigated and understood through the device physical behavior.

A first possible solution is given by assuming the electron or ion irradiation as

a key tool for off-line optimization of the device parameters. The results lead to increased holding voltage without affecting the thermal conditions. However, as the proposed solution might not be suitable for all products due the inefficiency in terms of cost, other studies have been performed in the direction of simpler layout modifications. The optimization of the snap-back condition has been proposed by means of doping profile splits, instead. A trade-off of both n-type and p-type implant layers has been studied, showing their impact on the conductance modulation and high injection regimes. A trade-off of the two implant contribution is studied and implemented through TCAD simulations, leading to future possible paths of research. The figure of merit of the SCR-LDMOS device has been finally compared with other technologies, highlighting the significant improvements on the overall performance of the proposed optimizations.

Bibliography

[1] G. Boselli, "Esd design challenges in state-of-the-art analog technologies," *Microelectronics Reliability*, vol. 52, no. 9, pp. 1769–1775, 2012, SPECIAL ISSUE 23rd EUROPEAN SYMPOSIUM ON THE RELIABILITY OF ELECTRON DEVICES, FAILURE PHYSICS AND ANALYSIS, ISSN: 0026-2714. DOI: https://doi.org/10.1016/j.microrel.2012.06.141. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0026271412003381.

[2] M.-D. Ker, C.-Y. Lin, and Y.-W. Hsiao, "Overview on esd protection designs of low-parasitic capacitance for rf ics in cmos technologies," *IEEE Transactions on Device and Materials Reliability*, vol. 11, no. 2, pp. 207–218, 2011. DOI: 10.1109/TDMR.2011.2106129.

[3] C. Duvvury and A. Amerasekera, "State-of-the-art issues for technology and circuit design of esd protection in cmos ics," *Semiconductor Science and Technology*, vol. 11, no. 6, p. 833, 1996. DOI: 10.1088/0268-1242/11/6/001. [Online]. Available: https://dx.doi.org/10.1088/0268-1242/11/6/001.

[4] E. Amerasekera and C. Duvvury, *ESD in Silicon Integrated Circuits* (Design And Measurement in Electronic Engineering). Wiley, 1995, ISBN: 9780471954811. [Online]. Available: https://books.google.it/books?id=dgNTAAAAMAAJ.

[5] *TCAD - Device Simulation kernel description*, https://www.synopsys.com/silicon/tcad/device-simulation.html, Accessed: 2023-08-02.

[6] *Research Gate, howpublished = https:https://www.researchgate.net, note = Accessed: 2023-08-02.*

[7] Synopsys TCAD, *Synopsys*, version 0.20.2, Feb. 19, 2010. [Online]. Available: https://www.synopsys.com/.

[8] D. C. Wunsch and R. R. Bell, "Determination of threshold failure levels of semiconductor diodes and transistors due to pulse voltages," *IEEE Transactions on Nuclear Science*, vol. 15, no. 6, pp. 244–259, 1968. DOI: 10.1109/TNS.1968.4325054.

[9] V. Dwyer, A. Franklin, and D. Campbell, "Electrostatic discharge thermal failure in semiconductor devices," *IEEE Transactions on Electron Devices*, vol. 37, no. 11, pp. 2381–2387, 1990. DOI: 10.1109/16.62296.

[10] M. Rudan *et al.*, *Physics of semiconductor devices*. Springer, 2015.

[11] G. Baccarani, M. R. Wordeman, and R. H. Dennard, "Generalized scaling theory and its application to a 1/4 micrometer mosfet design," *IEEE Transactions on Electron Devices*, vol. 31, no. 4, pp. 452–462, 1984.

[12] S. M. Sze, Y. Li, and K. K. Ng, *Physics of semiconductor devices*. John wiley & sons, 2021.

[13] T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr, "A review of hydrodynamic and energy-transport models for semiconductor device simulation," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 251–274, 2003. DOI: 10.1109/JPROC.2002.808150.

[14] D. R. Poirier and G. H. Geiger, "Conduction of heat in solids," in *Transport Phenomena in Materials Processing*. Cham: Springer International Publishing, 2016, pp. 281–327, ISBN: 978-3-319-48090-9. DOI: 10.1007/978-3-319-48090-9_9. [Online]. Available: https://doi.org/10.1007/978-3-319-48090-9_9.

[15] T. Goudon, V. Miljanović, and C. Schmeiser, "On the shockley–read–hall model: Generation-recombination in semiconductors," *SIAM Journal on Applied Mathematics*, vol. 67, no. 4, pp. 1183–1201, 2007.

[16] *TDK products kernel description*, https://product.tdk.com/en/contact/faq/capacitors-0093.html, Accessed: 2023-08-02.

[17] *Onsemi Website, howpublished = https://www.onsemi.com/design/resources/technical-documentation?notFound=AND9006-D.PDF*, Accessed: 2024-03-29.

[18] C. Duvvury and A. Amerasekera, "Esd issues for advanced cmos technologies," *Microelectronics Reliability*, vol. 36, no. 7, pp. 907–924, 1996, Reliability Physics of Advanced Electron Devices, ISSN: 0026-2714. DOI: https://doi.org/10.1016/0026-2714(96)00024-8. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0026271496000248.

[19] A. Delmas, A. Gendron, M. Bafleur, N. Nolhier, and C. Gill, "Transient voltage overshoots of high voltage esd protections based on bipolar transistors in smart power technology," in *2010 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, 2010, pp. 253–256. DOI: 10.1109/BIPOL.2010.5667967.

[20] Y. Cao, U. Glaser, J. Willemen, *et al.*, "Esd simulation with wunsch-bell based behavior modeling methodology," in *EOS/ESD Symposium Proceedings*, IEEE, 2011, pp. 1–10.

[21] P. Magnone, A. Petucco, N. Thevenet, and H. Abedini, "Simplified on-line monitoring system of mosfet on-resistance based on a semi-empirical model," in *2019 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2019, pp. 2746–2750. DOI: 10.1109/APEC.2019.8721843.

[22] G. Meneghesso, A. Tazzoli, F. A. Marino, M. Cordoni, and P. Colombo, "Development of a new high holding voltage scr-based esd protection structure," in *2008 IEEE International Reliability Physics Symposium*, 2008, pp. 3–8. DOI: 10.1109/RELPHY.2008.4558856.

[23] G. Notermans, F. Kuper, and J.-M. Luchies, "Using an scr as esd protection without latch-up danger," *Microelectronics Reliability*, vol. 37, no. 10, pp. 1457–1460, 1997, Reliability of Electron Devices, Failure Physics and Analysis, ISSN: 0026-2714. DOI: https://doi.org/10.1016/S0026-2714(97)00086-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0026271497000863.

[24] K.-I. Do, S.-H. Jin, B.-S. Lee, and Y.-S. Koo, "4h-sic-based esd protection design with optimization of segmented ligbt for high-voltage applications," *IEEE Journal of the Electron Devices Society*, vol. 9, pp. 1017–1023, 2021. DOI: 10.1109/JEDS.2021.3121212.

[25] S. Pendharkar, R. Teggatz, J. Devore, J. Carpenter, T. Efland, and C.-Y. Tsai, "Scr-ldmos. a novel ldmos device with esd robustness," in *12th International Symposium on Power Semiconductor Devices ICs. Proceedings (Cat. No.00CH37094)*, 2000, pp. 341–344. DOI: 10.1109/ISPSD.2000.856839.

[26] J. Goodrich, "Three chips in one: The history of the bcd integrated circuit," *IEEE Spectrum*, 11 Jun 2021. [Online]. Available: https://spectrum.ieee.org/three-chips-in-one-the-history-of-the-bcd-integrated-circuit.

[27] T. Imoto, K. Mawatari, K. Wakiyama, *et al.*, "A novel esd protection device structure for hv-mos ics," in *2009 IEEE International Reliability Physics Symposium*, 2009, pp. 663–668. DOI: 10.1109/IRPS.2009.5173326.

[28] B. K. Boksteen, A. Ferrara, A. Heringa, P. G. Steeneken, G. E. J. Koops, and R. J. E. Hueting, "Design optimization of field-plate assisted resurf devices," in *2013 25th International Symposium on Power Semiconductor Devices IC's (ISPSD)*, 2013, pp. 237–240. DOI: 10.1109/ISPSD.2013.6694460.

[29] S. Reggiani, G. Barone, E. Gnani, *et al.*, "Characterization and modeling of electrical stress degradation in sti-based integrated power devices," *Solid-State Electronics*, vol. 102, pp. 25–41, 2014, Selected papers from ESSDERC 2013, ISSN: 0038-1101. DOI: https://doi.org/10.1016/j.sse.2014.06.008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038110114001415.

[30] S. Poli, S. Reggiani, G. Baccarani, *et al.*, "Full understanding of hot-carrier-induced degradation in sti-based ldmos transistors in the impact-ionization operating regime," in *2011 IEEE 23rd International Symposium on Power Semiconductor Devices and ICs*, 2011, pp. 152–155. DOI: 10.1109/ISPSD.2011.5890813.

[31] A. N. Tallarico, S. Reggiani, R. Depetro, *et al.*, "Hot-carrier degradation in power ldmos: Selective locos- versus sti-based architecture," *IEEE Journal of the Electron Devices Society*, vol. 6, pp. 219–226, 2018. DOI: 10.1109/JEDS.2018.2792539.

[32] R. Hueting and R. van der Toorn, "Analysis of the kirk effect in silicon-based bipolar transistors with a nonuniform collector profile," *IEEE Transactions on Electron Devices*, vol. 52, no. 11, pp. 2489–2495, 2005. DOI: 10.1109/TED.2005.857176.

[33] *SPICE -Simulation Program with Integrated Circuit Emphasis kernel description*, https://https://www.analog.com/en/design-center/design-tools-and-calculators/ltspice-simulator.html, Accessed: 2023-22-02.

[34] G. Croce, A. Andreini, P. Galbiati, and C. Diazzi, "Bcd process technologies," in *Springer Handbook of Semiconductor Devices*, M. Rudan, R. Brunetti, and S. Reggiani, Eds. Cham: Springer International Publishing, 2023, pp. 67–116, ISBN: 978-3-030-79827-7. DOI: 10.1007/978-3-030-79827-7_3. [Online]. Available: https://doi.org/10.1007/978-3-030-79827-7_3.

[35] J.-H. Lee, S. Chen, Y. Tsai, *et al.*, "The influence of nbl layout and locos space on component esd and system level esd for hv-ldmos," in *Proceedings of the 19th International Symposium on Power Semiconductor Devices and IC's*, 2007, pp. 173–176. DOI: 10.1109/ISPSD.2007.4294960.

[36] S Lee and R Bashir, "Modeling and characterization of deep trench isolation structures," *Microelectronics Journal*, vol. 32, no. 4, pp. 295–300, 2001, ISSN: 0026-2692. DOI: https://doi.org/10.1016/S0026-2692(00)00148-8. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0026269200001488.

[37] L. Zunarelli, S. Reggiani, E. Gnani, R. Sankaralingam, M. Dissegna, and G. Boselli, "Tcad investigation of power-to-failure evaluation for ultrafast events in bjt-based esd protection cells," in *2022 IEEE International Reliability Physics Symposium (IRPS)*, 2022, pp. 6C.2–1–6C.2–6. DOI: 10.1109/IRPS48227.2022.9764432.

[38] S. Reggiani, M. Rudan, E. Gnani, and G. Baccarani, "Investigation about the high-temperature impact-ionization coefficient in silicon," in *Proceedings of the 30th European Solid-State Circuits Conference (IEEE Cat. No.04EX850)*, 2004, pp. 245–248. DOI: 10.1109/ESSDER.2004.1356535.

[39] T. Smedes and N. Guitard, "Harmful voltage overshoots due to turn-on behaviour of esd protections during fast transients," in *2007 29th Electrical Overstress/Electrostatic Discharge Symposium (EOS/ESD)*, 2007, 6A.4–1–6A.4–9. DOI: 10.1109/EOSESD.2007.4401774.

[40] C. Salamero, N. Nolhier, A. Gendron, M. Bafleur, P. Besse, and M. Zecri, "Tcad methodology for esd robustness prediction of smart power esd devices," *IEEE Transactions on Device and Materials Reliability*, vol. 6, no. 3, pp. 399–407, 2006. DOI: 10.1109/TDMR.2006.881483.

[41] H. Karaca, S. Holland, H.-M. Ritter, V. Kumar, G. Notermans, and D. Pogany, "3-d tcad methodology for simulating double-hysteresis filamentary i–v behavior and holding current in esd protection scrs," *IEEE Transactions on Electron Devices*, vol. 68, no. 9, pp. 4214–4222, 2021. DOI: 10.1109/TED.2021.3100301.

[42] M. Shrivastava and H. Gossner, "A review on the esd robustness of drain-extended mos devices," *IEEE Transactions on Device and Materials Reliability*, vol. 12, no. 4, pp. 615–625, 2012. DOI: 10.1109/TDMR.2012.2220358.

[43] S.-Y. Liu, W.-F. Sun, H.-W. Pan, H. Wang, and Q.-S. Qian, "A novel latch-up free scr-ldmos for power-rail esd clamp in half-bridge driver ic," in *2012 IEEE 11th International Conference on Solid-State and Integrated Circuit Technology*, 2012, pp. 1–3. DOI: 10.1109/ICSICT.2012.6466723.

[44] W. Song, R. Chen, Z. Tong, *et al.*, "Robust silicon-controlled rectifier with high-holding voltage for on-chip electrostatic protection," *IEEE Transactions on Electron Devices*, vol. 69, no. 2, pp. 696–703, 2022. DOI: 10.1109/TED.2021.3131536.

[45] S. Reggiani, E. Gnani, M. Rudan, *et al.*, "Measurement and modeling of the electron impact-ionization coefficient in silicon up to very high temperatures," *IEEE transactions on electron devices*, vol. 52, no. 10, pp. 2290–2299, 2005.

[46] Y. Cao, U. Glaser, S. Frei, and M. Stecher, "A failure levels study of non-snapback esd devices for automotive applications," in *2010 IEEE International Reliability Physics Symposium*, 2010, pp. 458–465. DOI: 10.1109/IRPS.2010.5488786.

[47] F. Ma, B. Zhang, Y. Han, *et al.*, "High holding voltage scr-ldmos stacking structure with ring-resistance-triggered technique," *IEEE Electron Device Letters*, vol. 34, no. 9, pp. 1178–1180, 2013. DOI: 10.1109/LED.2013.2272591.

[48] C. R. Zhan, P. Besse, J.-P. Laine, and A. Salles, "High-performance bi-directional scr developed on a 0.13um soi-based smart power technology for automotive applications," in *2017 39th Electrical Overstress/Electrostatic Discharge Symposium (EOS/ESD)*, 2017, pp. 1–7. DOI: 10.23919/EOSESD.2017.8073448.

[49] X. Bao and X. Jin, "An improved ldmos-scr design for high-voltage esd," in *2022 International EOS/ESD Symposium on Design and System (IEDS)*, 2022, pp. 1–4.

[50] Z. Chen, W. Lu, M. Wu, *et al.*, "A compact ldmos ddscr for hv esd protections with high robustness and reliability," *Solid-State Electronics*, vol. 161, p. 107 640, 2019, ISSN: 0038-1101. DOI: https://doi.org/10.1016/j.sse.2019.107640. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038110119302254.

[51] J.-I. Won, H.-D. Lee, K.-Y. Lee, K.-D. Kim, and Y.-S. Koo, "Analysis of the electrical characteristics of novel esd protection device with high holding voltage under various temperatures," in *TENCON 2009 - 2009 IEEE Region 10 Conference*, 2009, pp. 1–4. DOI: 10.1109/TENCON.2009.5395895.

[52] L. Zunarelli, L. Balestra, S. Reggiani, R. Sankaralingam, M. Dissegna, and G. Boselli, "Tcad study of the holding-voltage modulation in irradiated scr-ldmos for hv esd protection," in *2023 IEEE International Reliability Physics Symposium (IRPS)*, 2023, pp. 1–6. DOI: 10.1109/IRPS48203.2023.10118271.

[53] J. Vobecky, "Fast recovery diodes for high-current high- voltage insulated gate bipolar transistors," *IEEE Electron Device Letters*, vol. 43, no. 8, pp. 1311–1314, 2022. DOI: 10.1109/LED.2022.3187159.

[54] J. Vobeckỳ, H.-J. Schulze, P. Streit, *et al.*, "Silicon thyristors for ultrahigh power (gw) applications," *IEEE Transactions on Electron Devices*, vol. 64, no. 3, pp. 760–768, 2017.

[55] S. Brotherton and P Bradley, "Defect production and lifetime control in electron and γ-irradiated silicon," *Journal of Applied Physics*, vol. 53, no. 8, pp. 5720–5732, 1982.

[56] M. Bellini and J. Vobeckỳ, "Tcad simulations of irradiated power diodes over a wide temperature range," in *2011 International Conference on Simulation of Semiconductor Processes and Devices*, IEEE, 2011, pp. 183–186.

[57] P. Hazdra, J. Vobecký, H. Dorschner, and K. Brand, "Axial lifetime control in silicon power diodes by irradiation with protons, alphas, low- and high-energy electrons," *Microelectronics Journal*, vol. 35, no. 3, pp. 249–257, 2004, 6Th International Seminar on Power Semiconductors, ISSN: 0026-2692. DOI: https://doi.org/10.1016/S0026-2692(03)00194-0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0026269203001940.

[58] P. Hower, J. Lin, S. Pendharkar, *et al.*, "A rugged ldmos for lbc5 technology," in *Proceedings. ISPSD '05. The 17th International Symposium on Power Semiconductor Devices and ICs, 2005.*, 2005, pp. 327–330. DOI: 10.1109/ISPSD.2005.1488017.

[59] G. Boselli, A. Salman, J. Brodsky, and H. Kunz, "The relevance of long-duration tlp stress on system level esd design," in *Electrical Overstress/Electrostatic Discharge Symposium Proceedings 2010*, 2010, pp. 1–10.

[60] L. Cerati, L. Di Biccari, A. Andreini, C. Castiglione, and D. Blanc, "Power-to-failure investigation for pnp-based esd protections: From ns to ms," in *2013 35th Electrical Overstress/Electrostatic Discharge Symposium*, 2013, pp. 1–10.