UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MICROELETRÔNICA

ROBERT DETTENBORN

Parametric Analysis and Optimization of MOSFET Macromodels for ESD Circuit Simulation

Dissertation submitted in partial fulfillment
of the requirements for the degree of
Master of Microeletronics

Advisor: Prof. Dr. Marcelo de Oliveira Johann

Porto Alegre
December 2018

CIP — CATALOGING-IN-PUBLICATION

Dettenborn, Robert

Parametric Analysis and Optimization
of MOSFET Macromodels for ESD Circuit Simulation / Robert Dettenborn. – Porto Alegre: PGMICRO da UFRGS, 2018.

100 f.: il.

Dissertation (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Microeletrônica, Porto Alegre, BR–RS, 2018. Advisor: Marcelo de Oliveira Johann.

1. ESD. 2. Circuit simulation. 3. MOSFET macromodels. 4. Parametric analysis. 5. Optimization. I. Johann, Marcelo de Oliveira. II. Título.

“The best way to predict the future is to invent it.”

— Alan Kay

ACKNOWLEDGMENTS

I would like to thank my advisor Prof. Dr. Marcelo de Oliveira Johann for making this work possible, Eng. Everton Luis Granado Ghignatti Junior for the many advices, discussions and technical support, and Dr. Artur Vicente Pfeifer Coelho as well as Dr. Laurent Courcelle for providing the necessary resources at CEITEC S.A.

I dedicate this dissertation to my family and to the memory of my grandmother.

ABSTRACT

Electrostatic discharge (ESD) is a major reliability concern in semiconductor industry. An ESD event may degrade or destroy an integrated circuit (IC), impacting on production yields, manufacturing costs, product quality, product reliability and company profitability. Additionally, the breakdown voltages and failure currents of semiconductor devices are becoming lower with the technology scaling, placing severe constraints on robust IC design. As a result, effective on-chip ESD protection without compromising area and performance requirements is becoming a challenge in deeply-scaled technologies. In this context, circuit simulation can provide the required assistance in on-chip protection design, including robustness analysis of the circuits and performance prediction prior to silicon. However, modeling the MOSFET operation under ESD conditions for circuit simulation is still a challenging issue. The large current and voltage characteristics are typically not well covered by most standard SPICE compact models. To overcome such limitation, a practical modeling approach is highly desired. This work presents a study of MOSFET macromodels for ESD circuit simulation. First, it gives an overview of the device operation under ESD conditions. Then, it presents the evolution of MOSFET macromodels for ESD circuit simulation. Finally, a novel macromodel development approach based on parametric analysis and optimization is introduced.

Keywords: ESD. Circuit simulation. MOSFET macromodels. Parametric analysis. Optimization.

Análise Paramétrica e Otimização de Macromodelos para Simulação Elétrica de Descarga Eletrostática em MOSFETs

RESUMO

Descarga eletrostática (ESD) é uma importante preocupação em relação à confiabilidade de produtos na indústria de semicondutores. Uma descarga eletrostática pode degradar ou destruir circuitos integrados, afetando o rendimento da produção, os custos de fabricação, a qualidade e a confiabilidade dos produtos, bem como a lucratividade da empresa. Além disso, a redução das dimensões dos dispositivos com o avanço da tecnologia resulta em menores tensões de ruptura e correntes de falha, impondo severas restrições no projeto de circuitos integrados robustos. Com isso, o projeto de proteções intra-chip eficazes contra ESD, sem comprometer os requisitos de área e desempenho, está se tornando um desafio cada vez maior em tecnologias avançadas. Neste contexto, a simulação elétrica pode prover a assistência necessária ao projeto de proteções intra-chip, incluindo a análise de robustez dos circuitos e a estimativa de desempenho antes da fabricação em silício. No entanto, modelar a operação de MOSFETs sob condições de ESD para simulação elétrica ainda é um problema desafiador. O comportamento destes dispositivos sob altas correntes e tensões normalmente não é bem descrito pela maioria dos modelos compactos SPICE convencionais. Para contornar essa limitação, uma abordagem prática de modelagem é altamente desejada. Este trabalho apresenta um estudo acerca de macromodelos para simulação elétrica de ESD em MOSFETs. Primeiro, uma visão geral da operação de um MOSFET sob condições de ESD é fornecida. Em seguida, a evolução dos macromodelos para simulação elétrica de ESD em MOSFETs é apresentada. Por fim, uma nova abordagem para o desenvolvimento de macromodelos baseada em análise paramétrica e otimização é introduzida.

Palavras-chave: Descarga eletrostática, simulação elétrica, macromodelos de MOSFETs, análise paramétrica, otimização.

LIST OF ABBREVIATIONS AND ACRONYMS

ADE	Cadence Virtuoso Analog Design Environment (EDA tool)
BGA	Ball grid array
BJT	Bipolar junction transistor
BSIM	Berkeley short-channel insulated-gate field-effect transistor model
BVOX	Gate oxide breakdown voltage
CAD	Computer-aided design
CDM	Charged device model
CMOS	Complementary metal-oxide semiconductor
DC	Direct current (circuit analysis domain)
DIP	Dual in-line package
DNO	MunEDA WiCkeD Deterministic Nominal Optimization (EDA tool)
DUT	Device under test
EDA	Electronic design automation
EKV	Enz-Krummenacher-Vittoz (MOSFET compact model)
ESD	Electrostatic discharge
FDSOI	Fully-depleted silicon on insulator
GCNMOS	Gate-coupled n-type metal-oxide semiconductor
GGNMOS	Gate-grounded n-type metal-oxide semiconductor
GND	Ground (reference voltage)
GNO	MunEDA WiCkeD Global Nominal Optimization (EDA tool)
GPDK	Generic process design kit
HBM	Human body model
HDL	Hardware Description Language
HICUM	High current model (bipolar compact model)

I/O	Input/output
IC	Integrated circuit
LGA	Land grid array
MEXTRAM	Most exquisite transistor model
MM	Machine model
MOS	Metal-oxide semiconductor (common abbreviation of MOSFET)
MOSFET	Metal-oxide semiconductor field-effect transistor
NDR	Negative differential resistance
NMOS	N-type metal-oxide semiconductor
PMOS	P-type metal-oxide semiconductor
RF	Radio-frequency
RH	Relative humidity
SCR	Silicon-controlled rectifier
SPICE	Simulation program with integrated circuit emphasis
SQP	Sequential quadratic programming
STI	Shallow trench isolation
STNMOS	Substrate-triggered n-type metal-oxide semiconductor
TLP	Transmission-line pulse
UTBB	Ultra-thin body and buried oxide
VBIC	Vertical bipolar inter-company
VDD	Voltage drain drain (power supply voltage)

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

1.1 Electrostatic discharge

Electrostatic discharge (ESD) is defined as the rapid, spontaneous transfer of electrostatic charge between two bodies at different electrostatic potentials, either through contact or via an ionized ambient discharge (a spark) induced by a high electrostatic field (ESD Association, 2013). Means of creating electrostatic charge on a material include induction, ion bombardment, contact with another charged material or *triboelectric charging*[1]. Table 1.1 illustrates typical charge generation scenarios, their resulting voltage levels and the influence of the relative humidity (RH) on reducing charge accumulation.

Table 1.1: Examples of electrostatic charge generation and the typical resulting voltage levels.

Means of generation	*10-25% RH*	*65-90% RH*
Walking across carpet	35,000 V	1,500 V
Walking across vinyl tile	12,000 V	250 V
Worker at a bench	6,000 V	100 V
Poly bag picked up from bench	20,000 V	1,200 V
Chair with urethane foam	18,000 V	1,500 V

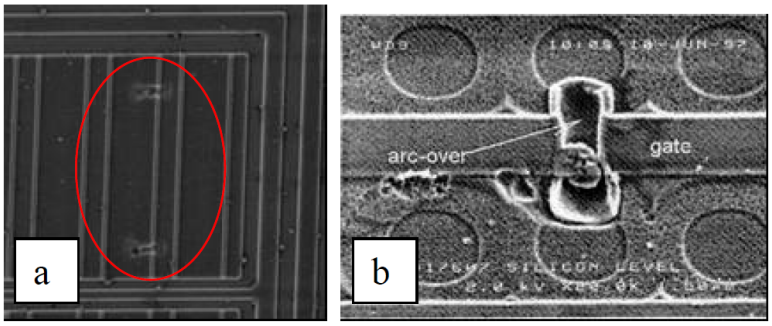
Source: ESD Association (2013)

1.2 Impact on integrated circuits

Electrostatic discharge may degrade or destroy semiconductor devices. In the first case, the device continues to perform its intended function. However, its operating life may be reduced (*latent failure*). In the second case, the device stops performing its intended function, partially or totally (*catastrophic failure*). Typical examples of ESD damage in CMOS integrated circuits include **metal melt**, **junction breakdown** and **oxide failure**. Figure 1.1 and Figure 1.2 illustrate common ESD failure signatures. Such failures naturally represent costs to semiconductor industry, impacting on production yields, manufacturing costs, product quality, product reliability and profitability.

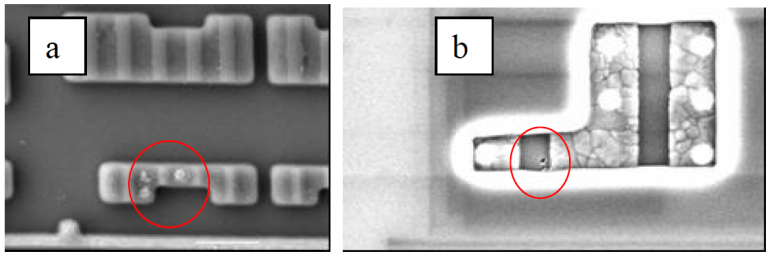
[1] *Triboelectric charging* is the effect of creating electrostatic charge by contact and separation of materials.

Figure 1.1: Failures in the drain junction of a MOS I/O transistor (a) and in the gate oxide of a core transistor (b).



Source: Industry Council on ESD Target Levels (2010)

Figure 1.2: Failures in two NMOS transistors (a) and in the gate oxide of an input transistor (b).



Source: Industry Council on ESD Target Levels (2010)

ESD damage in electronic devices is usually caused by one of the three basic ESD events below (ESD Association, 2013):

Discharge to the device The electrostatic charge is transferred from any charged conductor to the electronic device. A charged conductor, in this case, may be either the human body or a conductive object, such as a metallic tool or an automatic equipment. The models used to reproduce these events are, respectively, the *Human Body Model (HBM)* and the *Machine Model (MM)*. Figure 1.1 illustrates typical HBM-like failure signatures.

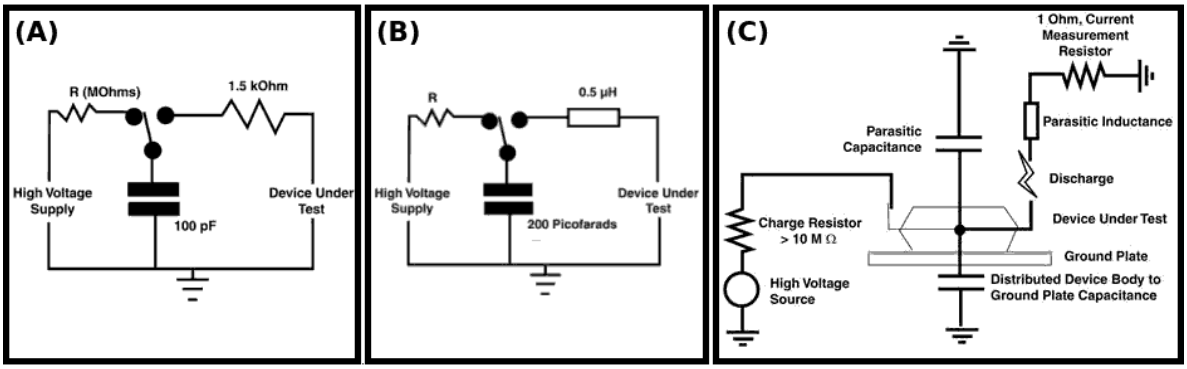
Discharge from the device Electronic devices may accumulate static charge themselves, for example, while they move across a surface or vibrate in a package. The electrostatic charge is then transferred from the electronic device to a conductor. The model used to reproduce this event is the *Charged Device Model (CDM)*. Figure 1.2 illustrates typical CDM-like failure signatures.

Field-induced discharge The electrostatic charge is induced on the electronic device when it is subjected to an electrostatic field associated to a charged object. If the device is then grounded while within the electrostatic field, the transfer of charge is the same described by the *Charged Device Model (CDM)*.

ESD Association, IEC, U.S. Department of Defense, JEDEC and JEITA are examples of

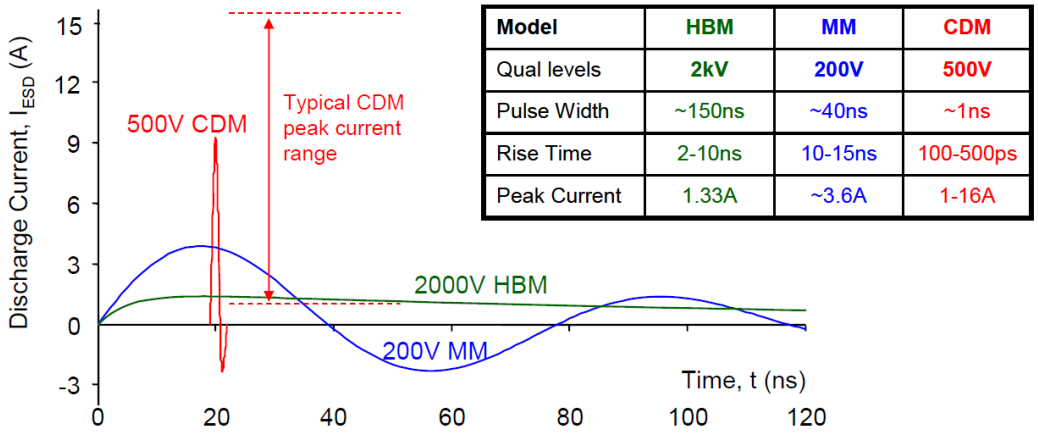
organizations that provide a significant amount of different standardized test methods to evaluate the ESD sensitivity levels of electronic devices. Figure 1.3 illustrates the test circuits[2] adopted by the ESD Association to reproduce ESD events described by HBM, MM and CDM. Figure 1.4 illustrates examples of typical current waveforms for each model. Recently, the industry has been driving efforts to remove the Machine Model (MM) test from device qualification requirements (Industry Council on ESD Target Levels, 2012) (DUVVURY et al., 2013) (ESD Association, 2014) (JEDEC, 2014). The Industry Council on ESD Target Levels provided data supporting that the MM qualification requirements are unnecessary when HBM and CDM specifications are both met (DUVVURY et al., 2013). These two primary models of ESD events have been proven to successfully reproduce over 99% of all ESD failure signatures found in field (ESD Association, 2014).

Figure 1.3: Typical HBM (a), MM (b) and CDM (c) test circuits.



Source: ESD Association (2014) (modified)

Figure 1.4: Current waveforms for HBM, MM and CDM ESD events.



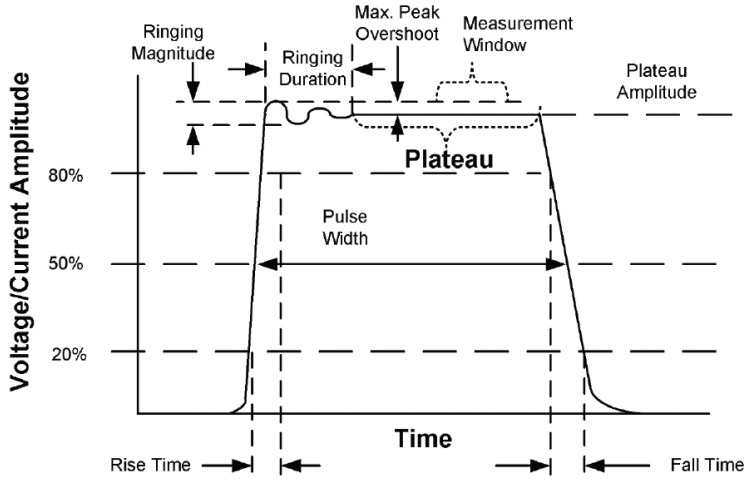
Model	HBM	MM	CDM
Qual levels	2kV	200V	500V
Pulse Width	~150ns	~40ns	~1ns
Rise Time	2-10ns	10-15ns	100-500ps
Peak Current	1.33A	~3.6A	1-16A

Source: Industry Council on ESD Target Levels (2010)

[2]Test circuits may differ in standards from one organization to another.

As mentioned above, there is a significant amount of different standardized test methods to evaluate the ESD sensitivity levels of electronic devices. However, the default method for characterizing the behavior of devices under ESD conditions is the Transmission-Line Pulse (TLP) testing (SMEDES, 2017). In contrast to HBM and CDM testing methods, TLP provides more than a pass/fail result. It also provides a quasi-static $I-V$ characteristic that describes the behavior of the device under test (DUT). The method basically consists in charging up and then discharging a transmission line (typically of 50 Ω) into a DUT, using pulses with duration, rise time and current levels that are relevant for ESD events. From measurements in a predefined stable part of the pulse, a quasi-static $I-V$ point can be determined. By repeating this charge/discharge procedure for multiple charge levels, a complete characteristic is established. Figure 1.5 illustrates a typical TLP pulse shape. In general, a low-current DC measurement is performed after each TLP pulse. A change in the DC measurement from one pulse to another is often an indication that the DUT has been damaged by the preceding pulse.

Figure 1.5: Typical TLP pulse shape. The annotations indicate relevant features that are defined in standard documents.



Source: Smedes (2017)

1.3 Trends and challenges

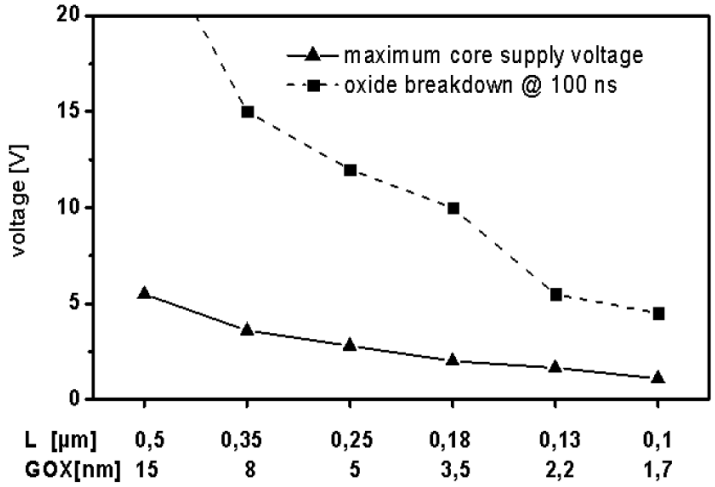
Technology scaling contributes to improve circuits speed, with shorter channel lengths and thinner gate oxides, as well as to reduce area costs. However, it places severe constraints on robust ESD design of integrated circuits:

Gate oxide is becoming thinner, resulting in lower breakdown voltages under ESD regime. Figures 1.6 and 1.7 show the trends of gate oxide breakdown voltages under HBM stress and CDM stress, respectively, with CMOS technology scaling. Consequently, the design of protection devices for the circuit inputs is becoming more difficult.

Drain junction Feature size is becoming smaller, source-to-drain distance is becoming shorter and integration density is increasing, resulting in lower drain junction breakdown voltages and lower failure currents under ESD regime (DUVVURY, 2008b). Figure 1.7 shows the trends of drain junction breakdown voltages under CDM stress with CMOS technology scaling. Consequently, the design of adequate protection devices for the circuit outputs is becoming more difficult.

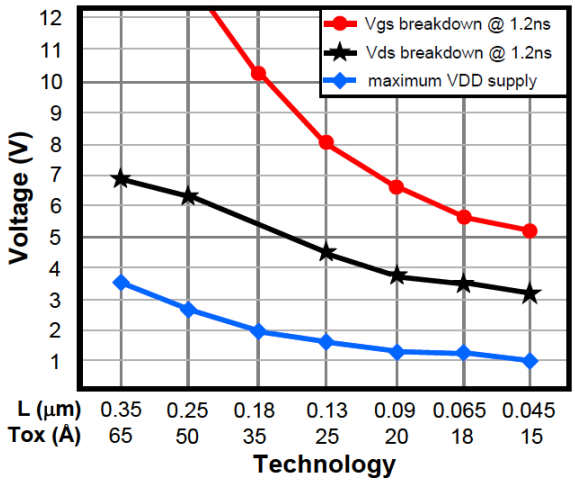
Metal interconnects are becoming thinner, resulting not only in reduced capacitance (usually desired for demanding performance requirements) but also resulting in increased resistance per square and reduced current density failure thresholds. As a consequence, resistive bus connections elevate the pad voltage at relatively low ESD currents (DUVVURY, 2008b). Trends of ESD current density for failure in metal interconnects with technology scaling are shown in Figures 1.8 (HBM) and 1.9 (CDM). As expected, failure currents are different for HBM and CDM discharges, since these events have different durations (pulse widths) and different current rise times, as seen in Figure 1.4.

Figure 1.6: HBM oxide breakdown voltage and core supply voltage as a function of technology scaling.



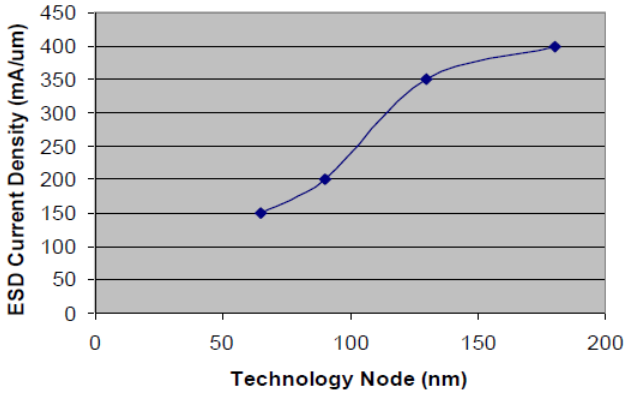
Source: Industry Council on ESD Target Levels (2011)

Figure 1.7: CDM oxide breakdown voltage (red line with circles), drain junction breakdown voltage (black line with stars) and core supply voltage as a function of technology scaling.



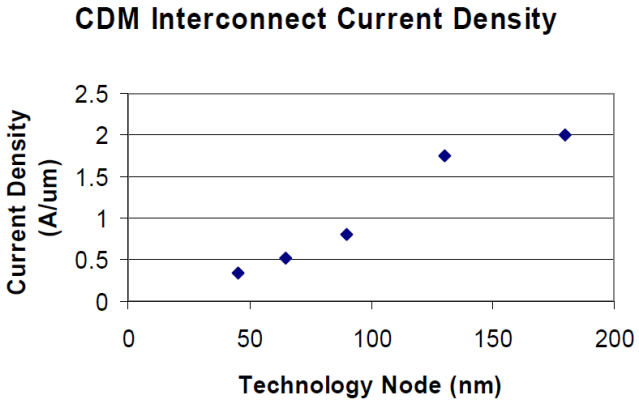
Source: Industry Council on ESD Target Levels (2010)

Figure 1.8: HBM current density for failure of metal interconnects as a function of technology scaling.



Source: Industry Council on ESD Target Levels (2011)

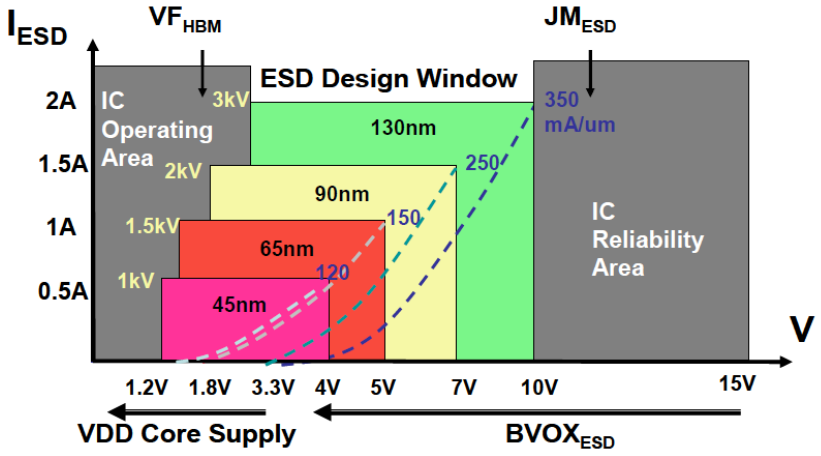
Figure 1.9: CDM current density for failure of metal interconnects as a function of technology scaling.



Source: Industry Council on ESD Target Levels (2010)

Technology scaling also impacts the **ESD design window**[3]. For instance, Figure 1.10 illustrates the trends in the ESD design window of High-Speed Serial Link designs at 15–20 Gbps. As the technology downscales, the gate oxide breakdown voltage ($BVOX_{ESD}$) reduces at a faster rate than the maximum VDD core supply voltage. It means that the voltage range available for ESD protection circuits to operate is becoming more narrow. Also, the current density failure thresholds for metal interconnects (JM_{ESD}), which are determined by the metal thickness, are becoming lower. The overall effect is a reduction of the practical ESD performance (VF_{HBM}) from 3 kV in the 130 nm node to eventually 1 kV in the 45 nm node. Figure 1.11 illustrates the ESD design window trends beyond the 65 nm node, where the ESD design margin is reaching the volt.

Figure 1.10: Impact of technology scaling on the ESD design window from 130 nm to 45 nm nodes.

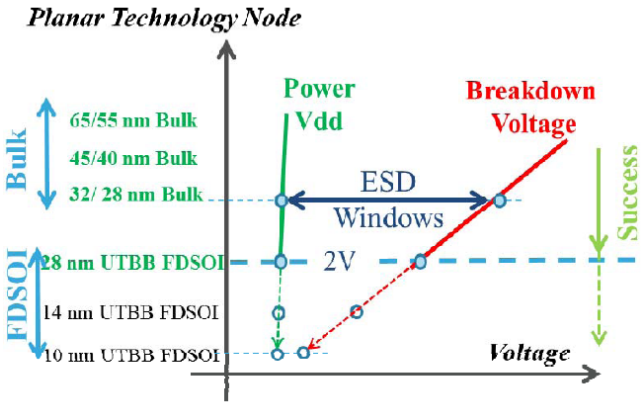


Source: Duvvury (2008b)

Effective ESD protection without compromising area and performance requirements is clearly becoming a challenge in deeply-scaled technologies. Figure 1.12 illustrates a projection of required area for ESD protection as the technology downscales. Beyond the 90 nm node, the required area to achieve 2 kV HBM ESD protection is becoming bigger than the area of the core circuit itself. In addition, package type (e.g. DIP, LGA and BGA), package size and number of circuit pins have significant impact on the protection against discharge currents at a

[3] *ESD design window* is a two-dimensional representation of the ESD protection design space for meeting a specific ESD robustness level. In the horizontal axis, the window is constrained by the application voltage and the ESD-regime gate oxide breakdown voltage (for input circuit protection) or drain junction breakdown voltage (for output circuit protection). In the vertical axis, the window is constrained by the maximum current carrying capacity of the ESD protection circuit. These constraints usually include a safety margin. ESD protection circuits must turn on beyond the application voltage and reach the ESD current specification before approaching the gate oxide or drain junction breakdown voltages.

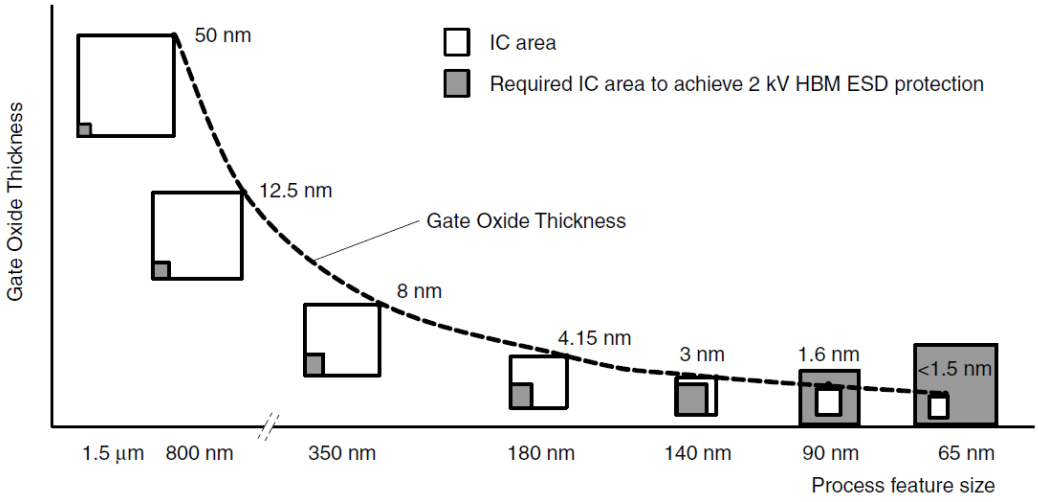
Figure 1.11: Impact of technology scaling on the ESD design window beyond the 65 nm node.



Source: Galy (2014)

given stress voltage, especially under CDM stress. The average package area and the number of circuit pins are increasing, driven by markets such as microprocessors and internet switching (Industry Council on ESD Target Levels, 2010). In general, larger is the package area (and its effective capacitance), higher are the CDM peak currents (DUVVURY, 2008a). It means that the I/O tolerance to CDM stress is being reached at lower voltage levels and, consequently, more robust ESD protection is required. The core circuit performance is also a concern. RF and high-speed digital I/O designs, for instance, can tolerate very little parasitic capacitance from the ESD protection devices (CAO et al., 2010) (KER; LIN; HSIAO, 2011). In analog design, another challenge expected in advanced technology nodes is the mismatch of ESD devices and ESD networks, which may lead to both local and global variation issues (VOLDMAN, 2012).

Figure 1.12: Required area to achieve 2 kV HBM ESD protection as the technology downscales.



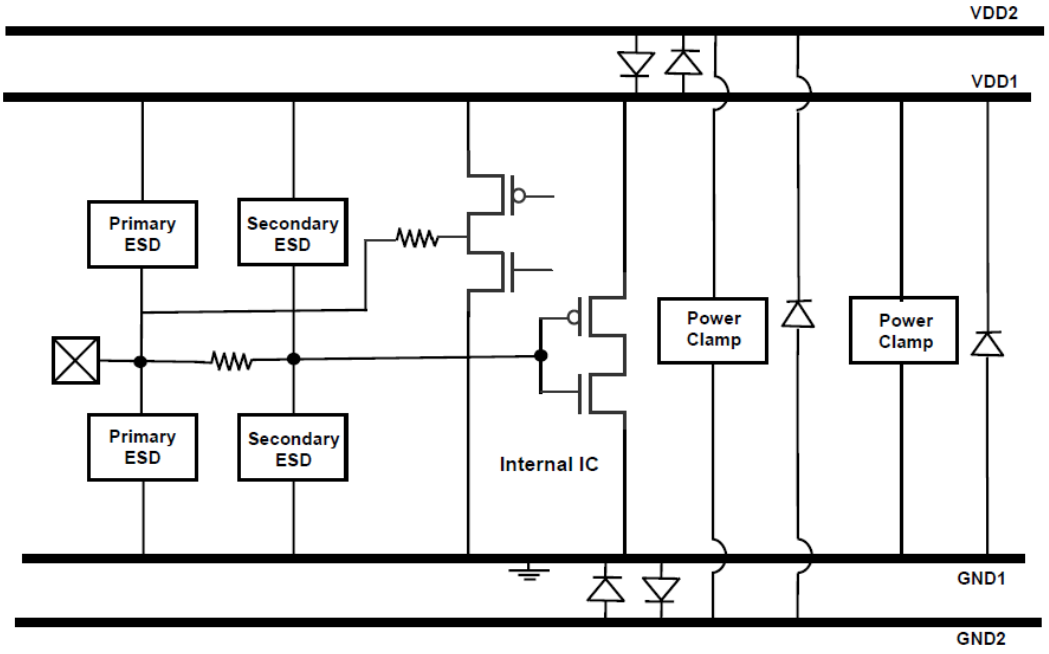
Source: NXP Semiconductors (2010)

Technology trends and their associated challenges to ESD protection severely increase the costs of semiconductor industry. These costs come from silicon area, circuit performance demands, resources, design respins and eventually from time-to-market. Clearly, product qualification levels for HBM and CDM might become impractical in advanced technology nodes. In this context, the Industry Council on ESD Target Levels has recently proposed a reduction from 2 kV to 1 kV for HBM (Industry Council on ESD Target Levels, 2011) and from 500 V to 250 V for CDM (Industry Council on ESD Target Levels, 2010) product qualification. Such new levels are supposed to be achievable with proper static control in production environments and modern ESD design methods.

1.4 On-chip ESD protection

Approaches to avoid ESD events and consequently increase manufacturing yields include proper handling of IC devices and the implementation of a controlled handling environment. However, an essential technique to reduce overall costs and improve the product reliability is the implementation of on-chip ESD protection circuits. As the breakdown voltages under ESD regime are in general very low (few volts) compared to the stress voltages (e.g. 250 V CDM, 8 kV HBM), the challenge is to hold the voltage surge below the technology breakdown voltage and shunt the discharge current from an I/O pin to a power supply pin (VDD or GND).

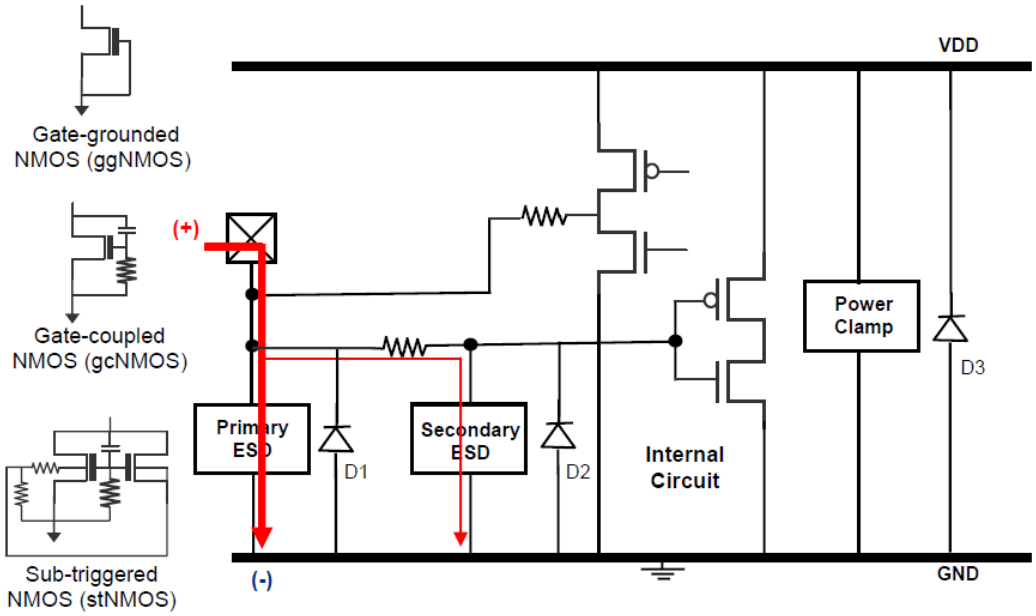
Figure 1.13: Generic ESD protection scheme for a bidirectional I/O circuit.



Source: Chun (2006)

Figure 1.13 illustrates a generic ESD protection scheme for a bidirectional I/O circuit. The purpose of the primary ESD devices is to limit the I/O pad voltage below the failure level of the output driver, shunting most of the ESD stress current to the power rails. The function of the secondary ESD devices with a series resistor between the primary and the secondary devices is to limit the gate voltage of the input receiver. It also aims to limit the current for low source impedance ESD stresses, thus being essential for CDM protection. The purpose of the power clamps and their parallel diodes is to provide a direct ESD current path between VDD and GND, protecting the devices of both the I/O circuit and the internal circuit against an ESD stress between these two power supply pins. Finally, the anti-parallel diodes between $VDD1$ and $VDD2$ and also between $GND1$ and $GND2$ aim to provide a direct ESD current path between two different power domains.

Figure 1.14: Pad-based protection scheme and examples of applicable ESD protection devices.



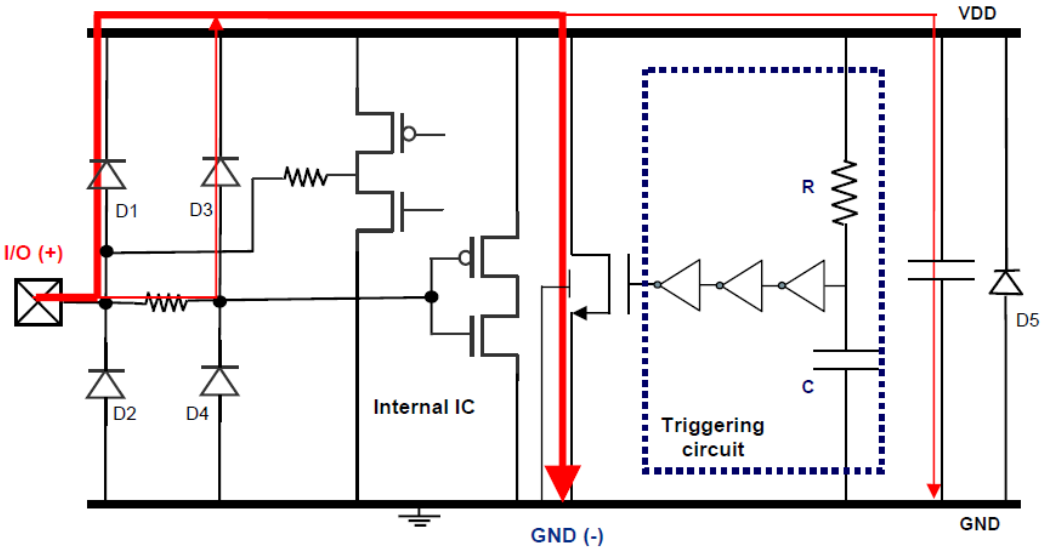
Source: Chun (2006)

Two common on-chip ESD protection approaches are the **pad-based protection** and the **rail-based protection** (CHUN, 2006). Figure 1.14 illustrates a pad-based protection scheme, where a positive ESD stress is applied from the I/O pad to GND. The ESD current, in this case, is directly shunted to GND by the primary and the secondary ESD devices. If a positive stress is applied from the I/O pad to VDD (instead of GND), the intended ESD current path also involves the diode $D3$. On the other hand, for a negative ESD stress applied from the I/O pad to GND, the intended ESD current path should involve only the diodes $D1$ and $D2$. If the negative stress is applied from the I/O pad to VDD (instead of GND), the intended

ESD current path consists not only of the diodes $D1$ and $D2$ but also of the power clamp. The function of the power clamp, in this case, is to hold the I/O pad voltage below the failure voltage of the I/O devices in parallel to it. Figure 1.14 also illustrates typical NMOS-based ESD protection devices used in pad-based configurations: the gate-grounded NMOS (GGNMOS[4]), the gate-coupled NMOS (GCNMOS) and the substrate-triggered NMOS (STNMOS). Another example of typical ESD protection device is the silicon-controlled rectifier (SCR), which has a very high current handling capability, resulting from the combination of two intrinsic bipolar transistors in a self-regenerative on-condition.

Figure 1.15 illustrates a rail-based protection scheme, where a positive ESD stress is applied from the I/O pad to GND. In this case, the ESD current is firstly redirected to the VDD power rail by the upper diodes $D1$ and $D3$, and then shunted to GND by the RC-triggered power supply clamp. If a positive stress is applied from the I/O pad to VDD (instead of GND), the intended ESD current path involves only the diodes $D1$ and $D3$. On the other hand, for a negative ESD stress applied from the I/O pad to GND, the intended ESD current path involves only the diodes $D2$ and $D4$. If the negative stress is applied from the I/O pad to VDD (instead of GND), the current flows from the VDD supply pad, passing through the power clamp and the lower diodes $D2$ and $D4$.

Figure 1.15: Rail-based ESD protection scheme.



Source: Chun (2006)

Both pad-based and rail-based ESD protection strategies have benefits and drawbacks. The design of pad-based ESD protection, for instance, is quite straightforward. However, the

[4]The *GGNMOS* is the most basic configuration among NMOS-based ESD protection devices. Its intrinsic *npn* lateral bipolar transistor has a large current handling capability which enables the device to bypass a significant amount of the ESD current.

snapback devices typically used are very sensitive to process variations, not being portable from one process to another (CHUN, 2006). Also, the compact circuit models of such snapback devices are not often provided, becoming difficult to verify the pad-based ESD protection effectiveness with circuit simulations. Rail-based ESD protection, on the other hand, often uses conventional devices (e.g. diodes and regular MOSFETs) which are much less susceptible to process variations, being relatively portable from one process to another. Also, as the compact circuit models of conventional devices are usually provided, the ESD protection circuit effectiveness can consequently be verified by circuit simulations.

1.5 ESD modeling and simulation

As explained in the previous sections, effective on-chip ESD protection without compromising area and performance requirements is becoming a challenge in deeply-scaled technologies. In this context, circuit simulation can provide the required assistance to on-chip ESD protection design, including robustness analysis of the circuits and performance prediction prior to silicon. However, modeling the MOSFET operation under ESD conditions for circuit simulation is still a challenging issue. The large current and voltage characteristics are typically not well covered by most standard SPICE compact models, such as BSIM3 (LIU et al., 1999) and BSIM4 (MORSHED et al., 2011). To overcome such limitation, a practical modeling approach is highly desired. For this purpose, **macromodels** can be very helpful.

A macromodel is a combination of conventional SPICE elements, optionally including math functions, to describe the observable *behavior* (not necessarily the implementation) of a device or circuit. Macromodels are often used to model devices that SPICE does not directly support. Most macromodels have many possible implementations. A classic example of a macromodel still in use today is the Boyle op-amp (BOYLE et al., 1974), which combines conventional SPICE elements with simple math functions to model the behavior of an op-amp. Instead of including all the transistors of the device (which would change from one op-amp design to another), simplified functions and parameters allow the adjustment of the model behavior.

The usage of macromodels to describe the $I-V$ characteristics of an ESD protection device for circuit simulation instead of using a hardware description language (HDL) such as Verilog-A, for example, reduces convergence issues. Also, there is no need to switch between different models and there is no need of a specific circuit simulator other than a regular SPICE simulator. With macromodels, both the efficacy of the ESD protection device and its influence

on the core circuits performance can be evaluated.

This work presents a study of MOSFET macromodels for ESD circuit simulation. First, Chapter 2 gives an overview of the device operation under ESD conditions. Then Chapter 3 presents the evolution of MOSFET macromodels for ESD circuit simulation. Finally, Chapter 4 introduces a novel macromodel development approach based on parametric analysis and optimization.

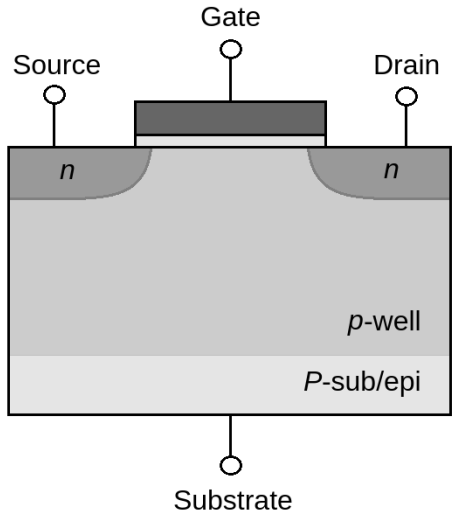
2 MOSFET OPERATION UNDER ESD CONDITIONS

During an ESD event, devices in the ESD path might operate outside their normal operating ranges. Thus, an understanding of devices operation under ESD stress conditions is essential to properly model their behavior. This chapter focuses on the MOSFET operation with emphasis on NMOS transistors. However, the description presented here is also applicable to PMOS transistors, with changes in the polarity of the majority and minority carriers.

2.1 Low-current operation

Figure 2.1 illustrates a simplified NMOS cross section schematic. When a positive gate-to-source voltage V_{GS} is applied, majority carriers of the *p* region below the gate are pushed away by the electric field, creating a depletion region. With further increase in V_{GS}, a significant amount of minority carriers of the *p* region are attracted towards the gate, eventually creating a conductive channel between the source and the drain terminals. The gate-to-source voltage at which a conductive channel between the source and the drain terminals is created is called *threshold voltage* (V_T).

Figure 2.1: NMOS transistor cross section schematic.



Source: the author

For low current densities, the typical operating regions of an NMOS device are **cut-off**, **linear** and **saturation**. While $V_{GS} \leq V_T$, the NMOS transistor is in the cut-off mode and the drain-to-source current I_{DS} is approximately zero. For $V_{GS} > V_T$, a conductive channel

is formed below the gate and the drain-to-source voltage V_{DS} determines whether the NMOS transistor is in the linear or in the saturation mode. When $V_{DS} \leq V_{GS} - V_T$, the NMOS transistor is in linear mode and the drain-to-source current I_{DS} is approximately

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} \left[(V_{GS} - V_T)\, V_{DS} - \frac{V_{DS}^2}{2} \right] (1 + \lambda V_{DS}) \tag{2.1}$$

where μ_n is the electron effective mobility, C_{ox} is the gate oxide capacitance per unit area, W is the channel width, L is the channel length and λ is a fitting parameter to consider effects such as channel length modulation, drain-induced barrier lowering or two-dimensional field distributions. In this case, the drain-to-source current I_{DS} is a linear function of the gate-to-source voltage V_{GS}. When $V_{DS} > V_{GS} - V_T$, the NMOS transistor is in the saturation mode and the drain-to-source current I_{DS} is a quadratic function of the gate-to-source voltage V_{GS}, given approximately by

$$I_{DS} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) \tag{2.2}$$

Cut-off, linear and saturation are the typical operating regions of the MOS transistor. Equations (2.1) and (2.2) are indeed relevant models to understand which are the main parameters that influence the NMOS drain-to-source current I_{DS}. However, the current equations in modern MOSFET compact models, such as BSIM3 and BSIM4, are usually more complex than that (LIU et al., 1999) (MORSHED et al., 2011). Also, during an ESD event, the high current and voltage levels may lead MOS transistors to operate outside their typical operating regions.

2.2 Avalanche breakdown

Under ESD conditions, the high voltage levels may lead MOS transistors to operate in the **avalanche** and in the **snapback** regions. Figure 2.2 illustrates the drain current I_D as a function of the drain voltage V_D for different gate voltages V_G, as well as the corresponding regions of operation. From Figure 2.2, it is possible to notice that for a gate voltage V_G greater than the threshold voltage V_T, the drain current contributes to reduce the drain voltage at which the device enters the avalanche and snapback regions.

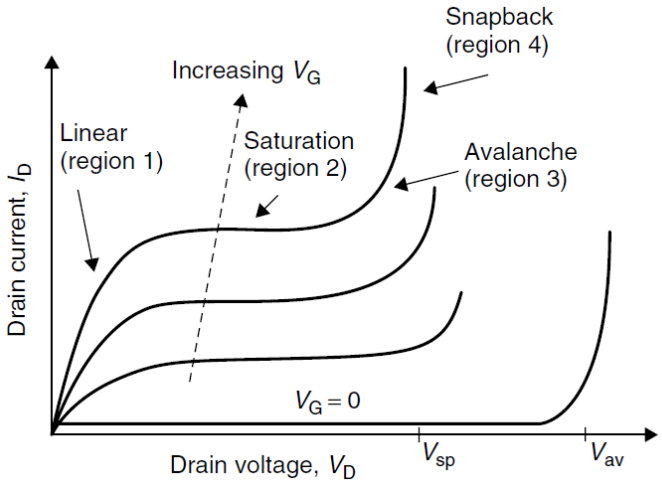
To simplify the understanding of the high-voltage operation of the NMOS device, let us assume gate, source, and substrate terminals are grounded (0 V) and an ESD event occurs, accumulating a significant amount of positive charges at the drain terminal in a very short time duration. As the electric potential at the drain terminal increases, majority carriers are pushed away from the reverse-biased drain-substrate junction, which becomes depleted. When the

potential drop across the depletion region reaches a few tenths of a volt, the reverse current is given by (AMERASEKERA; DUVVURY, 2002)

$$I_R = \frac{qADN_CN_V}{L_dN_B}\exp\left(-\frac{E_g}{kT}\right) + \frac{qW}{\tau_e}\sqrt{N_CN_V}\exp\left(-\frac{E_g}{2kT}\right) \tag{2.3}$$

where q is the electronic charge (1.602×10^{-19} C), A is the pn junction area (proportional to the diffusion area), D is the minority carrier diffusion coefficient, N_C and N_V are the density of states in the conduction band and in the valence band, respectively, L_d is the diffusion length, N_B is the background doping concentration, E_g is the energy bandgap (about 1.12 eV for Silicon), k is the Boltzmann's constant (1.38×10^{-23} JK^{-1}), T is the temperature, W is the width of the depletion region and τ_e is the effective carrier lifetime. In this case, the polarity of the voltage biasing the drain-substrate junction allows the flow of electrons from the *p* substrate to the *n* diffusion and holes from the *n* diffusion to the *p* substrate. As there are few electrons in the *p* substrate and few holes in the *n* diffusion (minority carriers), the reverse current of Equation (2.3) is usually very small.

Figure 2.2: $I-V$ curves of an NMOS transistor under high-voltage conditions.



Source: Amerasekera and Duvvury (2002)

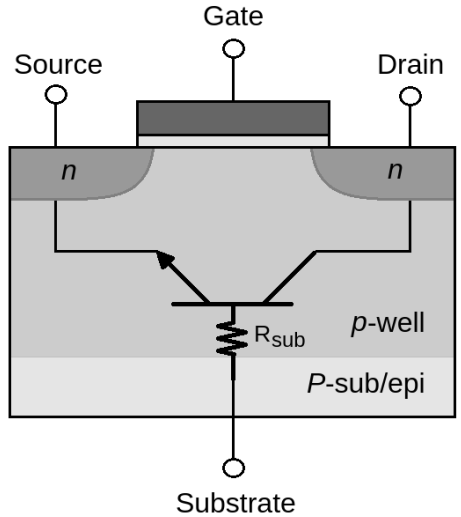
As the reverse voltage across the depletion layer increases even further, the resulting electric field E eventually reaches a critical value E_{crit}. Under such critical electric field, some accelerated electrons can impart enough kinetic energy in a collision with the lattice to lift electrons from the valence band into the conduction band, thereby creating new electron-hole pairs, which become free carriers. This process is known as **impact ionization**. The electrons and holes created by impact ionization are themselves also accelerated by the electric field and may also collide with the lattice, creating even more free carriers. An avalanche effect is then

triggered (region 3 in Figure 2.2). The carrier creation rate and the reverse current I_R rise abruptly. This process is known as **avalanche breakdown**. The critical electric field E_{crit} depends on the n and p doping concentrations. In a submicron process, E_{crit} is reached at about 10^5 Vcm^{-1} (AMERASEKERA; DUVVURY, 2002). It means that, in a rough calculation, impact ionization and avalanche breakdown can be triggered when approximately 10 V is applied in a depletion width of 1 μm. In the case of an ESD event, the resulting peak voltage may reach not only tens, but hundreds (CDM) or even thousands (HBM) of a volt.

2.3 Lateral bipolar transistor activation

Both NMOS and PMOS devices have an intrinsic lateral bipolar transistor, which eventually turns on under ESD conditions and dominates the device behavior. Figure 2.3 illustrates the schematic of a lateral *npn* transistor in an NMOS device cross section. When the npn turns on, part of the NMOS drain becomes the bipolar collector, part of the NMOS source becomes the bipolar emitter and part of the NMOS substrate becomes the bipolar base. The NMOS regions that will effectively compose the lateral bipolar structures will depend on factors such as the effective area of the pn junctions that will be biased as well as the current density distribution across the structures.

Figure 2.3: Schematic of the intrinsic lateral *npn* transistor in an NMOS device cross section.



Source: the author

During the avalanche breakdown, electrons are swept across the drain junction towards the drain contact, while holes drift towards the substrate contact, giving rise to a substrate current I_{sub}. An increase in I_{sub} also increases the voltage drop in the effective substrate resistance R_{sub} of Figure 2.3. When this voltage drop reaches the built-in voltage of the source-substrate junction (typically about 0.5 V), i.e.

$$V_{sub} = R_{sub}\, I_{sub} \geq 0.5\, V \tag{2.4}$$

the source-substrate junction becomes forward-biased. Electrons are then emitted from the source into the substrate and eventually reach the drain-substrate junction, depending on the distance between the n diffusions (equivalent to the base width of the intrinsic bipolar transistor) and the recombination rate in the substrate, which is dependent on the substrate hole concentration. When the electrons emitted from the source reach the reverse-biased junction, they contribute to the total drain current. With the drain-substrate junction operating in reverse bias condition and the source-substrate junction operating in forward bias condition, the intrinsic bipolar transistor is in the on-state, under active bias condition. The time it takes to the lateral npn to turn on is basically given by the **base transit time** (τ_B), which is dependent on the NMOS gate length L. Under active bias condition, the collector current is given by (AMERASEKERA; DUVVURY, 2002)

$$I_c = \frac{qn_i^2\tilde{D}}{N_B} A \exp\left(\frac{qV_{be}}{kT}\right) \tag{2.5}$$

where n_i is the intrinsic carrier concentration, $\tilde{D}$ is the effective minority carrier diffusion coefficient, N_B is the total number of impurities per unit area of the base, A is the base-emitter junction area and V_{be} is the base-emitter voltage. It is important to notice in Equation (2.5) that the collector current I_c (equivalent to the NMOS drain current) is an exponential function of V_{be} (equivalent to the voltage drop in the NMOS source-substrate junction).

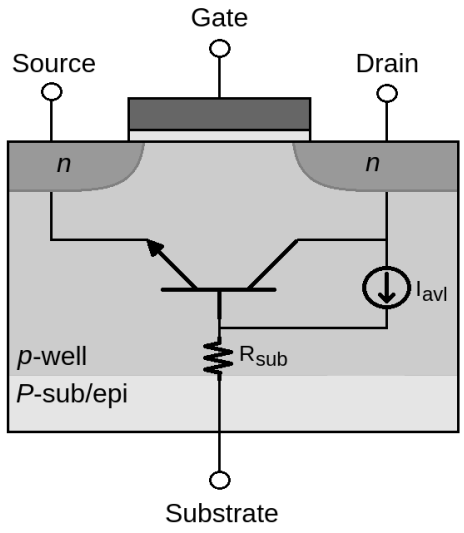
In this case, the lateral npn transistor operates in **self-triggering** or **self-biasing** mode. The collector-base junction itself, under avalanche breakdown, forward biases the emitter-base junction, self-triggering the transistor. Figure 2.4 illustrates the NMOS cross section schematic with its intrinsic bipolar transistor operating in the self-triggering mode. I_{avl} represents the current resulting from the avalanche effect in the collector-base (drain-substrate) junction. Its relation with an incident current I_{ESD} in the high-field region is given by (AMERASEKERA; DUVVURY, 2002)

$$I_{avl} = (M-1)I_{ESD} \tag{2.6}$$

where M is a multiplication factor in the collector-base (drain-substrate) depletion region. By definition, M is the ratio of the output current I_{out} to the input current I_{in} at the avalanching junction, i.e.

$$M = \frac{I_{out}}{I_{in}} \tag{2.7}$$

Figure 2.4: Avalanche current (I_{avl}) schematic.



Source: the author

An approximation for M may be obtained by Miller's empirical formulation (MILLER, 1957) (DUTTON, 1975)

$$M = \frac{1}{1 - \left(\frac{V_{cb}}{BV_{cb}}\right)^n} \tag{2.8}$$

where V_{cb} is the external applied collector-base voltage, BV_{cb} is the collector-base breakdown voltage and n is a fitting parameter ranging from 2 to 7, depending on the type of junction being considered. With n as a constant, this empirical expression is only accurate for relatively low collector current densities. In the case of large current densities, n may be described as a variable dependent on the collector current to obtain a more accurate M (DIVEKAR; LOVELACE, 1982) (KLOOSTERMAN; GRAAFF, 1989).

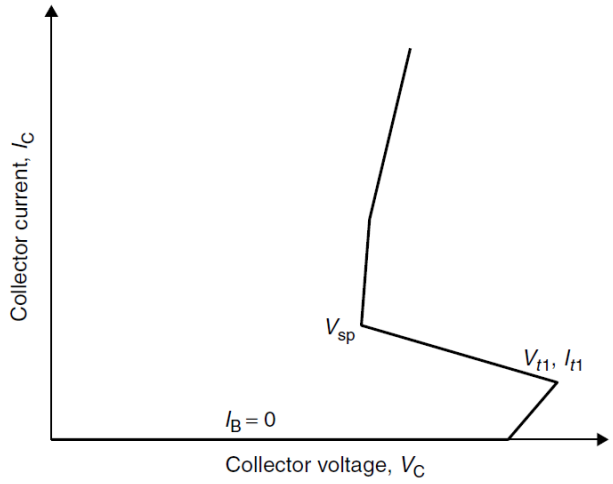
Before the lateral npn turns on, $I_{avl} = I_{sub}$. Once it turns on, the electrons emitted from the forward-biased emitter-base (NMOS source-substrate) junction that eventually reach the reverse-biased collector-base (NMOS drain-substrate) junction are also subjected to the high electric field in the depletion layer and potentially contribute to the generation of more electron-hole pairs by impact ionization, increasing even more the avalanche process and, consequently, the avalanche current I_{avl}.

2.4 Snapback and negative differential resistance

Figures 2.5 and 2.6 illustrate the $I-V$ curves of an npn transistor and an NMOS transistor, respectively, under ESD conditions. When the intrinsic lateral npn transistor turns on and the emitter-base electron current start to contribute to the avalanche process in the collector-base junction, the relation between the base current I_b and the emitter current I_e is given by (AMERASEKERA; DUVVURY, 2002)

$$I_b \propto (M-1)I_e \tag{2.9}$$

Figure 2.5: $I-V$ curve of an npn transistor in self-bias operation under high-current conditions.

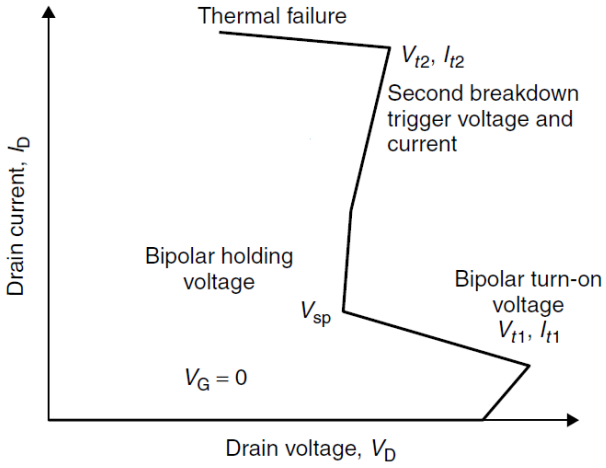


Source: Amerasekera and Duvvury (2002)

As the emitter current I_e increases, contributing to the avalanche multiplication and to the total current in the collector-base junction, the collector-base voltage V_{cb} and thus the multiplication factor M start to decrease. At this moment, a snapback characteristic is observed in the npn $I-V$ curve (Figure 2.5), and consequently in the NMOS $I-V$ curve (Figure 2.6). The snapback triggering voltage, i.e. the voltage at which V_c and V_D begins to decrease, is represented by V_{t1} in the Figures 2.5 and 2.6, respectively. Likewise, the snapback triggering current, i.e. the total collector current I_c at which V_c begins to decrease or, equivalently, the total drain current I_D at which V_D begins to decrease, is represented by I_{t1}. The condition for snapback is given by (AMERASEKERA; DUVVURY, 2002)

$$\beta\,(M-1) \geq 1 \tag{2.10}$$

Figure 2.6: $I-V$ curve of an NMOS transistor with $V_G = 0$ under high-current conditions.



Source: Amerasekera and Duvvury (2002) (modified)

where β is the npn common-emitter current gain. The snapback characteristic of the npn transistor is a consequence of the positive feedback loop between the avalanche generation of holes in the collector-base junction and the injection of electrons from the emitter-base junction into the npn base. At high-current conditions, injected electrons and generated holes change the carriers balance, especially in the p region, significantly changing the device conductivity. Such change in the conduction properties of a particular structure region can be defined as **conductivity modulation** (VASHCHENKO; SHIBKOV, 2010). In the case of the lateral npn transistor, the internal conductivity modulation results in a unstable **negative differential resistance** (NDR) effect between its terminals,

$$\frac{dR}{dt} < 0 \tag{2.11}$$

i.e. as the collector current I_c increases, the collector voltage V_c decreases,

$$\frac{\Delta V_c}{\Delta I_c} < 0\,. \tag{2.12}$$

Consequently, the electric field across the collector-base depletion region also decreases. The region between (V_{t1}, I_{t1}) and V_{sp} in the Figures 2.5 and 2.6 illustrates the $I-V$ characteristics of the snapback behavior and the NDR effect.

2.5 Low-impedance operation and thermal breakdown

The NDR effect saturates when the npn collector voltage V_c reaches the minimum level required to maintain the bipolar transistor in a stable on-condition. Such voltage level is denoted

by V_{sp} in the Figures 2.5 and 2.6. The $I-V$ curve now shows a positive resistance and the device provides a low-impedance path to the injected current.

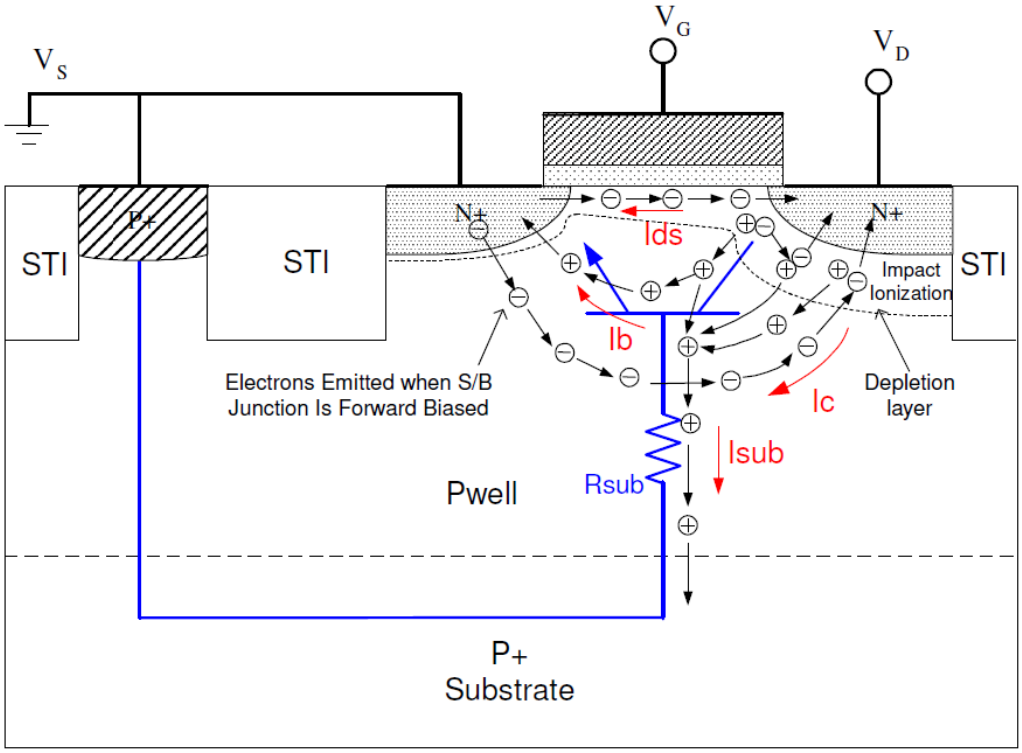
As the injected current increases, the temperature in the device also rises due to power dissipation, mostly in the reverse-biased collector-base (drain-substrate) junction where the current density and the electric field are both high. At a certain level, the temperature becomes high enough to thermally generate a significant amount of free carriers, which are also accelerated by the electric field in the reverse-biased junction, inducing a new avalanche process due to impact ionization. Once again, the free carriers density changes, especially in the base region of the bipolar (or in the substrate region of the NMOS), and the resulting conductivity modulation leads the device into another unstable differential negative resistance condition. The voltage drops for the second time. This condition is known as **second breakdown** or **thermal breakdown**. The voltage and the current at which the second breakdown takes place is denoted by V_{t2} and I_{t2}, respectively, in the Figure 2.6. With further increase in the current, the temperature in hot spots may quickly reach the semiconductor melt temperature and permanently damage the device (AMERASEKERA; DUVVURY, 2002).

2.6 Summary of the major ESD effects

Figure 2.7 schematically summarizes the major effects on an NMOS transistor under typical ESD conditions. In this example, we assume the gate is biased and an ESD event occurs at the drain terminal. As the gate-to-source voltage V_{GS} starts to increase and reaches the threshold voltage V_T, an *n* channel is formed below the gate and the device enters first the **linear** and then the **saturation** regions. The drain-to-source current is denoted by I_{ds} in Figure 2.7. With the high ESD-induced current levels, the drain capacitance rapidly charges up, generating a high electric field across the reverse-biased drain-substrate junction. Such high electric field eventually causes carriers generation in the depletion region due to **impact ionization**. These new carriers are also subjected to the high electric field and consequent impact ionization, generating even more carriers. With the resulting avalanche effect, the device enters the **avalanche region**. Electrons flow towards the drain contact while holes flow towards the substrate contact, giving rise to a substrate current (denoted by I_{sub} in Figure 2.7). When the resulting voltage drop across the effective substrate resistance (denoted by R_{sub} in Figure 2.7) reaches about 0.5 V, the source-substrate junction becomes forward-biased and the lateral npn transistor is then triggered, operating in self-biasing mode. The npn collector and base currents are denoted by I_c and I_b, respectively, in Figure 2.7. With the positive feedback loop between the

avalanche generation of holes in the collector-base junction and the injection of electrons from the emitter-base junction into the base, the device exhibits a **negative differential resistance** and enters the **snapback region**. In this region, the drain voltage decreases with an increase in the drain current until a stable holding voltage condition is reached. The device now starts to provide a low and positive impedance path to the ESD current. As the incident current increases even more, the temperature becomes high enough to thermally generate a significant amount of new carriers and the device eventually enters the **thermal** or **second breakdown region**. In this region, the temperature in hot spots may permanently damage the device.

Figure 2.7: NMOS transistor operation under typical ESD conditions.



Source: Zhou et al. (2005a)

3 MOSFET MACROMODELS FOR ESD CIRCUIT SIMULATION

Macromodels might not accurately reflect the target circuit behavior under every possible condition, since they might not necessarily include all available circuit parameters. Despite such limitation, under specific conditions, a macromodel may significantly reduce the model complexity and thus its simulation speed, as well as the modeling time, while maintaining a reasonable accuracy. Also, it may help to resolve convergence issues sometimes present in more complex and detailed models. In the context of this work, macromodels are particularly interesting to describe the large current and voltage characteristics of a MOSFET under ESD stress, which are typically not well covered by most standard SPICE compact models.

3.1 Modeling the avalanche breakdown with a parallel BJT and a current source

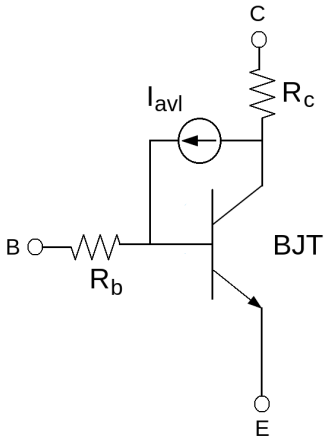
First efforts on modeling the avalanche breakdown for circuit simulation started in the early 70's (POON; MECKWOOD, 1972) (FOSSUM, 1973) to overcome a limitation of BJT compact models available at that time, especially of those based on Ebers-Moll (EBERS; MOLL, 1954) and Gummel-Poon (GUMMEL; POON, 1970) models. Figure 3.1 illustrates Dutton's bipolar equivalent circuit (DUTTON, 1975), where a voltage-dependent current source I_{avl} is added between the collector and the base resistances, R_c and R_b, respectively, of an Ebers-Moll-based compact model to include the avalanche breakdown effect. In this case, I_{avl} is given by

$$I_{avl} = K_1(M-1)I_{ES}\exp\left(\frac{qV_{be}}{kT}\right) \tag{3.1}$$

where K_1 is a fitting parameter, M is given by the Equation (2.8), I_{ES} is the collector intercept current, q is the electronic charge, V_{be} is the base-emitter voltage, k is the Boltzmann's constant and T is the temperature. Equation (3.1) shows the exponential dependency of the avalanche current on the base-emitter voltage.

Concerning the model implementation, the inclusion of the voltage-dependent current source between the collector and the base resistances required direct access to the internal structures of the BJT compact model adopted. Besides, the circuit simulation program had to be modified to include not only I_{avl} equation, but also its partial derivatives and the avalanche conductance and transconductance equations. Although the implementation was not straightforward, Dutton's model represented a significant contribution on modeling the avalanche breakdown of BJTs for circuit simulation.

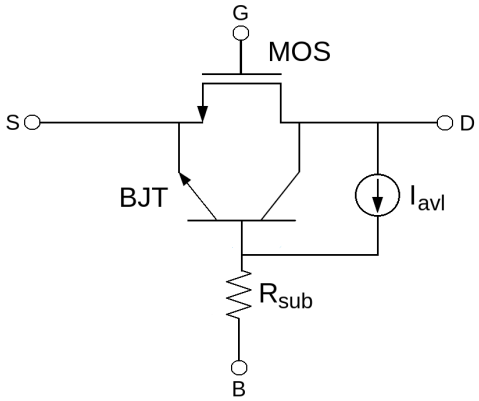
Figure 3.1: Modeling the avalanche breakdown of a BJT with a current source (I_{avl}).



Source: the author

Likewise, MOSFET compact models also had (and still have) limitations on modeling the device high-current operation. Figure 3.2 illustrates one of the first equivalent circuits adopted to model the avalanche breakdown of a short-channel MOSFET for circuit simulation (SUN et al., 1978). The role of the parallel BJT is to model the influence of the intrinsic lateral bipolar transistor on the NMOS device. The voltage-dependent current source I_{avl} aims to model the avalanche breakdown effect. R_{sub} intends to model the NMOS substrate resistance, which may also include the conductivity modulation effect if a current-dependent resistance is adopted. Later, analytical models including the high-current effects in MOSFETs were published by Hu and Chi (1982) and Hsu et al. (1982). However, despite the analytical results were comparable to measured data, these models were not implemented for circuit simulation. Actually, given the amount of equations and conditions, their implementation would require deep and complicated modifications in the MOSFET compact models and in the circuit simulator.

Figure 3.2: Modeling the avalanche breakdown of an NMOS transistor with a parallel BJT and a current source (I_{avl}).



Source: the author

Probably the first successful application of a short-channel MOSFET breakdown model for circuit simulation was performed by Pinto-Guedes and Chan (1988). Their model consisted in a modified version of the analytical model proposed by Hsu et al. (1982), where the impact ionization current of the substrate was replaced by an improved version of the impact ionization current model published by Mar, Li and Yu (1982). Also, a channel length dependence was included in the gain of the intrinsic lateral bipolar transistor to better describe the $I-V$ characteristics for a wider range of MOSFET channel lengths. Figure 3.3 illustrates the resulting equivalent circuit. Two current sources represent two components of the impact ionization current. I_{avl1} represents the current component due to avalanche multiplication of the MOSFET channel current, which is given by

$$I_{avl1} = (M_{CH} - 1)I_{CH} \tag{3.2}$$

where M_{CH} the is impact ionization multiplication factor for the channel current and I_{CH} is the MOSFET channel current. The other component, I_{avl2}, represents the current due to avalanche multiplication of the intrinsic lateral BJT emitter current, which is given by

$$I_{avl2} = \alpha\,(M_e - 1)I_e \tag{3.3}$$

where α is the bipolar common-base current gain, M_e is the impact ionization multiplication factor for the emitter current and I_e is the bipolar emitter current. The bipolar current gain α can be given by

$$\alpha = \gamma\,\alpha_T \tag{3.4}$$

where γ is the emitter injection efficiency and α_T is the base transport factor of the lateral BJT. However, to account for the channel length dependence of the model, an empirical expression for α was used by Pinto-Guedes and Chan, which is given by

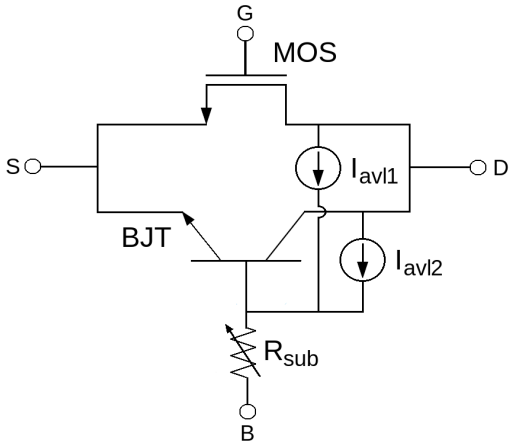
$$\alpha = \frac{\alpha_0}{1 + \frac{L}{L0}} \tag{3.5}$$

where α_0 and L_0 are two characterization constants and L is the effective channel length of the MOSFET.

Concerning R_{sub} in Figure 3.3, it is possible to use a current-dependent resistor to account for the conductivity modulation effect in the substrate, however Pinto-Guedes and Chan have neglected this effect and assumed R_{sub} as being constant to simplify the model. The implementation of the model basically consisted in adding the current components and the con-

ductances related to the intrinsic BJT directly into the circuit simulator. However, the task of implementing equations directly into the circuit simulator is not straightforward, especially considering the simulator source code is usually not available.

Figure 3.3: Modeling the avalanche breakdown of an NMOS transistor with a parallel BJT and two current sources (I_{avl1} and I_{avl2}).



Source: the author

Significant advances on modeling the avalanche breakdown with a parallel BJT and a current source have also been made by Amerasekera et al. (1996). Based on the equivalent circuit of Figure 3.2, they developed a set of equations which describes the high-current operation of a MOSFET and implemented these equations directly in a circuit simulator. In this model, the avalanche current I_{avl} is given by

$$I_{avl} = (M - 1)(I_{DS} + I_c) \tag{3.6}$$

where M is the avalanche multiplication factor in the high-field region of the drain, I_{DS} is the drain-to-source current (from the standard MOSFET compact model adopted) and I_c is the bipolar collector current. The avalanche multiplication factor in Equation (3.6) is given by

$$M = \frac{1}{1 - K_1 \exp\left(-\dfrac{K_2}{V_D - V_{D_{sat}}}\right)} \tag{3.7}$$

where K_1 and K_2 are fitting parameters, V_D is the applied drain voltage and $V_{D_{sat}}$ is the drain-source saturation voltage. The Miller's empirical formulation of M described by Equation (2.8) was not used by Amerasekera *et al.* because it does not take into account the effect of the gate

voltage V_G on the electric field at the drain junction. Such effect of V_G is then included through $V_{D_{sat}}$, which is given by

$$V_{D_{sat}} = \frac{V_G - V_T}{a_1 + b_1 \left(V_G - V_T\right)} \tag{3.8}$$

where V_T is the MOSFET threshold voltage, a_1 and b_1 are parameters of the standard MOSFET compact model adopted, with b_1 dependent on the channel length. The bipolar collector current I_c in Equation (3.6) is given by

$$I_c = I_{oc}\left[\exp\left(\frac{V_{be}}{V_{thermal}}\right) - \exp\left(\frac{V_{bc}}{V_{thermal}}\right)\right] \tag{3.9}$$

where I_{oc} is the reverse saturation current due to the diffusion of electrons in the base of the npn, V_{be} is the voltage across the base-emitter junction, $V_{thermal}$ is the thermal voltage (kT/q) and V_{bc} is the voltage across the base-collector junction. The bipolar base-emitter current I_b, on the other hand, is modeled by

$$I_b = I_{oe}\left[\exp\left(\frac{V_{be}}{V_{thermal}}\right) - 1\right] \tag{3.10}$$

where I_{oe} is the reverse saturation current due to the diffusion of holes in the emitter of the npn, V_{be} and $V_{thermal}$ are the same parameters of Equation (3.9). The substrate resistance in the circuit of Figure 3.2 is given by

$$R_{sub} = \frac{V_{be}}{I_{sub}} \tag{3.11}$$

where V_{be} is the voltage across the base-emitter junction and I_{sub} is the substrate current (measured at the snapback point).

It is important to observe the discontinuity of M in Equation (3.7). Such discontinuity may introduce simulation convergence issues. Therefore, it is preferably to use a continuous function instead, such as the one proposed by Lim et al. (1997)

$$M = \exp[K_1(V_D - V_{D_{sat}} - d_1)] + \exp[K_2(V_D - V_{D_{sat}} - d_2)] \tag{3.12}$$

where K_1, K_2, d_1, d_2 are fitting parameters, V_D is the applied drain voltage and $V_{D_{sat}}$ is the drain-to-source saturation voltage.

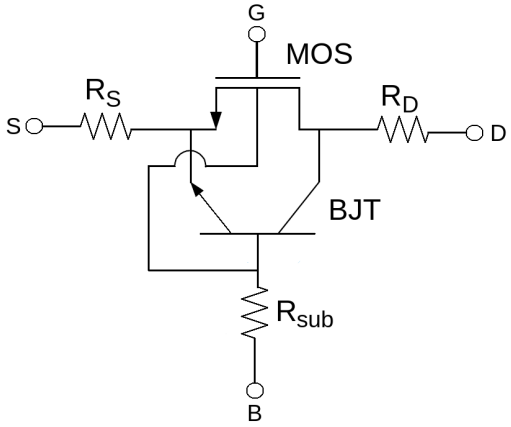
Last, it is also important to notice that the implementation of the equation set mentioned above in a circuit simulator is far from being a straightforward task. Besides, a parameter extraction methodology is required not only for the MOSFET compact model but also for K_1, K_2, I_{oc} and I_{oe} parameters of the equation set.

3.2 Modeling the avalanche breakdown with BSIM3 and VBIC compact models

The efforts on modeling the avalanche breakdown with a parallel BJT and a current source detailed in Section 3.1 were mostly motivated by the lack of such feature in the compact models available at that time, such as the BSIM models for the MOSFET and the Ebers-Moll and Gummel-Poon models for the BJT. The first popular BJT compact model to include the avalanche breakdown effect was VBIC (MCANDREW et al., 1996), by implementing the avalanche model proposed by Kloosterman and Graaff (1989). In fact, improvements in VBIC over Gummel-Poon compact model included not only a base-collector avalanche model but also a base-emitter breakdown model, a self-heating model, an improved depletion and diffusion charge model, an improved temperature model and an Early effect model based on depletion charge (MCANDREW et al., 2003). With the VBIC compact model release, the need of a current source to model the avalanche breakdown has been extinguished.

One of the first macromodels using advanced MOSFET and BJT compact models for ESD circuit simulation was proposed by Zhou et al. (2005b). Figure 3.4 illustrates the macromodel equivalent circuit, consisting of standard SPICE components only. The MOSFET is modeled with BSIM3v3 and the intrinsic BJT is modeled with VBIC. R_{sub} represents the substrate resistance, R_S and R_D represent the source and the drain resistances, respectively. In this case, BSIM3v3 and VBIC parameters were extracted using a regular methodology, while R_{sub} was extracted from snapback curves of GGNMOS structures measured with TLP.

Figure 3.4: Modeling the avalanche breakdown with BSIM3 and VBIC compact models.



Source: the author

The resulting drain-to-substrate current in the macromodel proposed by Zhou *et al.* is mainly composed by the substrate current of BSIM3v3 and the collector-base avalanche current

of VBIC. The substrate current in BSIM3v3, which includes the impact ionization effect, is modeled by (LIU et al., 1999)

$$I_{sub} = \frac{\alpha_0 + \alpha_1 L_{eff}}{L_{eff}}(V_{DS} - V_{DS_{eff}}) \exp\left(-\frac{\beta_0}{V_{DS} - V_{DS_{eff}}}\right) \cdot \frac{I_{DS_0}}{1 + \frac{R_{DS} I_{DS_0}}{V_{DS_{eff}}}} \left(1 + \frac{V_{DS} - V_{DS_{eff}}}{V_A}\right) \tag{3.13}$$

where α_0 is the first parameter of impact ionization current (fitting parameter), α_1 is the channel length scaling parameter of impact ionization current (fitting parameter), L_{eff} is the effective channel length of the MOSFET, V_{DS} is the drain-to-source voltage, $V_{DS_{eff}}$ is the effective drain-to-source voltage, which is modeled as a function that follows V_{DS} in the linear region and tends to $V_{D_{sat}}$ in the saturation region, β_0 is a V_{DS} dependent parameter of impact ionization current (fitting parameter), I_{DS_0} is the drain-to-source current without considering the impact ionization, R_{DS} is the parasitic drain-to-source resistance and V_A is the Early voltage due to the substrate current. It is possible to notice in Equation (3.13) the exponential dependency of the substrate current on the drain-to-source voltage.

The avalanche current in the collector-base junction is, in turn, modeled in VBIC by (MCANDREW et al., 1996)

$$I_{avl} = (I_{cc} - I_{bc}) A_{VC_1} (P_C - V_{bci}) \exp\left[-A_{VC_2}(P_C - V_{bci})^{ME-1}\right] \tag{3.14}$$

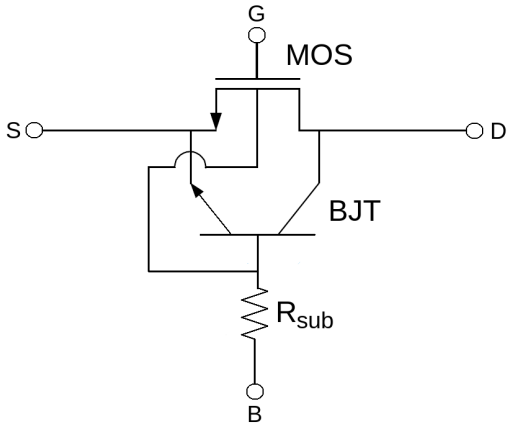
where I_{cc} is the forward transport current (or the collector current without the avalanche effect), I_{bc} is the base-collector component of the base current, A_{VC_1} and A_{VC_2} are fitting parameters, P_C is the built-in potential of the base-collector junction, V_{bci} is the intrinsic voltage across the base-collector junction and ME is the grading coefficient of the base-emitter junction.

The most significant contribution of the macromodel proposed by Zhou *et al.* is its straightforward CAD implementation using standard SPICE components only. It improves the macromodel availability since there is no need of implementing special equations directly in the circuit simulator. When compared to previous models discussed in Section 3.1, the improved BSIM3 and VBIC compact models reduce the convergence issues. VBIC also includes other relevant effects for ESD circuit simulation such as a self-heating model, an improved temperature model and an improved depletion and diffusion charge model. One of the major drawbacks of this macromodel is that VBIC is restricted to model the avalanche effect at low current densities only (weak avalanche) (BERKNER, 2002).

3.3 Snapback modeling with BSIM4 and MEXTRAM-based compact models

After introducing a macromodel using BSIM3v3 and VBIC compact models (detailed in Section 3.2), Zhou, Hajjar and Lisiak (2006) introduced a new macromodel using BSIM4 and a proprietary MEXTRAM-based BJT compact model instead. Figure 3.5 illustrates the equivalent circuit, which is similar to the equivalent circuit of Figure 3.4 but omitting the R_S and R_D resistances. In this case, the parameters of the BSIM4 and the MEXTRAM-based compact models were also extracted using a regular methodology and R_{sub} was also extracted from the snapback curves of GGNMOS structures measured with TLP.

Figure 3.5: Snapback modeling with BSIM4 and MEXTRAM-like compact models.



Source: the author

The total substrate current of this macromodel is the sum of the substrate terminal current in BSIM4 and the avalanche current in the MEXTRAM-based BJT compact model. The substrate current due to impact ionization in BSIM4 is the same as that in BSIM3v3, described by Equation (3.13) (MORSHED et al., 2011). What is new in BSIM4 and relevant to ESD circuit simulation is that, besides the substrate current due to impact ionization, the total substrate terminal current also includes the gate-induced drain leakage current I_{GIDL} and the gate-induced source leakage current I_{GISL}. I_{GIDL} in BSIM4 (using the compact model default settings) is modeled by (MORSHED et al., 2011)

$$I_{GIDL} = AGIDL\, W_{eff_{CJ}}\, N_f \frac{V_{DS} - V_{GSe} - EGIDL}{3\, T_{oxe}} \cdot \exp\left(-\frac{3\, T_{oxe}\, BGIDL}{V_{DS} - V_{GSe} - EGIDL}\right) \frac{{V_{DB}}^3}{CGIDL + {V_{DB}}^3} \tag{3.15}$$

where $AGIDL$ is the pre-exponential coefficient of I_{GIDL} (fitting parameter), $W_{eff_{CJ}}$ is the effective drain diffusion width for parasitics modeling, N_f is the number of device fingers,

V_{DS} is the drain-to-source voltage, V_{GSe} is the effective gate-to-source voltage (where "effective" means accounting for the poly depletion effect), $EGIDL$ is a model parameter for the band bending of I_{GIDL} (fitting parameter), T_{oxe} is the effective oxide thickness, $BGIDL$ is the exponential coefficient of I_{GIDL} (fitting parameter), V_{DB} is the drain-to-substrate voltage and $CGIDL$ accounts for the body-bias dependence of I_{GIDL} (fitting parameter). The gate-induced drain leakage current improves the accuracy of the total substrate current model, especially for $V_{GS} = 0$, which is the case when modeling a GGNMOS structure. The equation of I_{GISL} is similar to Equation (3.15), but changing "drain" to "source" (and vice versa) in all its parameters.

For the operation regime before the avalanche breakdown, I_{sub} described by Equation (3.13) and I_{GIDL} described by Equation (3.15) are dominant components of the total substrate current of the macromodel. When the avalanche breakdown begins, the BJT avalanche current becomes dominant. The avalanche current in the MEXTRAM-based BJT compact model is given by (ZHOU; HAJJAR; LISIAK, 2006)

$$I_{avl} = I_c \frac{KAVL(VJC - V_{bci})}{1 - MJC} \exp\left[-AVL(VJC - V_{bci})^{MJC-1}\right] \tag{3.16}$$

where I_c is the collector current without the avalanche effect (similar to $I_{cc} - I_{bc}$ in VBIC's Equation (3.14)), $KAVL$ and AVL are fitting parameters (similar to A_{VC_1} and A_{VC_2} in VBIC's Equation (3.14)), VJC is the built-in potential of the base-collector junction (equivalent to P_C in VBIC's Equation (3.14)), V_{bci} is the intrinsic voltage across the base-collector junction (equivalent to V_{bci} in VBIC's Equation (3.14)), MJC is the grading coefficient of the base-emitter junction (equivalent to ME in VBIC's Equation (3.14)). Given the similarities of Equation (3.16) to Equation (3.14), the MEXTRAM-based BJT model of Zhou, Hajjar and Lisiak should also produce similar results to VBIC model, i.e. a weak avalanche effect. Actually, the true MEXTRAM compact model is able to model both weak and high-current avalanche effect, including the snapback behavior (TOORN; PAASSCHENS; KLOOSTERMAN, 2012), resulting in a much more complicated avalanche model[1] than Equation (3.16).

The macromodel proposed by Zhou, Hajjar and Lisiak using a BSIM4 compact model for the MOSFET and a MEXTRAM-based compact model for the intrinsic BJT has a straightforward implementation. Compared to BSIM3, improvements in the total substrate terminal current model of BSIM4 include the gate-induced drain leakage and the gate-induced source leakage currents (in addition to the substrate current due to impact ionization, already present

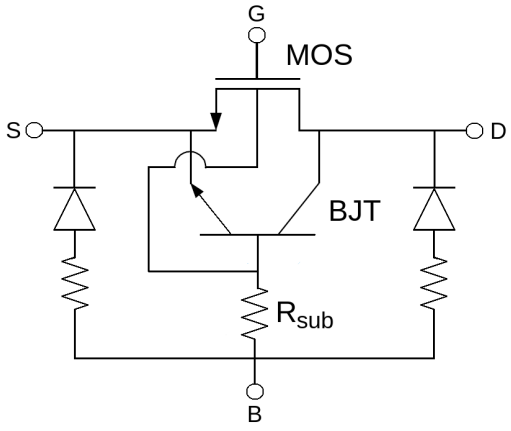
[1]The avalanche model of MEXTRAM is detailed in Appendix A.

in BSIM3). One of the major drawbacks of this macromodel is that the MEXTRAM-based compact model is proprietary and not widely available in circuit simulators. Also, its avalanche model described by Equation (3.16) is too similar to the weak avalanche model of VBIC described by Equation (3.14), which does not guarantee the proprietary MEXTRAM-based compact model is really able to simulate the high-current avalanche and snapback effects.

3.4 Modeling the impedance for negative drain-to-source voltage

ESD testing procedures usually consist not only in applying positive voltage stresses to each pair of the circuit pins, but also in applying the equivalent negative voltage stresses. Zhou et al. (2007) introduced a macromodel including the impedance for negative drain-to-source voltage. According to Zhou et al. (2007), to properly model the snapback characteristics of a typical NMOS-based ESD protection device, such as a GGNMOS, R_{sub} must be in the range of several hundred Ohms. However, when the drain-to-substrate junction is forward-biased, which is the case for a negative drain-to-substrate voltage, the typical measured resistance is only a few Ohms. Also, only a portion of the drain-to-substrate and source-to-substrate junctions contributes to the intrinsic BJT operation (AMERASEKERA et al., 1996). To account for such partial contribution, the drain-to-substrate and source-to-substrate junctions were partitioned into two separate diodes in the macromodel. Figure 3.6 illustrates the resulting equivalent circuit.

Figure 3.6: Modeling the impedance for negative V_{DS} with diodes and resistors.

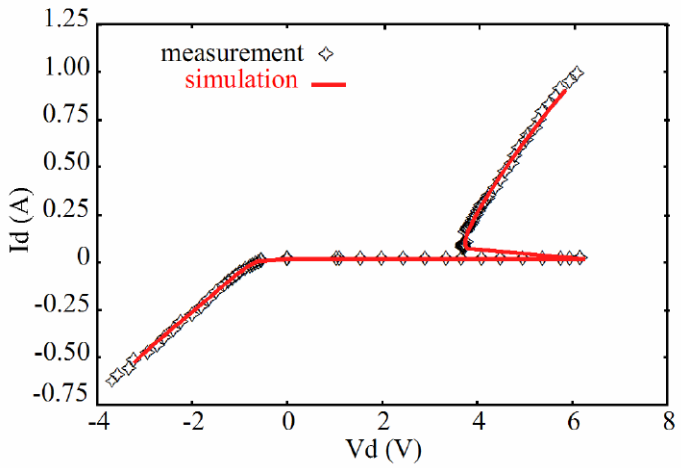


Source: the author

The NMOS transistor in Figure 3.6 is modeled with BSIM4 and its intrinsic bipolar transistor is modeled with the same proprietary MEXTRAM-based BJT compact model discussed

in Section 3.3. In a typical NMOS-based ESD protection device, the source and the substrate terminals are shorted together. In this case, when $V_{DS} > 0$, the influence of the two diodes of the macromodel is negligible. However, when $V_{DS} < 0$, the diode between the drain and the substrate becomes dominant, providing a low-impedance path to the current. Figure 3.7 shows the results obtained by Zhou *et al.* for both positive and negative drain-to-source voltages.

Figure 3.7: Comparison between simulation and measurement data of a GGNMOS device.



Source: Zhou and Hajjar (2008)

The main contribution of the macromodel presented by Zhou et al. (2007) is to introduce the modeling of the impedance for a negative drain-to-source voltage. Such feature was not considered in previous macromodels. However, the drawbacks of the proprietary MEXTRAM-based compact model are the same discussed in Section 3.3. It is not widely available in circuit simulators and its avalanche model is too similar to the weak avalanche model of VBIC, which does not guarantee its ability to effectively simulate the high-current avalanche and snapback effects.

4 AN OPTIMIZATION-BASED METHODOLOGY FOR MACROMODELS DESIGN

As discussed in Chapter 3, the first MOSFET macromodels for ESD circuit simulation were implemented using a voltage-dependent current source to include the avalanche breakdown effect, and to overcome a limitation of the MOSFET and BJT compact models available at that time. Such implementation required, in some cases, deep and complicated modifications in the compact models and in the circuit simulator, which is definitely not a straightforward task.

The usage of advanced compact models, such as BSIM4 and MEXTRAM, eliminated the need of using a voltage-dependent current source in the macromodel implementation. It also eliminated the need of modifying the circuit simulator code, which is usually not available to integrated circuit designers, especially in the case of commercial CAD tools. The task of implementing a macromodel became indeed more straightforward. However, the process parameters commonly extracted by the device manufacturers for both MOSFET and BJT compact models are typically not suitable to reproduce the corresponding devices operation under high-current and high-voltage conditions such as during an ESD event. In this context, the parameters of the macromodels available in the literature still need to be properly extracted or manually modified to fit the measured $I-V$ characteristics of the corresponding ESD protection device.

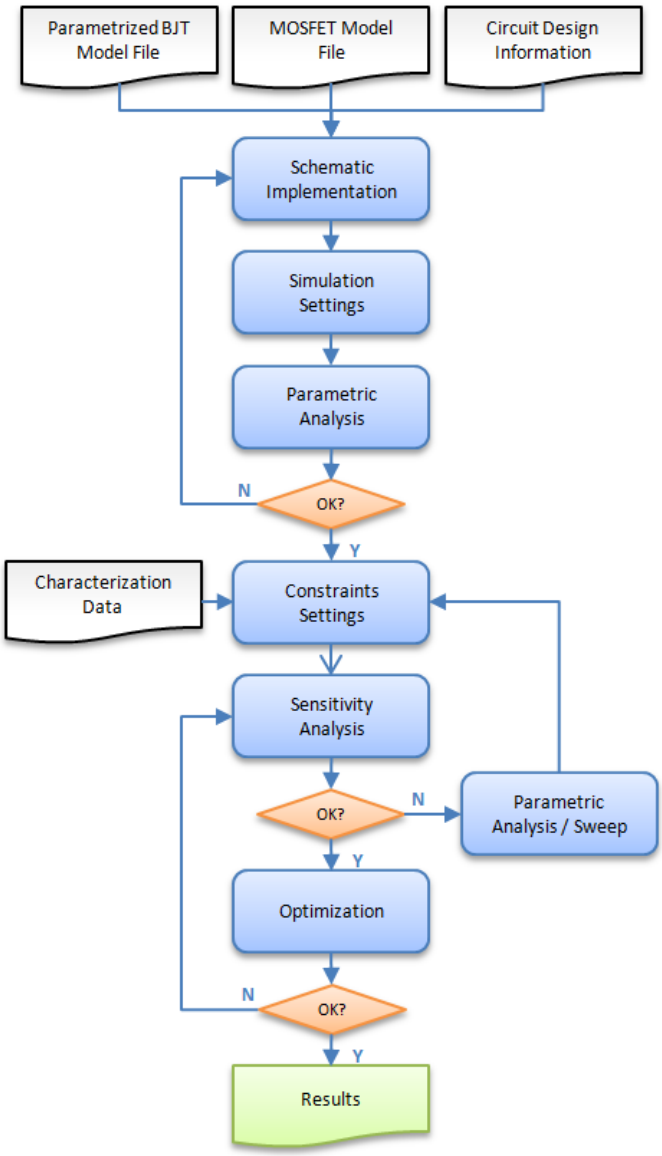
This chapter describes a methodology based on parametric analysis and optimization of MOSFET macromodels for ESD circuit simulation. The key idea behind the methodology presented here is the usage of stochastic-based and deterministic-based optimization algorithms to fit the $I-V$ characteristics of a MOSFET macromodel to the measured data of its corresponding ESD protection device. Such approach eliminates the need of any additional extraction of specific parameters for the macromodel implementation. As a case of study, a macromodel corresponding to a GGNMOS protection device (introduced in Chapter 1) will be optimized using the methodology.

4.1 Methodology overview

The idea of using optimization algorithms to fit the $I-V$ characteristics of a MOSFET macromodel to the measured data of its corresponding ESD protection device is actually not new. Napravnik and Jakovenko (2012) were probably the first to publish the usage of a differential evolutionary optimization algorithm to fit the $I-V$ characteristics of the macromodel presented by Amerasekera et al. (1996) and improved by Lim et al. (1997) to measured $I-V$

data. The MOSFET element of the macromodel adopted was implemented using a standard EKV compact model from a GPDK90 library, while the BJT, the voltage-dependent current source, and the R_{sub} equations were implemented in Verilog-A hardware description language. The BJT implementation was based on the VBIC compact model. The differential evolutionary optimization algorithm was developed and implemented by the authors themselves.

Figure 4.1: Flow chart of the methodology presented in this work.



Source: the author

Similarly to the work of Napravnik and Jakovenko (2012), the methodology presented in this work also focuses on the usage of optimization algorithms to fit the $I-V$ characteristics of MOSFET macromodels to measured $I-V$ data. However, the MOSFET macromodel

implemented in this work consists in standard SPICE elements only, instead of mixing hardware description language and SPICE compact models. The methodology includes a step of parametric analysis, which helps to understand the effect of each parameter and to improve the macromodel itself. A sensitivity analysis step helps to evaluate the impact of each parameter on target performances during the optimization steps. Last, two different optimization algorithms are adopted to fit the MOSFET macromodel $I-V$ characteristics to measured data. All tools and algorithms are commercially available, making the methodology as well as its implementation widely available to integrated circuit designers.

Figure 4.1 illustrates the optimization-based methodology presented in this work, whose major goal is to fit the $I-V$ characteristics of a MOSFET macromodel to the measured data of its corresponding ESD protection device. **Characterization data**, **circuit design information**, **MOSFET model file** and **parametrized BJT model file** are inputs in the flow chart of Figure 4.1 and are detailed in Section 4.2. The **schematic implementation**, **simulation settings** and **parametric analysis** steps are described in Section 4.3, while the **constraints settings** and **sensitivity analysis** steps are explained in Section 4.4. The **optimization** step, which is the core of the methodology, is detailed in Section 4.5. Finally, Section 4.6 illustrates a macromodel application example.

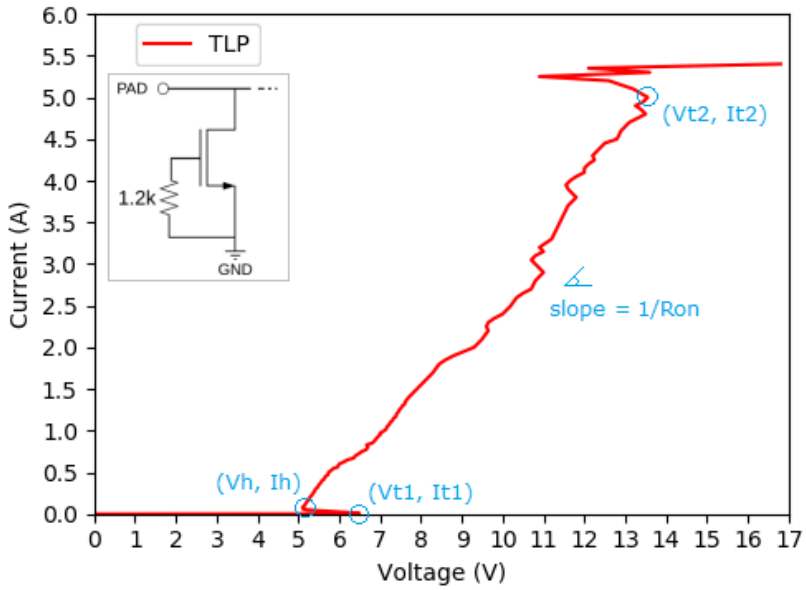
4.2 Characterization data, circuit design information and model files

The most essential input to the flow chart of Figure 4.1 is the $I-V$ characteristics (or **characterization data**) of the target ESD protection device. The **circuit design information** of the protection device is important to define which should be the equivalent macromodel configuration, which should be the MOSFET sizing in the macromodel and which MOSFET compact model should be adopted. The **MOSFET model file** in the flow chart of Figure 4.1 is a regular compact model with the standard process parameters provided by the foundry. On the other hand, the process parameters provided by the foundry for the BJT compact models are typically not suitable to describe the characteristics of the intrinsic lateral bipolar transistor of a MOSFET, as will be explained in this section. Therefore, these parameters are initially defined as variables in the **parametrized BJT model file** to be part of the optimization step in the flow.

The characterization data is the result of a testing procedure applied to the target ESD protection device to obtain its $I-V$ characteristics. The default method for characterizing the behavior of devices under ESD stress in the semiconductor industry is the Transmission-Line Pulse (TLP) testing, as already mentioned in Chapter 1. Figure 4.2 shows the TLP testing

results and the schematic of the GGNMOS device adopted as a case of study in this work. The snapback triggering voltage V_{t1} and triggering current I_{t1} correspond to 6.49 V and 9 mA, respectively. The device holding voltage V_h and holding current I_h are respectively 5.1 V and 80 mA. The second breakdown triggering voltage V_{t2} and triggering current I_{t2} correspond to 13.9 V and 5 A, respectively. R_{on} denotes the GGNMOS on-resistance, which is equivalent to the inverse of the slope of the curve between the holding point and the second breakdown triggering point. In this case, R_{on} is approximately 1.79 Ω.

Figure 4.2: $I-V$ characteristics and schematic of the reference GGNMOS ESD protection device.



Source: the author

Another input in the flow chart of Figure 4.1 is the circuit design information, which is important to define its equivalent macromodel configuration, the MOSFET sizing in the macromodel and which MOSFET compact model should be adopted. The schematic of the GGNMOS device adopted as a case of study is illustrated in Figure 4.2. The 1.2 kΩ gate-biasing resistor aims to reduce the device triggering voltage. The total width W of the NMOS channel is 350 μm, divided into 14 fingers of width 25 μm each. The NMOS channel length L is 0.4 μm. The device was manufactured using a CMOS 0.18 μm process technology. Based on this information, an equivalent macromodel should include at least an NMOS transistor with the same sizing and number of fingers described above, a resistor of 1.2 kΩ and, of course, a BJT to model the intrinsic lateral bipolar transistor of the MOSFET.

The MOSFET model file is also an input in the flow chart of Figure 4.1. Considering the MOSFET operation under ESD conditions explained in Chapter 2 and the corresponding

macromodels described in Chapter 3, let us assume the NMOS transistor dominates the operation of the equivalent macromodel under low-current conditions while the BJT dominates the operation under high-current conditions. If the NMOS transistor is expected to operate mostly under typical conditions, the process parameters of the CMOS 0.18 μm technology provided by the device manufacturer should be valid to describe its behavior. For the case of study, these process parameters along with a BSIM4v5 compact model compose the MOSFET model file.

The last input of the methodology flow chart is a parametrized BJT model file. Usually, the IC foundries provide the process parameters of compact models for the design of regular BJT devices, which are expected to operate within typical conditions. However, under the high-voltage and high-current conditions, such as during an ESD event, these process parameters would hardly provide a reasonable simulation result. Moreover, the physical structure of the intrinsic lateral bipolar transistor in a MOSFET-based ESD protection device is different from a regular BJT. First, the MOSFET of a primary ESD protection circuit has a very large width when compared to core circuit devices. For this reason, its layout is often designed with multiple fingers. In this case, each finger will have its corresponding intrinsic lateral bipolar transistor. Each intrinsic BJT, in turn, will have specific base dimensions because each MOSFET finger will have specific distances between its source, drain and substrate terminals. The electrical behavior of such physical structure becomes even more complex considering the fact that the intrinsic lateral bipolar transistors corresponding to each finger may not uniformly turn on during an ESD event. Last, fabrication process techniques that aim to improve the ESD robustness of a MOSFET-based protection device, such as *silicide blocking mask*, *ESD implantation* and *shallow trench isolation*, could also increase the complexity of an intrinsic lateral bipolar transistor behavior or even of the equivalent MOSFET macromodel itself. However, these techniques and their corresponding impact on modeling are not covered in this work.

For all the reasons mentioned above, the parametrized BJT model file in the flow chart of Figure 4.1 includes variables instead of fixed values for the compact model parameters. In fact, as will be demonstrated in the following sections, the BJT compact model parameters are primarily responsible for the high-current $I-V$ characteristics of the macromodel and, thus, are the main target of the optimization step. The BJT compact model adopted for the case of study presented in this work is a MEXTRAM v504, which includes the most complete model for high-current and high-voltage operation when compared to other common compact models such as Gummel-Poon, VBIC and HICUM (BERKNER, 2002). The contents of the parametrized BJT model file created for this work are presented below:

```
1     model custom_mextram bjt504 type=npn level=504
```

```
2     + tref=m_tref dta=m_dta exmod=m_exmod exphi=m_exphi exavl=m_exavl
3     + exsub=m_exsub is=m_is ik=m_ik ver=m_ver vef=m_vef bf=m_bf ibf=m_ibf
4     + mlf=m_mlf xibi=m_xibi izeb=m_izeb nzeb=m_nzeb bri=m_bri ibr=m_ibr
5     + vlr=m_vlr xext=m_xext wavl=m_wavl vavl=m_vavl sfh=m_sfh re=m_re rbc=m_rbc
6     + rbv=m_rbv rcc=m_rcc rcblx=m_rcblx rcbli=m_rcbli rcv=m_rcv scrcv=m_scrcv
7     + ihc=m_ihc axi=m_axi cje=m_cje vde=m_vde pe=m_pe xcje=m_xcje cbeo=m_cbeo
8     + cjc=m_cjc vdc=m_vdc pc=m_pc xp=m_xp mc=m_mc xcjc=m_xcjc cbco=m_cbco
9     + mtau=m_mtau taue=m_taue taub=m_taub tepi=m_tepi taur=m_taur deg=m_deg
10    + xrec=m_xrec aqbo=m_aqbo ae=m_ae ab=m_ab dais=m_dais aepi=m_aepi aex=m_aex
11    + ac=m_ac acbl=m_acbl dvgbf=m_dvgbf dvgbr=m_dvgbr vgb=m_vgb vgc=m_vgc
12    + vgj=m_vgj vgzeb=m_vgzeb avgeb=m_avgeb tvgeb=m_tvgeb dvgte=m_dvgte af=m_af
13    + kf=m_kf kfn=m_kfn kavl=m_kavl iss=m_iss icss=m_icss iks=m_iks cjs=m_cjs
14    + vds=m_vds ps=m_ps vgs=m_vgs as=m_as asub=m_asub
```

Line 1 above defines *custom_mextram* as the model name, *bjt504* as the compact model for the circuit simulator (in this case, MEXTRAM), *npn* as the type of the BJT compact model and *504* as the compact model level. From line 2 to line 14, all the other parameters of the compact model are defined as variables. A short description of all parameters above can be found in Annex A.

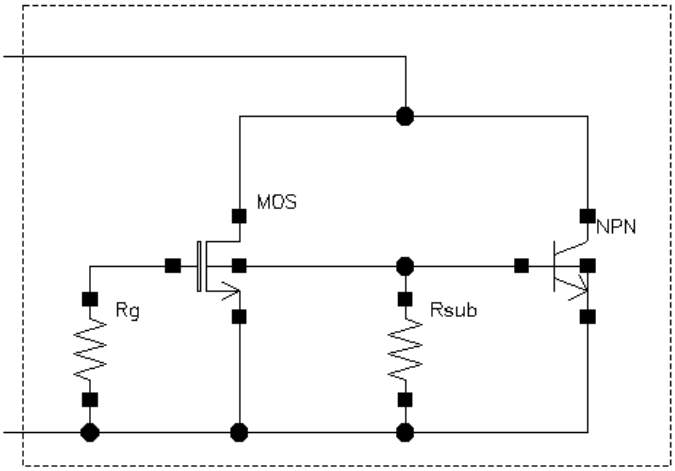
4.3 Schematic implementation, simulation settings and parametric analysis

This section describes the **schematic implementation**, the **simulation settings** and the **parametric analysis** steps of the methodology. As seen in Chapter 3, the equivalent macromodel of a GGNMOS protection device using only standard SPICE elements should include at least a MOSFET to describe the device low-current behavior and a BJT to describe the device high-current behavior. A MOSFET compact model with substrate current modeling feature is desirable, while a BJT with avalanche and snapback modeling features is essential. The compact models BSIM4v5 and MEXTRAM v504 include such features. Equation (3.13) in Chapter 3 describes the substrate current behavior due to impact ionization effect in BSIM4v5 (same equation of BSIM3v3). Its total substrate terminal current also includes the gate-induced drain leakage current and the gate-induced source leakage current effects described by Equation (3.15) also in Chapter 3. The avalanche current model of MEXTRAM v504 is detailed in Appendix A. The usage of resistors can also increase the macromodel flexibility.

Figure 4.3 illustrates the initial macromodel schematic, implemented with Cadence Virtuoso Schematic Editor. R_g is the 1.2 kΩ resistor of the target GGNMOS device, according to the circuit design information input described in Section 4.2. The arrangement of the remaining

elements (MOS, NPN and R_{sub}) is based on the work of Zhou, Hajjar and Lisiak (2006), detailed in Chapter 3. The MOSFET element uses a BSIM4v5 compact model with fixed CMOS 0.18 μm technology process parameters. Its channel length L was set to 0.4 μm and its total channel width W was set to 350 μm, divided into 14 fingers of 25 μm each. The BJT element uses a MEXTRAM v504 compact model with all its parameters (including the process-related ones) set as variables in the parametrized BJT model file, as explained in Section 4.2. R_{sub} element was also initially set as a variable parameter in the schematic.

Figure 4.3: Initial schematic implementation of the GGNMOS macromodel.



Source: the author

The **simulation settings** were defined in Cadence Virtuoso Analog Design Environment (ADE). All MEXTRAM v504 variable parameters were initially set to their compact model default values (see Annex B), except the $EXAVL$ flag which was set to "1" to enable the extended avalanche model explained in Appendix A. A current source was defined as an input stimuli for the macromodel and it was configured to range from 0.1 nA to 5 A in a DC sweep analysis[1]. 5 A is actually the maximum current before the target GGNMOS device enters the second breakdown (see Figure 4.2). Since MEXTRAM v504 does not include the second breakdown effect, it does not make sense trying to fit the macromodel $I-V$ characteristics above 5 A in this case. The circuit simulator was also configured to properly measure V_{t1}, I_{t1}, V_h, I_h and V_{t2}. The triggering voltage V_{t1} and current I_{t1} were set to be measured in the *first point* where the derivative of the macromodel voltage response to the input current is zero, while the holding voltage V_h and current I_h were set to be measured in the *second point* where the derivative of

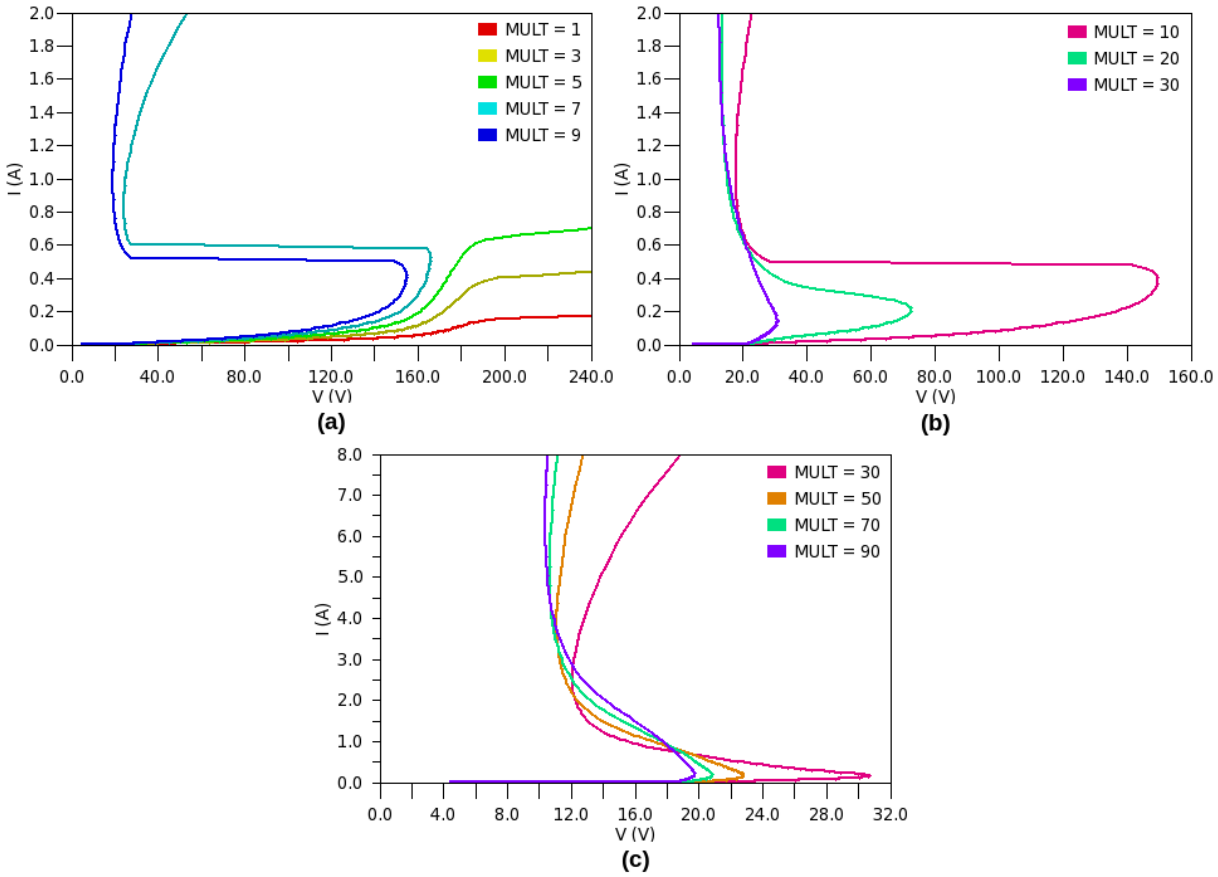
[1] A DC analysis (quiescent domain) should be suitable to compare the simulation results with TLP results since such testing method consists in measuring quasi-static $I-V$ points in a predefined stable part of the input pulse.

the macromodel voltage response to the input current is zero. The second breakdown triggering voltage V_{t2} is measured when the input current sweep reaches its maximum. The second breakdown current I_{t2} does not need to be measured since it is the maximum sweep current itself (5 A). The on-resistance R_{on} also does not need to be measured since it is the inverse of the slope of the curve between the resulting holding point (V_h, I_h) and the second breakdown triggering point (V_{t2}, I_{t2}). If these two points are properly optimized, R_{on} is also optimized. All these measurements are essential to the optimization step, as will be detailed in Sections 4.4 and 4.5.

The **parametric analysis** step consists in changing or sweeping a variable parameter to evaluate its effect on the circuit simulation results. This step was performed with ADE and Cadence Spectre Circuit Simulator. First of all, it is important to remind that, differently from the MOS transistor of the macromodel in Figure 4.3, the parallel *npn* transistor sizing is unknown and difficult to be determined, as explained in Section 4.2. Also, geometric scaling is not part of the MEXTRAM compact model (TOORN; PAASSCHENS; KLOOSTERMAN, 2012). The only parameter of the compact model able to provide some sort of geometric scaling is the multiplication factor parameter $MULT$, which gives the possibility of putting several transistors in parallel. It means that all currents, charges and noise current sources are multiplied by the value of $MULT$. On the other hand, all resistances of the compact model are divided by the value of $MULT$, while its flicker-noise coefficients scale with the multiplication factor according to specific equations. MEXTRAM allows a non-integer value of $MULT$ to increase the compact model flexibility. Given its impact on the model behavior, this parameter was firstly evaluated in the parametric analysis step.

Figure 4.4 shows the effect of $MULT$ on the macromodel $I-V$ characteristics. In (a), it is possible to observe that the avalanche and the snapback effects are only triggered when $MULT$ is greater than or equal to 7, for the initial macromodel parameter configuration. However, when the parameter is equal to 7, the triggering voltage V_{t1} is very high (more than 160 V). In (b), it is possible to observe that V_{t1} significantly decreases as $MULT$ increases to a few dozens. The holding current I_h, on the other hand, significantly increases with $MULT$, as can be observed in (c), where the input current source was exceptionally set to range from 0.1 nA to 8 A for a better visualization of the snapback curves. Any value of multiplication factor parameter can be chosen for further optimization of the macromodel as long as it triggers the snapback behavior of the macromodel. For this case of study, $MULT$ was set to a fixed value of 70, an intermediate value between 50 and 90, which provides one of the lowest triggering voltage of this initial parametric analysis, but still preserving a reasonable snapback characteristic.

Figure 4.4: Effect of $MULT$ parameter on the macromodel $I-V$ characteristics.

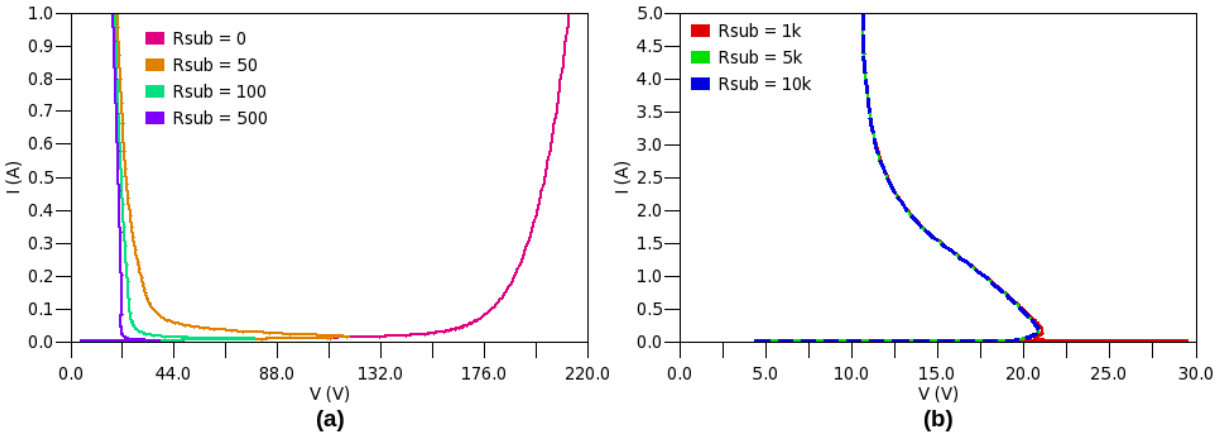


Source: the author

The second parameter subjected to parametric analysis was R_{sub}. Figure 4.5 (a) shows that for R_{sub} equal to zero (short circuit) there is no snapback triggering point, differently from when it assumes a value greater than zero. One could infer from Figure 4.5 (a) that the higher is the substrate resistance the lower is the triggering voltage. However, Figure 4.5 (b) shows a curious behavior. When R_{sub} is greater than a certain value (about 5 kΩ), the $I-V$ curve "saturates" with the same shape as if R_{sub} was removed from the macromodel. Also, the voltage peak (supposed triggering voltage) observed for lower R_{sub} values, such as 1 kΩ, disappear. In this sense, Figure 4.6 shows that such voltage peak occurs before the BJT avalanche begins to rise. Therefore, this voltage cannot be considered the triggering voltage V_{t1}. Its origin is probably in a specific biasing condition of the macromodel circuit elements and is certainly not related to the BJT avalanche breakdown. According to Zhou et al. (2007), R_{sub} should be in the range of several hundred ohms, but in this case such range would lead to incorrect V_{t1} measurements. Consequently, R_{sub} was removed from the macromodel. The schematic was then updated in the schematic implementation step as well as the corresponding modifications

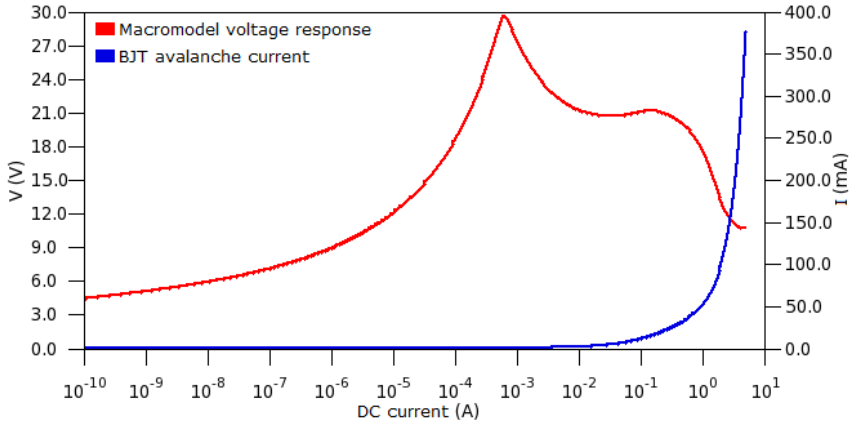
were performed in the simulation settings step of the methodology (closing the first *loop* in the flow chart of Figure 4.1).

Figure 4.5: Effect of R_{sub} (defined in Ω) on the macromodel $I-V$ characteristics.



Source: the author

Figure 4.6: Macromodel resulting voltage compared with the BJT avalanche current for R_{sub} = 1 kΩ.

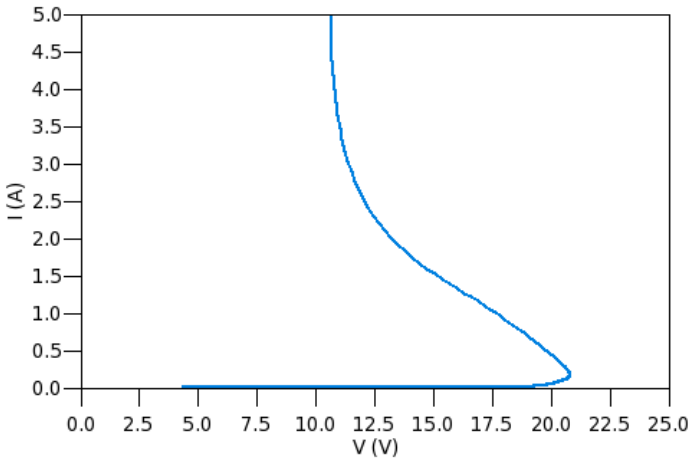


Source: the author

Figure 4.7 shows the resulting $I-V$ characteristics after the initial parametric analysis of $MULT$ and R_{sub}, for an input current ranging from 0.1 nA to 5 A. As defined in the simulation step, the the holding point (V_h, I_h) is measured in the second point where the derivative of the macromodel voltage response to the input current is zero. However, it is possible to observe in Figure 4.7 that such point can be measured more than once when the input current becomes higher than about 4.5 A. A wrong measurement of the holding point (or any other point of interest) can lead to undesired results in the optimization step. One easy way to overcome this

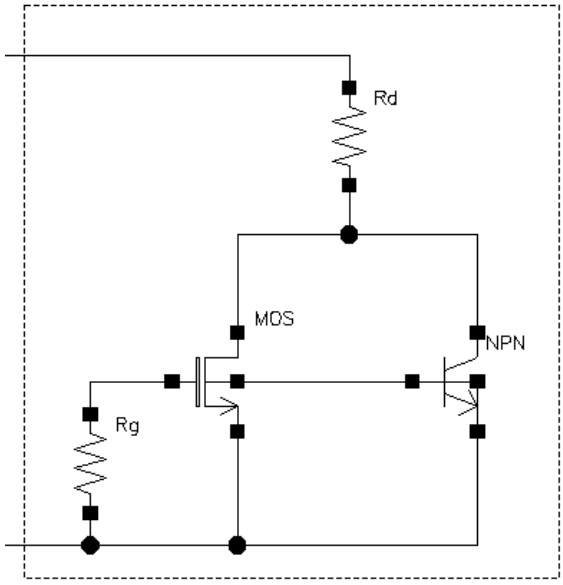
problem is intentionally increasing the on-resistance of the macromodel by adding a resistor element in series with the MOSFET and the BJT devices, as illustrated in Figure 4.8.

Figure 4.7: Macromodel $I-V$ characteristics for $MULT = 70$ and without R_{sub}.



Source: the author

Figure 4.8: Macromodel schematic without R_{sub} and with an additional series resistor R_d.

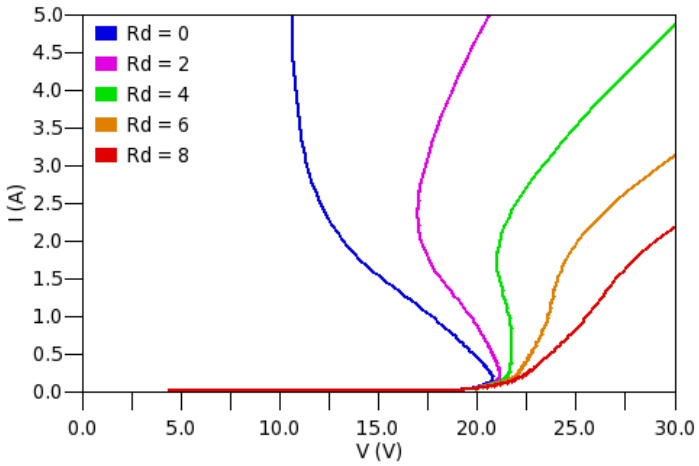


Source: the author

The on-resistance R_{on} of an ESD protection device should be as low as possible to provide a low-impedance path to the ESD current, preventing damage to the core circuit. R_{on} is typically a few ohms, and so should be the additional resistance of R_d in the macromodel. Figure 4.9 show a parametric analysis of R_d, which helps to define its initial value. It is possible to notice that when R_d is about 2 Ω, the S-shape of the $I-V$ curve is well preserved. In this case,

both the triggering and the holding points are clear and can be accurately measured. Therefore, R_d was initially set to 2 Ω for the next steps of the methodology.

Figure 4.9: Effect of R_d (defined in Ω) on the macromodel $I-V$ characteristics.



Source: the author

The schematic implementation, the simulation settings and the parametric analysis steps detailed in this section are important not only to define an initial parameter configuration for the optimization step, but also to improve the macromodel itself. The substrate resistance R_{sub}, included in all macromodels described in Chapter 3, was not necessary in the macromodel configuration presented in this section. Also, it was demonstrated that the S-shape of the $I-V$ curve initially obtained with the parametric analysis of $MULT$ can be improved by simply adding a series resistor to the macromodel, increasing its on-resistance and providing more accurate measurements of the triggering and the holding points.

4.4 Constraints settings and sensitivity analysis

This section describes the **constraints settings** and the **sensitivity analysis** steps of the methodology. During the optimization step, some parameters of the macromodel illustrated in Figure 4.8, such as the NMOS transistor parameters and the R_g resistance, are fixed (constant). Other parameters can be variable, such as the R_d resistance and part of the MEXTRAM compact model parameters. The constraints settings define which parameters are constant and which parameters are variable. In this step, an initial value is assigned to each variable parameter along with its corresponding upper and lower limits of variation for the optimization step. Also, the objective functions (or target performances) for the optimization algorithms are set.

The sensitivity analysis step, on the other hand, consists in evaluating how much each variable parameter affects each target performance. This step is also important to identify whether the upper and lower limits of variation for each parameter are properly set.

The constraints settings were implemented using MunEDA WiCkeD Constraint Editor. The resistance of R_g element and all the parameters of the MOS element (design parameters and compact model parameters) were set to a constant value for all steps of the methodology. The resistance of R_d element was initially set to a fixed value of 2 Ω, but it can become a variable parameter during the optimization steps to fit the macromodel on-resistance into the $I-V$ characteristics of the GGNMOS device if necessary. Part of the MEXTRAM compact model parameters were set to a constant value and part of them were configured as variables.

Table 4.1: List of constant and variable parameters defined for the MEXTRAM compact model.

General description	*Constant parameters*	*Variable parameters*
General parameters and flags	$MULT$, $LEVEL$, $EXMOD$, $EXPHI$, $EXAVL$, $EXSUB$	T_{ref}, DTA
Current parameters of the basic model	None	I_s, I_k, V_{er}, V_{ef}, β_f, I_{Bf}, m_{Lf}, XI_{B_1}, I_{zEB}, N_{zEB}, β_{ri}, I_{Br}, V_{Lr}, X_{ext}
Parameters of the avalanche model	None	W_{avl}, V_{avl}, S_{fH}
Resistances and epilayer parameters	None	R_E, R_{Bc}, R_{Bv}, R_{Cc}, R_{Cblx}, R_{Cbli}, R_{Cv}, SCR_{Cv}, I_{hc}, a_{x_i}
Depletion capacitances	C_{j_E}, XC_{j_E}, C_{BEO}, C_{j_C}, XC_{j_C}, C_{BCO}	V_{d_E}, p_E, V_{d_C}, p_C, X_p, m_C
Transit times	m_τ, τ_E, τ_B, τ_{epi}, τ_R	None
Parameters for the SiGe model features	None	dE_g, X_{rec}
Temperature model (mobility exponents and bandgap voltages)	$dV_{g\tau_E}$	$A_{Q_{B0}}$, A_E, A_B, A_{epi}, A_{ex}, A_C, A_{Cbl}, dA_{I_s}, $dV_{g\beta f}$, $dV_{g\beta r}$, V_{g_B}, V_{g_C}, V_{g_j}, V_{gZEB}, A_{VgEB}, T_{VgEB}
Noise parameters	A_f, K_f, K_{fN}, K_{avl}	None
Parameters specific for the four-terminal device	C_{j_S}, V_{d_S}, p_S	I_{Ss}, I_{CSs}, I_{ks}, V_{g_S}, A_S, A_{sub}

Source: the author

Table 4.1 presents the constant and variable parameters of MEXTRAM compact model defined in the constraints settings step. A short description of each parameter can be found

in Annex A. The $MULT$ parameter was set to 70 based on the parametric analysis results described in Section 4.3. $EXAVL$ flag was set to "1" to enable the extended modeling of the avalanche current detailed in Appendix A. All other parameters, either constant or variable, were initially set to their compact model default values, which can be found in Annex B. The constant parameters for depletion capacitances, transit times, temperature, noise and four-terminal device of Table 4.1 are actually not used in the DC model of MEXTRAM (TOORN; PAASSCHENS; KLOOSTERMAN, 2012), however they were also set to their compact model default values to avoid any unnecessary simulation issue. Given the amount of parameters, the upper and lower limits of variation defined in the constraints settings along with the initial value of both constant and variable parameters can be found in Appendix B. The target performances, i.e. the objective functions for the optimization algorithms are the points V_{t1}, I_{t1}, V_h, I_h and V_{t2} (I_{t2} is the maximum sweep current itself, 5 A). Their target values were set based on the characterization data detailed in Section 4.2.

The sensitivity analysis step basically consists in evaluating how much each variable parameter affects each target performance. It also helps to identify whether the upper and lower limits of variation for the parameters are properly set. In this work, this step was performed using MunEDA WiCkeD Sensitivity Analysis along with Cadence Spectre Circuit Simulator. MunEDA algorithms are proprietary, however a method to calculate the sensitivities is available, for example, in the work of Gomez (2012). Let $f(x)$ be an objective function (or target performance) with x as a variable parameter. A small change ∂x results in an objective function difference $\partial f = f(x + \partial x) - f(x)$. The parameter sensitivity S_x^f can then be calculated by

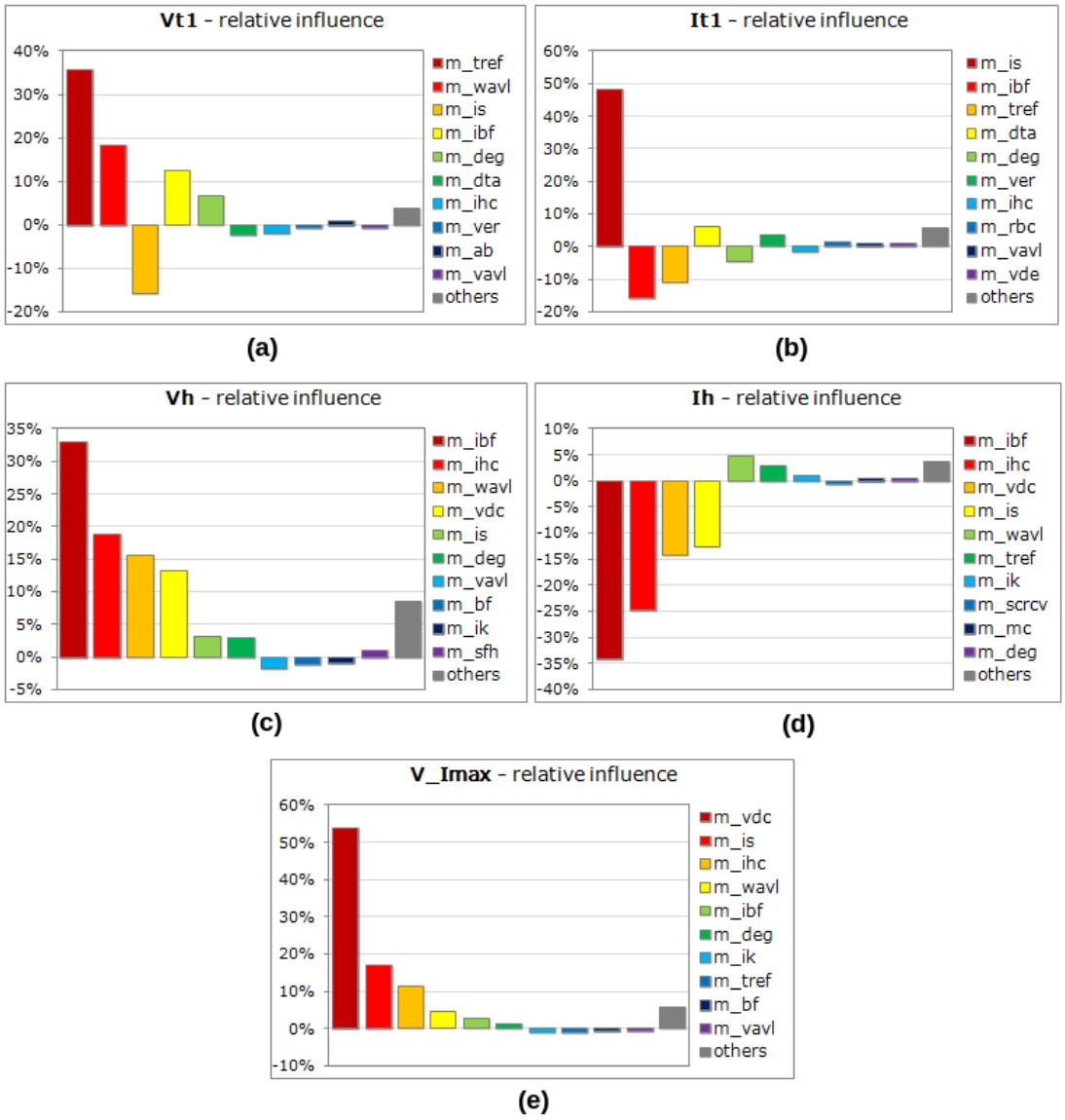
$$S_x^f = \frac{\partial f}{\partial x} \tag{4.1}$$

If the sensitivity of a parameter cannot be calculated, probably either the initial value of the parameter or the corresponding lower and upper limits of variation are not properly set in relation to the other parameters settings. An error in sensitivity calculation will be exemplified in Section 4.5.

Figure 4.10 illustrates the sensitivity analysis for the initial values of the macromodel parameters, their upper and lower limits of variation and the corresponding target performances defined in the constraints settings step. V_{t2} is denoted by $V_{I_{max}}$ in Figure 4.10 and in most of the illustrations of this chapter. The notation of the parameters in Figure 4.10 is the same of the parametrized BJT model file described in Section 4.2. Considering the five parameters with major relative influence on each target performance, it is possible to identify the general temperature parameters T_{ref} and DTA, the collector-emitter saturation current I_s, the saturation

current of the non-ideal forward base current I_{Bf}, the epilayer thickness of the avalanche model W_{avl}, the critical current for velocity saturation in the epilayer I_{hc}, the collector-base diffusion voltage V_{dC} and the bandgap difference over the base dE_g. It does not mean, however, that these parameters are the main responsible for the avalanche and snapback characteristics of the MEXTRAM compact model. In fact, the sensitivity analysis strongly depends on the upper and lower limits of variation for the parameters. The small change ∂x previously mentioned is proportional to how much a parameter is allowed to change. Therefore, the sensitivity analysis results of Figure 4.10 actually reflect the initial constraints settings and not necessarily the absolute impact of each parameter on the model behavior.

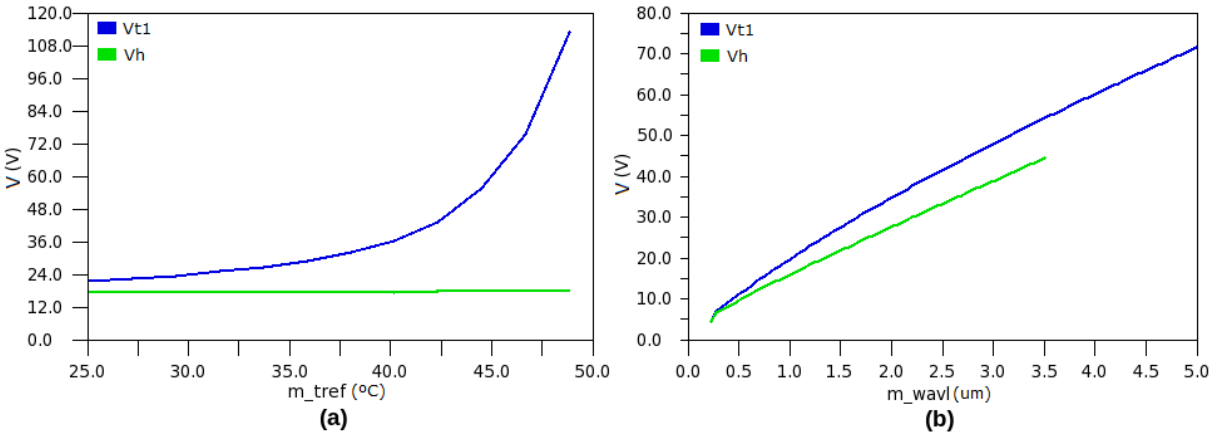
Figure 4.10: Results of sensitivity analysis for the initial constraints settings.



Source: the author

To validate the sensitivity analysis and evaluate the quantitative impact of a parameter on a specific target performance it is possible to perform the parametric analysis described in Section 4.3. For example, from Figure 4.10, T_{ref} has major relative influence on V_{t1}, but not on V_h. W_{avl}, on the other hand, have significant relative influence on both V_{t1} and V_h. The parametric analysis results shown in Figure 4.11 confirms these assertions. An increase in T_{ref} causes an exponential increase in V_{t1} and almost does not change V_h, while an increase in W_{avl} produces an increase in both V_{t1} and V_h almost at the same rate.

Figure 4.11: Effect of T_{ref} (a) and W_{avl} (b) parameters on V_{t1} and V_h performances.



Source: the author

4.5 Optimization of the macromodel I-V characteristics

This section describes the optimization step in the flow chart of Figure 4.1. The usage of two optimization algorithms to fit the $I-V$ characteristics of a MOSFET macromodel to the measured data of its corresponding ESD protection device is detailed. First, a stochastic-based algorithm is used to improve the probability of finding a global optimum for the optimization problem. When the first algorithm reaches a reasonable solution, a deterministic-based algorithm is used to improve such solution. In many optimization problems, the usage of only one of the algorithms mentioned above could be enough to find a reasonable result. However, given the non-linear properties of the optimization problem of this work, the usage of both algorithms has shown to be more effective, providing better and faster results.

The optimizations in this work were performed using the Global Nominal Optimization (GNO) and the Deterministic Nominal Optimization (DNO) tools from the MunEDA WiCkeD software package. Both tools use an external circuit simulator to determine the quality of

the optimization results. The Cadence Spectre Circuit Simulator was adopted for this purpose. The GNO tool implements stochastic-based algorithms, while the DNO tool implements deterministic-based algorithms. Stochastic approaches randomly sample on the objective function with a certain probability and can reach a global optimum at the price of a large number of performance evaluations. Numerical deterministic techniques, on the other hand, are mostly based on gradient information and typically can find a solution in a short time. However, for non-convex optimization problems, deterministic optimization methods might stuck in a local optimum and never converge to a global optimum (ZOU, 2009).

Table 4.2: Performances specifications for the Global Nominal Optimization.

Performance	*Lower bound*	*Target Value*	*Upper Bound*
V_{t1}	6.3 V	6.49 V	6.7 V
I_{t1}	0 A	9 mA	50 mA
V_h	4.8 V	5.1 V	5.4 V
I_h	50 mA	80 mA	150 mA
V_{t2}	4.8 V	18 V	30 V

Source: the author

Table 4.3: Algorithm and settings for the Global Nominal Optimization.

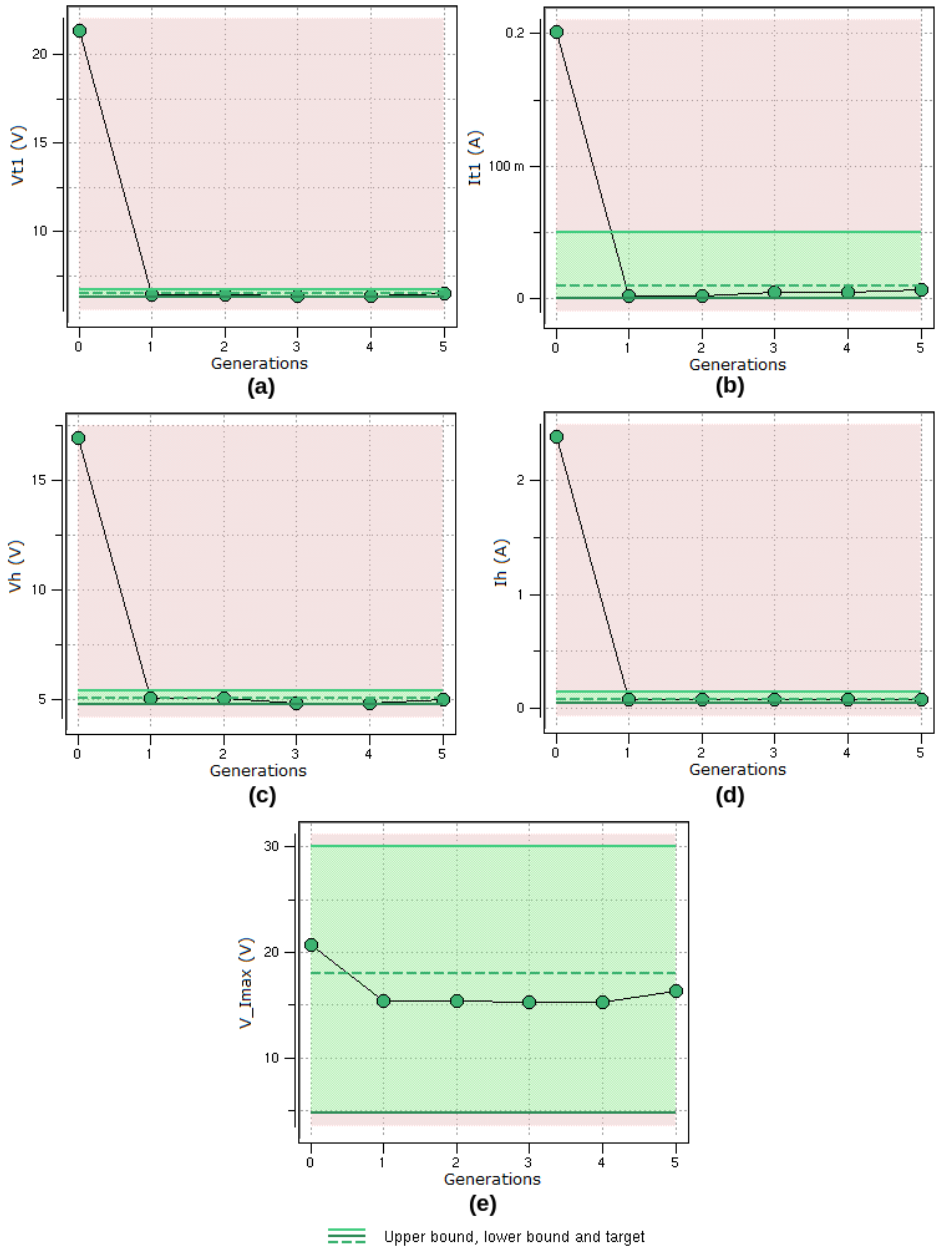
Algorithm	Stochastic
Number of generations	5
Generation size	10,000
Starting box size	100%
Final box size	10%

Source: the author

Table 4.2 presents the performances specifications for the Global Nominal Optimization tool. The target values were defined based on the characterization data of the GGNMOS device illustrated in Figure 4.2 of Section 4.2, except the target value of V_{t2}. The second breakdown triggering voltage of the GGNMOS device is approximately 13.65 V, however its specification for the GNO tool was set to a higher value because a series resistor R_d of 2 Ω was added to the macromodel after a parametric analysis step, increasing its on-resistance. Therefore, a V_{t2} of 13.65 V is not expected in this first optimization process. The lower and upper boundaries were set considering an acceptable (not optimal) solution. Table 4.3 illustrates the algorithm settings for the GNO tool. The *number of generations* defines the maximum number of sample

generations. After this number is reached, the optimization algorithm stops in any case. The *generation size* defines the number of samples in each sample generation. The *starting box size* defines the size of the initial sampling area around the initial value of the parameters in relation to their lower and upper limits of variation. The *final box size* defines the size of the sampling area for the last generation around the best point so far. For example, a final box size of 10% includes 10% of the range from the lower to upper limits of variation of the parameters in relation to their best value so far.

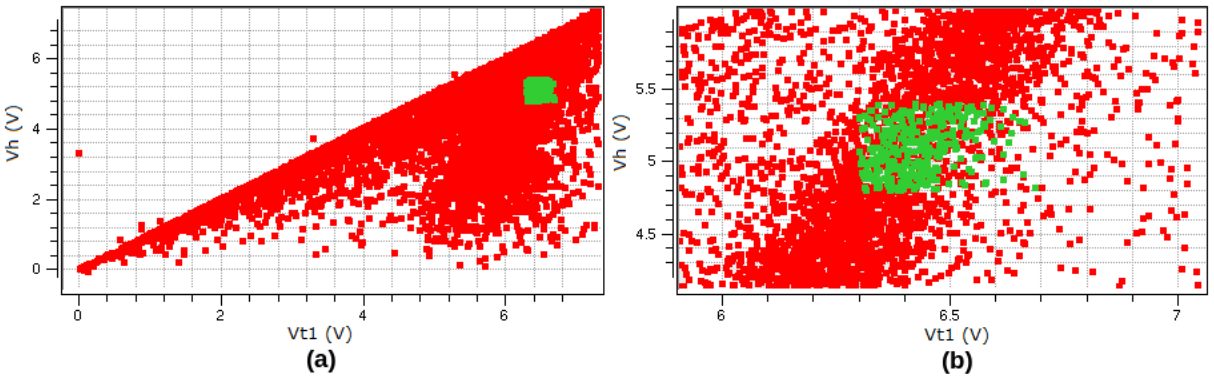
Figure 4.12: Evolution of the Global Nominal Optimization after 5 sample generations for each target performance.



Source: the author

Figure 4.12 illustrates the evolution of the stochastic-based algorithm after 5 sample generations. The dashed line represents the target performance values specified in Table 4.2. The green area which surrounds the dashed line represents their acceptable values within the specified lower and upper boundaries. First, the algorithm tries to find a solution inside the specified boundaries for the performances. After, it tries to approximate to the performance target value. It is possible to notice that the algorithm has reached an acceptable solution, very close to the target performances, after the first sample generation. There are two major reasons for this early solution. The first one is that a solution do exist inside the feasible region (or search space) constrained by the upper and lower limits of variation for the MEXTRAM parameters, defined in Appendix B. If the algorithm could not reach an acceptable solution using the configuration of Table 4.3, the next step would be to increase the upper and lower limits of variation for the parameters in the constraints settings described in Section 4.4. The second major reason is that a large generation size, such as 10,000 samples, increases a lot the probability of finding an acceptable solution (if it exists). Concerning the optimization running time, the stochastic-based algorithm took about 4 hours to simulate and evaluate 5 generations of 10,000 samples each using a 2.9 GHz Intel Xeon 6-core processor with 64 GB RAM.

Figure 4.13: A scatter plot example (a) (b) and the 10-best samples (c) of the Global Nominal Optimization.



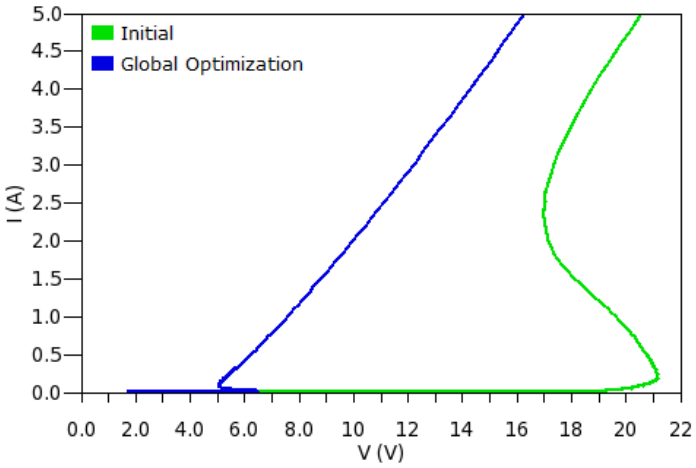
Nr.	IT1_dc ...	IVH_dc ...	V_imax_...	VH_dc...	VT1_dc ...
45087	6.73 m	78.52 m	16.31	5.05	6.47
45954	6.68 m	77.10 m	16.04	5.25	6.42
44982	6.57 m	80.85 m	15.84	5.12	6.49
41847	6.29 m	83.32 m	15.82	4.85	6.40
40571	6.20 m	78.58 m	15.93	5.12	6.44
46805	6.05 m	81.44 m	16.03	5.06	6.38
44750	5.86 m	77.94 m	15.92	5.09	6.44
43074	5.60 m	80.79 m	15.60	4.87	6.46
49513	5.55 m	80.82 m	15.69	5.21	6.43
47571	5.54 m	78.82 m	15.57	5.01	6.33

(c)

Source: the author

Besides the final solution illustrated in Figure 4.12, the GNO tool provides a scatter plot which displays the sample data in two variable performances. Figure 4.13 (a) shows the scatter plot for the V_{t1} and V_h performances and (b) focuses on the region where the performances are within the acceptable values defined in Table 4.2. The red square means that at least one performance specification was violated and the green square means that all performance specifications were fulfilled. The tool also provides a list of the best 100 samples. Figure 4.13 (c) illustrates the best 10 samples during this first optimization, including the sample number and the corresponding performance values. In both the scatter plot and the list of the best samples, it is possible to select any sample for further analysis and optimization. In this case, the best sample was selected. Figure 4.14 illustrates the resulting $I-V$ characteristics of the macromodel before and after the optimization with the stochastic-based algorithm of the GNO tool.

Figure 4.14: Macromodel $I-V$ characteristics before and after the Global Nominal Optimization.



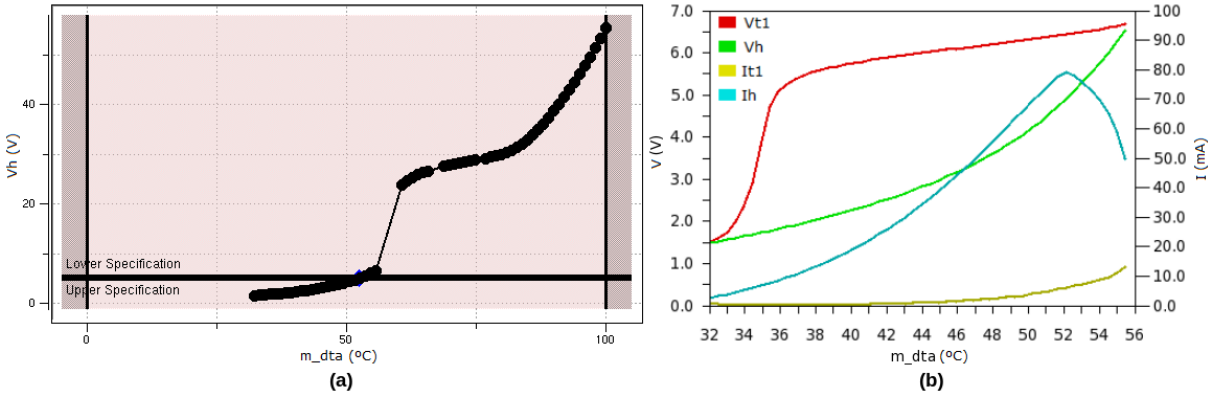
Source: the author

The macromodel $I-V$ characteristics obtained initially with the GNO tool are however still different from the GGNMOS characterization data, especially concerning the circuit on-resistance R_{on}. In this case, the next step according to the flow chart of Figure 4.1 is a sensitivity analysis. After executing such step for the set of parameters determined by the stochastic-based algorithm, the sensitivities of V_{t1}, I_{t1}, V_h and I_h performances to DTA parameter could not be calculated due to simulation error. A simulation error during a sensitivity analysis is usually a convergence issue, a division by zero, or any other issue that does not make possible to measure a target performance. As mentioned in Section 4.4, probably either the initial value of the parameter or the corresponding lower and upper limits of variation are not properly set in relation to the other parameters settings. In this case, many (maybe all) parameters were

changed by the optimization algorithm. With these new parameter values, one or more values within the initial limits of variation for the DTA parameter might be causing a simulation error. One way to check it is performing a parametric analysis or sweep as illustrated in the flow chart of Figure 4.1.

Figure 4.15 (a) illustrates a sweep of the DTA parameter within its initial limits of variation and the corresponding values of the V_h performance using MunEDA WiCkeD Parameter Sweep tool. It is possible to observe that V_h, as a function of DTA, is not continuous at every point in its domain [0 °C, 100 °C]. Due to the same kind of discontinuity, the sensitivities of performances V_{t1}, I_{t1}, V_h and I_h to DTA parameter could not be determined. In this case, it is necessary to narrow the lower and upper limits of variation for the parameter. Figure 4.15 (b) illustrates a parametric analysis which shows that all performances mentioned above are continuous when DTA ranges from 32 °C to 55.5 °C. With these new lower and upper limits added to the constraints settings, all sensitivities could be properly calculated, as illustrated in Figure 4.16.

Figure 4.15: DTA parameter sweep (a) and its parametric analysis (b).

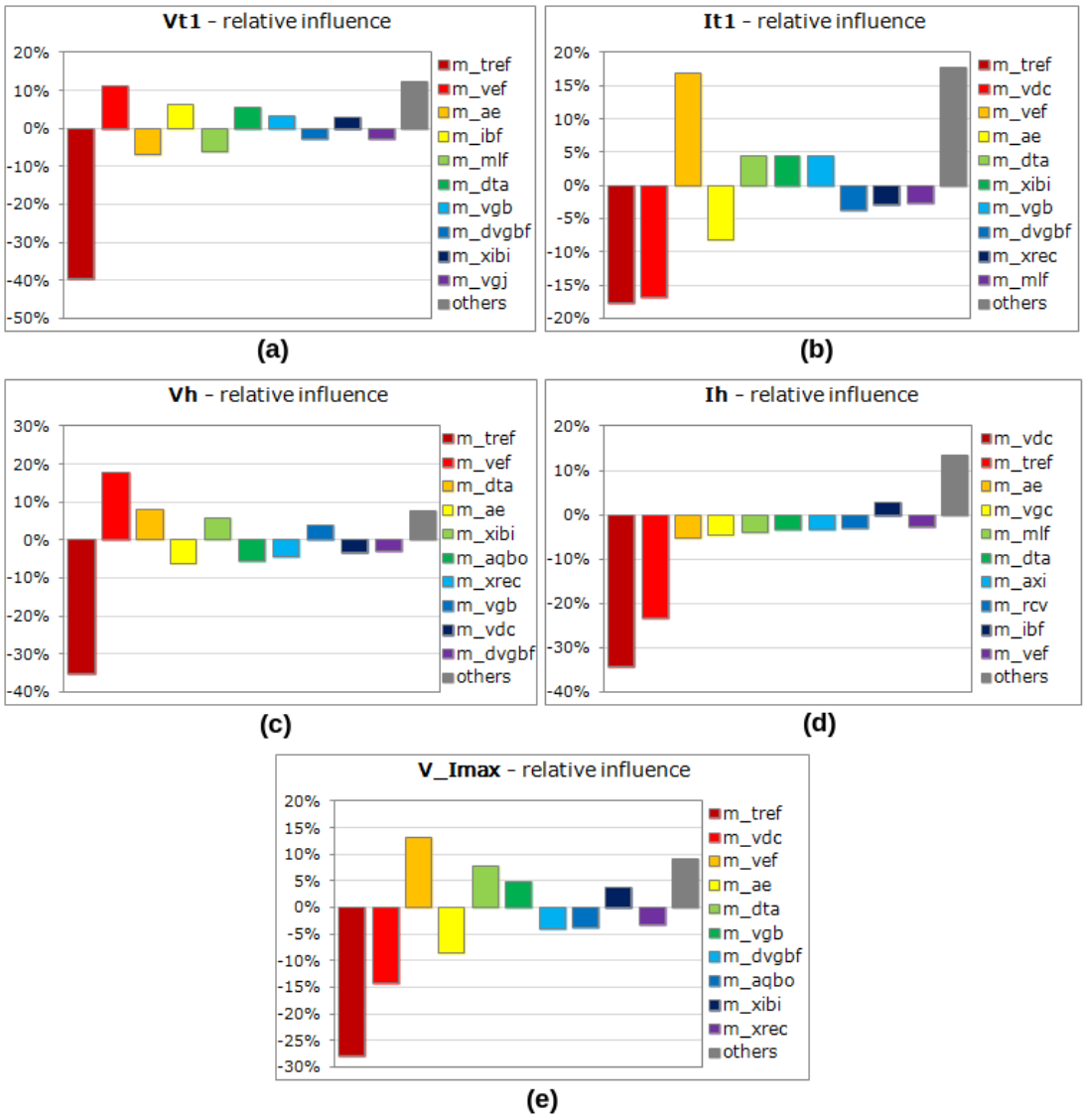


Source: the author

After properly calculating the sensitivity of the performances to all parameters, the next step is the optimization of the macromodel $I-V$ characteristics with the deterministic-based algorithm. In fact, it is essential to have all sensitivities properly calculated because the algorithms of the Deterministic Nominal Optimization tool are sensitivity-based. Also, as already mentioned, numerical deterministic techniques are mostly based on gradient information and, consequently, might stuck in a local optimum and never converge to a global optimum for non-convex optimization problems. For this reason, these optimization methods are more suitable when the macromodel $I-V$ characteristics seem to be reasonably close to the optimal solution (i.e. its reference $I-V$ curve). Besides resetting the lower and upper limits of variation for the

DTA parameter to 32 °C and 55.5 °C, respectively, R_d was also included in the Deterministic Nominal Optimization to improve the macromodel on-resistance. Its lower and upper limits of variation were respectively set to 1 Ω and 2 Ω.

Figure 4.16: Results of sensitivity analysis after the Global Nominal Optimization and with the new constraints settings for the DTA parameter.



Source: the author

Table 4.4 presents the performances specifications for the Deterministic Nominal Optimization tool. All target values were defined based on the characterization data of the reference GGNMOS device. The only difference from these specifications to the Global Nominal Optimization specifications of Table 4.2 is the target value of V_{t2}, which was reduced to 13.65 V after the inclusion of R_d as a variable parameter to improve the circuit on-resistance. Table 4.5 presents the algorithm settings for the DNO tool. The *algorithm* defined for this optimization

step is the SQP (sequential quadratic programming). It solves the optimization problem using quadratic sub-models of the performances in every iteration, which is suitable for this problem given the non-linear behavior of performances in relation to the macromodel parameters. The *number of iterations* defines the maximum number of iterations for the algorithm. After this number is reached, the optimization algorithm stops in any case. With the *adaptive sensitivity calculation*, the deviation of the parameters for sensitivity calculation are adapted to the performances behavior at each algorithm iteration.

Table 4.4: Performances specifications for the Deterministic Nominal Optimization.

Performance	*Lower bound*	*Target Value*	*Upper Bound*
V_{t1}	6.3 V	6.49 V	6.7 V
I_{t1}	0 A	9 mA	50 mA
V_h	4.8 V	5.1 V	5.4 V
I_h	50 mA	80 mA	150 mA
V_{t2}	4.8 V	13.65 V	30 V

Source: the author

Table 4.5: Algorithm and settings for the Deterministic Nominal Optimization.

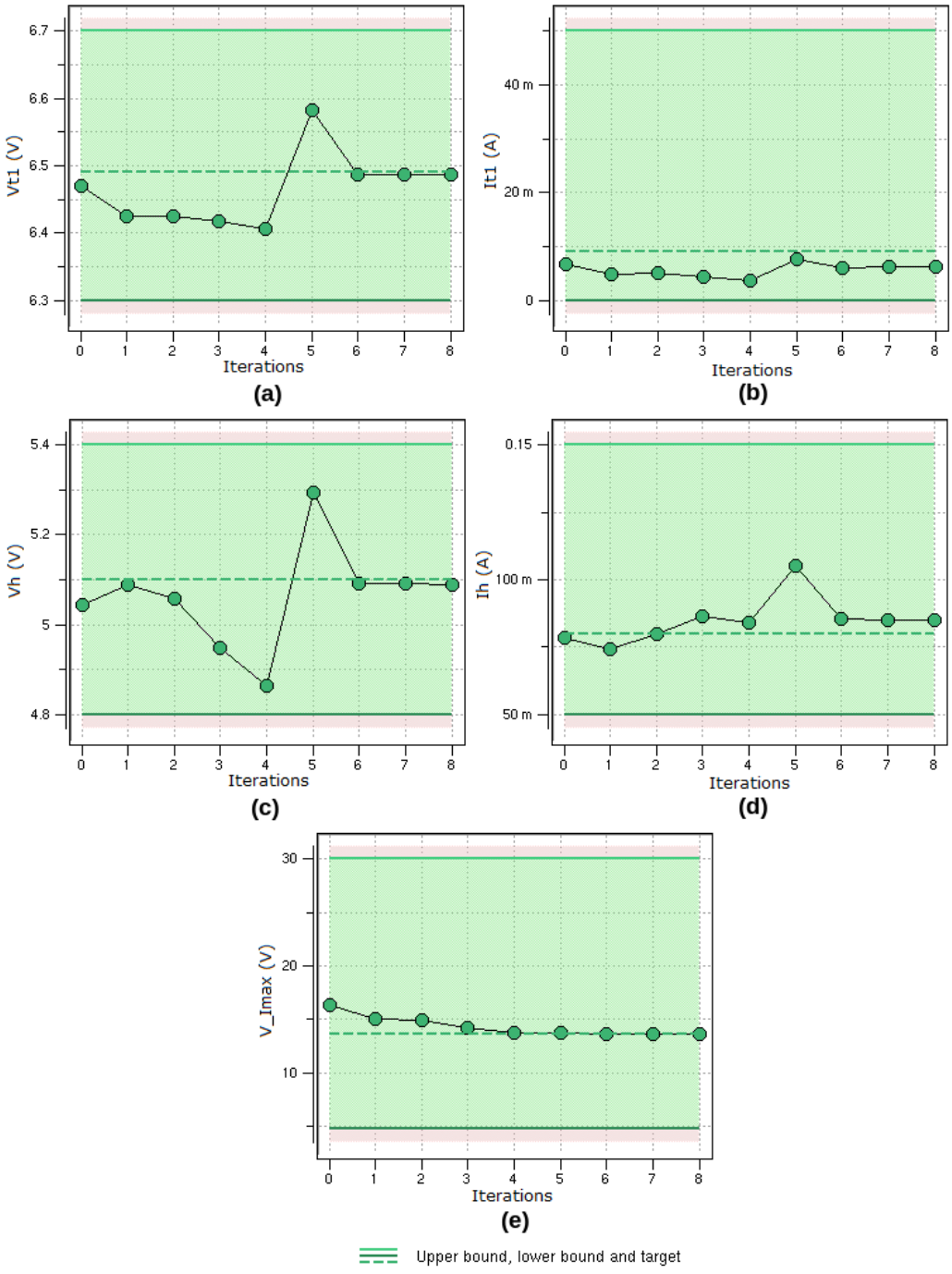
Algorithm	SQP
Number of iterations	10
Adaptive sensitivity calculation	Enabled

Source: the author

Figure 4.17 illustrates the evolution of the SQP algorithm after 8 iterations. In this case, the algorithm stopped before the specified 10 iterations because it was unable to provide further improvements. It is possible to notice that the currents I_{t1} and I_h were the harder performances to optimize. In the fifth iteration, for instance, when the performance I_{t1} was converging to the target value, the performances V_{t1}, V_h and I_h were significantly diverging from their targets. Anyway, the SQP algorithm has reached a solution which is very close to the target performances. Concerning the optimization running time, the SQP algorithm took about 2 hours to perform 8 iterations using a 2.9 GHz Intel Xeon 6-core processor with 64 GB RAM. Figure 4.18 illustrates the resulting $I-V$ characteristics of the macromodel before and after the optimization with the deterministic-based algorithm of the DNO tool. Comparing these results with the $I-V$ curve obtained with the stochastic-based algorithm, there was a minor improvement in the

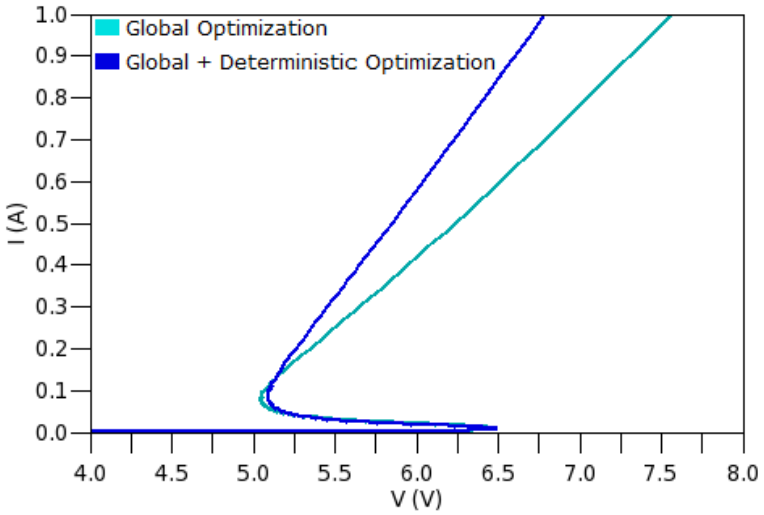
holding voltage V_h and a major improvement in the on-resistance R_{on}. The improvement in the on-resistance was expected with the inclusion of R_d as a variable parameter. In fact, the inclusion of R_d in the macromodel provided not only more accuracy in measuring the performances (as intended in Section 4.3), but also provided more flexibility for the macromodel on-resistance optimization.

Figure 4.17: Evolution of the Deterministic Nominal Optimization after 8 iterations for each target performance.



Source: the author

Figure 4.18: Macromodel $I-V$ characteristics before and after the Deterministic Nominal Optimization.



Source: the author

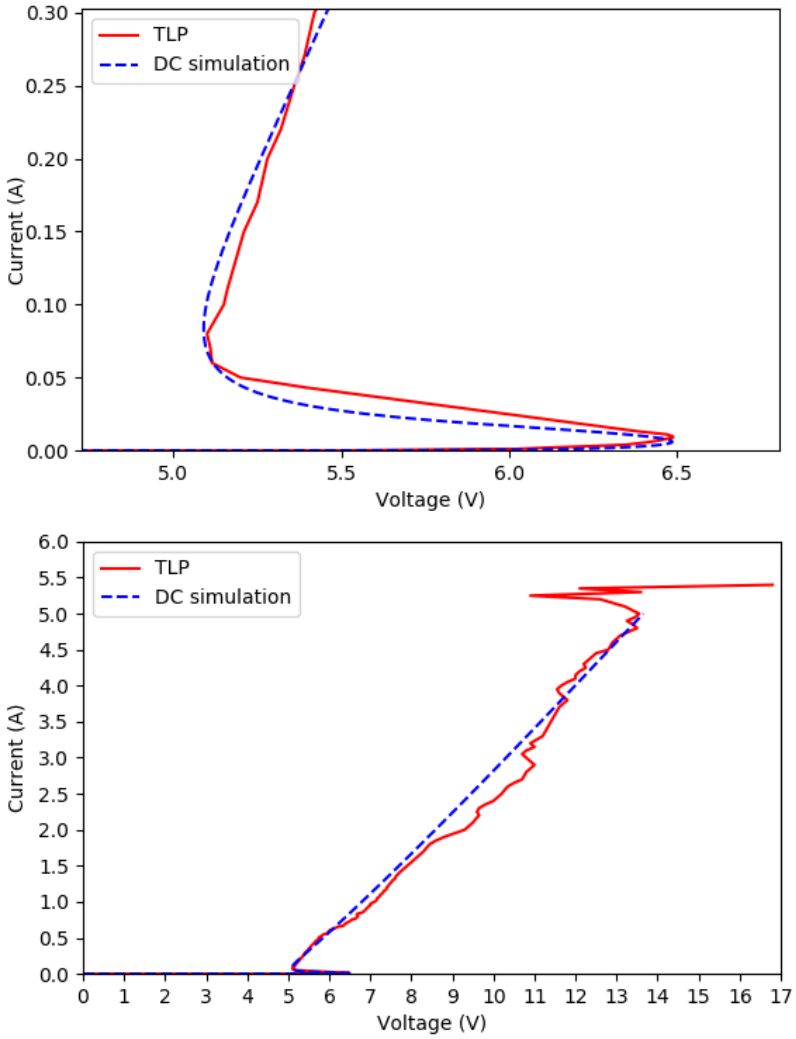
With no further improvements, the $I-V$ characteristics of the macromodel are optimized. Table 4.6 shows a comparison of the final performance values with their corresponding target performances during the optimization process. The most significant difference can be observed in I_{t1} performance. As mentioned above, I_{t1} was one of the hardest performance to optimize. While it was converging to its target value, the other performances were diverging from theirs. The final values of MEXTRAM parameters, on the other hand, can be found in Appendix B. Finally, Figure 4.19 compares the resulting macromodel $I-V$ characteristics with the TLP results of the corresponding GGNMOS ESD protection device.

Table 4.6: Final results after the optimization of the macromodel performances.

Performance	*Target Value*	*Final Value*	*Difference* (Δ)	*% Error*
V_{t1}	6.49 V	6.49 V	0 V	0.0 %
I_{t1}	9 mA	6.17 mA	-2.83 mA	-31.4 %
V_h	5.1 V	5.09 V	-0.01 V	-0.2 %
I_h	80 mA	85.16 mA	5.16 mA	6.5 %
V_{t2}	13.65 V	13.65 V	0 V	0.0 %

Source: the author

Figure 4.19: Comparison between the final macromodel $I-V$ characteristics and the TLP results of the corresponding GGNMOS ESD protection device.



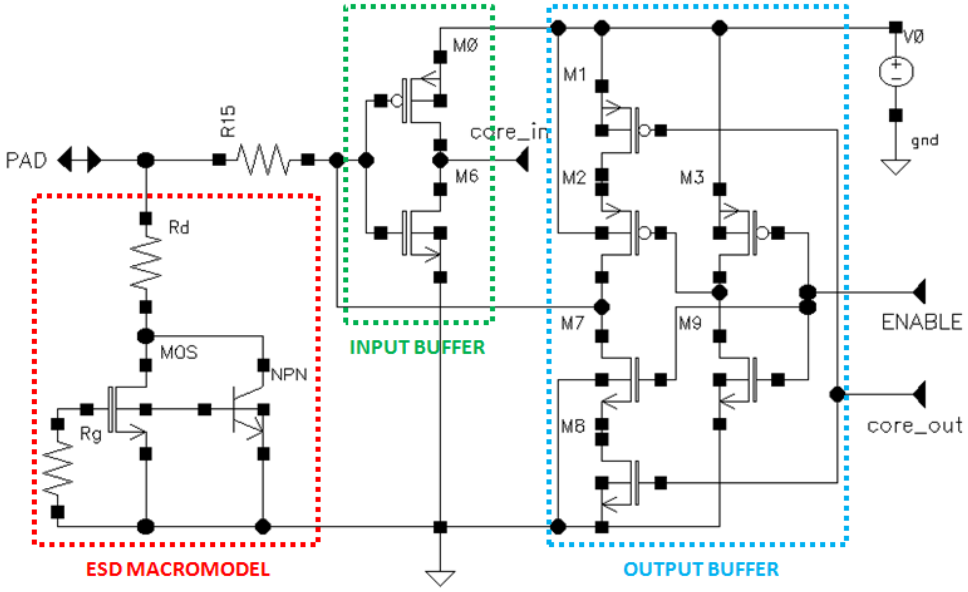
Source: the author

4.6 Macromodel application example

This section describes an application of the macromodel developed and optimized in this work to evaluate the efficacy of its corresponding ESD device in protecting the core circuits. First, the schematic of a bidirectional I/O circuit including the macromodel is presented. Then, the influence of the macromodel on the I/O circuit operation is briefly discussed. Last, the usage of automated ESD breakdown checks to evaluate the efficacy of the protection circuit is detailed.

Figure 4.20 illustrates the schematic implementation of the bidirectional I/O circuit. The input buffer is a regular inverter and the output buffer is a tri-state inverter with active-high control. All transistor compact models are BSIM4v5 with their parameters extracted from a 3 V CMOS 0.18 μm technology. To simplify the implementation, the channel length and width of the core circuit transistors were all set to their default technology values. The resistor R_{15} is typically used to limit the ESD current between a primary and a secondary ESD protection devices. Although only the primary ESD protection was implemented in this case, this resistor was kept to limit the current between the pad and the core transistors. Its resistance was set to 630 Ω.

Figure 4.20: Schematic of a bidirectional I/O circuit including the macromodel developed and optimized in this work.

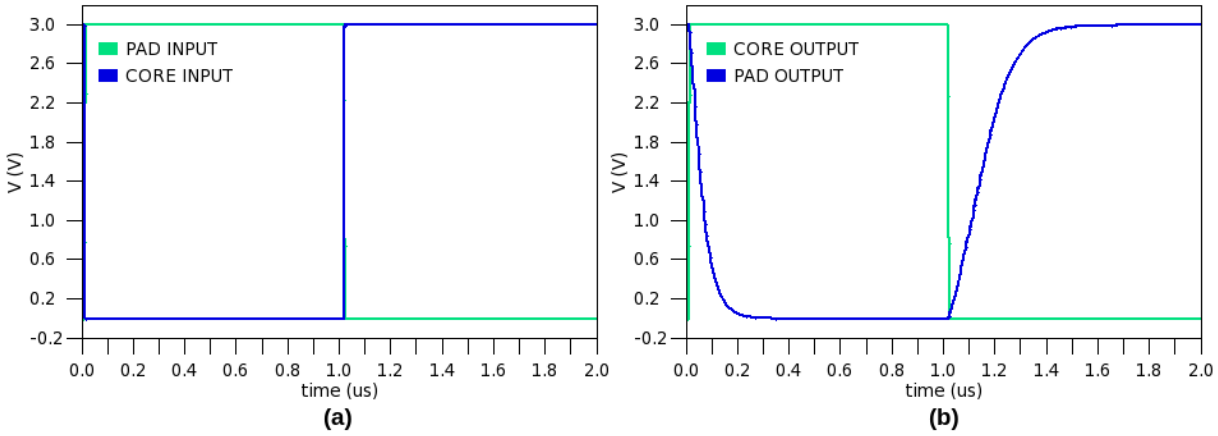


Source: the author

To evaluate the influence of the macromodel on the I/O circuit operation, a periodic-pulse voltage was applied to both input and output buffers to analyze their voltage responses. The rise time and the fall time of the periodic-pulse voltage were set to zero. Figure 4.21 illustrates the input buffer behavior in (a) and the output buffer behavior in (b). For the input buffer, the periodic-pulse voltage was applied to the I/O circuit pad and the response was measured in the *core_in* pin from the schematic of Figure 4.20. For the output buffer, the periodic-pulse voltage was applied to the *core_out* pin and the response was measured in the I/O circuit pad. It is possible to observe in Figure 4.21 a greater delay in the output buffer response. Such delay is actually caused mostly by the NMOS transistor of the ESD macromodel. Its large channel

width (350 μm) makes the transistor of the protection circuit act as a significant load capacitance for the output buffer. For this reason, the transistors of the output buffers are usually larger than the ones adopted in this example, since larger transistors are able to drive more current from the supply pins and, consequently, are able to charge faster the load capacitance of their corresponding outputs. Such increase in the circuit output load capacitance observed for the macromodel is also an issue commonly found during the physical implementation of the target ESD protection device.

Figure 4.21: Input buffer (a) and output buffer (b) responses to a periodic-pulse voltage.



Source: the author

The efficacy of the protection circuit was verified using the automated ESD breakdown checks of Cadence Virtuoso Analog Design Environment (ADE). The specifications for the ESD breakdown checks are listed in a file which is read by the circuit simulator during the simulations. Such specifications include the compact model name of the transistors, the corresponding voltage nodes to be checked, the minimum and the maximum voltage levels accepted and also the maximum duration of the voltage pulse. Any violation of the specifications is reported by the circuit simulator. Below, the breakdown checks model file created for this application example:

```
1     simulator lang=spectre
2
3     nmos_vds_esd assert mod=nmos param=vds min=-7.0 max=7.0 level=notice
4     nmos_vgs_esd assert mod=nmos param=vgs min=-7.0 max=7.0 level=notice
5     nmos_vgd_esd assert mod=nmos param=vgd min=-7.0 max=7.0 level=notice
6     nmos_vgb_esd assert mod=nmos param=vgb min=-7.0 max=7.0 level=notice
7     nmos_vdb_esd assert mod=nmos param=vdb min=-7.0 max=7.0 level=notice
8     nmos_vsb_esd assert mod=nmos param=vbs min=-7.0 max=7.0 level=notice
9
10    pmos_vds_esd assert mod=pmos param=vds min=-7.0 max=7.0 level=notice
```

```
11    pmos_vgs_esd assert mod=pmos param=vgs min=-7.0 max=7.0 level=notice
12    pmos_vgd_esd assert mod=pmos param=vgd min=-7.0 max=7.0 level=notice
13    pmos_vgb_esd assert mod=pmos param=vgb min=-7.0 max=7.0 level=notice
14    pmos_vdb_esd assert mod=pmos param=vdb min=-7.0 max=7.0 level=notice
15    pmos_vsb_esd assert mod=pmos param=vbs min=-7.0 max=7.0 level=notice
16
17    asserts_esd info what=assert where=rawfile
```

The minimum and the maximum voltage levels accepted were respectively set to -7 V and 7 V in the file description above, which are realistic breakdown values for the transistors of the 3 V CMOS 0.18 μm technology adopted. Actually, the breakdown voltages of a device depend not only on the voltage levels, but also on the voltage pulse duration, since the device breakdown is closely related to the amount of energy produced during an ESD event. However, the pulse duration is not specified in the file described above because all the checks were performed using a DC sweep analysis.

The ESD breakdown checks were performed with no VDD core supply voltage and with the terminals *core_in*, *core_out* and *ENABLE* open (floating). Initially, the input current applied to the pad of the bidirectional I/O circuit was set to sweep from 0 to 5 A. The maximum current flowing through the resistor R_{15} was 338 nA (very low). However, the resulting voltage at the core circuit input nodes was almost the same of the GGNMOS device $I-V$ characteristics shown in Figure 4.19. Since the resulting voltage was about 13.65 V at the maximum current applied, several devices violated the ESD breakdown specifications, as illustrated in Table 4.7. The *ESD breakdown check* column lists the voltage check in the corresponding node. For example, V_{DB} denotes the drain-to-bulk voltage check. The device names in the *failed devices* column are the same as in the schematic of Figure 4.20.

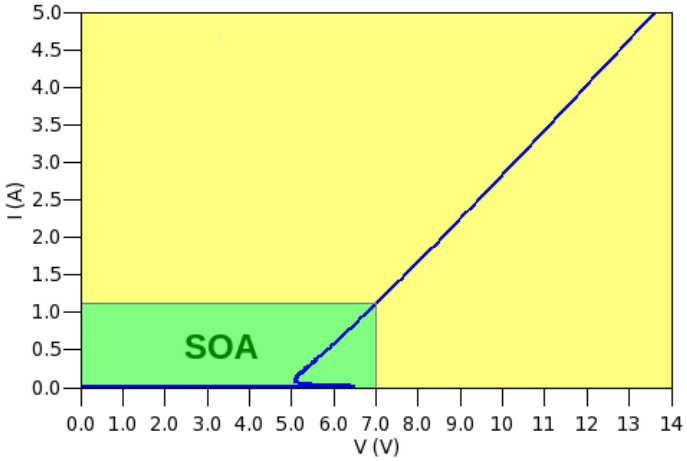
Table 4.7: List of failed devices for an input DC current sweep from 0 to 5 A.

ESD breakdown check	*Failed devices*
V_{DB}	M0, M7, M9
V_{DS}	M0, M7, M9
V_{GB}	M1, M3, M6
V_{GD}	M0, M1, M3, M6, M7, M9
V_{GS}	M1, M3, M6
V_{SB}	None

Source: the author

In the DC analysis described above, it was observed that the resulting voltage at the core circuit input nodes was almost the same of the protection circuit during the current sweep. Therefore, the current applied to the pad should not exceed a maximum value at which the resulting voltage at the protection circuit exceeds the breakdown voltage (7 V). Such maximum current value can be obtained from the $I-V$ characteristics of the protection device, as illustrated in Figure 4.22. In this case, the input current should not exceed about 1.112 A, otherwise the resulting voltage would exceed 7 V. Under such conditions, the bidirectional I/O circuit can be expected to operate without damage, i.e. in its *safe operating area* (SOA). Indeed, if we set the maximum DC sweep current to 1.112 A, all ESD breakdown checks pass.

Figure 4.22: Safe operating area (SOA) for the bidirectional I/O circuit (simulation).



Source: the author

According to Smedes (2017), TLP results derived from pulses of width 100 ns and rise time 10 ns describe well the device behavior for HBM-like events. A good rule of thumb is that a TLP current of 1 A corresponds to a 2 kV HBM voltage level for dissipation-driven failures (NOTERMANS; JONG; KUPER, 1998). Assuming these two assertions are true, the maximum current 1.112 A (current at which all ESD breakdown checks pass for the macromodel application example of this section) indicates that the target GGNMOS protection device is theoretically able to protect the core circuits for at least 2 kV HBM stresses.

5 CONCLUSIONS

Electrostatic discharge (ESD) is a major reliability concern in semiconductor industry. An ESD event may degrade or destroy an integrated circuit (IC), impacting on production yields, manufacturing costs, product quality, product reliability and company profitability. Additionally, the breakdown voltages and failure currents of semiconductor devices are becoming lower with the technology scaling, placing severe constraints on robust IC design. As a result, effective on-chip ESD protection without compromising area and performance requirements is becoming a challenge in deeply-scaled technologies. In this context, circuit simulation can provide the required assistance in on-chip protection design, including robustness analysis of the circuits and performance prediction prior to silicon. However, modeling the MOSFET operation under ESD conditions for circuit simulation is still a challenging issue. The large current and voltage characteristics are typically not well covered by most standard SPICE compact models, such as BSIM3 and BSIM4. To overcome such limitation, a practical modeling approach is highly desired.

Chapter 1 introduced the context and the motivation for this work. Typical examples of ESD damage in CMOS integrated circuits include metal melt, junction breakdown and oxide failure. With the technology scaling, gate oxide is becoming thinner, feature size is becoming smaller, source-to-drain distance is becoming shorter, integration density is increasing and metal interconnects are becoming thinner. As a result, the ESD breakdown voltages and currents are becoming lower and the design of on-chip ESD protection is becoming a challenge. Human Body Model (HBM) and Charged Device Model (CDM) are examples of models adopted to reproduce ESD events and also to qualify electronic devices. These two primary models have been proven to successfully reproduce over 99% of all ESD failure signatures found in field. However, the default method for characterizing the behavior of devices under ESD conditions is the Transmission-Line Pulse (TLP) testing. In contrast to HBM and CDM testing methods, TLP provides more than a pass/fail result. It also provides a quasi-static $I-V$ characteristic that describes the behavior of the device under test (DUT).

Chapter 2 explained the MOSFET operation under ESD conditions. Assuming an ESD event occurs at the drain terminal of an NMOS transistor, the resulting high current levels rapidly charges up the drain capacitance, generating a high electric field across the reverse-biased drain-substrate junction. Such high electric field eventually causes carriers generation in the depletion region due to impact ionization. These new carriers are also subjected to the high electric field and consequent impact ionization, generating even more carriers. With the

resulting avalanche effect, the device enters the avalanche region. Electrons flow towards the drain contact while holes flow towards the substrate contact, giving rise to a substrate current. When the resulting voltage drop across the effective substrate resistance reaches about 0.5 V, the source-substrate junction becomes forward-biased and the lateral npn transistor is then triggered, operating in self-biasing mode. With the positive feedback loop between the avalanche generation of holes in the collector-base junction and the injection of electrons from the emitter-base junction into the base, the device exhibits a negative differential resistance and enters the snapback region. In this region, the drain voltage decreases with an increase in the drain current until a stable holding voltage condition is reached. The device now starts to provide a low and positive impedance path to the ESD current. As the incident current increases even more, the temperature becomes high enough to thermally generate a significant amount of new carriers and the device eventually enters the thermal or second breakdown region. In this region, the temperature in hot spots may permanently damage the device.

Chapter 3 presented the evolution of MOSFET macromodels for ESD circuit simulation. First efforts on modeling the avalanche breakdown for circuit simulation started in the early 70's to overcome a limitation of BJT compact models available at that time, especially of those based on Ebers-Moll and Gummel-Poon models. A voltage-dependent current source was used to include the avalanche breakdown effect. Likewise, the first MOSFET macromodels were implemented using a regular MOSFET compact model with a parallel BJT compact model to describe the influence of the intrinsic lateral bipolar transistor, a voltage-dependent current source to model the avalanche breakdown effect and a resistor to model the substrate resistance of the device. However, the implementation of these macromodels was not straightforward, usually requiring deep and complicated modifications in the circuit simulator code (which is usually not available to integrated circuit designers) to include equation sets or additional elements into the compact models. Besides, a parameter extraction methodology was required not only for the MOSFET compact model but also for the parameters of the equation sets. Later, the usage of advanced MOSFET and BJT compact models, such as BSIM3, BSIM4, VBIC and MEXTRAM, eliminated the need of using a voltage-dependent current source to include the avalanche breakdown effect. The macromodel implementation using only standard SPICE components became easier as there was no need of modifying the circuit simulator code. However, the parameters of the macromodels still needed to be properly extracted or manually modified to fit the measured $I-V$ characteristics of the corresponding ESD protection device.

Chapter 4 proposed a novel methodology for the development and optimization of MOSFET macromodels for ESD circuit simulation. The methodology includes a step of parametric

analysis, which helps to understand the effect of each parameter and to improve the macromodel itself. A sensitivity analysis step helps to evaluate the impact of each parameter on target performances during the optimization steps. Last, stochastic-based and deterministic-based optimization algorithms were adopted to fit the MOSFET macromodel $I-V$ characteristics to measured data. With the parametric analysis step, the resistor element which was responsible for modeling the substrate resistance of the ESD device in previous macromodels available in the literature was removed from the schematic. In this case, the resistor was not properly describing the triggering voltage of the circuit. Also, a series resistor was included in the schematic to increase the macromodel flexibility in describing the on-resistance of the target ESD device. After, a stochastic-based algorithm was used to improve the probability of finding a global optimum for the optimization problem. When a reasonable solution was reached, a deterministic-based algorithm was used to improve such solution. In a few hours, the algorithms had reached a solution very close to their target performance specifications. The resulting macromodel $I-V$ characteristics were then compared with the TLP results of the corresponding ESD protection device.

An application of the macromodel developed and optimized in this work was described to evaluate the efficacy of its corresponding ESD device in protecting the core circuits. The schematic of a bidirectional I/O circuit including the macromodel was presented and the influence of the macromodel on the I/O circuit operation was briefly discussed. An increase in the load capacitance of the output circuits was observed in the presence of the macromodel, which is also an issue commonly found during the physical implementation of ESD protection devices. Then, the usage of automated ESD breakdown checks to evaluate the efficacy of the protection circuit was detailed. In the example, the macromodel was able to protect the core circuit up to a maximum current of 1.112 A, indicating that the target GGNMOS protection device would then be theoretically able to protect the bidirectional I/O circuit for at least 2 kV HBM stresses, according to Smedes (2017).

The major benefit of using MOSFET macromodels for circuit simulation is the possibility of simulating the influence of ESD protection circuits on the core circuits operation. The parasitic capacitance of an ESD protection device may degrade the performance of RF e high-speed digital I/O circuits. In this case, circuit simulation can provide the required assistance on the design of both the ESD protection and the I/O circuits to meet their performance requirements. The benefits of using the optimization-based methodology for ESD macromodels design presented in this work include its wide availability and its straightforward and fast implementation. The usage of only standard SPICE elements and commercial CAD tools makes

the methodology widely available to integrated circuits designers. Also, advanced and complete compact models were adopted. BSIM4 is a very complete MOSFET compact model and MEXTRAM is the most complete BJT compact model for high-current and high-voltage operation when compared to other common compact models such as Gummel-Poon, VBIC and HICUM (BERKNER, 2002). The methodology implementation is very simple compared to previous approaches that required deep and complicated modifications in the circuit simulator source code to include equations and additional elements. Also, it does not mix hardware description language with conventional SPICE elements, preserving several properties of the SPICE compact models which are important to simulate not only the ESD protection efficacy but also the core circuits functionality. The methodology also does not require any additional parameter extraction method specifically for the macromodel implementation as often required in previous approaches. Last, in a few hours the optimization algorithms adopted were able to reach a very reasonable $I-V$ characteristic for the macromodel.

In future works, the methodology could be improved. In this work the macromodel $I-V$ characteristics were optimized using a DC analysis (quiescent domain). However, several parameters of MEXTRAM compact model (e.g. depletion capacitances, transit times and noise parameters) are not considered in its DC model and then were not included in the optimization steps of the methodology. As a consequence, simulating the effect of the ESD protection circuits on the core circuits functionality would hardly provide accurate results in a transient analysis (time domain). In this sense, instead of using a DC current source during the optimization steps, it would be possible to use transient pulses with similar amplitude, duration and rise time of the reference TLP testing method. Another suggestion for future works is the adoption of the methodology to optimize macromodels for several different ESD protection devices, such as GGNMOS devices with different sizing and number of fingers, to obtain a scalable macromodel for ESD circuit simulation.

REFERENCES

AMERASEKERA, A.; DUVVURY, C. **ESD in Silicon Integrated Circuits**. 2nd. ed. Chichester, England: John Wiley & Sons, Ltd, 2002.

AMERASEKERA, A. et al. Modeling MOS snapback and parasitic bipolar action for circuit-level ESD and high current simulations. **Reliability Physics Symposium, 1996. 34th Annual Proceedings, IEEE International**, p. 318–326, May 1996.

BERKNER, J. Compact models for bipolar transistors. **European IC-CAP Device Modeling Workshop**, Mar 2002.

BOYLE, G. R. et al. Macromodeling of integrated circuit operational amplifiers. **Solid-State Circuits – Vol. 9(6), IEEE Journal of**, p. 353–364, Dec 1974.

CAO, S. et al. ESD design strategies for High-Speed Digital and RF circuits in deeply scaled silicon technologies. **Circuits and Systems I: Regular Papers, IEEE Transactions on**, p. 2301–2311, Sept 2010.

CHUN, J. H. ESD protection circuits for advanced CMOS technologies. **Ph.D. dissertation, Stanford University**, Jun 2006.

DIVEKAR, D. A.; LOVELACE, R. E. Modeling of avalanche generation current of bipolar junction transistors for computer circuit simulation. **Computer-Aided Design of Integrated Circuits and Systems – Vol. 1(3), IEEE Transactions on**, p. 112–116, Jul 1982.

DUTTON, R. W. Bipolar transistor modeling of avalanche generation for computer circuit simulation. **Electron Devices – Vol. 22(6), IEEE Transactions on**, p. 334–338, Jun 1975.

DUVVURY, C. ESD qualification changes for 45nm and beyond. **Electron Devices Meeting (IEDM), 2008 IEEE International**, p. 1–4, Dec 2008.

DUVVURY, C. Paradigm shift in ESD qualification. **Reliability Physics Symposium, 2008 IEEE International**, p. 1–2, Apr 2008.

DUVVURY, C. et al. **Discontinuing Use of the Machine Model for Device ESD Qualification**. 2013. Available at <http://www.jedec.org>. Access date: 2015-10-25.

EBERS, J. J.; MOLL, J. L. Large-signal behavior of junction transistors. **Institute of Radio Engineers – Vol. 42(12), IEEE Proceedings of the**, p. 1761–1772, Dec 1954.

ESD Association. **Fundamentals of Electrostatic Discharge – Part One – An Introduction to ESD**. 2013. Available at <http://www.esda.org>. Access date: 2015-10-28.

ESD Association. **Fundamentals of Electrostatic Discharge – Part Five – Device Sensitivity and Testing**. 2014. Available at <http://www.esda.org>. Access date: 2015-10-25.

FOSSUM, J. G. A bipolar device modeling technique applicable to computer-aided circuit analysis and design. **Electron Devices – Vol. 20(6), IEEE Transactions on**, p. 582–593, Jun 1973.

GALY, P. Electrostatic discharge (ESD) one real life event: Physical impact and protection challenges in advanced CMOS technologies. **Semiconductor Conference (CAS), 2014 International**, p. 31–34, Oct 2014.

GOMEZ, I. G. Optimization of analog circuits by applying evolutionary algorithms. **Ph.D. dissertation, National Institute for Astrophysics, Optics and Electronics**, Jul 2012.

GUMMEL, H. K.; POON, H. C. An integral charge control model of bipolar transistors. **The Bell System Technical Journal – Vol. 49(5)**, p. 827–852, Jun 1970.

HSU, F.-C. et al. An analytical breakdown model for short-channel MOSFETs. **Electron Devices – Vol. 29(11), IEEE Transactions on**, p. 1735–1740, Nov 1982.

HU, C.; CHI, M.-H. Second breakdown of vertical power MOSFETs. **Electron Devices – Vol. 29(8), IEEE Transactions on**, p. 1287–1293, Aug 1982.

Industry Council on ESD Target Levels. **White Paper 2: A Case for Lowering Component Level CDM ESD Specifications and Requirements – Rev 2.0**. 2010. Available at <http://www.esda.org>. Access date: 2015-10-10.

Industry Council on ESD Target Levels. **White Paper 1: A Case for Lowering Component Level HBM/MM ESD Specifications and Requirements – Rev. 3.0**. 2011. Available at <http://www.esdindustrycouncil.org>. Access date: 2015-10-11.

Industry Council on ESD Target Levels. **Machine Model Qualification Issues – March 2012 Version**. 2012. Available at <http://www.esdindustrycouncil.org>. Access date: 2015-10-25.

JEDEC. **Machine Model (MM) Qualification Issues – JEDEC Version May 2014**. 2014. Available at <http://www.jedec.org>. Access date: 2015-10-25.

KER, M. D.; LIN, C. Y.; HSIAO, Y. W. Overview on ESD protection designs of low-parasitic capacitance for RF ICs in CMOS technologies. **Device and Materials Reliability, IEEE Transactions on**, p. 207–218, Jan 2011.

KLOOSTERMAN, W. J.; GRAAFF, H. C. D. Avalanche multiplication in a compact bipolar transistor model for circuit simulation. **Electron Devices – Vol. 36(7), IEEE Transactions on**, p. 1376–1380, Jul 1989.

LIM, S. L. et al. A computationally stable quasi-empirical compact model for the simulation of MOS breakdown in ESD protection circuit design. **Simulation of Semiconductor Processes and Devices (SISPAD), 1997 International Conference on**, p. 161–164, Sept 1997.

LIU, W. et al. BSIM3v3.2.2 MOSFET model – users' manual. **University of California, Berkeley**, 1999.

MAR, J.; LI, S.-S.; YU, S.-Y. Substrate current modeling for circuit simulation. **Computer-Aided Design of Integrated Circuits and Systems – Vol. 1(4), IEEE Transactions on**, p. 183–186, Oct 1982.

MCANDREW, C. et al. BJT modeling with VBIC, basics and v1.3 updates. **2003 NSTI Nanotechnology Conference and Trade Show – Workshop on Compact Modeling – Vol. 2, Proceedings of the**, p. 278–281, Feb 2003.

MCANDREW, C. C. et al. VBIC95, the vertical bipolar inter-company model. **Solid-State Circuits – Vol. 31, 1996 IEEE Journal of**, p. 1476–1483, Oct 1996.

MILLER, S. L. Ionization rates for holes and electrons in silicon. **Physical Review – Vol. 105**, p. 1246–1249, Feb 1957.

MORSHED, T. H. et al. BSIM4v4.7 MOSFET model – users' manual. **University of California, Berkeley**, 2011.

NAPRAVNIK, T.; JAKOVENKO, J. ESD MOSFET model calibration by differential evolutionary optimization algorithm. **Electroscope, ISSN 1802-4564**, Dec 2012.

NOTERMANS, G.; JONG, P. D.; KUPER, F. Pitfalls when correlating TLP, HBM and MM testing. **Electrical Overstress/Electrostatic Discharge Symposium, IEEE Proceedings of the**, Oct 1998.

NXP Semiconductors. **AN10853 – ESD and EMC sensitivity of IC – Rev. 01 – 10 November 2010**. 2010. Available at <http://www.nxp.com/documents/application_note/AN10853.pdf>. Access date: 2015-11-01.

PINTO-GUEDES, M.; CHAN, P. C. A circuit simulation model for bipolar-induced breakdown in MOSFET. **Computer-Aided Design of Integrated Circuits and Systems – Vol. 7(2), IEEE Transactions on**, p. 289–294, Feb 1988.

POON, H. C.; MECKWOOD, J. C. Modeling of avalanche effect in integral charge control model. **Electron Devices – Vol. 19(1), IEEE Transactions on**, p. 90–97, Jan 1972.

SMEDES, T. Transmission line pulse testing: The indispensable tool for ESD characterization of devices, circuits and systems. **In Compliance Magazine**, Aug 2017.

SUN, E. et al. Breakdown mechanism in short-channel MOS transistors. **Electron Devices Meeting, 1978 IEEE International**, p. 478–482, Dec 1978.

TOORN, R.; PAASSCHENS, J. C. J.; KLOOSTERMAN, W. J. The MEXTRAM bipolar transistor model – level 504.11.0. **University of Technology, Delft**, Dec 2012.

VASHCHENKO, V. A.; SHIBKOV, A. **ESD Design for Analog Circuits**. 1st. ed. New York, NY, USA: Springer, 2010.

VOLDMAN, S. ESD and latchup considerations for analog and power applications. **Solid-State and Integrated Circuit Technology (ICSICT), 2012 IEEE 11th International Conference on**, p. 1–4, Oct 2012.

ZHOU, Y. et al. Modeling snapback and rise-time effects in TLP testing for ESD MOS devices using BSIM3 and VBIC models. **2005 NSTI Nanotechnology Conference and Trade Show – Workshop on Compact Modeling, Presentation at**, May 2005.

ZHOU, Y. et al. Modeling snapback and rise-time effects in TLP testing for ESD MOS devices using BSIM3 and VBIC models. **2005 NSTI Nanotechnology Conference and Trade Show – Workshop on Compact Modeling, Proceedings of the**, p. 199–202, Jan 2005.

ZHOU, Y.; HAJJAR, J. J. Source/drain junction partition in MOS snapback modeling for ESD simulation. **2008 NSTI Nanotechnology Conference and Trade Show – Workshop on Compact Modeling – Vol. 3, Proceedings of the**, p. 791–794, Jun 2008.

ZHOU, Y.; HAJJAR, J. J.; LISIAK, K. Compact modeling of on-chip ESD protection using standard MOS and BJT models. **Solid-State and Integrated Circuit Technology (ICSICT'06), 2006 8th International Conference on**, p. 1202–1205, Oct 2006.

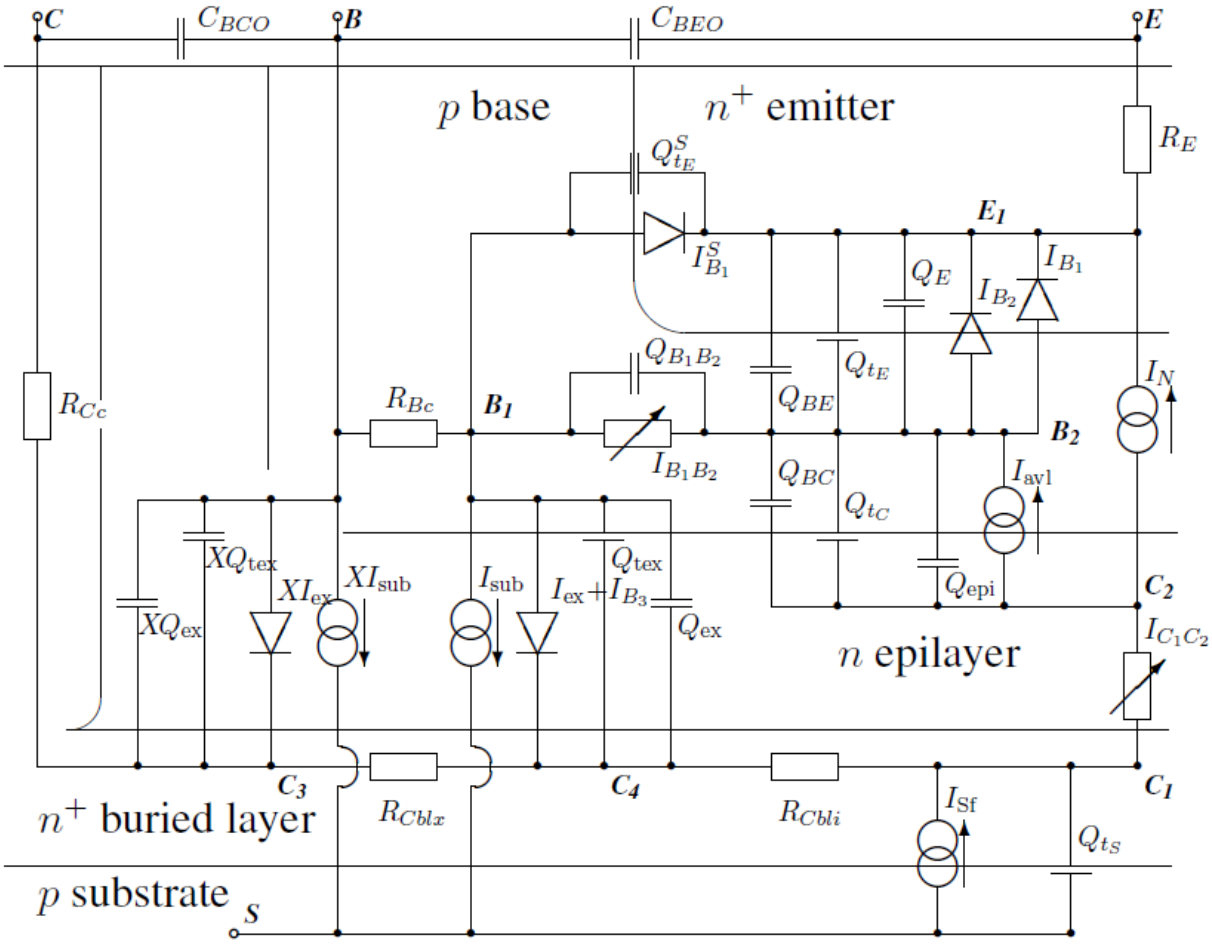
ZHOU, Y. et al. ESD simulation using compact models: from I/O cell to full chip. **Electron Devices and Solid-State Circuits (EDSSC), 2007 IEEE Conference on**, p. 53–58, Dec 2007.

ZOU, J. Hierarchical optimization of large-scale analog/mixed-signal circuits based-on pareto-optimal fronts. **Ph.D. dissertation, Technical University of Munich**, Jan 2009.

APPENDIX A — THE AVALANCHE CURRENT MODEL OF MEXTRAM

The MEXTRAM compact model, as any other bipolar transistor compact model, consists in describing the physical meaning and the equations for the various currents, charges and noise sources that form its equivalent circuit, which is given by the Figure A.1. In the context of this work, one of the most important part of the model is the avalanche current, denoted by I_{avl} in the Figure A.1. This section gives an overview of the avalanche current model of MEXTRAM version 504.12.

Figure A.1: MEXTRAM equivalent circuit for the npn transistor.



Source: Toorn, Paasschens and Kloosterman (2012)

MEXTRAM is the most complete BJT compact model for high-current and high-voltage operation when compared to other common compact models such as Gummel-Poon, VBIC and HICUM (BERKNER, 2002). It includes an optional feature to increase the avalanche current when the current density in the epilayer exceeds the doping level. The snapback behavior is

then modeled as well, giving a better physical description of the bipolar transistor. For such high current density, the current spreading in the collector region changes the electric field distribution and decreases the maximum electric field, impacting directly in the avalanche current generation. Also, when the current density in the epilayer exceeds the doping level, the output resistance can become negative, which may considerably degrade the convergence behavior of the model. For this reason, such optional feature can be switched on or off by setting a flag called $EXAVL$ (model parameter). Without using the extended avalanche model the output resistance can be very small, but it is always positive.

The avalanche current I_{avl} in MEXTRAM depends on the operating mode of the bipolar transistor. In reverse mode, the avalanche current I_{avl} is zero. In forward mode, I_{avl} is given by

$$I_{avl} = I_{C_1C_2} \frac{G_{EM}\, G_{max}}{G_{EM}\, G_{max} + G_{EM} + G_{max}} \tag{A.1}$$

where $I_{C_1C_2}$ is the collector epilayer resistance modeled as a current source (see Figure A.1), G_{EM} is the generation factor of the avalanche current and G_{max} is the maximum value allowed for the generation factor. $I_{C_1C_2}$ in Equation (A.1) is given by

$$I_{C_1C_2} = \frac{E_c + V_{C_1C_2}}{R_{CvT}} \tag{A.2}$$

where E_c is a voltage source which takes into account the decrease in resistance due to carriers injected from the base into the lightly-doped collector epilayer, $V_{C_1C_2}$ is the internal voltage between nodes C_1 and C_2 in the equivalent circuit of Figure A.1 (this voltage is calculated by the circuit simulator) and R_{CvT} is the variable part of the collector resistance, which is calculated after temperature scaling as a linear function of the resistance of the unmodulated epilayer R_{Cv} (model parameter) and as an exponential function of the temperature coefficient of the epilayer resistivity A_{epi} (model parameter).

The generation factor G_{EM} of the avalanche current in Equation (A.1) is given by

$$G_{EM} = \frac{A_n}{B_{nT}} E_M \lambda_D \left\{ \exp\left[-\frac{B_{nT}}{E_M} \right] - \exp\left[-\frac{B_{nT}}{E_M} \left(1 + \frac{W_{eff}}{\lambda_D} \right) \right] \right\} \tag{A.3}$$

where A_n is a material constant for impact ionization, B_{nT} is the result of another material constant for impact ionization (B_n) after temperature scaling, E_M is the maximum electric field at the base-collector junction, λ_D is the point in the space-charge region where the extrapolation of the electric field is zero and W_{eff} is the effective thickness of the epilayer. The material constants A_n and B_n are respectively 7.03×10^7 m^{-1} and 1.23×10^8 Vm^{-1} for an *npn* transistor

type. λ_D is given by

$$\lambda_D = \frac{E_M\, W_D}{2(E_M - E_{av})} \tag{A.4}$$

where W_D is the base-collector depletion layer thickness over which the electric field is important and E_{av} is the average electric field at the base-collector junction. When $E_M \simeq E_{av}$ the expression for λ_D will diverge. Hence, for $(1 - E_{av}/E_M) < 10^{-7}$, the following analytical limit for the generation factor is used instead of Equation (A.3)

$$G_{EM} = A_n\, W_{eff}\, \exp\left[-\frac{B_{nT}}{E_M}\right] \tag{A.5}$$

The generation factor G_{EM} of the avalanche current in Equation (A.1) may not exceed "1" and may not exceed G_{max}, given by

$$G_{max} = \frac{V_{thermal}}{I_{C_1C_2}(R_{BcT} + R_{B_2})} + \frac{q_B^I}{\beta_{fT}} + \frac{R_{ET}}{R_{BcT} + R_{B_2}} \tag{A.6}$$

where $V_{thermal}$ is the thermal voltage (kT/q), R_{BcT} is the parasitic resistance for the base, which is calculated after temperature scaling as a linear function of the constant part of the base resistance R_{Bc} (model parameter) and as an exponential function of the temperature coefficient of the extrinsic base resistivity A_{ex} (model parameter), R_{B_2} is the variable part of the base resistance and depends on the model parameters R_{Bv}, A_B, $A_{Q_{B0}}$, V_{er}, V_{ef}, V_{d_E}, p_E, V_{d_C}, p_C, X_p, V_{g_B} and V_{g_C} (described in Annex A) after temperature scaling, q_B^I is the base charge to take the high injection effect into account for the bipolar main current, β_{fT} is the bipolar forward current gain after temperature scaling, which depends on the model parameters β_f, A_E, A_B, $A_{Q_{B0}}$ and $dV_{g\beta f}$ (described in Annex A) and R_{ET} is the parasitic resistance for the emitter, which is calculated after temperature scaling as a linear function of the constant part of the emitter resistance R_E (model parameter) and as an exponential function of the temperature coefficient of the emitter resistivity A_E (model parameter). R_{Bc} and R_E are also shown in the equivalent circuit of Figure A.1.

When the $EXAVL$ flag is set to "0", the effective thickness of the epilayer in Equations A.3 and A.5 becomes

$$W_{eff} = W_{avl} \tag{A.7}$$

where W_{avl} is a model parameter (described in Annex A). Also, the maximum electric field at the base-collector junction in Equations A.3, A.4 and A.5 becomes

$$E_M = E_{av} + W_D\, \frac{V_{avl}}{{W_{avl}}^2} \left(1 - \frac{I_{C_1C_2}}{I_{hc} + I_{C_1C_2}}\right) \tag{A.8}$$

where V_{avl} is the voltage that determines the curvature of the avalanche current (model parameter) and I_{hc} is the critical current for hot-carrier behavior in the epilayer (model parameter).

When the $EXAVL$ flag is set to "1", the effective thickness of the epilayer in Equations A.3 and A.5 becomes

$$W_{eff} = W_{avl} \left(1 - \frac{x_i}{2\, W_{epi}} \right)^2 \tag{A.9}$$

where W_{epi} is the width of the collector epilayer and x_i/W_{epi} is the thickness of the injection region which is calculated from current $I_{C_1C_2}$ (Equation (A.2)), from junction biases given by the simulator and from the model parameters SCR_{Cv}, R_{Cv} and V_{d_C} (described in Annex A). Also, the maximum electric field at the base-collector junction in Equations A.3, A.4 and A.5 becomes

$$E_M = \frac{1}{2} \left(E_W + E_0 + \sqrt{(E_W - E_0)^2 + 0.1\, {E_{av}}^2\, \frac{I_{C_1C_2}}{I_{hc} + I_{C_1C_2}}} \right) \tag{A.10}$$

where E_W is given by

$$E_W = E_{av} - W_D\, \frac{V_{avl}}{{W_{avl}}^2} \left(\frac{1 + S_{fH}}{1 + 2\, S_{fH}} - \frac{I_{C_1C_2}}{I_{hc}\, SH_W} \right) \tag{A.11}$$

and E_0 is given by

$$E_0 = E_{av} + W_D\, \frac{V_{avl}}{{W_{avl}}^2} \left(1 - \frac{I_{C_1C_2}}{I_{hc} + I_{C_1C_2}} \right) . \tag{A.12}$$

S_{fH} in Equation (A.11) is the high-current spreading factor (model parameter) and SH_W is given by

$$SH_W = 1 + 2\, S_{fH} \left(1 + 2\, \frac{x_i}{W_{epi}} \right) . \tag{A.13}$$

APPENDIX B — OPTIMIZATION BOUNDARIES, INITIAL AND FINAL VALUES

This Appendix describes the initial values as well as the lower and upper limits of variation of each MEXTRAM parameter configured for the optimization process detailed in Chapter 4. It also describes the final values after performing all the optimizations and a difference between the final and the initial value of each parameter.

The initial values for the MEXTRAM parameters were defined based on their compact model default values (see Annex B), except for $MULT$ and $EXAVL$ parameters. $MULT$ was set to a fixed value (70) after the parametric analysis step and $EXAVL$ was set to "1" to enable the extended avalanche model explained in Appendix A. The lower and upper limits of variation of each MEXTRAM parameter were defined based on the compact model clipping values (see Annex B). For the parameters without clip low or clip high values, the lower and upper limits were defined by the author.

The following table presents the initial value, the lower and upper limits of variation, the final values and the difference (Δp) between the final and the initial value of each parameter. Upper and lower limits marked with "-" in their values means that the corresponding parameter keeps its initial value during the whole optimization process (constant parameter).

#	Symbol (Unit)	Lower	Initial (p_i)	Upper	Final (p_f)	Δp
			General parameters and flags			
0	$MULT$ (-)	-	70	-	70	0
1	$LEVEL$ (-)	-	504	-	504	0
2	T_{ref} (°C)	25	25	240	136.1	111.1
3	DTA (°C)	0	0	100	50.58	50.58
4	$EXMOD$ (-)	-	1	-	1	0
5	$EXPHI$ (-)	-	1	-	1	0
6	$EXAVL$ (-)	-	1	-	1	0
7	$EXSUB$ (-)	-	0	-	0	0
			Current parameters of the basic model			
8	I_s (A)	0	22 a	6 f	1.8 f	1.778 f
9	I_k (A)	1 p	100 m	2	1.222	1.122
10	V_{er} (V)	10 m	2.5	30	15.18	12.68
11	V_{ef} (V)	10 m	44	50	18.57	-25.43
12	β_f (-)	1	215	1 k	777.7	562.7

13	I_{Bf} (A)	0	2.7 f	200 p	99.92 p	$\sim$ 99.92 p
14	m_{Lf} (-)	100 m	2	10	3.617	1.617
15	XI_{B_1} (-)	0	0	1	658.6 m	658.6 m
16	I_{zEB} (A)	0	0	1 m	193.5 μ	193.5 μ
17	N_{zEB} (-)	0	22	40	19.48	-2.52
18	β_{ri} (-)	100 p	7	15	6.916	-84 m
19	I_{Br} (A)	0	1 f	3 n	1.558 n	$\sim$ 1.558 n
20	V_{Lr} (V)	0	200 m	2	1.264	1.064
21	X_{ext} (-)	0	630 m	2	1.326	696 m
			Parameters of the avalanche model			
22	W_{avl} (m)	1 n	1.1 μ	5 μ	3.363 μ	2.263 μ
23	V_{avl} (V)	10 m	3	10	3.118	118 m
24	S_{fH} (-)	0	300 m	1	885.3 m	585.3 m
			Resistances and epilayer parameters			
25	R_E (Ω)	1 m	5	10	4.338	-662 m
26	R_{Bc} (Ω)	1 m	23	160	116.2	93.2
27	R_{Bv} (Ω)	1 m	18	30	5.257	-12.743
28	R_{Cc} (Ω)	1 m	12	20	6.533	-5.467
29	R_{Cblx} (Ω)	0	0	80	44.15	44.15
30	R_{Cbli} (Ω)	0	0	80	39.73	39.73
31	R_{Cv} (Ω)	1 m	150	250	80.11	-69.89
32	SCR_{Cv} (Ω)	1 m	1.25 k	2 k	1.405 k	155
33	I_{hc} (A)	1 p	4 m	2	1.392	1.388
34	a_{x_i} (-)	20 m	300 m	2	422.9 m	122.9 m
			Depletion capacitances			
35	C_{j_E} (F)	-	73 f	-	73 f	0
36	V_{d_E} (V)	50 m	950 m	5	4.286	3.336
37	p_E (-)	10 m	400 m	990 m	285.5 m	-114.5 m
38	XC_{j_E} (-)	-	400 m	-	400 m	0
39	C_{BEO} (-)	-	0	-	0	0
40	C_{j_C} (F)	-	78 f	-	78 f	0
41	V_{d_C} (V)	50 m	680 m	2	1.04	360 m

42	p_C (-)	10 m	500 m	990 m	410.4 m	-89.6 m
43	X_p (-)	0	350 m	990 m	696.8 m	346.8 m
44	m_C (-)	0	500 m	5	1.727 m	1.227 m
45	XC_{jC} (-)	-	32 m	-	32 m	0
46	C_{BCO} (-)	-	0	-	0	0
	Transit times					
47	m_τ (-)	-	1	-	1	0
48	τ_E (s)	-	2 p	-	2 p	0
49	τ_B (s)	-	4.2 p	-	4.2 p	0
50	τ_{epi} (s)	-	41 p	-	41 p	0
51	τ_R (s)	-	520 p	-	520 p	0
	Parameters for the SiGe model features					
52	dE_g (eV)	0	0	600 m	289.2 m	289.2 m
53	X_{rec} (-)	0	0	2	740.5 m	740.5 m
	Temperature model (mobility exponents and bandgap voltages)					
54	$A_{Q_{B0}}$ (-)	0	333 m	20	163.5 μ	$\sim$ -332.8 m
55	A_E (-)	0	0	160	81.76	81.76
56	A_B (-)	0	1	120	73.75	72.75
57	A_{epi} (-)	0	2.5	120	79.31	76.81
58	A_{ex} (-)	0	620 m	160	38.96	38.34
59	A_C (-)	0	2	160	59.25	57.25
60	A_{Cbl} (-)	0	2	100	70.98	68.98
61	dA_{I_s} (-)	0	0	2	1.034	1.034
62	$dV_{g\beta f}$ (V)	0	50 m	2	406.4 m	356.4 m
63	$dV_{g\beta r}$ (V)	0	45 m	2	1.104	1.059
64	V_{g_B} (V)	100 m	1.17	2	1.471	301 m
65	V_{g_C} (V)	100 m	1.18	2	596.2 m	-583.8 m
66	V_{g_j} (V)	100 m	1.15	2	1.14	-10 m
67	V_{gZEB} (V)	100 m	1.15	2	1.319	169 m
68	A_{VgEB} (V/K)	100 n	473 μ	1	564.5 m	$\sim$ 564 m
69	T_{VgEB} (K)	0	636	2 k	1.407 k	771
70	$dV_{g\tau_E}$ (V)	-	50 m	-	50 m	0

Noise parameters						
71	A_f (-)	-	2	-	2	0
72	K_f (-)	-	20 p	-	20 p	0
73	K_{fN} (-)	-	20 p	-	20 p	0
74	K_{avl} (-)	-	0	-	0	0
Parameters specific for the four-terminal device						
75	I_{Ss} (A)	0	48 a	1 f	740 a	692 a
76	I_{CSs} (A)	-5	-1	5	-776.8 m	223.2 m
77	I_{ks} (A)	1 p	250 μ	1 m	285.9 μ	35.9 μ
78	C_{jS} (F)	-	315 f	-	315 f	0
79	V_{dS} (V)	-	620 m	-	620 m	0
80	p_S (-)	-	340 m	-	340 m	0
81	V_{gS} (V)	100 m	1.2	3	1.657	457 m
82	A_S (-)	0	1.58	20	17.16	15.58
83	A_{sub} (-)	0	2	100	40.48	38.48

ANNEX A — MEXTRAM PARAMETERS DESCRIPTION

The following table presents a short description of MEXTRAM parameters used in this work, as defined in Toorn, Paasschens and Kloosterman (2012). The parameters denoted with a "*" are not used in the DC model.

#	Symbol	Description (units)
		General parameters and flags
0	$MULT$	Multiplication factor (-)
1	$LEVEL$	Model level, must be set to 504 (-)
2	T_{ref}	Reference temperature. Default is $25°$C (°C)
3	DTA	Difference between the local ambient and global ambient temperatures (°C)
4	$EXMOD$	Flag for extended modeling of the reverse current gain (-)
5	$EXPHI$	* Flag for the distributed high-frequency effects in transients (-)
6	$EXAVL$	Flag for extended modeling of avalanche currents (-)
7	$EXSUB$	Flag for extended modeling of substrate currents (-)
		Current parameters of the basic model
8	I_s	Collector-emitter saturation current (A)
9	I_k	Collector-emitter high injection knee current (A)
10	V_{er}	Reverse Early voltage (V)
11	V_{ef}	Forward Early voltage (V)
12	β_f	Ideal forward current gain (-)
13	I_{Bf}	Saturation current of the non-ideal forward base current (A)
14	m_{Lf}	Non-ideality factor of the non-ideal forward base current (-)
15	XI_{B_1}	Part of ideal base current that belongs to the sidewall (-)
16	I_{zEB}	Pre-factor of emitter-base Zener tunneling current (A)
17	N_{zEB}	Coefficient of emitter-base Zener tunneling current (-)
18	β_{ri}	Ideal reverse current gain (-)
19	I_{Br}	Saturation current of the non-ideal reverse base current (A)
20	V_{Lr}	Cross-over voltage of the non-ideal reverse base current (V)
21	X_{ext}	Part of I_{ex}, Q_{tex}, Q_{ex} and I_{sub} that depends on V_{BC_3} instead of $V_{B_1C_4}$ (-)
		Parameters of the avalanche model
22	W_{avl}	Epilayer thickness used in weak-avalanche model (m)
23	V_{avl}	Voltage determining curvature of avalanche current (V)

24	S_{fH}	Current spreading factor of avalanche model (when $EXAVL = 1$) (-)
		Resistances and epilayer parameters
25	R_E	Emitter resistance (Ω)
26	R_{Bc}	Constant part of the base resistance (Ω)
27	R_{Bv}	Zero-bias value of the variable part of the base resistance (Ω)
28	R_{Cc}	Collector contact resistance (Ω)
29	R_{Cblx}	Resistance of the collector buried layer: extrinsic part (Ω)
30	R_{Cbli}	Resistance of the collector buried layer: intrinsic part (Ω)
31	R_{Cv}	Resistance of the unmodulated epilayer (Ω)
32	SCR_{Cv}	Space charge resistance of the epilayer (Ω)
33	I_{hc}	Critical current for velocity saturation in the epilayer (A)
34	a_{x_i}	Smoothness parameter for the onset of quasi-saturation (-)
		Depletion capacitances
35	C_{j_E}	* Zero-bias emitter-base depletion capacitance (F)
36	V_{d_E}	Emitter-base diffusion voltage (V)
37	p_E	Emitter-base grading coefficient (-)
38	XC_{j_E}	* Fraction of the emitter-base depletion capacitance that belongs to the sidewall (-)
39	C_{BEO}	* Emitter-base overlap capacitance (-)
40	C_{j_C}	* Zero-bias collector-base depletion capacitance (F)
41	V_{d_C}	Collector-base diffusion voltage (V)
42	p_C	Collector-base grading coefficient (-)
43	X_p	Constant part of C_{j_C} (-)
44	m_C	Coefficient for the current modulation of the collector-base depletion capacitance (-)
45	XC_{j_C}	* Fraction of the collector-base depletion capacitance under the emitter (-)
46	C_{BCO}	* Collector-base overlap capacitance (-)
		Transit times
47	m_τ	* Non-ideality factor of the emitter stored charge (-)
48	τ_E	* Minimum transit time of stored emitter charge (s)
49	τ_B	* Transit time of stored base charge (s)
50	τ_{epi}	* Transit time of stored epilayer charge (s)
51	τ_R	* Transit time of reverse extrinsic stored base charge (s)
		Parameters for the SiGe model features

52	dE_g	Bandgap difference over the base (eV)
53	X_{rec}	Pre-factor of the recombination part of I_{B_1} (-)
		Temperature model (mobility exponents and bandgap voltages)
54	$A_{Q_{B0}}$	Temperature coefficient of the zero-bias base charge (-)
55	A_E	Temperature coefficient of the resistivity of the emitter (-)
56	A_B	Temperature coefficient of the resistivity of the base (-)
57	A_{epi}	Temperature coefficient of the resistivity of the epilayer (-)
58	A_{ex}	Temperature coefficient of the resistivity of the extrinsic base (-)
59	A_C	Temperature coefficient of the resistivity of the collector contact (-)
60	A_{Cbl}	Temperature coefficient of the resistivity of the collector buried layer (-)
61	dA_{I_s}	Parameter for fine tuning of temperature dependence of collector-emitter saturation current (-)
62	$dV_{g\beta f}$	Band-gap voltage difference of forward current gain (V)
63	$dV_{g\beta r}$	Band-gap voltage difference of reverse current gain (V)
64	V_{g_B}	Band-gap voltage of the base (V)
65	V_{g_C}	Band-gap voltage of the collector (V)
66	V_{g_j}	Band-gap voltage recombination emitter-base junction (V)
67	V_{gZEB}	Band-gap at reference temperature relevant to the Zener effect in the emitter-base junction (V)
68	A_{VgEB}	Temperature scaling coefficient of emitter-base Zener tunneling current (V/K)
69	T_{VgEB}	Temperature scaling coefficient of emitter-base Zener tunneling current (K)
70	$dV_{g\tau_E}$	* Band-gap voltage difference of emitter stored charge (V)
		Noise parameters
71	A_f	* Exponent of the flicker-noise (-)
72	K_f	* Flicker-noise coefficient of the ideal base current (-)
73	K_{fN}	* Flicker-noise coefficient of the non-ideal base current (-)
74	K_{avl}	* Switch for white noise contribution due to avalanche (-)
		Parameters specific for the four-terminal device
75	I_{Ss}	Base-substrate saturation current (A)
76	I_{CSs}	Collector-substrate ideal saturation current (A)
77	I_{ks}	Base-substrate high injection knee current (A)

78	C_{jS}	* Zero-bias collector-substrate depletion capacitance (F)
79	V_{d_S}	* Collector-substrate diffusion voltage (V)
80	p_S	* Collector-substrate grading coefficient (-)
81	V_{gS}	Band-gap voltage of the substrate (V)
82	A_S	For a closed buried layer: $A_S = A_C$, and for an open buried layer: $A_S = A_{epi}$ (-)
83	A_{sub}	Temperature coefficient for mobility of minorities in the substrate (-)

ANNEX B — MEXTRAM PARAMETERS DEFAULT AND CLIPPING VALUES

The following table presents the **default values** and the **clipping values** of the MEXTRAM parameters used in this work, as defined in Toorn, Paasschens and Kloosterman (2012). The clipping values prevent the parameters to become outside a physically realistic range, or in a range that might create difficulties in the numerical evaluation of the model such as, for example, a division by zero. The default values come from a realistic transistor and are therefore a good indication of typical values.

#	Symbol	Default	Clip low	Clip high
		General parameters and flags		
0	$MULT$	1.0	0.0	-
1	$LEVEL$	504	-	-
2	T_{ref}	25.0	-273.0	-
3	DTA	0.0	-	-
4	$EXMOD$	1.0	0.0	2.0
5	$EXPHI$	1.0	0.0	1.0
6	$EXAVL$	0.0	0.0	1.0
7	$EXSUB$	0.0	0.0	1.0
		Current parameters of the basic model		
8	I_s	22.0×10^{-18}	0.0	-
9	I_k	0.1	1.0×10^{-12}	-
10	V_{er}	2.5	0.01	-
11	V_{ef}	44.0	0.01	-
12	β_f	215.0	1.0×10^{-4}	-
13	I_{Bf}	2.7×10^{-15}	0.0	-
14	m_{Lf}	2.0	0.1	-
15	XI_{B_1}	0.0	0.0	1.0
16	I_{zEB}	0.0	0.0	-
17	N_{zEB}	22.0	0.0	-
18	β_{ri}	7.0	1.0×10^{-10}	-
19	I_{Br}	1.0×10^{-15}	0.0	-
20	V_{Lr}	0.2	-	-

21	X_{ext}	0.63	0.0	1.0
Parameters of the avalanche model				
22	W_{avl}	1.1×10^{-6}	1.0×10^{-9}	-
23	V_{avl}	3.0	0.01	-
24	S_{fH}	0.3	0.0	-
Resistances and epilayer parameters				
25	R_E	5.0	1.0×10^{-3}	-
26	R_{Bc}	23.0	1.0×10^{-3}	-
27	R_{Bv}	18.0	1.0×10^{-3}	-
28	R_{Cc}	12.0	1.0×10^{-3}	-
29	R_{Cblx}	0.0	0.0	-
30	R_{Cbli}	0.0	0.0	-
31	R_{Cv}	150.0	1.0×10^{-3}	-
32	SCR_{Cv}	1250.0	1.0×10^{-3}	-
33	I_{hc}	4.0×10^{-3}	1.0×10^{-12}	-
34	a_{x_i}	0.3	0.02	-
Depletion capacitances				
35	C_{j_E}	73.0×10^{-15}	0.0	-
36	V_{d_E}	0.95	0.05	-
37	p_E	0.4	0.01	0.99
38	XC_{j_E}	0.4	0.0	1.0
39	C_{BEO}	0.0	0.0	-
40	C_{jC}	78.0×10^{-15}	0.0	-
41	V_{d_C}	0.68	0.05	-
42	p_C	0.5	0.01	0.99
43	X_p	0.35	0.0	0.99
44	m_C	0.5	0.0	1.0
45	XC_{j_C}	32.0×10^{-3}	0.0	1.0
46	C_{BCO}	0.0	0.0	-
Transit times				
47	m_τ	1.0	0.1	-
48	τ_E	2.0×10^{-12}	0.0	-

49	τ_B	4.2×10^{-12}	0.0	-
50	τ_{epi}	41.0×10^{-12}	0.0	-
51	τ_R	520.0×10^{-12}	0.0	-
Parameters for the SiGe model features				
52	dE_g	0.0	-	-
53	X_{rec}	0.0	0.0	-
Temperature model (mobility exponents and bandgap voltages)				
54	$A_{Q_{B0}}$	0.3	-	-
55	A_E	0.0	-	-
56	A_B	1.0	-	-
57	A_{epi}	2.5	-	-
58	A_{ex}	0.62	-	-
59	A_C	2.0	-	-
60	A_{Cbl}	2.0	0.0	-
61	dA_{I_s}	0.0	-	-
62	$dV_{g\beta f}$	50.0×10^{-3}	-	-
63	$dV_{g\beta r}$	45.0×10^{-3}	-	-
64	V_{g_B}	1.17	0.1	-
65	V_{g_C}	1.18	0.1	-
66	V_{g_j}	1.15	0.1	-
67	V_{gZEB}	1.15	0.1	-
68	A_{VgEB}	4.73×10^{-4}	-	-
69	T_{VgEB}	636.0	0.0	-
70	$dV_{g\tau_E}$	0.05	-	-
Noise parameters				
71	A_f	2.0	0.01	-
72	K_f	20.0×10^{-12}	0.0	-
73	K_{fN}	20.0×10^{-12}	0.0	-
74	K_{avl}	0.0[1]	0.0	1.0
Parameters specific for the four-terminal device				
75	I_{Ss}	48.0×10^{-18}	0.0	-

[1]The physical and therefore recommended value is $K_{avl} = 1$.

76	I_{CSs}	-1.0	-	-
77	I_{ks}	250.0×10^{-6}	1.0×10^{-12}	-
78	C_{js}	315.0×10^{-15}	0.0	-
79	V_{d_S}	0.62	0.05	-
80	p_S	0.34	0.01	0.99
81	V_{gs}	1.20	0.1	-
82	A_S	1.58	-	-
83	A_{sub}	2.0	-	-