ESD PROTECTION CIRCUITS FOR ADVANCED CMOS TECHNOLOGIES

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Jung-Hoon Chun

June 2006

© Copyright by Jung-Hoon Chun 2006
All Rights Reserved

I certify that I have read this dissertation and that, in my opinion, it is fully adequate, in scope and quality, as dissertation for the degree of Doctor of Philosophy.

_____

(Robert W. Dutton) Principal Advisor

I certify that I have read this dissertation and that, in my opinion, it is fully adequate, in scope and quality, as dissertation for the degree of Doctor of Philosophy.

_____

(Boris Murmann)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate, in scope and quality, as dissertation for the degree of Doctor of Philosophy.

_____

(Krishna Saraswat)

Approved for the University Committee on Graduate Studies

Abstract

Electrostatic Discharge (ESD) has become one of the most critical reliability issues in integrated circuits (ICs). Therefore, there have been extensive research efforts seeking to improve ESD protection capabilities; the goal is to achieve highly reliable IC products in the presence of ESD threats. However, the continuous scaling of technology and the introduction of new device concepts and materials into the mainstream CMOS technology has brought many new ESD challenges. In this dissertation a variety of ESD issues in advanced CMOS technology are addressed in breadth, covering topics that range from fundamental device physics to circuit design engineering, providing the guidelines needed to develop robust and transparent ESD protection circuits.

As the circuit performance has been improved and the bandwidth of circuit interfaces has reached into the Gigahertz regime, problems of ESD-to-circuit interactions have emerged as a major challenge. The impact of ESD parasitic elements on the performance of Gigahertz RF ICs is analyzed based on a simplified RF model of ESD protection devices. The present work is focused on the trade off between conventional RF ESD protection concepts that optimize in terms of capacitive loading, and the emerging concept of RF ESD co-design where ESD protection is tightly integrated into RF circuit design. To investigate co-design efforts and check the limitations of conventional low-C RF protection, 2 GHz narrowband and 5 GHz broadband LNA test circuits were developed, with protection level greater than the 2 kV HBM standard .

The ESD-to-circuit interaction problem of large signal distortion caused by ESD protection circuits is analyzed. ESD protection circuits typically contain a significant amount of nonlinear capacitance. At high frequencies and large amplitudes, this nonlinearity can degrade the signal integrity at the input pins of high performance mixed signal ICs, such as analog-to-digital converters (ADCs). This research provides a theoretical analysis of the problem as well as experimental results that quantify typical distortion levels introduced by state-of-the-art ESD protection structures. It is shown that with *SFDR* targets approaching 100 dB at frequencies near 100 MHz, the protection device can easily become a performance-limiting factor in the future.

This research also provides a guideline for selecting and implementing ESD protection devices in high-speed ADC applications with high linearity requirements.

Due to the demand of compact I/O design in highly integrated ASICs, design rules to avoid inadvertent ESD failures caused by the coupling between ESD protection devices and the desired circuit functions are required. In the context of developing design rules, a new failure phenomenon of PMOSFET pull-up devices under ESD conditions is analyzed. Localized turn-on of the parasitic PNP transistor can be caused by localized charge injection from the adjacent diodes into the body of the PMOSFET, significantly degrading the ESD robustness of PMOSFETs. The critical layout parameters affecting this problem are identified and design guidelines for avoiding this new PMOSFET failure mode are also suggested.

As the physical gate length of the CMOS devices enter the sub-100nm regime, the validity of conventional TCAD simulations and the underlying electrothermal physics face challenges due to new materials being used and more demanding operating conditions. To extend the use of commercial TCAD tools, ESD protection capabilities of new strained-Si/SiGe NMOS devices are investigated based on use of carefully calibrated parameters to account for electrothermal effects. Despite the low thermal conductivity of the buried SiGe layers, strained-Si devices show superior ESD protection capabilities compared to unstrained-Si (bulk-Si) devices due to the high bipolar current gain and increased impact ionization rate.

Finally, seeking a valid device simulation method for nano-scale devices, a new electrothermal model based on the advanced thermal physics is developed and implemented in the research simulation tool *PROPHET*. The optical and acoustic temperature systems are fully coupled to the hydrodynamic model. This electro-thermal model is validated based on simulations of a thin-body SOI NMOSFET. This system of model equations can capture most of the physical phenomena of the heat generation and conduction inside nano-scale devices with moderate numerical complexity. It is anticipated that the modeling concepts developed in this research may be applied in future ESD simulation studies.

Acknowledgments

I would like to thank many people who made my journey at Stanford rich and memorable. First and foremost, I would like to thank my advisor Professor Robert W. Dutton for his excellent guidance throughout my Ph.D. I am indebted to him for many advices about research and life in general. His steadfast devotion to his students and endless enthusiasm for research demonstrates an invaluable example.

I was fortunate to receive great support and guidance from my co-advisor Professor Boris Murmann. Under his oversight, I was able to complete a critical chapter of this dissertation. I am also grateful to his immediate and meticulous critique on the numerous drafts of all the chapters. His constant pursuit of excellence will continue to serve as a model for my professional life.

I would like to show my gratitude to Professor Krishna Saraswat for his thorough review of this thesis. I am also grateful to Professor Tom Kenny who served as the chair for my oral defense on very short notice.

Professor Zhiping Yu, Dr. Dan Yergeau, Dr. Klas Lilja, and Dr. Yang Liu have provided wonderful resources in the Technology CAD group. Their dedication and great engineering insight made it possible for me to meet various writing deadlines on time. I was also fortunate to have the warm hearted support from Fely Berrera, Miho Nishi, Xiaoxiang Liu, and Ann Guerra. They made my life much easier through administrative support.

I have benefited tremendously through interactions with great people in industry. Especially, I am very grateful to Texas Instruments for allowing me to use their resources in Dallas for more than one month. I am truly indebted to Dr. Charvaka Duvvury in Texas Instruments. Dr. Duvvury has brought up many interesting research topics and mentored me to solve those problems. I would like to thank Dr. Gianluca Boselli and Peter Rathfelder for sharing their ideas and generously providing invaluable data. I wish to express my thanks to Wolfgang Soldner, Dr. Martin Streibl and Dr. Herald Gossner in Infineon for the collaboration on RF ESD research. I also appreciate Dr. Louis Burgyearn and Dr. Tim Holtz in Fairchild Semiconductor for offering me the internship opportunity to work on their analog IC products.

I would like to thank Professor Min-Koo Han at Seoul National University, Professor Yearn-Ick Choi at Ajou University, and Dr. Byoung-Ha Park, Dr. Sung-Gi Yang at Samsung Electronics for their continuous encouragement.

I wrote most of my thesis while working full time at Rambus Inc. I would like to thank Rambus for providing a great environment. Especially I am grateful to Dr. Ken Chang, Dr. Fari Assaderaghi, Dr. Nhat Nguyen, and Yohan Frans.

I am very grateful to former and current Dutton group members. Without the early support of Changhoon Choi and Kwanghoon Oh, my graduate life would have been much tougher. Thanks also go out to Olof Tornblad, Taeyoung Oh, Xinyi Zhang, Choshu Ito, Yichang Lu (my Mandarin tutor!), Eric Pop, Jae Wook Kim, Minghui Han, Hai Lan, Reza Navid, Parastoo Nikaeen, Kyung Rok Kim, Tze Wee Chen, and Bin Wu. Members of other groups in the Center for Integrated Systems have also enriched my experience at Stanford: Haechang Lee, Moon-Jung Kim, Sungwoo Kim, Paul Wang Lee, Yangjin Oh, Echere Iroaga, Jason Hu, Justin Kim, Wei Xiong, Pedram Lajevardi, Jaeha Kim, Elad Alon, Ki-Young Nam, Sangmin Lee, Hoyeol Cho, Chi-On Chui, Sotirios Limotyrakis, Xinying Ding, and Hopil Bae.

My bible study group has always been a great source of strength. I would like to thank Jungmin Ha, Boyoon Choi, Adrian Suhermann, Dolly Gusnawan, Haekang Lee, Donghyun Kim and Yong Suk Moon. I wish to express the most sincere thanks to many friends in my church.

I have had great joy with Daeho Lee, Munyoung Cho, Hyunsik Park, Sunghe Kim, Raymond Deutch, my former roommates Byoungsok Jung and Donghyun Kim, many friends in SSHS alumni, Gumbamtae, and KSAS (I apologize for this categorizing that seems to be in the nature of acknowledgments). Without them, my life at Stanford would have been a lot less fun.

None of this would have been possible without the generous financial support from Stanford Graduate Fellowship, Korean Ministry of Information and Communication, and CIS Seed Fund.

Finally, special thanks go to my parents, my brother and sister, and my wife. I thank them for their unconditional support and tireless prayer. Their voices on the phone have always provided me with a home away from home. With all my love and respect, I dedicate this dissertation to my family and the memory of my Mom.

Because of the Lord's great love we are not consumed, for his compassions never fail.
They are new every morning; great is your faithfulness.

–Lamentations 3:22-23

Table of Contents

List of Tables

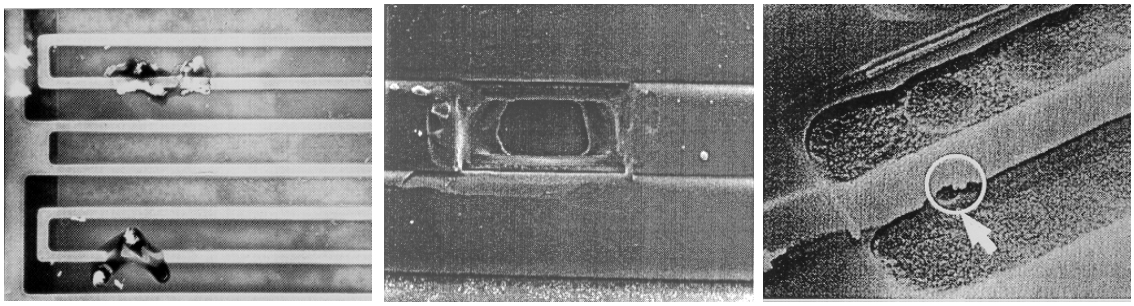
List of Figures

Chapter 1

INTRODUCTION

1.1 Background

Electrostatic Discharge (ESD) is a charge balancing process between two objects at different potential [1][2]. The phenomenon of ESD can often be observed in our daily lives. For example, static electricity can be generated due to the friction between different materials[1], and the accumulated electrostatic charge can spontaneously be transferred to the object at lower potential; either through a direct contact or through an induced electric field, just as when you reach a doorknob on one dry day. ESD events usually give a mild shock to human beings. However, if the same amount of ESD stress is injected into a microelectronic component, it could be detrimental.



(a) (b) (c)

Figure 1.1: ESD Failures in ICs: (a) junction breakdown, (b) metal/via damage, (c) gate oxide damage [3][4]

[1] This phenomenon is called a *triboelectric* charging. The potential induced by charges depends on the triboelectric property of materials.

ESD events often involve high voltage (~ several kV) and high current stress (1 - 10 A) on small electric devices. Despite the fact that ESD events are of very short duration (0.2 - 200 ns), the massive current/voltage pulses can give fatal damage to integrated circuits (ICs). Fig. 1.1 displays several examples of catastrophic ESD failures in the modern ICs, such as junction breakdown, molten metal/via effects and the gate oxide damage [3][4]. Besides these noticeable failures, minor levels of ESD shock can also generate latent problems within the devices, producing so-called 'walking wounded' devices [5].

In fact, a substantial number of IC failures are related to ESD/EOS[2] [6]-[8]. ESD/EOS damage is responsible for nearly one third of the failures of IC, and approximately 10 % of customer returns are exclusively due to ESD problems. Indisputably ESD is one of the most important quality and reliability concerns in the IC industry. Therefore, to make highly reliable IC products that are insensitive to ESD threats, ESD phenomenon must be well controlled through all phases of a device's life cycle. Despite the strong need for understanding of ESD phenomena, there has been a perception that ESD engineering is a 'black art'; ESD engineers used to provide solutions based on their experience without fundamental understanding of failure mechanisms. However, during the past twenty years, numerous studies have laid the foundation for a deeper understanding of the ESD phenomena, and based on that understanding, various on-chip ESD protection techniques have been developed [9].

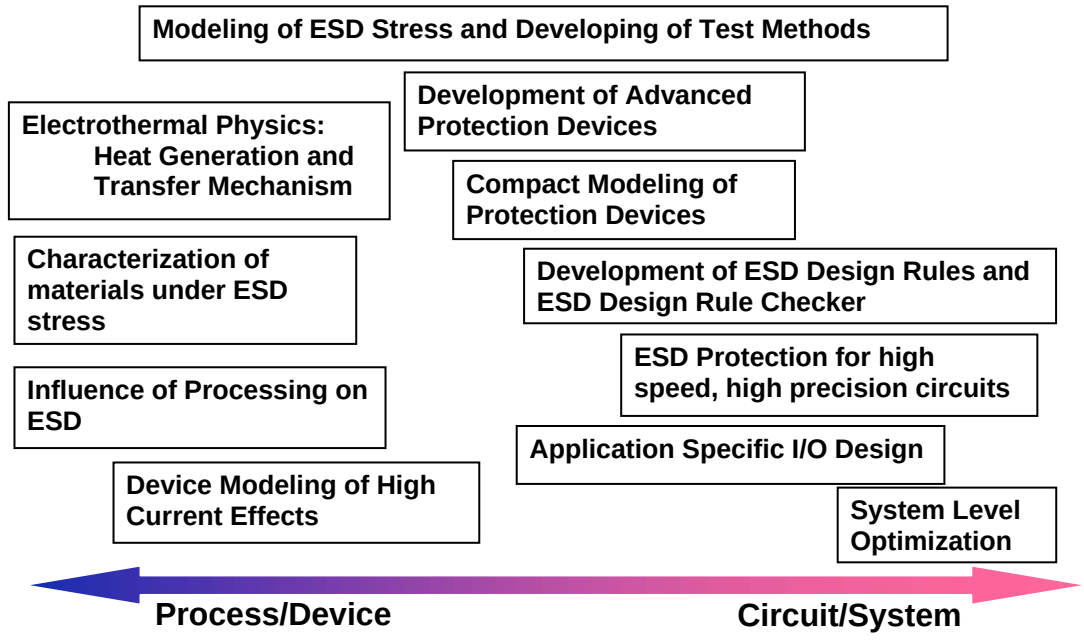


Figure 1.2: The scope of ESD research

[2] Electrical Overstress, ESD is a subclass of EOS

ESD studies can be regarded as a cross-discipline, bridging thermal, mechanical, and electrical physics. Fig. 1.2 summarizes the scope of ESD studies, which involve understanding from fundamental thermal physics to nanostructure devices, to circuits and systems. The details of this research agenda are explained below.

First, significant progress has been made to understand the fundamental aspects of different types of ESD phenomena and to investigate the implications of ESD events in IC technologies [10]. A variety of ESD phenomena have been categorized, leading to the development of several standardized models [11][12][13]. ESD events can be observed in various environments: silicon fabrication facilities, assembly and test sites, offices, home etc. Moreover, the various sources such as human beings, machines, and other harsh environments can cause ESD stress. Depending on the environment and the objects involved in the ESD events, the characteristics of ESD stress can vary considerably; the ESD failure mechanisms and signatures accordingly have variations. Proper ESD models can be of help in designing suitable protection schemes and in the development of test equipment which can effectively mimic the actual ESD events [14]-[17]. Details of ESD models and test methods will be presented in Chapter 2.

Second, the device characteristics under high current, high temperature conditions must be understood. The thermal physics of self-heating [18][19] and its impact on electrical characteristics have been thoroughly investigated using device simulation [20]-[23]. The impacts of geometrical changes on ESD performance have also been studied, contributing to the optimization of ESD protection devices [24][25]. However, the continuous device scaling has brought up the new physical phenomena of nano-scale conduction [26] and microscopic heat generation/transport [27]. These electro-thermal issues may be aggravated at the heating intensity that is common for ESD. On the other hand, as new materials are continuously introduced to main stream CMOS technology [28], there is an urgent need for experimental results including thermo-physical properties and verification of modeling. These experimental studies should be incorporated into the device simulation framework in order to predict the device performance under ESD conditions [29][30].

Third, the impact of process changes on ESD protection capabilities should be analyzed at each technology node. Technology scaling which leads to thinner gate oxides, shallower junctions, higher doping densities, narrower metal lines and vias, LDD, thin epi-substrates, shallow trench isolation (STI), silicided contacts etc. has had negative impact on ESD immunity. These challenges should be overcome by re-optimization of the protection devices or development of the novel designs [31][32]. Moreover, there is a growing need for an ESD protection scheme that is insensitive to technology variations.

Fourth, there have been many efforts to develop compact models of ESD protection devices containing the essential physics under ESD conditions [33]-[35]. To model MOS-based devices, the parasitic bipolar transistors within each device must be included in the compact models [36]. Circuit-level models of ESD protection devices have demonstrated good agreement with experimental data [35]. The SPICE-like circuit simulation models have several advantages compared to device-level simulations. For example, circuit level simulation takes much less time than device simulation, and the protected circuits can easily be included in ESD circuit simulation, giving insight about the behavior of both the ESD protection devices and the protected internal circuits under ESD conditions. On the other hand, the parasitic elements of ESD protection devices should be accurately modeled to predict their impacts on the performance of normal operating conditions. Modeling of the parasitic elements is gaining more importance as the operation frequency is rapidly increasing.

Fifth, there has been a growing need for ESD design rules and full-chip ESD design rule checker capabilities to support the complexity of modern VLSI chips [37]. The number of I/O pins is dramatically increasing, driving ESD protection circuitry to be more efficient and compact; therefore suitable design rules should be provided to make sure that the ESD protection device would successfully accomplish its job and to avoid an accidental coupling between ESD protection devices and internal circuits.

Sixth, recently the protection of high speed communication chips has posed serious challenges in ESD protection design, regarding ESD-to-circuit interactions [38]. While providing sufficient immunity to the ESD stresses, ESD protection devices should not affect the signal under normal operating conditions. However, the ESD protection devices introduce parasitic capacitance and resistance. The ESD parasitic elements can cause reflections of the signals and inefficient power transfer at the GHz frequency regime. The ESD devices can also generate noise or exacerbate the substrate noise coupling problems. These negative ESD-to-circuit impacts should be completely analyzed. Based on such analysis, proper ESD protection methodologies should be developed and a suitable strategy should be carefully selected, depending on the application of the ICs.

Seventh, along with the ESD-to-circuit interaction issue, ESD protection design requires more circuit/system level engineering. As the number of pads is dramatically increased, ESD protection devices take up a great portion of the overall chip area. To alleviate this area consumption problem, ESD protection elements should be minimized and some components such as power clamps between power-supplies need to be efficiently shared between I/O pads. For this purpose, we can benefit from system level optimization efforts [39]. On the other hand, there are

new challenges in some complicated applications such as multi-voltage ICs [40][41] and applications which need 'fail-safe[3]' capabilities [42].

1.2 Thesis Outline

Although significant progress has been made in understanding ESD and solving ESD related problems, there are still many challenges in ESD engineering of advanced CMOS technologies. This thesis is designed to solve some of the issues raised in the previous section. It aims to improve design guidelines for higher ESD reliability and to provide detailed insight into ESD behavior of advanced devices. It will cover various ESD problems recently brought up in nano-scale devices and high performance circuits as follows.

Chapter 2 presents a general overview of ESD studies to provide preliminary information for better understanding of this dissertation. ESD stress models, a variety of standard test methods and conventional on-chip ESD strategies are reviewed.

Chapter 3 investigates the impact of ESD protection devices on the performance of RF ICs and discusses design methodologies to minimize this impact. The RF performance degradation due to ESD protection devices is numerically analyzed. Recently reported studies which tried to overcome this ESD-to-circuit impact will also be reviewed. In Chapters 3.2 and 3.3[4], a 2 GHz narrow band low noise amplifier (LNA) and a 5 GHz broad band LNA will be demonstrated. Based on RF and ESD tests, we will discuss the concept of RF/ESD co-design and compare it to conventional RF ESD protection methodology that focuses on minimizing the device size.

Chapter 4 deals with another ESD-to-circuit impact. ESD protection circuits typically contain a significant amount of nonlinear capacitance. At high frequencies and large amplitudes, this nonlinearity can degrade the signal integrity at the input pins of high performance mixed signal ICs, such as analog-to-digital converters (ADCs). This study provides a theoretical analysis of this problem as well as experimental results that quantify typical distortion levels introduced by state-of-the-art ESD structures. This chapter also presents guidelines for designing ESD protection circuits suitable for high-speed, high-linearity applications.

In Chapter 5, studies of PMOS transistors in modern CMOS technologies are reviewed, focusing on the snapback and breakdown behavior of parasitic PNP BJTs in the high current regime. A new failure mechanism for PMOSFET devices under ESD conditions is also analyzed by

[3] often called 'hot-plug.'
[4] Chapters 3.2 and 3.2 are outcome of the collaboration with W. Soldner, M. Streibl, U. Hodel, M. Tiebout, H. Gossner, D. Schmitt-Landsiedel, C. Ito, and R. W. Dutton [50].

investigating various I/O structures in a 0.13 μm CMOS technology. Based on 2-D device simulations, the critical layout parameters affecting this problem are identified. Design guidelines for avoiding this new PMOSFET failure mode are also suggested.

In Chapter 6, various thermal issues of nano-scale devices are reviewed with an emphasis on the relevance to ESD simulations. After that, an example of an ESD electrothermal simulation will be demonstrated; the electrothermal characteristics of strained-Si MOSFETs operating in the high-current, high temperature regime is investigated using device/circuit mixed mode simulations. To take the nano-scale heat generation and phonon heat conduction into account, the optical and acoustic phonon temperature system for TCAD simulation is developed and self-consistently coupled with the hydrodynamic transport model. This electrothermal model is validated through simulations for a thin-body SOI NMOSFET.

Finally, Chapter 7 summarizes the contributions of this dissertation and discusses recommended research for present and future generations of IC technology design.

Chapter 2

ESD TEST & PROTECTION METHODOLOGIES

In this chapter, standardized ESD stress models and test methods are briefly overviewed to provide background information that is relevant to the dissertation. Two common on-chip ESD protection schemes are described and comparison between them is also presented.

2.1 ESD Stress Models and Test Methods

As discussed in Chapter 1, a variety of charged sources can be involved in ESD events. Several ESD models have been developed and standardized according to the nature of the ESD stress sources. Using these models, which are based on the physics of actual ESD events, ESD robustness of devices under different discharge processes can be characterized systematically. In reliability qualification used by the IC industry, three basic models are dominant: Human Body Model (HBM) [11], Machine Model (MM) [12], and Charged Device Model (CDM) [13]. Each of these is now summarized.

One of the most frequently observed ESD events is the transfer of electrostatic charge from a charged human body to an ESD sensitive device due to improper handling. The model developed to represent this event is the Human Body Model (HBM), which is the most classical and common industrial test methods. In the HBM, it is assumed that a certain amount of electrostatic

charge initially is stored on the body and the charge is transferred to an object through a finger when the physical contact between the charged human body and the object is made[5].

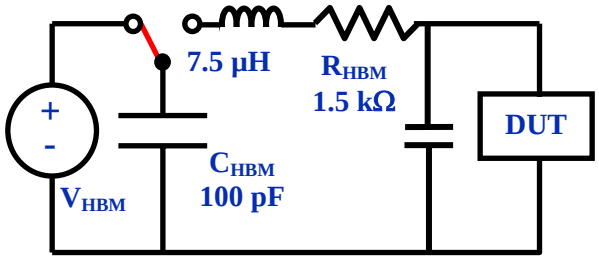


Figure 2.1: Human Body Model (HBM)

Fig. 2.1 shows a simplified equivalent circuit of HBM ESD conditions. It consists of a charging capacitor, and contact resistance between the charge source and the Device Under Test (DUT). In the HBM standard [14], the circuit component used to simulate the charged human body is a 100 pF capacitor and the resistance of the discharging path is 1500 Ω; it electrically looks like a current source if the DUT provides a current path of low resistivity. L_{HBM} (~0.75 µH) is the effective inductance of the discharge path in a real tester. HBM has the longest pulse among the three primary ESD models. The rise time of the HBM pulse is approximately 5-10 ns, and the decay time is ~150 ns[6].

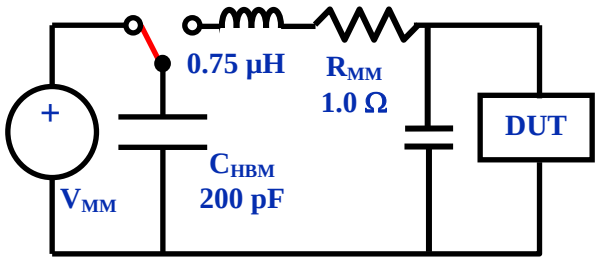


Figure 2.2: Machine Model (MM)

Another fundamental model widely used in the industry is the Machine Model (MM). Similar to HBM event, a capacitor (200 pf) which represents a conductive object such as a metallic handler is charged up to a high voltage and then discharged through the pins of an IC. In this

[5] HBM and MM tests involve two pins on a module. An ESD pulse is injected into one pin with another pin grounded and all other pins floating.
[6] An estimate of the rise time can be given by $2L_{HBM}/R_{HBM}$, and the characteristic decay time is associated with the time of the network, $R_{HBM}C_{HBM}$.

model, it is assumed that an arc discharge occurs between the charged source and the DUT [10]. An arc discharge fundamentally has a resistance of 10-20 Ω which is much lower than the R_{HBM} (~1500 Ω). Therefore, the MM response is more rapid than the HBM event, and has a form of bi-directional damped oscillation. The MM event also shows a significantly higher current than the HBM[7]. Since R_{MM} in Fig. 2.2 is very small, the parasitic capacitance and inductance of test equipment and the impedance of the DUT may noticeably change the current waveform; therefore, the standardization of MM [15] is more difficult comparing to that of HBM. However, the failure signature of the MM is generally the same as that of the HBM since discharge processes are similar in both types of ESD events. In many cases, MM ESD robustness could be guaranteed by the HBM test. Experimentally, the sustainable stress level of MM ESD (V_{MM}) has been found to be approximately 10 times lower than that of HBM ESD (V_{HBM}).

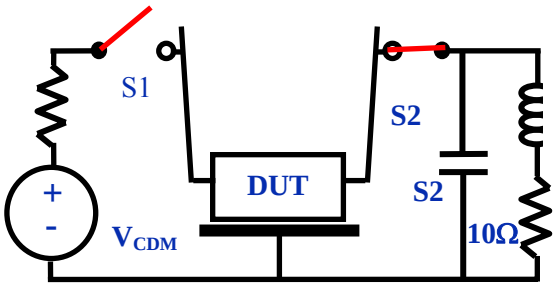


Figure 2.3: Charged Device Model (CDM)

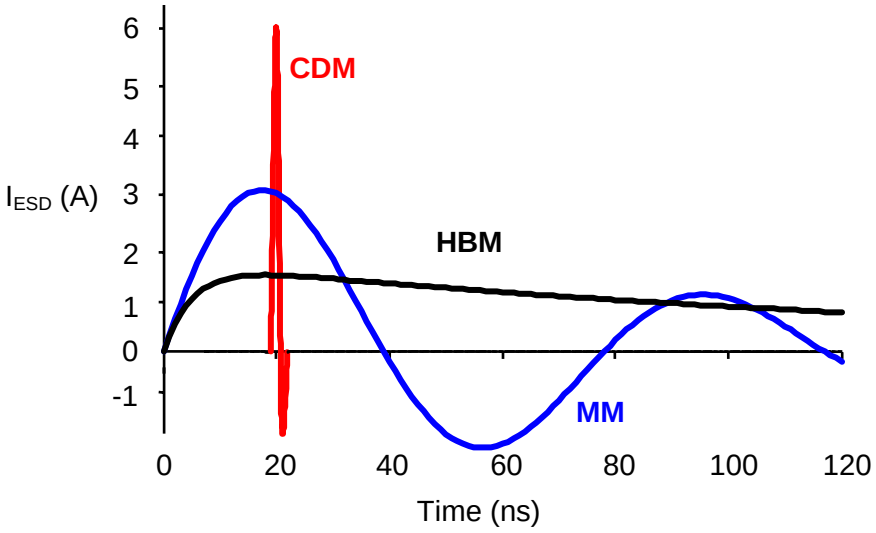


Figure 2.4: Idealized current waveforms for HBM, MM, and CDM ESD events

[7] The MM often is considered to be the worst case HBM.

The CDM event charges a chip and then discharges to ground out of a single pin. While the HBM and MM events occur between two pins, the discharging process of the CDM events involves a single pin on the module. The charging process can be initiated either by direct charging, or field-induced charging. Electrostatic charge can be accumulated on the device through triboelectric effects, which frequently occur when a device moves across a surface or vibrates in a package. The discharge process can be initiated as electric contact is made between the charged device and the discharging means such as automated handlers. In an actual test environment [16], a filed plate imposes a specified charge on the packaged module, and then a metallic probe touches a pin, forming a low resistive path from the pin to ground. As shown in Fig. 2.4, the CDM event occurs in an extremely short time interval[8], but generates very high current. CDM discharge occurs at less than 5 ns where typically the rise time of the event is of the order of 250 ps. Since the waveform of CDM events are different from HBM and MM, the failure signatures are not identical either. While thermal destruction is the primary failure mechanism in the HBM and MM, dielectric failure is the typical failure mode in the CDM type ESD.

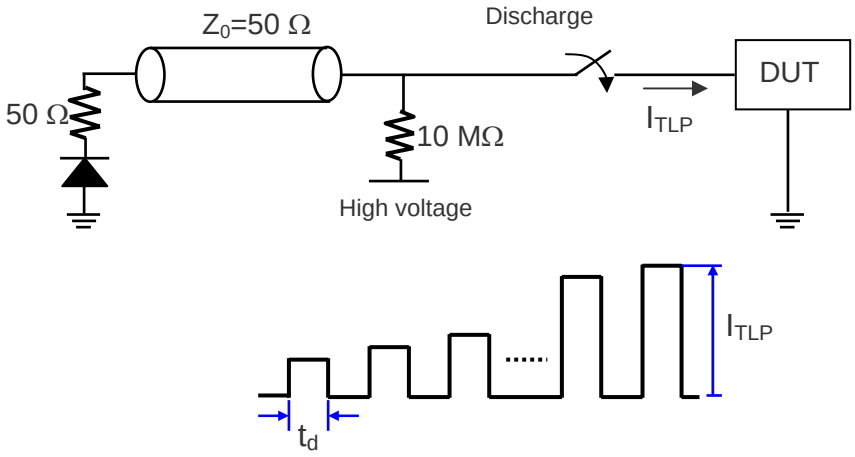


Figure 2.5: Circuit representation of the TLP setup and principle of pulsed characterization.

Besides the three primary ESD standards described above, the Transmission Line Pulsing (TLP) [17] technique has seen considerable applications for the development and characterization of ESD protection. ESD events occur within the time period of a few nanoseconds to hundred nanoseconds. Therefore, it is important to characterize behavior of devices on the ESD-relevant time scales. Since DC measurements can cause severe self-heating, and cannot address the

[8] The time scale of CDM is the order of the modern electronic circuit response times.

transient characteristics of devices under ESD conditions, pulsed measurement techniques are required for ESD characterization.

The basic concept of TLP is to apply a square wave to the DUT and then measure the current and voltage across the device. The schematic of a TLP system and the principle of pulsed characterization are illustrated in Fig. 2.5. A transmission line is charged to a specified test voltage and then discharged through the DUT when the switch closes. The pulse width of a transmission line pulse is determined by the length of the transmission line and the propagation velocity of the transmission line; therefore, the pulse width as well as amplitude of input voltage to the device can be controlled by the physical length of the transmission line and the initial voltage on the transmission line. In standard practice today, the TLP cable length is chosen to provide a TLP pulse width of 100 ns with less than 10 ns rise time[9].

To investigate and characterize ESD behavior of protection devices, the transmission line pulsing (TLP) have been used throughout this work[10].

ESD Model	Industry Standard
HBM	2000 V
MM	150 – 200 V
CDM	500 V

Table 2.1: Industry standard ESD protection levels

The requirement of ESD protection levels can be varied depending on the environment where products are controlled and handled. Table 2.1 summarizes the general industrial standard for the modern ICs. This standard has been adopted through observing the correlation between ESD test levels and actual ESD failure rate. It should be noted that some ICs require higher protection levels, and some ICs may allow lower ESD immunity levels.

[9] This choice of pulse width is determined to use TLP testing as an equivalent method to the HBM test. 30 ns pulse can be used for simulating MM. To investigate the devices under CDM, Very Fast TLP (VFTLP) with 1 – 3 ns pulse width can be employed.
[10] The ESD tests in Chapters 3 and 5 are performed using TLP.

2.2 On-Chip ESD Protection Methodologies

The ESD related reliability problems can occur during the manufacturing, shipping, receiving, and field handling of ICs or circuit boards. To increase manufacturing yields, reduce overall cost and improve the reliability of IC products, ESD events should be avoided or protected against. In general, there are two ways to reduce IC failures under ESD conditions. The first approach focuses on reducing the amount of ESD induced charges and redistributing them through proper handling of devices and controlling the handling environments. The second approach is to implement on-chip protection circuits in order to improve ESD robustness of the individual circuit components.

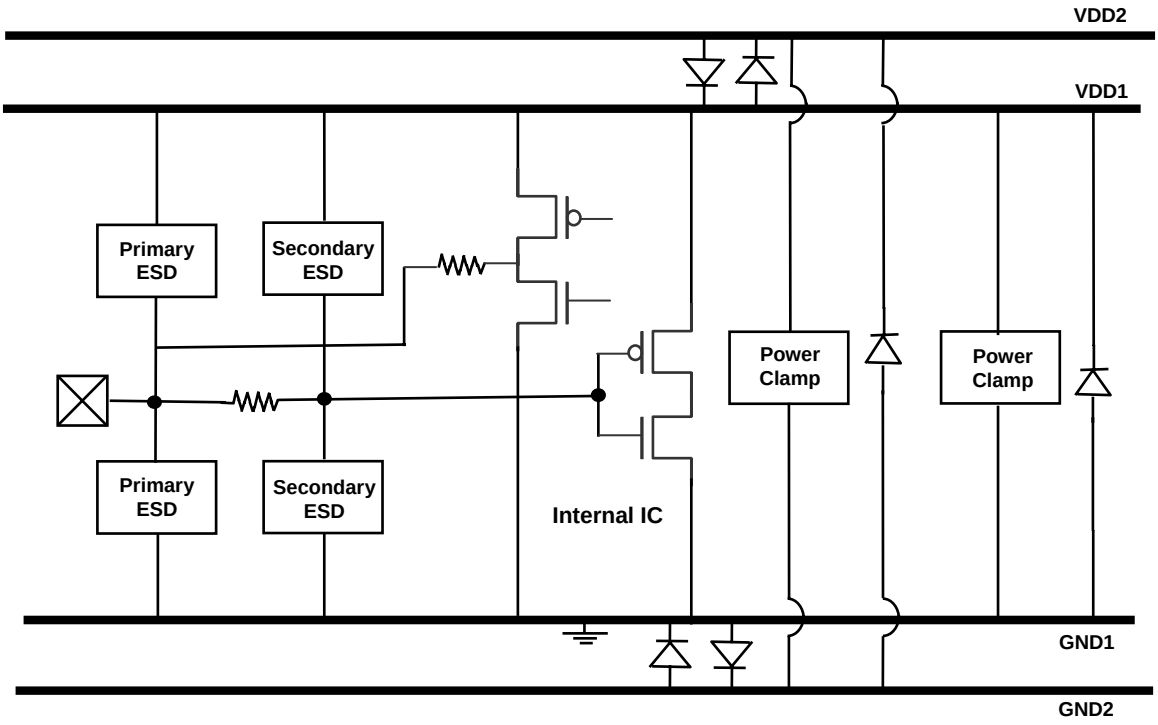


Figure 2.6: A general configuration of the ESD protection in a bidirectional I/O circuit.

On-chip ESD protection should provide a current path between any two pins of an integrated circuit. In Fig. 2.6, a generic bidirectional I/O circuit with ESD protection is illustrated. The function of the primary ESD is to limit the I/O pad voltage below the failure level of the output driver through bypassing most of the ESD stress current to the power rail. The secondary ESD device is often a downsized version of the primary ESD device. The secondary ESD device with a series resistor between the primary and secondary ESD can lower gate voltage of the input

receivers. This auxiliary protection with the secondary ESD device is indispensable for CDM protection. The power clamps can absorb direct ESD stress between V_{DD} and GND, protecting the devices in the I/O and internal circuits[11]. Finally, there are anti-parallel diodes between GND_1 and GND_2, and between V_{DD1} and V_{DD2}. These diodes provide current paths when ESD stress is induced between two different power domains.

Most ESD solutions rely on shunting charge from an I/O pin to a power supply (V_{DD} or GND), from which the charge can be distributed to other I/O pins or supplies. These solutions therefore fall into two general categories; pad-based ESD protection for ESD approaches that shunt current to GND and rail-based protection for those that shunt current to the V_{DD} power rail. In the following sections, these two common on-chip protection methods are presented in details and the advantages and drawbacks of each method are discussed.

2.2.1 Pad-based ESD Protection

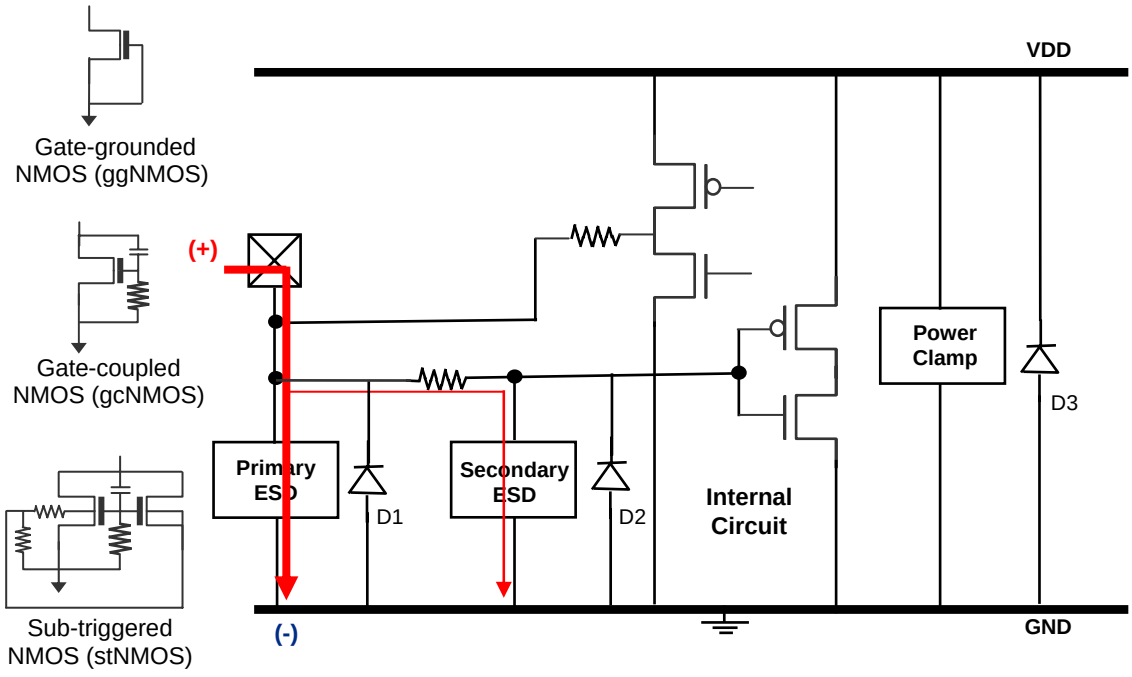


Figure 2.7: The concept of the pad-based ESD Protection and examples of protection devices in the pad-based method. The ESD current is directly shunted from the I/O pin to GND.

[11] As will be discussed in Section 2.2.2, the power clamp provides a crucial current path in rail-based protection

In the pad-based ESD protection approach, a primary ESD device exists on every I/O pad between the pad and the ground. In Fig. 2.7, the current path for a positive discharge from the I/O pad to GND is illustrated. The ESD current flows directly from the I/O pad to GND through the ESD devices; therefore, this protection scheme is called 'pad-based ESD protection.' In this scheme, the primary clamp alone must hold the I/O pad voltage below the failure voltage of I/O devices in parallel.

Between every pair of pins, there must be a solid ESD current path designed to carry high current for both possible polarities of the discharge. For example, when positive ESD stress is applied to the I/O pad with respect to V_{DD}, in addition to the primary and secondary ESD devices, the diode D3 is involved in the current path. Another example is a negative discharge from the I/O pad to V_{DD}. In this case, the ESD current path consists of the power clamp and the diodes[12] (D1 and D2).

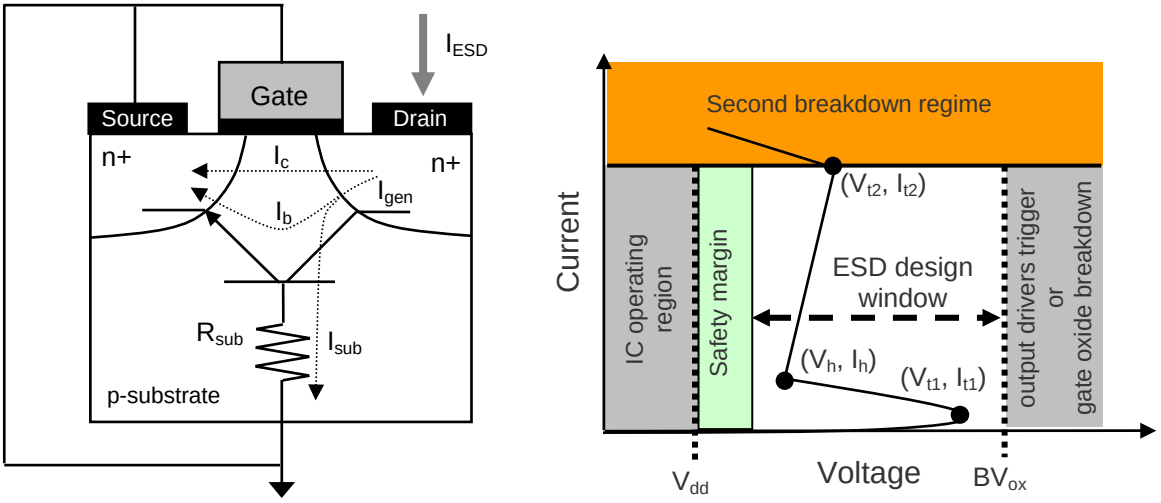


Figure 2.8: (a) Operation of the gate-grounded NMOS (ggNMOS) under ESD conditions [45] (I_{gen}: the generation current due to impact ionization, I_{sub}: the substrate current, I_c: the collector current, I_b: the base current), (b) *I-V* characteristics and ESD design window V_{t1}/I_{t1}: the triggering voltage/current, V_h/I_h: the holding voltage/current, V_{t2}/I_{t2}: the second breakdown triggering voltage/current, BV_{ox}: breakdown voltage of gate oxide).

In the pad-based scheme, ESD protection devices are usually snapback devices. In CMOS technology, NMOS-based devices have been widely used with several configurations such as the gate-grounded NMOS (ggNMOS), the gate-coupled NMOS (gcNMOS) [31][43], and the

[12] Parasitic P-body/N-diffusion diodes in NMOSFET drivers can be utilized for ESD protection diodes (D1 and D2 in Fig. 2.7). However, in SOI process, explicit diodes are required since the body resistance of a SOI NMOSFET is very high.

substrate-triggered NMOS (stNMOS) [32][44]. A silicon controlled rectifier (SCR) can also serve as an excellent ESD protection device since it has a very high current handling capability arising from the combination of two bipolar transistors in a self-regenerative on condition.

ESD protection with MOS-based devices makes use of parasitic bipolar transistors in MOSFETs [45]. The operation of ggNMOS under ESD is depicted in Fig. 2.8a. Once the ESD event occurs, the ESD induced current charges up the capacitance at the drain terminal of the ggNMOS; therefore, a high electric field builds up across substrate to drain junction. This high electric field results in impact ionization and finally an avalanche multiplication process is initiated. As avalanche generated holes flow towards the substrate (I_{sub}), the base potential of the lateral NPN transistor is elevated due to a voltage drop across the effective substrate resistance, R_{sub}. As the substrate current, I_{sub} increases with increased drain current, the base potential finally reaches the turn-on voltage of the lateral NPN[13] transistor, and then the NPN transistor triggers and enters a self-biasing mode. The drain voltage (current) at this triggering point is V_{t1} (I_{t1}). At this condition, ESD current flows through the substrate bulk and the current driving capability of the ggNMOS is significantly increased; hence, the on-resistance is significantly reduced and the device shows a snapback behavior as depicted in Fig. 2.8b.

As the drain current further increases, the ggNMOS enters the second breakdown regime at V_{t2} (I_{t2}). Second breakdown is generally caused by thermal runaway due to the negative thermal coefficient of bipolar transistors' on-resistance. The second breakdown results in irreversible thermal damage to the device; it gives rise to a large junction leakage current.

If the ESD protection devices are directly attached to the gate of the input receiver, then the triggering voltage should be lower than the gate oxide breakdown voltage, which is dependent on the duration of the ESD stress and the thickness of the gate oxide. In output drivers or bi-directional I/Os, the ESD clamp must keep parallel devices, the most sensitive of which is usually the nMOS output driver, from snapping back and failing. Therefore, an ideal clamp should trigger below the triggering voltage of the nMOS output driver and should hold the pad voltage below the holding voltage of the nMOS driver. To avoid accidental triggering of the ESD protection devices in normal operation, the ideal clamp should also limit the pad voltage to a level higher than the voltage seen on the pin during worst-case operating conditions. This safety margin must be determined regarding noise and voltage overshoot. These requirements are summarized in Fig. 2.8b showing the ESD design window.

[13] In Chapter 5, the experimental data of PNP transistors within PMOSFETs are provided. The operation mechanism of PMOSFETs under ESD conditions is basically same as that of NMOSFETs.

2.2.2 Rail-based ESD Protection

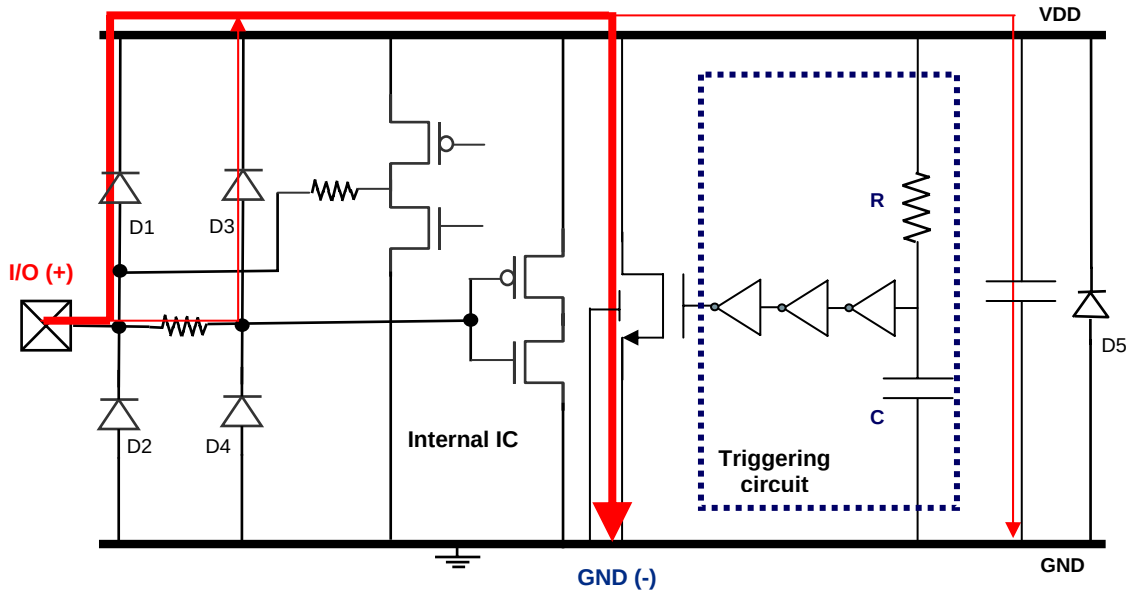


Figure 2.9: Concept of rail-based ESD Protection. ESD current is redirected to the V_{DD} power rail and then shunted to GND by a power clamp.

Fig. 2.9 shows an example of the rail-based protection in which the primary dual diodes[14] (D1 and D2), the secondary dual diodes (D3 and D4), the NMOSFET power supply clamp are employed. The power clamp provides a current path from the power supply pad to the ground pad during various ESD events. The simplest form of the RC-triggered power supply clamp is shown in Fig. 2.9, where the RC network and the inverter chain are designed such that the NMOSFET (M1) remains off during normal operation and turns on during an ESD event. This NMOSFET should be wide enough to handle the ESD current when it is on during an ESD event.

When the positive ESD stress is induced to the pad with respect to GND, there is no direct current path as in the pad-based protection in Fig. 2.7. However, as shown in Fig. 2.9, the current must flow onto the V_{DD} rail through the primary ESD device (D1 in Fig. 2.9) and through the V_{DD} power supply clamp (M1) to reach GND. Therefore, two devices in series (D1 and M1) must meet the voltage-limiting criterion to protect the internal circuits. For a negative pulse with respect to GND, the current passes through the lower diode (D2).

For a positive pulse with respect to V_{DD}, the current passes through the upper diodes (D1, D3) to the supply pin. For a negative pulse with respect to V_{DD}, the current enters the V_{DD} supply pin,

[14] In bulk CMOS processes, D1 and D3 in Fig. 2.9 are emitter-base diodes of vertical PNP transistors. However, in SOI processes, D1 and D3 are P-diffusion/N-body diodes.

flows through the power clamp, and then passes through the diodes (D2, D4) connected from GND to the pin.

The effectiveness of the I/O protection in a rail-based protection approach is determined by the placement of the power supply clamps; therefore the power supply clamps should carefully be distributed and the bus resistance from I/O pins to the power clamps should be accurately estimated.

2.2.3 Pad-based versus Rail-based Protection

In modern CMOS technologies, the selection decision often reduces to the two methods in most widespread use: either the pad-based strategy using an nMOS in lateral NPN snapback with supporting circuits to ensure uniform finger triggering, or the rail-based strategy using diodes (or vertical PNP) and power supply clamps.

As illustrated in Fig. 2.7, the design of the pad-based ESD protection is very straightforward, since only the protection device is involved in the most critical ESD events (positive discharge from the I/O to GND). In a compact design, ESD protection devices can be self-contained in I/O pads. In general, the MOS-based snapback device is immune to false triggering and robust if properly optimized. Moreover, it is suitable for fail-safe applications, because there is no diode between the I/O and VDD power rail.

However, the pad-based ESD scheme has several drawbacks. In the MOS-based snapback devices in advanced CMOS technologies, it is not easy to achieve acceptable I_{T2}. The advances in the process technologies have posed many challenges for the design of nMOS-based ESD protection circuitry. For example, LDD, silicided diffusion[15], thin epi substrate and shallow trench isolation (STI) have had major impacts, severely reducing I_{T2}. Especially, silicided diffusion has made the I_{T2} of normal ggNMOS virtually zero since it eliminates the ballasting resistance [22], exacerbating the non-uniform conduction between fingers and even within a finger. These obstacles have been overcome with aggressive device engineering, producing several advanced devices such as gate-coupled NMOS, substrate-triggered NMOS etc. However, in every generation of the process, it is expected to have a new challenging problem. In general, the characteristics of the snapback style devices are very sensitive to process variations, and they are not portable from one process to another. On the other hand, the compact circuit model of the snapback devices are not often provided; therefore, it is difficult to verify the ESD protection scheme with other protected

[15] With the addition of silicide blocking masks, the nMOS forms a robust, reliable and portable protection device. However, the drain region can considerably be enlarged to obtain sufficient ballasting resistance, resulting in large parasitic capacitance and high voltage drop under ESD conditions.

circuits in circuit simulations.

On the contrary, the rail-based ESD protection is much less susceptible to process variation, and relatively portable from one fabrication process to another. Since the rail-based scheme often uses conventional devices such as diodes[16] and MOSFETs, ESD protection circuit can be verified by SPICE simulations, and in general these devices are easily scalable to more advanced technology comparing to the snapback devices in the pad-based protection. Since any protection scheme must still include a clamp on the power supply, rail-based protection may take advantage of the power supply clamp.

However, as the number of I/O pins increases, additional power clamps are often needed only for ESD protection purpose, consuming large areas and possibly causing large leakage current. The power clamp has a risk of false triggering, especially in power-up sequence or in a large power supply noise environment; the engineering efforts to avoid false-triggering are also in need. On the other hand, because the rail-based scheme includes a forward diode to the VDD rail, it creates power supply sequencing issues and cannot be compatible with a fail-safe system unless special steps are taken to float V_{DD} bus of the power-downed IC.

Which strategy is better depends on many factors. If a process provides robust and reliable snapback devices, and then the pad-based protection can be a good candidate. In general, with an I_{T2} higher than 5 mA/μm, the nMOS in snapback can be very area efficient. However, below 0.25 μm technology, the design of a reliable nMOS ESD protection device demands substantial device engineering efforts without silicide blocking or extra ESD implantation to control the triggering voltage. Therefore, the pad-based protection strategy is adopted mostly in companies that have controllability for their internal fabrication facilities. On the other hand, the rail-based strategy is more suitable to the segments of the industry that rely on external foundries. If the process provides high beta vertical PNP device or low resistive diodes, then the rail-based protection can be a good candidate. In particular, if the regular shrinks or process change are expected, the rail-based protection is a better choice since it is less sensitive to process variations and relatively portable to a different process.

The issues in this thesis are relevant to both protection strategies. The examples in Chapters. 3 and 4 deal with both the pad-based and rail-based protection schemes, and the analysis about the parasitic component of ESD devices is applicable to both protection methodologies. The ESD protection scheme discussed in Chapter 5 is rail-based. However, it deals with the snapback of internal devices which is also a common issue in pad-based ESD protection. Chapter 6 is more

[16] In bulk CMOS processes, vertical PNP transistors must be accurately modeled for ESD simulation.

relevant to the pad-based ESD, since this chapter explores the ESD robustness of advanced devices in snapback, and thermal issues that are more aggravated in snapback devices.

Chapter 3

ESD PROTECTION FOR RADIO FREQUENCY INTEGRATED CIRCUITS

In multi-GHz RF ICs, ESD protection devices introduce considerable parasitic capacitance and resistance to inputs and outputs, thereby degrading the RF performance, such as input/output matching, gain, and noise figure.

In this Chapter, the impact of ESD protection devices on the performance of RF ICs is investigated and design methodologies to minimize this impact are discussed. In Section 3.1, the RF performance degradation due to ESD protection device is numerically analyzed. Recently reported studies to overcome this ESD-to-circuit impact are also reviewed. In Sections 3.2 and 3.3, a 2 GHz narrow band LNA (low noise amplifier) and a 5 GHz broad band LNA are demonstrated. Dual diode and SCR ESD protection devices with a variety of sizes are utilized in the LNAs. Performing the RF and ESD tests, we will discuss the concept of RF/ESD co-design and compare it to the conventional RF ESD protection methodology which focuses on minimizing the device size.

3.1 A Review of ESD Protection Design for RF ICs

As the demand for wireless RF and high speed mixed-signal systems is rapidly increasing, on-chip ESD protection design for these systems has posed a tremendous challenge. While providing sufficient immunity to the ESD stresses, ESD protection devices should not affect the

signal under normal operating conditions. However, the ESD protection devices introduce parasitic capacitances and resistances, and the capacitance associated with the ESD protection devices can be several pF. In the GHz frequency regime, the reactance due to this large capacitance becomes comparable to the characteristic impedance at the interfaces (typically 50 Ω), causing reflections of the signals, inefficient power transfer etc. The ESD devices can also generate noise or exacerbate the substrate noise coupling problem [46][47].

Because of these negative effects on the circuit performance, there used to be a "sign-off waiver" for the ESD protections of RF ICs, which means that no ESD protection or only limited size of ESD protection devices were installed at the inputs of RF ICs. However, due to the integration of today's complex mixed-system in CMOS technologies, there is no longer any differentiation between RF pins and digital pins; therefore, there should be no difference in their ESD performance. As the number of RF pins per device increases, the ESD robustness of the RF pins has become a critical factor in determining the yield ratio. Thus a good methodology to provide sufficient ESD protection capability with tolerable interference to the RF performances is required. To develop a good RF ESD protection scheme, at first the nature of ESD-to-circuit impact [47] should be understood.

3.1.1 RF Modeling of ESD Protection Devices

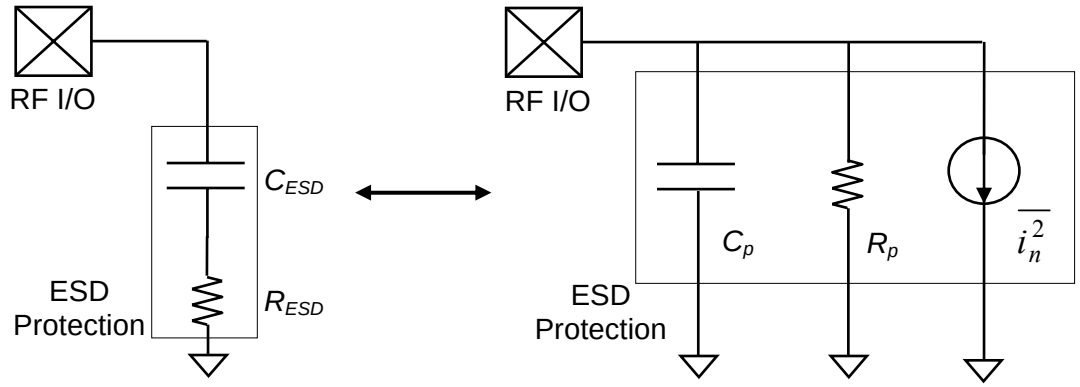


Figure 3.1: (a) A simplified model of ESD protection devices and (b) its parallel configuration

The compact models of ESD devices can be simplified to a series connection of a capacitor (C_{ESD}) and a resistor (R_{ESD}), as depicted in Fig. 3.1a. At a given frequency (ω), this series configuration

can be transformed to the parallel configuration in Fig. 3.1b. The quality factor (Q) is defined as in Eq. 3.1. Assuming Q >> 1, R_p and C_p in Fig. 3.1b can be expressed as in Eqs. 3.2 and 3.3.

$$Q = \frac{1}{\omega R_{ESD} C_{ESD}}, \qquad \text{Eq. 3.1)}$$

$$R_p \approx Q^2 R_{ESD} = \frac{Q}{\omega C_{ESD}} = \frac{1}{\omega^2 R_{ESD} C_{ESD}^2}, \qquad \text{Eq. 3.2)}$$

$$C_p \approx C_{ESD} \qquad \text{Eq. 3.3)}$$

In Fig. 3.1b, a thermal current-noise source ($\overline{i_n^2}$) is also included, and it can be described as follows,

$$\overline{i_n^2} = \frac{4kT\Delta f}{R_p} = 4kT\Delta f \omega^2 R_{ESD} C_{ESD}^2, \qquad \text{Eq. 3.4)}$$

The simple *RC* model in Fig. 3.1 is suitable for most of ESD devices, especially for the diode structures. More sophisticated models may be needed to deal with more complex devices. Fig. 3.2 shows an RF model for a gate-grounded NMOSFET, which has drain-gate capacitance (C_{dg}), parasitic drain resistance (R_d), gate resistance (R_g), drain-body capacitance (C_{db}), and substrate resistance (R_s). However, at a specific frequency level, this complex *RC* model can also be simplified to one resistance and one capacitance by successive impedance transformations.[17]

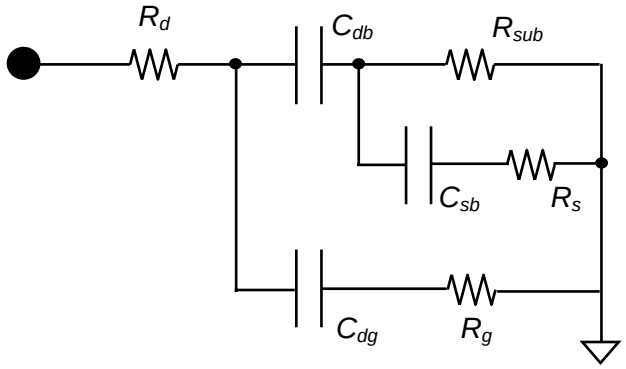


Figure 3.2: A compact RF model of gate-grounded MOSFETs.

[17] In the model in Fig. 3.2, the noise sources are not illustrated. The noise modeling can be accomplished based on MOSFET noise theories [47][48].

First, we will discuss the impact of R_p and C_p on the RF performance; the impact of the noise generated by the ESD protection devices will be discussed in Section 3.1.3. Depending on the type of on-chip ESD protection devices, C_{ESD} can range from several hundred fF to several pF. Recently, approximately 200 fF protection devices with a protection level higher than 2kV HBM have been reported [42]. In general, we can ignore the loading effect of R_{ESD}, because R_{ESD} is usually larger than 1 kΩ. However, the loading effect of parasitic C_{ESD} at the multi-GHz regime can significantly alter the input/output matching, degrading the power-transfer-efficiency in RF ICs. C_{ESD} also limits the bandwidth of high speed digital I/Os.

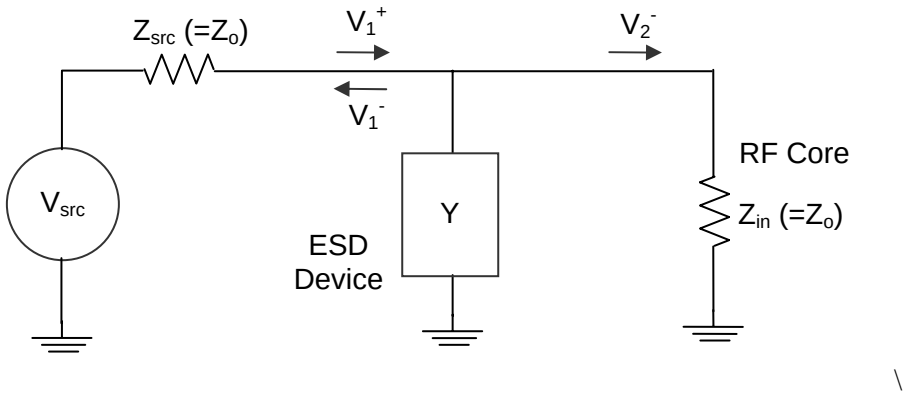


Figure 3.3: A two-port model of an RF I/O with an ESD protection device. The admittance, Y represents the ESD protection device.

Fig. 3.3 illustrates a generic example of RF I/O circuits with an ESD protection device. The RF input impedance (Z_{in}) is perfectly matched to the source impedance (Z_{src}), e.g. Z_o (typically 50 Ω); then ESD protection devices are added between the input source and the matching network. As mentioned before, at high frequencies, C_{ESD} alters the input matching and degrades S_{11} and S_{21}: S_{11} is the input refection s-parameter, the ratio of the reflected signal (V_1^-) to the incident signal (V_1^+); S_{21} is the forward transmission s-parameter, the ratio of the outgoing signal (V_2^+) to the incident signal (V_1^+). S_{11} and S_{21} for the system in Fig. 3.3 can be expressed as follows:

$$S_{11} = \frac{V_1^-}{V_1^+} = \frac{-Z_o Y}{2 + Z_o Y}, \qquad \text{Eq. 3.5)}$$

$$S_{21} = \frac{V_2^-}{V_1^+} = \frac{2}{2 + Z_o Y} \qquad \text{Eq. 3.6)}$$

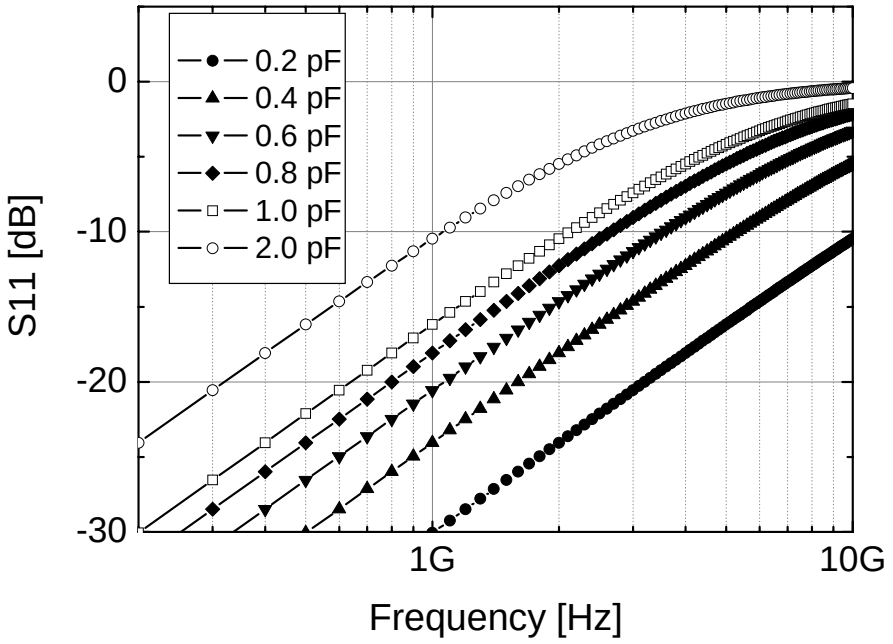


Figure 3.4: S_{11} of the two port system in Fig. 3.3 with a variety of C_{ESD}

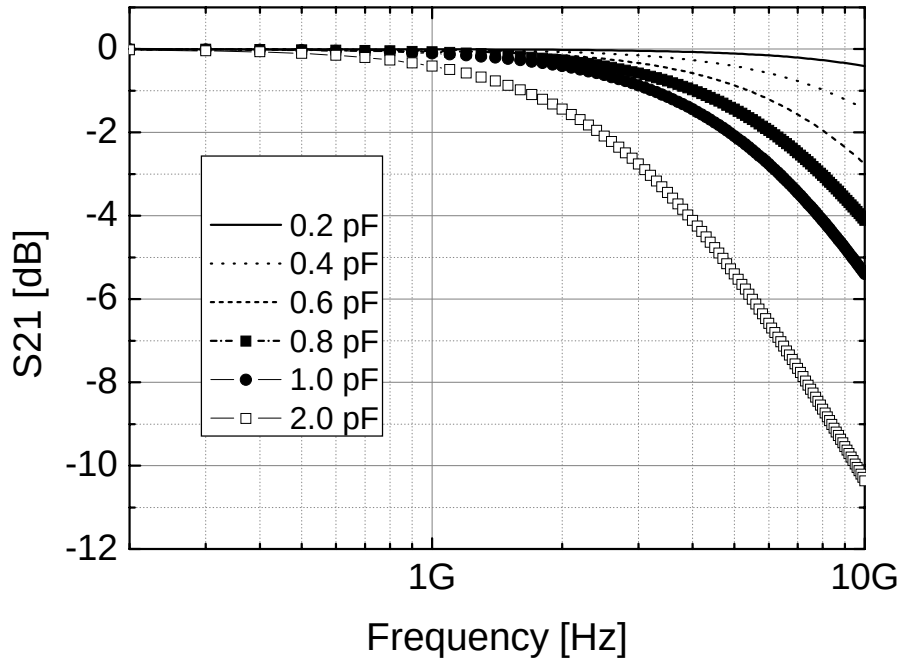


Figure 3.5: S_{21} of the two port system in Fig. 3.3 with a variety of C_{ESD}

S_{11} of the system in Fig. 3.3 is plotted in Fig. 3.4. With the C_{ESD} of 200 fF, S_{11} is below -10 dB even at a signal frequency of 10 GHz. However a 1 pF C_{ESD} causes a considerable signal reflection; resulting in -10 dB S_{11} at approximately 2.3 GHz. Fig. 3.5 shows the power loss represented by S_{21} versus frequency. With 200 fF C_{ESD}, the power loss is 0.54 dB at 10 GHz. However, if C_{ESD} is larger than 1 pF, the power loss increases to over 5.5 dB. That is, the power loss is extremely sensitive to the size of ESD protection devices at multi-GHz frequencies. This loading effect of C_{ESD} has been fully analyzed by Ito et al. [49] with a single ESD device and multiple distributed devices. It should be pointed out that for the analysis in Figs. 3–5, we have not included the capacitance and inductance of the package, which could be larger than pf and nH levels, respectively. However, assuming a certain package type and the relevant parasitic impedance, we can apply a similar approach to estimate the impact of ESD protection devices on input/output impedance matching and bandwidth limit.

3.1.2 ESD Protection Strategies for RF ICs

One of the straightforward methods to address the matching alteration problem due to C_{ESD} is minimizing the size of ESD protection devices. This method is often called ‘low-C’ ESD design. This ‘low-C’ method implies that the ESD device is sized to barely meet the ESD immunity requirements or sacrifice the ESD robustness to guarantee better RF performance. In this approach, it is possible that ESD protection circuits are selected from an ESD library and attached to I/Os perhaps after finishing the RF core circuit design, expecting the ESD protection devices have a negligible impact on the RF performance. Engineering work here is primarily focussed on the optimization of the ESD kit to minimize the parasitic capacitance.

The counterpart of the ‘low-C’ ESD design is the ‘ESD-RF co-design’ concept. A simple example of the co-design approach is to consider ESD protection devices to be a part of the matching network, and include a certain ESD protection device at the very beginning of RF circuit design. In many cases, the matching alteration problems can easily be solved using this approach [50]. Some drawbacks of the co-design approach are the high effort required for the RF modelling of ESD protection devices from the early phase of process development, as well as additional engineering expenses, since ESD developers need to be involved throughout the RF design phase.

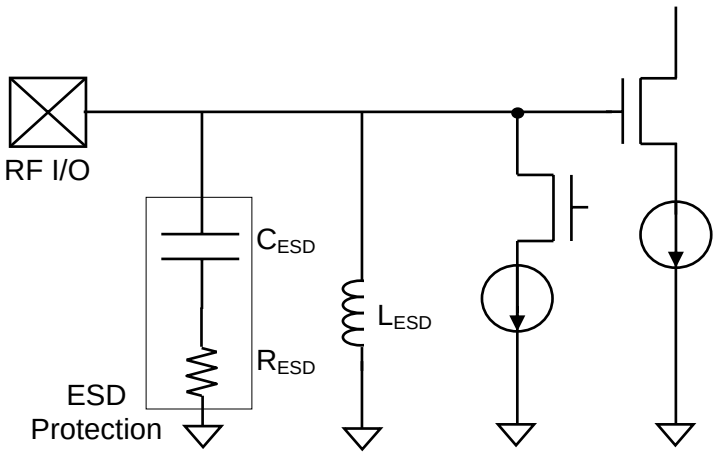


Figure 3.6: ESD cancellation scheme with an additional inductor

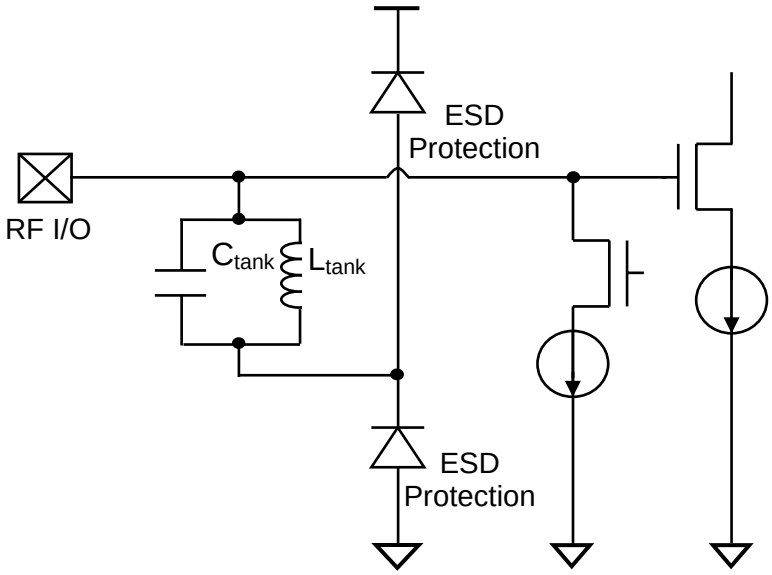


Figure 3.7: S_{21} ESD Isolation scheme with an 'LC-tank'

More aggressive examples of 'ESD-RF co-design,' such as 'ESD cancellation [51]' and 'ESD-isolation [52]' have been demonstrated. The 'ESD cancellation' scheme depicted in Fig. 3.6 has an additional on-chip inductor in parallel with the ESD protection devices. At a given RF signal frequency, this inductor resonates out the capacitance of the ESD protection device (C_{ESD}); therefore the ESD protection device is virtually invisible in the normal operations. The concept of 'ESD-isolation' is similar to 'ESD cancellation'. As illustrated in Fig. 3.7, an LC-tank is inserted between the input path and the ESD protection devices. The LC tank should be tuned so that the

resonant frequency is matched to the RF signal frequency; this 'hides' the ESD protection devices from the input path at the signal frequency.

Unfortunately, the 'ESD cancellation' and 'ESD isolation' schemes have several drawbacks or limitations. For example, in these protection schemes, at least one additional on-chip inductor is needed. This large on-chip inductor may be considered to be an excessive expense in improving ESD immunity. Furthermore, to minimize the signal loss due to the inductor, the quality factor of the inductor must be high, which is the case only in mature RF processes. A fundamental limitation is that the ESD protection schemes with inductors can be applied only to narrow-band RF ICs, since cancellation or isolation occurs at a certain frequency.

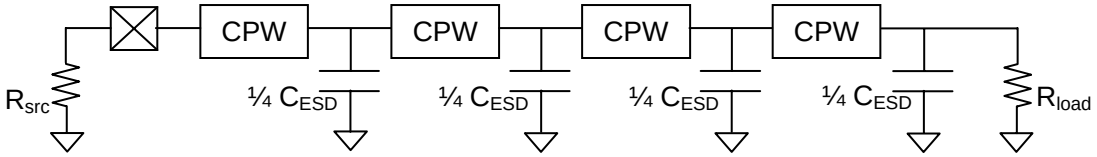


Figure 3.8: Four-segment distributed ESD protection system.

For broad band applications, several approaches have been proposed during the past few years. Kleveland et al. [53] demonstrated a distributed ESD protection system shown in Fig. 3.8. The four segments of ESD protection devices and the CPW (coplanar waveguide) compose an artificial transmission line with a characteristic impedance the same as the source and RF input impedance, thereby avoiding the impedance discontinuity due to a large single capacitance of the conventional ESD protection device. The CPW should be designed to obtain the desired characteristic impedance:

$$Z_O = \sqrt{\frac{L_{CPW}}{C_{ESD} + C_{CPW}}}, \qquad \text{Eq. 3.7)}$$

where L_{CPW} and C_{CPW} are the inductance and the capacitance of the CPW. In [53] MOS gate-grounded protection devices are employed. It is well known that non-uniform conduction of the grounded gate MOSFET can easily limit the ESD protection effectiveness unless the layout is carefully optimized with respect to symmetry; the ballasting resistors are placed in series with the ESD devices [1]. The ESD protection structure in Fig. 3.8 is likely to exacerbate this non-uniform conduction issue and a large current is shunted by only the first few segments close to the I/O pad. However in [53] the measurement results with HBM and CDM stresses demonstrated that

each small segment can effectively be turned on, which is probably due to the wide (~36 μm) metal lines between segments and the N-well ballasting resistors in series with the ESD protection devices. Ito et al. [49] quantified the impact of the distributed ESD protection scheme on the RF signal under normal operations and proposed an optimization methodology. It was also shown that a four-segment distributed ESD protection system can be beneficial for frequencies up to 10 GHz, suffering only a 0.02 dB loss between 0 ~ 10 GHz with a total capacitance of 200 fF. A drawback of this ESD protection scheme is its large size because it needs long CPWs between the segments to achieve a sufficient inductance. The structure demonstrated in [53] has 0.35 mm ~ 1.4 mm length; its application would be limited to the ICs with a few high-speed interfaces.

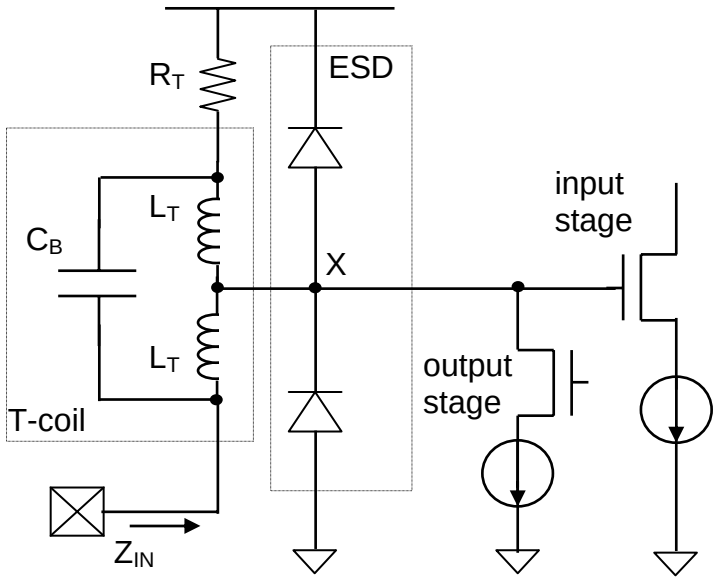


Figure 3.9: T-coil network for ESD protection [54][55]

For high-speed digital applications, the T-coil has been used in part of ESD protection structures [54] as illustrated in Fig. 3.9. With proper choice of L_T, C_B and the coupling coefficient between the two inductors with respect to R_T and the impedance at the node X, Z_{IN} can be equal to R_T over wide frequency range. Galal et al. [55] demonstrated that the T-coil network can provide a return loss of -20 dB at 10 GHz with negligible mid-band loss. However, the ESD test results show that HBM stress tolerance is only 800-1000 V although it has a large (1.2 pF) ESD protection device. In this ESD protection scheme, the on-chip inductors are involved in the ESD current path. The series resistance of the T-coils may cause this low ESD immunity, and the abrupt turning at each corner of the T-coil layout could be vulnerable due to electromigration under ESD conditions.

3.1.3 Noise Figure Degradation due to ESD Protection Devices

Table 3.1 summarizes the Q factors of the ESD protection devices reported in [51] and [56]. At 5 GHz, the measured Q of the gate-grounded NMOS transistor (ggNMOS) is higher than 13.3, and the Q of diodes is over 50.

Device Type		Technology	Frequency	Q	R_p (assuming 0.2 pF C_{ESD})
1	ggNMOS	0.25 µm CMOS	1.8 GHz	40	17.7 kΩ
			5.0 GHz	14.5	2.3 kΩ
2	ggNMOS (junction isolation)	0.35 µm BiCMOS	5.0 GHz	13.3	2.1 kΩ
3	ggNMOS (trench isolation)	0.35 µm BiCMOS	5.0 GHz	17.5	2.8 kΩ
4	Dual-diode	0.24 µm BiCMOS	5.0 GHz	50	8.0 kΩ
5	N+/Pwell diode	0.25 µm CMOS	1.8 GHz	145	64.1 kΩ
			5.0 GHz	52	8.3 kΩ
6	P+/Nwell diode	0.25 µm CMOS	1.8 GHz	184	81.3 kΩ
			5.0 GHz	66	10.5 kΩ

Table 3.1: Measured Q factors of ggNMOS transistors and diodes [51][56]

Assuming 200 fF of C_{ESD}, R_p (in Fig. 3.1 and Eq. 3.2) of the ggNMOS varies from 2.1 kΩ to 2.8 kΩ at 5 GHz. For the diodes, it reaches over 8 kΩ at the same frequency level. To see the impact of R_p on the RF performance, let us first assume that the ESD protection device is located right next to the pad and the impedance toward the internal circuit is matched to the source impedance (Z_O = 50 Ohm). In this case, the power loss can be formulated as in Eq. 3.8. Here, it is also assumed that the matching alteration of the reactance component is negligible or completely absorbed in the matching network.

$$Powerloss\ [dB] \approx 2 \cdot \log\left(\frac{2R_p + Z_O}{2R_p}\right) \qquad \text{Eq. 3.8)}$$

If an ESD protection device with $R_p >> Z_o$ is inserted as in Fig. 3.3, then the impact of R_p on the RF performance is negligible. For example, the second device in Table 1 has the Q of 13.3 at 5 GHz. Assuming 200 fF C_{ESD}, the calculated R_p is 2.1 kΩ. According to the Eq. 3.8, the power loss caused by 2.1 kΩ R_p with 50 Ohm *Zo* is only 0.02 dB. If the size is increased to 1 pF, the R_p is reduced to ~ 420 Ohm and the corresponding power loss is approximately 0.1 dB.

Including the thermal-noise of R_p itself ($\overline{V_{n,Rp}^2} = 4kTR_p$), the noise figure (*NF*) of the two port system depicted in Fig. 3.6 can be described as in Eq. 3.9 [57].

$$NF = 1 + \frac{R_s}{R_p}, \qquad \text{Eq. 3.9)}$$

here, R_s is the real part of the source impedance (Z_{src}) in Fig. 3.3.

With the assumption of Q >> 1, Eq. 3.9 can be rewritten with R_{ESD} and C_{ESD} as follows.

$$NF = 1 + \omega^2 C_{ESD}^2 R_s R_{ESD}, \qquad \text{Eq. 3.10)}$$

According to Eqs. 3.9 and 3.10, *NF* degradation due to the power loss and the self-generated noise can be negligible if C_{ESD} and R_{ESD} are small. For example, Wang et al. [47][58] reported that the ggNMOS transistor with 970 fF C_{ESD} causes *NF* to increase by only 1.52 %[18] (at 2.4 GHz) and a dual directional SCR with 130 fF C_{ESD} results in less than 0.03% NF degradation. Several studies [50][52][59][60] have reported the same order of noise degradation due to ESD protection devices.

However, other studies also reported that the NF of low noise amplifiers (LNA) can increase by ~ 38 % at 5 GHz with 0.41 pF ggNMOS [38], ~ 25 % at 2.4 GHz with 0.6 pF C_{ESD} [61], ~ 20% at 5.2 GHz with an inductor ESD protection (R_p of the inductor is 1.5 kΩ) [62].

The sources of this excessive noise with ESD protection could possibly be the following: 1) In the previous analysis of NF degradation, we assumed that the input matching is perfect over broad frequency ranges. However, due to the impedance mismatch at high frequencies, the input signal and the noise from the source are reflected as explained in Section 3.1.1 in terms of S_{11}. If the noise from the signal source is dominant in determining the total system *NF*, *NF* is not very sensitive to the S_{11} variation, since both the signal and the noise from the source are reflected

[18] This value was recalculated based on [47][58]. In these references, the percentage value of this NF degradation is 3.78 %., which was extracted from dB(*NF*) values.

together. However, the roll-off of the input signal and the noise from the source makes the noise of the subsequent stage more significant. This issue will be discussed again in details in Section 3.3 with the measurement results of a broadband LNA. 2) Besides the thermal noise of resistive components, depending on the device type there could be other noise sources such as shot noise etc. In [47], each noise element of various ESD protection devices was identified, and included in the noise analysis of LNA circuits. 3) Via the large size of ESD protection devices, the substrate noise or the power rail noise can be coupled to I/O pins, leak into the signal path. The coupling effects depend on the layouts, power bus design, and substrate type. This phenomenon cannot be overlooked in the present highly integrated mixed signal ICs. 4) In Fig. 3.3, it is assumed that the ESD protection devices are directly looking at the source impedance (= 50 Ω) and the impedance of the internal circuit is perfectly matched to the source impedance. However in many cases, the impedance at the node where the ESD protection device is located is not matched to 50 Ω. For instance, the inductance associated with the package could be a few nH, and the impedance-matching elements are often inserted between the pad and the ESD protection devices. In some cases like [63], the bond wires between the input pad and the ESD protection devices are actively used to increase the bandwidth. By these additional passive elements, the actual impedance which the ESD devices are looking at could be increased. In other words, Z_o in Eq. 3.8 is increased; therefore, the power loss due to ESD protection device can further be increased. 5) It is well known that the source impedance leading to the maximum power transfer is different from the source impedance for the minimum *NF* [64]. Depending on the design of LNAs, even the small change of input impedance can drive the circuit away from the noise optimal condition; therefore, the input impedance alteration due to ESD protection devices can cause *NF* to substantially increase, depending on the LNA design. This *NF* variation due to the small change of the source impedance can often be higher than the ESD-introducing *NF* degradation.

In this sub-section we have reviewed the limitations of low-C ESD protection in terms of bandwidth, gain loss, and noise figure degradation. For relatively low frequency applications (< 2-3 GHz), the low-C protection strategy can be a sufficient solution to meet the current industrial target for ESD immunity, regarding that 0.2 pF ESD protection devices can provide 2 kV HBM protection capability. However, at frequencies above 10 GHz, even 0.2 pF C_{ESD} can seriously degrade the RF performance; therefore, for extremely high frequency applications only co-design methodologies can provide the ESD protection capability without substantial degradation of RF performance. For the medium frequency range around 5 GHz, the trade-off between the low-C protection and rigorous co-design schemes must be considered. Given ESD immunity requirement, RF specifications, and the quality of RF passive components, RF/ESD designers

should carefully choose an ESD protection strategy. As the frequency goes up, the capacitance headroom for the ESD protection devices becomes smaller, however, the passive component such as on-chip inductors and high-Q MIM (metal-insulator-metal) capacitors can be sized smaller, which is more favorable for the co-design strategy. In many cases, by optimization and rearrangement of I/O circuits the ESD immunity can significantly be improved without intensive engineering efforts just on the ESD protection device itself [50].

To contextualize the concepts we have discussed so far, two LNA circuits with a variety of ESD protection devices will be analyzed in the following two sub-sections, focusing on the RF performance degradation due to the ESD parasitic elements. At first, a narrow-band LNA will be presented in Section 3.2, and the broad-band LNA will be discussed in Section 3.3. Both devices were fabricated using a 130 nm CMOS process and designed for on-wafer probing measurement. The case study with the narrow band LNA will describe how ESD consideration is involved in the early phase of RF circuit design in the spirit of RF/ESD co-design, while the broadband LNA study is more focused on the limitations of the low-C ESD design and demonstrating the transient-triggering SCR (TT-SCR) as a robust and compact ESD protection device.

3.2 Case Study 1: ESD Protection for Narrow-band RF ICs

3.2.1 Description of the NB-LNA

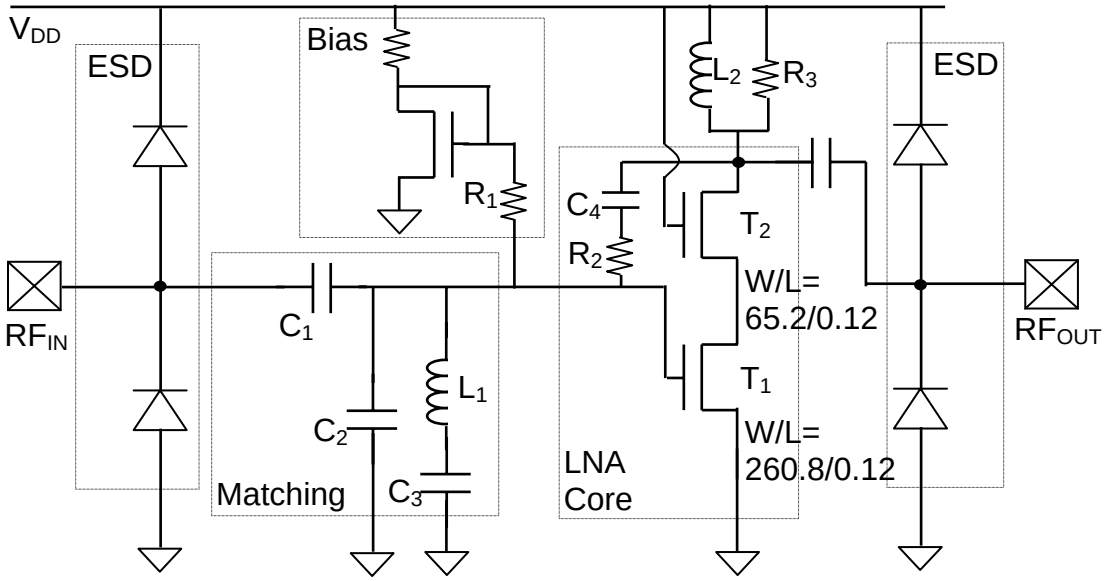


Figure 3.10: An ESD protected narrowband LNA

Fig. 3.10 shows the schematic of the narrow band LNA (NB-LNA) for Universal Mobile Telecommunication System (UMTS) application. This test circuit is fabricated in a 0.13 μm CMOS technology for ESD characterization with on-wafer measurement; therefore, high-Q bonding wires are not available.

The LNA core in Fig. 3.10 uses a cascode structure; the common source stage of the transistor T_1 provides high transconductance (Gm) with low Miller effect, while the common gate stage of T_2 boosts up the output impedance of T_1. Inductive degeneration at the source of the T1 is not employed, because the low Q of the spiral inductor compromises the noise figure (NF) and the required linearity specification (P_{1dB} > -25 dBm, IIP_3 > -15 dBm [57]) can be achieved without source degeneration.

The designed NB-LNA has an additional feedback path with a series connection of a capacitor (C_4) and a resistor (R_2). The feedback path improves the linearity and the input matching at the expense of gain, noise figure, and stability in a very moderate way. The resistor

and capacitor values of this feedback path should carefully be tuned so that the LNA is unconditionally stable [64].

The bias circuit is implemented with a diode-connected NMOSFET and two resistors. R_1, the resistor between the bias and core circuits, is as large as 40 kΩ to minimize the noise injection from the bias to the RF core.

The input matching network is based on the typical 'L-match [65]' with a small on-chip inductor (L_1) between the T_1 gate and V_{SS}, and a high-Q Metal-Insulator-Metal (MIM) capacitor (C_1) between the input and T_1. C_3 is employed to block the DC path from the gate of T_1 to V_{SS} through the inductor. The process used in this study was still under development and the on-chip inductors were not freely selectable, providing only discrete inductance values of L_1. Hence, in addition to the inductor (L_1), the two additional capacitors (C_2 and C_3) are used to finely tune the input matching. Because the cascode structure offers a very large intrinsic gain (above 20 dB), a high-Q Radio Frequency Choke (RFC) is unnecessary for the output load, and we can even reduce the effective Q by adding a parallel resistor (R_3) in parallel to the loading inductor (L_2), extending the output matching bandwidth.

The ESD protection devices for the NB-LNA are the typical dual-diodes (the N^+/P-well and P^+/N-well diodes). The bias dependent S-parameter models for these diodes are provided so that the ESD protection diodes could be included in the simulations of the core RF circuits. To see the impact of ESD protection devices on the RF performance, the size of the protection devices are varied: the larger one (full-C ESD) has 72 pairs of N^+/P-well and P^+/N-well diodes, and the smaller one (half-C ESD) has 36 pairs. The total capacitance of the full-C ESD is approximately 560 fF including the pad capacitance, and that of the half-C structure is 300 fF. For each ESD structure the capacitors in the matching network (C_1 and C_2) are properly sized so that the input reflection is below -10 dB.

3.2.2 RF Performance of NB-LNAs

The measured input and output reflection coefficients (S_{11} and S_{22}) as functions of frequency are plotted in Figs. 3.11 and 3.12. Both input and output matching around 2 GHz are successfully achieved with the two different sizes of the ESD protection devices, since the matching network is calibrated for each ESD protection device. Although the center frequency is shifted to the left by ~100 MHz, the input and output reflection coefficients at 2 GHz are still below -15 dB. Inaccurate RF modelling of the on-chip inductor is the most likely reason for this center frequency shift. The input bandwidth in terms of the input matching is approximately 300 MHz.

The S_{22} data in Fig. 3.12 show much broader output bandwidth, because the on-chip inductor, L_2 has low Q (approximately 5), and the additional resistor (R_3) further lowers the overall Q at the output node.

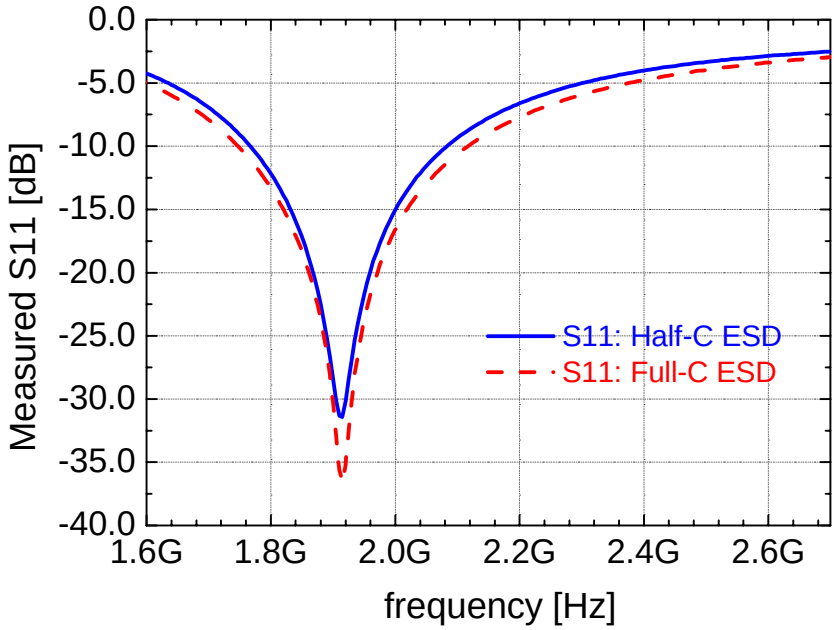


Figure 3.11: Measured input reflection coefficients, S_{11} vs. frequency

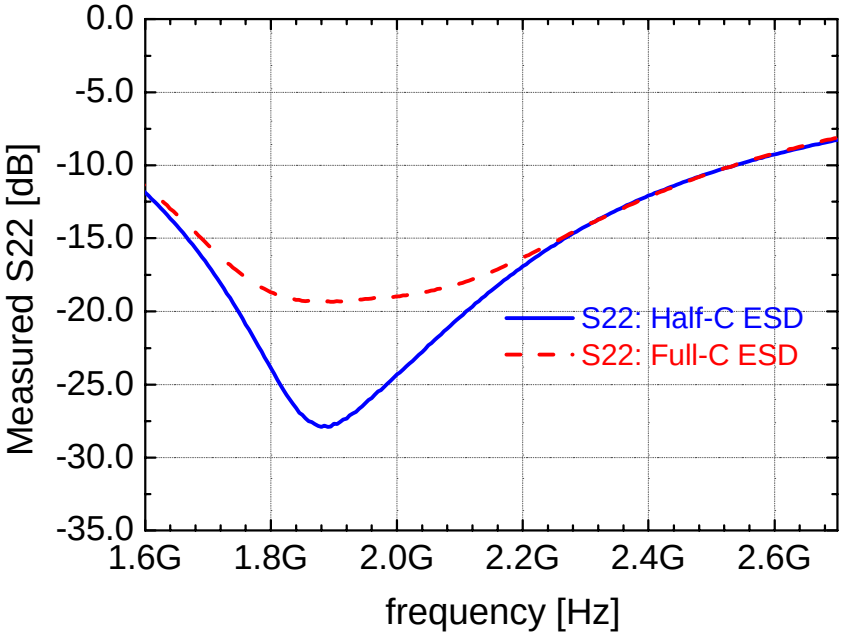


Figure 3.12: Measured output reflection coefficients, S_{22} vs. frequency

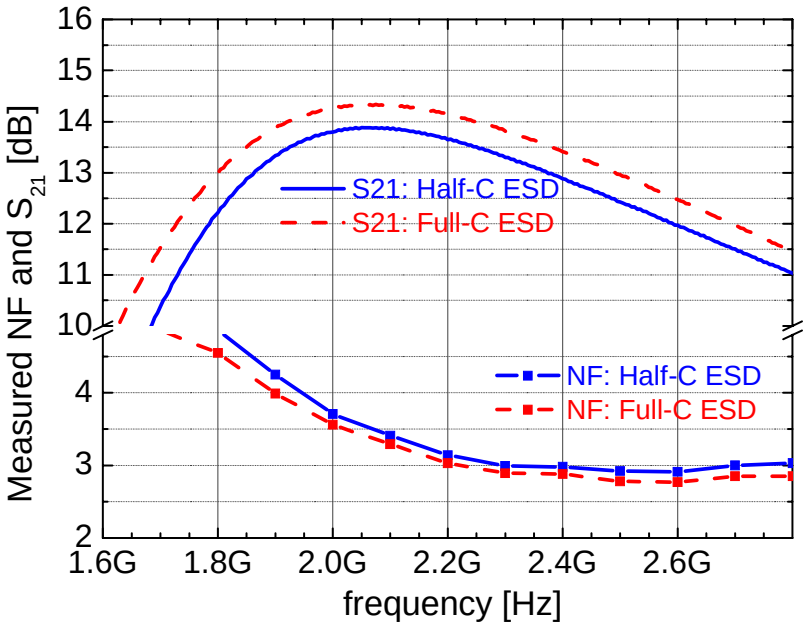


Figure 3.13: Measured gain (S_{21}) and noise figure (NF) of the NB-LNAs with full-C and half-C ESD protections.

The measured gain (S_{21}) and the noise figure (NF) of the NB-LNAs with both protection schemes are shown in Fig. 3.13. The targeted specification for the gain is over 13 dB, and for NF it is below 4 dB. As shown in Fig. 3.13, both design targets are comfortably met with the two different ESD structures. The characteristics of both devices exhibit a minor discrepancy which has an order of the PVT (process-voltage-temperature) variations. As the ESD structure size is doubled from half-C (300 fF) to full-C (560 fF), the gain (S_{21}) is improved by only 0.3 dB. On the other hand, the noise figure is degraded by ~ 0.2 dB. The degradation is approximately 4.7 %, comparable to the predicted number in Section 3.1 and other published data [66][67].

It turns out that in this specific application the RF performance degradation due to ESD protection devices can be tolerated within the targeted specification even if the LNA includes a fairly large size ESD protection (560 fF).

3.2.3 ESD Test Results

When ESD stress is injected into the I/O pins, the ESD protection devices should keep the voltage across the transistors below the critical level. In the employed 0.13 μm process, the gate oxide breakdown voltage measured with a transient stress is approximately 5-5.5 V, when the duration of the applied stress is same as that of the HBM ESD stress

A TLP (transmission line pulsing) ESD test was performed for the ESD characterization of the NB-LNA. In the test setup, the rise time of the pulse is ~7 ns, compatible with the HBM test specification. Both LNAs with half-C and full-C ESD structures pass a 1.4 A TLP test, which is the limitation of the test equipment. As discussed in Chapter 2, the 1.3 A TLP result is strongly correlated with 2 kV HBM hardness.

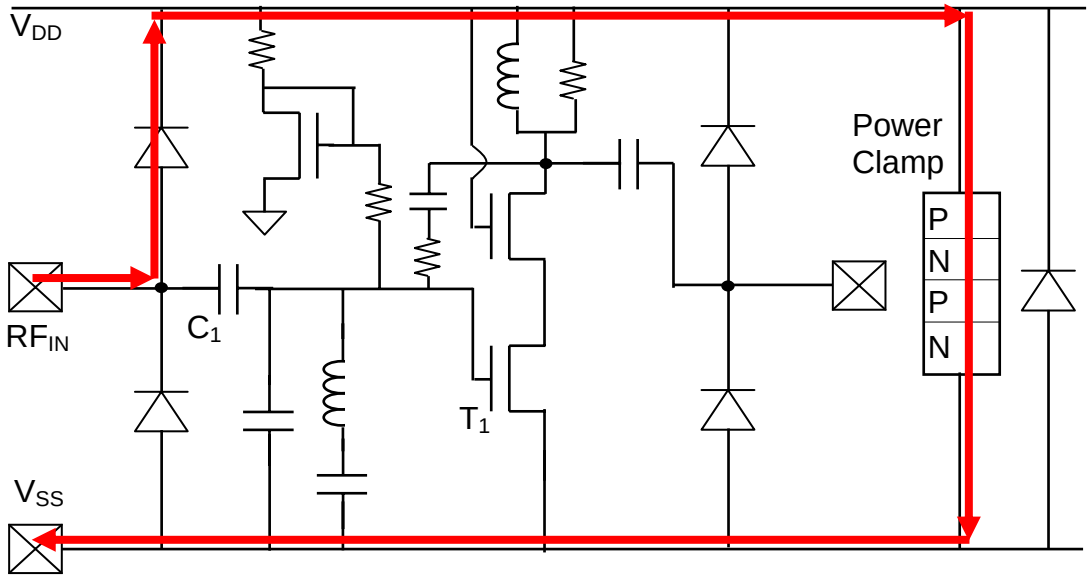


Figure 3.14: The ESD current path during a TLP test, RF_{IN} (+) V_{SS} (-)

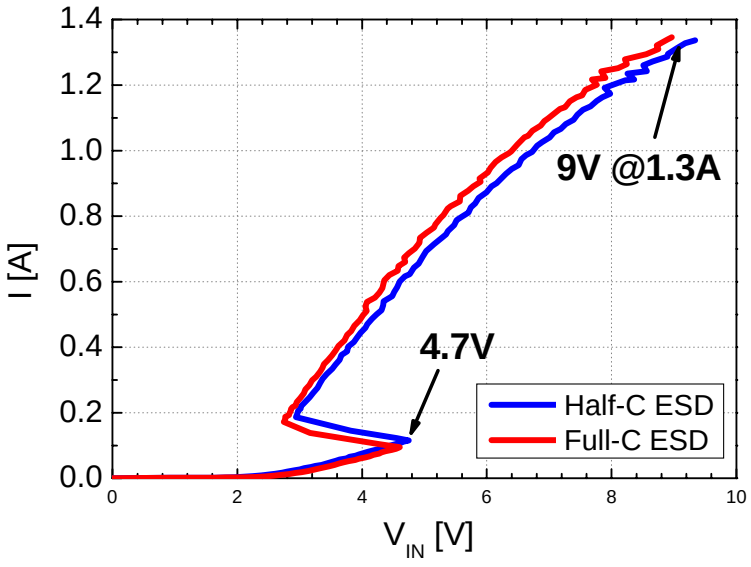


Figure 3.15: TLP test results of the NB-LNAs with the half-C and full-C ESD. The test condition is RF_{IN} (+) VSS (-).

Fig. 3.14 illustrates the most critical ESD test condition and the ESD current path during this ESD test event. The ESD charges are injected into the RF input pad (RF_{IN}) with reference to V_{SS}. The current first flows through the P+/N-well diode to V_{DD}, passes through the long V_{DD} power rail. Between the V_{DD} power rail and the V_{SS} ground rail there is a power clamp. In this study, a diode-triggered SCR is implemented as a power clamp. The ESD current is shunted by this power clamp and finally reaches the V_{SS} pad through the ground rail.

Fig. 3.15 shows the *I-V* curves of the ESD test illustrated in Fig. 3.14. The *I-V* curves exhibit snap-back behaviour, since the ESD current path involves the power clamp SCR. The half-C structure shows slightly higher voltage drop, however the discrepancies between the two *I-V* curves are practically negligible. The diode-triggered SCR itself has a triggering voltage of 3.8 V and a holding voltage of 1.2V; therefore, with the additional voltage drop of the P+/W-well diode and the power rail resistance, the triggering voltage of the total ESD current path reaches approximately 4.7 V as indicated in Fig. 3.15, and the holding voltage is close to 3 V. It turns out that the ESD current path is highly resistive (the differential resistance is approximately 4 Ω). According to [50], ~12 % of the total resistance comes from the P^+/N-well diode, 25 % from the SCR, and more than 60 % is due to the metal and via resistance of the power rails; this ESD design may be considered as a bad example of a rail-based ESD protection design.

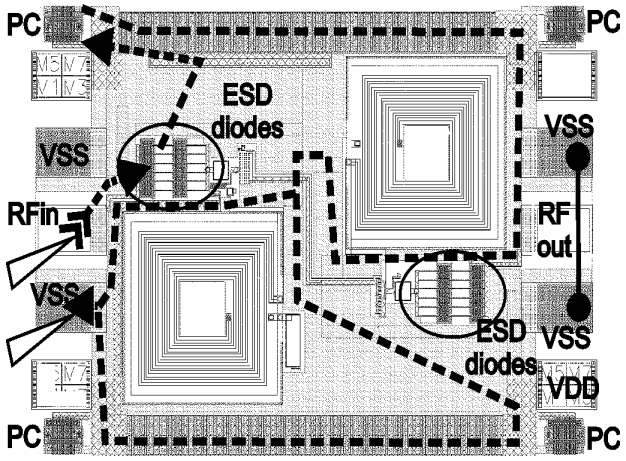


Figure 3.16: The Layout of the NB-LNA. The dashed line indicates the current path during an RF_{IN} (+) w.r.t. V_{SS} (-) ESD stress event. Reprinted from Proc. EOS/ESD Symposium 2005 [50]

The layout of the NB-LNA in Fig. 3.16 reveals the long ESD path responsible for the high resistance. Because of the high routing resistance, the voltage at the pad increase up to 9 V at 1.3 A, which is much higher than the gate oxide breakdown voltage (5.0-5.5 V). However, the LNAs interestingly survived the 1.4 A TLP test. The reason for this good ESD protection performance

can be found in the schematic in Fig. 3.14. As a part of input matching, an AC coupling capacitance is located adjacent to the input pad. Although the pad voltage increase up to 9 V during the transient ESD events, at the gate of T_1 transistor the voltage is divided by the C_1 and the effective impedance at the gate; therefore, the voltage between the gate of T_1 and V_{SS} can be held below 5 V. Basically, the coupling capacitor C_1 alleviates the ESD stress, which may be counted to be a benefit of RF/ESD co-design. Soldner et al. [50] also demonstrated that by reducing power rail resistance, the voltage at the RF_{IN} pad can be less than 4 V at 1.3 A.

The case study in this sub-section demonstrates that for this specific application (narrow-band LNAs for 2 GHz UMTS) a simple co-design approach is possible without aggressively reducing the size of the ESD protection devices. The dual diodes with 300 fF and 560 fF successfully sustained 1.4 A TLP stress (corresponding to > 2kV HBM stress), showing negligible degradation of the RF performance.

3.3 Case Study 2: ESD Protection for Broad-band RF ICs

3.3.1 Description of the BB-LNA

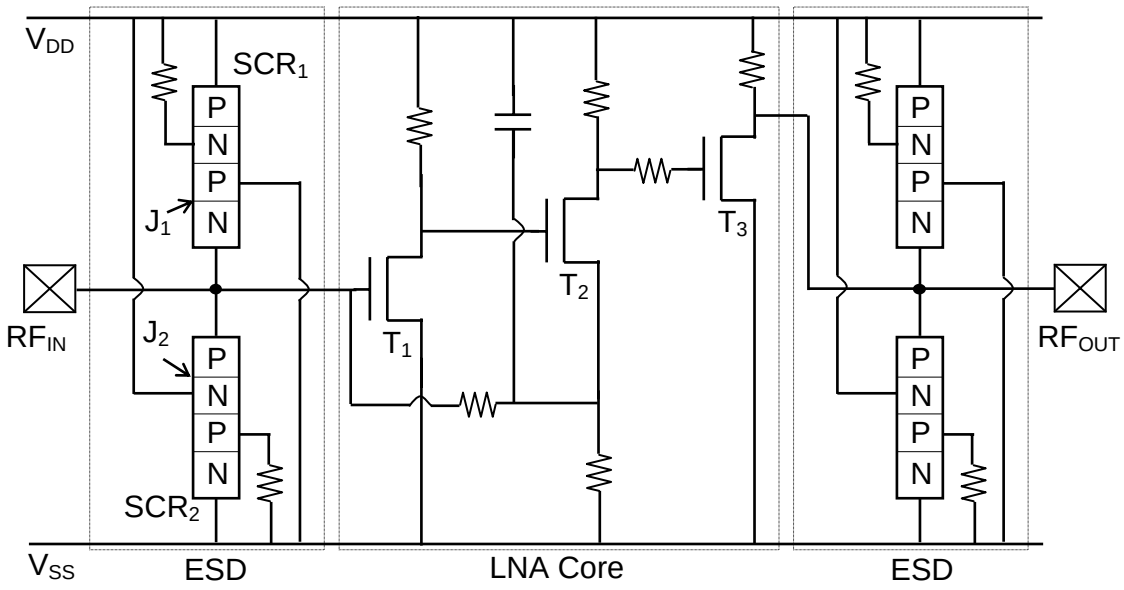


Figure 3.17: An ESD protected broad-band LNA

The schematic of a broad-band LNA (BB-LNA) is illustrated in Fig. 3.17. The RF core circuit was adopted from [68]. The RF core consists of three cascaded self-biased amplifier stages with a current-current feedback path from the second stage to the input, to extend the bandwidth and improve the linearity. The active power clamps and shunting capacitors between V_{DD} and V_{SS} are not drawn in Fig. 3.17. While the NB-LNA in Fig. 3.10 has an AC coupling capacitor that alleviates the ESD stress on the transistor at the front end, in the BB-LNA the gate of the MOSFET with a thin gate oxide is directly connected to the input pad.

In Fig. 3.17, transient-triggered silicon controlled rectifiers (TT-SCRs) [69][70] are used as ESD protection devices at both the input and output pads. Besides the TT-SCRs, typical dual diodes, similar to the protection devices in Section 3.2, and bipolar transistors are also implemented to investigate the limitations of a low-C protection strategy. In this section, the TT-SCR and the dual diodes will be discussed in depth; the details of the bipolar ESD protection devices can be found in [50]. During normal operation conditions, the SCRs are completely

turned off with the junctions J_1 and J_2 reverse-biased. The base of PNP (NPN) transistor within SCR_1 (SCR_2) is connected to V_{DD} (V_{SS}) with a finite resistance to prevent the SCRs from being accidentally triggered [69]. The triggering mechanism of the TT-SCRs will be described later in Section 3.3.3. The TT-SCR and diode structures occupy 650 and 350 μm^2, respectively. However, the FEOL (Front-End-Of-Line) capacitances associated with the TT-SCR (95 fF) are less than half of the dual diodes (180 fF). In the actual implementation of the LNAs, the additional BEOL (Back-End-Of-Line) capacitances of the TT-SCR and the dual diodes are estimated as 180 fF and 110 fF, respectively; therefore, the total capacitance of the TT-SCR (275 fF) is slightly smaller than that of the dual diodes (290 fF).

3.3.2 RF Performance of NB-LNAs

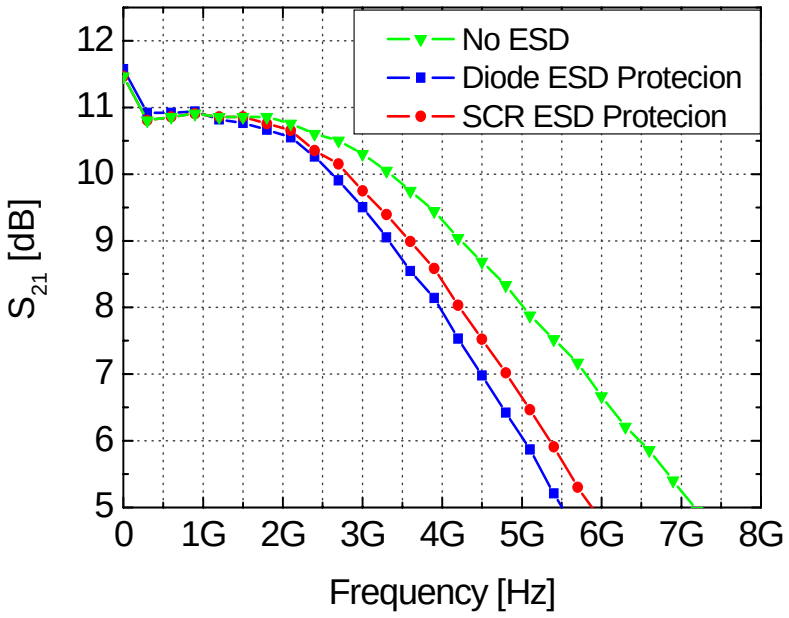


Figure 3.18: Measured gain (S_{21}) of the BB-LNAs.

Fig 3.18 shows the measured gain (S_{21}) of the BB-LNAs with a variety of ESD protection devices. The BB-LNAs consume 22.5 mW with a 1.5 V supply. The BB-LNA without ESD protection devices has a low-frequency gain of 10.9 dB and a bandwidth of 5.1 GHz. Attaching the ESD protection devices at the I/O pads, the 3-dB bandwidth of the BB-LNA is considerably reduced. The 3-dB bandwidth of the BB-LNA with the dual diode protection is reduced to ~ 4 GHz. The bandwidth with the TT-SCR is slightly higher, approximately 4.3 GHz. At 5 GHz, the gain reductions due to the TT-SCR and the dual diodes are 1.4 dB and 2.1 dB, respectively.

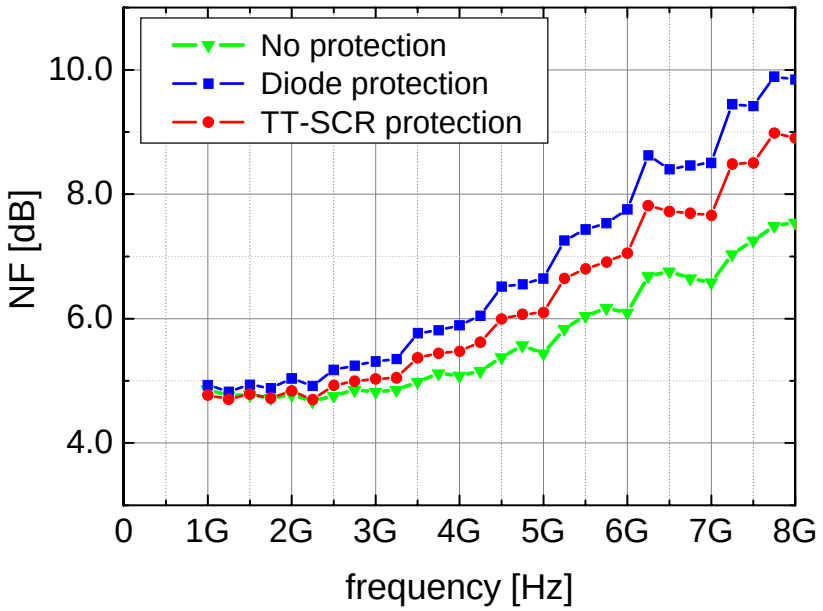


Figure 3.19: The noise figure (NF) of the BB-LNAs.

	No-ESD	Dual diodes	TT-SCRs
Gain(@ 5 GHz)	8.0 dB	5.9 dB (-2.1 dB)	6.6 dB (-1.4 dB)
NF (@ 5 GHz)	5.4 dB	6.6 dB (+1.2 dB)	6.0 dB (+0.6 dB)

Table 3.2: RF performance of the BB-LNAs with various ESD protection concepts.

The measured NF of the BB-LNAs is plotted in Fig. 3.19, and the RF performance degradation data due to the two different ESD protection devices are summarized in Table 3.2. The NF of the BB-LNA with the diode protection is 1.2 dB higher than that of the LNA without ESD protection, while the NF of the LNA with the TT-SCR is increased by 0.6 dB.

The NF degradation summarized in Table 3.2 is much higher than the predicted values in Section 3.1.3, which are based on the assumption that the NF degradation is caused by the lossy ESD elements. Due to the impedance mismatch at high frequencies, the input signal and the noise from the source are reflected as explained in Section 3.1.1 in terms of S_{11}. If the noise from the signal source is dominant in determining the total system NF, NF is not very sensitive to the S_{11} variation since both the signal and the noise from the source are reflected together. However, Fig.

3.19 shows that the NF of the BB-LNA is relatively high even at low frequencies[19], which means the noise generated in the amplifier has a significant impact in determining the overall NF. The roll-off of the input signal and the noise from the source makes the noise of the subsequent stage more significant. In other words, up to the frequency level where the noise from the source is still dominant compared to the input referred noise of the amplifier, the SNR (signal-to-noise-ratio) degradation could be negligible, resulting in minor increase in NF. However, at higher frequencies the noise generated in the amplifier becomes dominant, degrading the SNR; therefore, NF becomes more sensitive to the variation of the ESD protection device's size and the measurement frequency as in Fig. 3.19.

3.3.3 ESD Test Results

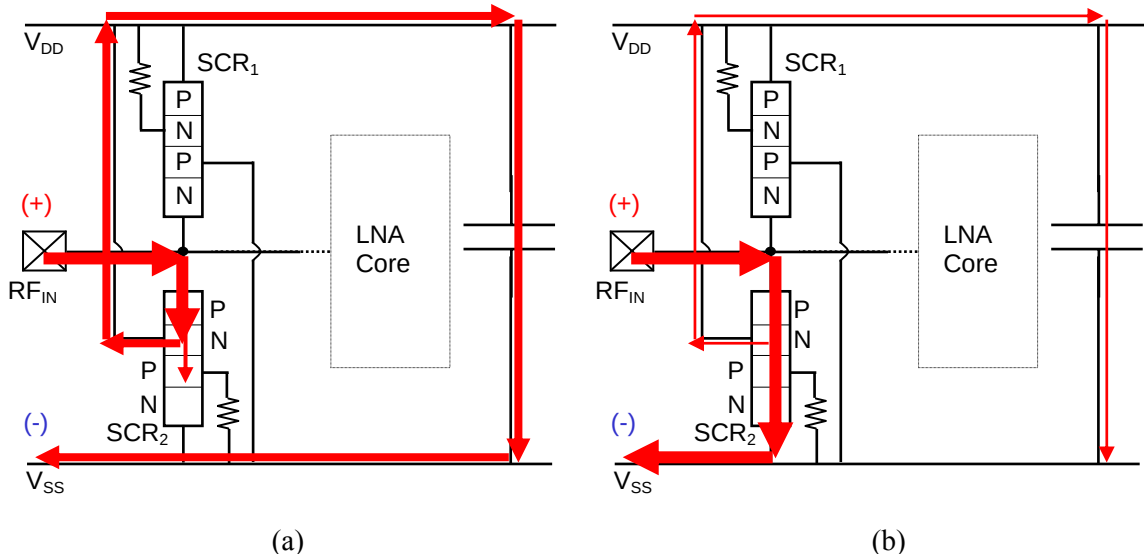


Figure 3.20: The operation of the TT-SCRs under ESD conditions, RF_{IN} (+) VSS (-). (a) Before the TT-SCR is triggered, and (b) after the TT-SCR is triggered.

In Fig. 3.20, the triggering mechanism of the TT-SCRs is illustrated. The large shunting capacitor[20] between the power rails is also depicted in Fig. 3.20. Fig. 3.20a shows the current path while the TT-SCR is being triggered. When the ESD current is injected to the RF_{IN} with respect

[19] In general, if the gain of each amplifier stage is high, the impact of the noise due to the elements at the amplifier output can be disregarded. However, the total gain of three amplifier stages is only 6~8 dB at 5 GHz; the gain of the 1st stage is not high enough to effectively screen the noise from its output.

[20] In most ICs, there are fairly large capacitors between power rails to filter the supply voltage. However, in this test chip, the large (~50 pF) capacitor is additionally used to provide a current path while the TT-SCR is being triggered.

to V_{SS}, at first this current flows via the first PN junction (the emitter/base junction of the PNP) of the SCR_2 and then the large shunting capacitor. In other words, the base potential of the PNP is held low due to the large shunting capacitor so that a certain number of holes are diffused from the emitter to the collector of the PNP. This hole current elevates the base potential of NPN, and thus finally triggers the thyristor as illustrated in Fig. 3.20b; resulting in the typical snapback behaviour.

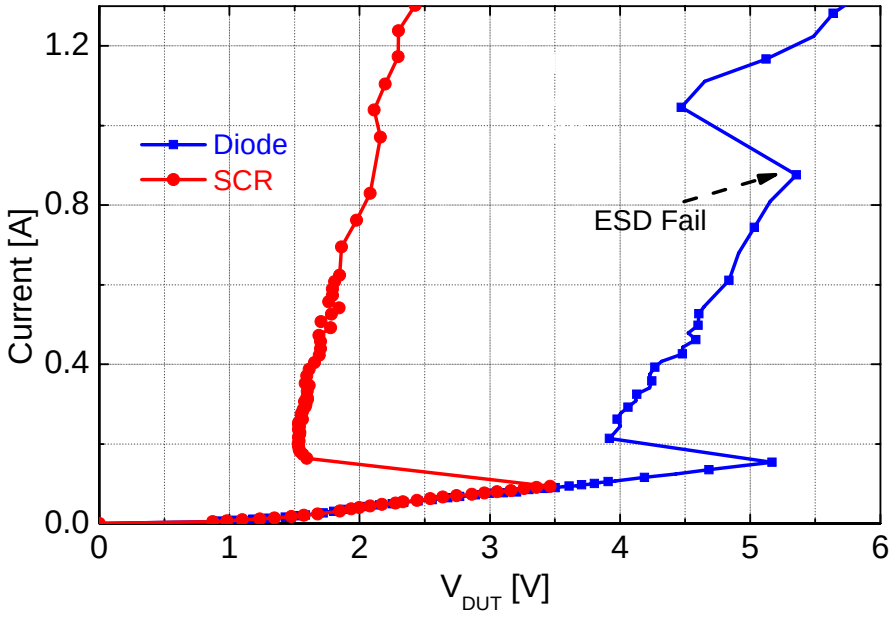


Figure 3.21: TLP test results of the BB-LNAs with the TT-SCRs and the diodes. The test condition is RF_{IN} (+) VSS (-).

In Fig. 3.21, the TLP test results of the TT-SCR and the dual diodes are compared. The ESD stress condition is same as Figs. 3.14 and 3.20. As explained above, the TT-SCR shows the snapback behaviour. The dual diode protection also shows the snapback behaviour, but it is due to the SCR-type power clamp shown in Fig. 3.14. The *I-V*-curve of the diode protection is similar to that of Fig. 3.15 in terms of the triggering voltage and the on-resistance. However, the second snapback is observed in Fig. 3.21. This second snapback indicates that the ESD failure occurs before the current reaches 0.9 A. In the NB-LNA tests, ESD stress of 1.4 A can be tolerated due to the AC coupling capacitor at the input, but in the BB-LNA the gate of the input MOSFET is directly exposed to the ESD stresses; failing at 0.85 A.

The TT-SCR shows relatively lower triggering voltage (3.5 V) and very low sustaining voltage (< 2.5 V @ 1.3 A). In this certain process, 5 V clamping voltage can be allowed for 2 kV HBM stresses; therefore, the SCR device can still be downsized. Given the on-resistance

extracted from Fig. 3.21, in [50] it is discussed that the minimum size of the SCR is approximately 130 fF to meet the 2 kV HBM requirement.

3.4 Summary

The impact of ESD parasitic elements on the performance of RF ICs has been analyzed with a simplified RF model of ESD protection devices. The degree of signal reflection and power loss is extremely sensitive to the size of ESD protection devices at multi-GHz frequencies; at 10 GHz, the input reflection S-parameter (S_{11}) becomes approximately -10 dB with 0.2 pF ESD devices. During the past few years, there have been many studies seeking to overcome this ESD-to-circuit impact, including optimization of the conventional ESD protection devices, co-design methods such as ESD cancellation/isolation and several broadband techniques etc. At relatively low frequencies, the low-C design methods can be well adopted since it is very simple and transparent to various RF designs. For extremely high frequency applications only co-design methodologies can provide the ESD protection capability without substantial degradation of RF performance. For the medium frequency range around 5 GHz, the trade-off between the low-C protection and rigorous co-design schemes must be considered.

To verify the co-design method, a 2 GHz narrowband LNA for UMTS application is implemented. For this specific application a simple co-design approach is possible without aggressively reducing the size of the ESD protection devices. The dual diodes with 300 fF and 560 fF successfully sustain ESD stress levels greater than 2 kV HBM, showing negligible degradation of the RF performance. A 5 GHz LNA is also developed to investigate the limitation of the low-C protection scheme. In this test bench, a transient triggered (TT) SCR circuit is employed for ESD protection. The TT-SCR shows relatively lower triggering voltage (3.5 V) and very low sustaining voltage (< 2.5 V @ 1.3 A) with approximately 280 fF capacitance. Although the TT-SCR considerably reduces the bandwidth of the LNA, it has been discussed that the TT-SCR device can be downsized to 130 fF while maintaining 2 kV HBM ESD immunity; it can be a good candidate for up to 5 GHz applications. Based on these case studies, advanced broad band techniques [49][53][54][55][63] are expected to take an essential role for the applications of > 5 GHz operation frequencies.

Chapter 4

SIGNAL DISTORTION DUE TO ESD PROTECTION DEVICES IN ANALOG-TO-DIGITAL CONVERTERS

Electrostatic discharge (ESD) protection circuits typically contain a significant amount of nonlinear capacitance. At high frequencies and large amplitudes, this nonlinearity can degrade the signal integrity at the input pins of high performance mixed signal ICs, such as analog-to-digital converters (ADCs). This Chapter provides a theoretical analysis of this problem as well as experimental results that quantify typical distortion levels introduced by state-of-the-art ESD structures. It is shown that with distortion targets nearing –100 dB at signal frequencies on the order of 100 MHz, ESD circuits will become a limiting factor in the future. In addition to these results, this study offers some brief guidelines for designing ESD protection circuits suitable for high-speed, high-linearity applications.

4.1 Introduction

The design of electrostatic discharge (ESD) protection circuitry has evolved as a new and significant challenge in mixed-signal circuits. Providing a high degree of ESD robustness, while minimizing ESD circuit-to-signal interaction, has become increasingly difficult due to the trend of increasing operating frequencies and precision. The associated challenges are most notable in

two distinct application areas. First, in Gigahertz RF circuits and Gb/sec digital communication systems, the mere presence of the parasitic capacitance introduced by ESD structures can severely limit the attainable bandwidth [38][47][49]. Second, in applications such as stand-alone ADCs, the nonlinearities associated with the parasitic capacitance threaten to degrade linearity performance at only moderate input frequencies, primarily due to the typically large signal amplitudes at the device pins. While the latter problem is well known among circuit designers, few papers discuss the issue. One noted exception is [71], which discusses the use of a Zener diode structure to mitigate the distortion issue. In this paper, we present results that correlate to more commonly used ESD structures in state-of-the-art ADCs.

In recent years, there has been a demand for high-speed ADCs that maintain good linearity performance up to very high input frequencies. For instance, wireless base stations that digitize the intermediate frequency band typically require a spurious free dynamic range (SFDR) greater than 80 dB, up to frequencies of several tens of MHz [72]. In future applications that may require even better performance, the distortion introduced by ESD structures can become a performance-limiting factor. Thus, achieving an *SFDR* of 90 dB at frequencies greater than 100 MHz generally requires judicious design and optimization of the ESD input structures.

In order to derive guidelines for the development and selection of ESD circuits, the following sections outline various approaches in estimating their harmonic distortion. Specifically, we present results from an analysis using the Volterra series method, Spice simulations, and measurement results using commercially available ADCs.

4.2 Volterra Series Analysis and Spice Simulations

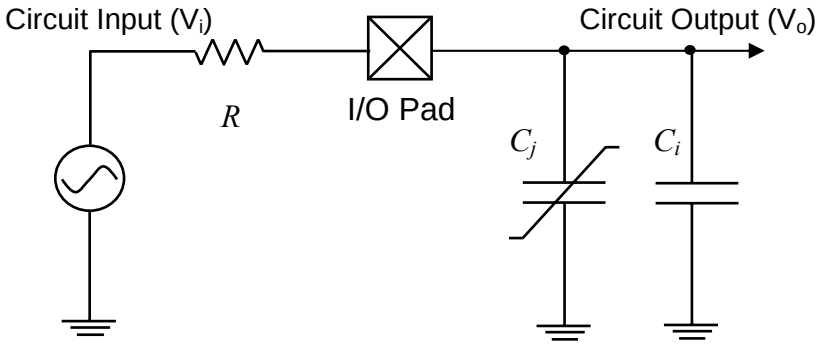


Figure 4.1: Setup for calculating the signal distortion due to ESD protection. The ESD device is modeled by a nonlinear junction capacitor (*Cj*).

The generic circuit model shown in Fig. 4.1 is suitable for an approximate distortion analysis of most conventional ESD structures. In this diagram, R represents the equivalent driving resistance of the input source. The capacitance C_i models the input capacitance of the ADC (e.g. sampling capacitor), and C_j represents the nonlinear capacitance of the ESD protection circuit. Assuming that C_j is a reverse biased junction capacitance, a basic equation for its voltage dependence is given by

$$C_j = \frac{C_{jo}}{\left(1+\frac{V_o}{\phi_0}\right)^M} \cong C_{jQ}\left[1 - M\frac{v_o}{V_Q+\phi_0} + \frac{1}{2}\left(M^2+M\right)\cdot\left(\frac{v_o}{V_Q+\phi_0}\right)^2\right]. \qquad \text{Eq. 4.1)}$$

The right-hand side of this expression follows from a Taylor expansion about the quiescent voltage V_Q, with v_o being the AC perturbance around this operating point. C_{jQ} corresponds to the junction capacitance at the quiescent point, and M is the so-called junction grading coefficient. Using this model, a Volterra series analysis [73] can be used to show that the second and third order harmonic distortions (HD_2 and HD_3) at the output are given by[21]

$$HD_2(f_{in}) = \frac{1}{2}M\frac{f_{in}}{f_1}\cdot\left|H(f_{in})\cdot H(2f_{in})\right|\cdot\left(\frac{\hat{v}_i}{V_Q+\phi_0}\right) \qquad \text{Eq. 4.2)}$$

$$HD_3(f_{in}) \cong \frac{1}{8}(M^2+M)\frac{f_{in}}{f_1}\cdot\left|H(f_{in})^2\cdot H(3f_{in})\right|\cdot\left(\frac{\hat{v}_i}{V_Q+\phi_0}\right)^2, \qquad \text{Eq. 4.3)}$$

where

$$H(f) = \frac{1}{1+j\frac{f}{f_0}}, \qquad \text{Eq. 4.4)}$$

and

$$f_0 = \frac{1}{2\pi R\left(C_i+C_{jQ}\right)} \qquad f_1 = \frac{1}{2\pi R C_{jQ}}. \qquad \text{Eq. 4.5)}$$

[21] The result given in Eq. 4.3 is based on the approximation that the so-called second-order interaction term is negligible. Here, this condition is satisfied as long as $f_{in} << f_1$, which is usually true in practice.

In the above expressions, $\hat{v}_i$ is the peak input amplitude, and *H(f)* above is simply the first order linear transfer function of the circuit. To avoid signal attenuation, the circuit usually must be designed, such that the maximum desired f_{in} is well below the corner frequency f_0. Consequently, the terms inside the magnitude operators in Eqs. 4.2 and 4.3 are typically close to unity over the frequency range of interest. This, in turn, suggests that low HD_2 and HD_3 values are only possible if $f_1 >> f_{in}$. Unfortunately, large f_1 dictates a small junction capacitance C_{jQ}, which corresponds to a small ESD device. Therefore, there is a fundamental tradeoff between ESD robustness and high frequency linearity.

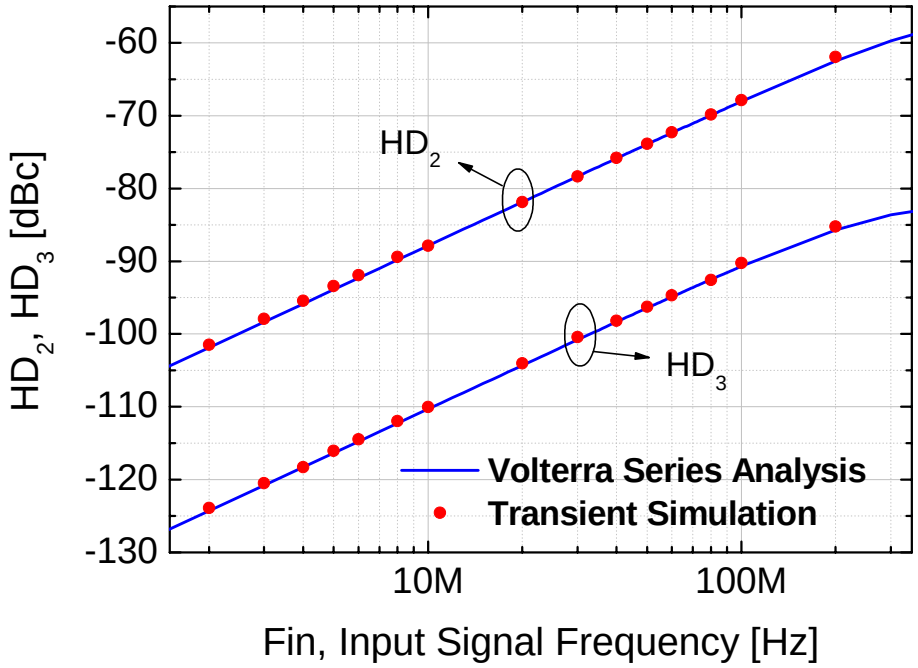


Figure 4.2: 2nd and 3rd order harmonic distortion extracted from Volterra series analysis (solid lines) and Spice transient simulation (dots).

For example, Fig. 4.2 plots the expressions given in Eq. 4.2 and Eq. 4.3 using $M = 0.3$, $C_{j0} = 1$ pF, $C_1 = 0$ pF, $\phi_0 = 0.7$ V, $V_Q = 1.5$ V, $\hat{v}_i = 0.5$ V, and R = 25 Ω (Thévenin equivalent of 50 Ω source resistance and 50 Ω termination). Also shown in Fig. 4.2 are data points from a Spice transient simulation, which confirm the validity of the derived expressions. As is evident from this result, only a moderate amount of junction capacitance can limit the attainable HD_3 performance at 100 MHz to values around -90 dB. Note that even though the HD_2 performance is significantly worse, second order nonlinearities are usually attenuated using fully differential configurations, and hence are relatively unimportant.

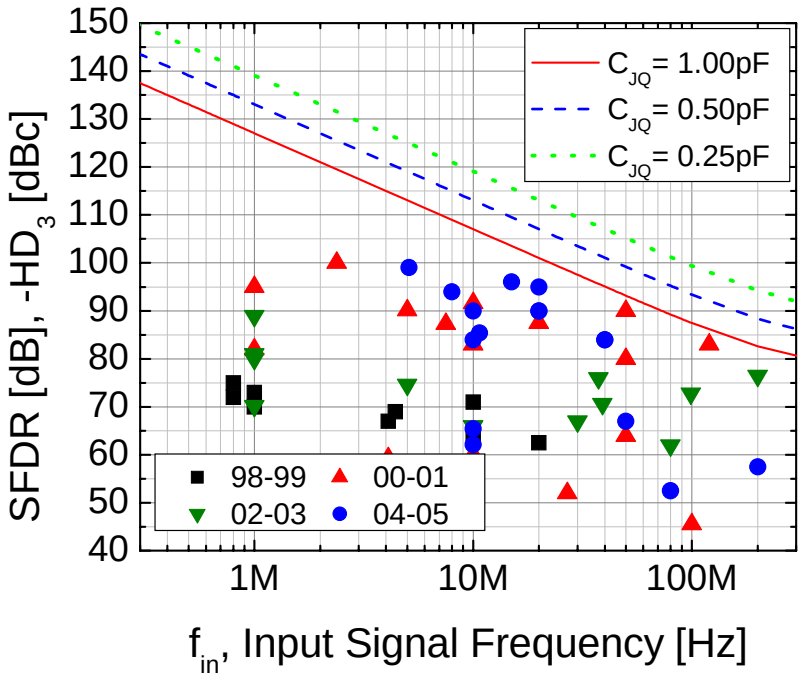


Figure 4.3: State-of-the-art SFDR performance of ADCs (data points from ISSCC 1998-2005) and theoretically estimated SFDR limits due to ESD protection (solid lines). The theoretical prediction assumes a simple reverse-biased junction model and single ended signals of 0.5-V amplitude.

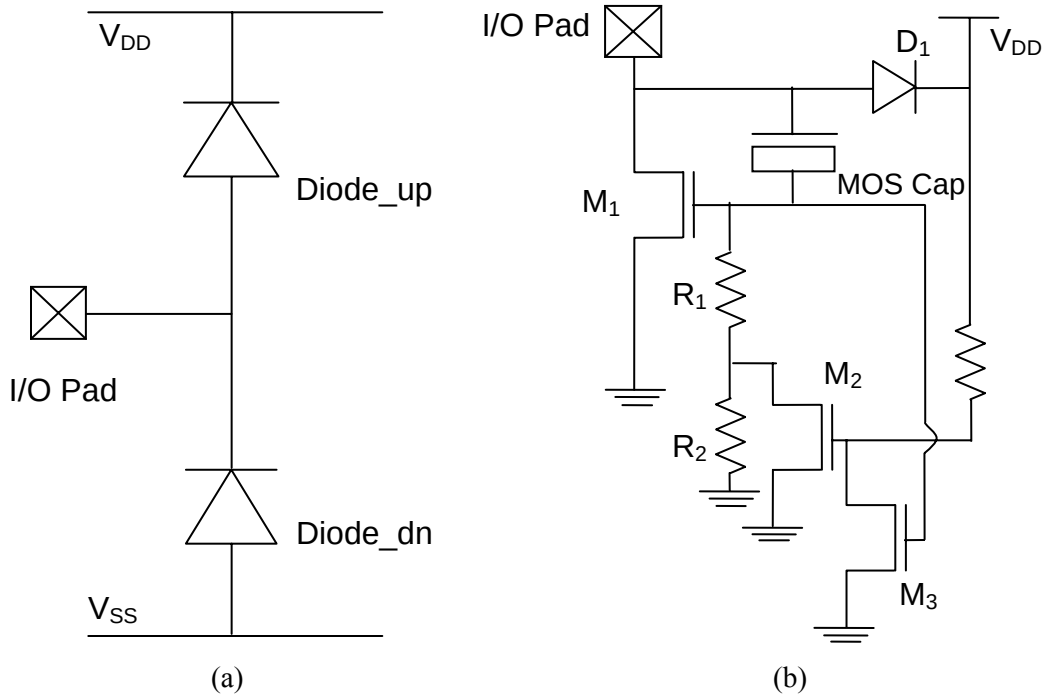


Figure 4.4: The measured ESD protection devices (a) conventional dual diode configuration in 'Device 1' (b) a gate-coupled NMOSFET in 'Device 2'

To contextualize, we compare our analytical result with experimental ADC data. Fig. 4.3 shows the estimated *SFDR* performance of junction-based ESD structures with measured ADC *SFDR* performance data[22] published at the International Solid-State Circuit Conference (ISSCC). For the analytical curves, we use the same parameters that were used to generate Fig. 4.2, except for C_{j0}, which is varied from 0.35 pF to 1.41 pF. C_{jQ} correspondingly changes from 0.25 pF to 1 pF. Furthermore, we assume that *SFDR* [dB] = - HD_3 [dB] as justified above.

According to this comparison, most of the recently developed ADCs still show a performance margin of approximately 3…5 dB below the ESD linearity limit, assuming 1 pF ESD devices. As the junction capacitance is reduced to 0.25 pF, the *SFDR* limit becomes approximately 100 dB at 100 MHz. That is, the distortion due to ESD protection can be effectively suppressed by reducing the size of the protection devices, as is often done in RF IC applications [50].

It should be pointed out that this Volterra series analysis with a simplified junction capacitance model is applicable to the majority of ESD protection devices in which a reverse-biased junction capacitance is the primary cause of its nonlinearity. The dual diode protection in Fig. 4.4a is a good example that falls in this category[23]. Other examples include ESD protection devices based on MOSFETs [1], such as the gate-coupled MOSFET and substrate-pumped MOSFET circuits. Fig. 4.4b shows a gate-coupled NMOSFET circuit. Although this protection device consists of a variety of components, such as C_{db} (drain-body junction capacitance of M_1), C_{gd} (gate-drain capacitance of M_1), resistors etc., the nonlinearity of the total capacitance is typically governed by the reverse-biased junction capacitance C_{db}. M_1 is completely turned off when V_{DD} is high; therefore, C_{gd} has a relatively small variation. Furthermore, MOS capacitance is also almost constant over the input signal range (further details are discussed in Section. 4.3).

4.3 Experimental Results and Discussions

To experimentally verify the impact of ESD protection circuits on high frequency linearity, we measured the distortion at the input pins of two commercially available ADCs. ‘Device 1’ is TI

[22] From Eq. 4.3, it is clear that HD_3, and thus *SFDR* depend on the applied input signal amplitude. Most of the data in Fig. 4.3 are based on a peak-to-peak differential input (V_{pp}) of 1-3 V. The shown analytical result corresponds to an input amplitude ($\hat{v}_i$) of 0.5 V (2 V_{pp} differential).

[23] To apply the Volterra series analysis to the dual diodes, v_o in Eq. 4.1 should be replaced by $-v_o$ for the reverse-biased junction capacitance of the ‘diode_up’. Therefore, the first order coefficient of the ‘diode_dn’ capacitance is compensated by that of the ‘diode_up’ capacitance, significantly improving HD_2. However, the second order terms are added, thus, HD_3 always becomes worse as the total junction capacitance increases.

ADS5221 [74]; a 12-bit, 65 MS/s ADC in 0.35 μm CMOS, with an *SFDR* of 85 dB at 32.5 MHz. The ESD protection circuit within 'Device 1' is a conventional 100 μm × 77 μm dual diode ESD circuit configuration as shown in Fig. 4.4a. 'Device 2' is TI ADS5500 [75]; a 14-bit, 125 MS/s ADC in 0.18 μm CMOS, with a typical *SFDR* of 82 dB at 100 MHz. The ESD cell within this IC is composed of a gate-coupled NMOSFET [31], as shown in Fig. 4.4b. The ESD cell of 'Device 2' occupies an area of 90 μm × 221 μm. The parasitic capacitances associated with the ESD protection devices are approximately 0.7 pF in both 'Device1' and 'Device 2.'

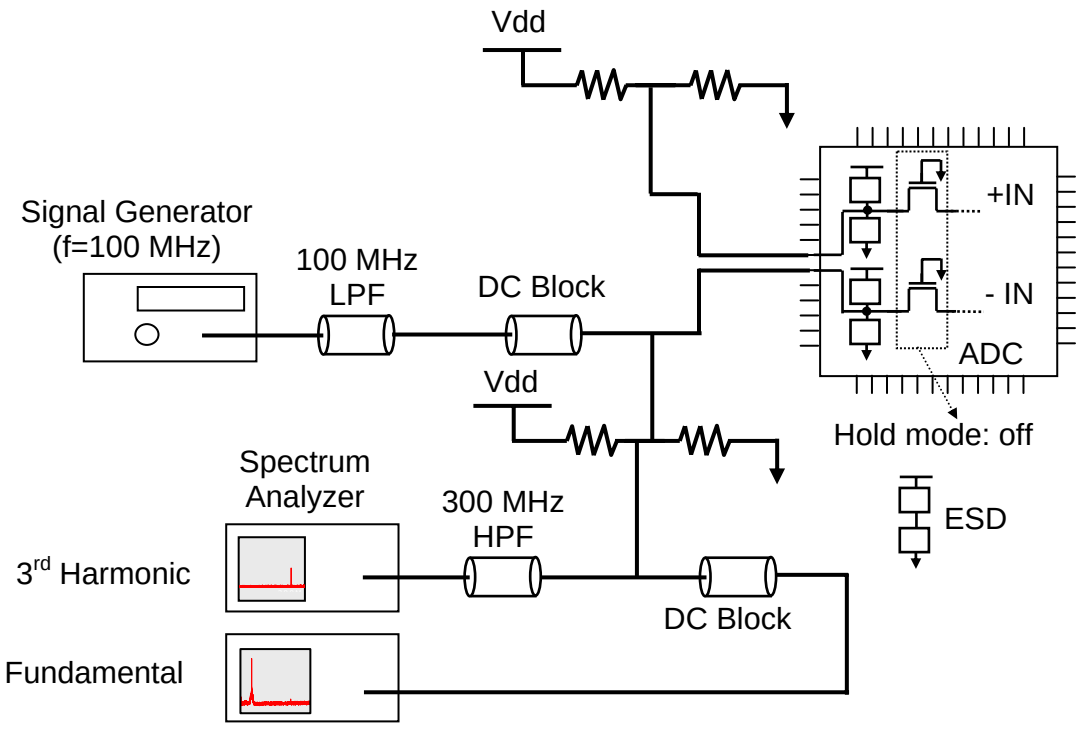


Figure 4.5: Experimental setup. A signal is applied to one of the ADC input pins through a low pass filter (LPF) and the harmonics are collected using a high pass filter (HPF).

The experimental setup is shown in Fig. 4.5. An Agilent E4432B signal generator and E4407 spectrum analyzer are used to generate and measure the harmonic signals. The ADCs are mounted on PCB boards, and external coaxial filters are connected as illustrated in Fig. 4.5. A single-ended sine wave is applied to one of the ADC input pins via a low pass filter (LPF) that removes the generator's harmonics. The harmonics caused by the ESD structures of the ADC

input pin are collected using high pass filters (HPF). For instance, a 100 MHz LPF is used to apply a pure 100 MHz signal and a 300 MHz HPF is selected to measure the resulting 3rd harmonic.

During this test, the ADC is placed into its 'hold mode,' which means that the internal sampling switches connected to the external pin are open. This precaution was taken to avoid measuring additional nonlinearities stemming from interactions with internal circuit components [76]. By measuring the harmonics without mounting the ADCs, we also confirmed that the signal distortion due to the PCB boards, the connectors and the coaxial filters are negligible. The impact of the package and the input pad are not separately measured, however the distortion due to these components should be minor comparing to the impact of the ESD protection devices.

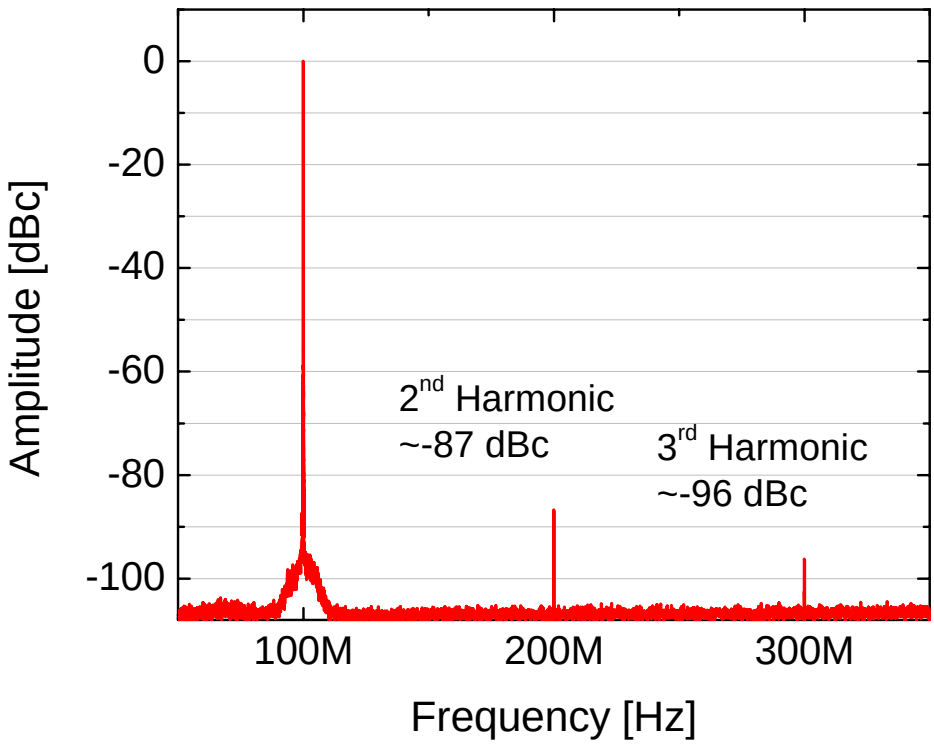


Figure 4.6: Measured signal at the input of 'Device 2'. The fundamental signal has an amplitude of 0.5 V.

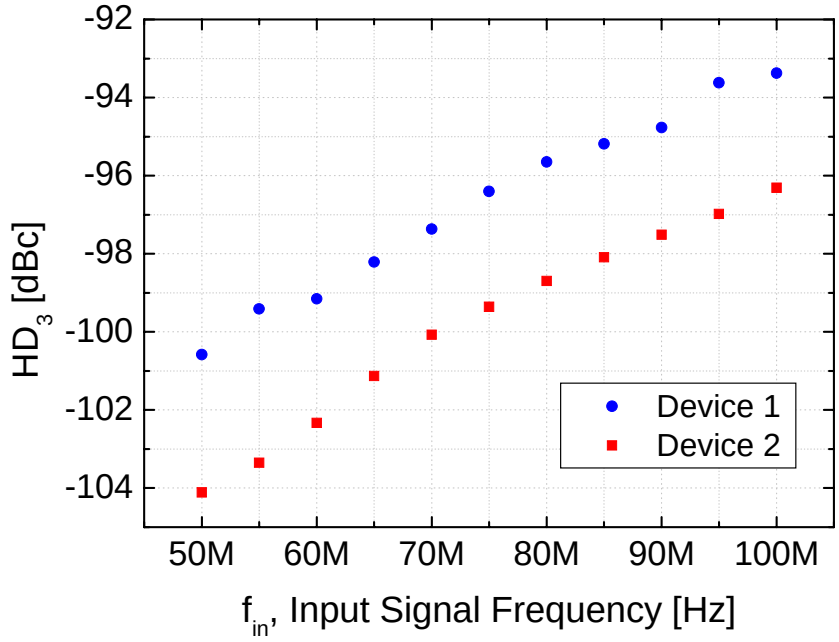


Figure 4.7: The measured 3rd order harmonics of 'Device 1' and 'Device 2' versus input frequency.

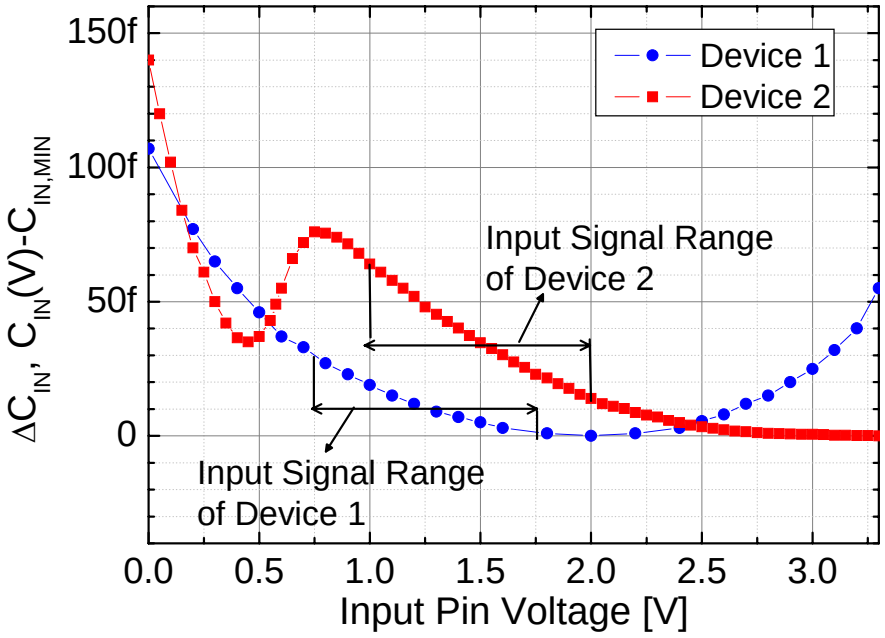


Figure 4.8: The measured input capacitance of 'Device 1' and 'Device 2' versus input voltage with reference to the minimum value of each capacitance.

Fig. 4.6 shows a typical spectrum at the input pad of 'Device 2.' Here, the DC bias is set to 1.5 V, and the signal amplitude is 0.5 V at an input frequency of 100 MHz. The measured magnitude of the 3rd order harmonic is -96 dBc. Fig. 4.7 shows the measured 3rd order harmonic of 'Device 1' and 'Device 2' versus the input frequency. In both cases, the distortion of the ESD device alone is well below the ADCs' overall performance targets (85 dB *SFDR* at 37.5 MHz, and 82 dB *SFDR* at 100 MHz).

Based on this data, it is evident that the linearity of the structure shown in Fig. 4.4b is superior to that of the basic protection circuit of Fig. 4.4a. In Fig. 4.8, the variations of the input pin capacitance vs. the input voltage are plotted. The variation is determined primarily by the ESD protection devices since the pad and package capacitances are almost constant over the large voltage range. As shown in Fig. 4.4a, the parasitic capacitance of 'Device 2' consists of the MOSFET drain junction capacitance parallel to a MOS capacitor. The MOS capacitor contributes about 25 % to the total capacitance of the structure. The total capacitance of 'Device 2' shows highly nonlinear characteristics when the input voltage varies from 0 V to 1 V, due to the large variation of the MOS capacitance up to its threshold voltage. However the MOS capacitance is almost constant above 1 V. Therefore, the capacitance of 'Device 2' is more linear compared to that of 'Device 1' over the input signal range of the ADC (0.75 - 1.75 V for 'Device 1' and 1.0 - 2.0 V for 'Device 2'). As seen in Fig. 4.7, this difference results in a 3-dB advantage in *SFDR*, even though the total parasitic capacitance of both devices is almost the same.

4.4 Summary

The signal distortion caused by ESD protection circuits is estimated using a variety of methods. The theoretical and experimental results indicate that the distortion from ESD protection circuits is typically more than 10 dB below the distortion introduced by state-of-the-art ADCs themselves. However, the numerical analysis also suggests that with *SFDR* targets approaching 100 dB at frequencies near 100 MHz, the protection device can easily become a performance-limiting factor in the future.

From this discussion, it follows that there are two basic options for mitigating signal distortion from ESD devices. First, distortion can be suppressed by minimizing the ESD device size. Reducing the size by one half improves the harmonic distortion performance by 6 dB. Second, instead of reducing the protection device size, which avoids sacrificing ESD robustness, one can attempt to make the effective ESD capacitance nearly linear by combining various types

of elements. An example of such a voltage-dependent compensation method has been described in [42].

Chapter 5

ESD FAILURE ANALISYS OF PMOS TRANSISTORS

In this chapter, the studies of PMOS transistors in sub-0.18 μm CMOS technologies are presented, focusing on the snapback and breakdown behavior of the parasitic PNP BJTs in the high current regime. A new failure mechanism of PMOSFET devices under ESD conditions is also analyzed by investigating various I/O structures in a 0.13 μm CMOS technology. Localized turn-on of the parasitic PNP transistor can be caused by localized charge injection from the adjacent diodes into the body of the PMOSFET, significantly degrading the ESD robustness of PMOSFETs. Based on 2-D device simulations the critical layout parameters affecting this problem are identified. Design guidelines for avoiding this new PMOSFET failure mode are also suggested.

5.1 Introduction

In the development of ESD protection devices, PMOS transistors have not drawn much attention due to the low β of the parasitic lateral PNP BJT, which results in small snapback and high on-resistance characteristics.

However, the performance of the parasitic PNP BJT in the snapback mode has continuously improved [77][78], as the features of CMOS technology have been scaled down [79]. Fig. 5.1

shows the I_{drain}-V_{drain} curves of the PMOS transistors of 180, 130, and 90 nm technology nodes [80]. The triggering voltage and the on-resistance have been dramatically changed; the triggering voltage has been reduced by ~ 1 V in each technology transition. The high-current β for a PMOS device in 90 nm technology becomes close to that of an NMOS in 130 nm technology [78][81]. Furthermore, the PMOS device shows more uniform conduction if n-well resistivity is sufficiently high, which allows reasonably high I_{T2} (> 4 mA/μm).

The improved performance of the PNP BJTs in PMOS transistors have shown promise that the PMOS transistors could be employed as active ESD clamping devices. However, the risk of inadvertent triggering also has been increased; raising new reliability issues [82].

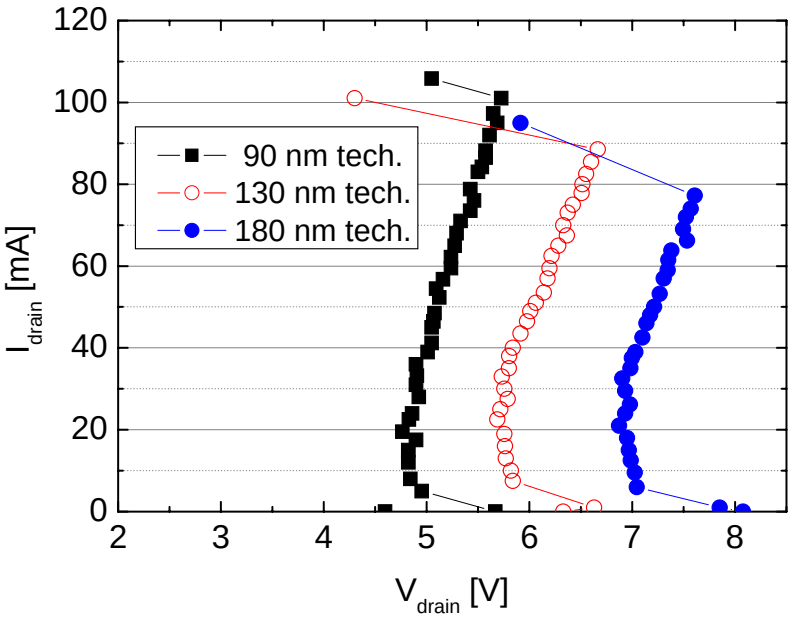


Figure 5.1: The I-V Characteristics of PMOSFEST in various technology nodes. The device width is 20 μm [80].

In the following section, the conventional triggering mechanism of the PMOSFETs in actual I/O structures is explained in detail. The design methodologies to minimize the reliability concerns related to the PMOSFETs are discussed, based on the study of Boselli et al. [78]. In Section 5.2, a recently observed PMOSFET failure phenomenon is presented [83]. The cause of this new failure will be analyzed in Sections 5.3 and 5.4. The new failure mechanism was examined using TLP (Transmission Line Pulsing) tests for the various I/O structures employing fully salicided, 3.3 V, 0.13 μm CMOS technology, and 2-D circuit/device mixed mode simulations. The design guidelines to prevent this PMOSFET failure are proposed in Section 5.5.

5.2 ESD Failures of PMOSFET Pull-up Devices

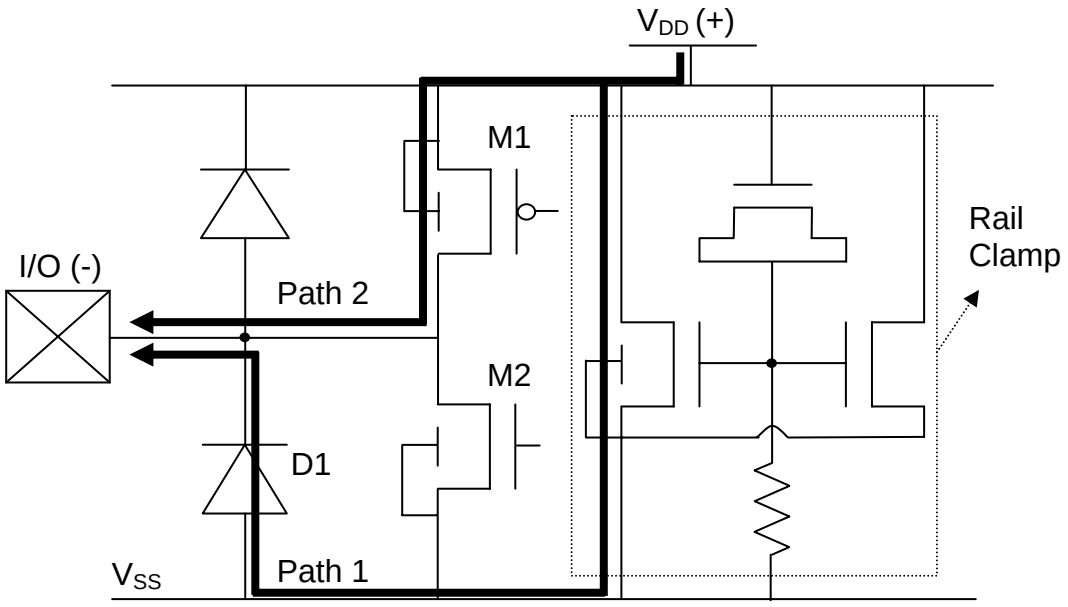


Figure 5.2: The generic configuration of an output driver with ESD protection circuit, where the pull-up PMOSFET has been reported to fail when a negative stress is applied between the I/O and V_{DD}.

Fig. 5.2 shows the schematic of a generic output driver, which includes an I/O pad, a pull-up PMOSFET (M1), a pull-down NMOSFET (M2), a rail clamp with gate-coupled and substrate-pump NMOS devices [32], and a negative strike diode (D1).

When the I/O is stressed negatively with respect to V_{DD} (or equivalently, V_{DD} positive with respect to I/O pad), the stress current is supposed to flow from the V_{DD} pad to the I/O, through the rail clamp and the negative strike diode (path 1 in Fig. 5.2). Depending on the effective on-resistance of path 1 and the applied ESD stress level, the voltage at the V_{DD} pad can rise high enough to force the pull-up PMOS into the snapback conduction mode, and path 2 in Fig. 5.2 starts to shunt current and compete with path 1. If the on-resistance of the PMOS in a snapback conduction mode is low enough to draw a significant amount of current from path 1, the PMOS can easily reach its second breakdown [78]. As mentioned in Section 5.1, the triggering voltage becomes lower as the channel length keeps shrinking; therefore the possibility of developing path 2 in Fig. 5.2 is also increasing.

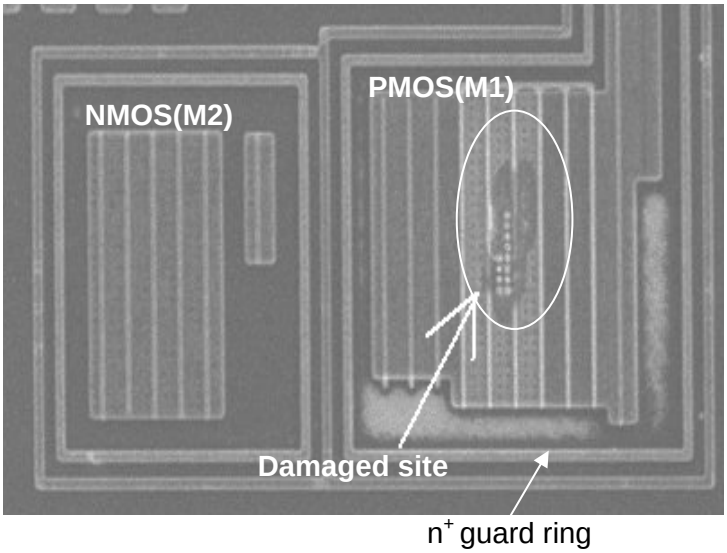


Figure 5.3: Location (circled) of the PMOS failure in an actual inverter. The pull-up PMOSFET (M1 in Fig. 5.2) and pull-down NMOSFET are shown.

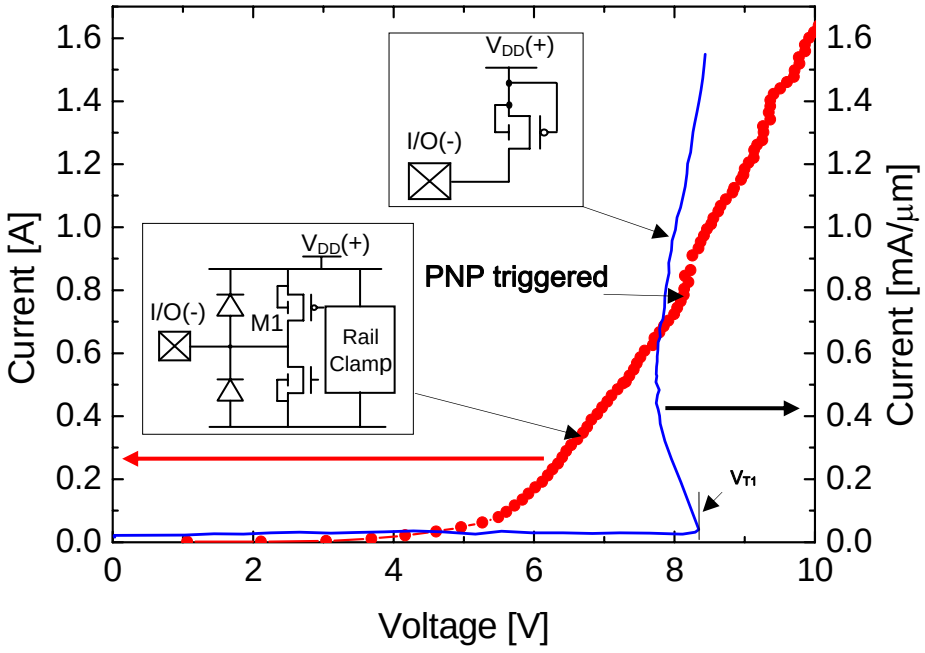


Figure 5.4: *I-V* curves generated by TLP tests of an output driver circuit and an isolated PMOS transistor. A large current (> 2.5 A) can be sustained by the output driver even after the stress voltage exceeds $V_{TI.}$ There is a distinct change of slope at I = 0.65 A where the parasitic PNP transistor is triggered on.

Fig. 5.3 presents an example of this PMOSFET failure in an actual output driver. As indicated by the circle, the damaged site is located at the center of the PMOSFET and shows the source-to-drain silicide filaments caused by the PNP triggering.

Boselli et al. [78] suggested various solutions to avoid this failure event. One of the simplest methods is to estimate the minimum PMOSFET width required to sustain the expected current level and then enlarge the PMOSFET more than this minimum width. A common circuit level solution is using the stacked devices for the pull-up PMOSFET; effectively doubling the triggering voltage. If the circuit performance can be compromised, a longer channel length device can be used instead of the minimum channel length device, or an additional resistance can be inserted in path 2 to impede the current flow. Some processes may provide a variety of device options such as devices with a low dose for the pocket and drain extension implants, which have a higher triggering voltage.

In Fig. 5.4, an example of a safe pull-up PMOSFET design is demonstrated. TLP measurement was done with V_{DD} positive with reference to the I/O. Note that the PMOSFET is properly sized to endure the expected stress level. Fig. 5.4 also illustrates an *Id-Vd* curve of an isolated PMOSFET, measured with connected gate and source. The isolated PMOSFET shows snapback behavior; the triggering voltage, V_{TI} is approximately 8.5 V. The complete I/O circuit depicted within Fig 5.5 has soft turn-on behavior inherited from the gate-coupled substrate-pump clamp [32].

The *I-V* characteristics of the complete I/O circuit also reveals that the voltage across the pull-up PMOS exceeds V_{TI} of the isolated PMOSFET, which means that the parasitic PNP of M1 is eventually triggered, and the stress current flows through both path 1 and path 2, as described in Fig. 5.2. As can be observed from Fig. 5.4, there is a distinct change of slope as the clamping voltage reaches the triggering voltage of the PMOSFET, leading to decreased on-resistance. In this I/O circuit, a large amount of current (> 2.5 A) can be sustained even after the stress voltage exceeds V_{TI}, since the PMOSFET is sized large enough to endure this level of ESD stress. It is also worth pointing out that, when this device eventually failed with extremely large current, the ESD damage was observed in the rail clamp, rather than in the PMOSFET (M1).

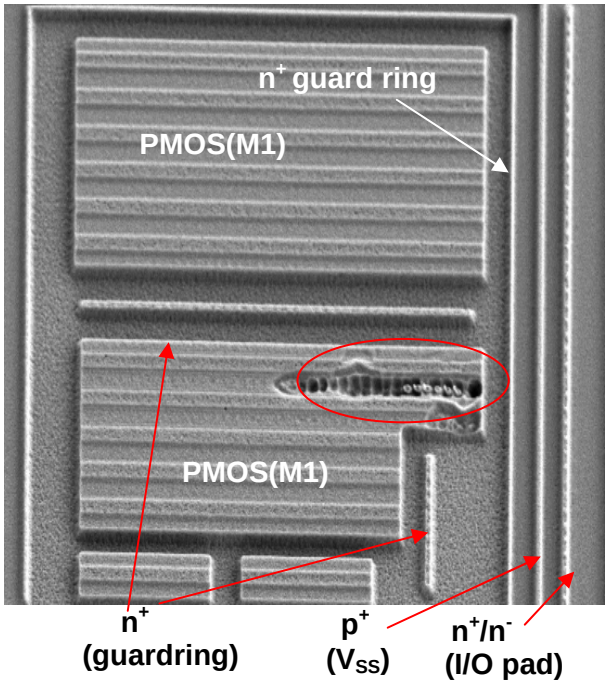


Figure 5.5 An example of PMOS failure in an actual I/O circuit. PMOS (M1 in Fig. 5.2) devices are located in two different blocks. A negative strike diode (D1 in Fig. 5.2), which consists of p+ (V_{SS}) and n+/n- (I/O pad) is on the far right. Between the two PMOS blocks and on the right side of the lower PMOS block, there are n+ diffusion lines connected to the guard ring. An ESD damaged site (the circled region) is observed close to the negative strike diode.

However, although the PMOSFET is sized large enough to endure the required stress level, the PMOSFET ESD damage phenomenon in certain I/O circuits are consistently encountered. The pattern of this behavior is quite different from that of the typical PMOSFET failure described in Fig. 5.3. Fig. 5.5 shows an example of new PMOSFET failure patterns. This specific I/O has a complex guard ring shape with a short n^+ stripe on the right side of the lower PMOSFET block and another n^+ stripe between the two PMOSFET blocks. The two vertical stripes on the far right are p+ diffusion (connected to Vss) and n+/n-well diffusion (connected to I/O pad), forming a negative strike diode (D1 in Fig. 5.2). It should be noted that the ESD damaged site with molten drain/source contacts [84][85] is observed close to the negative strike diode but not at the center as in Fig. 5.3. This failure pattern appeared consistently in certain I/O circuits with failing at 1.7 kV HBM condition despite the fact that they are designed to endure more than 4 kV. This fact indicates that there is a systematic effect which lowers ESD immunity level.

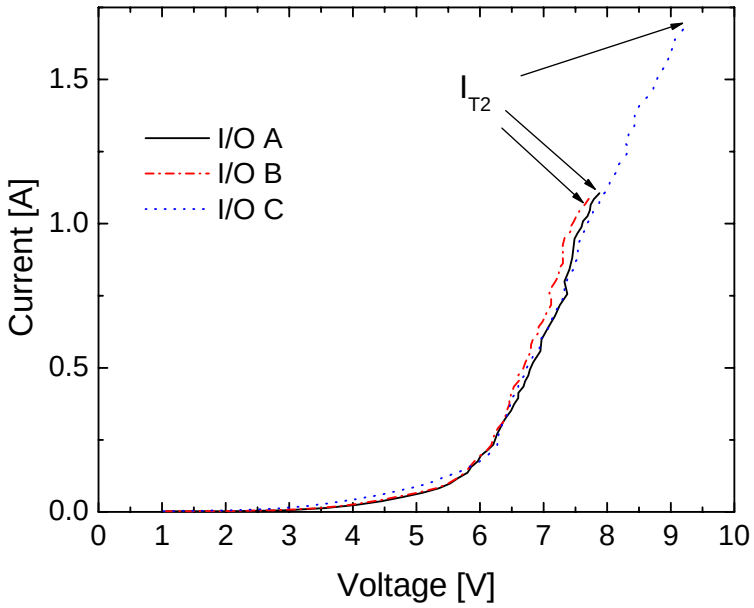


Figure 5.6: Measured TLP *IV* curves of various I/O structures up to their failure points.

	DR1	n^+ guard ring	HBM Fail Level
I/O A	0.8 µm	Incomplete	1.7 kV
I/O B	0.8 µm	Incomplete	1.8 kV
I/O C	> 0.8 µm	Incomplete	2.2 kV
I/O D	1.0 µm	Incomplete	2.6 kV
I/O E	22 µm	Complete	> 4 kV

Table 5.1: Layout information and ESD test results of various I/O circuits. DR1 is the distance between the negative strike diode and the guard ring (see the representation of layouts in Fig. 5.7).

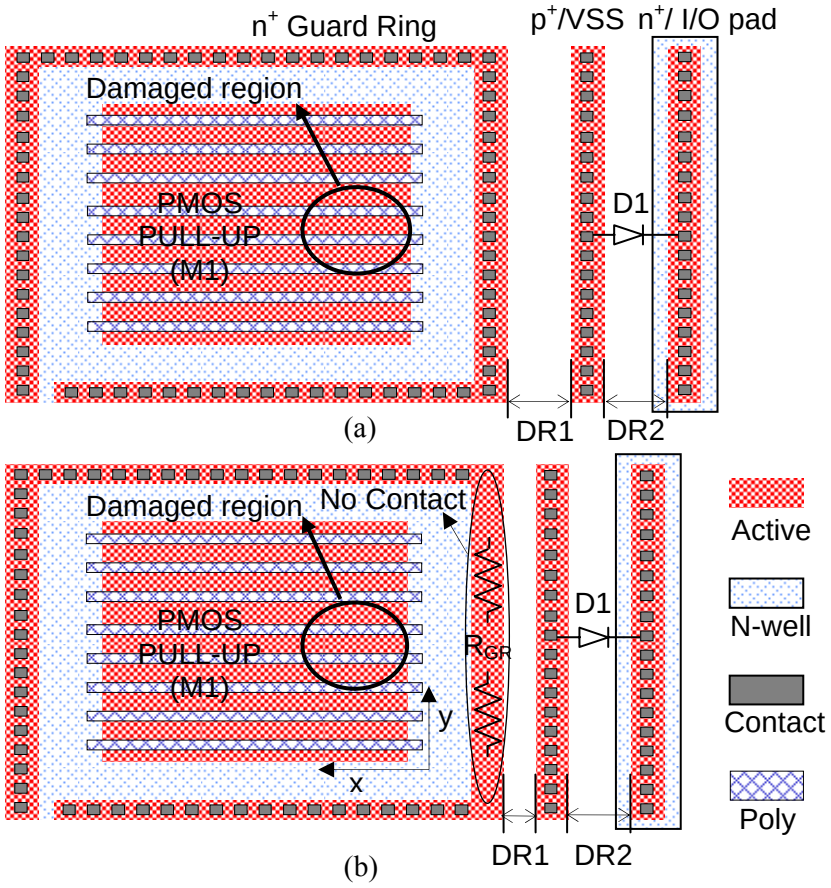


Figure 5.7: Simplified diagrams of the I/O layout. Only PMOS pull-up devices (M1 in Fig. 5.2) and negative strike diodes (D1 in Fig. 5.2) are illustrated. a) Its guard ring has metal contacts at every closure, b) metal contacts are missing at one stripe.

Fig. 5.6 shows TLP *I-V* curves of three different I/O circuits which have similar failure patterns displayed in Fig. 5.5. The design parameters and I_{T2} (translated to HBM failure level) of each device are listed in Table 5.1. In Fig. 5.7, schematics of the I/O layout are illustrated to facilitate the following discussions. The location of the ESD damaged spot is indicated in each schematic. The p-substrate to n-well diode (D1) is located adjacent to PMOS pull-up devices (M1), and ESD damage occurs in the vicinity of the diode and around the center of the gate fingers. The distance from p^+ V_{SS} to the n^+ guard ring is noted as DR1; the distance between p^+ V_{SS} and n^+ diffusion connected to I/O is DR2. The guard ring shown in Fig. 5.7a has metal contacts at every closure, but in Fig. 5.7b one stripe between the PMOS and diode does not have a metal contact to achieve a more compact design. The influence of this difference in the guard ring shape on ESD robustness will be discussed in detail in Section 5.4.

In Table 5.1, while the I_{T2} of I/O E (of which the measured *I-V* curve is not shown in Fig. 5.4) is higher than 2.5 A (~4 kV HBM), I/O A and B fail even before the voltage reaches V_{T1} of an isolated PMOSFET and show the HBM level lower than 2 kV. All device dimensions, such as the width of PMOSFETs and power clamps, are almost identical, whereas their layouts have some variations in the guard ring shape and the distance between the negative strike diode and the guard ring (referred to as DR1). For example, DR1 of I/O E is larger than 22 μm, but I/O A, B have 0.8 μm DR1. In I/O C, the PMOS block is not located parallel to the diode, and the shortest path from the diode to the guard ring around the PMOS block is about 0.8 μm long. The guard ring of I/O E is completely closed as in Fig. 5.3, while in other I/O circuits one side of the guard ring does not have metal contacts as illustrated in Fig. 5.7b.

Two observed facts should be emphasized here: first the damaged spots of I/O A,B,C and D are always located in the PMOS transistor near the negative strike diode, as can be seen in Fig. 5.5. Second, when the DR1 increases from 0.8 μm to 1.0 μm, the ESD failure level increases by approximately 1 kV. These TLP test results demonstrate that the distance between the negative strike diode and the guard ring (DR1), and the shape of the guard ring can be critical parameters determining ESD immunity of I/O circuits. The physics involved in this phenomenon will be clarified in the next section.

5.3 Analysis of a New Failure Mechanism

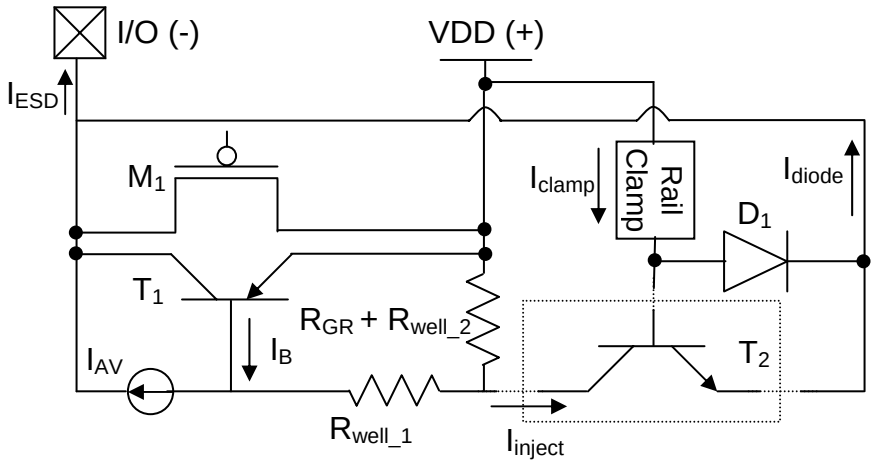


Figure 5.8: The schematic of the equivalent circuit under ESD conditions. The parasitic NPN BJT in the dotted box (T_2) should be considered only if a negative strike diode is placed close to the pull-up PMOSFET (M1).

Fig. 5.8 shows the schematic of the equivalent circuit of an I/O circuit under ESD conditions, which includes the devices involved in both path 1 and path 2 of Fig. 5.2. Here, R_{well_1} and R_{well_2} represent the resistance of n-well from the source/n-well junction to the n+ guard ring (see Fig. 5.9a). R_{GR} is the extra resistance caused by incomplete contact formation as illustrated in Fig. 5.7b. If the diode D1 is distant from the pull-up PMOS device, path 1 and path 2 can be analyzed independently. In this case, the parasitic PNP BJT (T1) is triggered through the self-biasing using the internal avalanche generation of carriers (I_{AV}) [86]. Here, R_{well_1}, R_{well_2} and R_{GR} act as effective body resistance of the PMOSFET. However, if these two devices are moved closer to one another, the analysis with D1 in Fig. 5.8 is no longer valid, and the parasitic NPN BJT (T2 in the dotted box of Fig. 5.8) should be considered instead of D1; this NPN transistor supplies the external triggering current (I_{inject}) to the base of PNP transistor. This explanation can be clarified by the simulation results shown in Fig. 5.9.

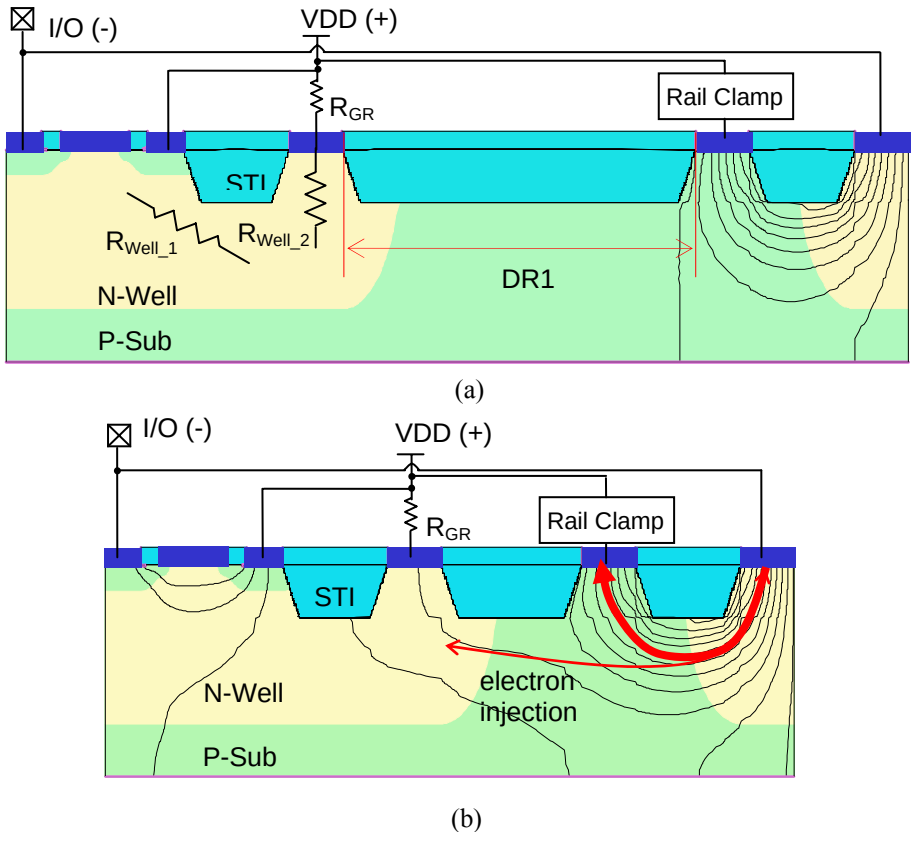


Figure 5.9: Current Flow lines under ESD conditions. a) The PMOS transistor is separated from the negative strike diode (DR1 = 3 µm), b) the PMOS is close to the diode, therefore, the two devices are electrically coupled (DR1 = 1 µm).

To investigate the influence of each layout parameter variation, circuit/device mixed mode simulations were performed. The device structure and external circuit components employed in the simulation are illustrated in Fig. 5.9. The rail clamp in Fig. 5.9 is implemented with a constant resistance, R_{clamp}, since the on-resistance of rail clamps can be considered to be constant in a wide high-current range [87]. When parasitic PNP transistors are triggered, the value of R_{clamp} is extracted from the measurement results. Fig. 5.9 also illustrates the simulated current flowlines of I/O devices represented by the equivalent circuit of Fig. 5.8. In Fig. 5.9a, the negative strike diode is located at a distance from the n^+ guard ring (DR1 ≈ 3 μm); therefore, most of the electrons emitted through the n-well/substrate junction from the I/O pad recombine with the holes from V_{SS} which is connected to the rail clamp. However, if the two devices are located close to one another (DR1 ≈ 1 μm), some of the electrons from the diode are injected into the n-well body of the PMOSFET, modulating the base potential of the PNP transistor (T1 in Fig. 5.8). This process results in the early turn-on of the parasitic PNP transistor. If R_{GR} increases (as in the case of Fig. 5.7b), the base potential is further modulated due to the voltage drop across the R_{GR}, resulting in lower triggering voltage.

These simulation results and the analysis of the equivalent circuit in Fig. 5.8 explain the cause of the PMOS failure reported in the previous section. The structure of Fig. 5.7b has several drawbacks compared to the guard ring structure of Fig. 5.7a. First, due to the smaller distance from the diode, the body potential of PMOSFET (the base potential of the PNP transistor) is modulated by the current injected from the diode. Second, incomplete guard ring metal causes spatial variations of the effective body resistance, which is reflected in R_{GR} of the equivalent circuit. These two drawbacks cause the non-uniform turn on of the PNP transistors and degrade the ESD immunity of I/O circuits. In the following section, the influence of this layout variation is discussed in detail.

5.4 Influence of Layout Parameter Variation

Considering the fact that the BJT is a current controlled device, we can formulate the triggering voltage for the case in which the PMOS and the negative strike diode are electrically coupled, with reference to the equivalent circuit in Fig. 5.8.

$$I_{inject} = I_{clamp} \cdot \beta \qquad \text{Eq. 5.1)}$$

$$V_{DD\_Pontential} = I_{clamp} \cdot R_{clamp} + V_{diode} = I_{inject} \cdot \frac{1}{\beta} \cdot R_{clamp} + V_{diode} \quad \text{Eq. 5.2)}$$

$$V_{body,\,trig} \approx (R_{GR} + R_{well\_2}) \cdot (I_{AV} - I_B + I_{inject,\,trig}) + R_{well\_1} \cdot (I_{AV} - I_B) \quad \text{Eq. 5.3)}$$

Here, β is the common emitter current gain of the NPN transistor (in the dotted box of Fig. 5.8). I_{inject} is the electron current from the negative strike diode into the n-well of the PMOS transistor. It is assumed that most of this electron current flows directly to the guard ring contacts through R_{well_2}, as shown in the current flowlines of Fig. 5.9. V_{diode} is the voltage drop through the negative strike diode and is a function of DR2 in Fig. 5.7. $I_{inject,\,trig}$ is I_{inject} at the PNP triggering point. $V_{body,\,trig}$ is the body potential near the source/body junction with reference to V_{DD} at the triggering point and can be considered to be a constant. In Eq. 5.3, $I_{inject,\,trig}$ is much larger than I_{AV} if the PMOS and the diode are located in proximity to one another, and then Eq. 5.2 can be approximated as follows.

$$VDD\_Potential\ @\ pnp\ \ triggering = V_{T1} \approx \frac{V_{body,\,trig}}{R_{GR} + R_{well\_2}} \cdot \frac{1}{\beta} \cdot R_{clamp} + V_{diode} \quad \text{Eq. 5.4)}$$

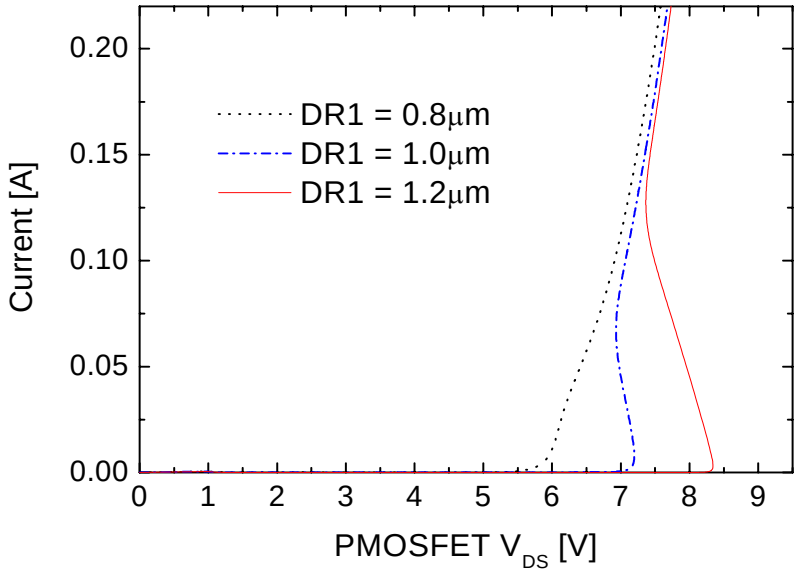


Figure 5.10: Simulated *Id-Vd* characteristics of a PMOS transistor in I/O circuits with various distances between the guard ring and the negative diode. DR1 lower than 1.2 μm considerably affects the triggering voltage of PNP transistors. Here, R_{GR} is 0 Ω and DR2 is 0.6 μm.

As the DR1 decreases, the common emitter current gain (β) of the parasitic NPN increases; hence, V_{T1} decreases as expected in Eq. 5.4. As can be seen in Fig. 5.10, the triggering point of parasitic PNP of the PMOSFET is a strong function of DR1. For instance, when DR1 is reduced

from 1.2 μm to 1.0 μm, V_{T1} decreases by ~ 1 V. The device with 0.8 μm DR1 is triggered around 5.7 V, without showing snapback behavior. However, if the DR1 is larger than 1.2 μm, V_{T1} is not sensitive to DR1 variation. In other words, the localized electron injection from the negative strike diode into the body of PMOSFETs significantly increases if the DR1 is less than a certain value and causes non-uniform conduction along the x-direction in Fig. 5.7; therefore, deteriorating the ESD robustness of I/O circuits.

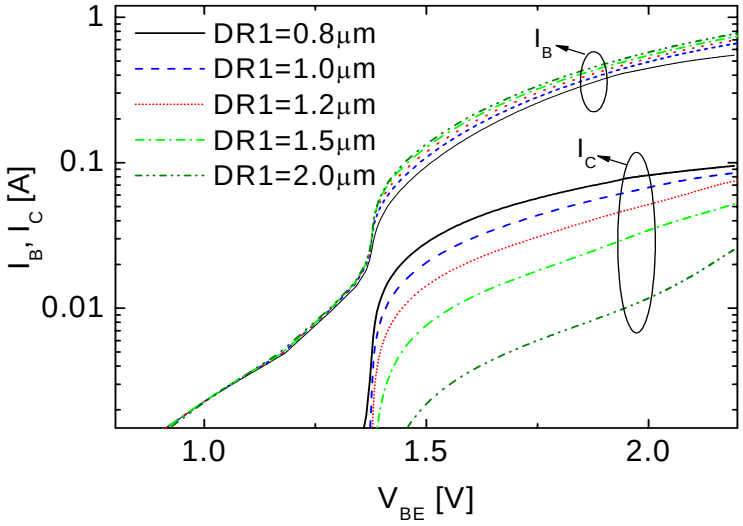


Figure 5.11: Simulated collector and base currents of the parasitic NPN transistor (T2 in Figure 5.8) as functions of the emitter-base voltage with various distances between the negative strike diode and the guard ring (DR1).

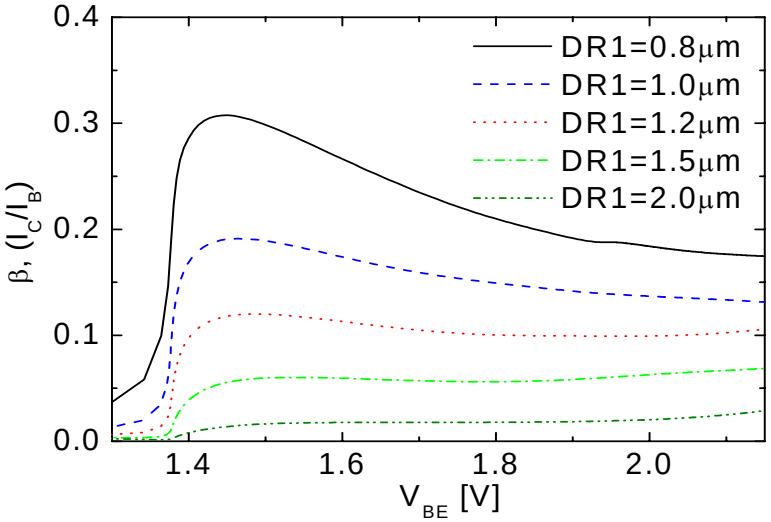


Figure 5.12: Simulated current gains as functions of emitter-base voltage with various distances between the negative strike diode and the guard ring (DR1).

To investigate the effect of DR1 variations on the characteristics of the parasitic NPN transistor, the collector current (I_C, same as I_{inject}), base current (I_B, same as I_{clamp}) and the static common-emitter current gain, β are plotted in Figs. 5.11 and 5.12. Since the doping concentration of the emitter (n-well) is not much higher than that of the base (p-substrate), the emitter efficiency (γ) is much smaller than 1. If the DR1 is larger than 1.5 μm, the base transport factor (α_T) is much less than unity; therefore, β is less than 0.1. However, the peak β increases over 0.1 as the DR1 decreases to 1.2 μm, and a large current is injected into the body of PMOSFET, causing the early turn-on of the PNP transistor.

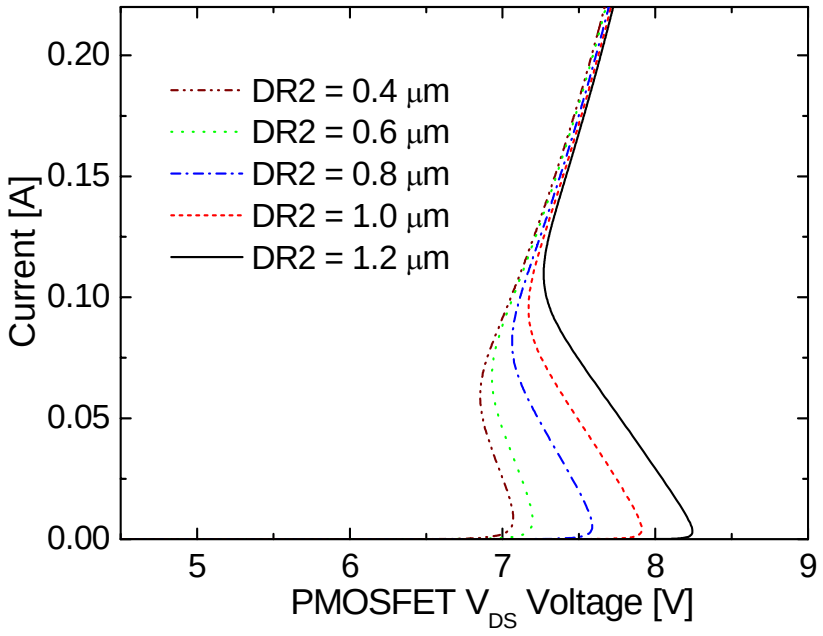


Figure 5.13: Simulated *Id-Vd* characteristics of the PMOS transistor in I/O circuits with various distances between the p+ diffusion and n+ diffusion of the negative diode (DR2). Here, R_{GR} is 0 Ω and DR1 is 1.0 μm.

Fig. 5.13 demonstrates that as the DR2 (the distance between p+ VSS and n+ diffusion connected to I/O pad) increases, the triggering voltage also increases, due to the decrease of T2's β and the increase of V_{diode} in Eq. 5.4. However, the increase of the DR2 also causes the on-resistance of path 1 in Fig. 5.2 to be increased; therefore, increasing the DR2 is not an efficient solution to mitigating the stress on the PMOS in path 2.

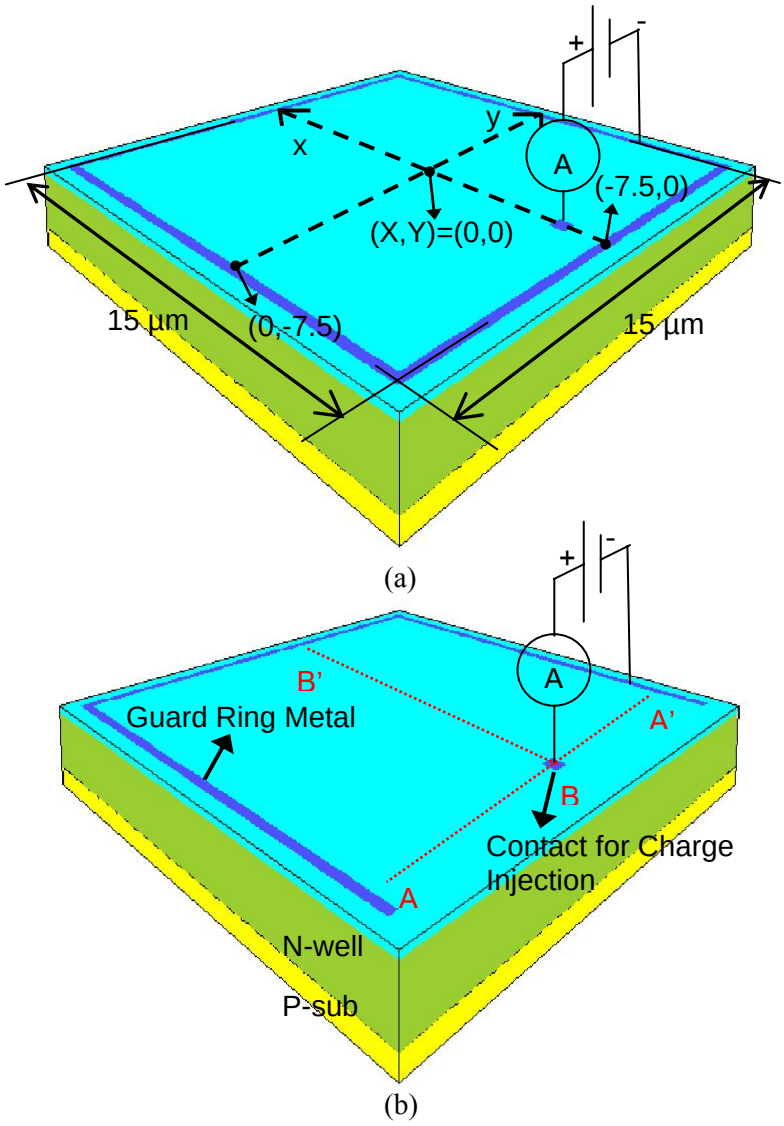


Figure 5.14: Extraction of spreading well resistance (R_{well}) (a) a closed guard ring as in Figure 5.7a, and (b) a U-shape guard ring as in Fig. 5.7b**.**

In the mixed mode simulations, a lumped resistor, R_{GR} is employed to reflect the finite conductivity of the n+ diffusion layer of the guard ring (see Fig. 5.7b). However, all segmented devices are also connected through the n-well as well as through n+ diffusions. The spreading resistance should be calculated for extracting an effective R_{GR}. As shown in Fig. 5.14, 3-dimensional simulations were performed to calculate the distributed well resistance and to investigate the effect of guard ring structure variation on the well resistance. In Fig. 5.14, two guard ring structures are shown: the one is the closed guard ring as in Fig. 5.7a; the other is the U-

type guard ring in which one stripe has no metal contacts but merely n+ diffusion as in Fig. 5.7b. In each structure, a contact is located within the guard ring closure for injecting electron current which flows to the guard ring. By measuring the voltage difference between this contact and the guard ring, we can extract effective R_{well} from a certain location within the guard ring closure to the guard ring.

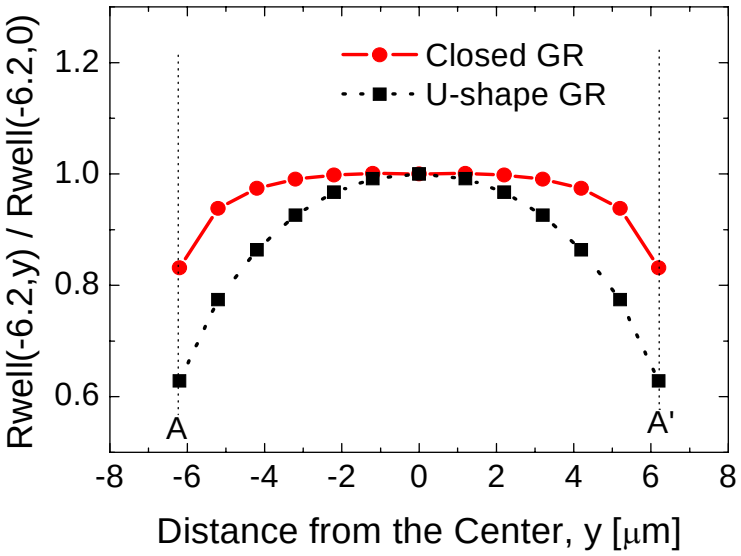


Figure 5.15: Variations in the spreading well resistance along the y-direction near guarding (x = -6.2 μm, along the line A-A'). The spreading well resistance is normalized with reference to the value at the center.

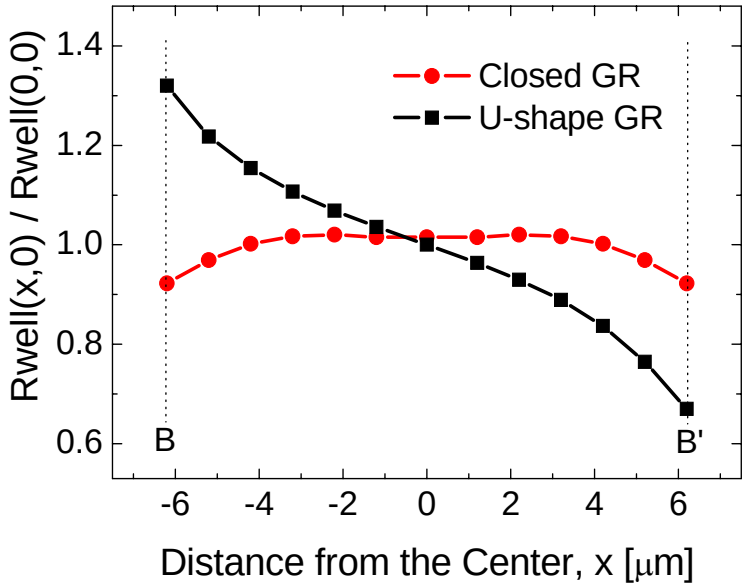


Figure 5.16: Variations in the spreading well resistance along the x-direction (y = 0 μm, along the B-B' line). The spreading well resistance is normalized with reference to the value at the center.

Fig. 5.15 shows R_{well} variations in two guard ring structures along the A-A' line (close to the negative strike diode) of Fig. 5.14b. The extracted resistance values are normalized with reference to the value at the middle of A-A'. At the location away from the center of A-A' by 3 μm, the R_{well} of the U-type guard ring structure decreases by ~7 %, while the variation of R_{well} inside the closed guard ring is less than 1 %. The variation of R_{well} along the line B-B' is also plotted in Fig. 5.16. As shown in Figs. 5.15 and 5.16, the device with a U-shape guard ring shows considerable variation in the spreading R_{well}, which may cause non-uniform conduction along both the x and y directions.

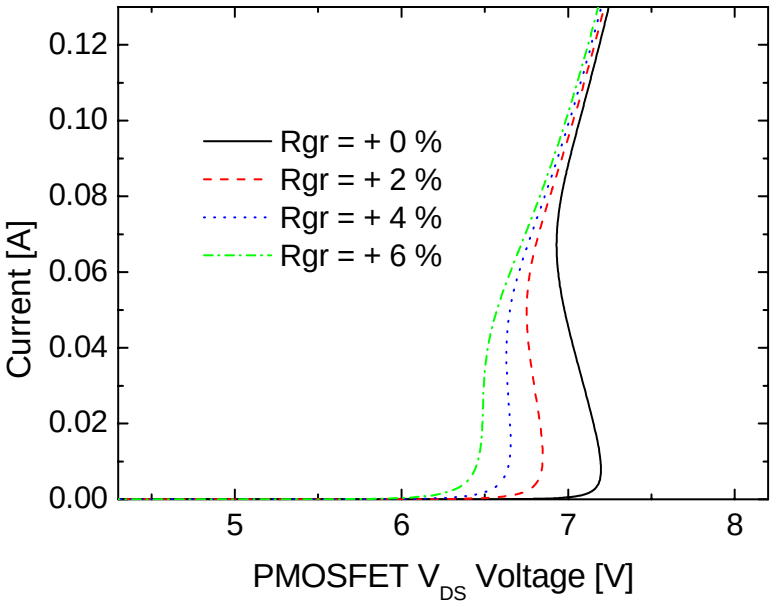


Figure 5.17: *Id-Vd* characteristics of PMOS transistors in I/O circuits with various external resistance between the body contact and the V_{DD}. DR1 and DR2 are 1.0 μm and 0.6 μm, respectively.

For a thorough investigation of the effect of this R_{well} variation on ESD immunity, 3-D ESD simulations with a variety of guard ring structures should be performed. However, in this study the effect of R_{well} variation is investigated using 2-D simulation by simply varying the lumped resistance, R_{GR} between V_{DD} and the n+ guard ring. In 2-D simulations, the well resistance, R_{well0} (= R_{well1}+R_{well2}, of Fig. 5.8) is extracted from the I_{sub} vs I_d curve [86][87] with zero R_{GR}; then a certain fraction of R_{well0} is applied to R_{GR}. As expected from Eq. 5.4, a higher R_{GR} causes the body potential of the PMOSFET to increase, while the PNP transistor is triggered at lower V_{T1}. Fig. 5.17 shows that the triggering voltage with an R_{GR} of 2% R_{well0} is lower than the triggering voltage with zero R_{GR} by ~ 0.4 V. This observation implies that the finite resistance of the n+

guard ring without a metal contact results in non-uniform potential distributions across the body of PMOS transistors and in the localized current conduction. For the simulations in Fig. 5.17, the DR1 and the DR2 are assumed to be 1 μm; there is a significant coupling between the negative strike diode and the PMOSFET as examined in Figs. 5.10 and 5.13. In this case the triggering voltage, V_{TI} is a strong function of R_{GR} according to Eq. 5.4. However, if the diode and the PMOSFET are completely separated (the DR1 and the DR2 are large), the impact of the R_{GR} variation on the V_{TI} becomes negligible.

5.5 Design Impact

This study demonstrates that the layout of the PMOS relative to diffused regions connected to the I/O pin, can play a role in the effectiveness of the overall protection at the pad. The best possible option is to move the PMOS layout more than 20 μm away from the I/O protection devices. However, in most cases this long distance may not be practical. In an ASIC environment various types of buffers would use the same protection cell at the pad, but the output NMOS and PMOS are laid out differently depending on the design features. Some buffers might tolerate a series resistor between the I/O pad and the PMOS. In such cases, the interference from the PMOS is not relevant, and the final protection level is determined by the current limited to the PMOS, or the failure current limit of the Vdd protection device itself. However, in common output buffer designs, a resistor is not tolerated, along with the fact that the PMOS device sizes differ depending on the designed buffer's performance. For example, a 4 mA buffer would have a relatively larger PMOS than for a 2 mA buffer. However, if the layout is ineffective, then the PMOS size does not matter. Once the layout is made more effective as described in Section 5.4, then the PMOS size would play a role in determining the overall protection level. In this case, for ESD purposes, it is desirable to make the 2 mA buffer PMOS look like the 4 mA buffer. This may be achieved with the introduction of a dummy PMOS as mentioned in [82].

Obviously the critical factor is the layout. Specifically, in an ASIC library environment, the placement of the protection devices and the latchup guard rings can be complex, due to the available macro pitch and the bus architecture. If the negative diode at the pad has to be placed in proximity to the PMOS, then it is essential to completely close the guard ring around the PMOS with full contacts and maintain DR1 in Fig. 5.7 to be > 2 μm. If this spacing can be >10 μm, then the guard ring shape should not matter.

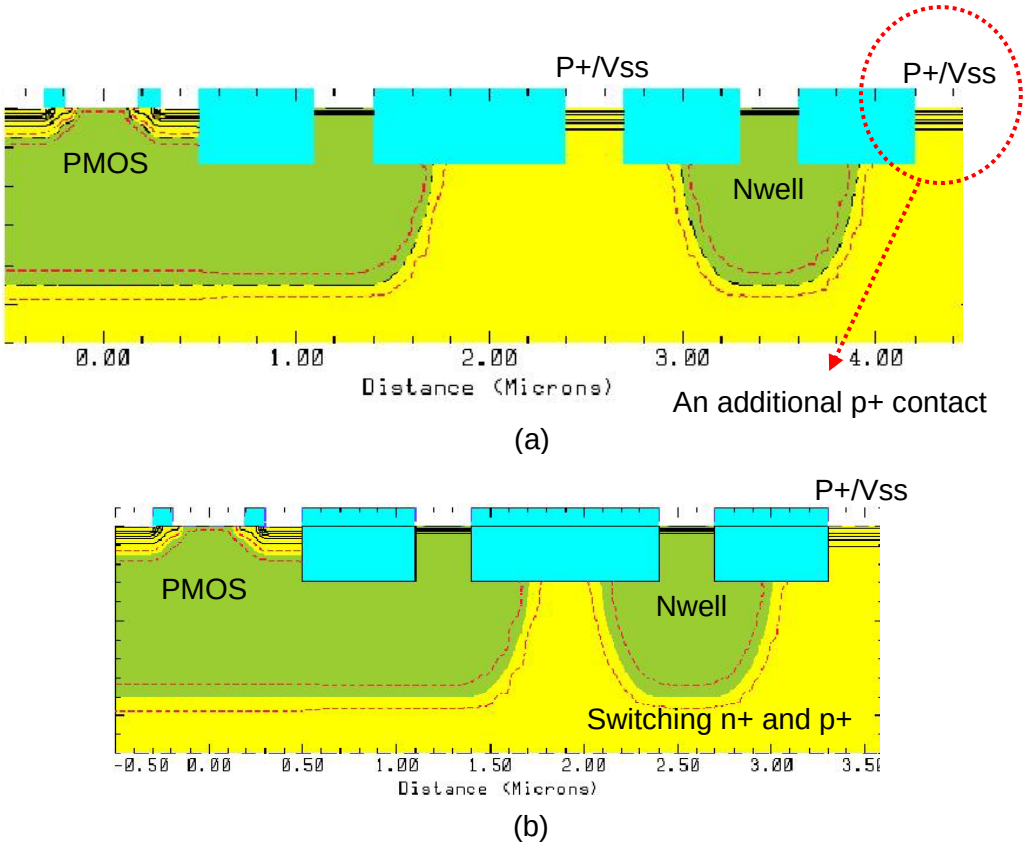


Figure 5.18: Possible layout modifications to reduce the charge injection from the negative-striking diode to the PMOSFET.

Simulation studies also suggest a few possible layout modifications. First, an additional p+ diffusion connected to Vss can be placed right to the n+/nwell, as shown in Fig. 5.18a. In this case, the number of electrons flowing to the left p+ diffusion is reduced by half, therefore electron injection to the body of PMOSFET also decreases. This solution may be more efficient for creating a compact design than increasing the space between the PMOSFET and diode. A second possible solution is switching the order of p+ diffusion and n+/nwell as shown in Fig. 5.18b. Even though n+/nwell is closer to the PMOSFET, actual p-base of NPN transistor is far from the base contact (p+ Vss); therefore, the base potential modulation of the parasitic NPN transistor is reduced and the coupling between the PMOSFET and diode can successfully be mitigated.

5.6 Summary

A new failure phenomenon of PMOSFET pull-up devices under ESD conditions has been reported and analyzed by investigating various I/O structures of 3.3 V, 0.13 μm CMOS

technology. The physical mechanism and the influence of layout parameters such as the distance between a negative strike diode and an n+ guard ring (DR1), guard ring shape, and effective resistance of the rail clamp, were investigated through Transmission Line Pulse (TLP) testing and device/circuit mixed mode simulations.

Localized turn-on of the parasitic PNP BJT of the PMOSFET pull-up device is caused by the localized charge injection into the body of PMOSFET from the negative strike diode. The most critical layout parameter for this failure is DR1. With the technology used in this study, values of the DR1 greater than 2 μm are recommended for a safe design. This phenomenon can be implicated in ESD failures of all future ASIC I/O circuits which require more compact layouts.

Chapter 6

ELECTRO-THERMAL SIMULATIONS OF ADVANCED DEVICES UNDER ESD CONDITIONS

The continuous scaling and introduction of new device concepts and materials to mainstream CMOS technologies have brought many new thermal issues. Thermal modeling of transistors and interconnects are gaining more importance, and electro-thermal simulations with full coupling between the electrical and thermal equations are playing an important role to predict the device performance not only under normal operations but also under ESD stress conditions.

In this section, various thermal issues of the nano-scale devices are briefly reviewed with an emphasis on the relevance to the ESD simulations. After that, an example of ESD electro-thermal simulation will be demonstrated in Section 6.2; the electro-thermal characteristics of strained-Si MOSFETs operating in the high-current, high temperature regime were investigated using device/circuit mixed mode simulations. The material parameters of strained-Si were calibrated for the device simulations. In particular the phonon mean-free-path of strained-Si with high electric fields was estimated based on a full-band Monte Carlo device simulation. Despite the low thermal conductivity of the buried SiGe layers, strained-Si devices show superior ESD protection capabilities compared to unstrained-Si (bulk-Si) devices due to the high bipolar current gain and increased impact ionization rate. In Section 6.3, to take the nano-scale heat generation and phonon heat conduction into account, the optical and acoustic phonon temperature system for Technology CAD simulation is developed and self-consistently coupled with the hydrodynamic

transport model. This electro-thermal model is validated through the simulations of a thin-body SOI NMOSFET.

6.1 Introduction

As the technology scaling continues, power dissipation per unit area is increasing despite the fact that power supply voltage is reduced. Power densities in the modern microprocessors are on the order of 100 W/cm^2, and they are expected to further increase according to the ITRS roadmap [88]. This massive heating often results in the localized region of high temperature, so called 'hot spots.' The dimensions of the hot spots at the chip level are usually millimeter-scale and the time-scale for heat conduction from the hot spots to the package is approximately 0.001 ~ 1 second; hence in the chip-level analysis, the microscopic details of heat generation and transport are not crucially taken into account and the thermal system is often considered to be in the steady-state. Studies of millimeter-scale hot spot phenomena have been focused on the system level engineering; power efficient design [89][90], active cooling [91][92] and the analysis of heat transport via interconnect, dielectrics, and packages [93].

However, the studies of heat generation and transport within a single transistor have also received much attention [27]. The heat transport from advanced devices such as SOI (silicon-on-insulator), GOI (germanium-on-insulator), SiGe HBT, strained-Si FET [94], FinFET [95][96] are worse than conventional bulk-Si devices. The materials introduced in these devices have low thermal conductivities, or the generated heat is confined by the surrounding SiO_2 which has approximately two orders lower thermal conductivity than bulk-Si; the heat removal of these nano-scale electronic devices becomes enormously complicated.

Material	Thermal Conductivity, κ_{th} ($Wm^{-1}K^{-1}$)
bulk Si	148
bulk Ge	60
$Si_{0.7}Ge_{0.3}$	8
SiO_2	1.4
Silicides	40

Table 6.1: Thermal conductivities of the various materials adopted in advanced CMOS processes.

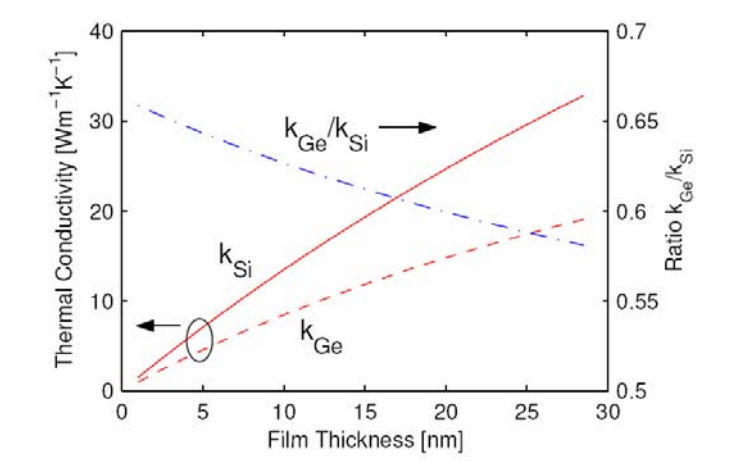


Figure 6.1: Estimated thermal conductivities of Si and Ge as functions of the film thicknesses [Reprinted from [99], courtesy E. Pop].

Table 6.1 shows the measured thermal conductivities of a variety of materials. For example, the thermal conductivity of bulk Ge is only 40 % of bulk Si. Furthermore, the thin film layers in the devices hinder heat transportation because of phonon boundary scattering [97][98]. Fig. 6.1 presents the estimated thermal conductivities of Si and Ge layers as functions of the layer thicknesses [99]. It is clearly shown that as the layer thickness decreases, the thermal conductivities are also dramatically reduced.

Most of studies on advanced devices have been focused on the enhancement of electrical characteristics, looking for the materials with better mobility that can achieve lower power dissipation with the same level of current driving capability. However, due to the drawbacks in heat transport discussed above, the degradation of the electrical characteristics due to self-heating could become sufficiently severe that it would counteract the advantages of the new materials and new device concepts. Hence, the advanced devices should be evaluated and optimized considering self-heating and its impact on the electrical characteristics [99]. In ESD investigations, thermal issues become even more aggravated since ESD events basically involve extremely high current and high temperature conditions [10][29][30].

As the channel resistance is continuously reduced, other resistive elements such as the drain/source extension resistance and the contact resistance between the silicide layer and doped Si source/drain regions become an appreciable portion of the total resistance [100]. These resistances are often a function of temperature [101]; the absence of the temperature-dependent model could result in significant errors in device and circuit simulations. To explore the self-heating impact of the contact resistance, Matsuzawa et al. [102] implemented a physical model of a silicide/Si hetero-junction, which is a unified model of thermionic emission and tunneling. In [103], to avoid the

complexity of the hetero-junction model, a numerical model of temperature-dependent contact resistance was proposed and implemented for TCAD simulations.

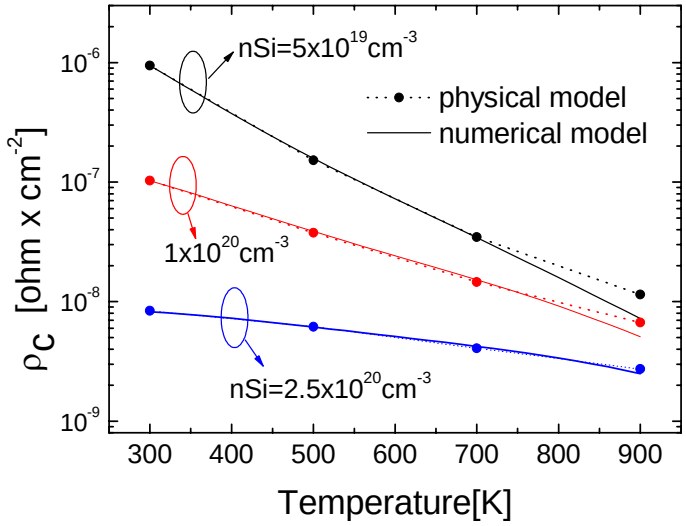


Figure 6.2: Comparison between the specific contact resistances (ρ_c) extracted from the hetero-junction simulation and those from the numerical model.

In Fig. 6.2, the specific contact resistance (ρ_C) values extracted from the complete hetero-junction model simulation [102] with a variety of ambient temperatures and doping concentrations are compared with the new numerical model [103]. Identical trends of doping and temperature dependence can be clearly observed.

Under ESD conditions, the contact resistance contributes even a more significant fraction of the total on-resistance since the conductivity of channel and substrate region is highly modulated. Moreover, recent studies [102][103] have shown that the self-heating caused by the contact resistance is also considerable under ESD condition. It is also known that the current distribution associated with the silicided contact system in the source/drain structure is strongly influenced by the contact resistance value [104][105]. Therefore, in order to achieve reliable electrothermal simulations of ESD, it is essential to trace the dynamic variation of contact resistance as a function of temperature within a very short time period. As shown in Fig. 6.2, ρ_C at temperatures approaching thermal failure is much less than that of room temperature so that the constant ρ_C over broad temperature range may cause an incorrect estimation of heat generation in the contact system. The electrothermal simulation using the numerical temperature-dependent contact resistance model described above has demonstrated that the variations of the specific contact resistance values can make a significant difference in self-heating and the resulting current distribution at the

silicon/silicide contacts under ESD conditions [103].

As shown in the preceded example, in order to predict the behavior of ESD protection devices under high temperature and high current conditions, it is essential to adequately model the temperature-dependence of critical electric properties. Especially, avalanche and thermal generation currents at extremely high temperatures should be correctly modeled, since the second breakdown under ESD conditions involves a large amount of heat generation. Hence, the coupling between the temperature and the current densities, mobility, impact ionization coefficients are all important in the simulation study of ESD phenomena. From now on, the basic electric and thermal equations for ESD simulations will be described and the coupling between them will be discussed in detail.

First, the current flow equations need to be expressed with thermal correction terms. For example, the conventional drift-diffusion current flow equations[24] [108] can be modified with a thermal correction term as follows [18][109][110]

$$\overrightarrow{J_n} = qn\mu_n \cdot \vec{E} + qD_n \vec{\nabla} n + qnD_n^T \cdot \vec{\nabla} T \,, \qquad \text{Eq. 6.1)}$$

$$\overrightarrow{J_p} = qn\mu_p \cdot \vec{E} - qD_p \vec{\nabla} p - qnD_p^T \cdot \vec{\nabla} T \qquad \text{Eq. 6.2)}$$

$\overrightarrow{J_n}$ ($\overrightarrow{J_p}$) is electron (hole) current density and μ_n (μ_p) is electron (hole) mobility. $\vec{E}$ is the electric field. The temperature dependence of the carrier mobility has been implemented using various empirical equations [111]-[113]. D_n and D_p are the diffusion constants of electrons and holes, the temperature dependency of D_n and D_p are indirectly defined using the Einstein relationship [108] to μ_n and μ_p. The 3rd term in Eqs. 6.1 and 6.2 represents the current flow due to the thermal gradients [114] with the thermal diffusion constants, D_n^T and D_p^T.

In the classical drift-diffusion transport equations, we assume that the carrier temperature is identical to the lattice temperature. However, in nano-scale devices the local carrier heating in the channel is substantial; hence, the carrier temperature can be significantly higher than the lattice temperature. In the hydrodynamic transport model[25], instead of using the thermal equilibrium approximation, we use the carrier energy balance equations to calculate the carrier temperature. In the hydrodynamic equations, the thermal diffusion terms in Eqs. 6.1 and 6.2 can be corrected using

[24] The complete drift-diffusion simulation couples the current flow equations in Eqs. 6.1 and 6.2, the continuity equations for electrons and holes, and the Poisson's equation to calculate the electrostatic potential.

[25] The hydrodynamic approach is similar to the drift-diffusion, but it solves the additional set of electron/hole energy balance equations to calculate the carrier densities, the carrier temperatures and the electrostatic potential [115].

the electron voltage ($u_n = k_B T_n / q$) and the hole voltage ($u_p = k_B T_p / q$), as follows.

$$\overrightarrow{J_n} = q\mu_n(u_n)[n\vec{E} + \overrightarrow{\nabla(u_n n)}], \quad \text{Eq. 6.3)}$$

$$\overrightarrow{J_p} = q\mu_p(u_p)[p\vec{E} - \overrightarrow{\nabla(u_p p)}] \quad \text{Eq. 6.4)}$$

T_n is electron temperature and T_p is hole temperature. The carrier velocity is controlled by the energy relaxation time rather than limited by saturation velocity; therefore the mobility model should be dependent on a carrier temperature. Eq. 6.5 is an example of carrier-temperature dependent mobility models [115].

$$\mu(T) = \frac{\mu_o}{1 + \frac{\alpha k_B (T_c - T_o)}{q}}, \quad \text{Eq. 6.5)}$$

Here, μ_0 is low-field mobility. T_c and T_o are the carrier and lattice temperatures. α is determined by the saturation velocity(v_s), low-field mobility and energy relaxation time (τ_ε), as follows.

$$\alpha = \frac{3\mu_o}{2v_s^{\,2}\tau_\varepsilon} \quad \text{Eq. 6.6)}$$

Another important parameter for ESD simulation is electron impact-ionization rate ($\alpha_{n,ii}$).

$$\alpha_{n,ii} = \alpha_{n,iio} \cdot \exp[-\frac{E_g(T)}{q\lambda_n E_{n,\|}}] \quad \text{Eq. 6.7)}$$

$\alpha_{n,iio}$ is a multiplicative factor of the electron ionization coefficient with a typical value of 7×10^5 cm^{-1} in bulk silicon and $E_{n,\|}$ is the electric field in the current flow direction. $E_g(T)$ is energy bandgap, a function of temperature. Due to the temperature dependency of the bandgap, the impact ionization rate is also strongly dependent on the temperature. λ_n is the mean free path of the electrons. The calibration of the impact ionization rate regarding λ_n will further be discussed in Section 6.2.1.

So far, we have discussed the current equations and thermal modeling of each term within the current equations. The other essential parts of electrothermal simulations are heat generation and transport models. The following equation is one of the widely used heat generation expressions [18][19][117].

$$H = \vec{J} \cdot \vec{E} + (R - G) \cdot (\varepsilon_g + 3k_B T) \quad \text{Eq. 6.8)}$$

Here, ε_g is the energy gap. Most of the generated heat can be considered to be simply from the Joule heating, $\vec{J} \cdot \vec{E}$. However, in the high temperature regime, the second term which represents the heating and cooling due to recombination and generation of charge carriers becomes significant. In the hydrodynamic system Eq. 6.8 can be replaced by the following sophisticated equation [27][118].

$$H = \frac{3}{2} k_B \frac{n(T_e - T_L)}{\tau_{e-L}} + (R - G) \cdot [\varepsilon_g + \frac{3}{2} k_B (T_e + T_L)] \quad \text{Eq. 6.9)}$$

In this expression, the heat generation is a function of the lattice temperature (T_L), the electron temperature (T_e), and the energy relaxation time between electron and lattice (τ_{e-L}). Here, it is assumed that electrons are majority carriers and holes are in thermal equilibrium with the lattice.

Eq. 6.10 shows a classical heat transport equation which basically describes Fourier law.

$$C_s \frac{\partial T}{\partial t} = \nabla \cdot (\kappa_s \nabla T) + H \quad \text{Eq. 6.10)}$$

$$\kappa_s = \frac{1}{3} C_s \bar{v} L \quad \text{Eq. 6.11)}$$

Here, C_s is the heat capacity per unit volume, and $\bar{v}$ is the average phonon velocity and L is the average phonon mean free path. In most of ESD simulation studies, Eq. 6.10 has been widely used. However, as the technology scaling goes on, the validity of this classical approach has been questioned as well. This issue will be discussed again in Section 6.3.

On the other hand, the temperature profiles under ESD conditions are strongly dependent on the thermal boundary conditions [119]. Ideally device simulations should be performed, including the contacts, via and dialectic/interconnect layers surrounding the device being analyzed. The volume included in simulations should also be large enough so that heat is dissipated into a semi-infinite medium. However, it is much more efficient, regarding the simulation time, to employ a thermal boundary resistance (R_{th}) [26] and terminate the simulation area with suitable boundary conditions.

$$R_{th} = \frac{(T - T_{ambient})}{\Lambda} \quad \text{Eq. 6.12)}$$

The temperature at the contact (T) is determined by the ambient temperature ($T_{ambient}$), R_{th}, and

[26] In the actual implementations, R_{th} can be either a lumped resistor or multiple distributed resistors over the contact area. For transient simulations, the distributed heat capacity (C_{th}) can also be used to improve accuracy in case the thermal equilibrium assumption is not valid.

the heat flux (Λ) which can concurrently be calculated by Eq. 6.10. As many new materials are introduced and the device geometries are continuously shrinking, we need better understanding and better measurements of the thermal resistance [27][120].

The completely coupled set of current flow equations and heat generation/transport equations discussed above has been successfully implemented in commercial TCAD simulators [20]. Using the circuit/device mixed mode simulations, the devices under ESD conditions have been rigorously analyzed, revealing the physics of the breakdown phenomena and facilitating the optimization of the ESD protection devices [22][24][25]. With careful calibration of electrical/thermal parameters, the set of the conventional drift-diffusion equations (Eqs. 6.1 and 6.2), the classical heat generation (Eq. 6.8) and diffusion equations (Eq. 6.10) can be used to replicate most of physical phenomena of ESD events [29][30]. Even the hydrodynamic approach is not often adopted to study the ESD devices for several reasons. First, it becomes very difficult to achieve global convergence in the computation with electron/hole current, carrier temperatures and lattice temperature, especially when the spatial and time variations of the carrier and lattice temperatures are large and the impact ionization rate is also huge, which is common under ESD conditions. On the other hand, ESD protection devices are not often implemented using minimum channel length devices. Therefore the drift-diffusion approach can still generate meaningful results for the analysis of the modern ESD protection devices and continues to be the most widely used analysis method.

However, a fundamental change has occurred as the device sizes are aggressively scaled down to nano-scale regime. In general, the hot spots near the drain region approximately tens of nanometers; therefore it is expected that the hot spots occupy more than 30 % of the channel in MOSFETs as the channel lengths are scaled down below 100 nm [122]. Phonon heat conduction from the hotpots governs the peak device temperature, which significantly affects the device conductance because the hot spots take a considerable portion of the entire channel. However, the hot spots are much smaller than the acoustic phonon mean free path which is approximately 100-300 nm in bulk Si. In this situation, ballistic emission of phonons from the hotspot can impede conduction and add a thermal resistance, leading to higher temperatures than those predicted by the classical diffusion model [123].

Moreover, a sub-continuum thermal effect is expected due to the nature of the phonon dispersion. High energy electrons scatter most effectively with the optical phonons, and transferred the energy to the optical phonons. Since the optical phonons are nearly-stationary, the energy transferred from the electrons persists in the hot region until decaying into the acoustic phonon. Hence, the heat transport is primarily governed by the acoustic phonons with relatively large group velocity. Due to the weak coupling between the optical phonons and acoustic phonons and the

difference in the group velocities, the phonon energy can be accumulated in the optical phonon branches, altering the electron-phonon scattering rate and eventually impeding the electron transport. To address the phonon dispersion issues, two separate temperatures for two groups of phonons, distinguished by their group velocities, were employed for the numerical analysis [124]. The two temperature analysis and its implication in device simulations will be explored in details in Sec. 6.3.

In heat transport simulations, the classical diffusion equations have been replaced by the phonon Boltzmann Transport Equations (BTE) [125], and the Monte Carlo method has been rigorously adopted to capture the detailed picture of electron-phonon interactions [126][127]. These approaches have achieved the detailed knowledge of the heat generation spectrum. However, this thermal analysis system still needs to be self-consistently coupled with electrical system and should be simplified for the simulations of complete transistor structures. It is expected that the improved electrothermal models based on the advanced thermal physics will lead to better understanding of ESD behavior in nano-scale regimes.

In the following section (Section 6.2), an example of ESD simulation with Stained-Si devices will be presented. In this example, the classical drift-diffusion and Fourier heat transport equations are used with carefully calibrated electrical/thermal parameters.

In the Section 6.3, to take the nano-scale heat generation and phonon heat conduction into account, the optical and acoustic phonon temperature system for TCAD simulation are demonstrated. This thermal system is self-consistently coupled with the hydrodynamic transport model, and validated through the simulations of a thin-body SOI NMOSFET.

6.2 Electro-thermal Simulations of Strained-Si Devices

Strained Si/relaxed $Si_{1-x}Ge_x$ MOSFET has been suggested as a promising device for beyond the 50 nm CMOS node. Because of its enhanced low-field electron mobility in the thin epi-Si layer, it can overcome the scaling limitations in the current driving capability. However, the low thermal conductivity of relaxed SiGe layers (~15 times lower than that of bulk-Si [128]) causes severe self-heating, substantially changing some electrical properties of a strained-Si/relaxed SiGe device in high current operation. In addition, since the thermal properties of transistors are critical for ESD and other reliability problems, studying the electro-thermal characteristics of strained-Si is of technological importance.

In this section, electrical and thermal characteristics of strained-Si are compared with bulk-Si, using device simulations with the phonon mean-free-path determined from the full-band

Monte Carlo analysis. The ESD protection capability of strained-Si devices is compared with bulk-Si devices by injecting the Human Body Model (HBM) and Machine Model (MM) ESD stress.

6.2.1 Calibration of Simulation Parameters

In Fig. 6.3, the device structures of a bulk Si NMOSFET and a strained-Si NMOSFET are depicted. Each device has gate length of 130 nm and gate oxide thickness of 2.5 nm. The strained-Si film layer is 15 nm thick and the source/drain junction depth is approximately 50 nm.

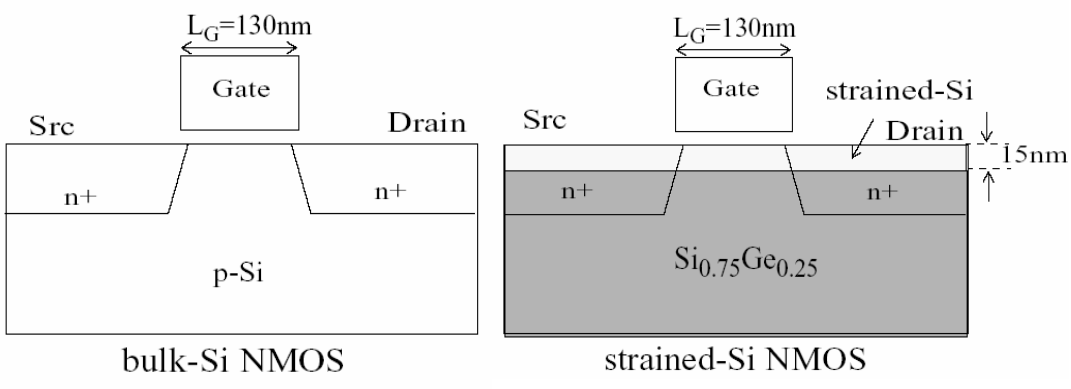


Figure 6.3: Device structures of bulk- and strained- Si NMOSFETs

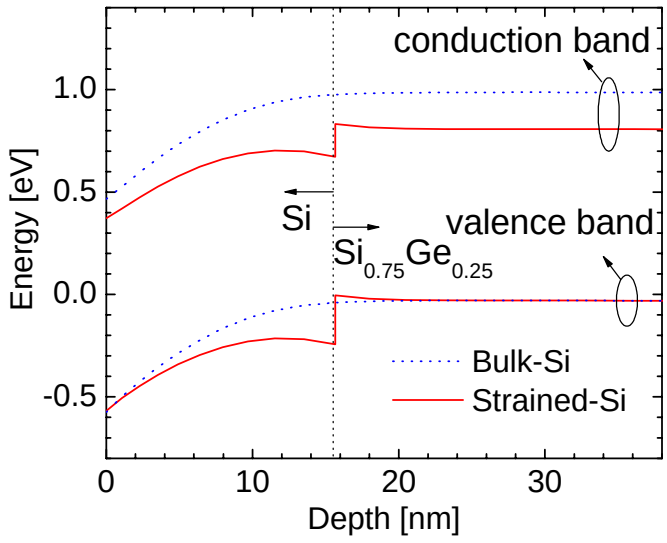


Figure 6.4: Simulated band-diagrams of strained- and bulk-Si devices along the depth direction at $V_G = 0$ V

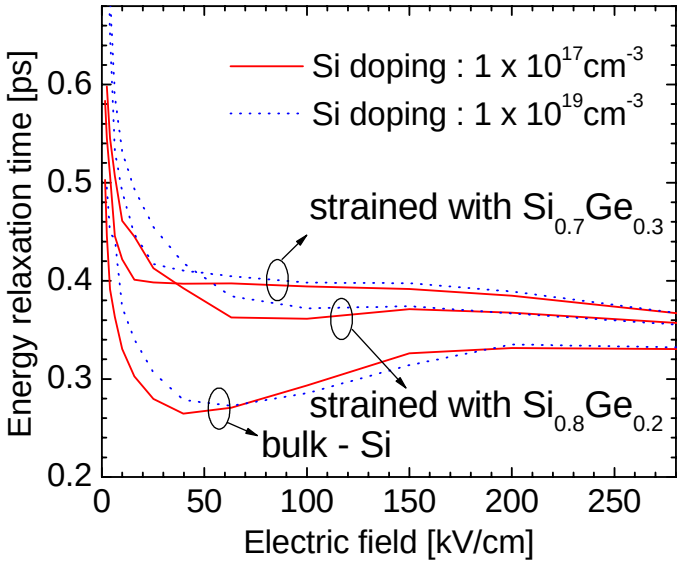


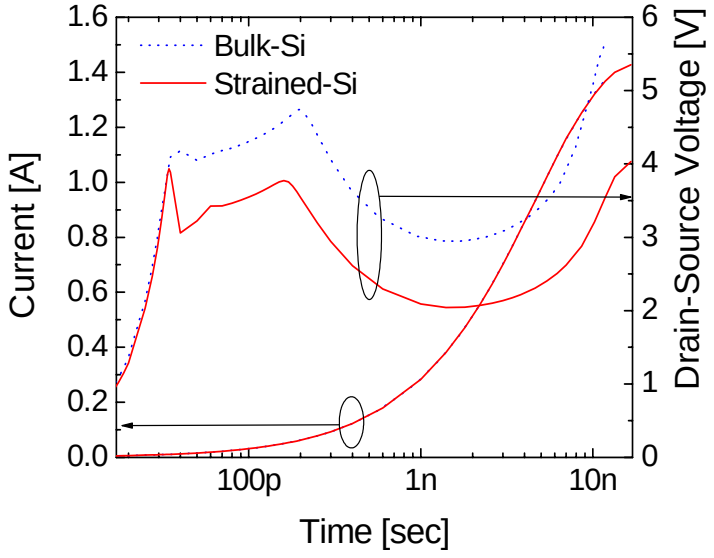
Figure 6.5: Electron energy relaxation times with respect to electric fields for strained- and bulk-Si.

The energy band structure of strained-Si devices is determined using material parameters based on Ref. [28][129]. Fig. 6.4 shows simulated energy band diagrams for strained- and bulk-Si devices [29][129]. At room temperature, the energy band gaps of strained-Si and $Si_{0.75}Ge_{0.25}$ are 0.98 and 0.90 eV, respectively, and the band gap of Si is 1.08 eV. Temperature dependent carrier mobility models and temperature dependent band gap models [20] are employed to account for thermal effects.

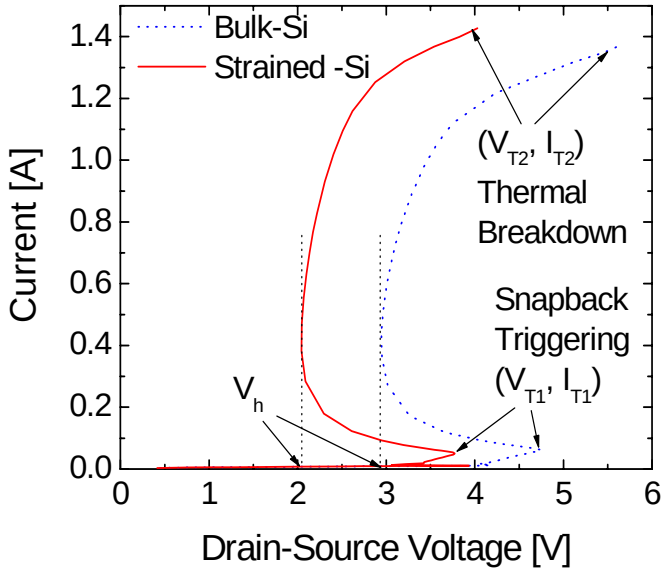
As shown in Eq. 6.7, the impact ionization rate is a function of energy bandgap and the mean free path of electrons. The mean-free-path of electrons (λ_n) is closely related to energy relaxation times in the scattering processes [131]; therefore, energy relaxation times for strained-Si were calculated using full-band Monte Carlo (FBMC) device simulations [132]. Fig. 6.5 shows the calculated energy relaxation times for electrons in strained-Si and bulk-Si with respect to the electric fields. In [28], the energy relaxation time was increased by roughly a factor of two from the bulk-Si value. However, the FBMC simulation results in Fig. 6.5 show that the increase in the energy relaxation time of the strained-Si, relative to bulk-Si, is reduced as the electric field increases. This implies that the mean-free-path (λ_n) in strained-Si for high current operation should be reduced from the value used in the lower electric field regime to investigate the electric characteristics under ESD conditions. In this study, λ_n of strained-Si on $Si_{0.75}Ge_{0.25}$ layer is assumed to increase by 50 % from that of bulk-Si; λ_n = 10 nm for bulk-Si and λ_n = 15 nm for strained-Si on $Si_{0.75}Ge_{0.25}$.

6.2.2 ESD Simulation Results and Discussions

Electro-thermal device simulations were performed for strained- and unstrained-Si (bulk) NMOSFETs based on the previously determined material parameters. Various levels of HBM, MM ESD pulses were injected into the drain of 100 μm wide NMOSFET, while the gate, source, and substrate are grounded.



(a)



(b)

Figure 6.6: (a) Simulated I_D and V_D as functions of time for strained- and bulk-Si NMOSFETs with the gate grounded. 2.4 kV HBM ESD stress is applied. (b) I_D *vs.* V_D curves for the strained- and bulk-Si NMOSFET with the gate grounded.

Fig. 6.6a shows simulated I_D and V_D curves as functions of time under 2.4 kV HBM stress. As can be seen here, the current rise time is ~10 ns and the current increases up to ~1.5 A. In Fig. 6.6b, I_D - V_D curves extracted from Fig. 6.6a are illustrated. The impact ionization rate and the current gain, β of the parasitic NPN transistor in strained-Si are higher than those of bulk-Si devices. The snapback voltage (V_{T1}) of strained-Si devices is, therefore, lower than that of bulk-Si devices and the hold voltage (V_h) - the minimum voltage required for the bipolar operation - is also lower for strained-Si devices. From an ESD perspective, a protection device with high β and low V_h in SiGe technology is advantageous for discharging the ESD stress current [133], since lower V_h generates less heat under high current conditions. When the current rises further above 0.5 A, lattice temperatures within the stressed devices significantly rise, reducing carrier mobility and increasing on-resistance. As both on-resistance and lattice temperatures increase further, the bandgap is significantly reduced, which triggers second breakdown (as indicated by V_{T2} and I_{T2} in Fig. 6.6b). The second breakdown triggering current (I_{T2}) of strained-Si devices is higher than that of bulk-Si devices by ~7 %.

ESD simulations with the HBM and MM stress reveal that strained-Si devices fail at 2.4 kV HBM and 150 V MM stress levels while bulk-Si devices fail at 2.0 kV HBM and 130 V MM. These simulation results imply that the power density and peak lattice temperatures during parasitic bipolar operations are lower for the strained-Si device than those for the bulk-Si device, even though thermal dissipation of strained-Si devices is much less compared to bulk-Si. In other words, the local overheating inside strained-Si NMOSFETs can be effectively suppressed because of the higher current gain (β) of the parasitic NPN transistors.

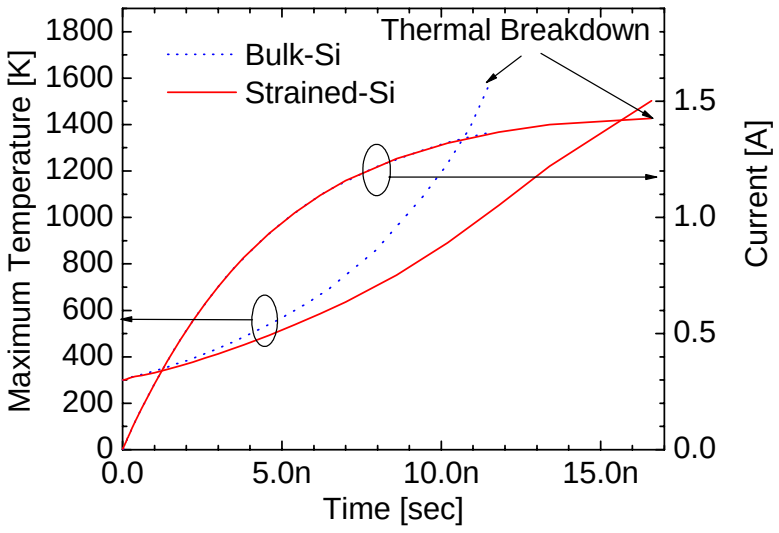


Figure 6.7: The peak temperatures within both strained-Si and bulk Si devices, as functions of time after 2.4 kV HBM ESD stress being initiated (The drain current is also plotted).

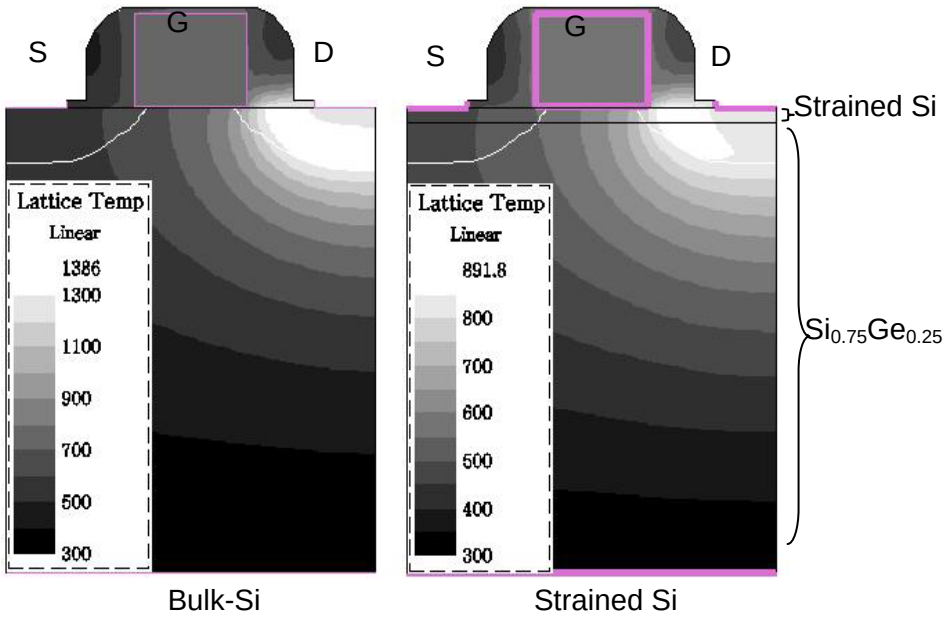


Figure 6.8: Temperature contours for bulk-Si and strained-Si devices at 10 ns after an ESD pulse is initiated.

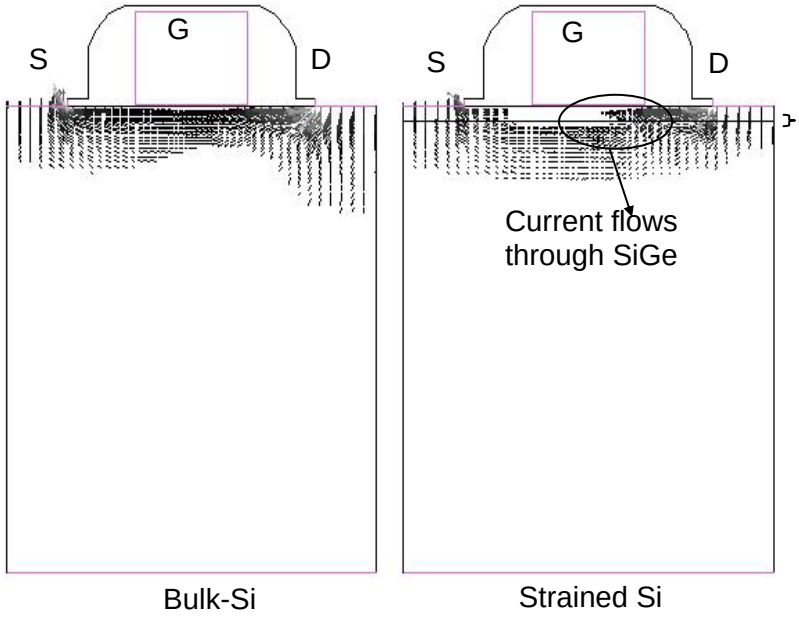


Figure 6.9: Current vector plots for bulk-Si and strained-Si devices at 10 ns after an ESD pulse is initiated.

Fig. 6.7 illustrates the peak temperatures as functions of time after 2.4 kV HBM ESD stress being initiated. The temperature in bulk-Si increases faster, while having almost the same I_D over the entire duration of the ESD event. Fig. 6.8 shows temperature contours for bulk-Si and

strained-Si devices at 10 ns when the current reaches ~1.3 A. As expected in Fig. 6.7, local heating near the drain region of bulk-Si is more severe. Hence, the peak lattice temperature of bulk-Si (~1350K) is much higher than that of strained-Si (~890 K). It is worth pointing out that in strained-Si devices under ESD conditions, current flows primarily through the SiGe layer, rather than through the surface layer as observed in Fig. 6.9. This phenomenon can suppress current localization at the Si surface, suggesting that once the parasitic bipolar transistor is turned on, uniform conduction can be achieved in strained-Si devices due to the high β of parasitic NPN transistors [133][134].

6.3 Electro-thermal Simulations With Optical and Acoustic Phonon Dispersion

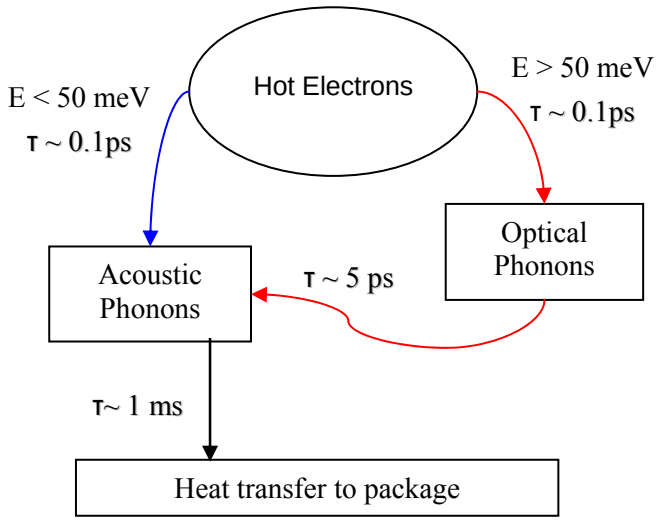


Figure 6.10: Energy transfer process diagram in silicon [122]. Electrons with high energy (> 50 meV) effectively interact with optical phonons (τ ~ 0.1 ps), and then optical phonons decay into the acoustic modes (τ ~ 5 ps).

As discussed in Section 6.1, heat generation and transport in the small *hotspot* involve the coupling of electrons and various modes of phonons. As illustrated in Fig. 6.10, high energy electrons (E > ~50 meV) scatter most effectively with optical phonons, however these phonons are relatively stationary. Therefore heat transport is mostly governed by acoustic phonons with large group velocity. The energy relaxation time between hot electrons and optical phonons is ~0.1 ps, while the relaxation time between optical phonons and acoustic phonons is ~5 ps. This

relaxation time difference results in the energy accumulation in optical phonons; the accumulated energy is highly localized due to the small group velocity of optical phonons. Heat generation and conduction mechanisms with the multi-mode phonon dispersion have been thoroughly analyzed via Monte Carlo method [27]. However, this thermal analysis system still needs to be self-consistently coupled with electrical system and should be simplified for the simulations of complete transistor structures.

Another critical thermal phenomenon is the phonon boundary scattering at interfaces. The phonon boundary scattering seriously impedes heat transport in thin semiconductor devices such as ultra thin body SOI devices, FinFETs, and nanowires [135]. In order to take into account this phenomenon, the thermal conductivity of thin silicon films was carefully measured [136], and Tornblad. et al. [98] proposed an anisotropic thermal conductance model extracted from a linearized form of the Boltzmann transport equation.

In order to investigate the impact of the nanoscale thermal phenomena described above, a thermal system with optical and acoustic phonon temperatures was implemented, using the PDE based device simulator *PROPHET* [137]. This thermal system is concurrently coupled to the hydrodynamic carrier transport model. The details of the hydrodynamic model implemented in *PROPHET* can be found in [115]. The anisotropic thermal conductance model is also employed to investigate the impact of phonon boundary scattering in thin film devices. To verify this electrothermal simulation system, device simulations with SOI NMOSFET devices with 40 nm gate length and 20 nm SOI thickness are performed.

6.3.1 Governing Equations

As discussed in Section 6.1, the classical heat diffusion equation (Eq. 6.10) cannot describe the heat transport mechanism on short length scales. At length scales of 10 ~ 100 nm[27], the phonon Boltzmann Transport Equation (BTE) may replace the classical heat diffusion equations, considering that phonons to be semi-classic particles. Furthermore, with relaxation time approximation, the phonon BTE can be simplified to the form of Eq. 6.13 [27][125].

$$\frac{\partial u}{\partial t} = \frac{u_o - u}{\tau_{ph}} - v \cdot \nabla u + H \qquad \text{Eq. 6.13)}$$

[27] Shorter than the acoustic phonon mean free path (a few hundred nanometers) and longer than the phonon wavelength (a few nanometers)

Here, u is the phonon energy density. τ_{ph} is the average phonon scattering time, and v is the phonon velocity.

The thermal equations implemented in this study are conceptually similar to Eq. 6.13. However the formulations are closer to the classical heat diffusion equation with the lattice temperature, the heat capacity, and the thermal conductivity. At first, the heat transport equation can be written with the carrier temperature (T_c) and lattice temperature (T_l) using the energy relaxation time approximation.

$$C_l \frac{\partial T_l}{\partial t} = \frac{3}{2} n \cdot k_B \cdot \frac{T_c - T_l}{\tau_{c-l}} + \nabla \cdot (\kappa \nabla T_l) \qquad \text{Eq. 6.14)}$$

Where, C_l is the heat capacity of the lattice. k_B is the Boltzmann constant and κ is the lattice thermal conductivity. The first term in the right side is the energy transfer from the carriers (electrons or holes) to the lattice and $\tau_{c\text{-}l}$ is the carrier energy relaxation time. With this equation, the hydrodynamic carrier transport model can be coupled to the heat transport models. However, Eq. 6.14 still cannot deal with the multi-mode energy transfer due to the phonon dispersion illustrated in Fig. 6.10.

In this study, the equations for energy transfer between carriers and phonons are implemented with two phonon temperatures (T_o for optical phonons and T_a for acoustic phonons) as follows,

$$C_o \frac{\partial T_o}{\partial t} = H + \nabla \cdot (\kappa_o \nabla T_o) - C_o \frac{T_o - T_a}{\tau_o} \qquad \text{Eq. 6.15)}$$

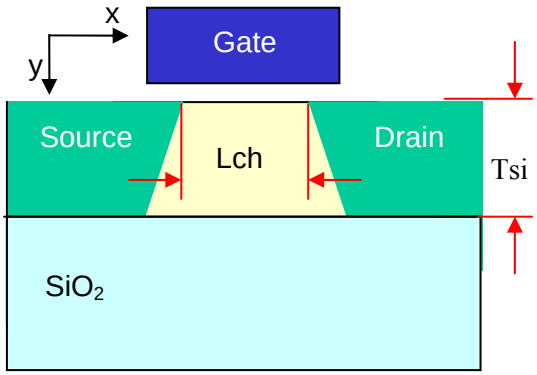
$$C_a \frac{\partial T_a}{\partial t} = \nabla \cdot (\kappa_a \nabla T_a) + C_o \frac{T_o - T_a}{\tau_o} \qquad \text{Eq. 6.16)}$$

Where, *Co (Ca)* is the heat capacity of the optical (acoustic) phonons. τ_o is the relaxation time for optical phonons (~5.0 ps). κ_a is the lattice thermal conductivity relevant to the propagating velocity of acoustic phonons, and κ_o is an adjustable parameter for the heat conduction through optical phonons. κ_o is much smaller than κ_a, since optical phonons are almost stationary. H is the energy transfer from electrons (or holes) to the optical phonons. H is almost same as the first term in the right side of Eq. 6.13, but it should be rewritten in terms of T_o.

$$H = \frac{3}{2} n \cdot k_B \cdot \frac{T_c - T_o}{\tau} \qquad \text{Eq. 6.17)}$$

This term can be replaced by a simple Joule heating, $\vec{J} \cdot \vec{E}$ in simulations with the drift-diffusion models.

Two-dimensional device simulations are performed with various combinations of the electrical and thermal models: 1) the hydrodynamic model with Eqs. 6.15 - 6.17 ('two phonon temperature model'), 2) hydrodynamic model with Eq. 6.14 ('one lattice temperature model'), and 3) drift-diffusion model with Eqs. 6.15 and 6.16, and $\vec{J} \cdot \vec{E}$ for the heat generation.



Gate length (Lch)	40 nm
Gate oxide thickness (Tox)	2 nm
SOI thickness (Tsi)	20 nm
BOX thickness	200 nm

Figure 6.11: The geometric information of the simulated thin-body SOI NMOSFET.

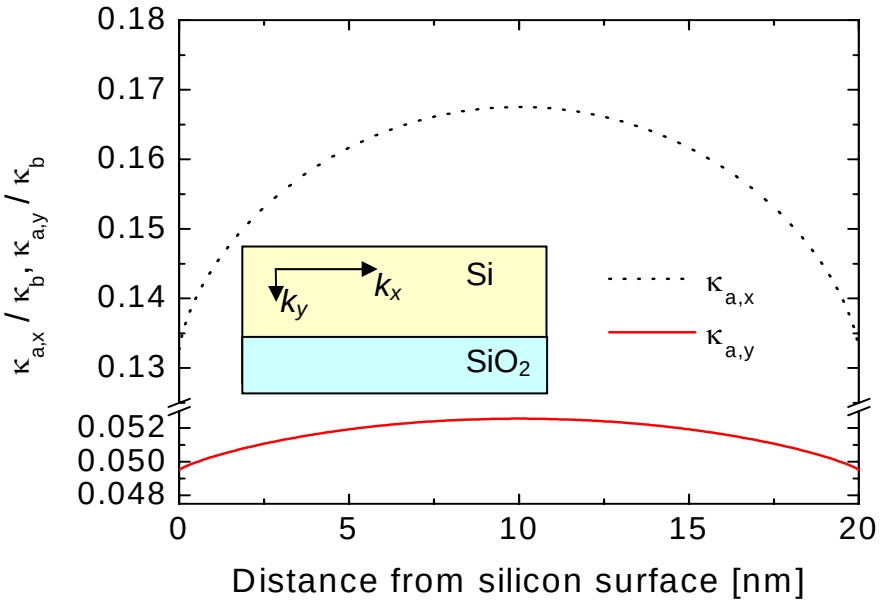


Figure 6.12: The anisotropic thermal conductivity in the 20 nm silicon film of the SOI NMOSFET in Fig. 6.11, normalized by the thermal conductivity at bulk silicon.

In order to investigate the impact of the optical and acoustic phonon systems, a thin body SOI MOSFET based on [138] is employed for the simulations. The geometric information of the device is listed in Fig. 6.11.

In Eq. 6.16, κ_a is modeled as an anisotropic thermal conductivity as follows [98],

$$\kappa_{a,x}\,(or\ \kappa_{a,y}) = \kappa_b \cdot \left(1 - \frac{1}{2}e^{-\left(\frac{y}{a\cdot\Lambda}\right)^b} - \frac{1}{2}e^{-\left(\frac{D-y}{a\cdot\Lambda}\right)^b}\right) \qquad \text{Eq. 6.18)}$$

Here, Λ is the mean-free-path of acoustic phonons. D is the thickness of films. κ_b is the thermal conductivity at bulk silicon. (a, b) is (0.35, 0.75) for $\kappa_{a,x}$, and (0.72, 0.95) for $\kappa_{a,y}$. y is the position with reference to one edge of the thin silicon film. As shown in Fig. 6.12, the calculated anisotropic thermal conductivity in the 20 nm silicon film of the SOI NMOSFET is reduced to 5 - 17 % of that of bulk silicon.

In order to predict electrical characteristics at high temperature, the temperature dependency of the low-field mobility and the electron saturation velocity models are modified according to [139][140]. The mobility model strongly depends on the carrier temperature. For setting proper thermal boundary conditions, source, drain, gate and substrate contacts are terminated with the finite thermal resistance as illustrated in Fig. 6.13.

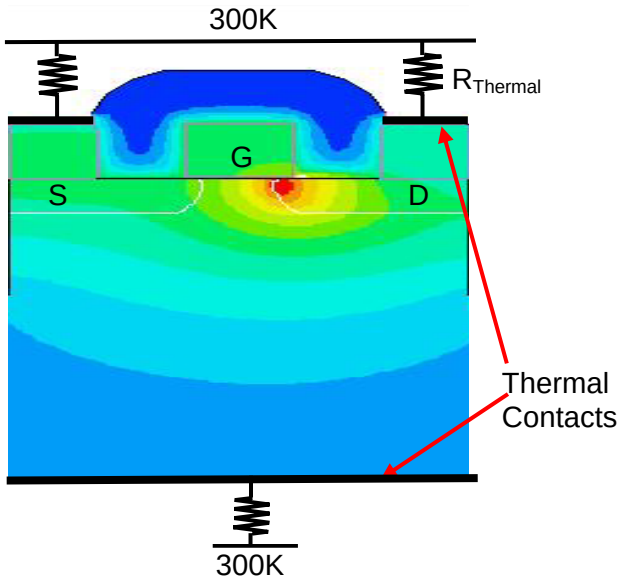


Figure 6.13: The thermal boundary conditions for the electro-thermal simulations.

6.3.2 Simulation Results and Discussions

Fig. 6.14 shows the electron energy along the channel at the silicon/gate oxide interface. The gate bias is set to 1 V, and the drain bias varies from 0.2 V to 1.0 V with 0.2 V steps. At drain bias higher than 0.4 V, ~75 % of the channel is occupied with electrons of > 50 meV, which means the energy relaxation from electrons to phonons is dominated by the interaction between electrons and optical phonons.

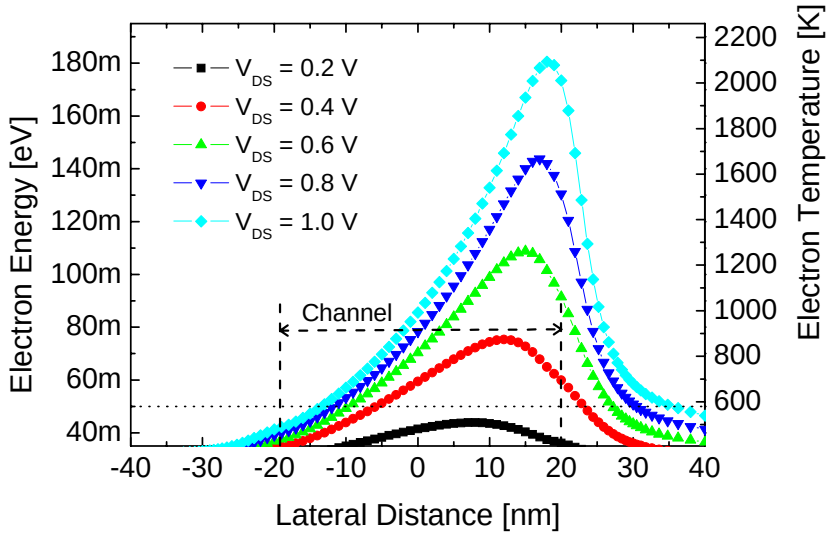


Figure 6.14: The electron energy along the channel at the silicon/gate oxide interface. V_{GS} is 1.0 V, and V_{DS} is 0.2 V~1.0 V. The physical channel is also indicated.

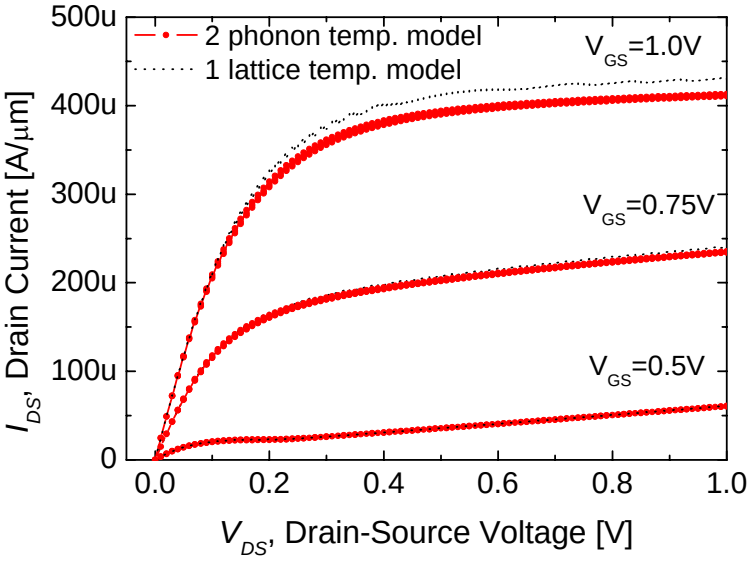


Figure 6.15: Simulated I_D vs. V_{DS} curves with various gate biases

To compare with 'the two phonon temperature model,' the simulation using just 'one lattice temperature model' (the hydrodynamic model with Eq. 6.14) is also performed. Fig. 6.16 shows the simulated *Id-Vd* curves. With V_{GS} of 0.5 V and 0.75 V, there is no noticeable change in *I-V* characteristics between two models. However, in the high current and voltage regime, we can see a reduction of the current driving capability. Drain current decreases by ~5 % at (V_{GS}, V_{DS}) = (1V, 1V). This reduction can be explained by the temperature profiles shown in Figs. 6.16–18.

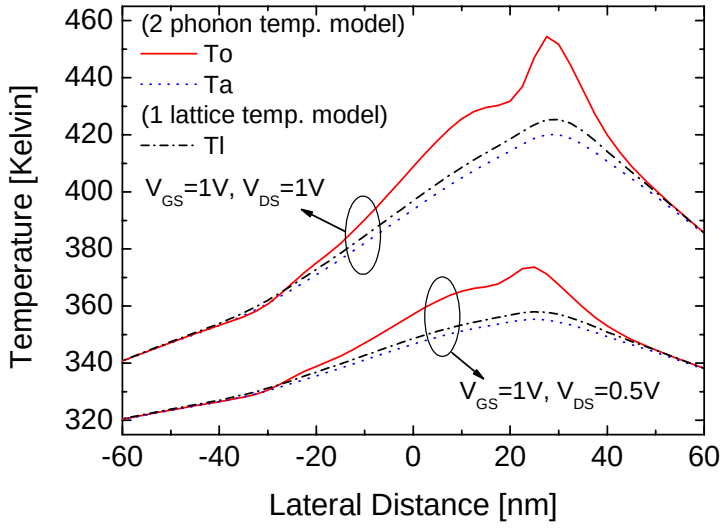


Figure 6.16: Optical (*To*) and acoustic (*Ta*) phonon temperature along the channel at the silicon/gate oxide interface. Lattice temperature (T_l) extracted from the simulation with 'one temperature model' is also plotted.

Fig. 6.16 shows the optical and acoustic phonon temperature profiles along the channel at the silicon/gate oxide interface. The lattice temperature profile extracted from the simulation with the 'one lattice temperature model' is also plotted. Because the energy relaxation time between hot electrons and optical phonons is much shorter than the relaxation time between optical phonons and acoustic phonons, the energy is accumulated in optical phonons. So, the optical phonon temperature is increased considerably over the acoustic phonon temperature throughout almost half of the channel, and has its peak inside the drain region. When V_{GS} and V_{DS} are 1 V, the peak of the optical phonon temperature is ~455 K, and the peak of the acoustic phonon temperature is ~415 K. The lattice temperature profile extracted from simulations with 'one lattice temperature model' is almost the same as the acoustic temperature profile. Due to the higher optical phonon temperature and increased rate of scattering, the electron mobility (and saturation velocity) is degraded, resulting in the reduced drain current in Fig. 6.15.

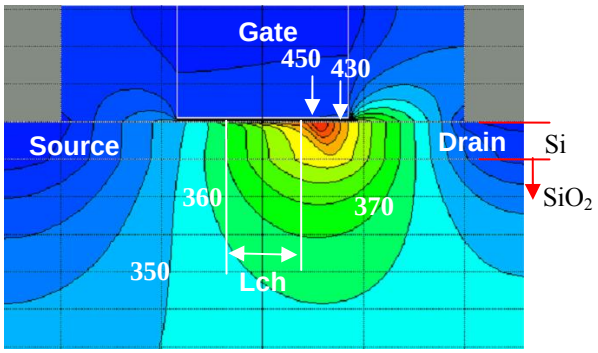


Figure 6.17: Two-dimensional contour plot of the optical phonon temperature. The interval between contours is 10 Kelvin. $V_{GS} = V_{DS} = 1$ V and $I_{DS} = \sim 0.4$ mA/μm.

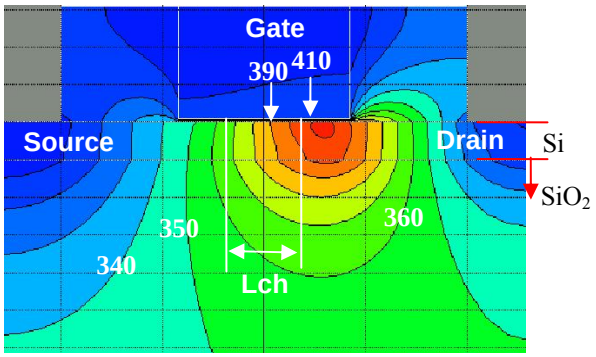


Figure 6.18: Two-dimensional contour plot of the acoustic phonon temperature. The interval between contours is 10 Kelvin. $V_{GS} = V_{DS} = 1$ V and $I_{DS} = \sim 0.4$ mA/μm.

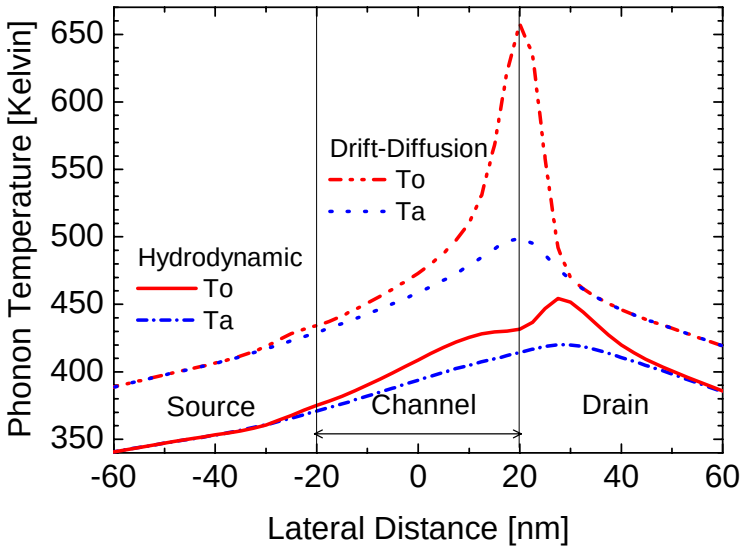


Figure 6.19: Comparison between the drift-diffusion models and hydrodynamic models. The heat generation in the drift-diffusion is implemented as $\vec{J} \cdot \vec{E}$.

Figs. 6.17 and 6.18 are two-dimensional contour plots for the optical and acoustic phonon temperatures. In each plot, the interval between contours is 10 Kelvin. It is clearly shown that the generated heat is confined mostly inside the silicon film since the thermal conductivity is dramatically reduced due to the phonon boundary scattering. The hotspot of the optical phonons is severely localized because of the low group velocity of the optical phonons, while the generated heat is transferred to the contacts mainly by the acoustic phonons.

Fig. 6.19 shows the comparison between the drift-diffusion models and hydrodynamic models. In the drift-diffusion models, the heat generation is dominated by $\vec{J} \cdot \vec{E}$, while in the hydrodynamic models it is due to the electron-phonon collision term as expressed in Eq. 6.17. A noticeable difference is the location where the temperature profile has its peak. The peak temperature with the heat generation term of $\vec{J} \cdot \vec{E}$, is located at the edge of the channel, where the electric field has its peak. However, the hydrodynamic model predicts that the peak temperature is shifted deep into the drain region. The result from the hydrodynamic model is more physical, because electrons travel a few mean-free-paths after gaining energy in high electric field region, and then release the energy to the lattice inside the drain region. It should be also pointed out that the heat generation by $\vec{J} \cdot \vec{E}$ is also more severely localized; therefore the peak temperature of the hotspot is higher in the drift-diffusion simulation results. These discrepancies between the drift-diffusion and hydrodynamic models are also demonstrated and verified by Monte Carlo simulations [27].

6.4 Summary

The semiconductor research community is pursuing transistor (FinFET, SOI, SiGe, thin oxide) and circuit geometries (3D-IC) that will dramatically amplify on-chip temperature differences, particularly for ESD. Hotspot issues in terms of microscopic heat generation and transport will be aggravated at the heating intensities common for ESD/EOS. However more work is needed to resolve the detailed electron and phonon transport phenomena under these conditions. There is an urgent need for experimental results including thermo-physical properties and verification of modeling.

Using a commercial TCAD tool with careful calibration of the electrothermal parameters, it has been demonstrated that strained-Si/SiGe NMOS devices show superior ESD protection capability compared to bulk-Si devices. Strained-Si devices have higher bipolar current gain and

impact ionization rates, compared to bulk-Si devices, due to narrower energy band gap and longer phonon mean-free-path. Due to the better performance of parasitic NPN transistors, the local overheating can effectively be suppressed in strained-Si devices so that the second breakdown triggering current for strained-Si devices is higher than that of bulk-Si devices by ~7 %. Furthermore, strained-Si devices are also expected to exhibit better current uniformity, since the main current path during the parasitic bipolar action is through the buried SiGe layer. Contrary to the case of SOI devices, which have similar self-heating and heat confinement problems, strained-Si NMOS devices can effectively be used for ESD protection devices.

The optical and acoustic temperature systems are fully coupled to the hydrodynamic model. This system can capture most physical phenomena of the heat generation and conduction inside nano-scale devices. Due to the relatively long energy relaxation time between optical phonons and acoustic phonons, the optical phonon temperature is increased considerably so that the electrical characteristics are noticeably affected. The extracted temperature profiles show qualitatively good agreement with the already published Monte Carlo simulation results.

Chapter 7

CONCLUSIONS

As the scaling of the CMOS technology continues, Electrostatic Discharge has become one of the most critical reliability issues. Therefore, there have been many efforts to improve the ESD protection capability and achieve highly reliable IC products against ESD threats. However, as the circuit performance also has been improved and the bandwidth of circuit interfaces has reached into the Gigahertz regime, problems of ESD-to-circuit interactions have emerged as a major challenge. Due to the demand of compact I/O design in highly integrated ASICs, it is also required to be cautious about the coupling between ESD protection devices and the desired circuit functions, and providing proper design rules to avoid inadvertent ESD failures. On the other hand, as the physical length of the CMOS devices enter sub-100nm regime, the validity of conventional TCAD simulations and the underlying electrothermal physics has been questioned.

7.1 Contributions

This dissertation has addressed the issues described above, both in breadth and depth, providing the guidelines to develop robust and transparent ESD protection circuits in advanced CMOS technologies.

The challenges associated with ESD-to-circuit interactions are most notable in two distinct application areas: Gigahertz RF circuits and Gb/sec digital communication systems. In this

dissertation, the impact of ESD parasitic elements on the performance of RF ICs has been analyzed with a simplified RF model of ESD protection devices. It is inferred that the signal reflection and power loss can be extremely sensitive to the size of ESD protection devices at multi-GHz frequencies. Many of studies to overcome this ESD-to-circuit impact, including optimization of the conventional ESD protection devices, co-design methods (ESD cancellation/isolation and several broadband techniques) have been reviewed. It is pointed out that at relatively lower frequencies, the low-C design methods can be effectively adopted since it is simple and transparent to various RF designs. However, for extremely high frequency applications only co-design methodologies can provide the needed ESD protection capabilities without substantial degradation of RF performance. For today's medium frequency range around 5 GHz, the trade-off between the low-C protection and rigorous co-design schemes must be considered. To verify the co-design method, a 2 GHz narrowband LNA for UMTS application is implemented. It is demonstrated that for this specific application the simple co-design approach is possible without aggressively reducing the size of the ESD protection devices. That is, the dual diodes with 300 fF and 560 fF successfully sustained > 2 kV HBM stress, showing negligible degradation of the RF performance. A 5 GHz LNA is also developed to investigate the limitation of the low-C protection scheme. A transient triggered (TT) SCR circuit is employed for ESD protection for this test bench. It is shown that TT-SCR can be a good candidate for the ESD protection of ~5 GHz applications. Based on these case studies, it is concluded that the advanced broad band techniques are expected to take an essential role for the applications of operation frequencies higher than 5 GHz.

Another important ESD-to-circuit interaction problem is analyzed; the large signal distortion caused by ESD protection circuits is estimated, using a variety of methods. The theoretical and experimental results indicate that the distortion from ESD protection circuits is typically more than 10 dB below the distortion introduced by state-of-the-art ADCs themselves. However, it is also suggested that with *SFDR* targets approaching 100 dB at frequencies near 100 MHz, the protection device can easily become a performance-limiting factor in the future.

From this discussion, it follows that there are two basic options for mitigating signal distortion from ESD devices. First, distortion can be suppressed by minimizing the ESD device size. Reducing the size by one-half improves the harmonic distortion performance by 6 dB. Second, instead of reducing the protection device size, which avoids sacrificing ESD robustness, one can attempt to make the effective ESD capacitance nearly linear by combining various types of capacitive elements. This study gives a guideline to select and implement ESD protection

devices in ADCs. The same analysis can be applied to other high frequency, large signal application such as RF power amplifiers.

In the context of developing design rules, a new failure phenomenon of PMOSFET pull-up devices under ESD conditions has been reported and analyzed by investigating various I/O structures of 3.3 V, 0.13 μm CMOS technology. The physical mechanism and the influence of layout parameters such as the distance between a negative strike diode and an n+ guard ring (DR1), guard ring shape, and effective resistance of the rail clamp, were investigated through Transmission Line Pulse (TLP) testing and device/circuit mixed mode simulations.

It is revealed that the localized turn-on of the parasitic PNP BJT of the PMOSFET pull-up device is caused by the localized charge injection into the body of PMOSFET from the negative strike diode. The most critical layout parameters are indicated and safe design rule constraints are suggested. This phenomenon can be implicated in ESD failures of all future ASIC I/O circuits which require more compact layouts. This type of modeling approach to establish ESD layout guidelines for technologies below the 90 nm node will increase in importance.

Using commercial TCAD tools with careful calibration of the electrothermal parameters, the ESD protection capability of the strained-Si/SiGe NMOS devices is investigated. Strained-Si devices have higher bipolar current gain and impact ionization rates, compared to bulk-Si devices, due to narrower energy band gap and longer phonon mean-free-path. Due to the better performance of parasitic NPN transistors, the local overheating can effectively be suppressed in strained-Si devices so that the second breakdown triggering current for strained-Si devices is higher than that of bulk-Si devices by ~7 %. Furthermore, strained-Si devices are also expected to exhibit better current uniformity, since the main current path during the parasitic bipolar action is through the buried SiGe layer. It is concluded that strained-Si/SiGe NMOS devices show superior ESD protection capability compared to bulk-Si devices

Finally, seeking a valid device simulation method for nano-scale devices with moderate numerical complexity, a new electrothermal model based on the advanced thermal physics is developed and implemented in the research simulation tool *PROPHET*. The optical and acoustic temperature systems are fully coupled to the hydrodynamic model. This system can capture most physical phenomena of the heat generation and conduction inside nano-scale devices. Due to the relatively long energy relaxation time between optical phonons and acoustic phonons, the optical phonon temperature is considerably increased so that the electrical characteristics are noticeably

affected. The extracted temperature profiles show qualitatively good agreement with the already published Monte Carlo simulation results. It is expected that the concept of this study will be implicated in the ESD simulation studies.

7.2 Suggested Future Work

Over the past few years, ESD issues have been regarded to be primarily the concern of the device engineers. However, more circuit level engineering is applied recently in addressing the ESD problems. Especially the protection of high speed chips such as > 5 GHz RF IC and 10 Gbps high speed links, is gaining more importance. As reviewed in Section 3.1, there have been several suggestions and borrowings of ideas from the microwave engineering community. Some of these ideas have already been employed in the industry, but some solutions might be lacking in practicality from an implementation point of view. Certainly, there is an urgent need for a protection circuit design methodology to protect these high-speed circuits, and there are also plenty of opportunities for further improvements and innovations.

Recently, there was a strong return of the rail-based ESD protection scheme which commonly employs dual-diodes and MOSFET power clamps between power rails. This movement is initiated since the device engineering of the pad-based ESD protection has been too challenging and hard to meet the time-to-market demand in the industry. As this protection scheme becomes one of the prevalent ESD solutions, more research efforts need to be focused on this methodology. The power clamps based on MOSFETs consume large areas, since the thin gate oxide devices usually are not allowed between the power rails. More efficient power clamp design and alternative device solutions are sorely needed from a circuit perspective. In addition, there are several issues that need to be given greater attention: reliable triggering of the power clamps and low leakage under normal operations, optimization of the diode structures and the power grid between the power clamps, accurate parasitic resistance extraction of the power grids etc.

Even though CDM stress has emerged as one of the biggest ESD threats, understanding of the nature of CDM is relatively poor. CDM failures have given many surprises, because the fast CDM stresses can easily penetrate into the internal circuits through capacitive and inductive coupling. In other words, CDM failures are not confined only within I/O circuits. A simulation methodology, which is able to cover protected internal circuits and indicate the circuits vulnerable to CDM stresses, should be developed. Regarding its high *dI/dt* nature, extraction of parasitic

inductance of the bus wires and investigation of its impact on ESD protection will be another interesting topic.

Due to the high integration in future generation ICs, the pad counts in a chip will be rapidly increasing. Accordingly, the portion of area that ESD protection circuits occupy is expected to increase; therefore, compact design of ESD protection devices is needed to alleviate this area consumption problem. This issue will be critical in 3D IC which may have a substantial number of I/Os on the top level. Instead of giving too conservative ESD design rules, tighter but safe design rules should be provided, which will possibly be achieved with precise prediction of ESD current paths and complete analysis of ESD failures.

As new transistors types (FinFET, SOI, SiGe, CNT), and circuit geometries (3D-IC) have emerged as possible alternatives to overcome the scaling limit of the conventional bulk-Si process, we need to address many new thermal issues which might be further aggravated under ESD conditions. There is a need for experimental results including thermo-physical properties and verification of modeling. Currently, there are not many reported ESD measurement. ESD measurement, simulation study, or numerical analysis is required.

The importance of thermal effects in terms of microscopic heat generation and transport has significant implications on ESD behavior since ESD is a phenomenon involving high temperature operation. However more work is needed to resolve the detailed electron and phonon transport phenomena under these conditions. In this dissertation, a simulation system with the optical and acoustic phonon temperatures is implemented to understand thermal effect in deep submicron technology. However, the coefficients in the implemented system should be tuned, reflecting the results from Monte Carlo simulations, and the validity of the equations should be reviewed and modified if necessary. For instance, in the current implementation, it has been assumed that electrons first interact with the optical phonons and transfer the energy only to those phonons, but some simulation results based on Monte Carlo simulations suggest that a certain portion of electron energy is transferred directly to acoustic phonons. These and other physically based models need to be further studied in the context of ESD applications.

Bibliography

[1] A. Amerasekera and C. Duvvury, *ESD in Silicon Integrated Circuits, 2nd edition*: Wiley, 2002.

[2] S. Dabral and T. J. Maloney, *Basic ESD and I/O Design*: Wiley, 1998.

[3] K. Banerjee, A. Amerasekera, G. Dixit, N. Cheung, and C. Hu, "Characterization of Contact and Via Failure under Short Duration High Pulsed Current Stress," in *Proceedings of International Reliability Physics Symposium*, pp. 216–220, 1996.

[4] M. D. Chaine, C. T. Liong, and H. F. San, "A Correlation Study Between Different Types of CDM Testers and Real Manufacturing In-Line Leakage Failures," *IEEE Trans. Components, Packaging, and Manufacturing Tech. Part A.* Vol. 18, No. 2, pp. 295–302, June, 1995.

[5] O. J. McAteer, R. E. Twist and R. C. Walker, "Latent ESD failures," in *Proceedings of the 4th EOS/ESD Symposium*, pp. 41-48, 1982.

[6] T. Green, "A review of EOS/ESD field failures in military equipment," in *Proceedings of the EOS/ESD Symposium*, pp. 7-14, 1988.

[7] R. Merrill and E. Issaq, "ESD design methodology," in *Proceedings of the EOS/ESD Symposium*, pp. 233-237, 1993.

[8] G. Wagner, J. M. Soden, and C. F. Hawkins, "Extent and cost of EOS/ESD damages in an IC manufacturing process," in *Proceedings of the EOS/ESD Symposium*, pp. 49-55, 1993.

[9] S. Beebe, "Characterization, Modeling and Design of ESD Protection Circuits," *Ph.D. dissertation*, Stanford University, 1998.

[10] S. Voldman, *ESD Physics and Devices*: John Wiley & Sons, Ltd., 2004.

[11] HBM-ESD Sensitivity Testing: Human Body Model (HBM)-Component Level, ESD Association Standard, S5.1, 1993.

[12] MM-ESD Sensitivity Testing: Machine Model (MM)-Component Level, ESD Association Standard, S5.2, 1994.

[13] CDM-ESD Sensitivity Testing: Charged Device Model (CDM)-Component Level, ESD Association Standard, D5.3, 1997.

[14] ESD Association WG 5.1, "ESD STM5.1-2001: Standard Test Method for Electrostatic discharge Sensitivity Testing- Human Body Model (HBM) Component Level," 2001.

[15] ESD Association WG 5.2, "ESD STM5.2-1998: Standard Test Method for Electrostatic discharge Sensitivity Testing- Machine Model (MM) Component Level," 1998.

[16] ESD Association, "ESD STM5.3.1-1999: Standard Test Method for Electrostatic discharge Sensitivity Testing- Charged Device Model (CDM) Component Level," 1999.

[17] TLP-ESD Sensitivity Testing: Transmission Line Pulse (TLP)-Component Level, ESD Association Standard Practice SP5.5, 2004.

[18] G. Wachutka, "Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling," *IEEE Trans. Computer Aided Design*, CAD-9, 1141–1149, 1990.

[19] U. Lindefelt, "Heat generation in semiconductor devices," *J. Applied Physics*, Vol. 75, pp. 942-957, 1994.

[20] *MEDICI : Two-Dimensional Device Simulation Program*, Synopsys Inc., 2002.

[21] W. Fichtner, "TCAD-based ESD on-chip protection design," in *Tech. Dig. IEDM*, 2001.

[22] K.-H. Oh, C. Duvvury, K. Banerjee, and R. W. Dutton, "Analysis of non-uniform ESD current distribution in deep submicron NMOS transistors," *IEEE Trans. Electron Devices*, Vol. 49, No. 10, 2002

[23] K. Esmark, *Device simulation of ESD protection elements*, Series of Microelectronics, Vol. 128, Hartung-Gorr Verlag Konstanz, 2002.

[24] K.-H. Oh, C. Duvvury, K. Banerjee, and R. W. Dutton, "Analysis of gate-bias-induced heating effects in deep-submicron ESD protection designs," *IEEE Trans. Device and Materials Reliability,* Vol. 2, No. 2, pp. 36 - 42, 2002.

[25] K.-H. Oh, C. Duvvury, K. Banerjee, and R. W. Dutton, "Impact of gate-to-contact spacing on ESD performance of salicided deep submicron NMOS transistors," *IEEE Trans. Electron Devices*, Vol. 49, No. 10, 2002

[26] M. Lundstrom, *Fundamentals of carrier transport,* 2nd edition: Cambridge University Press, 2000.

[27] E. Pop, "Self-Heating and Scaling of Thin Body Transistors," *Ph.D. dissertation*, Stanford University, Dec. 2004.

[28] K. Rim, J. L. Hoyt, and J. F. Gibbons, "Fabrication and analysis of deep submicron strained-Si N-MOSFETs," *IEEE Trans. Electron Dev*., Vol. 47, pp.1406-1415, 2000.

[29] C. Choi, J.-H. Chun, and R. W. Dutton, "Electro-thermal characteristics of strained-Si MOSFETs in high-current operation," *IEEE Trans. Electron Dev.* Vol. 51, 2004.

[30] J.-H. Chun, C. Choi, and R. W. Dutton, "Electrothermal simulations of strained-Si MOSFETs under ESD conditions," in *Proc. SISPAD*, 2004.

[31] C. Duvvury and C. Diaz, "Dynamic Gate-Coupled NMOS for Efficient Output ESD Protection," in *Proc. International Reliability Physics Symposium*, pp. 141-150, 1992

[32] C. Duvvury, S. Ramaswamy, A. Amerasekera, R. Cline, B. Andresen and V. Gupta, "Substrate Pump NMOS for ESD Protection Applications," in *Proceedings of EOS/ESD Symposium*, pp. 7-17, 1999.

[33] D. Scott, G. Giles, and J. Hall, "A lumped element model for simulation of ESD failures in silicided devices," in *Proceedings of EOS/ESD Symposium*, pp. 41–47, 1986.

[34] S. Ramaswamy, A. Amerasekera and M. Chang, "A unified substrate current model for weak and strong impact ionization in sub-0.25 micron nMOS devices," in *Tech. Dig. IEDM*, pp. 885-888, 1997.

[35] S. Ramaswamy, "Modeling Simulation and Design Guidelines for EOS/ESD Protection Circuits in CMOS Technologies," *Ph.D. dissertation*, University of Illinois at Urbana-Champaign, 1996.

[36] R. Dutton, "Bipolar transistor modeling of avalanche generation for computer circuit simulation," *IEEE Trans. Electron Dev.,* ED-22, pp. 334-338, 1975.

[37] P. Ngan, D. Oliver, T. Smedes, R. Gramacy, and C-K Wong, "A Automatic Layout Based Verification of Electrostatic Discharge Paths," in *Proceedings of EOS/ESD Symposium*, pp. 96-101, 2001.

[38] A. Wang, H. Feng, R. Zhan, H. Xie, G. Chen, Q. Wu, X. Guan, Z. Wang and C. Zhang, "A review on RF ESD protection design," *IEEE Trans. Electron Devices*, Vol. 52, No. 7, pp. 1304-1311, July 2005.

[39] M. Stockinger, J. W. Miller, M. G. Khazhinsky, C. A. Torres, J. C. Weldon, B. D. Preble, M. J. Bayer, M. Akers, and V. G. Karnat, "Boosted and Distributed Rail Clamp Networks for ESD Protection in Advanced CMOS Technologies," in *Proceedings of EOS/ESD Symposium*, 2003.

[40] S. Voldman and G. Gerosa, "Mixed-voltage-interface ESD protection circuits for advanced microprocessors in shallow trench isolation and LOCOS CMOS technology," in *Tech. Dig. IEDM*, 1994.

[41] M.-D. Ker, W.-J. Chang, C.-T. Wang, and W.-Y. Chen, "ESD Protection for Mixed-Voltage I/O in Low Voltage Thin-Oxide CMOS," in *Tech. Dig. ISSCC*, pp. 546–547, 2006.

[42] J. Lin, C. Duvvury, B. Haroun, I. Oguzman and A. Somayaji, "A Fail-Safe ESD Protection Circuit with 230 fF linear Capacitance for High-Speed/High-Precision 0.18 CMOS I/O Application," in *Tech. Digest IEDM*, pp. 102-106, 2002.

[43] J. Z. Chen, A. Amerasekera, and C. Duvvury, "Design methodology for optimizing gate driven ESD protection circuits in submicron CMOS processes," in *Proceedings of the EOS/ESD Symposium*, pp. 230-239, 1997.

[44] A. Amerasekera, C. Duvvury, V. Reddy, and M. Rodder, "Substrate triggering and salicide effects on ESD performance and protection circuit design in deep submicron CMOS processes," in *Tech. Dig. IEDM*, pp. 547-550, 1995.

[45] K.-H. Oh, "Investigation of ESD Performance in Advanced CMOS Technology," *Ph.D. dissertation*, Stanford University, 2002.

[46] S. H. Voldman, "The state of the art of electrostatic discharge protection: Physics, technology, circuits, design, simulation, and scaling," *IEEE J. Solid-State Circuits*, Vol. 34, pp. 1272-1282, Sept. 1999.

[47] K. Gong, H. Feng, R. Zhan and A. Wang, "A study of parasitic effects of ESD protection on RF ICs," *IEEE Trans. Mircrowave Theory and Techniques*, Vol. 50, No. 1, pp. 393-402, Jan. 2002.

[48] C. Enz and Y. Cheng, "MOS transistor modeling for RF IC design," *IEEE J. Solid-State Circuits*, Vol. 35, pp. 186–201, Feb. 2000.

[49] C. Ito, K. Banerjee and R. Dutton, "Analysis and design of distributed ESD protection circuits for high-speed mixed-signal and RF ICs," *IEEE Trans. Electron Devices*, Vol. 49, No. 8, pp. 1444-1454, Aug. 2002.

[50] W. Soldner, M. Streibl, U. Hodel, M. Tiebout, H. Gossner, D. Schmitt-Landsiedel, J. Chun, C. Ito and R. W. Dutton, "RF ESD Protection Strategies: Codesign vs. Low-C Protection," in *Proc. EOS/ESD Symposium*, Sept. 2005.

[51] S. Hyvonen, S. Joshi, and E. Rosenbaum, "Cancellation technique to provide ESD protection for multi-GHz RF inputs," *Electronics Lett*., Vol. 39, No. 3, pp. 284–286, 2003.

[52] S. Hyvonen, S. Joshi, and E. Rosenbaum, "Comprehensive ESD protection for RF inputs," in *Proc. EOS/ESD Symposium*, pp. 188–194, Sept. 2003

[53] B. Kleveland, T. J. Maloney, I. Morgan, L. Madden, T. H. Lee, and S. S. Wong, "Distributed ESD protection for high-speed integrated circuits," *IEEE Electron Device Lett*., Vol. 21, pp. 390–392, Aug. 2000.

[54] S. Galal and B. Razavi, "Broadband ESD protection circuits in CMOS technology," *IEEE J. Solid-State Circuits*, Vol. 38, No. 12, pp. 2334–2340, Dec. 2003.

[55] S. Galal and B. Razavi, "40-Gb/s amplifier and ESD protection circuit in 0.18- μm CMOS technology," *IEEE J. Solid-State Circuits*, Vol. 39, No. 12, pp. 2389-2396, Dec. 2004.

[56] D. Leenaerts and R. Velghe, "Bond pad and ESD protection structure for 0.25 um/ 0.18 μm RF-CMOS," *Proc. ESSCIRC*, pp. 569–572, Sept., 2003.

[57] B. Razavi, *RF Microelectronics*: Prentice Hall, A Simon & Schuster, 1998.

[58] A. Wang, H. G. Feng, R. Y. Zhan, G. Chen, and Q. Wu, "ESD protection design for RF integrated circuits: new challenges," in *Proc. IEEE Custom Integrated Circuits Conference*, pp. 411–418, 2002.

[59] G. Gramegna, M. Paparo, P. G. Erratico, and P. D. Vita, "A sub-1-dB NF 2.3-kV ESD-protected 900-MHz CMOS LNA," *IEEE J. Solid-State Circuits*, Vol. 36, No. 7, pp. 1010–1017, July 2001.

[60] D. Linten, S. Thijs, M. I. Natarajan, P. Wambacq, W. Jeamsaksiri, J. Ramos, A. Mercha, S. Jenei, S. Donnay, and S. Decoutere, "A 5-GHz fully integrated ESD-protected low-noise-amplifier in 90-nm RF CMOS," *IEEE J. Solid-State Circuits*, Vol. 40, No. 7, pp. 1434–1442, 2005.

[61] V. Chandrasekhar and K. Mayaram, "Analysis of CMOS RF LNAs with ESD protection," in *Proc. Int. Sym. Circuits and Systems*, pp. 799–802, May 2002.

[62] P. Leroux and M. Steyaert, "High-performance 5.2 GHz with on-chip inductor to provide ESD protection," *Electron. Lett*., Vol. 37, No. 7, pp. 467–469, 2001.

[63] S. Joshi and E. Rosenbaum, "ESD protection for broadband ICs (DC-20 GHz and beyond)," *Electronics Lett*. Vol. 39, No. 12, June 2003.

[64] G. Gonzalez, *Microwave transistor amplifiers: analysis and design,* 2nd edition: Prentice Hall, NJ, 1997.

[65] Thomas H. Lee, *The design of CMOS radio-frequency integrated circuits*: Cambridge University Press, 1998.

[66] H. Feng, K. Gong, and A. Wang, "A comparison study of ESD protection for RFIC's performance vs. parasitics," in *Proc. Radio Frequency Integrated Circuits Symp*., pp. 235–238, 2000.

[67] C. Richier, P. Salome, G. Mabboux, I. Zana, a. Juge, and P. Mortini,"Investigation on different ESD protection strategies devoted to 3.3 V RF applications (2 GHz) in a 0.18 CMOS process," in *Proc. EOS/ESD Symposium*, pp. 251-259, 2000.

[68] M. Tiebout, "LNA design for a fully integrated CMOS single chip UMTS Tranceiver," *Proc. ESSCIRC*, pp. 835, 2002.

[69] M.-D. Ker, "Complementary-SCR ESD protection circuit with interdigitated finger-type layout for input pads of submicron CMOS IC's," *IEEE Trans. Electron Devices*, Vol. 42, No. 7, July 1995.

[70] B. Keppnes, "ESD protection solutions for high voltage technologies," in *Proc. EOS/ESD Sym*, pp. 289–298, 2004.

[71] I. E. Opris, "Bootstrapped pad protection structures," *IEEE J. Solid-State Circuits*, pp. 301-302, Feb. 1998.

[72] H. Pan, M. Segami, M. Choi, C. Ling and A. A. Abidi, "A 3.3-V 12-b 50-MS/s A/D converter in 0.6-μm CMOS with over 80-dB SFDR," *IEEE J. Solid-State-Circuits*, pp. 1769-1780, Dec. 2000.

[73] P. Wambacq and W. M. Sansen, *Distortion Analysis of Analog Integrated Circuits*: Kluwer Academic Publishers, 1998.

[74] Texas Instruments, "12-Bit, 65MSPS, 3.3V Analog-to-Digital Converter (ADS5221)", 2004.

[75] Texas Instruments, "14-Bit, 125MSPS Analog-to-Digital Converter (ADS5500)", 2003.

[76] R. Reeder, "Frequency Domain Response of Switched-Capacitor ADCs," Application Note AN-742, Analog Devices, 2004.

[77] R. A. Ashton and Y. Smooha, "Characterization of a 0.16μm CMOS Technology using SEMATECH ESD Benchmarking Structures," in *Proceedings of 23rd EOS/ESD Symposium*, pp. 435-444, 2001.

[78] G. Boselli, C. Duvvury, and V. Reddy, "Efficient PNP Characteristics of PMOS Transistors in Sub-0.13μm ESD Protection Circuits," in *Proceedings of 24th EOS/ESD Symposium*, 2002

[79] C.-H. Choi, "Modeling of Nanoscale MOSFETs," *Ph.D. dissertation*, Stanford University, 2002.

[80] C. Duvvury and G. Boselli, "ESD and latch-up reliability for nanometer CMOS technologies," in *Tech. Dig. IEDM*, pp. 933–936, 2004.

[81] V. Gupta, A. Amerasekera, S. Ramaswamy, and A. Tsao, "ESD-related process effects in mixed-voltage sub-0.5um technologies," in *Proceedings of 20th EOS/ESD Symposium*, pp. 161–169, 1998.

[82] L.-M. Ting, C. Duvvury, O. Trevino, J. Schichl. And T. Diep, "Integration of TLP analysis for ESD Troubleshooting", in *Proceedings of 23rd EOS/ESD Symposium*, pp. 445-452, 2001.

[83] J.-H. Chun, C. Duvvury , G. Boselli, H. Kunz, and R.W. Dutton, "A PMOS Failure Caused by Localized Charge Injection," In *Proceedings of International Reliability Physics Symposium*, 2004.

[84] K. Oh, J. Chun, K. Banerjee, C. Duvvury, and R. W. Dutton, "Modeling of Temperature Dependent Contact Resistance for Analysis of ESD Reliability," In Proceedings *of International Reliability Physics Symposium*, pp. 226-234, 2003.

[85] A. Ameraskera, V. Gupta, K. Vasanth, and S. Ramaswamy, "Analysis of snapback behavior on the ESD capability of sub-0.20 μm NMOS", In *Proceedings of International Reliability Physics Symposium*, pp. 159-166, 1999.

[86] S. Ramaswamy, A. Amerasekera, and M. Chang "A unified substrate current model for weak and strong impact ionization in sub-0.25micron NMOS devices", in *Tech. Dig. IEDM*, pp. 885-888, 1997.

[87] X. Zhang, "Modeling and characterization of substrate resistance for deep submicron ESD protection devices," *Ph.D. dissertation*, Stanford University, 2002.

[88] International Technology Roadmap for Semiconductors (ITRS). [Online]. Available http://public.itrs.net

[89] S. Borkar, "Low power design challenges for the decade," in *Design Automation Conference*, pp. 293–296, Feb. 2001.

[90] C. C. Liu, J. Zhang, A. K. Dutta, and S. Tiwari, "Heating effects of clock drivers in bulk, SOI, and 3-D CMOS," *IEEE Electron Device Lett.*, Vol. 23, p. 716, 2002.

[91] J.-M. Koo, S. Im, L. Jiang, T. W. Kenny, J. G. Santiago, and K. E. Goodson, "VLSI hotspot cooling using two-phase microchannel convection," in *IMECE Tech. Dig.*, Nov. 2002.

[92] D. B. Tuckerman and R. F. W. Pease, "High-performance heat sinking for VLSI," *IEEE Electron Device Lett.*, Vol. 2, No. 5, pp. 126–129, 1981.

[93] R. Mahajan, R. Nair, V. Wakharkar, J. Swan, J. Tang, and G. Vandentop, "Emerging directions for packaging technologies," *Intel Technology J.*, Vol. 6, p. 62, 2002.

[94] K. Rim, J. L. Hoyt, and J. F. Gibbons, "Fabrication and analysis of deep submicron strained-Si N-MOSFETs," *IEEE Trans. Electron Dev.*, Vol. 47, pp.1406-1415, 2000.

[95] H.-S. P. Wong, K. K. Chan, and Y. Taur, "Self-aligned (top and bottom) gate double-gate MOSFET with 25 nm thick silicon channel," in *Tech. Dig. IEDM*, pp. 427 – 430, 1997.

[96] Y.-K. Choi, T.-J. King, *and C*. Hu, "Nanoscale CMOS spacer FinFET for the terabit era," *IEEE Electron Device Lett*., Vol. 23, No. 1, pp. 25–27, Jan 2002.

[97] M. Asheghi et al., "Thermal Conductivity Model for Thin Silicon-on-Insulator Layers at High Temperatures," *Int. SOI Conf.,* pp. 51-52, 2002.

[98] O. Tornblad et al., "Modeling and Simulation of Phonon Boundary Scattering in PDE-based Device Simulators," in *Proceedings of SISPAD*, pp. 58-61, 2000.

[99] E. Pop, C. O. Chui, S. Sinha, R. W. Dutton, and K. Goodson, "Electro-thermal comparison and performance optimization of thin-body SOI and GOI MOSFETs," in *Technical Digest IEDM*, 2004.

[100] S.-D. Kim, C.-M. Park, and J. C. S. Woo, "Advanced Model and Analysis of Series Resistance for CMOS Scaling into Nanometer Regime," *IEEE Trans*. *Electron Device*, Vol. 49, No. 3, pp. 467-472, 2002.

[101] K.-H. Oh, J.-H. Chun, K. Banerjee, C. Duvvury and R. W. Dutton, "Modeling of Temperature Dependent Contact Resistance for Analysis of ESD Reliability," in *Proc. Int. Reliability Physics Sym*, pp. 226-234, 2003.

[102] K. Matsuzawa, K. Uchida and A. Nishiyama, "A unified simulation of Schottky and ohmic contacts," *IEEE Trans. Electron Dev*, pp. 103-108, 2000.

[103] J.-H. Chun, Y. Liu, C. Duvvury, and R. W. Dutton, "Implementation of temperature dependent contact resistance model for the analysis of deep submicron devices under ESD," in *Tech. Dig*. *IEDM*, 2003.

[104] D. B. Scott, R. A. Chapman, C-C. Wei, S. S. Mahant-shetti, R. A. Haken, and T. C. Holloway, "Titanium disilicide contact resistivity and its impact on 1-μm CMOS circuit performance," *IEEE Trans. Electron Dev*., pp. 562-573, 1987.

[105] K.-H. Oh, K. Banerjee, C. Duvvury and R. W. Dutton*,* "Impact of Gate-to-contact Spacing on ESD Performance of Salicided Deep Submicron NMOS Transistors," *IEEE Trans. Electron Dev*., pp. 2183-2192, 2002.

[106] K. Esmark, *Device simulation of ESD protection elements*, Series of Microelectronics, Vol. 128, Hartung-Gorr Verlag Konstanz, 2002.

[107] K. Esmark, W. Stadler, H. Gossner, M. Wendel, X. Guggenmos, W. Fichtner, "Advanced 2D/3D ESD Device Simulation," in *Proc. EOS/ESD Symp*., pp. 420–429, 2000.

[108] S. M. Sze, *Physics of Semiconductor Devices,* 2nd edition: John Wiley & Sons, New York, 1981.

[109] R. Stratton, "Semiconductor current-flow equations (diffusion and degeneracy)," *IEEE Trans. Electron Dev*., ED-19, pp. 1288-1300, 1972.

[110] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*: Wien-New York: Springer-Verlag, 1984.

[111] D. M. Caughey and R. E. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proc. IEEE*, 55, pp. 2192–2193, 1967.

[112] N. D. Arora, J. R. Hauser, and D. J. Roulston, "Electron and hole mobilities in Silicon as a function concentration and temperature," *IEEE Trans. Electron Devices*, Vol. ED-29, pp. 292–295, Feb. 1982.

[113] C. Lombardi, S. Manzini, A. Saporito, and M. Vanzi, "A physically based mobility model for numerical simulation of nonplanar devices," *IEEE Trans. Computer-Aided Design*, Vol. 7, No. 11, pp. 1164–1170, Nov. 1988.

[114] T. H. Geballe and G. W. Hull, "Seeback effect in silicon," *Phys. Rev*., 98, pp. 940–947, 1955.

[115] T. Oh, "AC and Noise Analysis of Deep-submicron MOSFETs," *Ph.D. dissertation*, Stanford University, February, 2004.

[116] M. Lundstrom, *Fundamentals of carrier transport,* 2nd edition: Cambridge University Press, 2000.

[117] O. Tornblad, U. Lindefelt, and B. Breitholtz, "Heat generation in Si bipolar power devices: the relative importance of various contributions," Solid State Electronics, Vol. 39, No. 10, pp.1463–1471, 1996.

[118] J. Lai and A. Majumdar, "Concurrent Thermal and Electrical Modeling of Sub-micrometer Silicon Devices," *J. Appl. Phys.*, 79 (9), pp.7353-7361, 1996.

[119] A. Amerasekera, M.-C. Chang, J. Seitchik, A. Chatterjee, K. Mayaram and J.-H. Chun, "Self-heating effects in basic semiconductor structures," *IEEE Trans. Electron Dev*., ED-40, pp. 1836 – 1844, 1993.

[120] R. M. Costescu, M. A. Wall, and G. G. Cahill, "Thermal conductance of epitaxial interfaces," *Phys. Rev. B*, vol. 67, p. 54302, 2003.

[121] K.-H. Oh, C. Duvvury, K. Banerjee, and R. W. Dutton, "Analysis of non-uniform ESD current distribution in deep submicron NMOS transistors," *IEEE Trans. Electron Devices*, Vol. 49, No. 10, 2002

[122] E. Pop, K. Banerjee, P. G. Sverdrup, R. W. Dutton, and K. E. Goodson, "Localized heating effects and scaling of sub-0.18 micron CMOS devices," in *Tech. Dig. IEDM*, pp. 677–680, 2001.

[123] G. Chen, "Nonlocal and nonequilibrium heat conduction in the vicinity of nanoparticles," *J. Heat Transfer*, Vol. 118, pp. 539–545, 1996.

[124] S. Sinha and K. E. Goodson, “Phonon heat conduction from nanoscale hotspots in semiconductors,” in *Proc. Int. Heat Transfer Con*., pp. 573–578, 2002

[125] P. G. Sverdrup, Y. S. Ju, and K. E. Goodson, “Sub-continuum simulations of heat conduction in silicon-on-insulator transistors,” *J. Heat Transfer*, Vol. 123, pp. 130-137, 2001.

[126] S. Mazumder and A. Majumdar, “Monte Carlo study of phonon transport in solid thin films including dispersion and polarization,” *J. Heat Transfer*, Vol. 123, pp. 749–759, 2001.

[127] E. Pop, R. W. Dutton, and K. E. Goodson, “Analytic band Monte Carlo model for electron transport in Si including acoustic and optical dispersion,” *J. Appl. Phys*., Vol. 96, No. 7, Oct. 2004.

[128] S. T. Huxtable, A. Abramson, A. Majumdar, C.-L. Tien, G. Zeng, C. LaBounty, A. Shakouri, J. E. Bowers, and E. Croke, "Thermal conductivity of Si/SiGe and SiGe/SiGe superlattices," *Applied Physics Letters* , Vol. 80, pp. 1737-1739, 2002.

[129] J.-S. Goo, Q. Xiang, Y. Takamura, F. Arasnia, E. N. Paton, P. Besser, J. Pan, and M.-R. Lin, “Band offset induced threshold variation in strained-Si nMOSFETs,” *IEEE Electron Device Lett.*, Vol. 24, pp. 568-570, Sept. 2003.

[130] R. W. Dutton and C. Choi, “Technology limits and compact model for SiGe scaled FETs,” *Nano Tech. Conf.,* 2004.

[131] Y. Ohno, “Electron viscosity effects on electron drift velocity in silicon MOS inversion layers,” *IEEE Trans. Electron Devices*, Vol. 38, pp. 1889-1894, Aug. 1991.

[132] C. Jungemann, S. Keith, M. Bartels, and B. Meinerzhagen, “Efficient full-band Monte-Carlo simulation of silicon devices,” *Proc. IEICE*, Vol. E82-C, No. 6, pp. 870-879, 1999.

[133] S. Voldman, P. Juliano, N. Schmidt, R. Johnson, L. Lanzerotti, A. Joseph, C. Brennan, J. Dunn, E. Rosenbaum, and B. Meyerson, “Electrostatic discharge characterization of epitaxial-base silicon germanium heterojunction bipolar transistors,” in *Proc. EOS/ESD Symp.*, pp. 239-250, 2000.

[134] C.-H Choi, Z. Yu, and R. W. Dutton, “Elimination of nonsimultaneous triggering effects in finger-type ESD protection transistors using heterojunction buried layer,” in *Proc. SISPAD*, pp. 304-304, 1998.

[135] E. Pop and K. E. Goodson, "Thermal Phenomena in Nanoscale Transistors," *ITHERM 2004*, pp. 1-7, June 2004.

[136] M. Asheghi, B. Behkam, K. Yazdani, R. Joshi, K. E. Goodson, “Thermal Conductivity Model for Thin Silicon-on-Insulator Layers at High Temperatures,” *Int. SOI Conf.,* pp. 51-52, 2002.

[137] *PROPHET User Reference Guide*, MTAssociates, 2003.

[138] D. A. Antoniadis, I. J. Djomehri, K. M. Jackson, and S. Miller, "MIT Well Tempered Bulk-Si NMOSFET Device," MIT, MS, 1999, http://www-mtl.mit.edu/Well

[139] D. Esseni, M. Mastrapasqua, G. K. Celler, C. Fiegna, L. Selmi, E. Sangiorgi, "Low Field Electron and Hole Mobility of SOI Transistors Fabricated on Ultrathin Silicon Films for Deep Submicrometer Technology Application," *IEEE Trans. on Electron Devices*, Vol. 48, pp. 2842–2850, 2001.

[140] K. Uchida, H. Watanabe, A. Kinoshita, J. Koga, T. Numata, S. Takagi, "Experimental Study on Carrier Transport Mechanism in Ultrathin-Body SOI N- and P-MOSFETs with SOI Thickness Less than 5nm," in *Tech. Dig. IEDM*, pp.47-50, 2002.